Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765849

FACIAL EMOTION RECOGNITION SYSTEM

[1] Dr. F. Antony Xavier Bronson, [1]Dr. P. Dinesh Kumar, [1]Dr. G. Victo Sudha George [2]Mr. Sarath Rahul. S, [2]Mr. Siva Kumar. R, [2]Mr. Siva Sankar. V
[1]Professor, Department of CSE, Dr.M.G.R. Educational and Research Institute, Chennai
[2]Final Year B.Tech-CSE,Dr.M.G.R. Educational and Research Institute,Chennai.
antony.cse@drmgrdu.ac.in, dineshkumar.it@drmgrdu.ac.in ,
victosudhageorge@drmgrdu.ac.in, sarathrhl65@gmail.com , svsasi5@gmail.com ,
harisiva22062002@gmail.com .

Abstract -Facial expressions for emotion clarification have constantly been a smooth mission for humans, however fixing the same undertaking with computer algorithms is quite tough. Recent advances in computer vision and machine learning have made it possible to recognize emotions in images. In this paper, A new technology called facial emotion recognition using convolutional neural networks. It is based on a two-part Convolutional Neural Network (CNN). The first part removes the background from the image and the alternate part focuses on rooting thefacial pointvectors. In this model, expression vectors (EVs) are used to find 5 types of face regular expressions. Control data were obtained from a stored database of 10,000 images (154 subjects). Using an EV of length 24, able to accurately emphasize emotion with 96% accuracy. The two- position CNN runs successionally, and the last subcaste of the perceptron adjusts the weights and exponent values at each replication. This model differs from the commonly used single-level CNN strategy, which results in better accuracy. Additionally, the new background subtraction procedure applied prior to EV generation avoids many of the problems that can arise (e.g., distance from the camera). This Model was extensively tested with more than 750K images using extended Cohn–Kanade expression, Caltech faces, CMU and NIST datasets. This model emotion detection to be useful in many applications such as predictive learning of students, lie detectors, etc.

Keyword: CNN, Haar-cascades, Machine Learning, Open-CV, Python.

I. Introduction

Facial expressions play an important role in emotion recognition and are used toidentify people with non-verbal communication processes. They are second only to tone of voice and are very important in everyday emotional communication. They are also emotional indicators that allow a man to express his emotional state. People can instantly recognize a person's emotional state. As a result, information about facial expressions is often used in automatic emotion recognition systems. The purpose of the research presented in this article is to recognize seven basic emotional states based on facial expressions: neutral, joy, surprise, anger, sadness, fear, and disgust. It is a technology used for analyzing sentiments by different sources, such as pictures and videos.

In recent years, the widespread adoption of cameras and technological advances in bioanalysis, machine learning, and pattern recognition have played a major role in the development of FER technology. From tech giants like NEC or Google to smaller ones like Affectiva or Eyeris, many companies are investing in this technology, a testament to its growing importance. There are also several initiatives under the Horizon20201 Research and Innovation Program exploring the use of this technology. FER analysis consists of three steps: a) face detection, b) facial expression detection, and c) expression classification for emotional states. Emotion detection is based on analyzing the location of landmarks on the face (e.g., nose tip, eyebrows).The video also analyzes changes in these positions to reveal the contraction of facial muscle groups (Ko 2018). Facial expressions can be algorithmically divided into basic emotions (e.g., anger, disgust, fear, joy, sadness, surprise) and complex emotions (e.g., happy-sad, happy-surprise, glad-disgust, sad). - scary, sad- angry, sadly surprised).) (Du et al. 2014). In other cases, facial expressions may be related to physiological or mental conditions (such as fatigue or boredom). Images or video sources used as input to the FER algorithm can range from surveillance cameras to cameras placed next to advertising screens in stores, social media and streaming services, or even personal devices. FER can also be combined with biometrics. Accuracy can be improved by using techniques that analyze different types of sources, such as speech, text, health data from sensors, or blood flow patterns derived from images. The potential uses of FERs cover a wide range of applications, examples of which are listed below in groups by application.The human face is the most exposed part of the body and computer vision systems (usually cameras) can be used to analyze facial images for emotion recognition. Changes in lighting and head position are major factors affecting the quality of a camera-based emotion recognition system. Methods based on 2D image analysis are particularly sensitive to these factors. Methods of implementing 3D facial models are much more promising. Microsoft Kinect was used for 3D face modeling in the experiment mainly because of its low cost and ease of operation. The Kinect has a lower scan resolution, but a relatively high image alignment speed (30 fps). It has an infrared emitter and two cameras. One of the cameras records visible light and the other works with infrared light and is used to measure depth. Infrared light reflected from the user's body creates a 3D model of the face.

II. Related Works

One of the most used testbeds is the CK database. Active Appearance Models (AAMs) and a linear support vector machine (SVM) classifier are used to leave one out subject cross validation for both AU and emotion detection for the posed data. Emotes and AU tags, enhanced image data, and tracked landmarks will be available July 2010[1]. OpenCV is used to coding facial recognition is now easier. Gather the face, the images of the face of the persons that want to be identified. Feed the face data and respective names of each to the recognizer, it can train the recognizer. Feed the new face data of a people and check the face recognizer recognizes them [3]. Understanding emotions using text, voice, and verbal expression data. Facial images also provide a constructive opportunity to interpret and analyze a person's emotional problems. The emotion classification can be focused on artificial neural networks and Support Vector Machines [4]. The real-time face tracker output can be defined and extracted with two new types of features that are eccentricity and linear features.This feature can be used to train the machine learning classifier [5]. Machine learning algorithm are used for recognition and classification of different classes of face emotions by training of different set of images. Implementation of herein algorithm would give a several regions of identification, psychological researches and many real-worldproblems [6]. CNNConvolutional neural networks are developed for the facial recognitionof emotion expressionsand it is classified into seven basic categories, CNN calculates features by learning automatically [7]. Combines AU (animation units) and FPP (feature point location) features tracked by Kinect to capture deformation of the 3D mesh during facial expressions. For emotion recognition using these real-time facial expression features, we propose a convergence algorithm based on IEP (Improved Emotional Profiles) and Maximum Confidence [9].

III. Requirement Analysis

Detecting human emotions is implemented in many areas where additional protection or information about a person is required. This can be seen as the second step towards face recognition. You may need to set up a second layer of security where emotions are also detected along with faces. This can be useful to ensure that the person looking at the camera is not just a 2D image. Another important area that confirms the importance of emotion detection is business promotion. Most businesses thrive on customer response to all of their products and services. If AI systems can capture and identify emotions in real time based on a user's image or video, they can determine whether a customer liked a product or offer.Identified that security is the main reason for identifying a person. This can be based on fingerprint matching, voice recognition, passwords, retina detection, and more. Revealing a person's intentions can also be important in preventing threats. This can be useful in vulnerable locations such as airports, concerts, and large public gatherings, where many security breaches have occurred in recent years. Human emotions can be divided into fear, contempt, disgust, anger, surprise, sadness, joy, and neutrality. These feelings are very analytical. Distortion of the facial muscles is very small and detecting these differences can be very difficult. This is because even a small difference in facial expression changes. Also, since emotions are highly context-dependent, the same person or different people may have different expressions of the same emotion. While focus only on the parts of the face where emotion is most evident, such as the corners of the mouth and eyes, how to extract and classify these gestures remains an important problem. Neural networks and machine learning have been used for these tasks, with good results. Machine learning algorithms have proven very useful for pattern recognition and classification. The most important aspect of a machine learning algorithm is its features. This article examines how features are extracted and modified for algorithms such as support vector machines. Compare algorithms and feature extraction methods from different papers. The human emotion dataset can be a great example to study the reliability and performance of classification algorithms and how they perform on different types of datasets. Typically, face recognition algorithms are applied to images or captured frames prior to feature extraction to detect emotions. The emotion detection steps can be summarized as follows.

1) Dataset preprocessing
2) Face detection
3) Feature extraction
4) Classification based on the features

IV. Methodology

Convolutional Neural Networks (CNNs) are the most common way to analyze images. CNNs differ from multilayer perceptrons (MLPs) in that they have hidden layers called convolutional

layers. The proposed method is based on a two-level CNN structure. The first recommended level is Background Removal, which is used to extract emotion from images. Here, the primary expression vector (EV) is extracted using the conventional CNN network module. Expression vectors (EVs) are created by tracking important points on the face. EVs are directly related to changes in expression. EVs are obtained using a perceptron base block applied to face images with background removed. In the proposed FERC model, there is also a non-convolutional perceptron layer as a final step. Each convolutional layer receives an input (or image), transforms it, and outputs it to the next layer. This transformation is a convolutional operation. All convolutional layers used can detect patterns. Four filters were used in each convolutional layer. The input images fed into the first part of the CNN (which is used to remove the background) usually consist of faces along with shapes, edges, textures and objects. At the beginning of convolutional
layer 1, edge detector, circle detector and corner detector filters are used. When a face is detected, the second part, the CNN filter, captures facial features such as eyes, ears, lips, nose, and cheeks. The edge detection filter used by this layer. The second part of the CNN is for example [0.25, 0.17, 0.9; 0.89, 0.36, 0.63; 0.7, 0.24, 0.82]. This number is initially chosen between 0 and 1. These numbers are optimized for EV detection based on good information obtained from the supervised training dataset. Here, optimized the filter values using minimal error decoding. When the filter is tuned by supervised learning, it is applied to the face with the background removed (i.e., the output image of the first part of the CNN) to detect other parts of the face (i.e., eyes, lips, nose, ears).)., etc.)

A. Image Acquisition:Performing image acquisition in image processing is always the first step in the workflow sequence because, without an image, no processing is possible. After the image has been obtained, various methods of processing can be applied to the image to perform the many different vision tasks. There are various ways to acquire image such as with the help of a camera or scanner. Acquired image should retain all the features.
B. Preliminary processing: The main purpose of preprocessing is to improve the appearance of images and improve handling of datasets. Image preprocessing, also known as image restoration, is the act of correcting distortions, distortions, and noise that occur during image processing. Interpolation is a technique often used for operations such as scaling, rotation, compression, and geometric correction. Denoising is an important step in processing. However, noise effects segmentation and pattern matching.

C. Binarization: The resulting image is RGB color. Instead of processing the three components R (red), G (green), and B (blue), it converts to grayscale because it only carries intensity information that is easy to process. Takes the RGB value of each pixel and outputs a single value representing the brightness of that pixel.

D. Edge detection: Edge detection is the name of a set of mathematical methods for identifying points in a digital image where the brightness of the image changes abruptly or, more formally, has such continuity. Points where the brightness of an image changes dramatically are usually made up of a series of curved segments called edges. Edge detection is an image processing technique that finds the edges of objects in an image. It works by detecting breaks in brightness.

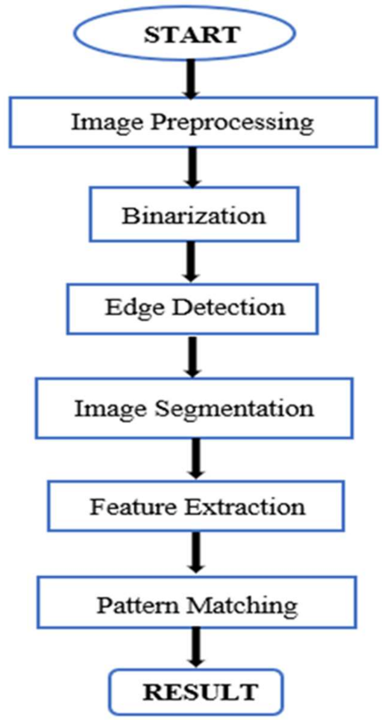


Fig.1: Design Flow

E. Image Segmentating: It is the process of dividing a digital image into parts. The purpose of segmentation is to simplify and/or change the image representation to something more meaningful and easily parseable. Image segmentation is commonly used to detect objects and boundaries (lines, curves, etc.) in an image. Image segmentation algorithms are usually based on one of two main properties of image intensity values: 1) Tearing: based on sharp intensity changes, such as edges in an image. 2) Similarity: Based on dividing the image into similar regions according to a set of predefined criteria.

F. Feature Extracting: It is a special form of size reduction. If the inputs to the algorithm are too large to handle and are expected to be highly redundant, the inputs are transformed into a set of reduced expression functions. Converting input data into a set of features is called feature extraction. If the extracted features are chosen carefully, the feature set is expected to extract relevant information from the input to perform the desired task using this reduced representation instead of the full-size input.

V. IMPELEMENTATION

This work uses a static approach using feature extraction and emotion recognition using machine learning. The focus is on feature extraction using Python and image processing libraries and using machine learning algorithms for prediction. The implementation is divided into three parts. The first part is image preprocessing and face detection. Face recognition uses built-in methods available in the dlib library. When a face is detected, it extracts regions of interest and important facial features. There are various features that can be used to detect emotions. This task focuses on points on the face, such as the eyes, mouth, and eyebrows.

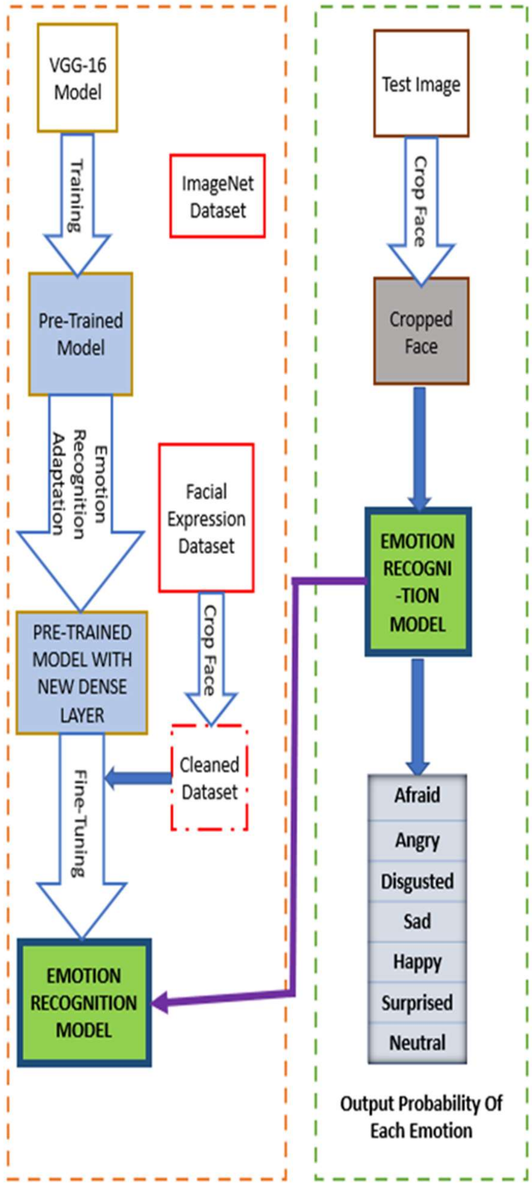


Fig. 2: Facial Emotion recognition system architecture diagram

I have a problem of classifying multiple classes rather than multiple labels. A feature set can belong to multiple labels, but there are some differences because it belongs to only one distinct class. The extracted facial features are used with SVM todetect multi-class emotions. SVM is one of the widely used and generally accepted sentiment classification algorithms. Our database has a total of 7 classes to classify. To compare the results of different algorithms,compared the results with logistic regression and random forest.

A. Setting up the database

The image files for the CK+ database is in different directories and sub-directories based on the person and session number. Also, not all the images depict emotion; only 327 files have one of the emotions depicted from 1-7. All the files were of type portable networks graphic file(.png). The emotion labels are in the different directory but with the same name as image files. A small utility function in java which used the emotion file name to pick up the correct image from the directory and copy it in our final dataset folder. Appended the name of the emotion file to the image file name. Thus, while parsing the file in our program will have the emotion label for that file.

B. Image Processing Pipeline

Face detection will be the first and most important part of the processing pipeline. Even though the image contains only frontal expression data of the face, the face had to be defined before further processing. Once a face is detected, it becomes easier to identify regions of interest and extract features.

Fig.3: Face Detection

For face detection, many algorithms like Haar-cascades from OpenCV. Finally settled for face detector based on histogram of oriented gradients from Dlib library. HoG descriptors along with SVM are used to identify the face from the image. Images are converted to grayscale and resized for uniformity. OpenCV's Haar cascades for face recognition. As a result, a face detector based on the gradient direction histogram from the Dlib library. In Fig. 3 shows the Face detection in image preprocessing.

C. Algorithm

Haar cascade pseudo code

Choose f (highest acceptable false positive rate per layer) and d (Lowest acceptable detection rate per layer)

Let the Ftarget is target overall false positive rate

Let the P is a set of positive examples

Let the N is a set of negative examples

Let the Fo = 1, Do=1, and i=0 (Fo: overall false positive rate at layer 0, Do: adequate detection rate at layer 0, and i: is the current layer)

While Fi>Ftarget (Fi: total false positive rate at layer i):

it+ (layer escalating by 1)

ni=0; Fi = Fi-1 (ni: negative example i):

While Fi > f*Fi-1:

ni++ (check a next negative example)

Use P and N to train with AdaBoost to make a xml (classifier)

Check the outcome of new classifier for Fi and Do

Decrease threshold for new classifier to adjust detection rater >= d*Fi-1

N= empty

If Fi>Ftarget, use the current classifier and false detection to set N

D. Landmarks

Face landmarks are very important and can be used for face detection and recognition. The same guidelines can be used for expressions. The Dlib library has 68 facial landmark detection functions that determine the location of 68 facial landmarks. Use the dlib library to extract the (x, y) coordinates of each face point. These 68 points can be divided into specific areas such as the left eye, right eye, left eyebrow, right eyebrow, mouth, nose, and chin. In Fig. 4 shows the image of landmarks on the face.

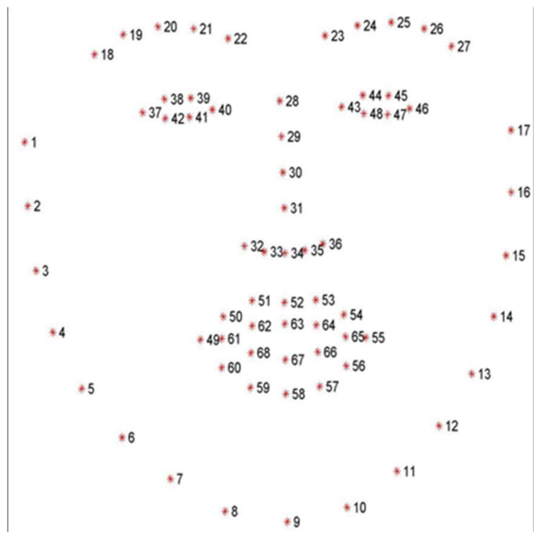


Fig.4: Face Landmarks

HoG descriptors along with SVMs are used to identify faces in images. Images are converted to grayscale and scaled for uniformity.

E. Facial feature extraction

To extract facial features, dlib's 68 landmark facial feature predictor was used. The face detection algorithm returns a window (x, y, width, height) that is the face detected. Detected faces are passed to feature prediction algorithms. The predict function returns 68 points for the eyes (left and right), mouth, eyebrows (left and right), nose, and chin. Used a numpy array to transform the 68 points into an array of 68 x and y coordinates representing their locations. These are the facial features used to predict emotions.

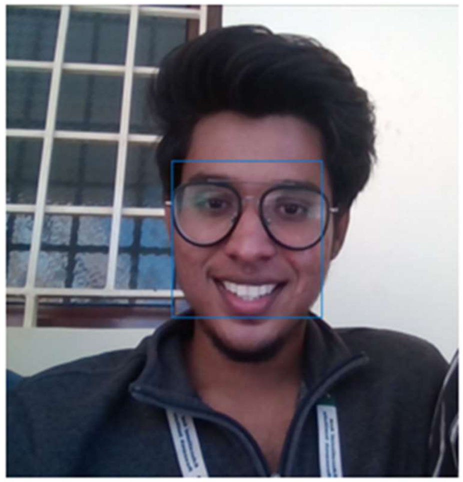
Easier to get landmarks in numpy array format. Also, the index of each feature in, specific features rather than the entire set. The 28 feature points are divided into 1 to 17 points for the chin and 49 to 68 points for the mouth. For example, if you want to ignore the jaws, you can convert the function to a numpy array, giving the x and y coordinates of the jaws as 0. Calculated the distance and area of polygons for some facial landmarks.

Fig. 5: Facial landmarks

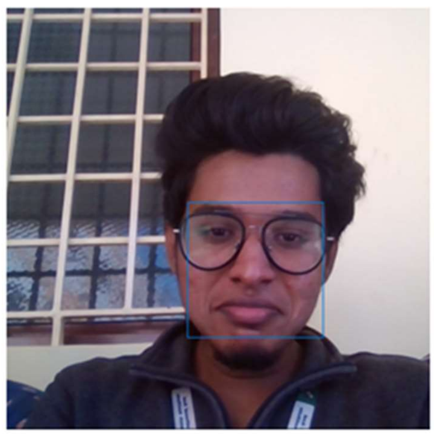
Python Pipeline A dataset of 327 files was stored in a directory and each file was processed to generate a set of features. Upon receiving the file, I parsed the file name to extract sentiment labels. Sentiment labels have been added to the list of labels to configure multiclass target variables. Images were processed for face recognition and feature prediction. Functions derived from each file were added to a list and later converted to 327 * 68 * 2 numpy arrays. It also had a target class in the form of a numpy array. The same process was applied to the RaFD database.

VI. Result

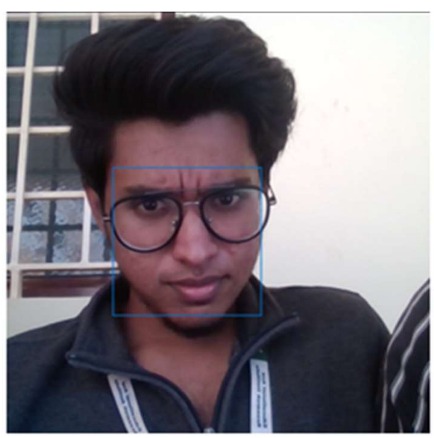
These results were obtained for the MLP classifier and "natural" data segregation (subject independent) across all users. Experiments were performed at the user's fixed location in relation to the Kinect device under identical conditions. In Fig. 6 output images of the emotion are detected and classified as angry, neutral, happy, sad.



HAPPY



SAD



ANGRY

Fig. 6. Sample Emotions

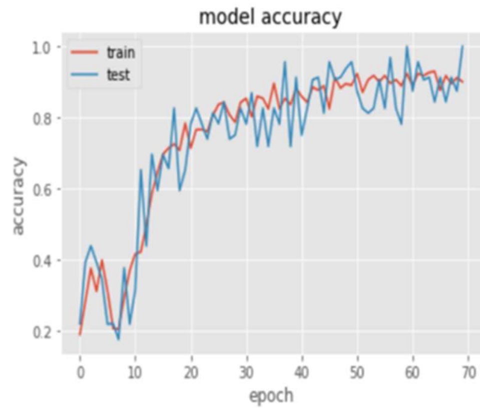


Fig. 7: Accuracy graph

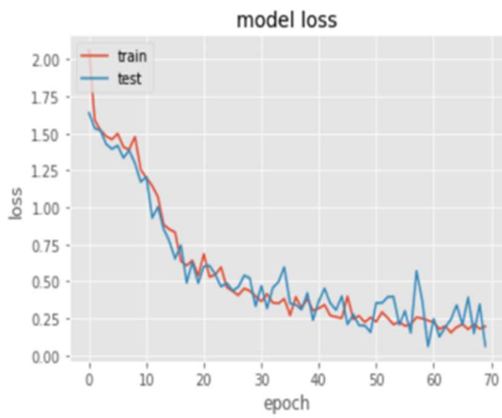


Fig. 8: Accuracy/Loss Graph

Experiments performed on seven emotional states achieved very good classification accuracy of emotion (96% with random data isolation) and satisfactory classification accuracy (73% with "natural" data isolation). In Fig. 7 it shows the accuracy graph shows the accuracy of the model and in Fig. 8 shows the loss of the model.

VII. Conclusion

The classification accuracy was affected by how users reproduced specific facial expressions. In real-world situations, the accuracy of classification can be influenced by many additional factors. When experiencing real emotions, facial expressions can vary greatly and may reveal more or less.

VIII. Reference

[1] Con-Canade Enhanced Dataset (CK+): Complete Expression Dataset for Action Units and Emotion-Determined Expressions. Proceedings of the Third International Workshop on CVPR for the Analysis of Human Communication Behavior (CVPR4HB 2010), San Francisco, USA, 94-101.

[2]"Convolutional Neural Networks for Automatic Face Recognition", International Conference on Applied Systems Innovation (ICASI), 2017, Sapporo, 2017, p. 814-817. 2: 10.1109/ICASI.2017.7988558.
[3] R. Raja, ―Face detection using OpenCV and Python: A beginner ‘s guide, ‖ 2017.
[4] "Advancements and recent trends in emotion recognition using facial image analysis and machine learning models," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017
[5] Recognition novel method for geometrical facial features extraction," 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 2014.
[6] "Reliable Face Recognition Method and Emotion Detection System Using Support Vector Machine", 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, 2016.
[7] Ratliff M. S., Patterson E., Emotion recognition using facial expressions with active appearance models, Proceedings of the Third IASTED International Conference on Human Computer Interaction, ACTA Press, Anaheim, CA, USA, 2008, 138–143.
[8] Tian Y. I., Kanade T., Cohn J. F., Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), no. 2, 97–115.
[9] Mao Q., Pan X., Zhan Y., Shen X., Using Kinect for real-time emotion recognition via facial expressions, Frontiers Inf Technol Electronic Eng, 16 (2015), no. 4, 272–282.
[10] Li B. Y. L., Mian A. S., Liu W., Krishna A., Using Kinect for face recognition under varying poses, expressions, illumination and disguise, 2013 IEEE.
[11] Workshop on Applications of Computer Vision (WACV), 2013, 186–192.
[12] Lundqvist D., Flykt A., Öhman A., The KarolinskaDirected Emotional Faces- KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, no. 1998.
[13] Koelstra S., Patras I., Fusion of facial expressions and EEG for implicit affective tagging, Image and Vision Computing, 31 (2013), no. 2, 164–174.
[14]Ahlberg J., CANDIDE-3 - An Updated Parameterised Face, 2001.
[15] Li B. Y. L., Mian A. S., Liu W., Krishna A., Using Kinect for face recognition under varying poses, expressions, illumination and disguise, 2013 IEEE Workshop on Applications of Computer Vision (WACV), 2013, 186–192.