AUTOMATIC FACIAL EXPRESSION RECOGNITION BASED ON SPATIAL AND TEMPORAL SEQUENCING

[1]Dr. D.B.K. Kamesh, Professor, Dept. of CSE, Malla Reddy Engineering College for Women, Maismmaguda, Hyderabad.

[2]Gadde Srichandrika, [3]Gade Mary Rithika Reddy, [4]Irakam Bhavya Sri

[2,3,4] Student, Malla Reddy Engineering College for Women, Maismmaguda, Hyderabad

Abstract:

Face expression recognition is becoming more popular as artificial intelligence advances. Emotional recognition is critical in interaction technology. Nonverbal elements account for two-thirds of communication in interaction technology, while verbal elements account for one-third. A technique known as facial emotion recognition is used to recognize facial expressions (FER). The way a person expresses their feelings, as well as their mental state or point of view, is greatly influenced by their facial expressions. To uncover the most fundamental human emotions, this study combines gender analysis and age estimation. Joyful, sad, angry, scared, astonished, and neutral facial expressions are examples of basic emotions. Here's an example of a real-time facial expression.

Introduction:

A face detection process involves dividing an image into two classes: one with targets (faces), and the other with clutter (background). There are similarities across faces, but they differ in terms of age, skin tone, and facial expression, making this challenging. Different lighting situations, picture quality, and geometries, as well as the potential for partial occlusion and disguise, add another layer of complexity to the issue. Any face should be detectable by a face detector in any backdrop situation and under any set of illumination circumstances. Two jobs can be separated from face detection analysis. The first task is a classification task, which receives input from a few randomly chosen areas of the image and produces a binary value of yes or no, indicating whether or not there are any faces present. The second task is the face localization task, which outputs the location of any face or faces inside an image as some bounding box or boxes with a bounding radius (x, y, width, height). By applying automatic facial expressions, intelligent robots can be created. These bots can be utilized in a variety of settings, including service desks and interactive games. According to Ekman, there are six universal expressions. They are wrath, grief, surprise, disgust, and happiness. These expressions can be identified by observing variations in the face. By raising the corners of the mouth and tightening the eyelids, we might say, for instance, that a person is happy. Facial expression variations can reveal a person's psychological moods, social communication, and intentions. Automatic facial expression identification is widely used in numerous applications, such as human emotion analysis, natural human-computer interaction,

picture retrieval, and talking robots. Since humans consider facial expressions to be among the most natural and effective ways to communicate their intentions and feelings, face recognition using histogram of oriented gradients using CNN detection has become a significant topic in the technology community. The system's final step is facial expression recognition. The training process for expression recognition systems mainly consists of three steps: feature learning, classifier development, and feature selection. The first step is the feature learning stage, followed by the feature selection stage and the classifier development stage. After the feature learning stage, only learnt changes in facial expressions among all features are extracted. The best characteristics, as determined through feature selection, then serve as a representation of facial expression. They should strive to minimize the intra class variances of phrases in addition to maximizing inter class variety. In addition to maximining interclass variation, expressions should also reduce intraclass variation. Minimizing the intra-class variation of expressions is difficult since the same expressions of different people in an image are distant from one another in pixel space. YOLO, SDD, RCNN, and Faster RCNN are methods for facial detection.

Literature Survey: [Fayyaz Ali et al.,] The FER2013 database, which includes seven types of facial expressions including surprise, fear, anger, neutrality, sadness, disgust, and happiness throughout the previous few decades, is used in this study to offer a novel method for facial expression recognition. Numerous applications have been created for feature extraction and inference in the study of techniques to recognise facial expressions. However, the considerable intraclass diversity makes it still difficult. The accuracy of both handcrafted and leaned aspects, such as HOG, was thoroughly examined. The two models that were put out in this study were the Histogram of Oriented Gradients-based Deep Convolutional Neural Network and the FER using Deep Convolutional Neural Network (FER-CNN) (FER-HOGCNN). The FER-CNN model set's training and testing accuracy was 98%, 72%, and similarly, the losses were 0.02 and 2.02 correspondingly. On the other hand, the FER-HOGCNN model's training and testing accuracy were set at 97% and 70%, respectively. Losses were 0.04 and 2.04, similarly. Findings: It was discovered that while the FER-HOGCNN model's accuracy is good overall, it is not significantly superior to the Simple FER-CNN model. Due to the low image quality and restricted dimensions of the dataset, the HOG loses certain crucial characteristics during training and testing. Application/Improvements: The study aids in the development of the FER System's image processing capabilities. In addition, this work will be expanded in the future in order to combine the LBP and HOG operator with Deep Learning models to extract the key features from images.

[J. R. Barr et al.,] We present a method for determining the social network structure of

individuals appearing in a collection of video clips. Individuals are unknown and cannot be matched to known enrolments. By grouping similar-looking faces from different videos, an identity cluster representing an individual is formed. A node in the social network represents each identity cluster. If the faces from two nodes appear together in one or more video frames, they are linked. Our approach employs a novel active clustering technique to generate more accurate identity clusters based on user feedback about ambiguously matched faces. The final output consists of one or more network structures representing the social group(s) and a list of people who may connect multiple social groups. Our findings show that the proposed clustering algorithm and network analysis techniques work. [Ryu et al.,] LDTP (local directional ternary pattern) is a new face descriptor for facial expression recognition. LDTP uses directional information and ternary patterns to take advantage of the robustness of edge patterns in the edge region while overcoming the weaknesses of edge-based methods in smooth regions to efficiently encode information of emotion-related features (i.e., eyes, brows, upper nose, and mouth).

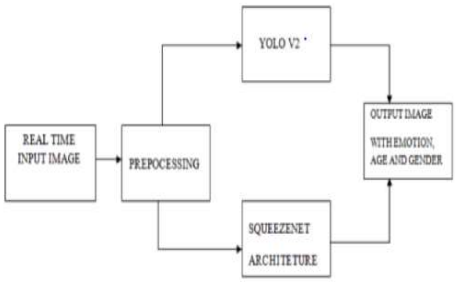
Existing System:

This system proposes a method for recognizing facial expressions based on a novel facial decomposition. First, using facial landmarks detected by the IntraFace algorithm, seven regions of interest (ROI) representing the main components of the face (left eyebrow, right eyebrow, left eye, right eye, between eyebrows, nose, and mouth) are extracted. The features are then extracted using various local descriptors such as LBP, CLBP, LTP, and Dynamic LTP. Finally, to complete the recognition task, the feature vector representing the face image is fed into a multiclass support vector machine. The proposed method outperforms state-of-the-art methods based on other facial decompositions, according to experimental results on two public datasets. The accuracy of the prediction is 75.8%.

Proposed System:

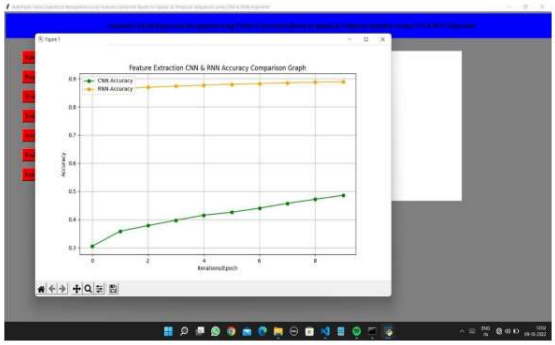
With a combination of gender classification and age estimation, our system aims to identify basic human emotions. Basic emotions include happy, sad, angry, fear, surprise, and neutral emotions on the face. A real-time facial emotion recognition system based on You Look Only Once (YOLO) version 2 architecture and a squeeze net architecture is proposed here. The yolo architecture is a system for detecting objects in real time. It is used in this application to identify and detect faces in real time. For accuracy, these images are captured using anchor boxes. The squeeze net architecture is used for gender classification and age estimation. It provides significant, accurate object detection and extracts high-level features that aid in the classification of images and the detection of objects. With many hidden layers and cross validation in the neural network, both architectures produce more accurate results than other methods. The precision is 88.89%.

Architecture: The DFD is also known as a bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data, various processing performed on this data, and output data generated by the system. 2. One of the most important modelling tools is the data flow diagram (DFD). It's used to simulate the system's components. The system process, the data used by the process, an external entity that interacts with the system, and the information that flows in the system are the components. 3. DFD depicts how information flows through the system and how it is transformed by a series of transformations. It is a graphical representation of information flow and the transformations that occur as data moves from input to output. 4. DFD is also referred to as a bubble chart. A DFD can be used to represent any level of abstraction in a system. DFD can be divided into levels corresponding to increasing information flow and functional detail.



Results:

The process of converting a user-oriented description of input into a computer-based system is known as input design. This design is critical for avoiding errors in the data input process and directing management in the correct direction for obtaining accurate information from the computerised system. 2.It is accomplished by designing user-friendly data entry screens that can handle large amounts of data. The goal of designing input is to make data entry easier and error-free. The data entry screen is set up in such a way that all data manipulations are possible. It also has record viewing capabilities. 3. When the data is entered, it will be validated. Data can be entered by using screens. Appropriate messages are delivered when necessary, preventing the user from becoming mired in instant gratification. Thus, the goal of input design is to create an easy-to-follow input layout.



Conclusion & Future Scope:

The use of machines in society has grown dramatically in recent decades. Machines are now used in a wide range of industries. As their exposure to humans grows, their interactions must become more natural and smooth. To accomplish this, machines must be equipped with the ability to comprehend their surroundings. In particular, the intentions of a human being. Because every expression is a combination of emotions, emotion recognition remains a difficult and complex problem in computer science. For more accurate and efficient facial expression recognition, we proposed an efficient real-time

facial expression recognition system that combines two algorithms, yolo version 2 and squeeze net architecture based on deep neural networks. The future scope could be an action taken in response to emotional recognition. If the system detects a sad emotion, it can play a song, tell a joke, or send a message to the user's best friend. This could be the next step in AI, where the system can understand, comprehend, and react to the user's feelings and emotions. This helps to close the gap between machines and humans. We can also have an interactive keyboard where users can simply use the app and the app will identify the emotion and convert it to the desired emoticon.

References:

[1] Jumani, S.Z., Ali, F., Guriro, S., Kandhro, I.A., Khan, A. and Zaidi, A., 2019. Facial Expression Recognition with Histogram of Oriented Gradientsusing CNN. Indian Journal of Science and Technology, 12, p.24.

[2] J. R. Barr, L. A. Cament, K. W. Bowyer, and P. J. Flynn. Active clustering with ensembles for social structure extraction. In Winter Conference on Applications of Computer Vision, pages 969–976, 2014

[3] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[4] R. Goh, L. Liu, X. Liu, and T.Chen. The CMU face in action (FIA) database. In International Conference on Analysis and Modelling of Faces and Gestures, pages 255–263. 2005.

[5] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In Computer Vision and Pattern Recognition, pages 529–534, 2011

[6] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patchbased probabilistic image quality assessment for face selection and improved video- based face recognition. In Computer Vision and Pattern Recognition Workshops, pages 74–81, 2011.

[7] B. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras. In Biometrics: Theory Applications and Systems, pages 1–8, 2013

[8] N. D. Kalka, B. Maze, J. A. Duncan, K. A. O Connor, S. Elliott, K. Hebert, J. Bryan, and A. K. Jain. IJB–S: IARPA Janus Surveillance Video Benchmark.