

Prediction of Human Facial Expression using Deep Learning

V. D. Punjabi¹, Harshal Mali², Suchita Birari³, Rohini Patil⁴, Aparna Patil⁵, Shubhangi Mahajan⁶

¹Assistant Professor, Dept. of Information Technology, R. C. Patel Institute of Technology, Maharashtra, India

^{2,3,4,5,6}Student, Dept. of Information Technology, R. C. Patel Institute of Technology, Maharashtra, India

Abstract - Facial Expression conveys non-verbal cues, which plays an important roles in interpersonal relations. The Facial Expression Recognition system is the process of identifying the emotional state of a person. Captured image is compared with the trained dataset available in database and then emotional state of the image will be displayed. The results of emotion recognition of seven emotional states (neutral, joy, sadness, surprise, anger, fear, disgust) based on facial expressions. The goal of the system completes using deep learning, with the help of CNN we can archive the goal with higher results. The system will take image or frame as an input and by providing the image to the model the model will perform the preprocessing and feature selection after that it will be predict the emotional state.

Key Words: Deep Learning, K-nn Algorithm, MLP Classifier, Computer vision, kinect in emotion recognition

1. INTRODUCTION

Recognizing human expressions and emotions has drawn the attention of researchers, as the capability of recognizing one's expressions helps in human-computer interaction, to right advertising campaigns, and crowning with an augmented and enhanced human communication, by amending the emotional intelligence ("EQ") of humans. There are many ways to inspect the recognition of human expressions, ranging from facial expressions, body posture, voice tone etc. In this, we have focused on facial expression recognition. Facial Emotion Recognition (FER) is a thriving research area in which lots of advancements like automatic translation systems, machine to human interaction are happening in industries. In contrast the focus to survey and review various facial extraction features, emotional databases, classifier algorithms and so on. This is organized as follows.

1.1 Emotion Recognition

Facial Recognition is the technology that deals with methods and techniques to identify the emotions from the facial expression. Various technological developments in the area of Machine Learning and artificial Intelligence, made the emotion recognition easier. It is expected that expressions can be next communication medium with computers. A Need for automatic emotion recognition from facial expression increases tremendously. Research work in this area mainly concentrates on identifying human emotions from videos or from acoustic information. Most of the research work recognizes and matches faces but they have not used convolutional neural networks to infuse emotions from images. Emotion Recognition deals with the investigation of identifying emotions, techniques and methods used for identifying. Emotions can be identified from facial expressions, speech signals etc. Enormous methods have been adapted to infer the emotions such as machine learning, neural networks, artificial intelligence, emotional intelligence. Emotion Recognition is drawing its importance in research which is primary to solve many problems. The primary requirement of Emotion Recognition from facial expressions is a difficult task in emotional Intelligence where images are given as an input for the systems.

1.2 Facial Emotion Recognition

Facial Emotion Recognition is research area which tries to identify the emotion from the human facial expression. The surveys states that developments in emotion recognition makes the complex systems simpler. FER has many applications which is discussed later. Emotion Recognition is the challenging task because emotions may vary depending on the environment, appearance, culture, face reaction which leads to ambiguous data. Survey on Facial emotion recognition helps a lot in exploring facial emotion recognition [2].

1.3 Deep Learning

Deep Learning is machine learning technique which models the data that are designed to do a particular task. Deep learning in neural networks has wide applications in the area of image recognition, classification, decision making, pattern recognition etc. Other deep Learning techniques like multimodal deep learning is used for feature selection, image recognition etc. [3].

1.4 Applications of Facial Emotion Recognition

Emotion Recognition is used in BPO's for identifying calls based on their emotions. Emotion Recognition serves as the identifier for conversational analysis for identifying the unsatisfied customer, customer satisfaction so on. FER is used in car board system depending on information of the mentality of the driver can be provided to the system to initiate his/her and the customer safety[7].

2. LITERATURE SURVEY

Koelstra S. et al.(2013) The explosion of user-generated, untagged multimedia data in recent years, generates a strong need for efficient search and retrieval of this data. The predominant method for content-based tagging is through slow, labour-intensive manual annotation. Consequently, automatic tagging is currently a subject of intensive research. However, it is clear that the process will not be fully automated in the foreseeable future. They propose to involve the user and investigate methods for implicit tagging, wherein users' responses to the interaction with the multimedia content are analysed in order to generate descriptive tags. A multi-modal approach that analyses both facial expressions and Electroencephalography(EEG) signals for the generation of affective tags. They perform classification and regression in the valence-arousal space and present results for both feature-level and decision- level fusion. They demonstrate improvement in the results when using both modalities, suggesting the modalities contain complementary information [2].

Deepak Grimier et al.(2013) Face detection is concerned with finding whether or not a face exists in a given image; if face is present, it returns the image location and face size. The advance of computer technology in recent years has facilitated the development of real- time vision modules that can interact with humans. Many existing systems use a window-based classifier that scans the entire image for the presence of the human face and such systems suffers from scale variation, pose variation, Illumination changes, etc. Here, they propose a lighting insensitive face detection method based upon the edge and skin tone information of the input color image[4].

Ratliff M. S.et al. Recognizing emotion using facial expressions is a key element in human communication. A framework for the classification of emotional states, based on still images of the face. The technique present involves the creation of an active appearance model (AAM) trained on face images from a publicly available database to represent shape and texture variation key to expression recognition. Parameters from the AAM are used as features for a classification scheme that is able to successfully identify faces related to the six universal emotions.

Kanade T. et al. Automatic expression analysis systems attempt to recognize a small set of prototypic expressions, such as happiness, anger, surprise, and fear. Such prototypic expressions, however, occur rather infrequently. Human emotions and intentions are more often communicated by changes in one or a few discrete facial features. They develop an automatic face analysis (AFA) system to analyze facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features in a nearly frontal-view face image sequence.

Mao Q. et al. Emotion recognition via facial expressions (ERFE) has attracted a great deal of interest with recent advances in artificial intelligence and pattern recognition. Most studies are based on 2D images, and their performance is usually computationally expensive. A real-time emotion recognition approach based on both 2D and 3D facial expression features captured by Kinect sensors. To capture the deformation of the 3D mesh during facial expression, it combine the features of animation units (AUs) and feature point positions (FPPs) tracked by Kinect. A fusion algorithm based on improved emotional proles (IEPs) and maximum confidence is proposed to recognize emotions with these real-time facial expression features. Experiments on both an emotion dataset and a real-time video show the superior performance of our method. In a real-time ERFE approach based on animation units (AUs) and feature point positions (FPPs) tracked by a Kinect sensor to recognize six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) and neutral.

Lundqvist D. et al.(1998)It has been argued that visual subliminal processing of face-inherent sex information might depend on the retino-collicular projection. In the current study, they tested this possibility. Based on the known exclusive sensitivity of the retino-collicular projection for low spatial frequencies (LSF), they derived the following predictions. First, masked and, therefore, subliminal high-spatial frequency (HSF) face primes would not be processed, but, secondly, masked unfiltered face primes, thirdly, masked LSF face primes, and, fourthly, unmasked HSF primes Should be processed and should lead to a congruence effect that is, faster responses (better performance) in congruent conditions, with face primes and face targets of the same sex, as compared to incongruent conditions, with face primes and face targets of Different sexes.

Tian Y. I. et al. Most automatic expression analysis systems attempt to recognize a small set of prototypic expressions, such as happiness, anger, surprise, and fear. Such prototypic expressions, however, occur rather infrequently. Human emotions and intentions are more often communicated by changes in one or a few discrete facial features. They develop an automatic face analysis (AFA) system to analyze facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal-view face image sequence. The AFA

system recognizes fine-grained changes in facial expression into action units (AU) of the Facial Action Coding System (FACS), instead of a few prototypic expressions.

Li B. Y. L. et al. an algorithm that uses a low resolution 3D sensor for robust face recognition under challenging conditions. A preprocessing algorithm is proposed which exploits the facial symmetry at the 3D point cloud level to obtain a canonical frontal view, shape and texture, of the faces irrespective of their initial pose. This algorithm also fills holes and smooths the noisy depth data produced by the low resolution sensor. The canonical depth map and texture of a query face are then sparse approximated from separate dictionaries learned from training data. The texture is transformed from the RGB to Discriminant Color Space before sparse coding and the reconstruction errors from the two sparse coding steps are added for individual identities in the dictionary.

3. METHODOLOGY

3.1 Method:

In the classification process, six action units (AU), calculated by the Kinect device, were used as features. Figure 3.1 shows exemplary values of AU for facial expressions of one of the participant. The images correspond to emotional states (ES): neutral, joy, surprise, anger, sadness, fear, disgust.

ES	neutral	joy	surprise	anger	sadness	fear	disgust
AU0	-0.21	0.77	-0.10	0.30	0.17	-0.11	0.91
AU1	-0.06	0.09	0.60	-0.07	-0.04	0.20	0.13
AU2	-0.25	1.00	-0.49	0.06	-0.37	-0.60	0.88
AU3	-0.21	0.00	-0.13	0.04	-0.09	-0.17	0.00
AU4	-0.04	-0.47	0.58	-0.19	-0.02	0.38	-0.32
AU5	-0.23	-0.30	0.10	-0.34	-0.27	-0.02	-0.39

Figure 3.1: The Facial Expressions and Corresponding AUs.

Source: <https://www.sciencedirect.com/science/article/pii/S1877050917305264>

An exemplary distributions of AU2 and AU5 obtained. Even preliminary analysis of the distribution of the features allows us to see the ability to distinguish some emotions. It decided to test the possibility of automatic recognition of emotions using AU. The experiments were conducted using k-NN and MLP classifiers [13].

3.2 Face Registration

Face Registration is a computer technology being used in a variety of applications that identifies human faces in digital images. In this face registration step, faces are first located in the image using some set of landmark points called "face localization" or "face detection". These detected faces are then geometrically normalized to match some template image in a process called "face registration".

3.3 Facial Feature Extraction

Facial Features extraction is an important step in face recognition and is defined as the Process of locating specific regions, points, landmarks, or curves/contours in a given 2-D image or a 3D range image. In this feature extraction step, a numerical feature vector is generated from the resulting registered image. Common features that can be extracted are-

1. Lips
2. Eyes
3. Eyebrows
4. Nose tip

3.4 Classifiers

A. K-NN Classifier

The K-nearest neighbors (KNN) algorithm is a type of supervised machine learning algorithms. KNN is extremely easy to implement in its most basic form, and yet performs quite complex classification tasks. It is a lazy learning algorithm since it

doesn't have a specialized training phase. Rather, it uses all of the data for training while classifying a new data point or instance. KNN is a non-parametric learning algorithm, which means that it doesn't assume anything about the underlying data. This is an extremely useful feature since most of the real world data doesn't really follow any theoretical assumption e.g. linear-separability, uniform distribution, etc. [14].

B. MLP Classifier

A multilayer perceptron (MLP) is a class of feed forward artificial neural network. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable. Multilayer perceptron are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer [14].

3.5 Support Vector Machine (SVM)

SVM is one of the supervised learning classifiers and also has non-linear features. SVM gives a generalized performance since it uses statistical learning theory. SVM gives optimum performance for limited size dataset. Different kernel functions can be tried to get the optimum performance of SVM classifier. It provides an interactive and dynamic approach. To get high-performance efficiency, selection of kernel function plays an important role. Support vector machines have previously been successfully employed in a variety of classification applications, including identity and text recognition as well as DNA micro array data analysis. SVM is widely used in various pattern recognition tasks. SVM is a state-of-the-art machine learning approach based on the modern statistical learning theory. SVM can achieve a near optimum separation among classes. SVMs is trained to perform facial expression classification using the features proposed. In general, SVM are the maximal hyper plane classification method that relies on results from statistical learning theory to guarantee high generalization performance. Kernel functions are employed to efficiently map input data which may not be linearly separable to a high dimensional feature space where linear methods can then be applied. SVMs exhibit good classification accuracy even when only a modest amount of training data is available, making them particularly suitable to a dynamic, interactive approach to expression recognition [10].

An ideal separation is achieved when the hyper plane and the training data of any class is the largest. This separating hyper plane works as the decision surface. SVM has been successfully employed for a number of classification tasks such as text categorization, genetic analysis and face detection.

3.6 Proposed System

Face recognition system can be formulated as following phases [15].

1. Image Data

Images used for facial expression recognition are still images or image sequence. Two dimensional facial images are used. Two dimensional human facial images pixel and intensities are used as input data set.

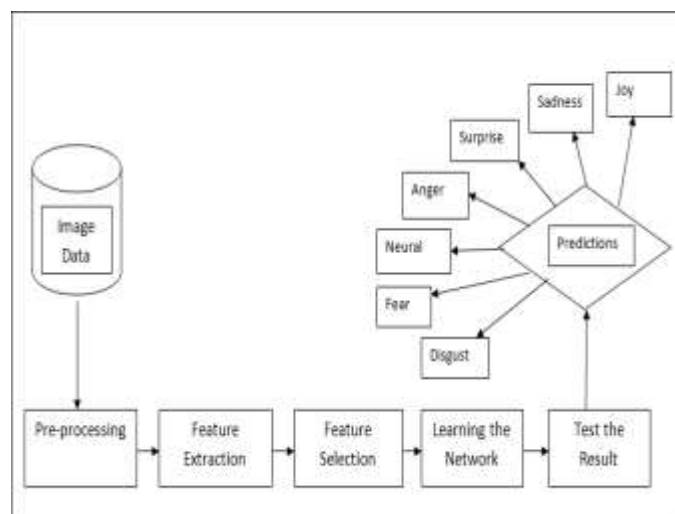


Figure 3.2: Face Recognition Phases

Facial image acquisition and measures for image quality and their usage. Section on image acquisition for face recognition discusses issues in lighting, sensor, lens, blur issues, which impact short-range biometrics but are more pronounced in long-range biometrics. The design of controlled experiments for long-range face and why they are needed. Some of the weather and atmospheric effects that occur for long-range imaging, with numerous of examples. Measurements of "system quality," including image-quality measures and their use in prediction of face recognition algorithm. This section also introduces the concept of failure prediction and techniques for analyzing different "quality" measures. The section ends with a discussion of post-recognition "failure prediction" and its potential role as a feedback mechanism in acquisition. Each section includes a collection of open-ended questions to challenge the reader to think about the concepts more deeply. For some of the question, answer them after they are introduced; others are left as an exercise for the reader.

2. Pre-Processing

Pre-processing of acquired images is important step to get efficient result. It includes noise removal and normalization against the variation of pixel position or brightness. PCA technique is used for normalization and resizing of image pixel in $N \times N$ matrix. Face recognition has become a popular area of research in computer vision, it is typically used in network security systems and access control systems but it is also useful in other multimedia information processing areas. Performance of the face verification system depends on many conditions. One of the most problematic is varying illumination condition. They discuss the preprocessing method to solve one of the common problems in face images, due to a real capture system i.e. lighting variations. The different stages include gamma correction, Difference of Gaussian (DOG) filtering and contrast equalization. Gamma correction enhances the local dynamic range of the image in dark or shadowed regions while compressing it in bright regions and is determined by the value of. DOG filtering is a grey scale image enhancement algorithm that eliminates the shadowing effects. Contrast equalization rescales the image intensities to standardize a robust measure of overall intensity variations. The technique has been applied to Yale-B data sets, Face Recognition Grand Challenge (FRGC) version 2 and a real time created data set.

3. Segmentation

Segmentation is concerned with the demarcation of image portions conveying relevant facial information. Shape, Motion, Color, Texture and Spatial configuration of the face or its components. Covariance Matrix of dimensions $N^2 \times N^2$ to determine eigenvectors and eigenvalues. Present for face segmentation and facial landmark detection in range images. Our goal was to develop an automatic process to be embedded in a face recognition system using only depth information as input. To this end, our segmentation approach combines edge detection, region clustering, and shape analysis to extract the face region, and our landmark detection approach combines surface curvature information and depth relief curves to find the nose and eye landmarks. The experiments were performed using the two available versions of the Face Recognition Grand Challenge database and the BU-3DFE database, in order to validate our proposed methodology and its advantages for 3D face recognition purposes.

4. Classification

Classification technique is used to classify image category. In this Eigen vectors they are compared with training set to classify the image of different moods. This is the last process to get desired output. The intent of the classification process is to categorize all pixels in a digital image into one of several land cover classes, or "themes". This categorized data may then be used to produce thematic maps of the land cover present in an image. Normally, multispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization.

3.7 Applications

A facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication. They are a primary means of conveying social information between humans, but they also occur in most other mammals and some other animal species. (For a discussion of the controversies on these claims, see Fridlund and Russell Fernandez Dols.

1. With facial expression recognition systems computer will be able to assess the human expression depending on their effective state.
2. Facial expression recognition has practical significance it has very broad application prospects such as user-friendly interface between man and machine, humanistic design of goods and emotional robot etc.
3. The facial expression recognition system has been used in communication to make the answer machine more interactive with people.

4. The facial expression recognition system applied in different areas of life such as security and surveillance, they can predict the offender or criminal's behavior by analyzing the images of their faces that are captured by the control camcorder.

4. CONCLUSION

Prediction of human facial expression using deep learning are important in human communication and interactions. Also, they are used as an important tool in behavioral studies and in medical rehabilitation. Facial image detection techniques provide a fast and practical approach for non-invasive detection. The purpose to present this concept was to develop an intelligent system for facial image expression classification using neural networks.

REFERENCES

1. Przyby lo J., Automatyczne rozpoznawanie element Pow mimiki twarzyw obrazie analizaich przydatnoP sci do sterowania, rozprawa doktorska, Akademia GPornicz- Hutnicza, KrakPow, 2008.
2. Koelstra S., Patras I., Fusion of facial expressions and EEG for implicit affective tagging, Image and Vision Computing, 31 (2013), no. 2, 164-174
3. <https://www.investopedia.com/terms/d/deep-learning.asp>
4. Ekman P., Friesen W., Facial Action Coding System, Consulting Psychologists Press, Stanford University, Palo Alto, 1977.
5. <https://towardsdatascience.com/wtf-is- image-classification-8e78a8235acb>
6. <https://www.zeolearn.com/magazine/fiv e-important-techniques-that-you- shouldknow-about-deep-learning>.
7. Ahlberg J., CANDIDE-3 - An Updated Parameterised Face, 2001.
8. Microsoft Kinect, <https://msdn.microsoft.com/en- us/library/jj131033.aspx>.
9. Deepak Ghimire, Emotion recognition using facial expressions with active appearance models,2013, 812-840.
10. Ratli 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.
11. Kanade T., Cohn J. F., The Karolinska Directed Emotional Faces -KDEF, 1998,.41
12. 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.
13. Lundqvist D., Flykt A., • Ohman A., The Karolinska Directed Emotional Faces KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, no. 1998,.
14. Ogiela M. R., Tadeusiewicz R., Pattern recognition, clustering and classification applied to selected medical images, Studies in Computational Intelligence, 84 (2008), no. 117-151.
15. <https://www.sciencedaily.com/re leases/2018/05/180531131102.html>
16. 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
17. 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