PalArch's Journal of Archaeology of Egypt / Egyptology

A NOVE FACIAL EXPRESSION RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

K. Sai Krishna[1], B. Rajashekar[2], M. Kiran Kumar[3]

[1,2,3]Department of Electronics and Communication Engineering, Anurag University

Ghatkesar, Medchal (Dist), INDIA

Corresponding Author: Email: [3]kirankumarece@cvsr.ac.in

K.Sai Krishna, B.Rajashekar, M.Kiran Kumar. A Nove Facial Expression Recognition System Using Convolutional Neural Networks -- Palarch's Journal of Archaeology of Egypt/Egyptology 18(9), 1040-1048. ISSN 1567-214x

Keywords: Convolutional Neural Network (CNN) model, Facial Expression Recognition (FER), geometric features.

ABSTRACT:

The more significant, effective and non-verbal method of communication is the Facial expression using the emotions. In the machine learning field, Facial Expression Recognition (FER) is an important subject. Particularly, Facial recognition system has become the foremost significant research topic in biometrics. The training of system with a small dataset using the available data has a significant impact which can negatively affects the performance. Hence a novel facial expression recognition system using convolutional neural network (CNN) is proposed in this paper. Here, a novel method of extracting geometric features that uses simple techniques for calculating the facial components to obtain the robustness of any pose variation is employed. The huge amount of data is trained in this FER system by using a deep learning architecture of Convolutional Neural Network (CNN) model. Various types of filtering techniques have been applied for augmenting the little amount of small business employee images. An attempt can also be made to determine which data augmentation options had the greatest impact on facial recognition, so facial recognition by training each image with many features on a new augmented data set is performed with non-real time application. The proposed facial expression recognition system is validated with these obtained features using the augmented emotion data set of Cohn Kanade + (CK+) and provides the six types of emotion results with high accuracy than all of the other FER system technologies.

INTRODUCTION

In present trend of developing technologies with their innovation and usage all over the wor may lead to the most of the people spending their time to interact with computers and mobile phones. The huge amounts of software interfaces existed are the primordial, concise and nonverbal though they can play essential

roles in our lives. The Human-Computer interaction (HCI) can be improved greatly by the addition of recognizing emotional expressions to anticipate the feelings and emotional state of users. Human-Computer interaction systems have been seen as one of the fastest growing and most attractive areas and also human-robot interaction (HRI) systems and health systems have developed remarkably and rapidly [1] [7].

People often use various types of signs to convey their emotions like with voice, hand gestures and Facial expressions. In total approximately up to 50% of human communication is represented with the facial expressions, whereas 7% expressions of emotions are represented by the other forms such as verbal communication [2]. Taking facial expressions into account in a human-robot interaction system therefore allows to successfully simulating the natural interactions. In examine the behavioral patterns of people in order to enhance their mental emotional states an efficient facial expression recognition (FER) system can help effectively a lot. There are two major stages present in a facial expression recognition system: they are feature extraction and then classification carried out with the classifiers. In the determination of recognition outcome feature extraction has a significant part. The optimal working of even a best classifier would also be fails if extracted features are bad. More numbers of features were generated in the majority of feature extraction cases. Selection of necessary and important features which ensures the criteria for facial expression recognition is the main role. Accuracy of the system can be increased additionally by supply the system using these selected features [3][4].

In this paper, a novel method of facial recognition system using convolution neural networks is presented, in which six feature types of facial emotional expressions are extracted from an augmented data set of Cohn Kanade (CK+) using the deep learning based approach of convolution neural networks with the geometric feature calculations. The geometric feature calculation employs a feature extraction based on shape of face that means it uses the shape of facial components like eyes, eyebrows, nose, mouth, etc. The similarity between two expressions and the variations of facial expression are the major problems of this facial expression recognition system. In addition, the geometric properties are resistant to variation in the position of the face, skin, size and orientation of the head.

Fer Approaches Based On Deep Learning

In the communication of people face has a significant part. The people communication is contributing with non-verbal information collected from the facial expressions and movements. Several applications can be facilitated by the reorganization of facial expressions from the face images in the human-computer interaction area. Over the past two decades, prospects and developments in the field of multimedia signal processing have drawn the attention of several computer vision researchers to focus on the problems of facial expression recognition. The evidence for categorization of basic facial expressions has been provided in the late 1970s by the Ekman's fundamental studies. Angry, disgust, fear, happy, sad, surprise and neutral represent the basic categories of facial expressions which were categorized based on the above

studies. Ekman and Friesen introduced the Facial Action Coding System (FACS) for recognizing the facial expressions where the action units are implemented to describe the facial movements. Number of researches has been motivated by this work to study facial expressions in a 2-Dimensional technology using image and video processing. In this, several facial expressions can be classified by the measurement of number of facial movements and by tracing the facial features. Those seven basic expressions were used in the recently developed systems as the fundamentals for the analysis and reorganization of facial expressions.

Deep learning Model

The requirement of the deep learning model can be addressed by the two most important approaches such as the interpretable model and the explanatory model. The input modification methods are contained in the study of Interpretable models which combine the internal deep neural network activations by semantic concepts aimed to discover common mid-level visual patterns that appear in image acquisitions. The generation of explanation through the bounded boxes or texts and methods for generation of visualizations at precise pixel level is contained in the study of explanation model [5].

Facial Expression Recognition

Extracting the necessary features from a single image is the one of the most critical steps in the designing that describes the physical phenomena of facial expression system. The Local Gabor features, Weber Local Descriptor and Histograms of oriented gradients are the most widely used feature descriptors for the facial expression recognition by the conventional studies on their activities. Though, the resulted features from the extraction generally have a large dimension. Therefore most general reduction methods of non-negative matrix factorization, local binary patterns and a principal component analysis are used. The interrelationship across the hidden layers are obtained by the deep learning model being study the way of extracting features with its ability than other traditional models. The way of extracting proper features for the facial expression recognition activities can automatically studied by this Deep learning model. The facial expression recognition system uses the auto-encoder model as feature descriptor. Convolutional neural networks (CNNs) are not only suitable as feature extractors, but also as classifiers. The most widely used models are the Dynamic Bayesian Grid, Gaussian Mix model, Support Vector Machines and Deep neural networks. In the extraction of image level features from the sequences of facial image, Hidden Markov models are used [9].

The Multi-layer Perception Model, Support Vector Machines and k-Nearest Neighbors classifiers have been used conventionally by the most of the researches for proposing their techniques. The texture, face landmark features, gradient feature mapping, Eigen vectors, Histogram of Oriented Gradients, etc. are the hand crafted features which were used by those classifiers [6][8]. The above feature can be extracted by using the methods of Local Binary Patterns, Gabor filters, Eigen Faces, Principal Component Analysis and Linear Discriminate Analysis. Generally, CNN also uses these functions however in its

own method. The differentiation is that traditional methods require us to manually extract features (also known as "handcraft features"), while CNNs can automatically learn these features on their own.

Fer System Using Cnn

The basic idea behind the implementation of proposed system is to obtain training data for use on the CNN model, which is designed to obtain a large number of samples from the face images containing a small number of samples in the source dataset without disturbing the integrity of the original image. This will reveal how the non-real-time face recognition system is affected by the classification performance. Figure (1) shows the frame work of proposed facial expression recognition system using convolution neural network.

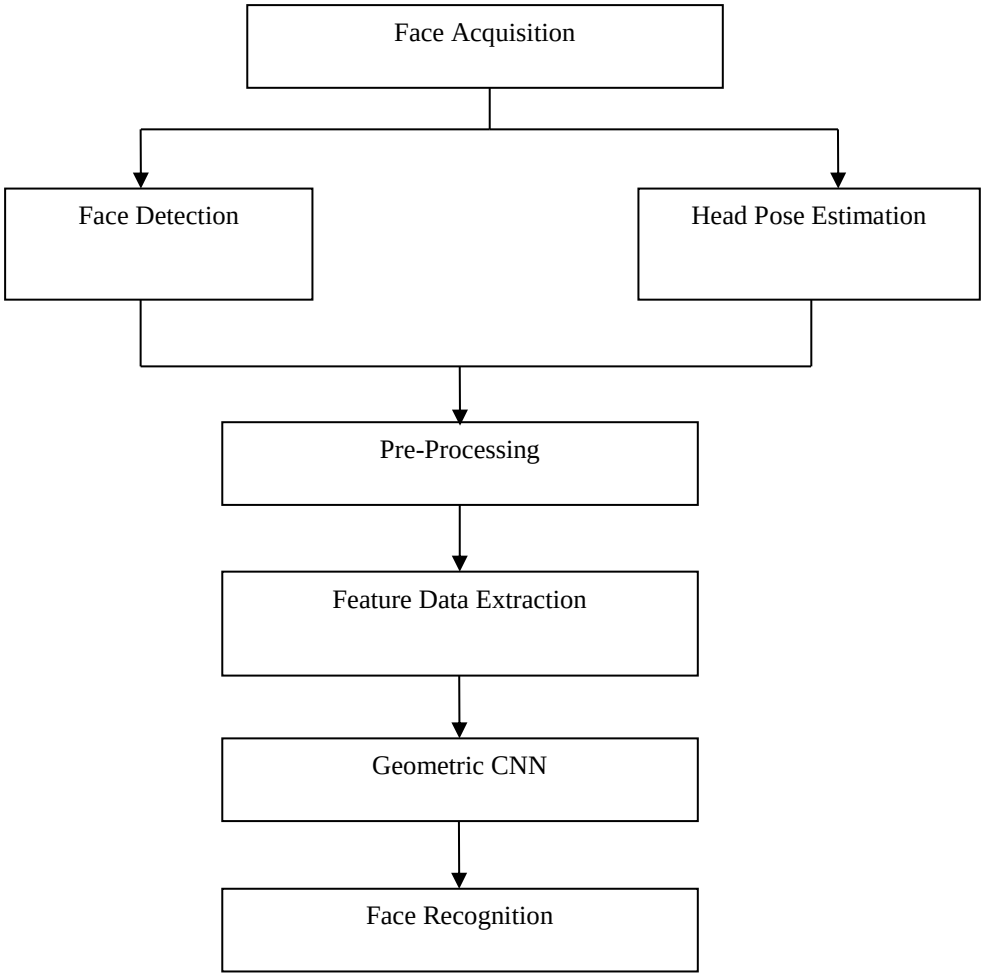

Fig. 1: Frame Work of Facial Expression Recognition System Using CNN

Face Acquisition

There is a necessary of gathering the data in the first phase of implementing proposed system together with the stated multimodal data formats. The best and efficient source of data is the mini-mental state examination dataset is but the data is documented merely under laboratory-controlled conditions. The

equipment that capable of recording various types of data is needed to gather the data in real time environments and people should simply use it if they were requested to participate in experiment sessions. In general, data can be simple captured and acquired by using the infrared-based tracking sensors, eye tracking sensors and electroencephalogram signal based headphones. Furthermore, the facial expressions of multiple subjects are being natural and spontaneous to be as precise as real-time environment and as minimum as six types of basic expressions are recorded by each.

Pre-processing of Images

The pre-processing method is categorized into two steps: Face detection and head Pose estimation. Each stage of processing can be explained below in detail. The pre-processing method proceeds with performing the data augmentation process in order to get sufficient training samples. Basically, generating the adequate data samples as of from original images is called as data augmentation.

Face Detection

The major concern of this Facial expression recognition system is the face detection. Even though input image has been gathered from the benchmark facial expression dataset, there exists unnecessary background information in the facial image which is unrelated to the expression recognition. Therefore, accuracy of the facial recognition results that must eliminate as much unrelated background information as possible will provide the exact Facial expression recognition. The proposed system implements the frequently used Viola_Jones framework for facial the recognition.

Head Pose Estimation

As stated earlier, the majority of the approaches existed before are focused on controlled conditions in the laboratory where faces are usually in the front view. In the real time environment, however, front view is not obtainable for all time and therefore represents a challenge in recognizing facial expressions. Various aspects of the views have been taken into account by the only few approaches but with little accuracy. The variations in the facial shapes are captured by extracting the geometry features from the wrap transformation of facial marks and the system is allowed to track the facial components by extracting of dynamic face textures using the features of the by Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP). By the Acted Facial Expressions in the wild dataset, this technique attains an accuracy of 46.8%. A win 7 operating system having a Core-i7 Central processing unit with the 64-bit is used to perform the proposed technique in 27 msec. Hence, this technique can be used in real time applications with exclusive of graphic processing units. The design of facial expression recognition system challenging to facial movements and variations can be done by using two visual features and an audio feature vector. Whenever the faces are not in front view, spontaneous smile is recognized in the normal situations. The relationship between the image fields and the strength of the characterized smile is extracted, which depends on the posture of the head and the accuracy rate can be improved by training the

augmented data sets with the combination of various convolution networks using a multi label approach.

The accuracy rates of 94% with the labelled Faces in the Wild dataset and 92% with support vector machine dataset along with improved speed can be achieved with the development of ensemble algorithm by choosing only a small number of networks. The head posture problem can be overcome with the help of combination of different features obtained from different sensors. A higher performance can be demonstrated with the proposed system in real time applications with the runtime of 9msec on a computer with a 4.2GHz Core i7 CPU.

Feature Data Extraction

Feature extraction is an important step in a facial expression recognition system for the recognition of emotions. As the features extracted are able to reduce the variations within the class and enhance the variations between the classes, this facial expression recognition system can perform dynamically. An accurate and effective process of recognition can be ensured with a better representation of features. The appearance based and the geometric based are the two kinds of features based facial expression recognition models. First, the facial expression recognition system relating to the extraction of position, angle and shapes of several facial components such as eye, nose, ear and mouth which were the local facial features then the resultant vectored features are presented according to the geometric relationship. Later on, this system relating to the development of active appearance model and use the facial textured information as the vectored feature. As the appearance-based model is a complicated process to perform in a real time application, the geometric based models are considered as a best and most widely used model than appearance-based model which can achieve a higher rate of facial expression recognition along with more accuracy rate.

Geometric CNN

Feature extraction of facial geometry components, active appearance model and the convolution neural network based facial expression classifier are combinedly used together for the development of facial expression recognition system. The frame work of this facial expression recognition system is shown in the figure (1). The stilled images are the inputs of this system which consisting of a particular face only. The active appearance model which was illustrated in the earlier stages is used for the detection of geometrical points on the face and 68numbers of geometrical points on a face are detected at its output. All these facial geometric points are divided into five basic components and the facial geometric points under the process of Feature extraction of facial geometry components are only used that influence the deformation of facial expression in the eyes, inner eyebrows, eyebrows, nose and mouth. Then in a next step, resultant vectors are given to a convolution neural network based facial expression classifier and the labels of emotion classes are the output for this classifier. Deep learning based convolution neural networks are the classifiers employed in this proposed work for the facial expression reorganization as part of our doctoral dissertation. Since the ambiguity in the

values of facial components from different subjects can be tolerated by these convolution neural networks classifiers, they are selected as the best classifiers for this system.

Deep learning architectures include CNN. CNN is a special kind of convolution neural network. The main feature of this method is that it allows learning characteristics at different levels with a provision of a high level input data abstract representation. Therefore, it can be applied to the network without any attribute extraction. Hence a convolution neural network is an end-to-end classifier. Formation of feature extractors based on the data used for training is the additional benefit. The deep learning based convolution neural network model with dropout only used to extract the functionality for the facial expression recognition activity. The accuracy of the proposed convolution neural network model is 94%. The outputs are considered as the extracted features from the quadrants grouping layer.

RESULTS

The Keras based framework is used in this for evaluating the performance of proposed Facial Expression Recognition System in a Linux platform. The CUDA Deep Neural Network (cuDNN) library along with the frame work of NVIDIA CUDA 6.5 having a NVIDIA GTX 1080 GPU (8GB) standard is used to perform these experiments. A benchmarking dataset of Cohn Kanade (CK+**)** that consisting of Red Green Blue-Dataset (RGBD) face images is used to enable effective and efficient assessments. The description about data set used here is given below.

Cohn Kanade (CK+) Dataset

This data set has around a 600 completely interpreted sequences which characterize the angry, disgust, fear, happiness, sadness, surprise and neutral expressions corresponding to nearly 125 male and female subjects. Just six fundamental expressions which include angry, disgust, fear, happiness, sadness and are used in the evaluation process. Since the every sequence in the dataset starts by the neutral expression and then continues with peak expressions, last frame is chosen in the every sequence. Therefore, approximately 80-120 samples are chosen for the each and every expression. The number of samples regarding to each expression can be increased 50 times by using the Data augmentation (usage of simple operations such as rotation, translation and skewing). At last, a 10-fold cross-validation can be implemented for the evaluation.

Moreover, every sequence in this the CK + dataset contains Action Unit (AU) labels. It has to be taken into consideration that neutral facial images have no AUs and the remaining every sequence have AUs in the last three frames. The relationship between the emotion labels of the AUs and their distribution can be found by performing the pre-processing stage. AUs are more likely to appear with some emotion labels for example anger emotions as AU 4, 7, 17, 23, 24 and Happy emotion as AU 1, 2, 4, 5, 20, 25 as shown in Figure (2).

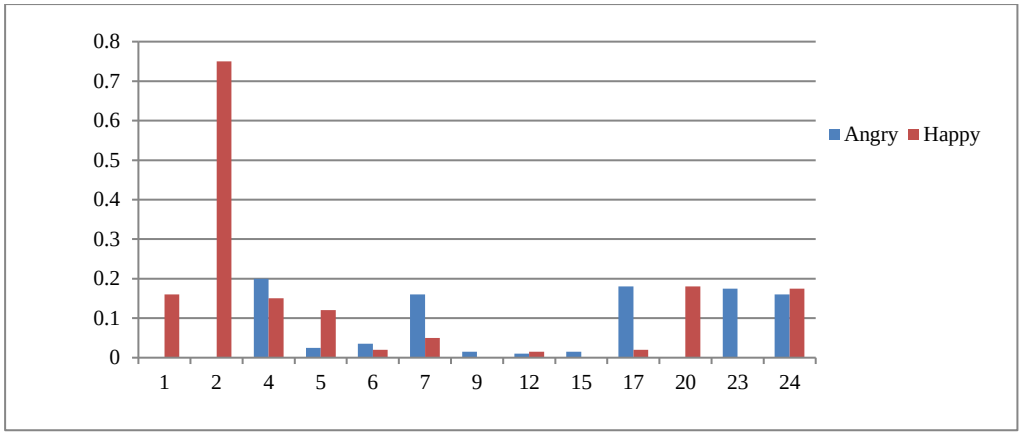

Fig. 2: Distribution of AUs In Angry and Happy Emotions

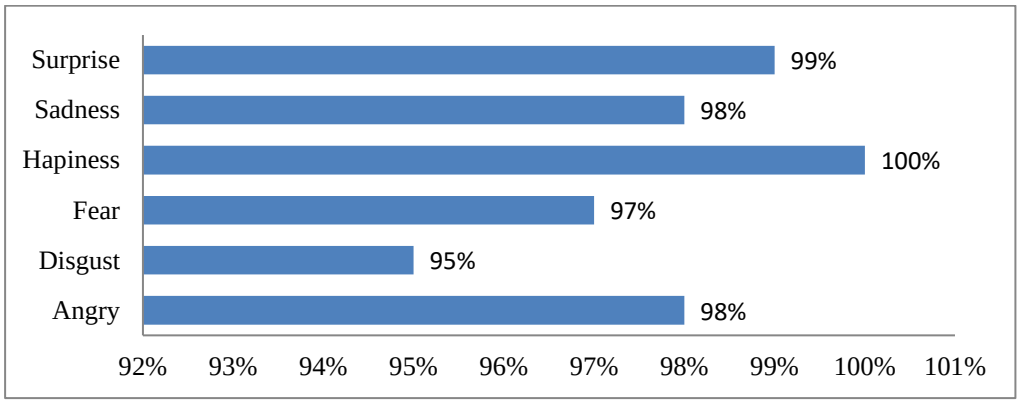

Fig. 3: Recognition Results of Six Basic Facial Expressions

The proposed geometry feature expression values are measured in the experiment on a dataset to calculate the accuracy rate of Convolutional Neural Network classifier. The fuzzy-based facial components inference system is utilized here for testing those geometry feature expressions that works with classification of fuzzy linguistic values for every facial component. As shown from the result (Figure (3)), the values of geometric features validity have been confirmed by the applied expert in the CK + data set.

Table 1: Accuracy Comparison of Different FER Approaches

S. No	Classifier	Accuracy Rate
1	FCM	81%
2	FIS	75%
3	SVM	92%
4	Proposed CNN Geometry FER	94%

CONCLUSION

In particular, in order to extract better attributes in image analysis, traditional feature extraction and an alternative to artificial neural network methods, Convolutional Neural Network (CNN) models, which are deep learning models, have started to be developed. Basically, facial expressions are the collection of

facial emotions or movements and the proposed deep learning based convolutional neural network model can classify these emotions from the facial images. The proposed Facial Expression Recognition system proceeds with a process of feature extraction that extract the essential features based on the face geometry components. The basic idea behind the proposed system is to provide simple calculation method for recognizing facial geometry features which represents the values of image coordinates. It was illustrated from the experimental results that the proposed method has high performance of operation with more accuracy rate in recognition of facial expressions compared to the other classifier models.

REFERENCES

S. Singh and F. Nasoz, "Facial Expression Recognition with Convolutional Neural Networks," 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020, pp. 0324-0328, doi: 10.1109/CCWC47524.2020.90312

Hyeoncheol Kim and Sunbin Kim, "Deep Explanation Model for Facial Expression Recognition Through Facial Action Coding Unit", IEEE Inte. Confe. on Big Data and Smart Computing (Big Comp), year-2019.

S. Xie and H. Hu, "Facial Expression Recognition Using Hierarchical Features With Deep Comprehensive Multipatches Aggregation Convolutional Neural Networks," in IEEE Transactions on Multimedia, vol. 21, no. 1, pp. 211-220, Jan. 2019, doi: 10.1109/TMM.2018.2844085.

Min Li, Huiquan Zhang, Chao Qi, Qiushi Wang, Huailing Zhang, Zhifan Gao and Jinling Xing, "Facial Expressions Recognition Based on Cognition and Mapped Binary Pattern", IEEE Access, Volume: 6, Year: 2018.

Z. Xie, Y. Li, X. Wang, W. Cai, J. Rao and Z. Liu, "Convolutional Neural Networks for Facial Expression Recognition with Few Training Samples," 2018 37th Chinese Control Conference (CCC), 2018, pp. 9540-9544, doi: 10.23919/ChiCC.2018.8483159.

Bowen Pan, Qiang Ji, Shangfei Wang and Huaping Chen, "Thermal Augmented Expression Recognition", IEEE Transactions on Cybernetics, Volume: 48, Issue: 7, Year: 2018.

J.G. Trafton and P. Baxter, "Cognitive Architectures for Human-robot Interactions", 2014 ACM/IEEE proceedings of Inte. Conf. HRI, Bi., ele., feld, pp: 504-505, March, 2016.

Jeff Donahue, Trevor Darrell, Lisa Anne Hendricks, Marcus Rohrbach, Zeynep Akata and Bernt Schiele, "Generating visual explanations", Max Welling, Nicu Sebe, Bastian Leibe and Jiri Matas editors, In European Conference on Computer Vision (ECCV), year-2016.

Kiran Kumar Mandrumaka, M. Mounika, Fazal Noorbasha, "Design and Testing of 16 bit Carry Save Adder using Reconfigurable LFSR", IJATCSE, Volume 9, No.5, September - October 2020.

C.-C Loy, P. Luo, Z. Zhang and X. Tang, "Learning social relation traits from face images", IEEE Proceedings of the International Conference on Computer Vision, Santiago, Chile, 13-16, pp. 3631–3639, December, 2015.