IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

Facial Emotion Recognition using CNN

Subina Thasneem P.A.[1], Santhi P.[2]

PG Student, Computer Science and Engineering, IES College under Kerala Technological University,Thrissur,India[1]

Assistant Professor, Computer Science and Engineering, IES college under Kerala Technological University,

Thrissur, india[2]

Abstract: The ability to analyse facial expressions plays a major role in non-verbal communication. If a someone only analyses what a person's mouth says and ignores what the person's face says, then we can only have a part of the story. Humans were the only ones who could distinguish between expressions but not anymore, with advancing technology our computers can learn how to detect emotions as well. This report is a guide to facial expression recognition software using OpenCV, Keras, Tensorflow and CNN, by implementing a program in Python it has become possible to build an algorithm that performs detection, extraction, and evaluation of these facial expressions for automatic recognition of human emotion in real-time. The main features of the face are considered for the detection of facial expressions. To determine the different emotions, the variations in each of the main features are used. To detect and classify different classes of emotions, machine learning algorithms are used by training different sets of images. This paper discusses a real-time emotion classification of a facial expression into one among the seven universal human expressions: Anger, Disgust, Fear, Happy, Neutral, Sad, Surprise by the implementation of a real-time vision system that can classify emotions.

Keywords: CNN, Facial Emotion Recognition, FACS, EMA.

I. INTRODUCTION

Human emotions and facial expressions have quite important role for communication in the society and there are lots of research for understanding the relation between the emotions and corresponding facial expressions for decades. With the increase of the usage of social media and photo/video-based applications, the visual data in the Internet grows exponentially. Therefore, the community finds ways to use this data in different kinds of psychological and social experiments. Facial expression recognition is one of the ways that can be used in different kind of research areas such as psychological behavior examination, human-machine interactions and human services applications.

Convolutional neural networks are designed for taking advantage of the two dimensional structure of an input image. Therefore this method can fit better than other learning methods for facial expression recognition problem. Another benefit of CNN is that it is easier to train and have fewer parameters than fully connected networks.

There are two main features of this work. First, in this project, different regularization and data augmentation techniques are used with the convolutional neural network and the study has important role in order to observe that the effect of the configuration of the convolutional neural network layers and used fine-tuning and training techniques. Second, the real time test interface provides the researchers an opportunity to understand the convolutional neural network structure and test the trained network with online data in order to see the effect of the network and how successful the network with these configuration and training techniques.

This paper provides an overview of a designed method for facial expression classification and different techniques for data extraction, augmentation and classification based on CNN architecture. The rest of the paper is organized as follows: Section II provides background information related to facial expression classification with CNN structure from the literature, Section III describes the proposed method for facial expression classification and techniques that are used, Section IV gives an experimental results and discussion, Section V concludes the paper.

II. LITERATURE SURVEY

The face is the most expressive and communicative part ofa human being [1]. It's able to transmit many emotionswithout saying a word. Facial expression recognitionidentifies emotion from face image, it is a manifestation of theactivity and personality of a human. In the 20th century, the American psychologists Ekman and Friesen [2] defined six basics' emotions (anger, fear, disgust, sadness, surprise and happiness), which are the same across cultures.

Facial expression recognition has brought much attentionin the past years due to its impact in clinical practice, sociable robotics and education. According to diverse research, emotion plays an important role in education. Currently, a

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

teacher use exams, questionnaires and observations as sourcesof feedback but these classical methods often come with lowefficiency. Using facial expression of students the teacher canadjust their strategy and their instructional materials to help foster learning of students.

The purpose of this article is to implement emotion recognition in education by realizing an automatic system that analyze students' facial expressions based on Convolutional Neural Network (CNN), which is a deep learning algorithm that are widely used in images classification. It consist of a multistage image processing to extract feature representations. Our system includes three phases: face detection,normalization and emotion recognition that should be one ofthese seven emotions: neutral, anger, fear, sadness, happiness,surprise and disgust.

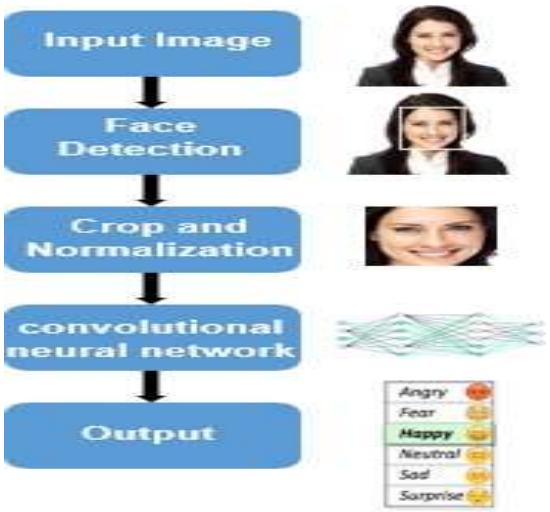

Figure:. The structure of our facial expression recognition system

Facial expression indicates human emotion in communication. Facial expression recognition, as the key technology of the emotional computing system, is not only used to human-computer interaction, but also extended to the field of interactive game platforms, safe driving, smart recommendation and auxiliary medical care, etc. It is showing a good application prospect in various fields.

The FER usually includes three steps: the face detection, the feature extraction and the expression classification. Among them, the feature extraction of facial expression plays a key role in the expression recognition system, which affects the recognition accuracy. Recently, the research on facial expression feature extraction is becoming a hot topic, and many methods have been proposed, such as Active Shape Model (ASM) , Active Appearance Models (AAM) , Local Binary Pattern (LBP) , Gabor Filter, principal component analysis (PCA), and Histograms of oriented gradients (HOG) [6], etc. All methods mentioned above need be operated by people to extract features, which leads to that some of the original features are lost. In the field of machine learning, data-driven feature learning algorithm has been proposed that is first to use deep learning to extract features. Unlike traditional machine learning algorithms, deep learning can automatically extract facial features without human participation. The essential features can be characterized autonomously from the sample data by the multilayered deep neural network. The performance of extracting features used deep learning method is better than that of traditional machine learning algorithms, and it is published by many literatures .

In 2006, Master of Machine Learning and Professor Hinton in University of Toronto in Canada first proposed the theory of deep learning, and published a paper with respect to using deep-structured neural network models to achieve dimensionality reduction in Science. Hinton believes that many hidden layers in artificial neural networks have excellent capabilities to learn features, which may be more conducive to visualization and classification. Hinton also thinks that the difficulty with respect to deep neural network training can be effectively overcome through layer-by-layer initialization. Deep learning is a class of methods, which is generally designed, for training deep-structured models. The deep structure model represents features with multi-layers. It has stronger characterization capabilities than that of the shallow structure model. Typical deep learning models include Convolutional Neural Networks (CNN) , Deep Belief Networks (DBN) , Stacked Auto-encoder (SAE), Recursive Neural Networks (RNN) and so on. CNN is a deep neural network containing input layer, convolutional layers, pooling layers, fully connected layer and output layer. It is inspired from brain neuroscience, and imitates the process to handle visual information with two types of neural cells, simple cells and complex cells, in the visual cortex. The CNN contains basic convolution operations and pooling operations. The convolution operation is to simulate simple cells and the pooling operation is to simulate complex cells. In CNN, the subblocks (local receptive areas) are the input with respect to the lowest layer of the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

hierarchical structure in the image, and the information is transmitted to different layers through layer by layer. The most significant feature of the observed data is obtained by a digital filter in every layer.

This paper presents a facial expression recognition method based on CNN. Firstly, face detection and preprocessing are performed on the facial expression image. After that, the features of the facial expression are extracted using some trainable convolution kernels. Then, the extracted face expression feature is reduced by the maximum pooling method. Finally, the Softmax classifier is used to categorize facial expression images, and facial expressions are classified into seven types of facial expression: happiness, surprise, sadness, anger, fear, disgust and neutrality.

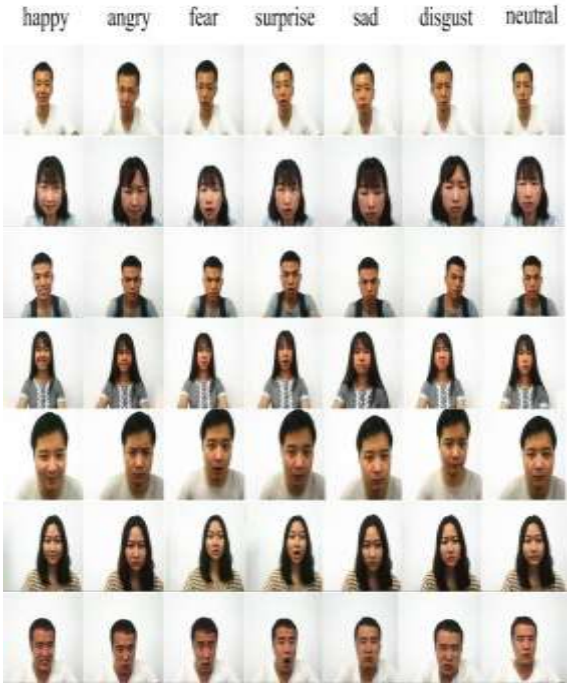

Figure . Samples of the seven basic expression images in our own facial expression database

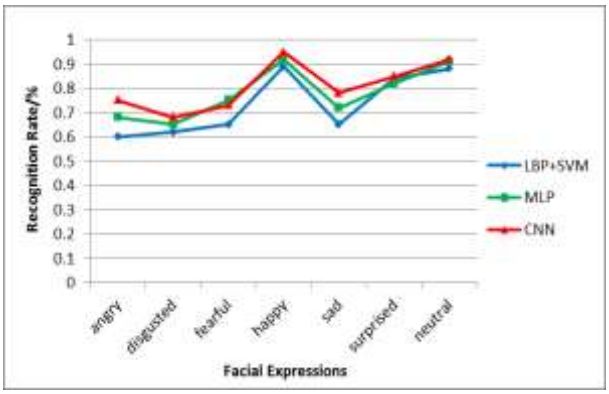

Figure . Recogniton rate of different facial expression recognitonmethods

The CNN structure is conductive to parallel learning with sharing weight of the connection neurons in the featuremap.

CNN has parallel data processing capabilities, which accelerates the training of CNNs and significantly reduces training time. It takes much less time with the GPU than that with CPU in training the CNN, because GPU can process the data in parallel. There is almost no difference in recognition rate between the two methods. Due to the network structure parameters, the training set and the test set are the same, the recognition rate is very close.

As social beings, human often socialize with verbal and non verbal in interacting with others. In non-verbal language it is indicated by facial expressions to interact [1]. Studies that have been conducted show that compared to vocal and verbal that can provide 45% of information, facial expressions can convey up to 55% of information in interacting [2].

Emotion is a state of feeling that affects behavior and makes changes in mind and physiology in humans. Emotions become one of the important aspects in life, stimulation of reactions to emotions can come from outside or from within

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

each individual. Emotions become important in life, especially in the relationship between humans as a form of attention, love and even malice, anger and other feelings [3]. Emotional aspects are managed by structures in the systemlimbic, which is facial recognition through facial expression, where the limbic system has a role to back up emotions [4]. According to Folkert the limbic system with Amygdala structure is one component of the limbic system for emotional stimulation known by changes in facial muscles with the formation of very responsive facial expressions . The study of facial expression has been developed since the Aristotelian era which studied the outer appearance of humans. Facial expressions as a form of gesture describe special expressions that have been tested by scientist Charles Darwin who say if facial cues are forms of universal emotion . According to Darwin, some of these signals are based on human expressive actions due to instinctive behavior rather than learning outcomes. Paul Ekman, an American psychologist identifies six categories of emotional classifications; happy, sad, surprised, angry, scared and disgusted . Emotional changes occur in the face, namely the forehead, eyebrows, eyes, nose, mouth, cheeks. Reported by expressions through facial expressions are used as a means of convincing someone, knowing the level of one's personality and being able to realize misunderstandings among individuals involved in the culture of communication. Through verbal communication one can lie, but through expressions it will be difficult to cover up his lies. A person will feel nervous when looking at his interlocutor if he is lying. Analysis of facial expression is related to the visual recognition of changes in the characteristics and movements of his face. The basic component in the analysis of facial expressions is face detection, feature extraction and facial expression recognition . Based on this research background, the researcher proposes to do machine learning focusing on knowing the accuracy of detection of 6 basic facial expressions and 1 neutral expression in real time with Android as a media capture automatically. This study used Viola Jones approach as face position detection and deep learning approach, Convolutional Neural Network, as an extractor of facial expression image features and the next step is to classify facial expressions with an output in the form of expressions predictions of happy, sad, surprised, disgusted, neutral, angry and afraid.

Some studies with CNN in recognizing faces produce quite good accuracy with reduced filter sizes. The development of convolutional neural networks to produce recognition of facial expressions with CNN, but still using datasets on image products . The research on facial expression is done by if Gabor's function influences the accuracy of the feature extraction of expression recognition, but the sad expression class is completely unrecognizable at all. In landmark face detection carried out the initial steps for the analysis of facial expressions and emotional recognition using Constrained Local Model (CLM) only limited to local detector optimization. Many attempts have been made to find efficient solutions improving the recognition accuracy and machine learning methods, CNN can improve expression recognition accuracy in biometric system . The detection of expression is also done in research by utilizing the expression of baby's image, but the classification with feature extraction from autoencoder is strongly influenced by the quality of face detection and landmark detection phases. Meanwhile, research utilize the pre-trained CNN models for recoginition three expressions (satisfied, neutral and disappointed) to find out what their experience with the restaurant concept. Another study is for expression detection by comparing the combination of Gabor features with Haar and Gabor features with Landmark features in the database with the best classification prediction results are happy class expressions but have not yet reached the diversity of models or variations of each expression.

Facial expression is the common signal for all humans to convey the mood. There are many attempts to make an automatic facial expression analysis tools as it has application in many fields such as robotics, medicine, driving assist systems, and lie detector. Since the twentieth century, Ekman et al. defined seven basic emotions, irrespective of culture in which a human grows with the seven expressions (anger, feared, happy, sad, contempt , disgust, and surprise). In a recent study on the facial recognition technology (FERET) dataset, Sajid et al. found out the impact of facial asymmetry as a marker of age estimation. Their finding states that right face asymmetry is better compared to the left face asymmetry. Face pose appearance is still a big issue with face detection. Ratyal et al. provided the solution for variability in facial pose appearance. They have used three-dimensional pose invariant approach using subject-specific descriptors . There are many issues like excessive makeup pose and expression which are solved using convolutional networks. Recently, researchers have made extraordinary accomplishment in facial expression detection, which led to improvements in neuroscience and cognitive science that drive the advancement of research, in the field of facial expression. Also, the development in computer vision and machine learning makes emotion identification much more accurate and accessible to the general population. As a result, facial expression recognition is growing rapidly as a sub-field of image processing. Some of the possible applications are human–computer interaction , psychiatric observations, drunk driver recognition, and the most important is lie detector.

Convolutional neural network (CNN) is the most popular way of analyzing images. CNN is different from a multi-layer perceptron (MLP) as they have hidden layers, called convolutional layers. The proposed method is based on a two-level CNN framework. The first level recommended is background removal, used to extract emotions from an image, as shown in Figure. Here, the conventional CNN network module is used to extract primary expressional vector (EV). The expressional vector (EV) is generated by tracking down relevant facial points of importance. EV is directly related to changes in expression. The EV is obtained using a basic perceptron unit applied on a background-removed face image. In the proposed FERC model, we also have a non-convolutional perceptron layer as the last stage. Each of the convolutional layers receives the input data (or image), transforms it, and then outputs it to the next level. This transformation is convolution operation, as shown in Fig. 2. All the convolutional layers used are capable of pattern

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

detection. Within each convolutional layer, four filters were used. The input image fed to the first-part CNN (used for background removal) generally consists of shapes, edges, textures, and objects along with the face. The edge detector, circle detector, and corner detector filters are used at the start of the convolutional layer 1. Once the face has been detected, the second-part CNN filter catches facial features, such as eyes, ears, lips, nose, and cheeks. The edge detection filters used in this layer are shown in Fig. 3a. The second-part CNN consists of layers with 3 × 3 kernel matrix, e.g., [0.25, 0.17, 0.9; 0.89, 0.36, 0.63; 0.7, 0.24, 0.82]. These numbers are selected between 0 and 1 initially. These numbers are optimized for EV detection, based on the ground truth we had, in the supervisory training dataset. Here, we used minimum error decoding to optimize filter values. Once the filter is tuned by supervisory learning, it is then applied to the background-removed face (i.e., on the output image of the first-part CNN), for detection of different facial parts (e.g., eye, lips. nose, ears, etc.) To generate the EV matrix, in all 24 various facial features are extracted. The EV feature vector is nothing but values of normalized Euclidian distance between each face part.

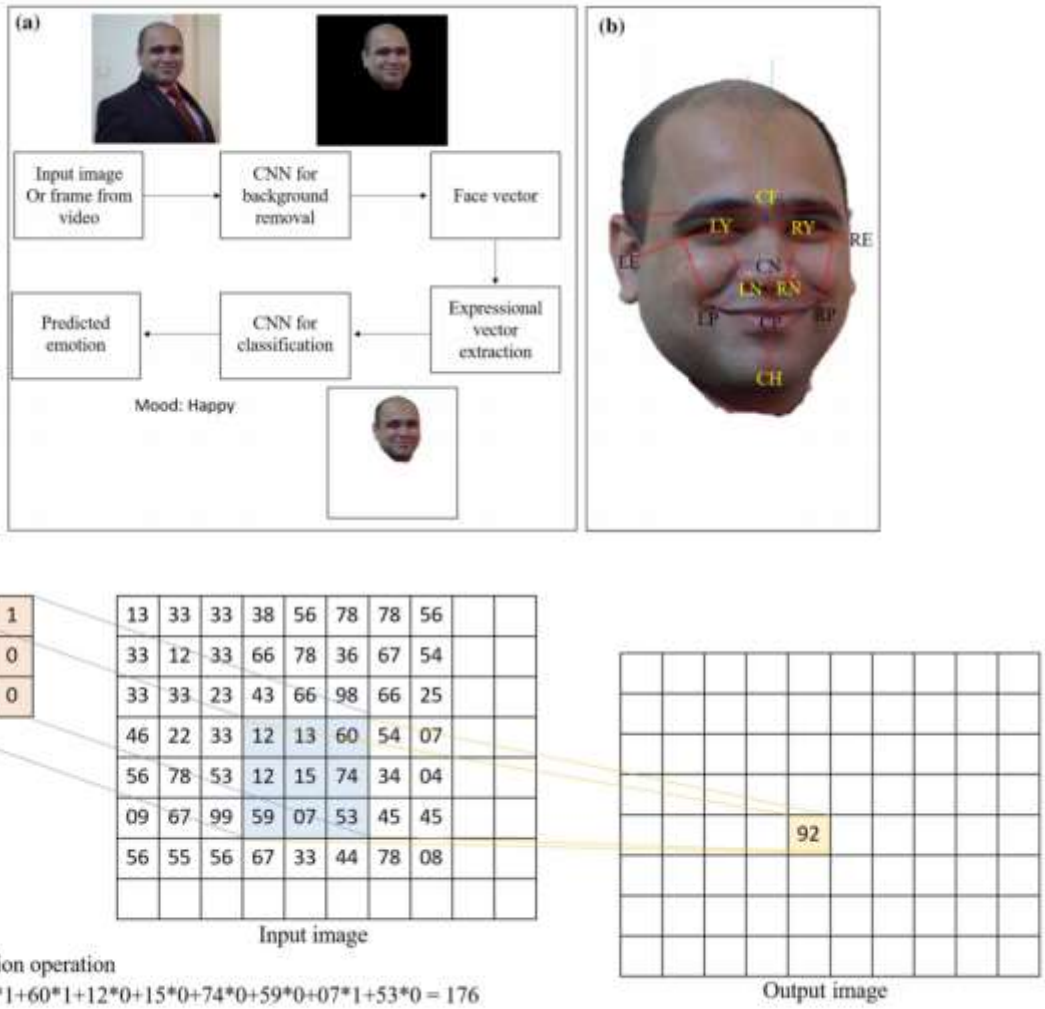

Fig. 2 Convolution filter operation with the 3 × 3 kernel. Each pixel from the input image and its eight neighboring pixels are multiplied with the corresponding value in the kernel matrix, and finally, all multiplied values are added together to achieve the final output value

FERC is a novel way of facial emotion detection that uses the advantages of CNN and supervised learning (feasible due to big data). The main advantage of the FERC algorithm is that it works with different orientations (less than 30°) due to the unique 24 digit long EV feature matrix. The background removal added a great advantage in accurately determining the emotions. FERC could be the starting step, for many of the emotion-based applications such as lie detector and also mood-based learning for students, etc.

III.DISCUSSION

The CNN designed in this paper has a good recognition performance in recognizing human face expression. The CNN model is a multilayer perceptron, which has the

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105167

characteristics of local perception, hierarchical structure, combination of feature extraction and classification. The expression recognition system based on CNN proposed in this paper can directly use the pixel value of the facial expression images as input. It obtain more abstract expression feature by learning the facial expression images autonomously. The process of complex artificial feature extraction in traditional facial expression recognition is avoided, and feature extraction and expression classification are performed simultaneously. Due to the connections between neurons are not fully connected, the same connections weights of neurons of feature map are used in one layer, and the complexity of the network model is reduced greatly. The pooling layer enhances the robustness of the CNN and can deal with some distortion of the image.
According to the experimental data and experimental conclusions, the next research work will further optimize the structure of CNN to improve the speed and accuracy of facial expression recognition.

IV. CONCLUSION

In this study, CNN based facial expression recognition method is designed with Kaggle Facial Expression Recognition Dataset and an accuracy of 61.8% is achieved in the classification of seven different categories. In order to have a proper dataset, preprocessing and data augmentation techniques are applied to the raw grayscale image data. Convolutional neural network with 20 layers are used and different types of facial expressions are classified.
In design process, we have faced different difficult problems and overcome each of them by applying required solution steps. Although the resolution of the sample images are quite low, from only 48×48 pixels, we managed to design a convolutional neural network which is capable of classifying seven different group effectively both offline testing and with real time data.

To further improve the performance, we will apply different feature extraction methods with more powerful machine and increase the size of the images; we will try to achieve higher accuracies at the end.

ACKNOWLEDGMENT

This paper and the research behind it would not have been possible without the exceptional support of my supervisor, Santhi P. I am also grateful for the insightful comments offered by my colleagues allowed me to continue my research with the book much longer than I could have hoped. Finally I would like to thank my college and university for the unconditional support for completing this work..

REFERENCES

[1] M. N. Dailey, G. W. Cottrell, C. Padgett, and R. Adolphs, “Empath: A neural network that categorizes facial expressions,” J. Cogn. Neurosci., vol. 14, no. 8, pp. 1158–1173, 2002.
[2] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biol. Cybern., vol. 36, no. 4, pp. 193–202, 1980.
[3] Y. LeCun, “Object Recongnition with Gradient-based Learning,” Shape, Contour Group. Comput. Vis., vol. 91, no. 0, pp. 399–404, 2017.
[4] P. Ekman and W. V Friesen, “Facial action coding system: A technique for the measurement of facial movement.,” CA Consult. Psychol. Press. Ellsworth, PC, Smith, CA (1988). From Apprais. to Emot. Differ. among unpleasant Feel. Motiv. Emot., vol. 12, pp. 271–302, 1978.
[5] H. Da Cunha Santiago, T. I. Ren, and G. D. C. Cavalcanti, “Facial expression Recognition based on Motion Estimation,” in Proceedings of the International Joint Conference on Neural Networks, 2016, vol. 2016-Octob, pp. 1617–1624.
[6] M. Singh, A. Majumder, and L. Behera, “Facial expressions recognition system using Bayesian inference,” in Proceedings of the International Joint Conference on Neural Networks, 2014.
[7] M. Schmidt, M. Schels, and F. Schwenker, “A Hidden Markov Model Based Approach for Facial Expression Recognition in Image Sequences,” in Artificial Neural Networks in Pattern Recognition, 2010, pp. 149–160.
[8] I. J. Goodfellow et al., “Challenges in representation learning: A report on three machine learning contests,” Neural Networks, vol. 64, pp. 59–63, 2015.
[9] R. Shindjalova, K. Prodanova, and V. Svechtarov, “Modeling data for tilted implants in grafted with bio-oss maxillary sinuses using logistic regression,” AIP Conf. Proc., vol. 1631, pp. 58–62, 2014
[10] A. Raghuvanshi and V. Choksi, “Facial Expression Recognition
[11] with Convolutional Neural Networks,” CS231n Course Proj., 2016.
[12] D. Lekhak, "A Facial Expression Recognition System Using
Convolutional Neural Network," Tribhuwan University, Institute of Engineering.,