

FACIAL EXPRESSION RECOGNITION –A REVIEW

¹Megha Chandran¹, Dr. Naveen. S.²

¹PG Student, Dept. of Electronics and Communication, LBS Institute of Technology for Women, Kerala, India

²Assistant Professor Dept. of Electronics and Communication, LBS Institute of Technology for Women, Kerala

Abstract - Advances in computer vision have created a great impact in the technological world. Facial expression is an important tool to communicate one's emotions non verbally Facial Expression Recognition also uses computer vision technology and this paper is a study on how this technology and algorithms are used in attaining the accurate result of emotion recognition. This is also an attempt of how deep learning can create great performance in emotion recognition.

automatic facial expressions recognition technology has also found new dimensions in the fields of animation film and synthesis of face images. Emotional expressions are very important in human communication. They mediate interaction between people, enhance and provide more clarification to the meaning instead of words or sentences.

1. INTRODUCTION

In human society, facial expressions play a significant role. Facial expressions contain lots of information that cannot be conveyed by voice in during communication between people. Facial expressions can be used not only to express our emotions but also to provide signs during social interactions such as level of interest, the desire to take a speaking turn and to provide continuous feedback on the understanding of the information conveyed so that it can help in obtaining the overall audience review in any interactive sessions. As a result of the potential applications in various field automatic facial expressions recognition has become a significant topic in computer vision research in near years . The automatic classification of facial expressions has got high significance in the field of human-computer interaction technology, and also has potential application value in many fields such as computer-aided training, medical field and distance education. Automated facial expression recognition finds application in areas like psychology to HCI (human computer interaction) to HRI (human-robot interaction). Psychological applications include autism early intervention techniques, etc. HRI and HCI make computers and robots behave more human-like. In automated learning, the computer identifies the cognitive state of the patient. After identifying the expressions the automated system can take a decision to manage the mood as if a patient is sad student is gloomy, it might tell a joke, or play a song. The application of automatic facial expressions recognition technology can be used in the fields of animation film and synthesis of face image. Facial expressions serves very important role in human communication. They improve interaction between people, enrich and give more clarification to the meaning of words or sentences. The traditional electronic intelligent teaching has the drawback that the interaction between students and teachers about teaching feedback can only be delivered via voice information. Emotional calculation based on facial expressions can be used to measure the feedback of the traditional electronic teaching. In the medical field, the automatic recognition of facial expressions can help the doctor know the patient's feelings in time The application of

1.1 Deep learning

Deep learning falls in the family of artificial neural networks which is the main component of machine learning. Convolutional neural networks have gained popularity in solving any image data challenge. It is a type of ANN which accepts images as inputs They allows for modifications to the architecture as there are sparse connection between layers. Weights are shared between these layers. They can be used in video analytics and image processing application. Data that has spatial relationships is appropriate for applying to CNN.

CNNs contain convolutional layers. Spatial relationship is preserved in this network. Parameter sharing is the main advantage of CNNs.

Facial expressions recognition (FER) systems consist of face detection, facial feature extraction and expression classification. American psychologists Ekman and Friesen defined six basic human facial expressions, like happy, angry, surprise, disgust, etc and developed facial action coding system (FACS) based on action units (AUs) to describe facial expressions. Researches on automatic facial expression recognition systems are based on methods such as principal component analysis (PCA) , linear discriminant analysis (LDA) , elastic bunch graph matching (EBGM) , independent component analysis (ICA) , two-dimensional principal component analysis (2D-PCA) , artificial neural networks ANNs, embedded hidden Markov models (EHMM) , Gabor wavelets etc. Even though these algorithms could improve the accuracy and speed of automatic facial expressions recognition, they are also affected by changes in light, pose, aging, alignment. In the process of facial expression recognition, the facial expression features are extracted from the facial images and then processed. There methods are based one two main approaches: the overall template matching system and the system based on geometric features. In the overall system, the template can be a pixel or a vector. In the geometric feature system, the principal component analysis and the multi-layer neural network or convolutional neural networks are widely used to obtain the feature representation of the face, and to detect the main geometrical feature points in the main part of the face like distance between two eyes in the image data. Feature-based

approaches are computationally larger than template-based methods, but are not sensitive to scale, size, head orientation and face position.

Challenges in facial expression recognition include the existence of partially occluded faces. Occlusions changes can occur in the occludes and their positions. The occlusions includes due to hair, glasses, scarf, breathing mask, hands, arms, food, and other objects that come across the face. They may block the eye, mouth, nose, and other part of the face. The variability of occlusions in data may result in the reduction of recognition of accuracy. This issue of occlusion, is tackled by Yong Li, Jiabei Zeng, Shiguang Shan, Xilin Chen, [1] by proposing a Convolutional Neural Network with attention mechanism (ACNN), in a way to mimic how human recognize the facial expression. People recognizes the facial expressions based on certain patches of the face. When some regions of the face are blocked (e.g., the lower left cheek), they may judge the expression according to the symmetric part of face (e.g., the lower right cheek), or other highly related facial regions (e.g., regions around the eyes or mouth). Getting inspiration from intuition, ACNN automatically perceives the blocked facial patches and pays attention mainly to the unblocked and informative patches. ACNN has the ability to focus on distinctive as well as unobstructed regions in facial image.

Automatic facial expression recognition (FER), aims to analyse and understand human facial behaviour. FER methods can be divided into two categories: video sequence-based methods (dynamic) and image-based methods (static). Due to the strong feature learning ability of deep neural networks, deep neural network based FER method in video sequences is achieved by Shiqing Zhang, Xianzhang Pan, Yueli Cui, Xiaoming Zhao, and Limei Liu [2] using a hybrid deep learning model. This hybrid deep learning model contains three deep models. The first two deep models are deep Convolutional Neural Networks (CNNs), a spatial CNN network processing static facial images and a temporal CNN network processing optical flow images. These two CNNs are separately used to learn high-level spatial features and temporal features on the divided video segments. The third deep model is a deep fusion network built with a Deep Belief Network (DBN) model, which is trained to jointly learn a discriminative spatio-temporal segment-level feature representation. They adopted a linear Support Vector Machine (SVM) to perform facial expression classification tasks in video sequences. A linear Support Vector Machine (SVM) was used to perform facial expression classification tasks in video sequences. Chao Qi, Min Li, Qiushi Wang, Huiquan Zhang, Jinling Xing, Zhifan Gao, Huailing Zhang [3] extracted the facial contours by seeking the largest island method based on local binary pattern (LBP) and they segmented the facial regions by establishing pseudo3D model in which the gray values of the images are kept as its z axis. Multi-Dimensional Scaling (MDS) and the approximation of Earth Mover's Distance (EMD) is used to reduce the dimension of data and used to train the convolutional neural network model to predict the

expressions, and also choose the different feature regions or combinations as the reference to find out the best distinguished region or combination and compare with the conclusions in cognitive neurology. They also compared traditional basic emotion model and dimension space model. The face detection is done by finding the specific location of the face. As the complex background of the image affects the accuracy of algorithms, in automatic expression recognition human facial contour region extraction is particularly important. The selected LBP operator to describe the change of the local region texture, to extract the texture information from the local region of the gray image as it has been used in texture classification, image retrieval, face recognition and other fields and applications. According to Cognitive psychology studies, for human beings, the attention of the various facial areas is not the same, in the process of emotional recognition especially when considering eyes and mouth areas. The face was divided into six regions that reflect the commonality and characteristics of the human face. The six regions chosen were forehead, eyes, nose, mouth, left cheek and right cheek. Even though LBP operator is robust to monotonic gray level changes and computational efficiency, they possess some limitations. They are is sensitive to the noise as only the symbol characteristics difference between the central pixel and the neighborhood pixel is considered while computing. LBP inevitably loses some information due to these reasons. Since the number of patterns is increasing with the number of sampling points, too many patterns are not conducive to identify and access textures. In a high dimensional space, transformation approach is used to transform to dimensions to a low dimensional subspace. Isometric Mapping (Isomap) is a classic algorithm used.

In the paper proposed by Lucy Nwosu¹, Hui Wang, Jiang Lu¹, Ishaq Unwala¹, Xiaokun Yang¹ and Ting Zhang [4], a simple CNN architecture is used to predict basic expressions from facial parts. Viola Jones method for face and facial regions detection/extraction is combined with convolutional neural network for feature extraction and classification. Facial parts are used as input to reduce the size of input given to CNN and reached a conclusion that FER processing time when compared to other CNN architectures that used the whole face for facial expression recognition is less. For this they proposed two-channel CNN architecture for feature extraction and expression recognition.

A new architecture with Multi-channel Convolutional Neural Network (MCCNN) is proposed for recognizing facial expressions. An unsupervised training is done on a single channel as a Convolutional Auto encoder (CAE). A standard CNN is also used. Information from both channels is converged in a fully connected layer and is then used for classification. Hongying Meng, Nadia Bianchi-Berthouze, Yangdong Deng, Jinkuang [5] tried to model the dynamics that characterize naturalistic expressions by working on a novel two-stage automatic system to continuously predict affective dimension values from facial expression videos. This was attained with two stages where in the first stage,

traditional regression methods used to classify each individual video frame, and in the second stage, a time-delay neural network (TDNN) to model the temporal relationships between consecutive predictions. The two-stage approach separated the emotional state dynamics modeling from an individual emotional state prediction step based on input features. Their results showed that the temporal information used by the TDNN is not biased by the high variability between features of consecutive frames and allows the network to more easily exploit the slow changing dynamics between emotional states so that it can be used for continuous dimensional emotion prediction from facial expression image sequences. TDNN is a neural network model with the capability of capturing the dynamic relationship between consecutive observations. The instant of an emotional expression (e.g., a video frame) is classified by taking into account not only the input features describing that instant, but also the input features describing the previous instants, i.e., how the expression evolved over time to the current state. The number of past instants considered, as the delay is set as a parameter of the network. The advantages they quoted were, dealing of regression problems instead of categorizations due to its nature, and secondly the usage of past knowledge gathered in real time, rather than analysing the full sequence.

2. CONCLUSION

This paper highlights the application of deep learning in automatic facial expression recognition systems. Comparisons of different works on facial expression recognition systems and relevance are reviewed.

REFERENCES

- [1] Zhang, Shiqing, et al. "Learning affective video features for facial expression recognition via hybrid deep learning." *IEEE Access* 7 (2019): 32297-32304.
- [2] Li, Yong, et al. "Occlusion aware facial expression recognition using CNN with attention mechanism." *IEEE Transactions on Image Processing* 28.5 (2018): 2439-2450.
- [3] Qi, Chao, et al. "Facial expressions recognition based on cognition and mapped binary patterns." *IEEE Access* 6 (2018): 18795-18803.
- [4] Nwosu, Lucy, et al. "Deep convolutional neural network for facial expression recognition using facial parts." 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2017.
- [5] Hamester, Dennis, Pablo Barros, and Stefan Wermter. "Face expression recognition with a 2-channel convolutional neural network." 2015 international joint conference on neural networks (IJCNN). IEEE, 2015.
- [6] Meng, Hongying, et al. "Time-delay neural network for continuous emotional dimension prediction from facial expression sequences." *IEEE transactions on cybernetics* 46.4 (2015): 916-929.
- [7] Aliaa A A Youssif "Automatic Facial Expression Recognition system based on geometric and appearance features" *Computer and information Science*, Vol. 4, No. 2, March 2011.