REALTIME EMOTION DETECTION USING KERAS

Ponvizhi P[1], Vijay Sai R[2]

[1]Student ,K.S.Rangasamy college of technology,Tiruchengode,Tamilnadu,India

[2]Assistant Professor ,K.S.Rangasamy college of technology,Tiruchengode,Tamilnadu,India

ABSTRACT

In many automated device applications, such as robotics education, artificial intelligence, and Defence, recognition of facial expressions plays a major role. It is difficult to precisely recognize facial expressions. Approaches to solve the problem of FER (Facial Expression Recognition) can be divided into 1) single static images and 2) image sequences. Similar methods, historically, Researchers used the Multi-layer Perceptron Model, k-Nearest Neighbours, Help Vector Machines to solve FER. Features such as Local Binary Patterns, Eigenfaces, Face-landmark features, and Texture features were derived from these methods. Among all these strategies, Neural Networks have gained a lot of popularity and are commonly used for Oh. FER. Due to their casual architecture and ability to provide good results without the need for manual feature extraction from raw image data, CNNs (Convolutionary Neural Networks) have recently gained popularity in the field of deep learning. This paper focuses on a study of different CNN-based facial expression recognition techniques. It involves state-of-the-art techniques proposed by various researchers. The paper also illustrates the steps needed for FER to use CNN. This paper also provides an overview of methods focused on CNN and problems that need focus when selecting CNN to solve FER.

Keywords- convolutional neural network; understanding of facial expression; deep learning

1. INTRODUCTION

Facial expression is one of the methods of non-verbal communication between individuals and plays an important role in conveying non-verbal signals as a communicative source. These signs may understand a person's mood/mental state and communicate a weighted means of communication and spoken words. Understanding actions through the identification of facial expression (FER) is a natural way that plays a vital role in meaningful communication and social interaction. Awareness of human behavior has many applications in different areas of everyday life, such as entertainment, protection, and healthcare. It can be used for patient monitoring in hospitals, for suspicious person identification in video surveillance, During vehicle driving and video games, lie identification in the investigation and mood classification. FER robots have recently been equipped to evaluate the mood of workers and consumers in large-scale businesses, making contact between humans and computers more feasible. Researchers have presented various FER techniques with considerable precision, but identifying facial expression of faces captured from different angles and from different nationalities is still a daunting task. One of the beneficial sources for the study of human attitude and actions is facial expression. Psychological research has shown that facial expression characteristics are located around the mouth, nose, and Eyes that are important to FER. Several methods for FER in still images and video data were suggested in the literature where different classifiers were used, including SVM, KNN, decision tree, neural networks, Bayesian network, and rule-based classifiers. The key objective of the approaches described above is to provide a trade-off between processing time and efficient accuracy of recognition. Pooling max. The mainstream methods discussed above are limited to frontal facial images only and do not work precisely for video data where faces appear with varying lighting conditions at different angles. The proposed framework addresses these issues with the following main contributions:

We used the Viola-jones algorithm for face detection in this study. In each video frame, face detection and recognition of multiple faces are computationally costly. We therefore used an effective KLT tracking algorithm, which can track the recognized face in the video accurately.

Owing to the similar properties of color and shapes of all human faces, extraction of visual features for facial recognition is very difficult. The faces of individuals are depicted in our technique using HOG

characteristics, which can accurately capture information for robust facial recognition from the facial directions, edges, and intensities fed to the SVM classifier.

The mainstream CNN models have high precision, but they are not effective to function on low processing equipment due to the large number of layers. Also, it takes a large amount of data to train a precise CNN model. We proposed an efficient CNN model for FER to solve these challenges, which includes only six convolutionary layers and applied comprehensive data augmentation techniques to the facial expression dataset to resolve the problem of lack of data, the presence of faces from different views, and video data lightening conditions.

The goal of the proposed framework is to examine the conduct of a person based on recognition of facial expression. We then carried out a rationale, Therefore, in a video for behavioral comprehension, we carried out a reasoning process on the statistical frequency matrisx of the facial expression of the person.

2. LITERATURE REVIEW

We review and address the state-of-the-art approaches to micro-expression recognition briefly in this chapter. There are two sub-sections to this literature review: we first begin our discussions on hand-designed methods, followed by the latest methods focused on learning.

A function descriptor has been proposed by Lu et al. to uniquely define the term by fusing histograms of the motion boundary. By fusing the differentials of horizontal and vertical components of optical flow, the feature descriptor is created. For the classification of micro-expressions, the generated feature vector is fed into SVM. In a video using spatio-temporal pressure on the face caused by non-rigid movements, Shreve et al. suggested a coherent approach to recognize both macro and micro facial expressions. To differentiate between the macro and micro gestures, they have measured the strain magnitude from different facial regions such as chin, jaw, cheek, forehead, etc. Pfister et al. have used the spatio-temporal local texture descriptors and various classifiers to recognize the micro-expressions. They proposed a temporal interpolation model to intercept the problem of variable video lengths. Zhao et al. have proposed the Local Binary Pattern - Three Orthogonal Planes (LBP-TOP) feature for facial expression recognition using dynamic texture to exploit the spatio-temporal information in very compact form. Wang et al. have proposed Local Binary Pattern with Six Intersection Points to tackle the problem of redundancy from the LBP-TOP. They fed the extracted characteristics for facial expression recognition into the SVM classifier. In reality, the LBP-TOP lacks adequate features since three orthogonal frames reflect video. The spatio-temporal completed local quantized patterns that measure the vector quantization and use the codebook to learn more complex patterns have been suggested by Huang et al. The optical flow field computed over various sub-regions of the face image has recently been used by Liu et al. to identify facial microexpression. They also computed the directional mean optical flow function to minimize the dimensionality of the feature vector. They fed into the SVM classifier the extracted characteristics for facial expression recognition.

In fact, because three orthogonal frames mirror video, the LBP-TOP lacks adequate features. Huang et al. have proposed the spatio-temporal completed local quantized patterns that calculate the vector quantization and use the codebook to learn more complex patterns. Liu et al. have recently used the optical flow field computed over different sub-regions of the face image to define facial micro expression. To minimize the dimensionality of the feature vector, they also computed the directional mean optical flow function. They fed the extracted characteristics for facial expression recognition into the SVM classifier. In fact, the LBP-TOP lacks adequate features because three orthogonal frames mirror video. The spatio-temporal completed local quantized patterns that measure the vector quantization and use the codebook to learn more complex patterns have been suggested by Huang et al. The optical flow field computed over various sub-regions of the face image has recently been used by Liu et al. to describe facial microexpression. They also computed the directional mean optical flow function to minimize the dimensionality of the feature vector. The Convolutionary Neural Network (CNN) is the new deep learning trend to solve vision problems where the input is pictures or images. A method for classifying micro-expression by extracting the optical flow attribute from the reference frame of a micro-expression video has been suggested by Liong et al. Then, for expression classification, the extracted optical flow features are fed into a 2D CNN model. In order to produce the synthetic images used to train the deep convolutionary neural network (DCNN) to identify micro-expressions, Takalkar et al. have used data augmentation techniques. The Convolutionary Neural Network (CNN) is the latest trend in deep learning to solve vision issues where pictures or images are the input. Liong et al. have proposed a method for classifying micro-expression by extracting the optical flow attribute from the reference frame of a micro-expression film. Then, the extracted optical flow features are fed into a 2D CNN model for expression classification. Takalkar et al. have used data augmentation techniques in order to generate the synthetic images used to train the deep convolutionary neural network (DCNN) to classify micro-expressions. Peng et al. fine tuned ImageNet's pre-trained CNN over micro and macro-expression recognition facial expression datasets. For video-based micro-

expression recognition, Li et al. have implemented a 3D flow-based convolutional neural network model. Using this network for minute facial gestures.

3. Proposed methodology

Peng et al. fine tuned the pre-trained CNN of ImageNet over facial expression recognition datasets of micro and macro-expression. Li et al. have introduced a 3D flow-based convolutional neural network model for video-based micro-expression recognition. For minute facial expressions, use this network. Secondly, using the SVM classifier (if registered), the detected face FD is recognized, otherwise we must first register the face to our DF database. Using the proposed CNN model, we find the facial expression FER after identification of the detected face. Finally, using his facial expression FER for the entire video V, we measure the action statistics FS of the known face FR.

3.1. Face detection and tracking

Face detection is a key step in the suggested idea and, depending on this step, there are two other processes. We finally selected the Viola-Jones algorithm, which works perfectly for face detection, after investigating several face detection algorithms based on color, motion and other distinct features. The Voila-Jones algorithm consists of four phases: selection of hair features, development of an integral image, Adaboost training, and classifier cascading. Some common features that are extracted are included in all human faces, and the most significant features are selected in the first stage using the collection of hair features. In the proposed idea, face detection is a main step and, depending on this step, there are two other procedures. After researching multiple face detection algorithms based on color, motion and other distinct features, we finally selected the Viola-Jones algorithm, which works perfectly for face detection. The Voila-Jones algorithm consists of four phases: hair feature selection, integral image creation, Adaboost training, and cascading of classifiers. In all human faces, some common features that are extracted are included, and the most important features are selected using the selection of hair features in the first level. Face detection is a main step in the proposed idea and, depending on this step, there are two other procedures. We finally selected the Viola-Jones algorithm, which works perfectly for face detection, after studying several face detection algorithms based on color, motion and other distinct characteristics. The Voila-Jones algorithm consists of four phases: selection of hair features, development of integral images, Adaboost training, and classifier cascading. Some common features that are extracted are included in all human faces, and the most important features are selected using the first level's collection of hair characteristics.

3.2. Face registration and recognition

3.2.1 Feature extraction

The extraction of the HOG feature was suggested by Dallas et al. Compared to other existing features, it has demonstrated state-of-the-art efficiency. The HOG, like SIFT and SURF, is a regular global descriptor and all these descriptors are used for object representation. HOG collects data from each pixel's face directions, edges, and visibility. Used to measure horizontal and vertical gradients. In bins, each cell retains knowledge about edge orientations. In bins, each cell retains knowledge about edge orientations. A number of gradient orientations are expressed by the bin.

3.2.2 Facial Recognition

Via the maximized margin principle, SVM discovers the optimal dividing hyperplane between two groups. It is essentially a form of binary classification; therefore, for multi-class classification, we used its one-versusrest (1VR) version. N different binary classifiers for N number of categories are constructed by the 1VR process. The kth binary classifier is trained as positive samples using the training sample from the kth class and the remaining N-1 group training data as negative samples. When checking the classifier, the binary classifier that gives the highest output value is predicted by the mark. The benefit of the SVM approach is that hyperplane modeling only deals with support vectors rather than the entire training dataset, so the size of the training dataset is typically not a concern. We collected twenty images with different angles and illuminations for the registration of a new face, as provided. HOG face characteristics are extracted from twenty photos of each person and are used for 1VR SVM classifier training while we used the actual videos of TV drama series for testing.

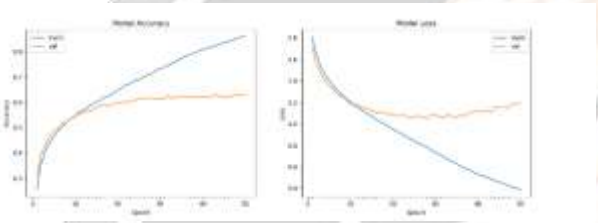
4.1. Data augmentation

In the suggested context, data augmentation techniques are used to make the KDEF dataset rotational, size, and noise invariant for such complex scenarios. It is a common fact that on a large amount of training data, deep learning techniques have successful results. We used the KDEF FER dataset to train our CNN model, but this dataset contains far less data that motivates us to incorporate approaches to data augmentation to make the

dataset adequate for CNN model training. "We used an open source library known as "Augmenter" for image augmentation. It offers a vast range of services that are supplemented in various automated ways. This library aims to make augmentation less error-prone, more reproducible, effective, and easily functional for classification tasks. Rotation of images at different angles, flipping of images, Gaussian blur, sharpening, embossing and skewing of images with various parameters are the various techniques used in our augmentation framework.

4.2 The proposed CNN architecture

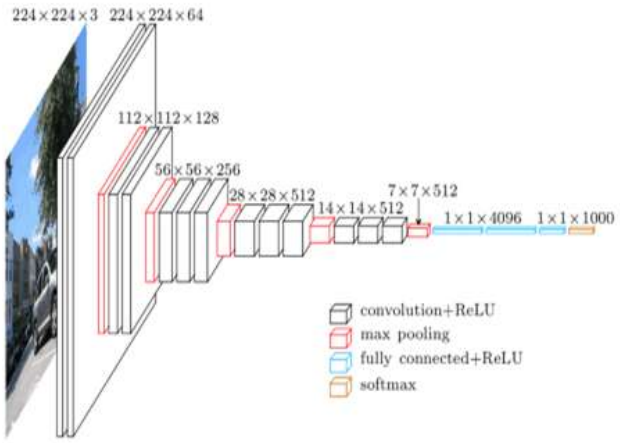
Since 2012, CNN has demonstrated the capacity to identify images on a large-scale ImageNet dataset. Several CNN image classification architectures and many other areas of image and video analytics have been developed by researchers, such as video summary, image retrieval, and action recognition. Thus, inspired by these findings, we used CNNs for the difficult task of emotion recognition in our proposed system. The architecture proposed follows the structure of the standard AlexNet model, but its input and weight sizes vary. CNN has demonstrated the ability to recognize pictures on a large-scale ImageNet dataset since 2012. Researchers have developed many CNN image classification architectures and several other fields of image and video analytics, such as video summary, image retrieval, and identification of actions. Thus, motivated by these results, we used CNNs in our proposed method for the difficult task of emotion recognition. The proposed architecture follows the typical AlexNet model's structure, but its input and weight sizes differ. With 128 kernels of size 3 ?? 3 ?? 32, Conv2 takes the output of pool1 as input and filters it. The pool2 layer reduces the processed layer size to 128 x 16 x 16. Without overriding the pooling layers, the conv3, conv4, and conv5 layers are connected with each other. In a map of 1 ?? 1 ?? 7 features, the conv6 with special 1 ?? 1 filters followed by average pooling collects the facial information. For emotion classification, the AVG pool output is transferred to the SoftMax classifier.



The test accuracy reached 63.2% in 25 epochs.

5. Experimental evaluation

The proposed system is tested experimentally in this section using subjective judgment and quantitative data. Using the KDEF dataset, the facial expression model of the proposed system is trained and evaluated and the face registration model is validated with our own generated dataset. The experimental environment consists of a Caffe system for CNN training on the setup of the Nvidia-DIGITS frontend with the GTX270 GPU. In MATLAB, the testing of the overall concept is carried out. In addition, by performing a survey on individuals of different ages via a questionnaire, the subjective study was carried out while quantitative results were determined using the uncertainty matrix and overall accuracy.



CNN architecture

5.1 KDEF dataset

One of the most daunting facial expression image sets is the KDEF dataset. For human facial emotions, it contains a net number of 4900 images. There are a total of 70 subjects in this dataset, 35 gathered from females and 35 from males, with participants between the ages of 20 and 30. The dataset is more complex since each expression is taken from five different angles, which makes it difficult from a different point of view to identify visual facial expression patterns. Some sample images from the KDEF dataset are illustrated. Only face regions are trained in the proposed CNN model because the overall concept relies on the face area alone. Therefore, with the Viola-jones algorithm, we automatically cut faces from all database images for training. We selected the KDEF dataset because its selection is adequate to train a deep learning model for each category and its data is also well suited to the proposed idea because of the variance in the angle of capture of the images. As with data from TV shows, in various situations from different backgrounds, the actors have different emotions. In comparison to this, only front-view images consist of other data sets for facial expression analysis. In the suggested CNN model, only face regions are trained because the overall definition relies on the face area alone. Therefore, for training, we automatically cut faces from all database images with the Viola-jones algorithm. We chose the KDEF dataset because its collection is adequate to train a deep learning model for each group, and because of the variance in the angle of capture of the images, its data is also well suited to the proposed concept. As with data from TV shows, the actors have different feelings in different circumstances from diverse backgrounds. In addition to this, other data sets for facial expression analysis consist of only front-view images. Only face regions are trained in the suggested CNN model because the overall description relies on the face area alone. Hence, with the Viola-jones algorithm, we automatically cut faces from all database images for training. We chose the KDEF dataset because its selection is appropriate for each group to train a deep learning model, and its data is also well suited to the proposed definition due to the variance in the angle of capture of the images. As with data from TV shows, in various situations from varied backgrounds, the actors have different feelings. In addition to this, only front-view images consist of other data sets for facial expression analysis. Overall, it is clear that the proposed CNN increases efficiency with a KDEF dataset of 5.19 percent. The Ada Boost, Deep Conv Auto Encoder, and TLCNN are, however, very close behind. The other two well-known CNN models, VGG-16 and Resnet-50, offer far less precision with 73.6 percent and 76.0 percent respectively, using transfer learning on KDEF datasets. Low level characteristics, on the other hand, i.e. 79 percent precision has been achieved with SVM classifier SURF features.

5.2. Our own dataset formation

Overall, with a KDEF dataset of 5.19 percent, it is evident from Table 3 that the proposed CNN improves performance. However, very close behind are the Ada Boost, the Deep Conv Auto Encoder, and TLCNN. With 73.6 percent and 76.0 percent, respectively, the other two well-known CNN models, VGG-16 and Resnet-50, give much less accuracy using transfer learning on KDEF datasets. Low level features, on the other hand, have been achieved with SVM classifier SURF characteristics, i.e. 79 percent accuracy. "Extras" is a British humorous TV series made, written and directed by Ricky and Stephen and co-produced by the BBC and HBO. These series are very long and have so many episodes in each season as well, so only the representative and interesting ten episodes with salient events and moments from both TV series were selected by the proposed structure. With a low resolution of 360 ?? 480 and 30 frames/s, they are downloaded in mp4 format.

CONCLUSION

Face recognition is still a challenging problem in the field of computer vision. It has received a great deal of attention over the past years because of its several applications in various domains. Although there is strong research effort in this area, face recognition systems are far from ideal to perform adequately in all situations form real world. Paper presented a brief survey of issues methods and applications in area of face recognition. There is much work to be done in order to realise methods that reflect how humans recognise faces and optimally make use of the temporal evolution of the appearance of the face for recognition.

This Concept provided a proposed model to solve the problems of emotion recognition based on facial recognition in virtual learning environments, and the efficiency and accuracy are considered at the same time. Using HAAR Cascades to detect eyes and mouth and identify all kinds of emotion through the neural network method, the combination of efficiency and accuracy is achieved. It can be applied to real distance education. The application of emotion recognition in virtual learning environments is a much-researched topic. In addition to the change of uncertainty factors makes teachers and students face pattern is more complex, so the emotion recognition in the online learning network application mode is a very challenging topic.

REFERENCES

[1]. Bargh, J. A. (2010), "The four horsemen of automaticity": Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer Jr., T. K. Srull (Eds.), Handbook of social cognition (2nd ed., pp. 1– 40). Hillsdale, NJ: Erlbaum.

[2]. B. Abboud, F. Davoine, and M. Dang(2004), "Facial expression recognition and synthesis based on an appearance model," Signal Process.: Image Commun., vol. 19, no. 8, pp. 723–740.

[3]. Bo Wu, Haizhou Ai, Chang Huang(2013), "LUT-Based Adaboost for Gender Classification", in Int'l Conf. on Audio- and VideoBased Biometric Person Authentication, Guildford, UK, pp. 104-110.

[4]. De Houwer, J., Moors, A. (2012)," How to define and examine implicit processes". In R. Proctor,J. Capaldi (Eds.), "Implicit and explicit processes in the psychology of science "(pp. 183–198). New York:

[5]. G. Zhao,M. Pietikainen,(2017), "Dynamic Texture Recognition using Local Binary Patterns with an Application to Facial Expressions," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915-928.

[6]. G. Shakhnarovich, P. Viola and B. Moghaddam(2017), "A Unified Learning Framework for Real Time Face Detection and Classification", in IEEE Proc. Int'l Conf. on FG, Grenoble, France, pp. 14-21.

[7]. Kahneman, D., Treisman, A. (2010), "Changing views of attention and automaticity". In R. Parasuraman , R. Davies (Eds.), Varieties of attention (pp. 29 – 61). New York: Academic Press.

[8]. Klinnert, M. D., Campos, J. J., Sorce, J. F., Emde, R. N., & Svejda, M. (2015),"Emotions as behavior regulators". In R. Plutchik , H. Kellerman(Eds.), Emotions: Theory, research, and experience (Vol. 2, pp. 57–86). New York: Academic Press.

[9]. L. Ma and K. Khorasani(2019), "Facial expression recognition using constructive feedforward neural networks," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 3, pp. 1588–1595.

[10]. Maja Pantic, and Leon J.M. Rothkrantz,(2017), "Automatic Analysis of Facial Expressions: The State of Art", IEEE Trans. on PAMI, VOL. 22, NO. 12, pp. 1424-1445.

[11]. M.J. Lyons, J.Budynek, and S. Akamatsu(2012), "Automatic Classification of Single Facial Images", IEEE Trans. on PAMI, VOL. 21, NO. 12, pp.1357-1362.

[12]. M. Pantic and L. J. M. Rothkrantz,(2013), "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445.

[13]. P. Viola and M. Jones(2014), "Rapid Object Detection using a Boosted Cascade of Simple Features", in IEEE Proc. Int'l Conf. on CVPR, Hawaii, USA, pp. 511-518.

[14]. R. E. Schapire and Y. Singer(2009), "Improved boosting algorithms using confidence-rated predictions", Machine Learning, VOL. 37, NO. 3, pp. 297-336.

[15]. Shan, C., Gong, S., McOwan, P. W. (2015), "Robust facial expression recognition using local binary patterns". IEEE International Conference on (Vol. 2, pp. II-370).

[16]. S. Akamatsu, M. Kamachi, et al.(2005), "Coding Facial Expressions with Gabor Wavelets", in IEEE Proc. Int'l Conf. on FG, Nara, pp. 200-205.

[17]. T. Kanade, J. F. Cohn, and Y. L. Tian(2010), "A comprehensive database for facial expression analysis", in IEEE Proc. Int'l Conf. on FG, Grenoble, France, pp 46-53.

[18]. T. Otsuka, and J. Ohya,(2018), "Spotting Segments Displaying Facial Expressions from Image Sequences Using HMM", in IEEE Proc. Int'l Conf. on FG, Nara, Japan, pp. 442-447.

[19]. Tong WANG, Haizhou AI, Gaofeng HUANG(2015), "A TwoStage Approach to Automatic Face Alignment, in Proc". SPIE International Symposium on Multispectral Image Processing and Pattern Recognition, Beijing, China, pp. 558-563.

[20]. Y. Gao, M. Leung, S. Hui, and M. Tananda(2003), "Facial expression recognition from line-based caricatures," IEEE Trans. Syst., Man, Cybern. A: Syst. Humans, vol. 33, no. 3, pp. 407–412.