THINK INDIA (Quarterly Journal)

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019

Effective Classroom Monitoring by Facial Expression Recognition and Ensemble Learning

THINK INDIA (Quarterly Journal)

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019

Dr. Seema Kolkur, SourishVaghulade,GauravTejwani,YashVazirani
Dept of Comp Engg,Thadomal Shahani Engineering College,Mumbai, India
kolkur.seema@gmail.com,sourishvaghulade@gmail.com,gaurav.tejwani.97@gmail.com,vaziraniyash98@gmail.com

Abstract—Face Detection and Recognition of facial expressions is crucial to many applications in real time. A great deal of improvement is made in these areas due to recent concepts of deep learning and abundant datasets available online. The purpose of this paper is to propose a deep learning model along with ensemble learning to further improve the results of face detection and recognition system .Mere presence of a student in a classroom does not merit his/her attention towards the material taught in it. So, the trained model is further extended to a full-fledged classroom monitoring system which can help both students and teachers for effective teaching learning process. The system can recognize students facial expression and based on that can determine how long each student was focusing in class. Also the system can automatically mark a student present in a class. Our system delivers promising results.

Keywords—Deep Learning, Ensemble Learning, Expression Analysis, Face Recognition,

I. Introduction

Facial Expression Recognition (FER) is an important process for humans to facilitate understanding, communication and analysis. It consists of locating faces in the scene (object recognition), extracting useful features from the detected face regions, analyzing and classifying the features into different categories [1]. There are numerous applications of expression recognition in Medicine, E-learning, Monitoring, Entertainment, Law, Marketing, etc [2]. Today Deep Learning is widely used for image classification problems. Convolutional Neural Network (CNN) and transfer learning techniques together have delivered fantastic results in many applications. We used Ensemble Learning along with deep architectures of CNN to further improve the results. We apply the resultant model for effective classroom monitoring.

II. State Of The Art

A Facial expression recognition system typically performs following subtasks [3].

A. Face Detection

Various methods have been proposed in literature for face identification. In the case of static images, the most commonly used technique is called Viola-Jones, which achieves fast and reliable detection for frontal faces [4]. Among other localization techniques, there is a neural network-based face detection solution by Rowley et al [5] and a statistical method for 3D object detection applied to faces and cars, by Schneiderman[6].

Face tracking in image sequences uses other types of approaches, which rely on constructing 3D face models. Some popular examples are the 3D Candide face model [7] and Piecewise Bezier Volume Deformation tracker (PBVD) [8]

B. Feature Extraction Algorithms

The next and most important step is feature extraction, which can determine the performance, efficiencyand scalability of the system. The main goal is mapping the face pixels into a higher-level representation, in order to capture the most relevant properties of the image and reduce the dimension of data.

There are three types of approaches that appear in the literature, which depend on data and the goal of the system.

1. *Geometric or feature based techniques*

These are concerned with identifying specific areas or landmarks of the face. These methods are computationallymore expensive, but these can also be more robust and accurate, especially if there is variation in size or orientation. An example would be Active Shape Models, also known as ASM, which is popular for face and medical imaging applications. These statistical models learn the shape of objects and iteratively get adjusted on a new one (face). However, they can be highly sensitive to image brightness or noise. Improved results are achieved with Appearance Active Models (AAM), a more elaborated version of ASM which also incorporates texture information for building an object model.

2. *Appearance or holistic methods*

These methods do not treat the face as individual parts, but analyses the face as a whole. One of the most popular algorithms in the literature is the Gabor Wavelets, which can achieve excellent results in recognizing facial expressions. An interesting system developed by Bartlett et al in 2003 [9] uses this method and it has been deployed on several platforms. A major downside for real-time applications is the high computational complexity and memory storage, even though it is usually combined with a dimension reduction technique. An alternative approach, originally used for texture analysis but which recently gained popularity in the current context, is Local Binary Patterns (LBP)[11]. This technique has a great advantage in terms of time complexity while exhibiting high discriminating capabilities and tolerance against illumination changes. Either Eigenfaces or Fisherfaces are used for higher level representation in face recognition algorithms. LBPH algorithm tries to find the local structure of an image, by comparing each pixel with its neighboring pixels.

3. *Hybrid techniques*

These are perhaps the best in tackling feature extraction, which consists of a combination of the previous methods. Geometric procedures such as AAM are being

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019

used for automatic identification of important facial areas on which holistic methods, such as LBP, are applied.

C) Classification Algorithms

Once the features are detected, the given face needs to be classified as belonging to one of the pre-defined expression classes. There is a large variety of classifiers that are used in literature and choosing which one to use depends on criteria, such as type and size of data, computational complexity, the importance of robustness and overall outcome. One of the most popular methods is Support Vector Machines, greatly used for their results and high generalization capabilities, but suited for binary classification problems. Alternatively, powerful, flexible and capable of training complex functions are Artificial Neural Networks, which are also naturally multi-class algorithms, or Random Forests. Cohen et al. [11] suggest using dynamic classifiers, such as Hidden Markov Models. This method is proposed for person-dependent systems, as it is more sensitive to temporal pattern changes, in the case of videos.

Recently CNNs are successfully utilized for feature extraction and inference FER ([13]– [18]). We used the concept of ensemble learning [5] along with deep networks to further improve the expression recognition results. Instead of using a single CNN model two similar CNN models are used, one working on original image and other on mirrored image. We also applied the trained model for the effective classroom monitoring.

III. Methodology

A. Dataset

FER2013 is a large, publicly available FER dataset consisting of 35,887 face crops [19]. All images are grayscale and have a resolution of 48 by 48 pixels (Fig.1). These are divided in different categories namely anger, disgust, fear, happiness, sadness, surprise, as well as neutral. The dataset is challenging as the depicted faces vary significantly in terms of person age, face pose, and other factors reflecting realistic conditions. The dataset is split into training, and test sets with 70:30 ratio respectively. Basic expression labels are provided for all samples. The human accuracy on this dataset is around 65.5%. Only 5 of the 7 expression classes have been used due to lack of sufficient images for the excluded 2 classes. Image distribution for these 5 expression classes is as indicated in the table 1.

Fig. 1 Example faces in FER2013

Table 1. Image dataset

Emotions	**No. of images (48x48)**
Neutral	6198
Happy	8988
Sad	6077
Angry	4759
Surprise	4002
Total	**30024**

B. Model Architecture

Fig. 2 shows the architecture of deep network model used. The resultant model is ensemble of two similar models applied to an original image and it's mirrored image. Final prediction is made by concatenating results of two. As it can be noted four layers of convolution layers are used with different filter sizes, ReLU activation function and batch normalization. Max pooling of size 2*2 is used after the second and third layer. This arrangement is followed by 3 dense layers. Same set of layers are stacked in a model working on a mirror image. As there are 5 emotion classes last layer is a softmax layer of size 5. A dropout of 0.25 is used after some layers to avoid over-fitting.

We got accuracy of 70% on FER13 test dataset as shown by confusion matrix in fig. 3. Some results from the test dataset are depicted in fig. 4 along with the label predicted.

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019

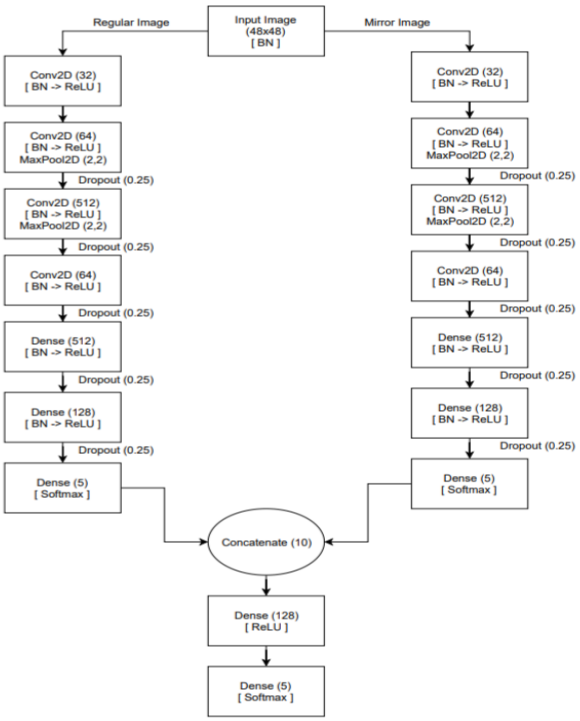

Fig. 2 Architecture of model network

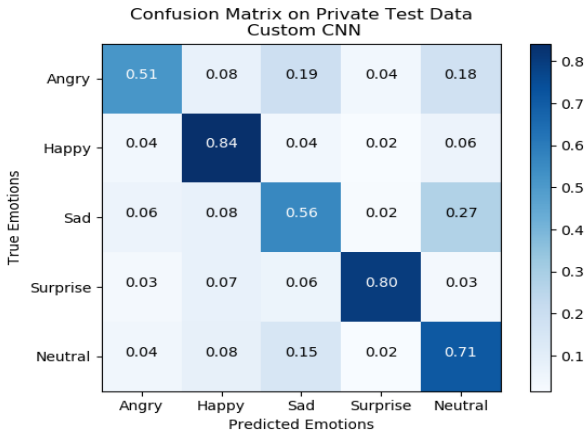

Fig 3. Confusion Matrix for Custom CNN

neutral angry

Fig 4. Result of Custom CNN on FER2013 dataset

IV. EFFECTIVE CLASSROOM MONITORING

Mere presence of a student in a classroom does not merit his/her attention towards the material taught in it. A Chinese school (Hangzhou Number 11 High School) uses facial recognition technology to monitor how attentive students are in class [20]. A good monitoring system helps students to be more attentive, teachers more adaptive, and derive various meaningful insights using data analysis.

We extended our trained model into complete monitoring system with a simple architecture as shown in fig. 5.

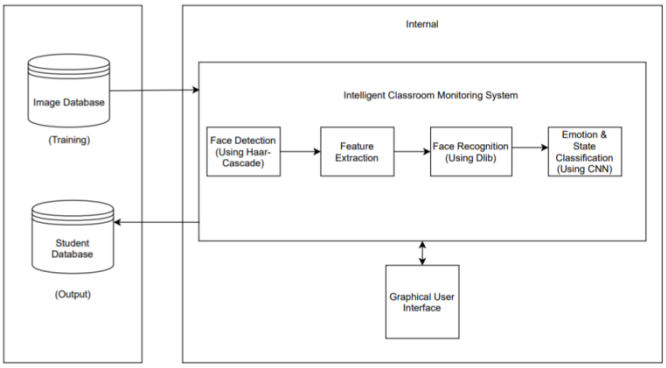

Fig 5. Proposed Monitoring System Architecture

The system mainly focuses on the recognition of human expressions and assigning the respective state the student is in(focused/distracted) and to identity the student (from student database). This is done using a webcam or an external camera. The user initiates the monitoring process by click of a button on the GUI which starts collecting data from single or multiple cameras pointed away from the user towards the classroom. The live feed of the camera will be visible on a bigger device like a laptop/desktop on which the predictions are being made. This feed is broken down frame by frame and passed through the pipeline of trained algorithms which in turn detect the faces of the students along with their identities, expressions and states.To detect a face from every feed Haar Cascade is used [4].

A student is assumed to be ‘focused’ if the expression is neutral for predefined some time. Other facial expressions are considered to be ‘distracted’ for the student. Based on the

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019

presence of a student's image on screen the system calculates his/her attendance.

The system identifies different facial expressions for the students, and that information is then fed into a computer which assesses if they are enjoying lessons or if their minds are wandering. If it concludes that the student is distracted with other thoughts during class, it will send a notification to the teacher to take action.

Some results of real time classroom monitoring are displayed in fig. 6. As can be found the system correctly detects faces of all students and captures their emotions.

Fig 6. Real time result of proposed model

V. CONCLUSION

This paper proposes a system for facial expression recognition involving the domains of computer vision and machine learning. The facial expression of a student, after being captured, goes through multiple stages of processing such as face detection, feature extraction, and classification. The knowledge obtained can be used for effective classroom management. The system displays student's identity along with the time he/she was focused in the class. It also calculates attendance of the student in a class.

References

[1] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 37, no. 6, pp. 1113–1133, 2015

[2] M. V. B. Martinez, "Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition", in Advances in Face Detection and Facial Image Analysis. Springer, 2016, pp. 63–100.

[3] Sohn, K., Jung, D. Y., Lee, H., & Hero, A. O. (2011),"Efficient learning of sparse, distributed, convolutional feature representations for object recognition", 13th International Conference on Computer Vision (ICCV), pp. 2643-2650

[4] Huang, F., &LeCun, Y. (2006),"Large-scale learning with SVM and convolutional nets for generic object categorization",In *Proc. Computer Vision and Pattern Recognition Conference (CVPR'06)*

[5] Leon J.M. Rothkrantz Maja Pantic,"Automatic analysis of facial expressions: the state of the art", IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1424–1445, 2000.

[6] Paul Viola and MichaelJ. Jones,"Robust real-time face detection", International Journal of Computer Vision, 57(2):137–154, 2004.

[7] Takeo Kanade Henry A. Rowley, ShumeetBaluja,"Neural network-based face detection", IEEE Transactions on Pattern Analysis and Machine Intelligence , 20(1): 23–38, 1998.

[8] Takeo Kanade Henry Schneiderman,"A statistical method for 3d object detection applied to faces and cars". 1:746–751 vol.1, 2000.

[9] J¨orgenAhlberg,"Candide-3 – an updated parameterized face', Technical Report LiTH-ISY-R-2326, Dept. of Electrical Engineering,Link¨oping University, Sweden, 2001.

[10] Isabelle Guyon and André Elisseeff, "An Introduction to Feature Extraction", 1 ClopiNet, 955 Creston Rd., Berkeley, CA 94708, USA.

[11] S. L. Happy, A. George, and A. Routray, "A real time facial expression classification system using Local Binary Patterns", 4th International Conference on Intelligent Human Computer Interaction (IHCI), 2012.

[12] Hai Tao and THOMAS S Huang,"A piecewise bezier volume deformation model and its applications in facial motion capture", Series in machine perception and artificial intelligence, 52:39–56, 2002

[13] Y. Tang, "Deep Learning using Support Vector Machines," in International Conference on Machine Learning (ICML) Workshops, 2013.

[14] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in ACM International Conference on Multimodal Interaction (MMI), 2015, pp. 435–442.

[15] B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, and S.-Y. Lee, "Fusing Aligned and Non-Aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach," in IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops, 2016, pp. 48–57

[16] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, pp. 1–17, 2016.

[17] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going Deeper in Facial Expression Recognition using Deep Neural Networks," CoRR, vol. 1511, 2015.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[19] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015.

[20] https://www.engadget.com/2018/05/17/chinese-school-facial-recognition-kids-attention/ last retrievedon 12/08/2019

ISSN:0971-1260
Vol-22-Issue-4-October-December-2019