Attendance System Based on Dynamic Face Recognition

Shizhen Huang[1], Haonan Luo[2]
College of Physics & Information Engineering
Fuzhou University
Fuzhou, Fujian P.R. China
hs501@qq.com, 414889405@qq.com

***Abstract*—A video-based attendance system is designed by using the method of real-time face recognition. The system supports multi-user attendance and face liveness detection at the same time. The system can automatically collect face data, which will be saved in the database as well as attendance results. The face detection part of the system is based on MTCNN (Multitask Convolutional Neural Network) algorithm, and the face recognition part is based on FaceNet algorithm. The algorithm implementation is based on TensorFlow framework, and the face liveness detection part is based on ERT (Ensemble of Regression Tree) algorithm, which can judge whether the user blinks. The attendance system is written in Python language, and the user interface is designed by Qt library. The experimental results show that the system achieves a good performance in real-time face recognition. The false accept rate and false rejection rate of face recognition are within 2%, and the recognition rate can be stable at 20 FPS.**

Keywords—MTCNN; FaceNet; ERT; attendance system; face recognition

I. Introduction

Nowadays, attendance system is of great significance in enterprises, schools, governments and other places where personnel management is needed. Attendance through fingerprint recognition requires queuing for identification, which consumes a lot of time. In the case of finger injury, the accuracy of fingerprint recognition will be greatly reduced, and the fingerprint can also be forged by others. By scanning the ID card for attendance, the identity of the cardholder cannot be verified, which will also produce fraudulent attendance behavior. By using iris recognition for attendance, the recognition speed is slow, and the consumption of time and equipment cost is large. Checking attendance by mobile phone location is similar to scanning ID card, which cannot confirm the user's identity, and the location information can also be forged.

With the continuous development of machine learning and artificial intelligence technology, the methods of face detection, face recognition and face landmarks detection have changed greatly. As an important biological feature, human face has been widely used in the attendance system. The dynamic face recognition technology eliminates the need for users to stop and wait for verification. The user only needs to appear within the scope of video surveillance, and the system can automatically recognize it. Because of its real-time and convenience, this technology has become a hot research direction of attendance system [1]. In this paper, an attendance system based on dynamic and multi-face recognition is designed. To prevent users from using photos for attendance, the system incorporates the ability to recognize blinking movements. Finally, for user convenience, we also designed a user interaction interface for the attendance system.

II. Algorithm Implementation

A. Multitask Convolutional Neural Network

The multitask convolutional neural network [2] is used to detect the face and the landmark of the face, including eyes, nose and mouth. In the first stage, we exploit a fully convolutional network, called Proposal Network (P-Net), to generate candidate boxes of face. Similar candidate boxes will be merged by non-maximum suppression (NMS). In the second stage, we use another CNN, called Refine Network (R-Net), which will eliminate a large number of false candidates. In the last stage, the face detection results are improved, and the network will output five facial landmarks' positions. According to the actual needs of attendance system, we improve the algorithm by modifying some parameters and thresholds. As shown in Figure 1, the improved algorithm uses 1.9 seconds to detect all faces, which meets the real-time requirements of attendance system.

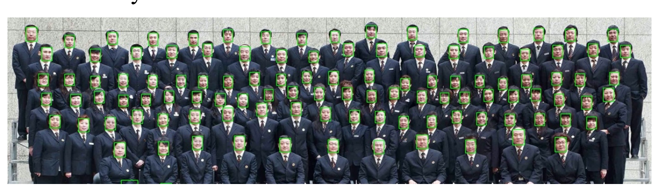

Figure 1. Face detection by MTCNN

B. FaceNet

FaceNet is a deep learning neural network proposed by Google in 2015 [3], which is mainly used in face recognition. FaceNet algorithm can transform the face image into a 128-dimensional feature vector. Euclidean distance is used to represent the distance between two vectors. The closer the distance is, the higher the similarity between the two face images is. By mapping the face image to the vector then calculating the distance between the vectors, the complexity of the algorithm is greatly reduced.

978-1-7281-9761-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CISCE50729.2020.00081

The algorithm uses triplet loss function to train the neural network. As shown in Figure 2, Anchor, positive sample (the same person as Anchor) and negative sample (not the same person as Anchor) are all 128-dimensional vectors. Before training, the distance between them is uncertain. After training, the distance between Anchor and positive sample will be reduced as much as possible, while the distance with negative sample will be expanded as much as possible.

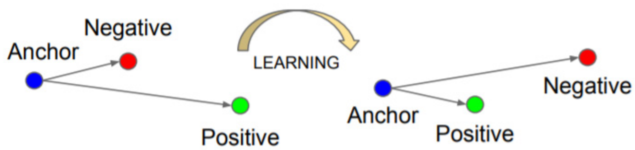

Figure 2. The Triplet Loss Function

Siamese network [4] is a kind of network which can realize one-time learning. When new face images are added to the database, there is no need to retrain the model. This network solves the problem of continuously collecting new face images in the attendance system. Based on Siamese network, the accuracy of FaceNet model can reach 99.6%. In order to improve the operation speed, the detected face image is scaled to 160 × 160, and the pre whitening process is carried out, and the threshold value of the algorithm is adjusted.

C. *Gradient Boosting Decision Tree*

Gradient boosting decision tree algorithm is a face alignment algorithm based on regression tree. By establishing a cascaded residual regression tree, the face shape can be regressed from the current shape to the standard shape step by step. Based on this algorithm, we can detect 68 landmarks of human face. When the face has been detected, it only takes 4ms to detect the landmarks of a single face [5].

Detecting whether the user is executing the action specified by the system is the most widely used and the lowest cost face living detection method. By combining blinking, nodding, looking to the left, looking to the right, opening mouth and other instructions at the same time, it can effectively judge whether a real person is standing in front of the camera.

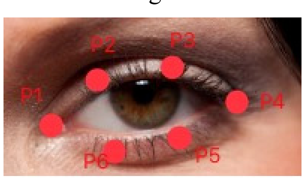

Figure 3. landmarks in each eye

Based on 68 landmarks of human face, there are 6 landmarks in each eye as shown in Figure 3. According to the aspect ratio of the eyes, the user's actions of closing, opening and blinking can be judged, so as to realize face living detection. The calculation formula of aspect ratio is as follows:

$$ARE = \frac{\| p_2 - p_6 \| + \| p_3 - p_5 \|}{2 \| p_1 - p_4 \|} \qquad (1)$$

III. ATTENDANCE SYSTEM

As shown in Figure 4, the interface consists of four buttons, a video display window, an edit box, and two information display boxes. The information display boxes are used to display the system information and the list of attendees that day. The edit box is used to enter the user's name to start the face collection module. The *Face Collect* button is used to start the process for automatically collect face data when the requirements are all met. The *Start* button is used to start whole system. The *Database* button is used to output the face data already in the database. The *Stop* button terminates the attendance or face collection process.

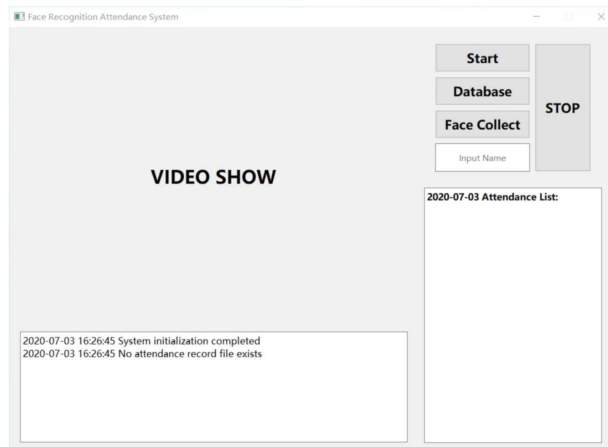

Figure 4. User Interface

A. *Face Collect*

As shown in Figure 5, face collection is a key part of attendance system. The quality of face data directly affects the speed and accuracy of face recognition. The traditional face collection method is not able to judge the quality of the information. In this module, we use a variety of methods to achieve the automatic collection of face data, which needs to meet the following conditions:

1. The picture needs to be stable enough. Based on the frame difference method [6], we first obtain the gray image of two adjacent frames, and then take the absolute value after subtracting the pixel value of the two images, and finally binarize the result. When the ratio of 255 pixels to the total number of pixels is less than 25%, the picture is stable enough.

2. Meet the minimum face size. The smallest face bounding box is displayed in the video display using a red frame. The green bounding box is the face image detected by MTCNN. It can only be judged that the face size meets the minimum size requirement if the green box surrounds the red box.

3. Eyes open. Based on the face landmarks detected by GBDT algorithm, the aspect ratio of two eyes can be calculated. When the average aspect ratio is greater than 2.5, the eyes are judged to be open.

When the face image meets the above conditions, the system will automatically collect the face image, then invoke the FaceNet network to convert the face image into a 128-

dimensional vector and save it in the database. By using this method, the space required for data storage is greatly reduced, and the original 160 *160 *3 face image is converted to a digital rent storing 128 integers. This accelerates the comparison of face vectors for subsequent data calls.

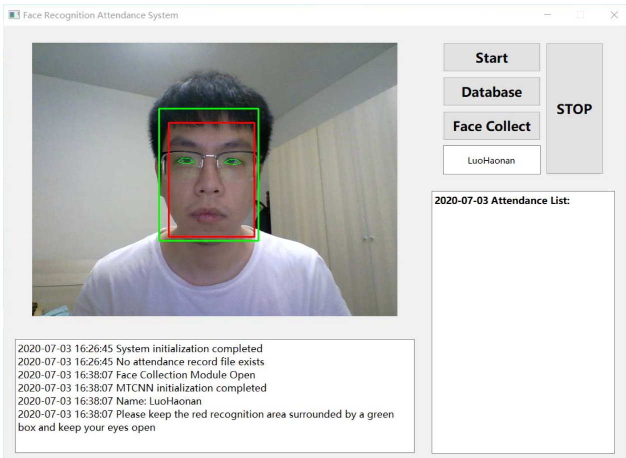

Figure 5. Face Collect

B. *Real-time attendance*

When the attendance system is turned on, the video display window will display the picture captured by the camera in real time, and start face detection. The identification process is shown in Figure 6.

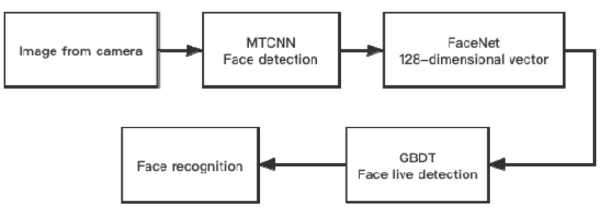

Figure 6. Identify Process

The size of the picture captured by the camera is 640 × 480. When a face is detected, the system converts the face image to a 128-dimensional vector and calculates the Euclidean distance with each face vector that already exists in the local database, then output the minimum Euclidean distance. If the Euclidean distance obtained is less than the set threshold, the corresponding name will be displayed, otherwise the "unknown" will be displayed. The identified person will enter the face living detection status. If the person blinks, the attendance will be successful and the data display window on the right will show the person's name.

The system supports multiple people appearing in video pictures, as well as face recognition and face living detection. As shown in Fig. 7, a mobile phone picture named "Li Yupeng" on the right and a real person named "Luo Haonan" on the left appear in the video at the same time. After face living recognition, the user named "Luo Haonan" successfully attended, while the user named " Li Yupeng" was recognized, but did not appear in the attendance list.

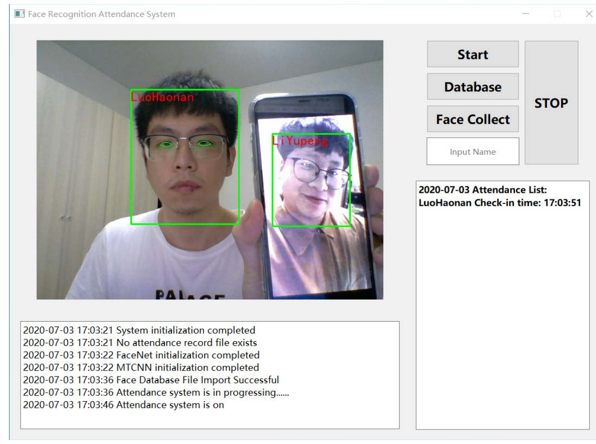

Figure 7. Face Living Detection

C. *Data Management*

The system saves the collected face data and attendance results in the local database. Each face data is stored as a 128-dimensional vector in the database. The attendance system generates a database file daily to hold attendance results. When the attendance system is turned on, the list of attendees who have attended today is read automatically. Everyone can attend only once per attendance period.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Usually false accept rate and false rejection rate [7] are commonly used to evaluate the performance of face recognition system. False accept rate refers to the probability of recognizing two different faces as the same person, which belongs to the matching between different classes. False rejection rate refers to the probability that two identical faces but be recognized as different people, which belongs to the matching within the same class. Their formulas are as follows:

$$FAR = \frac{N_{FA}}{N_{IRA}} \times 100\% \tag{2}$$

$$FRR = \frac{N_{FR}}{N_{GRA}} \times 100\% \tag{3}$$

N_{FA} is the number of times to identify two different people as the same person; N_{IRA} is the number of inter-class tests; N_{FR} is the number of times to identify the same two people as a different person; N_{GRA} is the number of in-class tests.

The face recognition module is implemented by judging the Euclidean distance between two faces. When the Euclidean distance is greater than a certain threshold, system think two faces are not the same person. Modifying the threshold will result in different rates of false accept and false rejection, which will have a significant impact on the attendance system.

We collect 500 different face images and 100 photos of the same person in different states. After each threshold modification, 124750 inter-class tests and 4950 intra-class tests were performed. The threshold increased from 0.3 to 0.9, with an increase of 0.05 each time. The results of false accept rate and false rejection rate are shown in Figure 8.

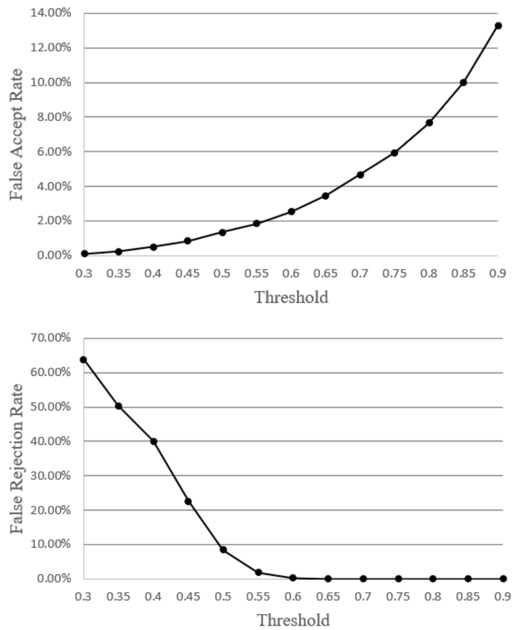

Figure 8. False Accept Rate and False Rejection Rate

From the above experimental results, we can see that when the threshold value is 0.55, the false accept rate is 1.85%, and the false rejection rate is 1.92%. In the actual application of attendance system, the threshold can be reduced. We can minimize the false accept rate, and increase the false rejection rate. The performance of the attendance system is optimal when the threshold is set to 0.50 in practice.

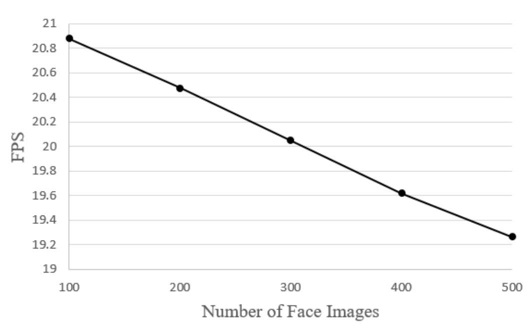

Figure 9. Identification Speed Test Results

In recognition, it is necessary to compare the detected face with each existing face in the database, so the number of faces already in the database will affect the speed. We have tested the system at 16GB RAM, and the PC side of Intel® Core (TM) i7-4720HQ runs. We don't use GPU to speed up the calculation of the neural network. Based on this condition, the identification speed test results of the attendance system are shown in Figure 9.

When the number of faces in the database is 100, the identification speed can reach 21FPS. When the database is expanded to 500 face pictures, the system can still maintain the speed of 19FPS, so the system has good real-time performance.

V. CONCLUSION

This paper designs a dynamic face recognition attendance system with living detection based on MTCNN, FaceNet and ERT neural networks. After optimizing the MTCNN algorithm for face detection, the detection rate can reach 33 fps. This paper uses FaceNet neural network to convert face image into a 128-dimensional vector. The Euclidean distance between the detected face and local database face is calculated to realize the function of face recognition. Combined with GBDT algorithm, the activity state of human eyes is analyzed to complete the live detection function. Based on the above three networks, an attendance system with user interaction interface has been developed using Python language. The attendance system can be used in schools, companies, enterprises and other scenarios that require multiple people to attend at the same time. The results show that both the false accept rate and the false rejection rate of face recognition are within 2%. The recognition rate can be stable around 20 FPS, and the performance of the attendance system is good.

ACKNOWLEDGMENT

This research was financially supported by the key project of Science and Technology Department of Fujian Province, No. 2017H0017.

REFERENCES

[1] E. Jose, M. Greeshma, “Face recognition based surveillance system using FaceNet and MTCNN on Jetson TX2,” 2019 5th International Conference on Advanced Computing & Communication Systems, 2019.

[2] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Processing Letters, vol.23, 2016, pp. 1499-1503.

[3] F. Schroff, D. Kalenichenko and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 815-823.

[4] J. Yuan, H. Guo, Z. Jin, H. Jin, X. Zhang and J. Luo, “One-shot learning for fine-grained relation extraction via convolutional siamese neural network,” 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 2194-2199.

[5] V. Kazemi, J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[6] Ling Zhang, Li-Min Chen, “Application of an improved frame-difference method based on video in traffic flow measurement,” Journal of Chongqing University, 2004.

[7] Xingjie Zhu, Yong Xu, “Low false reject rate and false accept rate multi-step fire detection method,” Optik - International Journal for Light and Electron Optics, vol.124, 2013, pp. 6636-6641.