Dormitory Management System Based on Face Recognition

Yu Yang[1,3] and Liang Ye[1,2,3(✉)]

[1] Department of Information and Communication Engineering, Harbin Institute of Technology, No. 2 Yikuang Street, Harbin 150080, China
yeliang@hit.edu.cn
[2] Physiological Signal Analysis Team, University of Oulu, Pentti Kaiteran katu 1, 90014 Oulu, Finland
[3] Science and Technology on Communication Networks Laboratory, Shijiazhuang, China

Abstract. This paper explores the application of face recognition system in dormitory management, and designs an EXE software, which realizes the function of entering and leaving dormitory through face voucher, storing the information of students, and automatically updating relevant data. The face recognition module includes three functions: face image recognition and interception, face alignment, face feature extraction and face verification. The traditional method of HOG is used for recognition and interception, and the method of gray-scale processing and gamma normalization is used for preprocessing to reduce the influence of light. The face alignment uses the 68-point landmarks and similarity transformation to align the face. The face feature extraction uses the pre-trained FaceNet neural network to extract the 128-d feature vector of the face, which is stored in the database or compared with the face in the database to output the Euclidean distance, then compare and find the most possible person according to the distance. The database module includes two functions: storing face information and displaying all information of students MySQL is used to build the database. An automatic dormitory information management system is established through the interconnection interface between MySQL and Python.

Keywords: Face recognition · HOG · FaceNet · MySQL · UI

1 Introduction

At present, with the rapid development of artificial intelligence, the era of AI has quietly arrived, and "showing your face" has gradually become a new trend. With the higher demand of quick and accurate methods of verification, Face recognition is becoming more and more common.

Face recognition is a biometric technology based on human's face feature information. A series of related techniques for face recognition are used to collect images or video streams containing faces with cameras, and to automatically detect and track faces in images, and then to recognize the detected faces. These techniques are also called portrait recognition or facial recognition. 2018 is an important node for the

S. Shi et al. (Eds.): AICON 2020, LNICST 356, pp. 350–361, 2021.
https://doi.org/10.1007/978-3-030-69066-3_31

common application of face recognition technology in China, marking the final arrival of the era of face recognition [1].

Except the field of security and finance, face recognition has been widely used in many other fields, such as transportation, education, medical treatment, police, e-commerce and so on. In order to further grasp the opportunities brought by face recognition technology, China has issued a series of policies to support the research about this new technique.

2 Background

Artificial intelligence (AI) is a advanced and new field. Face recognition is one of the representative directions of AI, which is widely used in our production and life, with high performance of real-time and awareness. Its application is common and its status in access control, payment, authorization and smart-home is increasingly high. At present, the research and development of face recognition is developing rapid and the jurisdiction is getting there and higher, but the use has not been so popular, many occasions that can apply face recognition systems have not installed that for a series of reasons.

Face recognition study became popular in the early 1990s [2] following the introduction of the historical Eigenface method. After this, many holistic approaches are put forward. However, a well-known problem about these holistic approaches is that these methods fail to address the unexpectable facial changes that deviate from their prior assumptions. So local-feature-based face recognition appeared to solve this problem. As the time flies, techniques like learning-based descriptors, deep learning descriptors appeared and the accuracy was dramatically boosted to 99.8% in just several years [3].

The purpose of this paper is to find an quick and accurate method which do not have the demand of systems of high performance to manage students' dormitory in a fast and efficient way. It has three parts as follows:

(1) Face recognition module: FR module includes three functions: face image recognition and interception, face alignment, and face feature extraction and face verification. The traditional method called Histogram of Oriented Gradient is used to detect and intercept a face, and the pre-processing uses the method of gamma normalization to reduce the influence of illumination. Face alignment using a python package called Dlib to align the 68 special point of the face. Face feature extraction adopts pre-trained Facenet neural network to extract 128 dimensional feature vector of face and to store it in database or compare with face in database and calculate Euclidean distance, so as to compare the most similar person according to this distance.
(2) Database module: the database module contains two functions: storing face information and displaying all information of members. Database is established by MySQL database.

(3) UI and software module: the software's user interface is designed by QT designer and encapsulated into a software system that can run by each device. The hardware demand is low and the related software environment is no longer needed.

3 Work Methodology

3.1 FR Module

3.1.1 Face Recognition and Interception

HOG features are called Histogram of Oriented Gradient [4]. The directional gradient histogram feature is a very effective image feature descriptor, which is widely used in various detection and recognition scenes. The algorithm can describe the image features by calculating the gradient of the local image and counting its directional gradient histogram, and can describe the edge texture information of a specific target well. HOG basic idea is that even without a clear understanding of the relative gradient or boundary attitude, the appearance and shape of an object can often be represented by a local strong gradient change or edge direction. For algorithm implementation, this idea is implemented by dividing the image window into small blocks, which are called cell. Units For each pixel in each cell, the one-dimensional gradient histogram is accumulated. In order to have better invariance to different illumination, control normalization is also an important step before using these local data. Normalization can be achieved by accumulating local histogram "energy" in a larger region and using this energy to normalize all cell. This larger region is called block. The algorithm of HOG are as follows:

Algorithm 1:

1. Grayscale processing.
2. Gamma normalization.

   Compute the value of each pixel with the function as follows:

$$I(x, y) = I(x, y)^{gamma} \tag{1}$$

   gamma usually take 0.5

3. Calculating the gradient of each pixel of the image (including size and direction).

   One dimensional gradient of pixel (x, y) in the image can be calculated as follows:

$$\begin{aligned} G_x(x, y) &= H(x+1, y) - H(x-1, y) \\ G_y(x, y) &= H(x, y+1) - H(x, y-1) \end{aligned} \tag{2}$$

$G_x(x, y), G_y(x, y), H(x, y)$ represents the gradient of pixel (x, y) in x orientation, y orientation and the value of the pixel. Then use the one dimensional gradient to calculate the amplitude and phase of each pixel with the following function:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2}$$
$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{3}$$

4. Segmenting of images into small area called cells (each 16 pixels).
5. Building gradient direction histograms for each cell.

Divide 180 into several bins. Then use these bins as abscissa to do weighted vote for the histograms. The weight are related to the amplitude of each pixel which are calculated by a statistical method called tri-linear interpolation.

Combining cells into large intervals (block), and normalizing gradient histograms within blocks.

First combine several adjacent cells into a larger area called block, then use a function to normalize the histograms within each block. There are four methods to normalize the histograms, this paper use L2-norm normalization as follows:

$$L2 - norm, v \rightarrow v/\sqrt{\|v\|_2^2 + \varepsilon^2} \tag{4}$$

v represents the value of each bin.

After the extraction of HOG features, these features can be used to train a SVM [5] classifier to detect a human face.

3.1.2 Face Alignment

Face alignment module is based on the Dlib deep learning library. According to the 68-points face landmarks [6], by finding and multiplying a similarity transformation matrix, the face features can be aligned to the average face. The algorithm can complete the face alignment in a few milliseconds and meet the fairly high accuracy, which can meet the efficiency and accuracy requirements of this program. A practical implementation of the algorithm is provided by the Dlib deep learning library, which can be easily applied to the program.

3.1.3 Face Verification

A deep network structure called Facenet [7] proposed in 2015 is used in this system. The biggest innovation of the network should be to propose a different loss function. What's more, instead of using a classifier, we use the distance of points in the eigenspace to indicate whether the two images are the same class. The network structure is as follows (Fig. 1):

1. Batch are sent to a deep architecture, such as a convolutional neural network. Google Inception ResNet V1 [8] is selected here.
2. 2.After CNN, each feature **x** should be L2-normalized to make, so that the feature of all images is mapped to a hypersphere.
3. Then send the feature into an embedding process, so the image x is mapped to 128-dimensional Euclidean space by function $f(\mathbf{x})$.

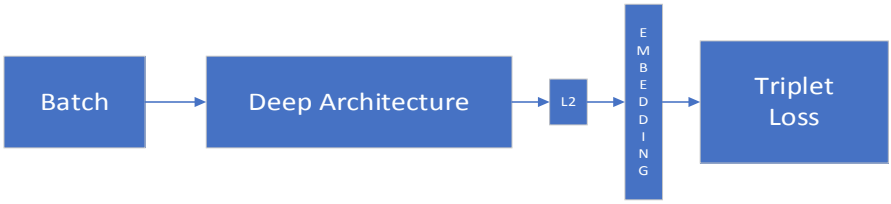


Fig. 1. Net structure

4. After the embedding layer, we optimize the weight of CNN with these features, and here we propose a new loss function, triplet loss function, which is the biggest innovation of the network.

The triple loss function is a loss function with three sample inputs, containing an anchor sample, a positive sample and a negative sample. By training the classification model with LDA method, the intra-class feature distance is getting closer and the inter-class feature distance is getting far away. In order to ensure that the feature distance between the anchor sample and the intra-class image (positive sample) is closer and the feature distance from the inter-class image (negative sample) is farther, we need the triplet loss function to realize it.

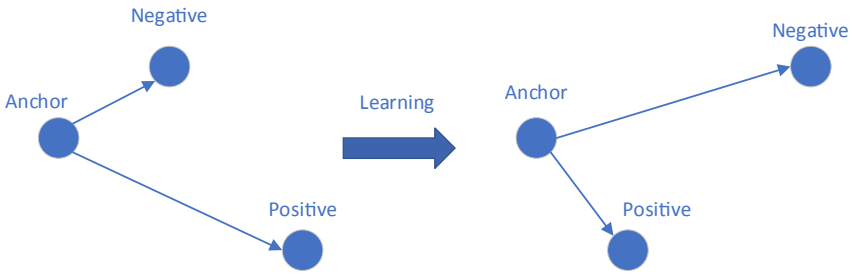


Fig. 2. Training ideas

A constraint can be constructed:

$$\left\|f\left(x_i^a\right) - f\left(x_i^p\right)\right\|_2^2 + \alpha < \left\|f\left(x_i^a\right) - f\left(x_i^n\right)\right\|_2^2, \forall\left(x_i^a, x_i^p, x_i^n\right) \in \Gamma \tag{5}$$

The variable α is the distance tolerance between the intra-class pair and the inter-class pair, which can be set artificially. Γ is the set of all triples in the training set.

Then write the formula 5 as a loss (optimization) function, and the model is optimized by reducing the value of the loss function. The loss function can be expressed as:

$$L = \sum_{i}^{N} \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+ \tag{6}$$

By using the transfer learning method, the FaceNet model weights trained by 13000 pictures are transferred to generate 128-dimensional feature vectors of faces. Through comparing the Euclidean distance between the input image and images in the database, we can judge who is the most similar owner of the input face image. The Euclidean distance formula is as follows:

$$\sqrt{\sum_{i=0}^{127} (\vec{x}_i - \vec{y}_i)^2} \tag{7}$$

$(\mathbf{x}, \mathbf{y})$ are the 128 dimensional features of two faces.

3.2 Database Module

Database module is a place to store and display personnel information and face information. To achieve the following functions:

1. Storage and Automatic Updating of Dormitory Personnel Information.
2. Supporting easy to query, modify and import information manually.

MySQL is a relational database management system. Based on the MySQL database, we can establish an quick and high-capacity database module which satisfies the demand above.

4 Results

4.1 Test Platform

Hardware environment: CPU: Intel (R) Core (TM) i5-5257U.
GPU: Intel (R) Iris (TM) Graphics 6100 which does not support CUDA.
Software environment: operating system: Windows10.
Development language: Python.
IDE: Pycharm.

4.2 Variable Configuration and Data Set

The sizes of each element in FR system are as follows:

OpenCV detection window: WinSize = 128 * 64 pixels, the step size of sliding in the image is 16 pixels (both horizontal and vertical).

Block: BlockSize = 64 * 64 pixels, the step size of sliding in the detection window is 16 pixels (both horizontal and vertical).

Cell: Cell size = 16 * 16 pixels.

The training data set of face verification system is selected from CAS-PEAL data set [9] randomly. This test selects 500 images of 50 different people.

4.3 Test Result

After configuring and training, this section carries on a whole analysis to the system, mainly analyzes the face detection accuracy, the recognition speed, the face recognition accuracy rate, the database adds and deletes the check operation speed. LFW [10] (LableFaces in the Wild) data set is used for the test data set.

4.3.1 Face Recognition Performance Test

A trained SVM classifier is used to detect a part of the sampled picture in the LFW data set three times, and the three results are recorded in Table 1:

Table 1. FR test results

Number of tests	Verification accuracy	Recognition time
1	0.971	48.33 s
2	0.972	54.89 s
3	0.971	33.84 s

As a result, the average recognition time of 1454 pictures is about 40 s, the recognition accuracy is about 97.1%, the recognition performance is ideal, and the face detection can be realized quickly and accurately.

4.3.2 Face Verification Performance Test

The LFW data set is first aligned with the face, and the input layer image with FaceNet size should be 160 * 160 pixels. Then the FaceNet is used to verify all the images of the LFW data set, and the total face detection time, face detection accuracy, and the results are as follows:

```
Model directory: C:\Users\apple1\PycharmProjects\evaluation\facenet\src\models\20180402-114759
Metagraph file: model-20180402-114759.meta
Checkpoint file: model-20180402-114759.ckpt-275
Runnning forward pass on LFW images
............
Accuracy: 0.98467+-0.00407
Validation rate: 0.90567+-0.01995 @ FAR=0.00067
Area Under Curve (AUC): 0.998
Equal Error Rate (EER): 0.015
```

Fig. 3. Face verification result

It can be seen that the accuracy rate is 98.46 $\pm$ 0.004% and is at a high level. The system can accurately distinguish faces and meet the demand of practical application.

4.3.3 Database Performance Testing

Write 10000 pieces of data into the table, then delete them in batches. The records of the run time of the three programs are as shown in Table 2.

Table 2. Database test results

Number of tests	Runtime
1	129.0 s
2	118.7 s
3	126.5 s

As can be seen, the average time of adding and deleting 10000 pieces of information in the database is about 123 s, and it supports batch deletion, batch import, which also has very fast speed and very convenient operation.

4.3.4 UI Performance

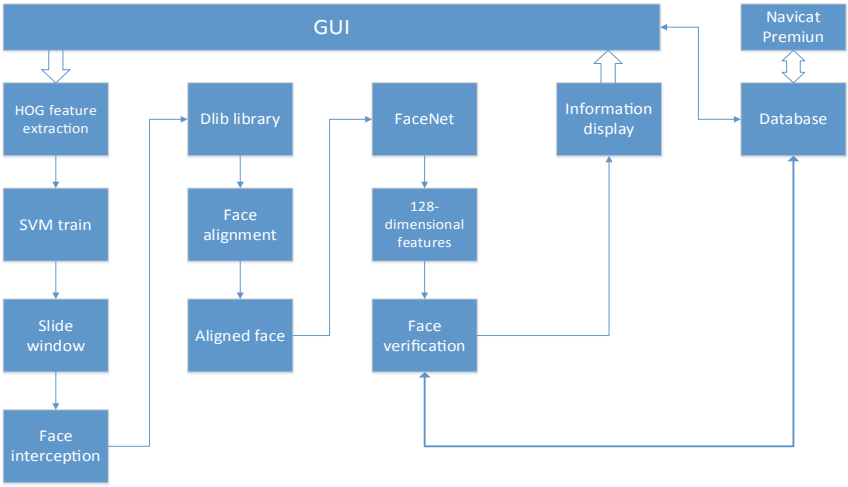


Fig. 4. UI functions

Open the encapsulated software, the overall user interface function is shown in Fig. 4 and design is shown in Fig. 5(a). After turning on the camera, click on the recognition button for real-time recognition. The recognition result of strangers is shown in 5(b), and the recognition result of students in the database and database changes are shown in Fig. 5(c)(d)(e).

The error of the password and the error of the unfamiliar face are all reported normally according to Fig. 6(a)(b). The results of the registration of the stranger's face

(a)whole UI

(b)recognize stranger

(c)recognize person in the database

对象 students @apartment (APAR...

开始事务 文本 · 筛选 排序 导入 导出

number	name	sex	regist_time	state
160200000	trump	男	2020-04-14	离寝
160200101	zhangziyin	女	2020-04-14	离寝
160201010	liyifan	女	2020-04-14	在寝
160201031	yuanmingzhi	男	2020-04-14	在寝
160200131	zhouyou	男	2020-04-14	在寝
160200500	zhoujielun	男	2020-05-29	离寝

(d) original database

对象 students @apartment (APAR...

开始事务 文本 · 筛选 排序 导入 导出

number	name	sex	regist_time	state
160200000	trump	男	2020-04-14	离寝
160200101	zhangziyin	女	2020-04-14	离寝
160201010	liyifan	女	2020-04-14	在寝
160201031	yuanmingzhi	男	2020-04-14	在寝
160200131	zhouyou	男	2020-04-14	在寝
160200500	zhoujielun	男	2020-05-29	在寝

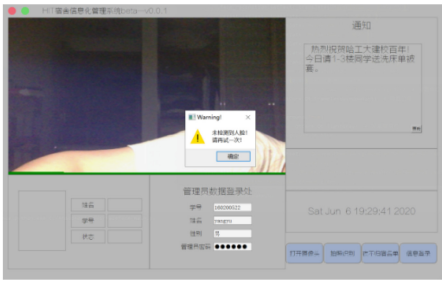
(e)changed database

Fig. 5. FR results and changes of and database in recognition library

are shown in Fig. 7(a)(b). Click recognition again to identify the newly registered face and it's shown that the recognition is successful by Fig. 7(c).

Because the test time is uncertain, the time can not be used as the trigger signal, so a button is used to simulate the time which is the deadline of backing dormitory, and the click button generates the list of people who are not in bed, as shown in Fig. 8.

(a)error password



(b)no face

Fig. 6. Error situation

(a)register new face

对象 students @apartment (APAR...

开始事务 文本 · 筛选 排序 导入 导出

number	name	sex	regist_time	state
160200000	trump	男	2020-04-14	离寝
160200101	zhangziyin	女	2020-04-14	离寝
160201010	liyifan	女	2020-04-14	在寝
160201031	yuanmingzhi	男	2020-04-14	在寝
160200131	zhouyou	男	2020-04-14	在寝
160200500	zhoujielun	男	2020-05-29	离寝
160200522	yangyu	男	2020-06-05	在寝

(b)changed database

(c)recognize newly registered

Fig. 7. Identification of registered information

Fig. 8. Single generation

At this point, the whole system function verification and analysis, the system successfully completed all the above functions.

5 Conclusion

This paper focuses on the design and implementation of a dormitory information management system based on face recognition, and describes the research of related algorithms with machine learning and deep neural network as the core, then finish the design and implementation of a dormitory information management system based on these theories.

The system is targeted at providing a more economical and effective new management mode for the traditional dormitories that use manual management, and only have computers of low performance in parallel computing. The computer does not need high performance GPU, and the application portability is strong, the traditional PC machine with camera can run this software without complex configuration due to this software's strong compatibility.

6 Further Study

The accuracy of the program is greatly affected by the clarity, distance and illumination of the camera. To further improve the accuracy, it is necessary to collect more images of all registered dormitory members and retrain the FaceNet of pre-training, so as to improve the ability to distinguish the members of the dormitory and the faces of strangers. By training with more data, the CNN can fit the real result more precisely. By collecting training set more suitable and consider the influence of clarity, distance and illumination in, it is possible to use this system commonly in some schools with old infrastructure and help them realize the smart education more conveniently.

Acknowledgement. This paper was funded by the National Natural Science Foundation of China under grant number 41861134010, the Key Laboratory of Information Transmission and Distribution Technology of Communication Network (HHX20641X002), National Key R&D Program of China (No.2018YFC0807101).

References

1. Wei, J., Jian-Qi, Z., Xiang, Z.: Face recognition method based on support vector machine and particle swarm optimization. Expert Syst. Appl. **38**(4), 4390–4393 (2011)
2. Wang, M., Deng, W.: Deep face recognition: a survey. arXiv (2018)
3. Sun, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. Adv. Neural Inf. Process. Syst. **27**, 1988–1996 (2014)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE (2005)
5. Saunders, C., Stitson, M.O., Weston, J.: Support vector machine. Comput. Sci. **1**(4), 1–28 (2002)
6. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2013)
7. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2015)
8. Szegedy, C., Ioffe, S., Vanhoucke, V., et al.: Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning (2016)
9. Gao, W., Cao, B., Shan, S., et al.: The CAS-PEAL large-scale chinese face database and baseline evaluations. IEEE Trans. Syst. Man Cybern. **38**(1), 149–161 (2008)
10. Huang, G., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database forstudying face recognition in unconstrained environments. Technical report (2008)