Calculation of the Rate of Taking Lessons Based on Face Recognition and Head Pose Estimation

Yidan Tang [1, a], Baosheng Wang[1, b], Yizhi Shen[1, c] and Ziyang Zong[1, d]

[1]School of Information and Control, Nanjing University of Information Science & Technology, Nanjing ,Jiangsu,210044,China;

[a]1404966650@qq.com, [b]420388393@qq.com, [c]3265346603@qq.com, [d]realtyd@163.com

Abstract

Classroom is a place for students to acquire knowledge. However, students' attention is easily distracted by various multimedia devices, resulting in the low raise-head rate. In order to create a sound learning atmosphere in the future, it is particularly important to study the rate of taking lessons. Firstly, each student in the classroom is identified by the face recognition technology which based on deep residual network. Active Appearance Model is used to estimate every student's head posture to decide whether its pitch angle is less than the threshold value, and then the individual rate of taking lessons and the rate of taking lessons of the whole class are calculated. Finally, experiments are carried out to verify the accuracy of the calculation method.

Keywords

Face recognition;Active Appearance Model;Head Pose Estimation; the rate of taking lessons.

1. Introduction

Traditionally, the rate of taking lessons is based on manual statistics, which is time-consuming and laborious. The method of image processing is used to calculate the rate of taking lessons directly by analyzing the video captured by the classroom camera. This method is fast and efficient, once applied to practical teaching, it will bring great convenience to educators. Therefore, with the continuous development of intelligent monitoring equipment, aiming at the application requirements of efficient classroom teaching, this paper proposes a method of calculating the rate of taking lessons based on face recognition and head pose estimation.

2. Face Recognition Technology

Face recognition system mainly includes two parts[错误!未找到引用源。], namely, face detection and face verification. Face detection is to determine whether there is a face in a picture, and if there is a face, find the location of the face and circle it out. Face verification is to determine whether the face in two pictures is the same person.

Deep learning has been more and more widely used in computer vision, and its effect has been greatly improved compared with traditional methods. ResNet[错误!未找到引用源。] proposed by He Kaiming et al. introduced the residual network structure. Through the residual network, the depth of network layer design can reach 1000 layers, and the final network classification effect is very good.

Dlib is a C++ toolkit that contains machine learning algorithms commonly used to solve practical problems. It has been widely used in industry and academia, including various VR devices, automation devices, embedded devices, mobile phones and so on. Dlib is open source and can be used free of charge in any application. In this paper, we use the face detection method provided in Dlib, and use the depth residual network to realize real-time face recognition.

2.1 Face Detection

The principle of face detection algorithm is to use windows with different sizes and locations to slide on the image to be detected, and then to judge whether there is a face in the window. Before in-depth learning, the common method is feature extraction + ensemble learning classifier, such as popular Haar feature + Adaboost cascade classifier. The method of HOG (Histogram of Oriented Gradient)[3] + regression tree is used in Dlib. Dlib also uses convolutional neural network for face detection, and its effect is better than the integrated learning method of HOG.

2.2 Face Checking

Verification is what we often call "classification". After the camera detects the face in the image, the system compares the face with the locally preserved face image to determine who the face is. The classification is divided into two parts:

(1) Feature vector extraction, this paper uses the interface of ResNet model trained in Dlib, which generates a 128-dimensional face feature vector.

(2) Distance matching. After obtaining the feature vectors, the Euclidean distance is used to match the feature vectors with the local face feature vectors, and the nearest neighbor classifier is used to get the label of the samples.

3. Head Pose Estimation

In the field of computer vision, head pose estimation algorithm has great practical value[4]. As shown in Figure 1, there are three parameters describing the direction of the head[5]: yaw of horizontal rotation, Pitch of vertical rotation and Roll of left and right rotation. Normally, the range of head movement of normal adults is - 40.9 to 63.3 for Roll, 60.4 to 69.6 for Pitch and - 79.8 to 75.3 for Yaw.

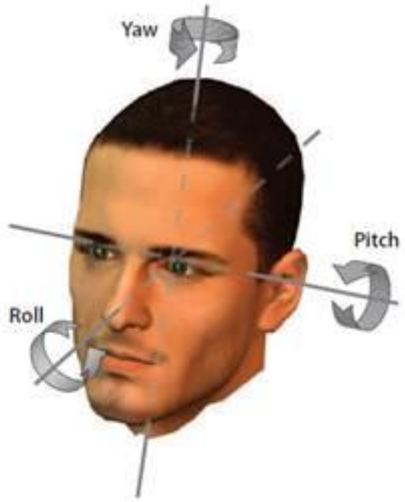

Figure 1. Head Pose Definition Diagram

3.1 Head Pose Estimation Method

Active Appearance Model (AAM)[6] is a typical flexible model. AAM can approach face image iteratively and gradually, so it can better overcome the influence of various detection errors and obtain relatively accurate head posture. AAM has a great advantage in speed, and is generally suitable for occasions with large amount of calculation. College students usually have 30 to 50 students in class, so the head posture estimation using AAM algorithm can meet the requirements of this paper.

AAM is based on Active Shape Model (ASM)[7]. It further builds statistical model of texture (shape-independent image obtained by deforming face image into average shape). Then it integrates shape and texture statistical models into an appearance model.

3.1.1 ASM

ASM is an algorithm based on Point Distribution Model (PDM).The basic principle of PDM is that the geometric shape of an object can be represented by a shape vector formed by the coordinates of several key feature points in series. Figure 2 shows a face image with 65 key feature points calibrated.

Figure 2. Face labeling

The ASM algorithm generates the shape model of the object by parameterized sampling shape, and then uses PCA (Principal Component Analysis)[8]method to establish the motion model describing the shape feature points. The specific steps of ASM are as follows:

(1) Firstly, K key feature points of each training sample are recorded manually.

(2) Then a shape vector is composed of K key feature points calibrated in a graph.

$$\alpha_i = \left(x_1^i, y_1^i, x_2^i, y_2^i, \ldots, x_k^i, y_k^i\right), i = 1,2,\ldots,n \quad (1)$$

Among them, $\left(x_j^i, y_j^i\right)$ represents the coordinates of the jth feature point on the 1st training sample, n represents the number of training samples, and N training samples constitute n shape vectors.

Finally, the shape vector is processed by PCA.

3.1.2 AMM

The difference between AAM and ASM is that AAM not only uses face shape information, but also calculates face texture information to find the relationship between shape and texture. Texture information generally refers to the pixels of the object. AAM is mainly divided into two parts: AAM modeling and AM matching calculation. AAM modeling is used to build the active appearance model of objects. The so-called appearance model refers to the face model built on the basis of ASM shape model combined with the texture information of the extracted target face. The steps of AAM algorithm are as follows:

(1) Firstly, PCA method is used to describe the dynamic changes of shape feature points, which are used to display the positions between face feature points.

(2) Secondly, an energy function is defined by the mean square deviation between the sample of AAM model and the input image, which is used to evaluate the matching degree of AAM. In the process of face location and matching, the linear expression of the model is combined with the effective matching algorithm to change the parameter group of the model, so as to control the position change of the shape feature points to generate a new AAM model instance.

(3) Finally, the parameters of the model are updated by using the energy function values obtained before, and the energy function is minimized iteratively and matched between the model instance and the input image. The final position of shape feature points describes the position of feature points in the current face image. The AAM algorithm has developed from the original gradient descent algorithm to the linear hypothesis based matching calculation, and then to the reverse combination AAM matching algorithm based on Lucas-Kanade[9] algorithm, which has greatly improved both in efficiency and matching accuracy.

4. Experiment

Before the beginning of the experiment, the photos of each student in the classes to be recognized are stored in the local catalogue for face verification. Each face feature vector is labeled in pinyin with student number and name. Using the interface of ResNet model, each photo will be returned to a 128-dimensional face feature vector for face comparison.

Firstly, the video obtained in class captures a frame of picture every second, and then detects the face of the picture.Secondly, if there are faces, we use ResNet's interface to get face feature vectors and compare them with the corresponding features of N identities registered in the database one by one. We can find a feature with the highest similarity value of input features, and compare the maximum similarity value with the preset threshold value.If it is greater than the threshold, the identity corresponding to the feature is returned. In this paper, the similarity threshold is set to 90%. Then, if the identity of the classmate has been identified, the head pose is estimated to calculate the pitch angle and then determine the head-up or head-down state.

5. Conclusion

In this paper, we use the deep residual network face recognition technology to detect the rate of taking lessons, change the original way of using surveillance video human monitoring, intelligent machine to determine whether to bow, and output clear and clear results. On the one hand, it can reduce the waste of manpower and material resources, avoid the influence of classroom discipline on teachers'energy; at the same time, it can be easy to teach, and can get the rate of each person's attendance and the rate of the whole class in the whole teaching activities, and generate the schedule of each student's attendance.

Acknowledgements

College Students` Practice Innovation Training Program of Nuist: 201810300037Z.

References

[1] Wu Jinyi, Zhou Delong. Overview of Face Recognition Methods [J].// Computer Applied Rese arch, 2009, 26 (09): 3205-3209.
[2] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Confer ence on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.
[3] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// IEEE Compute r Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 200 5:886-893.
[4] Zuo Yanchao. Research and application of head posture tracking technology based on computer vision [D].//Northern University of Technology, 2017.
[5] Chen Shuming, Chen Meiling. Review of Head Posture Estimation Technology [J].//Journal of Quanzhou Normal University,2015,33(06): 78-85.
[6] Cootes T F, Edwards G J, Taylor C J. Active appearance models[C]// European Conference on Computer Vision. Springer Berlin Heidelberg, 1998:484-498.
[7] Cootes T F,Taylor C J,Cooper D,Graham J.Active Shape Models-Their Training and Applicatio n[J]Computer Vision and Image Understand-ing, 1995:38-59.
[8] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. C-hemometrics & Intelligent L aboratory Systems, 1987, 2(1):37-52.
[9] Baker S, Matthews I. Lucas-Kanade 20 Years On: A Unifying Framework[J]. International Jour nal of Computer Vision, 2004, 56(3):221-255.