ISSN: 2664-9640

Research on Evaluation Method of Students' Concentration in Smart Classroom Based on Head Pose Recognition

Wenlong Su

School of Computer, Guangdong University of Science and Technology, Dongguan 523083 China)

Abstract

Head pose recognition is one of the long-term research topics in the field of computer vision and pattern recognition, and its application potential is huge. Aiming at the application scenario of head pose in smart classroom, this paper proposes a student focus evaluation method based on head pose recognition. Multi task convolution neural network is used to locate and extract facial feature points, and the human head pose angle is calculated based on PNP algorithm. Using this method, a classroom focus evaluation system is designed, which has certain theoretical guiding significance for classroom teaching and school wisdom classroom research.

Keywords

Head pose recognition; Focus assessment; Smart classroom.

1. Background

How to objectively quantify and evaluate the teaching quality of college classroom and the learning effect of students is not only an important task for college teaching and management, but also the basic requirement of smart classroom and smart campus. In recent years, with deep learning, many breakthroughs have been made in machine vision, target detection and feature extraction. More and more algorithms and models are applied to various classroom behavior detection. This paper proposes a concentration discrimination method based on head posture and a combination of multiple detection methods to evaluate students' classroom concentration and learning effect.

2. Focus Assessment Method

2.1. Basic Steps

This paper adopts a focus discrimination method based on the combination of head posture and facial expression recognition. Firstly, the angle of students' head posture is used as the main index of students' focus discrimination, and then their facial expressions are recognized. Finally, the results of the two indexes are evaluated, and a student focus discrimination system is established. The basic steps of the assessment are as follows:

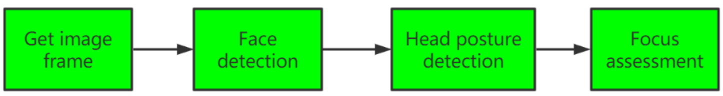

Fig. 1 Basic steps of focus assessment

(1)Get image frame

The fixed camera is used to collect and record the classroom performance of classroom students, and the image frames in the video are extracted for the next face feature detection and focus judgment.

(2)Face detection

ISSN: 2664-9640

Face detection and face alignment are preconditions for accurate estimation of head pose [1]. In this paper, face region detection and face key point detection are put together through multi task convolutional neural network (MTCNN) to detect and locate students' face images.

(3) Head posture detection

After obtaining the facial signs of the face, the head pose is estimated, and the head pose is estimated by using the method of feature point location (PNP, perspective-n-point).

(4) Focus assessment

According to the students' head posture information, combined with expression, posture and other ways, an attention detection system is designed to calculate and evaluate the students' classroom concentration.

Next, this paper will focus on three aspects: head pose detection and calculation, and student attention evaluation method.

2.2. Head posture detection

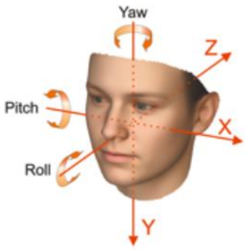

Fig. 5 Head angle evaluation coordinates

After obtaining the facial signs of the human face, the head pose is estimated, and the PNP algorithm is used to calculate the head pose deflection angle. In this method, the Euler angle is used to evaluate the human head pose. Figure 5 shows the human head image taken by the monocular camera, which can be expressed by the Euler angle, namely yaw angle, pitch angle The rotation angle (roll) is three angles to understand the spatial posture of the human head.

The attitude estimation problem faced in this paper is usually the perspective-n-point PNP problem, which aims to find out the attitude of the object for the calibrated camera; Among them, n 3D points of the known object and their corresponding 2D projection in the picture. It is solved according to the transformation relationship matrix between the coordinates of n points of the target in the 3D world coordinate system and the point set corresponding to the projection into the 2D image coordinate system.

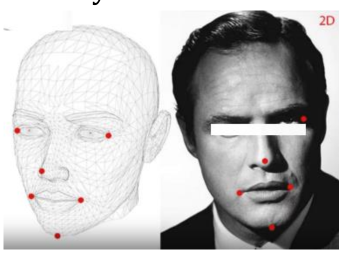

Fig. 6 Correspondence between 2D and 3D pose face key points

In order to calculate the 3D pose of the object in the picture, you need to know the 2D coordinates x and Y in the picture. For the face, the key points of the face are: nose tip, chin, left corner of the left eye, right corner of the right eye, left corner of the mouth and right corner of the mouth. You also need to know the coordinate points of 3D:

nose tip：0.0, 0.0, 0.0

chin：0.0, −330.0, −65.0

left corner of the left eye：−225.0f, 170.0f, −135.0

right corner of the right eye：225.0, 170.0, −135.0

ISSN: 2664-9640

left corner of the mouth： −150.0, −150.0, −125.0

right corner of the mouth： 150.0, −150.0, −125.0

This involves the conversion of world coordinates and camera coordinates, as shown in Figure 7 below:

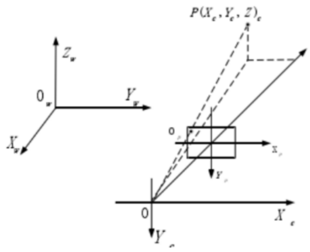

Fig.7 Relationship between world coordinate system and camera coordinate system

$$
\begin{aligned}
s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} 1/dX & 0 & u_0 \\ 0 & 1/dY & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \mathbf{M}_1\mathbf{M}_2\mathbf{X}_w = \mathbf{M}\mathbf{X}_w
\end{aligned} \tag{4}
$$

Where $m1, m2$ are internal and external parameters of the camera.

3. Analysis of experimental results

In order to verify the correctness and effectiveness of pose estimation, the open Biwi Kinect data set [3] obtained from Microsoft Kinect statistics is used to test the accuracy of head pose estimation method. The data set has more than 1500 images and about 20 images and video clips with different attitudes. The experimental hardware environment is Intel Core i9-9920x processor, 3.5 GHz, DDR4 memory 16GB, and the software environment is 64 bit. The visual studio environment under windows10 system uses openCV library.

Firstly, the accuracy of face detection and location is tested. Samples with a certain angle offset can be well detected and marked.

Fig. 8 Face detection results

3.1. Head pose estimation and analysis

The Biwi Kinect data set is used to calculate the head attitude rotation angle of the sample to obtain the estimated values of pitch angle, yaw angle and rotation angle. The difference between the estimated value and the marked value in the data set is taken as the error of the estimated value. The following figure shows the annotation results of some data sets.

ISSN: 2664-9640

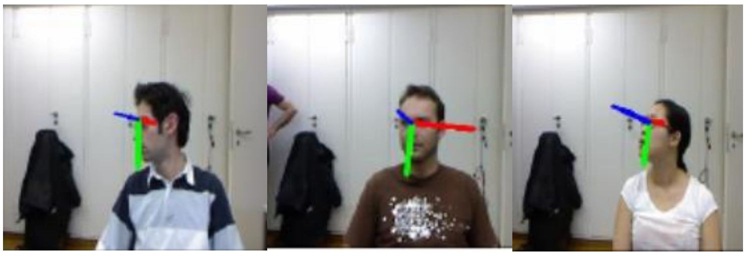

Fig. 9 Head pose estimation results

3.2. Focus assessment and analysis

For the above head posture evaluation methods, the author develops a set of system to evaluate the concentration. In order to verify the detection effect of this system and method, the author obtains part of the classroom learning process from the Internet to test the students' daily learning classroom behaviors, such as listening carefully, looking down at their mobile phones, speaking and looking around.

Fig.10 Primary school classroom concentration test

Fig. 11 College classroom concentration test

The system labels and counts the behaviors with large head posture deviation and long time in the existing data set. However, due to the small number of samples, this method still has less detection and false detection when there are a large number of students, and the future research needs to be further improved.

4. Summary

This paper presents a method to evaluate students' attention based on the combination of head pose analysis and multiple feature recognition. Firstly, the multi task convolution neural network is used to locate and extract the facial feature points, and the human head pose angle is calculated based on PNP algorithm. Finally, the focus evaluation method is used to evaluate the students' focus. Through the experimental test, the proposed method can effectively evaluate students' classroom concentration, and has a certain theoretical guiding significance for classroom teaching and school wisdom classroom research.

References

[1] Ranjan R, Patel V M, Chellappa R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41(1):121-135.

[2] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

[3] Papazov C, Marks T K, Jones M. Real-time 3D head pose and facial landmark estimation from depth images using triangular surface patch features[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15) . Boston: IEEE, 2015: 4722-4730.