Hindawi
Scientific Programming
Volume 2021, Article ID 7183278, 10 pages
https://doi.org/10.1155/2021/7183278

Research Article

Innovative Research on the Development of Online Education Mode of Internet Thinking Based on the Discrimination of Learning Attention under the Analysis of Head Posture

Su Song and Fangzheng Wang

Public Basic College, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu 224000, China

Correspondence should be addressed to Fangzheng Wang; 11406@jsmc.edu.cn

Received 13 August 2021; Revised 18 September 2021; Accepted 24 September 2021; Published 30 November 2021

Academic Editor: Bai Yuan Ding

With the rapid development of Internet technology and the popularity of 5G and broadband, online education in China, especially mobile online education, is in full swing. Based on the development status of online education in China, this paper analyzes the innovative application of learning attention discrimination based on head posture analysis in the development of online education mode of Internet thinking. Learning attention is an important factor of students' learning efficiency, which directly affects students' learning effect. In order to effectively monitor students' learning attention in online teaching, a method of distinguishing students' learning attention based on head posture recognition is proposed. In the tracking process, as long as the head angle of the current frame is close to the head angle of the key frame in a certain scale model, the visual angle apparent model can reduce the error accumulation in large-scale tracking. A Dynamic Bayesian Network (DBN) model is used to reason students' Learning Attention Goal (LAG), which combines the relationships among multiple LAGs, multiple students' positions, multicamera face images, and so on. We measure the head posture through the similarity vector between the face image and multiple face categories without explicitly calculating the specific head posture value. The test results show that the proposed model can effectively detect students' learning attention and has a good application prospect.

1. Introduction

Traditional education can no longer meet the needs of modern education. The cost of computer production is declining, and the modern education mode with computer as the medium is more and more widely used in education. China's online education has been on the rise, making many offline education and training institutions vigorously develop online education, and even many Internet companies that did not pay attention to the field of education began to enter the field of online education [1]. China's online education is so full of vitality that everyone applauds it, but its development cannot be smooth sailing, and it is bounded by many problems. How to solve these problems is the basic requirement to promote the further development of online education.

Since the explosive birth of information transmission technology in the 19th century, people have never stopped exploring the application of various new technologies in the field of education. Root et al. define online education management system as "a software application for classroom education and after-school training, involving educational administration, information transmission, report generation and learning effect tracking" [2]. However, Lockee et al. define online education management system as "the integration of networked tools to support online learning" [3]. Zhang J. and Zhang F. discuss the development trend and form of online higher education through the research based on relevance theory, and emphasize that Massive Open Online Courses (MOOCs) are not only an online classroom, but also the interaction between teachers and students [4]. Fehr introduces the practical application of MOOCs,

analyzes their existing problems, and holds that "There is no doubt that the development of MOOCs is of great significance to higher education all over the world [5]. It provides opportunities for the online development of traditional higher education in a flexible and convenient way, but it cannot be considered that virtual education can replace traditional education. In disciplines such as medicine and architecture, the advantages of traditional higher education model are still incomparable." By analyzing the development data of online higher education in the United States from 2002 to 2014, Mcauliffe et al. preliminarily explore the development direction and path of online higher education in China, and form its own theoretical model. In a word, scholars at home and abroad started their research on online higher education late, mostly focusing on analyzing foreign literature. Even if a few models are put forward, their feasibility and suitability are open to question [6].

Attention in visual learning is related to head posture and eye sight direction. The research shows that [7, 8], in many cases, it is enough to analyze students' Learning Attention Goal (LAG) through head posture. Because students are not used to staring at a certain goal with slanting eyes for a long time, they will turn their heads to face the goal. Therefore, this paper uses the head posture to analyze the LAG of students, so we propose a Dynamic Bayesian Network (DBN) model to deduce the LAG of students. The head pose is measured by the similarity vector between the face image and multiple face categories without explicitly calculating the specific head pose value. The observation of probability model includes face image and face position under multicameras. We collected test data in the teaching environment, and the experimental results show that our model is effective.

Structure of This Paper. The first section mainly introduces the research background, significance and main innovations of this paper. Section two presents a summary of related research. This paper introduces the research status of the key technologies (head pose estimation, face feature points, and attention recognition). In section three, the discrimination process of learning attention based on head posture analysis is mainly studied. Section four analyzes and discusses experimental results. Finally, the paper summarizes the full text, analyzes the existing problems in the current methods, and looks forward to the future research directions.

2. Related Work

In recent years, more and more researchers began to study the problem of visual LAG recognition. Bce et al. study the problem of students' LAG recognition in a small round table environment, in which an omnidirectional camera is placed on the conference table [9]. Later, they studied the method of identifying LAG in the environment of multiple remote cameras [4]. Hamrah et al. also study the problem of LAG recognition in different conference environments [10]. However, in the meeting environment, students mainly sit in fixed seats, and their bodies do not move much. Zhou et al. also monitor students' LAG in outdoor environment, but the range of their head posture is limited to face posture close to the front. These tasks mainly deal with the analysis of multiple LAGs in fixed positions or single LAGs in multistudent positions [11]. However, our application environment includes multiple student locations and multiple LAGs. Zhao proposed a head pose recognition algorithm based on template matching technology. For each recognition object, multiple head images in different poses are extracted as sample images, and each image is marked with corresponding pose parameters [12]. Yeager et al. found that in most cases, students' attention target behavior can be obtained by analyzing the head posture angle [13]. Yfka et al. detect face feature points through random cascade regression tree, and use N-point perspective algorithm to estimate head posture, thus realizing the visualization of students' learning attention [14].

Hirata and Kusatake establish a human-computer interaction system by using the learning attention detection model, which can effectively judge the position and head posture of the target person in the multiperson environment [15]. Guo et al. study the problem of attention target recognition in different conference environments. However, in the conference environment, users mainly sit in fixed seats, and their bodies do not move much [16]. Lu and Yanmin study the problem of attention target recognition in outdoor environment. They mainly analyze whether passers-by watched posters on the wall [17]. Xiao et al. introduce the run-length matrix of binary pattern into the random feature selection of random tree, which improves the classification ability of single decision tree and achieves better recognition rate for multiclassification discrete head pose estimation [18]. Ren et al. define the distance between eyes and screen and the range of head posture by calibrating the system, and use a single light source to track the gaze direction of eyeballs [19]. In the literature [20], three-dimensional faces are recognized by combining stereo vision information such as rotation and pitch, and the learning attention direction of eye sight can be accurately tracked through the details of eye images.

Through the study of the abovementioned related literatures, it is found that most of the research methods have certain requirements for equipment. Because of the low resolution of the individual's face in the classroom scene, and the illumination change, occlusion, and large posture change in the environment, it is very difficult to learn attention recognition. Therefore, we adopt a noninvasive learning attention recognition method based on the head posture direction, and recognize the LAG of many people at the same time in the large classroom scene.

3. Research Method

3.1. Analysis of Students' Learning Attention. The main purpose of this study is to distinguish students' learning attention, and to determine the direction of students' learning attention by estimating whether students' eyes are concentrated in the blackboard area according to their head posture.

As shown in Figure 1, when students' eyes are focused on a certain point in the blackboard, such as P_1, students are considered to be focused on learning. On the contrary, when students' eyes deviate from the blackboard area for a long time, such as P_2, students are considered to be distracted in learning. Under normal circumstances, people are not used to looking at the target they pay attention to with oblique eyes. Therefore, the rotation direction of head posture can be regarded as the line of sight of students approximately to analyze students' learning attention.

According to the classroom environment, a coordinate system is established, which takes the center point on the blackboard as the coordinate origin, the horizontal right direction of the origin as the X-axis positive direction, the vertical origin direction as the Y-axis positive direction, and the vertical XY plane pointing to students as the Z-axis positive direction. According to the students' head line of sight reaching the edge of blackboard, it is regarded as the criterion of students' abnormal behavior, as shown in Figure 2.

$\alpha_1, \alpha_2, \beta_1, \beta_2$ is the threshold of abnormal head deflection of students, and α_1, α_2 is the rotation range of students in θ_{Yaw} direction; β_1, β_2 is the rotation range of students' θ_{Pitch} direction. When the rotation range of the head exceeds the threshold, it can be considered that the students' sight is outside the blackboard area, and it is judged that the learning attention is distracted.

Assume that the blackboard has a length of h, a width of d, and a head center coordinate of $F(x, y, z)$. When the students sit in the first row of the classroom and look at the left and right edges of the blackboard at points B and D shown in Figure 1, it is the maximum rotation range of the students' heads in the θ_{Yaw} direction, which is written as

$$\begin{aligned} \alpha_1 &= -\arctan\left(\frac{h}{z}\right)\alpha_2 \\ &= \arctan\left(\frac{h}{z}\right). \end{aligned} \tag{1}$$

When the student sits at point C and looks at the upper and lower edges of the blackboard, it is the maximum rotation range of the student's head in the θ_{Pitch} direction, which is written as

$$\begin{aligned} \beta_1 &= -\arctan\left(\frac{y}{z}\right)\beta_2 \\ &= \arctan\left(\frac{\mathrm{d} - y}{z}\right). \end{aligned} \tag{2}$$

According to the actual teaching environment, assuming that the center point of the head coincides with the eyes, and the height of the adult students' eyes from the ground is 1.3 m, the head rotation range of the students is determined to be θ_{Pitch} direction $[-7^\circ, 28^\circ]$, θ_{Yaw} direction $[-48^\circ, 48^\circ]$.

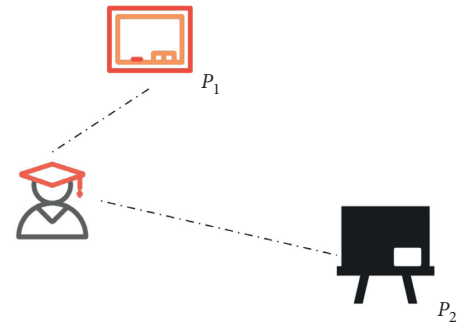

Figure 1: Students' learning attention.

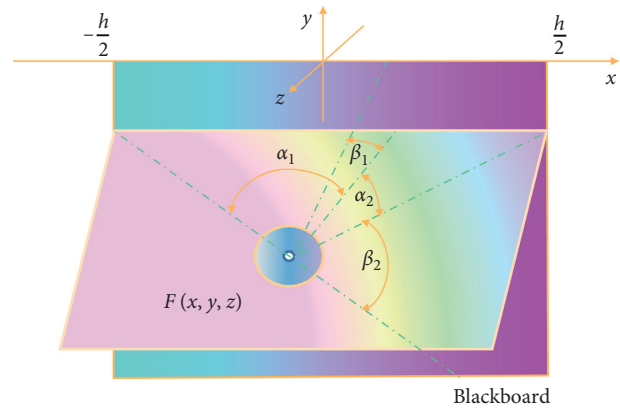

Figure 2: Head deflection range.

3.2. Discrimination Process of Learning Attention Based on Head Posture Analysis

3.2.1. Visualization of Students' Learning Attention Based on Head Pose Estimation. Combining the advantages of eye tracker and single camera learning attention analysis system, this paper proposes a visual analysis method of students' learning attention based on head pose estimation of single image, and constructs a corresponding visual analysis system of students' learning attention [21]. In this paper, the front camera installed in the middle position above the blackboard is used to record the students' lectures, and then the method shown in Figure 3 is used to estimate the students' head posture. Finally, the students' eyes are projected to the teacher's lecture video recorded by the rear camera by mathematical deduction.

As shown in Figure 3, this method mainly consists of the following six steps.

(1) Acquisition of data (video frame): the classroom teaching video is acquired by LifeCam camera of Microsoft 1080p, and the video frames are separated.

(2) Camera calibration: in order to improve the accuracy of head pose recognition, it is necessary to use a convenient and accurate calibration method to calibrate camera parameters.

(3) Face detection: use the disclosed face detector to detect faces from video frames [22].

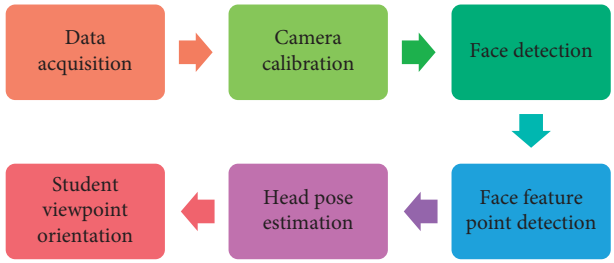

Figure 3: Working principle and algorithm flow of visual analysis of students' learning attention.

(4) Face feature point detection: this paper uses random cascade regression tree to obtain the coordinate information of 19 face feature points, which is used to provide two-dimensional information in solution.

Students' viewpoint positioning: according to the rotation and translation matrix information of head posture, the students' viewpoint is projected to the teacher's lecture video shot by the rear camera by using the transformation relationship of spatial coordinates, so as to realize the visual display of students' learning attention.

Combining the oculomotor and the single-camera learning attention analysis system, the random cascade regression tree is used to locate the face feature points, and a rigid model obtained by statistical measurement is introduced as the 3D face approximation. The students' eyes are projected onto the video images taught by teachers, and the visual analysis of students' learning attention is realized.

3.2.2. Visual Angle Apparent Model

(1) Key Frame Adjustment. After tracking the current frame t, in order to calculate the attitude parameters of $t+1$ frame, the current frame t becomes a new key frame in the model, and the expectation and covariance of X should be expanded accordingly. Because x_{t+1} is unknown at this moment, the expectation and covariance of X are extended as follows:

$$E_X^{t+1} = \left[0; E_X^t\right],$$
$$\left[\left(\sigma_X^{t+1}\right)^2\right]^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \left[\left(\sigma_X^t\right)^2\right]^{-1} \end{bmatrix}. \tag{3}$$

Here, E_X^t and $(\sigma_X^t)^2$ represent the expectation and covariance of X after the tracking of the current frame t is completed, and E_X^{t+1}, $(\sigma_X^{t+1})^2$ represents the extended expectation and covariance of X.

To reduce the number of key frames in the model, only one key frame is selected from each perspective. If the attitude parameters of one key frame are very close to those of other key frames in the model, the key frame will be removed from the model [23]. Since the previous frame $t-1$ is always used as the reference frame of the current frame t, the previous frame $t-1$ is likely to be removed from the model after the tracking of the current frame t is completed. Let the largest one of the three rotation angles of the previous frame be ω_{t-1}, and when the corresponding angle ω_i of a certain key frame i in the model satisfies the condition

$$\left\|\omega_{i-1} - \omega_i\right\| \le \tau, \tag{4}$$

remove the previous frame $t-1$ from the model, where τ is the threshold determined by the number of key frames in the model. The removal process is completed by deleting the corresponding rows and columns in E_X^t, $(\sigma_X^t)^2$.

(2) *Multiscale Visual Angle Apparent Model.* When the head movement range is small, the visual angle apparent model can effectively reduce the tracking error. Visual angle apparent model when the head moves in a small range is also called single-scale visual angle apparent model, but when the head moves in a large range, it will exceed the effective range of single-scale visual angle apparent model. There are two main reasons for this phenomenon: first, the scale transformation of the head image is large when moving in a large range; second, the apparent model parameters themselves contain large error accumulation due to long-term movement, which leads to the tracking result deviating from the true value [24]. Therefore, each visual angle apparent model is only valid within a certain range.

When the head movement range is large, multiple visual angle apparent models can be used to cooperate with each other to reduce the error accumulation caused by large-scale movement. In the specific implementation, the effective range of each visual angle apparent model is defined as a space neighborhood around its initial frame head position.

Let m represent the initial frame of the current apparent model, D_m represent the distance between the head of the initial frame and the camera, and the distance D_t between the head of the current frame and the camera satisfies the condition

$$\left\|\mathrm{D}_t - \mathrm{D}_m\right\| \ge \eta. \tag{5}$$

The tracking method will generate a new apparent model, where η is a predefined threshold. The current frame t will become the initial frame of the new model. Multiple visual angle apparent models which are continuous in space are called multiscale visual angle apparent models.

On the one hand, the multiscale visual angle apparent model can effectively solve the error accumulation problem of tracking large-scale motion, especially large-scale forward and backward motion. On the other hand, because each scale visual angle apparent model is only responsible for tracking motion in a small range, it is beneficial to the application of the model. For example, when the head leaves the camera view and re-enters, the head posture can be quickly recovered by using the multiscale visual angle apparent model.

3.2.3. DBN Model for Analyzing Visual LAG. We propose a DBN model to reason students' attention goals. Reasoning students' attention goals by calculating the maximum posterior probability. The model integrates the relationships among multiattention targets, multistudent positions, and multicamera face images, and conducts joint reasoning.

(1) Overview of Models. A Dynamic Bayesian Network (DBN) model for analyzing LAG proposed in this paper [25] is shown in Figure 4.

In the model, we have the following:

Implicit variable F_t represents the LAG of students at time t, and its value is M possible lags.

Implicit variable C_t^i represents the pose category of the face image shot by camera i at time t, and its value is K face pose category.

The observation variable Z_t^i represents the face image taken by the camera i at time t.

The observation variable L_t^i represents the horizontal position of the face image shot by the camera i at time t.

The model combines multicamera information to analyze LAG more accurately. For example, when students stand in area 2 and area 3, both cameras can capture students. In some cases, the images obtained by a single camera may not accurately analyze the LAG of students. At this time, through the information of another camera, it may be easier to analyze students' goals.

(2) Description of Each Part of the Model. According to the probability model, the joint probability distribution among all variables can be written as follows:

$$P\left(F_{1:T}, C_{1:T}^{1:R}, L_{1:T}^{1:R}, Z_{1:T}^{1:R}\right) = \prod_{t=1}^{T} P\left(F_t|F_{t-1}\right) \prod_{i=1}^{R} P\left(C_t^i|C_{t-1}^i, L_t^i, F_t\right) P\left(Z_t^i|C_t^i\right). \tag{6}$$

Among them, $P(F_t|F_{t-1})$ stands for the purpose of transition probability matrix between different LAGs at adjacent moments, which is to enhance time smoothness.

$P(C_t^i|C_{t-1}^i, L_t^i, F_t)$ represents the probability dependence of face pose category C_t^i on LAG F_t, face position L_t^i, and face pose category C_{t-1}^i at the previous moment. This is the core of this model, which describes the probability dependence among multiple LAGs, multiple student positions, and faces obtained by multiple cameras.

$P(Z_t^i|C_t^i)$is the likelihood of face observation. The face posture class C_t^i is known, and the likelihood represents the probability that the face observation Z_t^i is generated by the face posture class, as shown in

$$P\left(Z_t^i|C_t^i = k\right) = \frac{1}{\Lambda} \exp\left(\frac{-d^2\left(Z_t^i, M_k\right)}{\sigma^2}\right). \tag{7}$$

Here, Λ is the normalization factor, M_k represents the image subspace $C_t^i = k$ of the face pose category, and $d^2(Z_t^i, M_k)$ represents the distance from the face image to the image subspace, such as the reconstruction error when the image is projected to the subspace.

(3) Model Reasoning. The analysis problem of LAG is regarded as the reasoning problem of probability model. Given the observations Z and L, we hope to deduce the hidden variables $F$and C. That is, our objective function is to maximize the following joint probability density distribution:

$$(\widehat{F}, \widehat{C}) = \arg\max_{F,C} p(F, C, Z, L). \tag{8}$$

We use the approximate reasoning algorithm proposed in [26] to minimize the cost function "free energy." The free energy uses a simple probability density distribution $Q(h)$ to approximate the true posterior probability density distribution $P(h|v)$ where h and v represent hidden variable (F, C) and observed variable (Z, L), respectively. Then, $Q(h)$ is used to calculate the objective function with the following formula:

$$Q(h) = \prod_{t=1}^{T} Q\left(F_t|F_{t-1}\right) \prod_{i=1}^{R} Q\left(C_t^i|C_{t-1}^i, F_t\right). \tag{9}$$

According to the method of [11], free energy can be written as

$$E = \int_h Q(h) \mathrm{In} \frac{Q(h)}{P(h, v)}. \tag{10}$$

Given the state of hidden variables F and C at time $t-1$, by minimizing E, we can get

$$\frac{\partial E_t}{\partial Q\left(C_t^i|C_{t-1}^i, F_t\right)} = 0 \Rightarrow Q\left(C_t^i|C_{t-1}^i, F_t\right) \propto P\left(C_t^i|C_{t-1}^i, L_t^i, F_t\right) P\left(Z_t^i|C_t^i\right), \tag{11}$$

$$\frac{\partial E_t}{\partial Q\left(F_t|F_{t-1}\right)} = 0 \Rightarrow Q\left(F_t|F_{t-1}\right) \propto P\left(F_t|F_{t-1}\right) \prod_{i=1}^{R} \prod_{C_{t-1}^i=1}^{K} \prod_{C_t^i=1}^{K} \left(P\left(C_t^i|C_{t-1}^i, L_t^i, F_t\right) P\left(Z_t^i|C_t^i\right)\right)^{Q\left(C_{t-1}^i|F_{t-1}\right)}. \tag{12}$$

In formula (12), it can be calculated as follows:

$$Q\left(C_{t-1}^i|F_t\right) = \int_{F_{t-1}, Q_{t-1}^i} Q\left(C_t^i|C_{t-1}^i, F_t\right) Q\left(C_{t-1}^i|, F_{t-1}\right) Q\left(F_{t-1}\right). \tag{13}$$

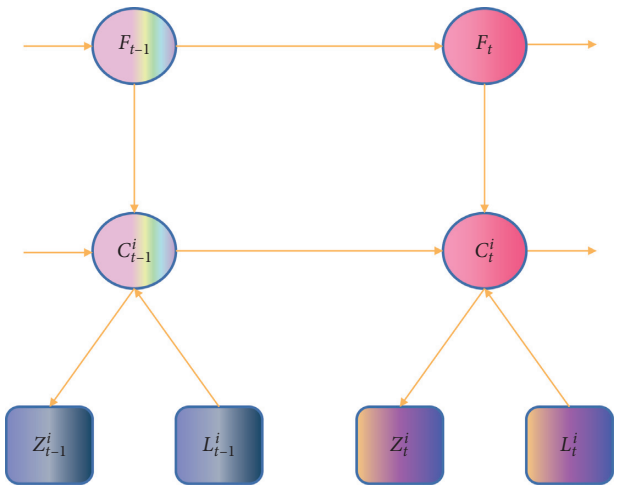

FIGURE 4: DBN model. The subscript t indicates the frame number, and the superscript i indicates the camera number.

Then, the probability density distribution can be calculated by iterative method:

$$Q(F_t) = \int_{F_{t-1}} Q(F_t|F_{t-1})Q(F_{t-1}),$$
$$Q(C^i_t) = \int_{F_t} Q(C^i_t|F_t)Q(F_t). \quad (14)$$

Finally, the LAG with the highest probability is the final reasoning result of the model.

$$\text{LAG} = \arg\max_{F_t} Q(F_t). \quad (15)$$

4. Analysis and Discussion of Experimental Results

4.1. The Efficiency Test of Visual Analysis of Learning Attention in This Paper. This paper transforms a conference room into a small classroom. The subjects sat 6 m in front of the blackboard and looked at the teachers who were writing the edition books. A front camera was installed directly above the blackboard to monitor the subjects' learning status. The rear camera is installed behind the subjects to monitor the teaching situation of teachers. In order to ensure the test accuracy, the camera was calibrated during installation.

Based on the captured video frame images, using the head pose estimation method proposed in this paper, the three-dimensional angle information and three-dimensional coordinate axis (total six-dimensional information) of the subject's head pose are calculated, in which the six-dimensional information is marked at the tip of the nose for convenience of display. Using the derived visualization method of students' learning attention based on head posture, the physical position of the gaze point of the subject can be calculated and marked on the captured video frame image.

The 1080P high-definition video taken by the front/rear camera is visually analyzed for single-person learning attention, and the video duration is 2 minutes. At the same time, in order to test the parallel acceleration performance of the visual analysis method of learning attention in this paper, 1~4 physical threads of i5-4570 (4-core) CPU are used to process two videos in series/parallel. The experimental system uses 32 GB memory and TitanX graphics card (12 GB memory) to ensure that memory and graphics card do not become the performance bottleneck of hardware system.

Firstly, the two videos with a duration of 2 min are subjected to five steps, including frame image reading, face detection, face feature point detection, head pose estimation, and learning attention visualization, and the running time of each step and the total running time of the whole algorithm are obtained, respectively. In this paper, we first use a single thread to get the running time of serial computing. Based on this, we use 2~4 threads to get the running time of parallel computing to observe the effect of parallel acceleration (see Figure 5).

The following conclusions can be drawn from all the data listed in Figure 5.

When one thread is used for serial processing of visual analysis of students' learning attention, the processing speed of 1080P single face video is very slow, and it takes 711.58 ms to process one frame image.

Comparing the running time of each step of serial processing for visual analysis of students' learning attention, we can find that the two steps of face detection and face feature point detection are the most time consuming, which are 541.33 ms and 122.36 ms, respectively. Therefore, the key to improve the visual analysis speed of learning attention lies in how to reduce the time consumption of face detection and face feature point detection.

Using 2~4 threads for parallel computation of visual analysis of learning attention, it is found that parallel computation can effectively reduce the time-consuming of face detection, but for the other 4 steps (reading frames, face feature point detection, head pose estimation, and learning attention visualization), the acceleration effect is not obvious. In order to analyze the attention of 25–50 students in classroom teaching in real time by using 4 KB high-definition cameras in practical applications in the future, large-scale parallel processing must be carried out by computing clusters with many node machines. This method uses a single image for head pose estimation, which will be completely suitable for parallel processing of many node machines.

4.2. Head Posture Tracking Experiment Using the Visual Angle Apparent Model

4.2.1. Experiment 1. Evaluate the tracking results of visual angle apparent model when the body moves back and forth in a large range. In the experiment, a video sequence was recorded at a speed of 12 Hz by using a Digiclops stereo camera, in which the subjects moved from a position about 0.5 m away from the camera to a position about 1.5 m away from the camera along the z-axis. During the movement, the subjects constantly changed various head postures, and the rotation angle of their heads around three coordinate axes ranged from −45° to 45°. According to the effective range of

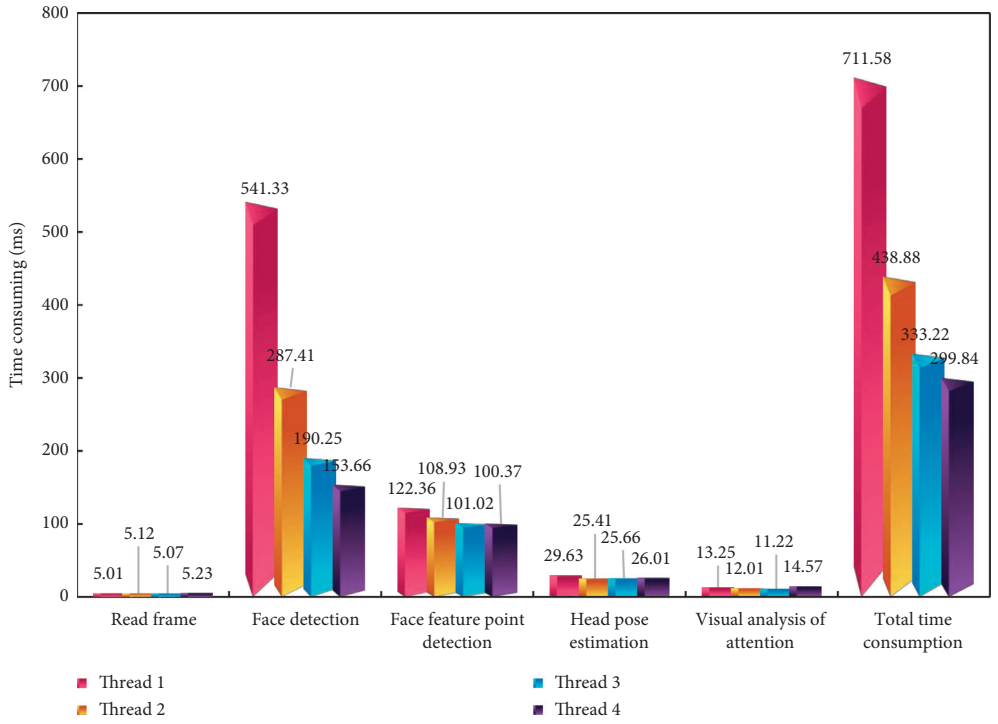

Figure 5: The average acceleration effect of multithreading and the average time consumption of each step of operation in this paper.

each scale visual angle apparent model, the tracking method constructs three scale visual angle apparent models, and their initial frames are about 0.5, 0.9, and 1.3 m away from the camera, respectively.

4.2.2. Experiment 2. Firstly, the tracking results of the proposed method are tested when the head leaves the camera view and then re-enters. In the experiment, another video sequence was recorded at a speed of 12 m by using the Digiclops stereo camera, in which the subjects left the camera's perspective when they were about 1.2 m away from the camera, and then re-entered the camera's perspective when they were about 0.8 m away from the camera.

4.2.3. Experiment 3. In order to further evaluate the performance of the proposed method, the tracking error of this method was measured in Experiment 3. The real head posture parameter data is obtained by using the motion sensing sensor pciBIRD. Three video sequences were recorded in the experiment, and the head pose parameters of each frame were obtained by pciBIRD. All three video sequences were recorded at a speed of 12 Hz, with an average length of 1020 frames. During recording, the motion of the tracked object was similar to that of Experiment 1, and the parameter settings were the same as those of Experiment 1.

The average tracking error (average error of three angles) using video sequence one is shown in Figure 6, in which only the tracking error of the tracked object moving from about

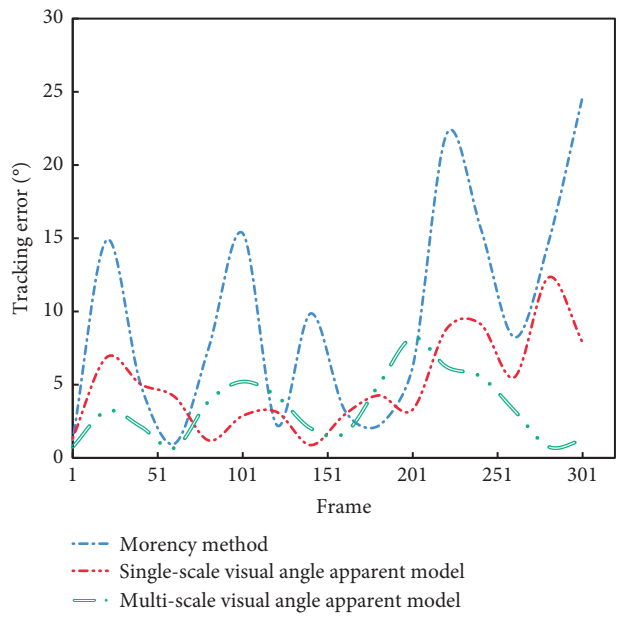

Figure 6: Error accumulation in tracking process.

0.9 m away from the camera to about 1.3 m away from the camera is displayed, with a total length of 320 frames.

Therefore, this method can accurately track attitude parameters (4° root mean square error), while Morency's method has a larger error (8° root mean square error), and the maximum error can reach 20°. Figure 6 also shows the error when using the single-scale visual angle apparent model. It can be seen that when the head movement exceeds the effective range of the single-scale visual angle model, the tracking error is larger.

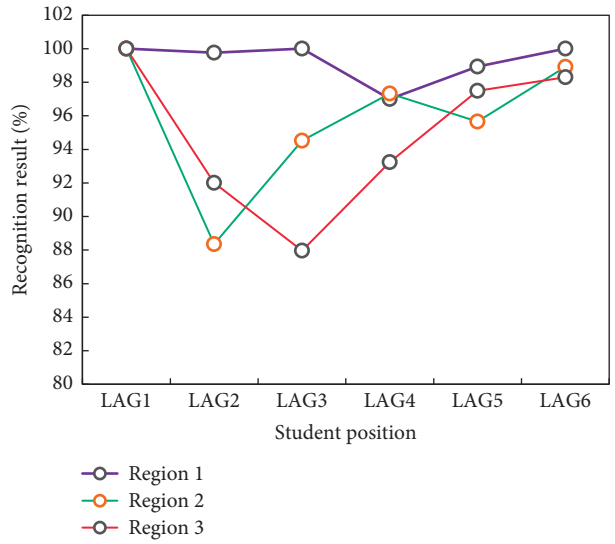

Figure 7: LAG recognition result: the training set and the test set are the same.

4.3. LAG Recognition. We evaluate the validity and extensibility of the model through two sets of experiments. In the first group of experiments, we used the data of all 8 people for training and testing. In the second group of experiments, the method of row-by-row cross-validation was adopted, that is, training with the data of 7 people and testing with the data of another person. A total of 8 rounds were run, and each person's data was taken as the test set. Figures 7 and 8 are the results of two groups of experiments, respectively. The percentage in the figure is the recognition accuracy obtained by dividing the number of correctly recognized frames by the total number of frames. Accuracy evaluation is only performed on manually marked video clips. In the second group of experiments, we only consider the results of the test set video, and give the average results of 8 rounds of experiments.

As can be seen from the above figure, the result is very good. This is because the students in the training set and the test set are the same. The results of the second group of experiments are not as good as those of the first group. This is understandable, because the students in the test set have not appeared in the training set, and the appearance and illumination of different students are quite different. It can be seen that when students stand in area 1 and watch LAG4, 5, and 6, the accuracy is not very high. They often do not mistakenly identify as adjacent LAG. This is because the distance between LAG4, 5, and 6 is relatively short, and they are far away from Area 1. Therefore, when students look at these different targets, they often only turn their heads very slightly, and sometimes they only turn their eyeballs instead of their heads. If these conditions are ruled out, the experimental results are acceptable considering the difficulty of the data captured in real scenes.

4.4. Discriminant Analysis of Students' Learning Attention. Based on the above analysis, the students' learning attention is analyzed and studied by the criterion of

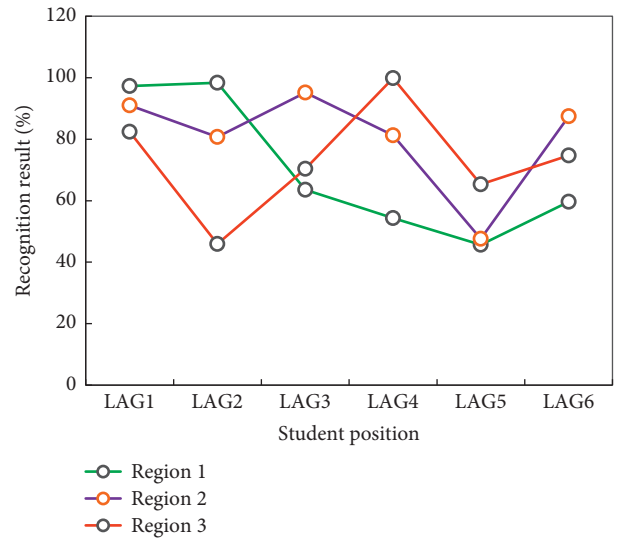

Figure 8: LAG recognition result: cross-validation by row one method.

students' learning attention. In order to verify the detection effect of this method, it is designed that learners imitate students' traditional classroom learning process, and test students' daily learning classroom behaviors such as irregular listening carefully, looking down at their mobile phones, and looking around. The following steps were taken: selecting a high-definition camera with 12 million pixels as an acquisition tool, fixing the camera 2 m in front of the learner, detecting the learning attention of the learner by analyzing the sampled video, and recording the learning process of the learner within 60 s under ordinary illumination.

Implementation process of the algorithm: when students sit in front of the camera, the camera will record the students' learning situation, and then each frame image of the students' learning process will be detected by the

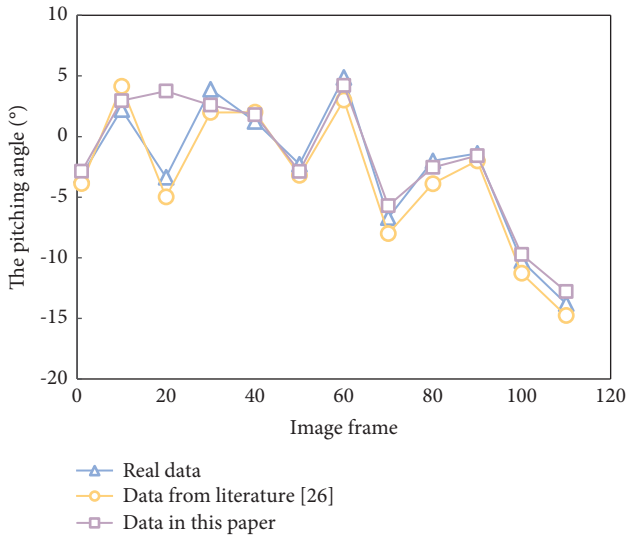

Figure 9: Comparison results of pitch angles of different algorithms.

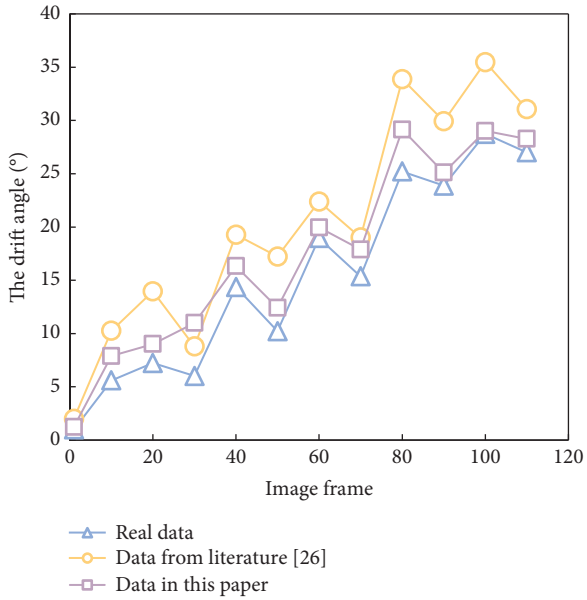

Figure 10: Comparison results of yaw angles of different algorithms.

learning attention detection system, and the information of the students' head rotation will be recorded, in which one corner out of range will be recorded as 1, and the one without exceeding will be recorded as 0. First, if the record is 1 continuously within 2 seconds, it will be judged as learning distraction and counted. Secondly, the ratio of the sum recorded as 1 per unit time to the total time will be calculated and the students' classroom learning attention will be output. From this, we can count the number of distractions in learning and the situation of students' learning attention per unit time. Some typical behaviors are shown in Figures 9–11.

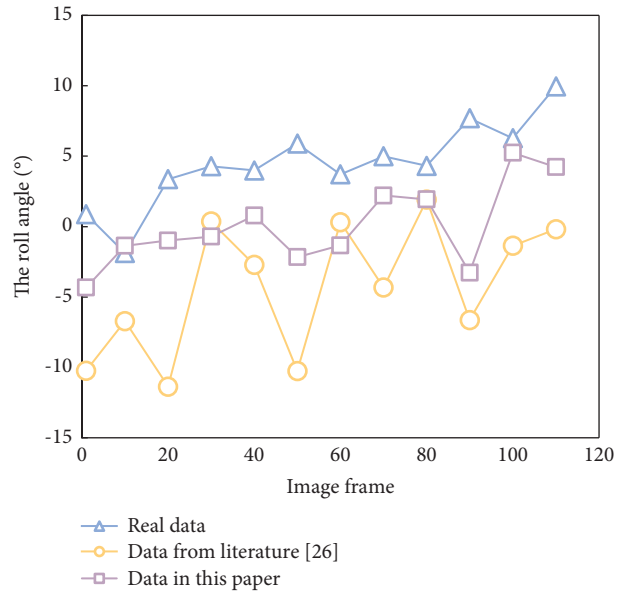

Figure 11: Comparison results of roll angles of different algorithms.

As shown in the above figures, the displacement deviation of the head posture in $X/Y/Z$ direction by this method is within the acceptable range. The maximum error of displacement estimation in the x-axis and y-axis is less than 40 mm and 50 mm, respectively, and the maximum error of displacement estimation in z-axis is less than 230 mm. In this paper, the discriminant analysis of students' learning attention is accurate, which can basically coincide with the calibration curve, and the deviation is small.

5. Conclusion

Online education is a new trend in the development of education, which will bring profound changes in educational concepts, educational systems, teaching methods, personnel training models, etc., and play a positive role in deepening education and teaching reform, improving education quality, and promoting education equity. In this paper, a discrimination method of learning attention based on head posture analysis is proposed. By selecting key frames with different head postures and generating them online, besides attaching posture parameters to the key frames, the head region is accurately extracted from each key frame as the head perspective, and the key frames are combined into a multiscale visual angle apparent model according to the spatial distribution. A DBN model is proposed to reason students' LAG. The model integrates the relationships among multi-LAG, multistudent positions, and multicamera face images, and conducts joint reasoning. We measure the head posture by the similarity vector between the face image and multiple face categories without explicitly calculating the specific head posture value. We collected test data in the teaching environment, and the experimental results show that our model can get better results.

In the current experiment, when the user looks at the distant visual attention target, the attitude measurement is inaccurate due to the small difference of images. In the future, we will consider fusing motion information to detect the change of the user's visual attention target. The visual attention target in this paper is several screens on the wall. In the future, we will consider extending the visual attention target to more places, such as different areas on the workbench.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] G. Otte, "Online instruction as local education: cuny's online baccalaureate," *Journal of Asynchronous Learning Networks*, vol. 11, no. 1, pp. 9–14, 2019.

[2] W. B. Root and R. A. Rehfeldt, "Towards a modern-day teaching machine: the synthesis of programmed instruction and online education," *Psychological Record*, vol. 71, no. 1, pp. 85–94, 2021.

[3] B. B. Lockee, "Online education in the post-covid era," *Nature Electronics*, vol. 4, no. 1, pp. 5-6, 2021.

[4] J. Zhang and F. Zhang, "Empirical research on factors affecting users' willingness to continuously use online education platform," *Value Engineering*, vol. 38, no. 5, pp. 134–137, 2019.

[5] E. Fehr, "Ecologies of faith in a digital age: spiritual growth through online education," *Growth: The Journal of the Association for Christians in Student Development*, vol. 18, no. 18, p. 11, 2019.

[6] D. Mcauliffe, "Challenges for best practice in online social work education," *Australian Social Work*, vol. 72, no. 1, pp. 110–112, 2019.

[7] V. Kearney, J. W. Chan, T. Wang et al., "Dosegan: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation," *Scientific Reports*, vol. 10, no. 1, Article ID 11073, 2020.

[8] H. Datta, A. Hestvik, N. Vidal et al., "Automaticity of speech processing in early bilingual adults and children," *Bilingualism: Language and Cognition*, vol. 23, no. 2, pp. 429–445, 2020.

[9] M. Bce, S. Staal, and A. Bulling, "How far are we from quantifying visual attention in mobile hci?" *IEEE Pervasive Computing*, vol. 19, no. 2, pp. 46–55, 2020.

[10] R. Hamrah, R. R. Warier, and A. K. Sanyal, "Finite-time stable estimator for attitude motion in the presence of bias in angular velocity measurements," *Automatica*, vol. 132, no. 2, Article ID 109815, 2021.

[11] C. Zhou, Q. Chen, Z. Li, B. Zhao, Y. Xu, and Y. Qin, "Aspect category detection based on attention mechanism and Bi-directional LSTM," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 37, no. 3, pp. 558–564, 2019.

[12] I. Zhao, "An improved attitude estimation algorithm based on mpu9250 for quadrotor," *Mechanical Engineer*, vol. 5, pp. 36–39, 2019.

[13] M. Yeager, B. Gregory, C. Key, and M. Todd, "On using robust mahalanobis distance estimations for feature discrimination in a damage detection scenario," *Structural Health Monitoring*, vol. 18, no. 1, pp. 245–253, 2019.

[14] B. Yfka, B. Lyn, and B. Qzy, "Fluorescent probes for detection of biothiols based on "aromatic nucleophilic substitution-rearrangement" mechanism," *Chinese Chemical Letters*, vol. 30, no. 10, pp. 1791–1798, 2019.

[15] M. Hirata and N. Kusatake, "How cattle discriminate between green and dead forages accessible by head and neck movements by means of senses: reliance on vision varies with the distance to the forages," *Animal Cognition*, vol. 23, no. 2, pp. 405–414, 2020.

[16] Y. Guo, J. Zhang, and W. Lian, "Study on learning attention discrimination based on head posture," *Science Technology and Engineering*, vol. 20, no. 14, pp. 5688–5695, 2020.

[17] Y. Lu and L. Yanmin, "Design of online classroom attention evaluation system based on realsense," *china medical education technology*, vol. 34, no. 3, pp. 82–86, 2020.

[18] S. Xiao and S. Nan, "Head pose estimation of 3D point cloud based on deep learning," *Computer Applications*, vol. 40, no. 4, pp. 72–77, 2020.

[19] Y. Ren, S. Su, K. Guan, W. Fu, and X. Zhang, "Detection and fatigue analysis based on oculogram and head posture signals," *Acta changchun university of science and technology: Natural Science Edition*, vol. 1, pp. 38–44, 2020.

[20] S. Li, "An analysis of effective strategies to maintain the attention of primary school students in classroom learning," *Chinese class*, vol. 9, p. 65, 2020.

[21] L. Liang, T. Zhang, and H. Wei, "Head pose estimation based on multiscale convolution neural network," *Advances in Laser and Optoelectronics*, vol. 56, no. 13, 2019.

[22] J. Cui and J. Wang, "Facial expression recognition model based on enhanced head pose estimation," *Computer Science*, vol. 46, no. 6, pp. 328–333, 2019.

[23] X. Liu, "Analysis of children's attention in preschool education," *Encyclopedia Forum Electronic Journal*, vol. 3, pp. 662-663, 2020.

[24] Z. Zeng, X. Lu, S. Xu, and M. Chen, "False comment detection based on deep learning model of multi-layer attention mechanism," *Computer Applications and Software*, vol. 37, no. 5, pp. 183–188, 2020.

[25] T. Wu and C. Chunping, "Emotion analysis model of long-term and short-term memory based on attention crossover with position weight," *Computer Applications*, vol. 39, no. 8, pp. 2198–2203, 2019.

[26] J. Hu, W. Zhang, and S. Chen, "Research on attribution of attention deficit in college students fragmentation learning-qualitative analysis based on grounded theory," *Audio-visual Education Research*, vol. 40, no. 12, pp. 36–43, 2019.