Heliyon 10 (2024) e37365

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

How attentive is a student in class? A concentration evaluation method based on head Euler angle

Zexiao Huang [a], Ran Zhuo [b], Fei Gao [a,*]

[a] College of Computer, Zhejiang University of Technology, Hangzhou 310023, China
[b] Zhejiang Vocational College of Special Education, Hangzhou 310023, China

ARTICLE INFO

Keywords:
Head up rate
Head front-facing rate
Head Euler angle
Concentration evaluation

ABSTRACT

How attentive is a student in class? It is an interesting and challenging question in the field of education research. In this paper, a concentration evaluation method based on head Euler angle is proposed. First, a fine large-small-interval video sampling strategy is advocated and concepts of dynamic threshold and action amplitude coefficient are defined, which are fused together to get the head up rate. Second, head front-facing rate is computed through analyzing the yaw angle, which is obtained from a proposed spatial physical model of camera and student seat position. Third, for any given time period, class concentration of a student is derived by fusing and normalizing the head up rate and head front-facing rate based on Euclidean-distance. Finally, accuracy experiments are conducted, where the proposed method achieves an average accuracy of 89.8% compared with the average score from the questionnaire given by 22 reviewers. Also, robust experiments on several head pose estimation models are facilitated, which achieves an average accuracy of 98.43% on recorded video dataset, indicates insensitivity to the head pose estimation model used for data collection and verifies that the proposed method is independent on any specific head pose estimation model. The experimental results show that the proposed method is accurate, effective and robust for the concentration evaluation in class.

1. Introduction

How attentive is a student in class and how to evaluate the concentration? It is an interesting and challenging question in the field of education research. Concentration is one of the key metrics that are used to evaluate how deep and how long an individual focuses on a job. How to evaluate concentration has become as one of the research hotspots. Human head serves as a crucial source for receiving external information, and the body's response to the external world is often most intuitively represented by the data related to head movements. Consequently, current methods of concentration evaluation were mainly developed around the analysis of head-related data, which can be classified into four categories as follows: head pose estimation [1–4], facial expression analysis [5–8], gaze direction analysis [9–11], and multimodal fusion [12–16]. In the method of head pose estimation, data such as head Euler angles, head contours, and changes in head posture are mainly employed to evaluate the concentration. Although it offers rapid and intuitive evaluation through scoring head behaviors, it is still coarse and is hard to facilitate the quantitative assessment. On the other hand, although the methods of facial expression analysis and gaze direction analysis focus on finer details and can measure the concentration quantitatively, they are not suitable for the scenarios with minimal variation in facial expressions and gaze direction,

* Corresponding author.
E-mail address: feig@zjut.edu.cn (F. Gao).

https://doi.org/10.1016/j.heliyon.2024.e37365
Received 25 August 2024; Received in revised form 30 August 2024; Accepted 2 September 2024

Available online 12 September 2024

such as classroom, meeting, and working environment, etc. Besides, they need higher precision in data collection. Although multimodal fusion methods have achieved certain results, they still lack an explanatory mechanism for the correlation between fused pose data and concentration and show poor performance in practical applications.

In various scenarios, class concentration evaluation of a student makes significant importance and plays an essential role in the understanding of their learning status [17]. As for the aforementioned discussion, there are still some limitations like big error, poor generalization and difficulty in quantification, which make current methods be unsuitable for the scenario like classroom where individuals change little in facial expressions and gaze direction. Lee [18] using X-ray photometry and EEG analysis to explore the correlation between head posture and brain attention levels. Positive correlations were found between increased range of motion of the cervicospinal area and the mental resting state score and between increased flexion and extension motions and mental concentration score, which indicates that the student's head posture are positively correlated with the mental concentration score. Jha [19] explored the relationship between head and eye movements and the target object of focus while studying driver concentration. The experiment demonstrated that gaze direction is positively correlated with the driver's concentration. From the perspective of feasibility and practicality, Li [20] argued that information derived from head Euler angles can be effectively implemented with lower pixel resolutions (compared to detecting gaze direction and facial expressions) and can be directly applied using existing classroom monitoring equipment. In a classroom setting, changes in students' head posture can significantly reflect their level of concentration. Among these, whether the head is raised and whether the head is facing forward are two key postures closely associated with concentration. Hence, in this paper, a concentration evaluation method based on head Euler angle is proposed, which focuses on the computation of head up and head front-facing data. The main contributions are as follows:

1) A head up rate calculation method based on pitch angle and dynamic threshold is proposed to evaluate the proportion of the time of student heading up. The method obtains the head pitch angle through pose estimation and calculates a dynamic threshold for pitch angle based on a large-small-interval video sampling strategy. Then, the head up rate for each interval is calculated according to the dynamic threshold. Experimental results verify the high accuracy and practical value.

2) A head front-facing rate calculation method based on yaw angle and spatial pose model is presented to evaluate the proportion of the time of student facing forward. The method establishes a spatial position model of camera and student seat position, derives the facial orientation according to the yaw angle, computes the included angle between the facial orientation and the line of camera and student seat position, and finally calculates the head front-facing rate.

3) A normalized Euclidean-distance-based concentration evaluation method is set up to evaluate the overall concentration level of a student in class. In the method, head up rate and head front-facing rate are integrated and normalized through employing Euclidean-distance. The experimental results validate the rationality of the proposed method.

The remainder of this paper is organized as follows: section 2 briefly reviews the related work; section 3 outlines the proposed method; section 4 shows the experimental results, and section 5 gives the conclusion.

2. Related work

Concentration evaluation is an interesting and challenging task. Current methods of concentration evaluation mainly focus on the analysis of head-related data, which can be classified into four types, i.e. head pose estimation, facial expression analysis, gaze direction analysis, and multimodal fusion.

2.1. Head pose estimation based method

Head pose estimation methods rely on head pose estimation models. Abate et al. [21] provided a detailed overview of the latest techniques and applications in HPE, particularly in aspects related to attention recognition, such as driver attention detection. By comparing the advantages and disadvantages of different methods, the article offered valuable insights for selecting appropriate HPE technologies for attention recognition in our study. Head pose estimation methods typically involve obtaining information such as head Euler angle and facial key points through head pose estimation models, followed by estimating concentration through regression modeling. Xu et al. [1] utilized a head pose estimation model to acquire users' head Euler angles and corrected angular errors induced by spatial disparities to derive concentration scores. Mustafa et al. [2] detected student engagement based on head pose and eye aspect ratio by employing a binary classification method to evaluate students' concentration states. Zhao et al. [3] calculated pitch angle from facial key point information and used a binary classification method with set thresholds to count instances of head up movements, obtaining the head up rate. Dai et al. [4] trained a classification network using a self-constructed dataset of head movement images to track the rate of students lifting their heads. Li et al. [22] evaluated students' attention states by measuring the Euler angles of their facial poses, and then integrated the concentration states of each student to obtain a curve of the overall concentration score over time, which describes the collective attention state of the entire classroom. However, these methods overlook the subtle distinctions in head movements in different directions, and binary classification fails to normalize the level of concentration, leading to significant errors. Calculations are relatively rough due to factors such as distortion, face deviation, and differences in human-machine heights being disregarded.

2.2. Facial expression analysis based method

Facial expression analysis methods typically classify facial expressions and assess student concentration levels based on deep learning techniques. Sümer et al. [5] proposed a concentration evaluation method based on attention-embedded facial emotion

features. However, the method requires extensive pre-labeling of large facial expression datasets, which makes it impractical for practical scenarios. Thomas et al. [6] fused facial expressions, head poses, and eye gaze data to analyze student concentration levels, but it suffers from drawbacks such as high computational complexity and weak inter-modality data correlations. Sharma et al. [7] integrated emotional analysis, eye tracking, and head movements to provide three categories of concentration assessments for each time segment of student learning. However, it primarily targets online education and lacks quantification of concentration evaluation metrics. Buono et al. [8] trained a student concentration assessment model based on LSTM using facial behavior data to predict the engagement levels of facial actions, gaze, and head poses. Kawamura et al. [23] calculated the Jaccard coefficient based on students' facial reactions to teachers' actions as an indicator for evaluating concentration, which used SVM to classify each action and the level of the Jaccard coefficient, and obtained a weighted calculation model for concentration to evaluate students' concentration. Nevertheless, due to challenges such as the difficulty in collecting facial expression datasets and weak associations between concentration and facial expressions, facial expression analysis methods face significant challenges.

2.3. Gaze direction analysis based method

Eye gaze direction methods typically use gaze collection detection devices such as eye trackers to determine the focal points of the eyes and establish a focus model for head concentration. Yang et al. [9] analyzed teachers' gaze fixation using eye trackers and its relation to students' classroom listening performance, experiment shows that a balanced gaze fixation rate contributes to better task focus. Shogo et al. [10] proposed a user concentration calculation method based on eye gaze direction. However, this method performs poorly in open environments, and participants need to wear eye-tracking devices; otherwise, the direction of their gaze is difficult to capture using other devices such as cameras, which could potentially impact attention tasks potentially. Ramachandra et al. [11] developed an intelligent assessment system based on eye gaze. This system uses a camera in front of the screen to detect students' eye gaze direction and record the duration of eye fixation. Based on these fixation durations, heat maps are generated to evaluate the areas of focus for students' attention. Lin et al. [24] employed facial expression scores, visual focus rate, and task proficiency as evaluation indicators to measure students' learning engagement in head-mounted virtual interactions. However, this method is limited when facial expressions are obscured, leading to the loss of some facial data. The authors did not consider the impact of the pitch angle on facial expression recognition. Meanwhile, due to the requirement for each participant to wear eye-tracking equipment and the inherent uncertainty in eye gaze behavior, eye gaze methods are difficult to apply effectively in real-world classroom attention assessment.

2.4. Multimodal fusion based method

Considering that changes in concentration state result from the combined effect of bodily actions, many researchers employ multimodal data fusion. Patricia et al. [12] fused multiple modalities such as gaze direction, head pose, and facial expressions to train a posture recognition model for inferring student concentration. Sandeep et al. [13] designed an online learning concentration assessment model based on graph convolutional neural networks. This model evaluates students' attention by utilizing facial expressions and gaze direction information from videos. Experimental results suggest a positive correlation between attention and academic performance. Peng et al. [14] proposed a student concentration evaluation mechanism based on multiple reactions, this method divided students' actions into head pose estimation, facial expression recognition, and multiple response estimation, however the accuracy of concentration feedback in simulated experiments was only 70%. Xie et al. [15] fused facial expressions, head poses, and eye-mouth coordinates to train a BP neural network for student concentration evaluation. Prakhar et al. [16] calculated concentration through weighted averages of facial expressions, emotions, and student survey data. However, the limitations of multimodal fusion methods lie in the difficulty of establishing logical relationships between multimodal data and real concentration levels, as well as the lack of appropriate methods for quantifying concentration. Moreover, due to significant distortions and high workload associated with cross-modal data fusion, their application in practical scenarios is challenging.

2.5. Real-time concentration evaluation method

Although many deep learning-based methods have made progress in recent years, their limitation lies in the need for manual annotation and classification of data. Only qualitative classification and regression of attention can be performed through these methods, rather than accurately measuring the quantitative level of attention. Guo et al. [25] proposed an improved ViViT network for online classroom concentration detection. Liao et al. [26] utilized facial convolutional networks and long short-term global attention networks to extract students' spatiotemporal facial information for concentration calculation. However, the use of publicly available datasets led to low accuracy in concentration computation, and there was no practical analysis of students' facial poses in concentration tasks scenarios. Due to prolonged sitting, students' bodily posture, facial expressions, and other behaviors may be naturally influenced by bodily functions. Therefore, the above-mentioned attention assessment model should reduce redundant feature information. Moreover, excessive integration of multimodal information and multiple categorical variables in complex formulas diminish the model's robustness.

In summary, there are still some limitations of current methods concentration evaluation, such as big error, poor generalization, difficulty in quantification, etc. Actually, it is difficult to express the class concentration with discrete values. It would be more reasonable to express concentration with a normalized continuous range within [0, 1.0], where 0 indicates complete lack of concentration and 1.0 represents full concentration. Then, how to design a rational concentration evaluation method to meet this representation

assumption? Hence, a concentration evaluation method based on head Euler angle is proposed, which is described in detail in the following sections.

3. Method

Considering that class teaching typically lasts for certain duration, it is more meaningful to evaluate students' sustained concentration in class. Therefore, concentration evaluation values are set within the range of [0, 1.0], where 0 represents complete lack of concentration, such as constantly looking around or being engrossed in a mobile phone, and 1.0 represents full concentration, such as consistently looking up at the blackboard or facing forward throughout. To achieve this, a class concentration evaluation model is constructed as shown in Fig. 1. First, a head pose estimation model is used to obtain the pitch and yaw angles of a student. Then, a fine large-small-interval video sampling strategy is advocated to analyze the pitch angle and the head up rate P_{ij} for each small interval is calculated. By establishing a spatial physical model of camera and student seat position and analyzing the yaw angle, the head front-facing rate Y is computed for each frame. Convert the interval position of the head up rate P_{ij} into the frame sequence position to obtain the head up rate P and head front-facing rate Y for any time period. Finally, head up rate P and front-facing rate Y are fused to be fitted and normalized using Euclidean-distance to derive the concentration C. After applying a linear transformation function $f(x)$ to P and Y, and projecting the resulting two-dimensional real coordinate points (Y', P') onto the coordinate system region H using function g, a scatter plot representing the direction of head concentration is obtained. The density of scatter points reflects the concentration of students' head orientation. In the model, how to calculate the head up rate, head front-facing rate and concentration is the key.

3.1. Head up rate

The head up rate refers to the proportion of time that students spend looking up compared to the total time, which serves as an indicator of the degree of vertical concentration of the students' heads. Since human attention cannot be sustained for long time, individuals may experience fatigue after a certain period, thus leading to actions such as looking up or down. Considering the variability of human concentration, a fine large-small-interval video sampling strategy is proposed as illustrated in Fig. 2, where, the class video sequence is divided into M large intervals, each containing N small intervals, each large interval has a duration of T and each small interval has a duration of t. Analyzing the average pitch angle within each small interval and the median pitch angle within the large interval yields the head up rate P_{ij}, where $i = 1, 2, ..., M$ and $j = 1, 2, ..., N$. The strategy facilitates a more detailed understanding of students' head postures during different time intervals.

As shown in Fig. 3, the camera is positioned along the central line of the classroom directly above the podium to capture the students' head pitch angles, with the pitch angle direction defined as follows: downward head movement is considered negative, while upward head movement is considered positive.

Upon careful observation, two scenarios are identified within the sampling intervals: 1) intervals where both upward and downward head movements occur simultaneously, as illustrated in Fig. 4a; 2) intervals where only upward or downward head movements occur, as depicted in Fig. 4b. It is evident from Fig. 4 that there is a strong correlation between the pitch angle and head movements. Therefore, when calculating the head up rate, both the pitch angle and the time series are essential factors. Moreover, considering that assessing head movements requires considering the length of time intervals and the angle changes between intervals, the proposed fine large-small-interval video sampling strategy is adopted. Additionally, a dynamic threshold, q_i, is defined to serve as a reference value for comparing upward and downward head movements within each small interval of the i-th large interval, as shown in Equation (1).

$$q_i = \begin{cases} \mathrm{med}(SP_i) & \text{if } \left|\Delta(SP_i)\right| \geq \theta \\ \mathrm{med}(S) & \text{else} \end{cases} \tag{1}$$

Where, SP_i represents the pitch angle sequence of the i-th large interval, S represents the pitch angle sequence of the entire video, and $S = \{SP_i \mid i = 1, 2, \ldots, M\}$. med is a function that takes the median of a given sequence, ΔSP_i represents the pitch angle amplitude difference of the i-th large interval, and θ denotes the threshold for head up and head down actions.

The magnitude of head movements reflects the degree of head lifting or lowering when a student is tilting the head up or down. When a student is in a frontal-facing posture, the magnitude of head movements is relatively small, and the head up rate tends to be close to the median level within the large interval. Therefore, the magnitude of head movements in the small interval when the head is tilted up or down also needs to be considered. To solve this, action magnitude coefficient, a_{ij}, is defined as shown in Equation (2).

$$a_{ij} = \begin{cases} \left|\dfrac{p_{ij} - q_i}{\max(SP_i) - q_i}\right| & \text{if } p_{ij} \geq q_i \\ \left|\dfrac{p_{ij} - q_i}{\min(SP_i) - q_i}\right| & \text{else} \end{cases} \tag{2}$$

$$P_{ij} = \frac{1}{1 + e^{-a_{ij}(p_{ij} - q_i)}} \tag{3}$$

Then, the analysis of the specific meaning of action magnitude coefficient a_{ij} in conjunction with Fig. 4 is described in conjunction with Fig. 4. As shown in Fig. 4a, when both head up and head down actions occur within an interval, we take two small intervals

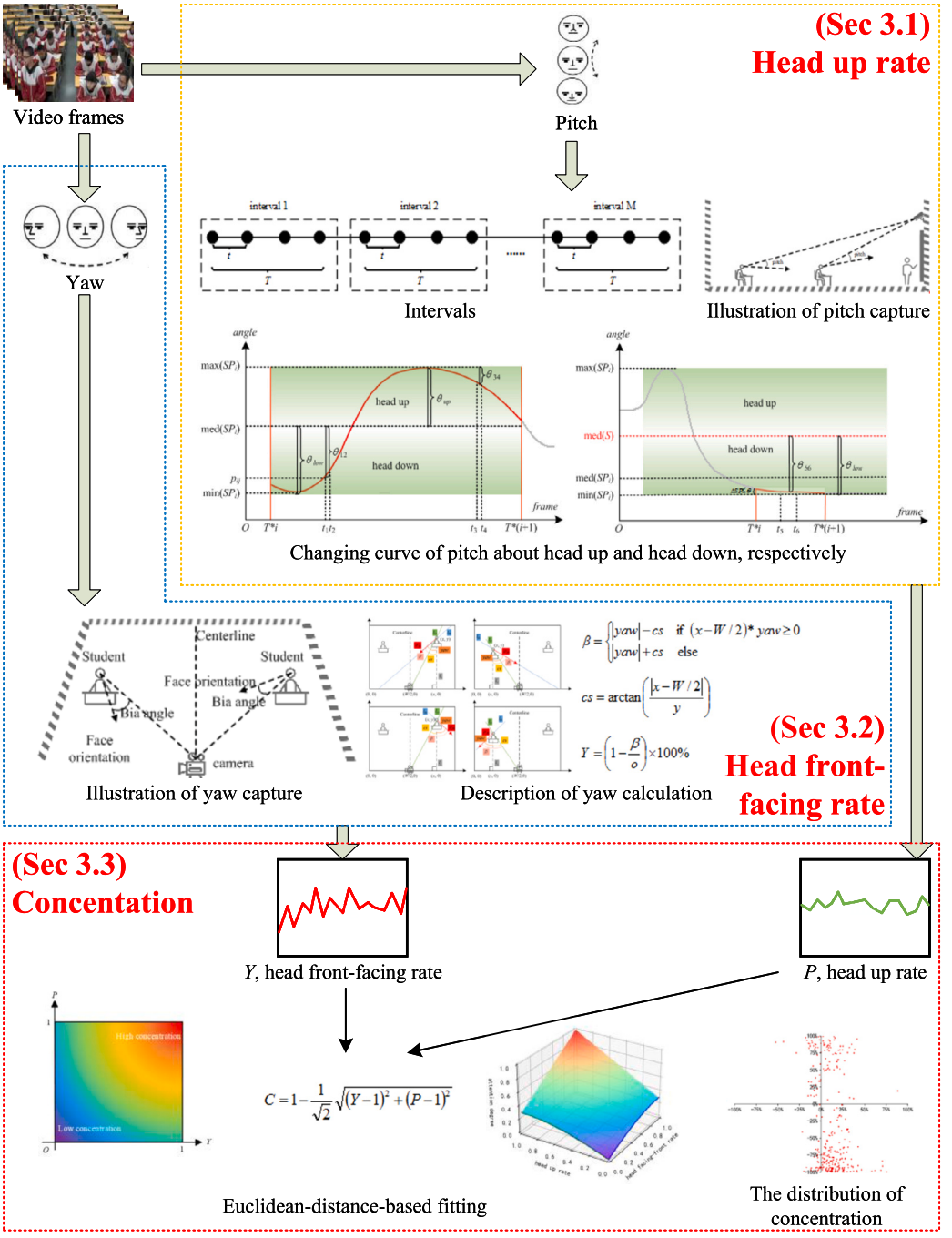

Fig. 1. Overview of the proposed concentration evaluation method.

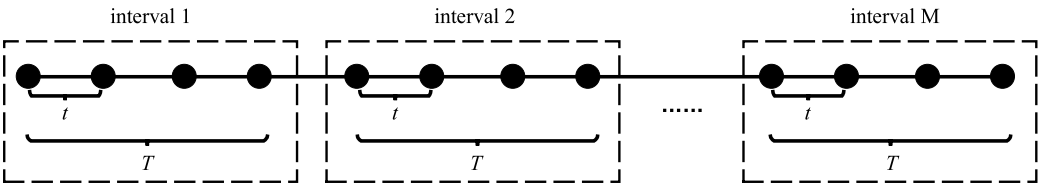

Fig. 2. Illustration of the proposed sampling strategy.

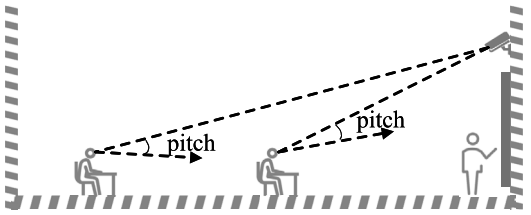

Fig. 3. Illustration of pitch capture.

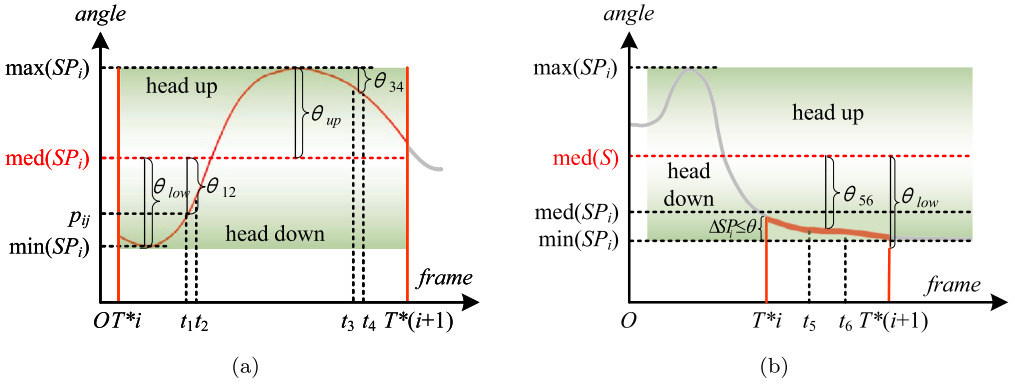

Fig. 4. Sample diagram of pitch angle curve.

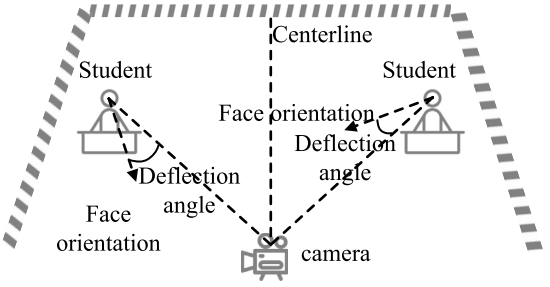

Fig. 5. Illustration of yaw capture.

corresponding to time periods $t_1 \sim t_2$ and $t_3 \sim t_4$. The difference between the average pitch angle of the small interval and the median of the pitch angle in the large interval, denoted as θ_{12} and θ_{34} respectively. During the time period $t_1 \sim t_2$, the student is in a head down state, so the action magnitude coefficient $a_{t_1t_2} = \theta_{12}/\theta_{low}$, while during the time period $t_3 \sim t_4$, the student is in a head up state, thus $a_{t3t4} = \theta_{34}/\theta_{up}$. As illustrated in Fig. 4b, if only a head down action occurs within the interval, we take a small interval corresponding to the time period $t_5 \sim t_6$. The difference between the average pitch angle of the small interval and the median of the pitch angle in the video sequence, denoted as θ_{56}. In this case, the action magnitude coefficient $a_{t_5t_6} = \theta_{56}/\theta_{low}$. Moreover, according to Equation (2), it can be understood that the mathematical interpretation of a_{ij} is the percentage of the average pitch angle p_{ij} within the small interval relative to the maximum pitch angle of the head's directional motion. The larger the action magnitude within the small interval is, the stronger correlation between the dynamic threshold and head movements is, thus a_{ij} approaches 1. Hence, based on the different situations of head movements within the small interval, the head up rate P_{ij} for each small interval can be calculated according to Equation (3).

3.2. Head front-facing rate

The head front-facing rate refers to the angle formed by the student's frontal face and the line connecting the student's head to the camera's optical center which expressed as a ratio to the maximum allowable deviation angle, relative to the maximum set yaw angle. It serves as an indicator that reflects the level of horizontal focus of the student's head. In real classroom scenarios, students' attention may inadvertently be influenced by their surroundings, resulting in lateral movement of the head.

As depicted in Fig. 5, camera is positioned along the midpoint of the classroom, directly above the podium (following the camera pose shown in Fig. 3). The camera captures the yaw angle of the student's head, with the direction of the yaw angle defined as positive for leftward head tilting and negative for rightward head tilting.

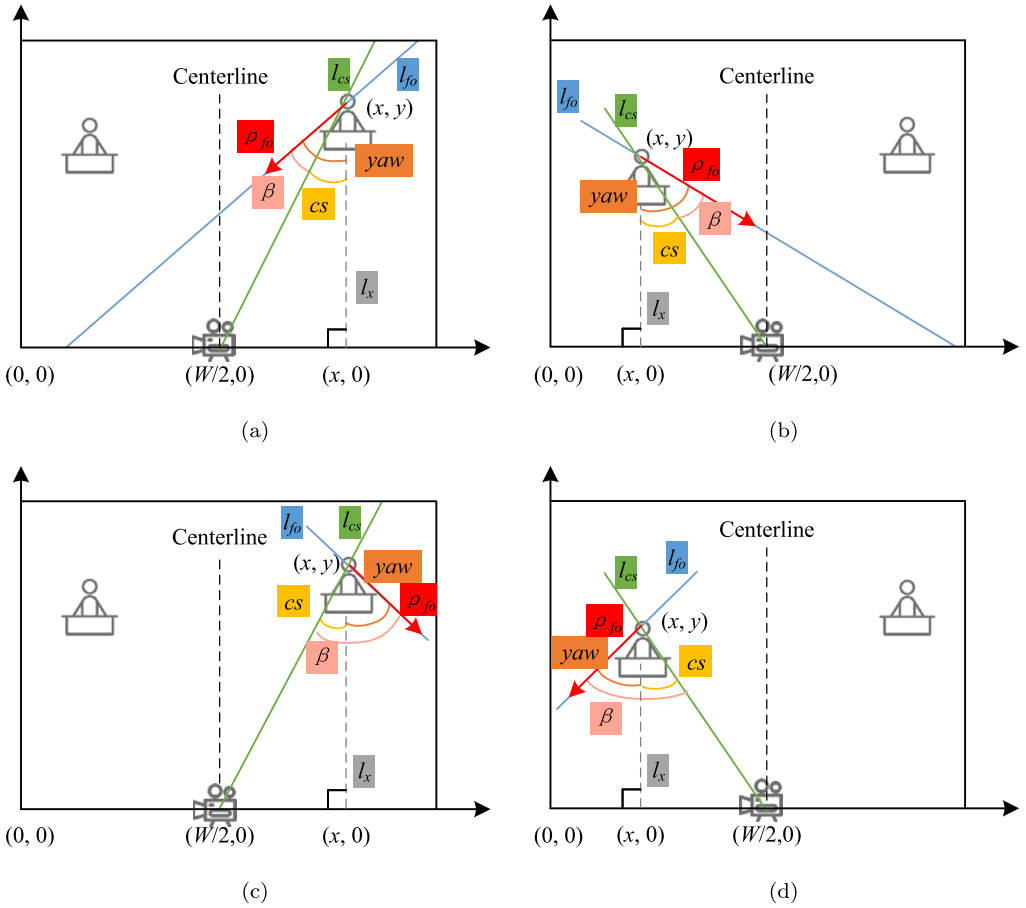

Fig. 6. Description of yaw calculation.

Due to variations in seats and differences in facial orientation, the direction of head focus may vary. Through deep observation, it can be found that when students are focused, their head front-facing direction should be towards the front of the classroom. Four scenarios of student head orientations within the sampling intervals are observed: 1) the student is seated to the right of the camera with the head tilted to the left, as shown in Fig. 6a; 2) the student is seated to the right of the camera with the head tilted to the right, as shown in Fig. 6b; 3) the student is seated to the left of the camera with the head tilted to the left, as shown in Fig. 6c; 4) the student is seated to the left of the camera with the head tilted to the right, as shown in Fig. 6d. According to the spatial pose model depicted in Fig. 6 between the camera and student seating positions, when the student's head is tilted left or right, the yaw angle of the head forms an angle β with the line connecting the head and the camera's optical center. When the student's head is facing the camera directly, the yaw angle β is 0°, and when the student's head deviates from facing the camera, the yaw angle β is greater than 0°. Therefore, yaw angle is considered as a factor in the evaluation of head front-facing rate, where the threshold O is set as the maximum yaw angle threshold.

As shown in Fig. 6, head front-facing rate Y is calculated for any given time interval based on the camera position $(W/2, 0)$ and the coordinates of the student's head (x, y) according to Equation (4), Equation (5) and Equation (6), where, l_{cs} represents the line connecting the camera to the student's head, ρ_{fo} denotes the ray indicating the student's head front-facing direction, and f_{fo} is the line extended from ρ_{fo}. Clearly, l_{fo} passes through the common point (x, y) with l_{cs}, and $(x, 0)$ denotes the intersection point of the line l_x, perpendicular to the line l_{cs} at point (x, y), with the X-axis.

$$\beta = \begin{cases} |yaw| - cs & \text{if } (x - W/2) \geq 0 \\ |yaw| + cs & \text{else} \end{cases} \tag{4}$$

$$cs = arctan\left(\frac{|x - W/2|}{y}\right) \tag{5}$$

$$Y = \left(1 - \frac{\beta}{O}\right) \times 100\% \tag{6}$$

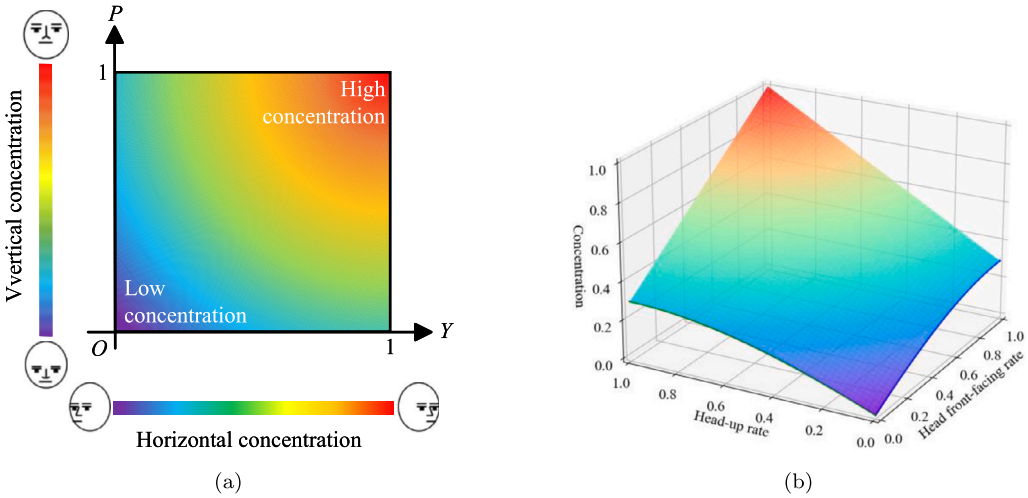

Fig. 7. Illustration of concentration fitting.

Where, cs represents the angle formed between the line l_{cs} and the perpendicular line l_x, where O denotes the maximum value set for the deviation angle in the experiment.

3.3. Euclidean-distance-based concentration

In the data analysis, distance calculation is often used to measure the similarity between any two data points. If the distance between two points is small, it indicates that they are close together in space, namely they are similar. When both the head up rate and head front-facing rate of a student are at high levels, the concentration is high; conversely, the concentration is low. Similarly, distance can also be used to measure the dissimilarity between two points. In this case, a smaller distance may indicate a smaller difference between the two points, while a larger distance indicates a greater difference. Therefore, from the spatial relationship shown in Fig. 7, it can be observed that the head up rate and head front-facing rate are highly correlated with concentration. Here, the head front-facing rate represents the distance from the head horizontally to high concentration, while the head up rate represents the distance from the head vertically to high concentration. Therefore, head up rate and head front-facing rate are transformed into points (Y, P) in a Cartesian coordinate system, where Y represents the head front-facing rate that indicates the concentration in the horizontal direction, and P represents the head up rate that indicates the concentration in the vertical direction. When (Y, P) is closer to the point (1, 1), the student's concentration is higher; conversely, the concentration is lower.

Based on the aforementioned analysis, a normalized Euclidean-distance-based concentration evaluation method is proposed, i.e., given the head up rate P and head front-facing rate Y for any time interval, the normalized concentration C can be computed according to Equation (7). Obviously, $C \in [0, 1.0]$.

$$C = 1 - \frac{1}{\sqrt{2}}\sqrt{(Y-1)^2 + (P-1)^2} \tag{7}$$

4. Experiment

This section validates the effectiveness, accuracy, robustness, and other performance metrics of the proposed method through the following subsections: section 4.1 introduces the way and method of experimental dataset collection; section 4.2, 4.3, 4.4, and 4.5 respectively verify the effectiveness, accuracy, precision, and robustness of the focus measurement; the impact of the large and small intervals proposed in section 3 on the calculation of focus is discussed. In the experiments, θ is set to 5° and O is set to 90°.

4.1. Experimental preparation

The experimental dataset was recorded to verify the effectiveness of the method through the scene setup shown in Fig. 8. A camera was placed directly above the volunteer, with Fig. 8a defining the directions of head movements. As shown in Fig. 8b, the volunteers oriented their faces towards pre-marked positions on the wall. By changing the direction of the face, the head up rate (P) and head front-facing rate (Y) of the volunteers were controlled. The concentration was obtained by adjusting the duration for which the face orientation was maintained. In addition to the wall markings, the head Euler angles were calculated using the YOLO head detection model and the FSANet [27] head pose estimation model during data collection, with real-time head pose calibration provided by the camera.

Fig. 9 depicts the distribution of actions and indicators of the dataset used in our testing, which includes 22 segments. Among them, Fig. 9a includes six segments that solely capture head up and head down actions, which are used to test the accuracy of the

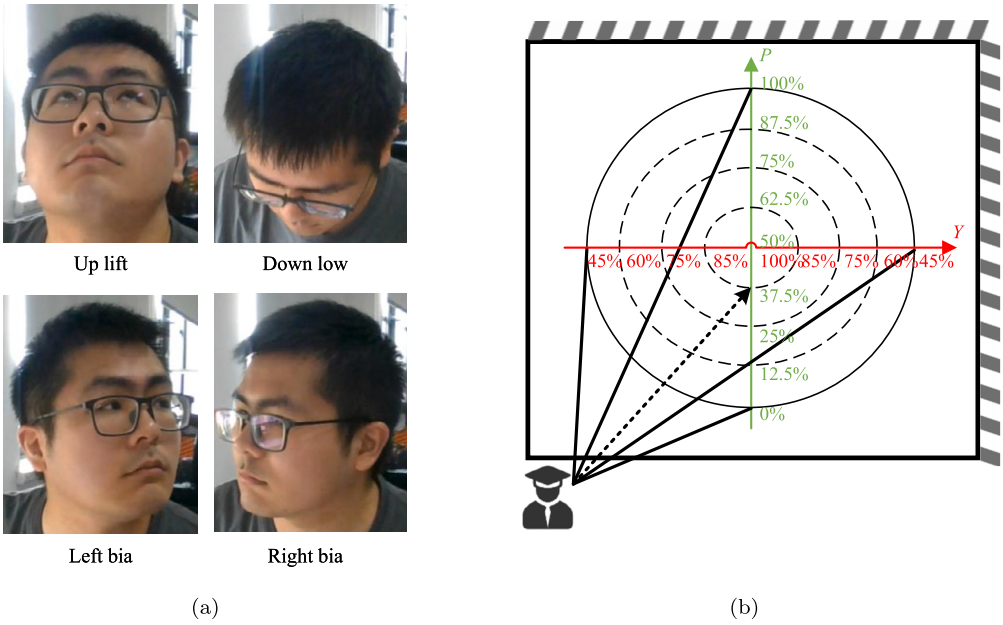

Fig. 8. Definition of student head movements.

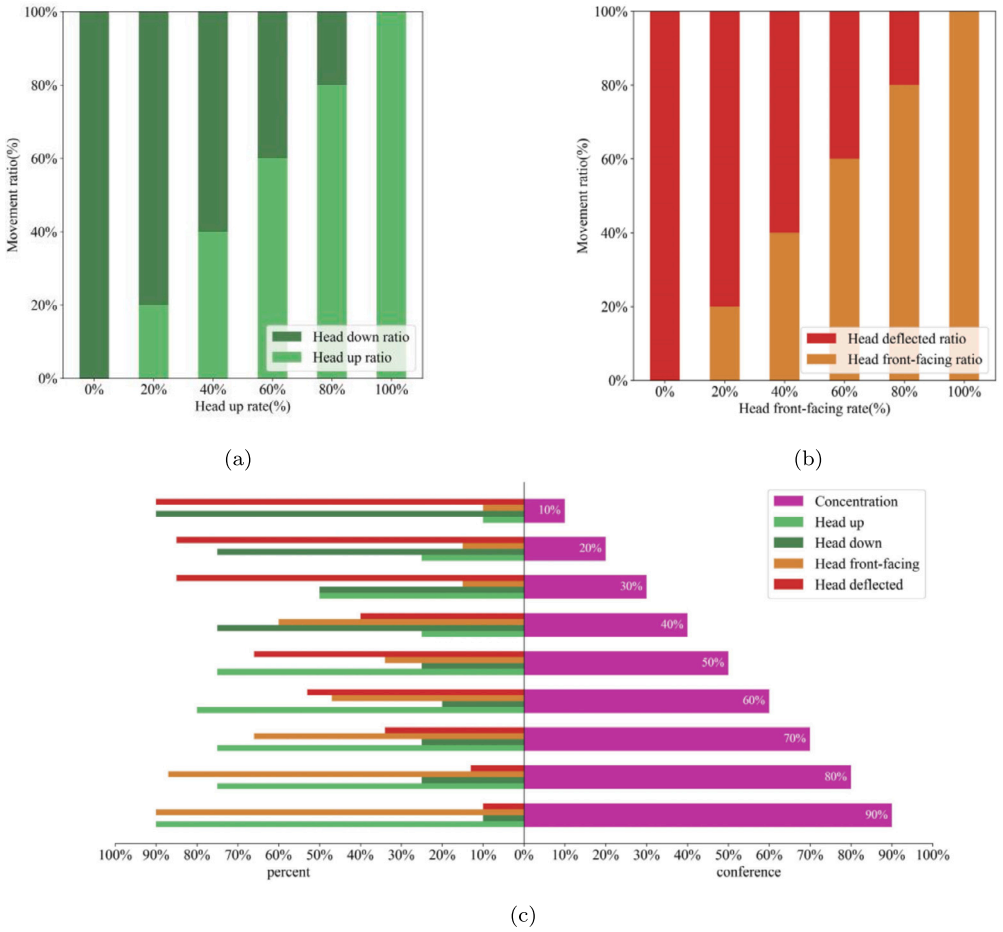

Fig. 9. Experimental dataset metrics and head movement ratios.

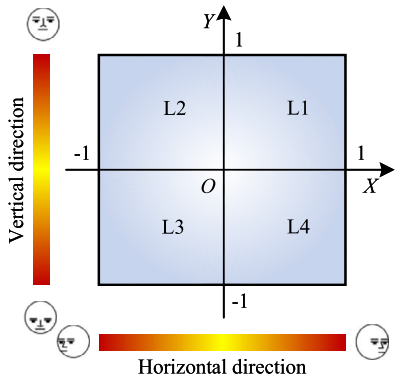

Fig. 10. Distribution diagram of head concentration orientation.

Table 1
Head focus direction description.

Region Identification	L1	L2	L3	L4
Head Orientation	Upper left	Upper right	Lower right	Lower left

head up rate. Fig. 9b consists of six segments exclusively capturing head front-facing and head deflected actions, which aim at testing the accuracy of the head front-facing rate. Lastly, Fig. 9c comprises nine segments that simultaneously record head up and head front-facing actions, which are used to evaluate the accuracy of concentration.

4.2. Verification of the correlation between head up rate, head front-facing rate and concentration

The following experiment was conducted to verify the correlation between the head up rate and the head front-facing rate and the concentration of students: for any given time interval, the head front-facing rate Y is obtained, and the current video interval is sampled frame by frame to derive the head up P. Converting point (Y, P) to (Y', P') through a linear transformation function $f(x) = 2(x - 1)$, followed by mapping the point (Y', P') onto the plane's Cartesian coordinate system region $H = \{(Y', P') \mid Y' \in [-1.0, 1.0], P' \in [-1.0, 1.0]\}$ using the function $g(Y', P') : \mathbb{R}^2 \rightarrow H$. The explanation of scatter diagram in region H is shown in Fig. 10 and Table 1. Its density reflects the focus direction of students' heads. The area with dense scatter points is the action often made by students' heads in the video sequence.

As shown in Fig. 11, the influence of the head up rate and the head front-facing rate on concentration was analyzed separately to obtain the direction points of head movements in the video segment. In Fig. 11a, the head primarily performs nodding and bowing actions, with significant changes in the head up rate, which in turn affects concentration. In Fig. 11b, the head mainly focuses on two lateral areas and continuously switches attention between them, with changes in the head front-facing rate influencing concentration. In section 3.3, we combine the head up rate and the head front-facing rate to obtain concentration through Equation (7), integrating these two factors. Through this ablation experiment, the correlation between the head up rate, the head front-facing rate, and concentration is demonstrated.

4.3. Concentration accuracy verification

According to Equation (7), it is evident that the proposed concentration evaluation method is related to head up rate and head front-facing rate. Therefore, to validate concentration, the accuracy of head up rate and head front-facing rate need to be verified. As illustrated in Fig. 9, the X-axis represents the true values of head up rate and head front-facing rate in the test videos, while the Y-axis represents the experimental calculation results. Five sets of videos are used to test the accuracy of head up rate and head front-facing rate separately. The experimental results in Fig. 12a indicate that certain errors may arise when the head up rate is at the boundary, but overall, the method exhibits high robustness under other conditions. Conversely, the experimental results in Fig. 12b demonstrate that head front-facing rate consistently exhibits high robustness at all times.

In the head up rate test, particular attention is given to the potential scenarios mentioned earlier, where both head up and head down actions could occur in the video sequence, as well as cases where only head up or head down actions are present. The results in Fig. 13 validate the accuracy of the head up rate calculation method. In Fig. 13a, where head up occurs for 2 minutes and head down for 3 minutes in the video, the head up rate curve aligns with expectations. Similarly, in Fig. 13b, where head up occurs throughout the entire 5 minutes, the head up rate curve aligns with expectations.

The principle of head front-facing rate calculation is mathematically proven in Fig. 6. Fig. 14 illustrates the impact of yaw angle variations on head front-facing rate across different video sequences. When the yaw angle is small, the head front-facing rate is high, consistent with experimental expectations. Conversely, when the yaw angle is large, the head front-facing rate is low, also consistent with experimental expectations.

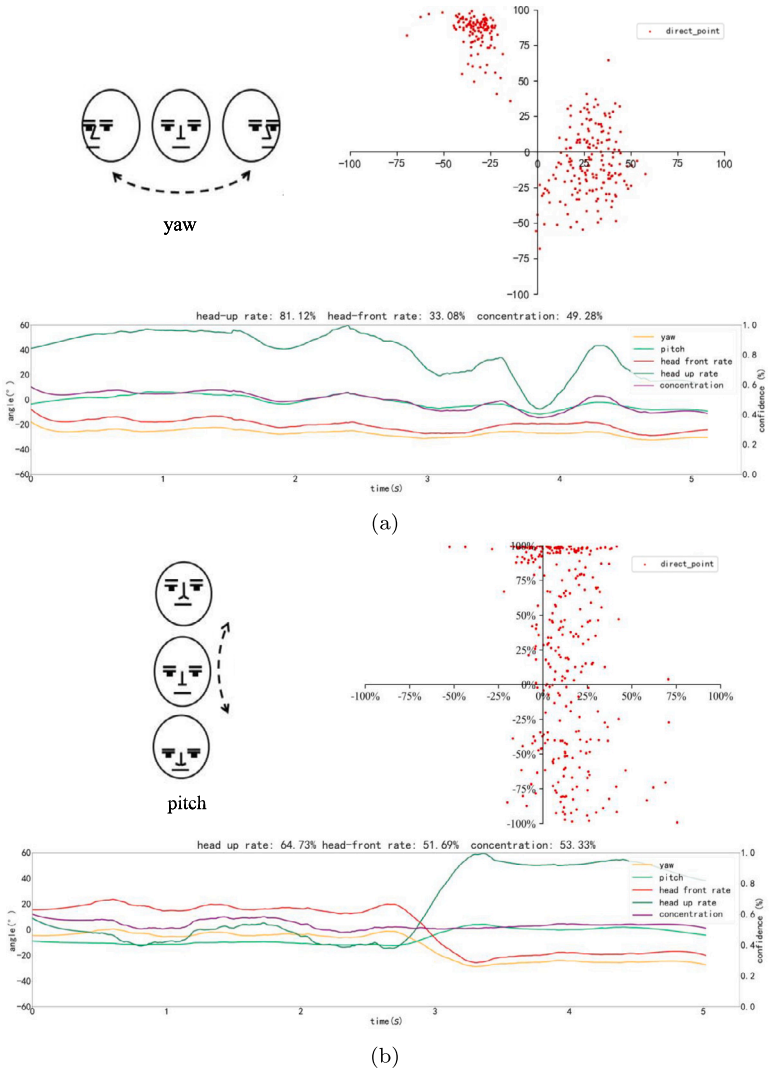

Fig. 11. Correlation analysis between head movements and concentration.

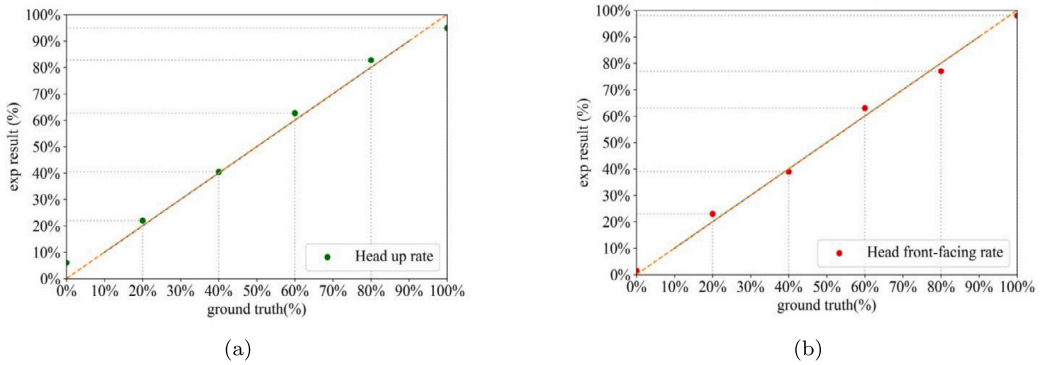

Fig. 12. Comparison of head up rate and head front-facing rate accuracy.

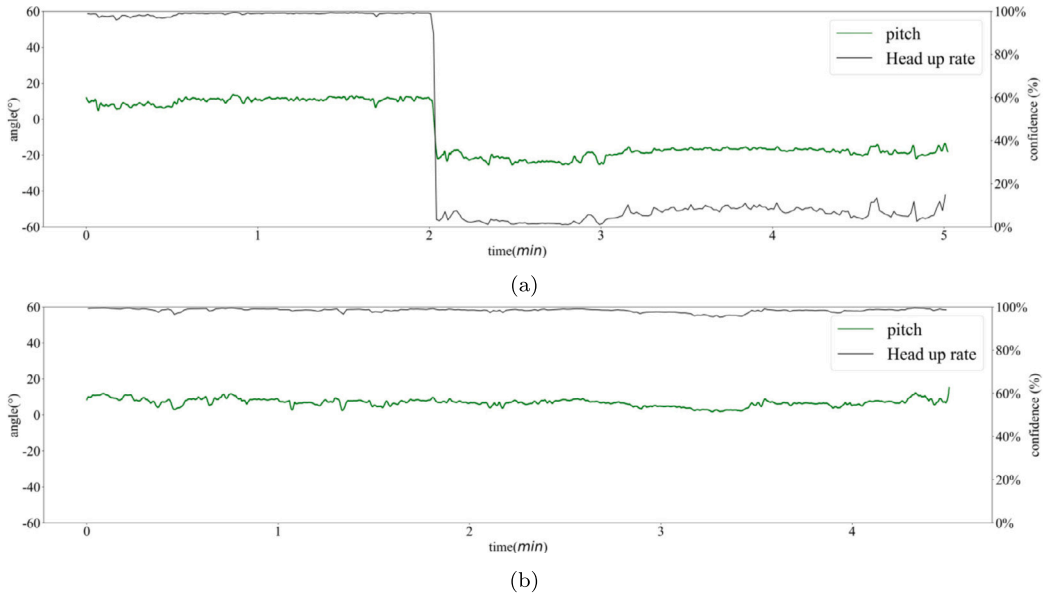

Fig. 13. Head up rate accuracy verification.

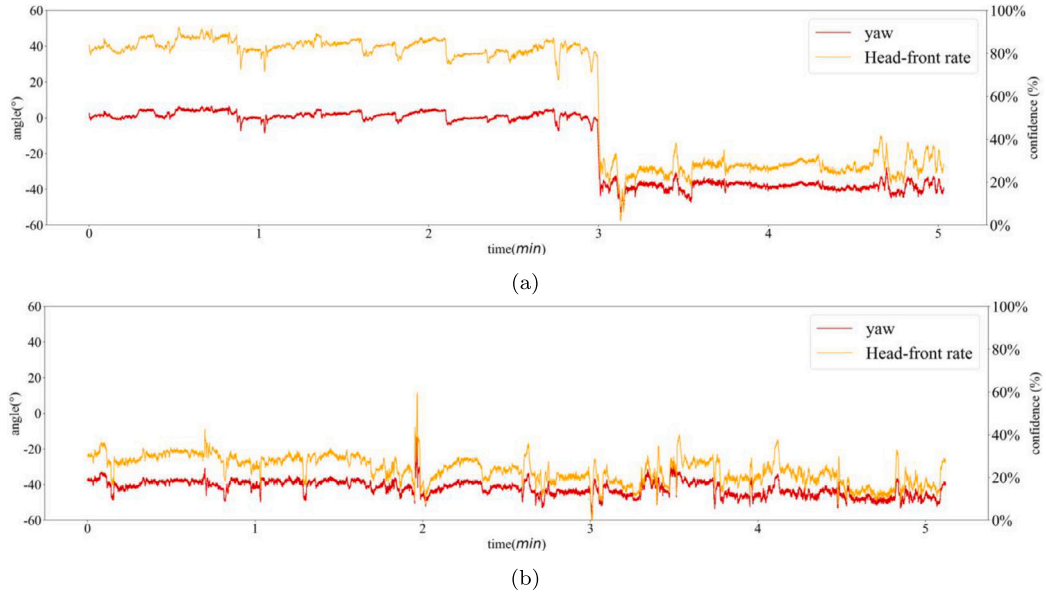

Fig. 14. Head front-facing rate accuracy verification.

4.4. Feasibility and accuracy of concentration detection

Validation of Concentration Detection Effectiveness in Single-Person Online Learning Scenarios through Front-Facing Computer Camera. Fig. 15 visualizes the real-time concentration results, which include three types of information: face no. (#face0, 1, 2, . . .), head Euler angles (y:yaw, p:pitch, r:roll), and concentration (Y:head front-facing rate, P:head up rate, C: concentration). By comparing with the concentration action fitting illustration in Fig. 7, it is evident that concentration is strongly correlated with the head up rate and the head front-facing rate. Changes in head movements indicate variations in the individual's engagement with the focused task.

In the recorded classroom videos, the concentration of multiple individuals was detected. Video analysis was conducted using the YOLO-face model to obtain face bounding boxes for each frame, and the head pose estimation model was used to derive the Euler angles for each face box. This resulted in a collection of video face Euler angles. By employing a concentration evaluation method, the concentration for each face was calculated, and the real-time results are displayed in Fig. 16. As observed in Fig. 16, in actual classroom settings, the concentration levels measured by the proposed method are generally in alignment with objective observations.

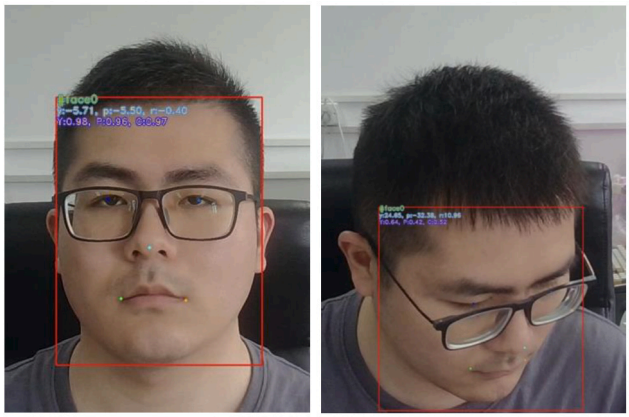

Fig. 15. Front-facing camera single-subject concentration detection.

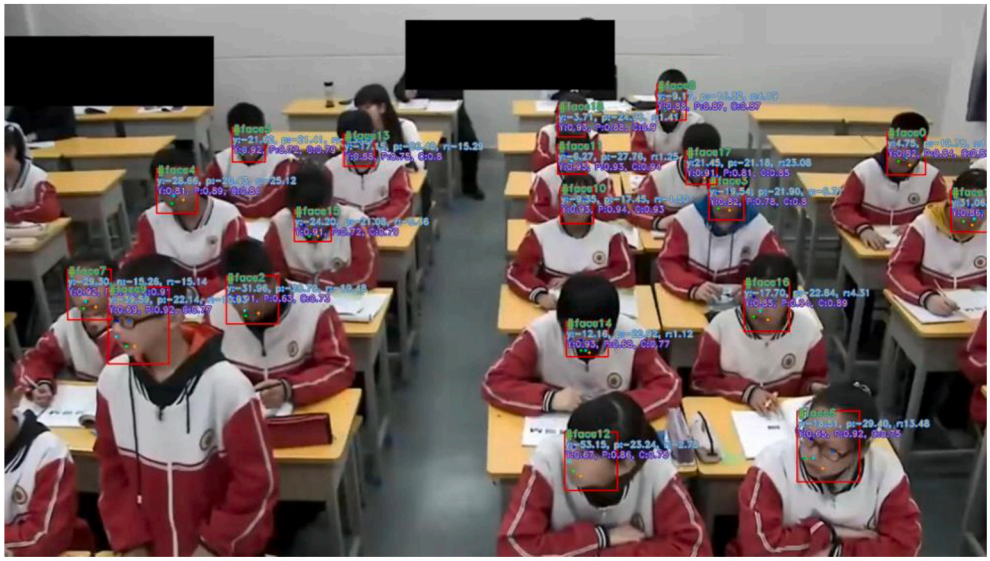

Fig. 16. Classroom video multiple-subject concentration detection.

Table 2
Experimental results of accuracy comparison.

Student No.	1	2	3	4	5	6	7	8	9	10
Ground Truth	0.763	0.651	0.852	0.843	0.752	0.651	0.864	0.623	0.879	0.682
Ours	0.682	0.778	0.892	0.753	0.872	0.728	0.782	0.656	0.842	0.749
Accuracy	89.40%	80.50%	95.30%	89.30%	84.00%	88.20%	90.50%	94.70%	95.80%	90.20%

To verify the accuracy of the method for assessing concentration, we used publicly available classroom videos for verification. We invited 22 volunteers, consisting of students and teachers, to observe the concentration levels of 10 students during a 5-minute classroom video. Each volunteer scored the concentration of each student, and the average score from the manual evaluations was used as the objective standard for the students' concentration levels. These scores were then compared with those calculated by the proposed method, and the comparison between the scores calculated by our method and the average scores from the questionnaires is presented in Table 2. As shown in Table 2, using the average result of manual evaluation as the true value of student concentration, the algorithm achieves an average accuracy of approximately 89.8%.

4.5. Robustness on head pose estimation models

To evaluate the robustness of the proposed method on head pose estimation models, popular head pose estimation models are utilized from the past decade to provide head Euler angle data. Experiments are conducted on nine videos with concentrations ranging from 0.1 to 0.9. The results are depicted in Fig. 17 and Table 3.

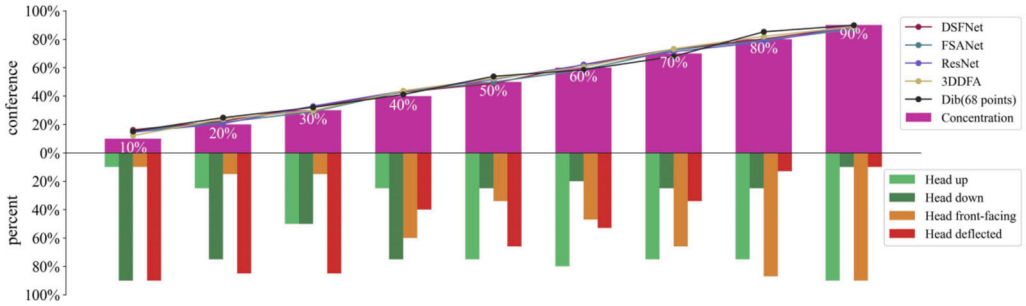

Fig. 17. Model robustness experimental results.

Table 3
Comparison of results from different head pose estimation models.

Video Information				HPE Model Computation Results				
Video ID	*P*	*Y*	*C*	DSFNet [28]	FSANet [27]	ResNet [29]	3DDFA [30]	Dib [31]
1	10%	10%	0.10	0.1621	0.1533	0.1451	0.1211	0.1521
2	25%	15%	0.20	0.2231	0.2145	0.2056	0.2385	0.2478
3	50%	15%	0.30	0.3245	0.2897	0.3291	0.2972	0.3198
4	25%	60%	0.40	0.4276	0.4321	0.4353	0.4367	0.4109
5	75%	34%	0.50	0.4885	0.4982	0.5163	0.5231	0.5382
6	80%	47%	0.60	0.6214	0.5851	0.6197	0.6051	0.5843
7	75%	66%	0.70	0.7312	0.7241	0.7114	0.7321	0.6787
8	75%	87%	0.80	0.8012	0.7921	0.7812	0.8198	0.8521
9	90%	90%	0.90	0.8871	0.8721	0.8921	0.8876	0.8991
MAE				3.25	5.07	6.155	7.39	4.91
Average Accuracy of *C*				98.15%	99.32%	98.495	98.21%	97.98%

The upper portion of Fig. 17 illustrates the concentration values obtained from each head pose estimation model compared to the actual concentration of the videos, while the lower part displays the percentages of head up, head down, head front-facing, and head deflected actions in the corresponding videos. Each test video featured distinct head movements, ensuring our method's ability to handle varied head pose scenarios. Concentration values are manually annotated for each test video prior to experimentation. Head Euler angles are collected using different head pose estimation models, and concentrations were calculated accordingly.

Table 3 presents the information of all simulated videos used in the experiment on the left side and the concentration calculation results of the selected head pose estimation models on the right side. MAE (Mean Absolute Error) at the bottom of the table measures the ability of the head pose estimation models to predict head Euler angles. Average accuracy of the concentration represents the average accuracy of concentration calculation results compared to the ground concentration values for each test video. Comparing the average accuracy of concentration, it is evident that the performance and error variance of the following models are better when the concentration is not close to the boundaries (far from 0 and 1.0). However, when the concentration approaches the boundaries (close to 0 or close to 1.0), students tend to maintain the same posture, resulting in either low or high head up rate and head front-facing rate. In this situation, the factors influencing attention calculation shift from the students to the errors inherent in the head pose estimation model (MAE), consequently affecting the computed head up rate and head front-facing rate. Nevertheless, overall, the proposed method achieves high average accuracy of concentration between 0.1 and 0.9 and demonstrates robustness without sensitivity to specific head pose estimation model performances.

4.6. Selection of hyperparameters for interval sizes

This section compares the effects of different lengths of interval sizes on the calculation results of head up rate and concentration. By comparing different lengths of large and small intervals, we obtained the comparative experimental results as shown in Fig. 18. Based on different lengths of large intervals (T) and small intervals (t), we obtained the corresponding head-raising rate and concentration level. Bilinear interpolation is employed to fit each set of experimental results' scatter points into surfaces. Then, optimal solution planes (grey) are delineated for each set, with the points closest to the planes considered as optimal solutions for each set of experimental results. The optimal solution for each set of experimental results is determined by selecting the point closest to the plane. Summarizing the experimental results show that setting the large interval length (T) to approximately 20s to 30s and the small interval length (t) to around 3.0s is preferred.

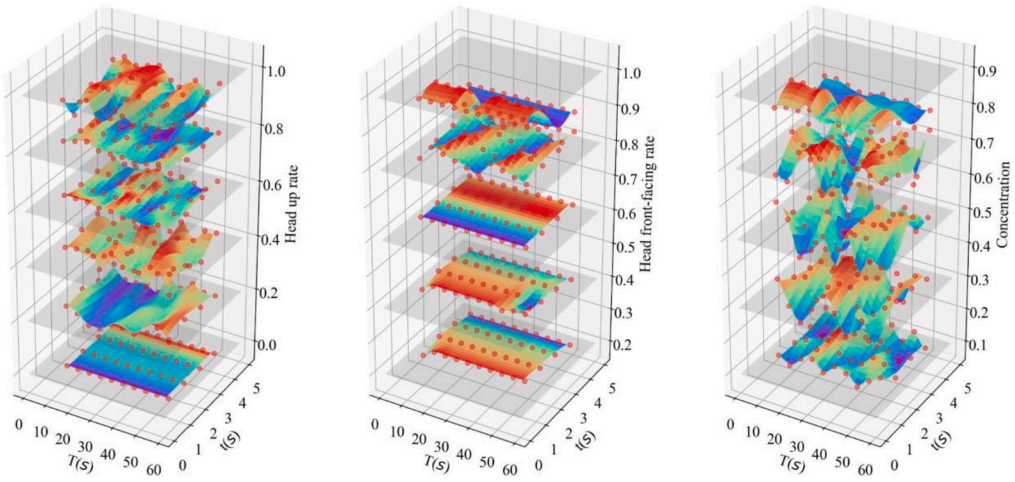

Fig. 18. Interval optimal solution comparisons.

5. Conclusions

In this paper, presents a novel concentration evaluation method is proposed. Head Euler angles, video information, and head position coordinates are collected to obtain two crucial indicators, i.e. head up rate and head front-facing rate. Subsequently, the concentration level based on Euclidean-distance measurement is normalized. Through experiments, the accuracy of the method, the correlation between student head poses and concentration levels, the robustness of existing head pose models, and the selection of hyper parameters for video sampling intervals are verified. The experimental results indicate that the proposed method can accommodate the diversity of student head movements, thereby mitigating issues such as threshold adjustments and pattern recognition flaws present in existing methods, suggesting promising prospects for concentration evaluation in classroom teaching scenarios.

However, there are still some limitations. For instance, the method needs further optimization to adapt to more complex scenarios, factors such as teacher guidance, classroom layout, and subject matter need more consideration, etc. Additionally, the introduction of more advanced sensor technology and machine learning methods could potentially enhance the precision of the method. To enhance the precision and robust, those are part of our ongoing work.

CRediT authorship contribution statement

Zexiao Huang: Writing – original draft, Visualization, Validation, Software, Methodology. **Ran Zhuo:** Resources, Investigation, Funding acquisition, Formal analysis, Data curation. **Fei Gao:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is being supported by the National Key Research and Development Project of China under Grant No. 2020AAA0104001, the Zhejiang Provincial Science and Technology Planning Key Project of China under Grant No. 2021C03129 and 2021C01194, and the Zhejiang Provincial Natural Science Foundation of China under Grant No. LTGG23F020003.

References

[1] X. Xu, X. Teng, Classroom attention analysis based on multiple Euler angles constraint and head pose estimation, in: MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part I 26, Springer, 2020, pp. 329–340.
[2] M.U. Uçar, E. Özdemir, Recognizing students and detecting student engagement with real-time image processing, Electronics 11 (9) (2022) 1500.
[3] W. Zhao, S. Jia, Q. Xue, X. Li, Z. Xiao, Calculation method of classroom head up rate based on head pose estimation, in: 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), IEEE, 2022, pp. 372–376.
[4] Y. Dai, H. Zhao, X. Zhang, Evaluation method of teaching effect based on visual calculation in classroom environment, in: 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2021, pp. 1821–1825.
[5] Ö. Sümer, P. Goldberg, S. D'Mello, P. Gerjets, U. Trautwein, E. Kasneci, Multimodal engagement analysis from facial videos in the classroom, IEEE Trans. Affect. Comput. 14 (2) (2021) 1012–1027.

[6] C. Thomas, D.B. Jayagopi, Predicting student engagement in classrooms using facial behavioral cues, in: Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education, 2017, pp. 33–40.
[7] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, S.R. Khanal, M.C. Reis, J. Barroso, V.M. de Jesus Filipe, Student engagement detection using emotion analysis, eye tracking and head movement with machine learning, in: International Conference on Technology and Innovation in Learning, Teaching and Education, Springer, 2022, pp. 52–68.
[8] P. Buono, B. De Carolis, F. D'Errico, N. Macchiarulo, G. Palestra, Assessing student engagement from facial behavior in on-line learning, Multimed. Tools Appl. 82 (9) (2023) 12859–12877.
[9] J. Yang, C. Liu, Y. Zhang, Q. Yu, Z. Pi, The teacher's eye gaze in university classrooms: evidence from a field study, Innov. Educ. Teach. Int. 60 (1) (2023) 4–14.
[10] S. Hamachi, P. Supitayakul, Z. Yücel, A. Monden, Investigation of the relation between task engagement and eye gaze, in: 2022 13th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter), IEEE, 2022, pp. 163–167.
[11] C.K. Ramachandra, A. Joseph, Ieyegase: an intelligent eye gaze-based assessment system for deeper insights into learner performance, Sensors 21 (20) (2021) 6783.
[12] P. Goldberg, Ö. Sümer, K. Stürmer, W. Wagner, R. Göllner, P. Gerjets, E. Kasneci, U. Trautwein, Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction, Educ. Psychol. Rev. 33 (2021) 27–49.
[13] S. Mandia, K. Singh, R. Mitharwal, Recognition of student engagement in classroom from affective states, Int. J. Multimed. Inf. Retr. 12 (2) (2023) 18.
[14] Y. Peng, M. Kikuchi, T. Ozono, Development and experiment of classroom engagement evaluation mechanism during real-time online courses, in: International Conference on Artificial Intelligence in Education, Springer, 2023, pp. 590–601.
[15] N. Xie, Z. Liu, Z. Li, W. Pang, B. Lu, Student engagement detection in online environment using computer vision and multi-dimensional feature fusion, Multimed. Syst. 29 (6) (2023) 3559–3577.
[16] P. Bhardwaj, P. Gupta, H. Panwar, M.K. Siddiqui, R. Morales-Menendez, A. Bhaik, Application of deep learning on student engagement in e-learning environments, Comput. Electr. Eng. 93 (2021) 107277.
[17] K. Vassie, M. Richardson, Effect of self-adjustable masking noise on open-plan office worker's concentration, task performance and attitudes, Appl. Acoust. 119 (2017) 119–127.
[18] Y. Lee, W. Gong, J. Jeon, Correlations between forward head posture, range of motion of cervicospinal area, resting state, and concentrations of the brain, J. Phys. Ther. Sci. 23 (3) (2011) 481–484.
[19] S. Jha, C. Busso, Analyzing the relationship between head pose and gaze to model driver visual attention, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 2157–2162.
[20] S. Li, Y. Dai, K. Hirota, Z. Zuo, A students' concentration evaluation algorithm based on facial attitude recognition via classroom surveillance video, J. Adv. Comput. Intell. Intell. Inform. 24 (7) (2020) 891–899.
[21] A.F. Abate, C. Bisogni, A. Castiglione, M. Nappi, Head pose estimation: an extensive survey on recent techniques and applications, Pattern Recogn., J. Pattern Recogn. Soc. (2022) 127.
[22] S. Li, Y. Dai, K. Hirota, Z. Zuo, A Students' Concentration Evaluation Algorithm Based on Facial Attitude Recognition via Classroom Surveillance Video, vol. 24, Fuji Technology Press Ltd., 2020, pp. 891–899.
[23] R. Kawamura, K. Murase, Concentration estimation in e-learning based on learner's facial reaction to teacher's action, in: Proceedings of the 25th International Conference on Intelligent User Interfaces Companion, 2020, pp. 103–104.
[24] Y. Lin, Y. Lan, S. Wang, A Method for Evaluating the Learning Concentration in Head-Mounted Virtual Reality Interaction, vol. 27, Springer, 2023, pp. 863–885.
[25] Z. Guo, Z. Zhou, J. Pan, Y. Liang, Engagement recognition in online learning based on an improved video vision transformer, in: 2023 International Joint Conference on Neural Networks (IJCNN), IEEE, 2023, pp. 1–8.
[26] J. Liao, Y. Liang, J. Pan, Deep facial spatiotemporal network for engagement prediction in online learning, Appl. Intell. 51 (2021) 6609–6621.
[27] T.-Y. Yang, Y.-T. Chen, Y.-Y. Lin, Y.-Y. Chuang, Fsa-net: learning fine-grained structure aggregation for head pose estimation from a single image, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1087–1096.
[28] H. Li, B. Wang, Y. Cheng, M. Kankanhalli, R.T. Tan, Dsfnet: dual space fusion network for occlusion-robust 3d dense face alignment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 4531–4540.
[29] N. Ruiz, E. Chong, J.M. Rehg, Fine-grained head pose estimation without keypoints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2074–2083.
[30] X. Zhu, Z. Lei, X. Liu, H. Shi, S.Z. Li, Face alignment across large poses: a 3d solution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 146–155.
[31] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.