Teacher Attention Analysis Method Based on Face Three-Dimensional Feature Point Coordinate Transformation and Viewpoint Visualization and Applications

Bing Gong(✉)

Information Engineering School, Eurasia University, Xi'an, People's Republic of China
gongbing@eurasia.edu

Abstract. In the process of quantitative analysis of classroom teaching, it is often necessary to evaluate individual indicators related to attention such as "whether teachers' teaching attracts students, whether they stimulate learning enthusiasm" and "whether there is eye contact between teachers and students". The teacher's attention to the students reflects the teacher's teaching state and teaching psychology, and meets the requirements of observability, repeatability and interpretability of quantitative analysis of the classroom teaching process. In this paper, the method of mutual conversion and mapping between natural coordinates and classroom coordinates is used to project the face normal of the student's nose tip to the teacher's teaching video shot by the rear camera. The intersection of the rays in the classroom coordinate system on the camera plane must be calculated first, and then calculate the mapping point coordinates of the intersection in the teacher coordinate system. Finally, the viewpoint visualization in classroom teaching is realized.

Keywords: classroom teaching · face three-dimensional feature point · coordinate transformation · viewpoint visualization · teacher attention analysis

1 Introduction

Due to the lack of tools and software to automatically track the gaze of teachers and students in the process of quantitative analysis of classroom teaching, at present, the evaluation teachers enter the class or watch the classroom inspection videos to manually conduct individual evaluations related to attention. Evaluate. At the same time, limited by the professional level and energy of the evaluation teachers, it is difficult for the evaluation teachers to quantitatively analyse and evaluate the attention of the teachers and all the students in the class [7].

As shown in Fig. 1, by analyzing the posture of the teacher's head (back of the head), the teacher is looking at the students on the right side of the center to teach. Guide, listen to the class with one's own computer as the carrier, and this is a very common phenomenon in many classroom teaching at this time. Combined with the gaze

Z. Zhan et al. (Eds.): ICBDIE 2022, AHCS 5, pp. 536–545, 2023.
https://doi.org/10.2991/978-94-6463-034-3_55

Fig. 1. A classroom teaching scene in university education (Original by the author)

tracking of other students, it can be seen that the teacher's teaching still attracts the students' attention. At the same time, combined with the gaze tracking in the video stream, it can be seen that there is also eye contact between teachers and students. All in all, the teacher's teaching can stimulate students' enthusiasm for learning.

In the traditional classroom teaching of primary and secondary schools, classroom teaching is mainly based on teachers' teaching, supplemented by other forms such as classroom discussions [6]. The main source of teacher-student interaction is not asking and answering questions, but the eye-to-eye interaction between teachers and students. Generally speaking, teachers will pay more attention to high-achieving students, but less attention to the outstanding students; compared with the outstanding students, the students with better grades are more eager to have frequent eye contact with the teacher; at the same time, they are not interested in the teaching content [4]. In contrast, interested students will pay more attention to the blackboard and the teacher, and in return, the teacher will have more eye contact with interested students. When the teacher asks a question, students who are interested in the question or have mastered the knowledge point will raise their hands, eager to make eye contact with the teacher, and hope to express themselves, while students who are not interested or have not mastered the improvement knowledge point hardly dare to look up at the teacher [3].

Therefore, for classroom evaluation, the use of eye interaction-oriented attention analysis can effectively evaluate whether students are interested in and mastered the teaching content, and the attention analysis technology that tracks the gaze of many teachers and students in the classroom is very important and practical significance.

2 Visual Analysis of Teacher Attention

As the first classroom teaching mode with thousands of years of development history, it is still an irreplaceable teaching form now and in the foreseeable future. Therefore, how to analyse the first classroom teaching mode and realize personalized learning diagnosis based on the analysis results, learning path planning, personalized learning plan recommendation, teaching status visualization, learning intervention and other specific applications have become the current development direction and research focus and difficulty of quantitative learning analysis.

This paper focuses on the current blank areas of academic research and the areas that the industry urgently needs to break through, and uses the all-round high-definition video capture collectors, distributed pickups and other informatization methods of the Smart Education Laboratory built by Xi'an Eurasia University to capture the posture and behaviour of classroom students and psychological external manifestations, extract the factors that have causality, strong correlation and weak correlation to the teaching scene and behaviour, and carry out full-process, all-round, automatic, and non-perceptual information processing for the first classroom teaching panorama, and analyse causal and correlational factors. Using various technologies such as quantitative classroom teaching analysis based on attention analysis and quantitative classroom teaching analysis based on facial recognition, the high-definition video of classroom teaching is recognized frame by frame, and all participants participating in the classroom teaching process are identified. The level of investment in teaching and learning, and generate judgment models to provide benchmarks for further big data analysis.

The teacher's attention is visually analysed by using dual cameras. The rear camera estimates the teacher's head posture for gaze tracking, and the front camera is responsible for projecting the teacher's gaze onto the video plane for visualization of the teacher's attention [1].

Based on the visual analysis results of the teacher's attention, one or more classes can be counted as "the number of times, duration, and total time that each student has received the teacher's attention" to evaluate whether the teacher treats all students in the class equally or not. Focus on some or individual students [2].

Based on the visual analysis results of the attention of teachers and students, the eye contact between teachers and students can be analysed, that is, at a certain moment, when the teacher and a certain student both project their eyes to each other, it can be considered that the teacher and the student The students have made eye contact, and the number of eye contact in one or more classes is counted. It can be concluded that: (1) The more the number of eye contact between the teacher and all the students, the more students are willing to make eye contact with the teacher, indicating that the teacher has a good affinity, and the teaching effect is better; (2) If a student has more eye contact with the teacher, it indicates that the student has a stronger desire to learn.

Quantitative analysis of attention in the classroom teaching process requires a long-distance attention analysis of teachers and many students in the classroom to evaluate the attention distribution of teachers and students. Through the attention distribution of teachers, it can be concluded whether they paid fair attention to all the students tested in the classroom during the teaching process, rather than only paying attention to some students. Whether a student or all students pay attention to the content of the teacher's teaching indirectly evaluates whether the teacher's teaching stimulates students' enthusiasm for learning, and whether the content can catch the students' attention.

In the current educational environment, learners can only understand their learning effects through academic performance evaluations or other written evaluations given by teachers, lacking in-depth understanding of their own physiology or psychology, and unable to correctly understand their own learning [8]. Learning habits after points of interest. It is difficult for the tested students to clearly and intuitively understand their own learning situation during the learning process. Therefore, when learning problems occur, they are often unable to accurately focus on the root of the problem, or cannot accurately understand the root of the problem. For example, when some of the tested students have unsatisfactory academic performance due to lack of concentration, the individual tested students may not be aware of the problem of distraction. Therefore, the test students need tools to help them improve their self-assessment, and use data facts to help the test students understand their personal study habits. Through the inquiry of personal data, establish self-awareness and evaluation of personal learning attitude and other aspects, understand their own learning situation, conduct self-reflection, and actively adjust during the class process, form good study habits, and improve learning efficiency [5].

For teachers, although it is possible to make subjective judgments on students' classroom performance through the accumulation of senior teaching experience, this method is very dependent on teachers' teaching experience. Experienced backbone teachers may make more accurate judgments, but young teachers lack experience due to their lack of experience, there are certain difficulties in judging students' classroom performance [9]. At the same time, although many schools are currently developing the "small class teaching model", there is still a one-to-many phenomenon for classroom teaching. It is difficult for teachers to pay attention to each student's learning in the classroom. Therefore, the original analysis is insufficient and the empirical approach has certain limitations. It is important for teachers to fully and properly mobilize learners' attention. Only by mastering the characteristics of learners' attention can we more effectively grasp the situation of each learner, better organize the teaching content, and optimize the teaching strategy.

When the tested students are studying in the classroom, the computer can monitor the learners' current head posture in real time through technical means, and analyze the data of their attention. Complete the attention analysis of the student based on the attention situation data. At the same time, the collection of data should be completed without affecting classroom teaching, so as to ensure the non-perceptual, full-process, and fully automatic implementation of the analysis process.

3 Teachers' Attention Visualization Analysis Method

Due to the objective and inestimable characteristics of the classroom scene, such as the cluttered background and the large venue, and due to the limitation that the current camera equipment uses high-definition images, it is still not enough to meet the requirements of individual eye tracking, so it is necessary to directly detect and estimate through machine vision. It is difficult to get out the teacher's line of sight, but in the teaching scene, the teacher's head posture can basically be used to replace the general direction of his line of sight, so the visual analysis of attention for the quantification of the classroom teaching process is based on the teacher's head posture of.

Figure 2 shows the smart classroom laboratory of Xi'an Eurasia College. The teacher teaches about 10 students in the classroom in front of a blackboard with a width of about 4 m. In order to analyze and summarize the teaching skills of teachers and monitor the students facing the back in group teaching, three high-definition cameras are installed at point B above the middle of the rear wall (opposite the blackboard). At the same time, in order to ensure various teaching forms (such as grouping etc.) to monitor the teaching process of all classroom participants, and two high-definition cameras are installed on both sides of the wall, with a total of 9 high-definition cameras in the whole classroom.

Based on the video image data captured by the front and rear cameras shown in Fig. 2, we mainly use the following six functional modules for visual analysis of attention in the classroom teaching process.

(1) Data (video frame) acquisition: The classroom teaching video is collected through the 1080p high-definition PTZ camera based on the H.264 high-definition compression algorithm, and the video frames are separated to obtain image samples for attention analysis.
(2) Camera calibration: In order to improve the accuracy of head gesture recognition, it is necessary to use a convenient and accurate calibration method to calibrate the camera parameters.
(3) Face detection: Detect faces from discrete video frames using publicly available face detection algorithms.
(4) Face feature point detection: This section uses random cascade regression to obtain the coordinate information of 11 face feature points, which is used to provide two-dimensional information in the PnP solution.
(5) Head pose estimation: Based on a standard face model obtained by a statistical measurement, the mapping relationship between 2D/3D is obtained by solving PnP, and the rotation and translation matrix of the head pose is output.
(6) Teacher's viewpoint positioning: According to the rotation and translation matrix information of the head posture, the student's viewpoint is projected into the teacher's teaching video shot by the rear camera by using the transformation relationship of the spatial coordinates, so as to realize the visual display of the teacher's teaching attention.

The photography principle of the camera can be expressed by the following formula.

$$m = \lambda C[Rt]M \tag{1}$$

where λ is the scale factor, m represents the feature point coordinates in the image coordinate system, C represents the camera internal parameter matrix, R represents the rotation matrix, t represents the translation matrix, and M represents the three-dimensional coordinates of the face transition point in the target coordinate system.

First, use the calibration board to calibrate the camera to obtain the camera's internal parameter matrix C and scale factor μ, and then obtain the face feature point coordinate matrix m in the image coordinate system, and a standard face model obtained by statistical measurement is known, that is, the coordinates M of the face feature points in the human head coordinate system are known. Through these known conditions, the rotation matrix R and the translation matrix t can be obtained by using the above formula, and the rotation matrix and translation matrix can be obtained to know the rotation information of the teacher's head relative to the camera and the teacher's head in the actual coordinate system of the cameras position. The obtained information can further visualize the teacher's attention, that is, the viewpoint.

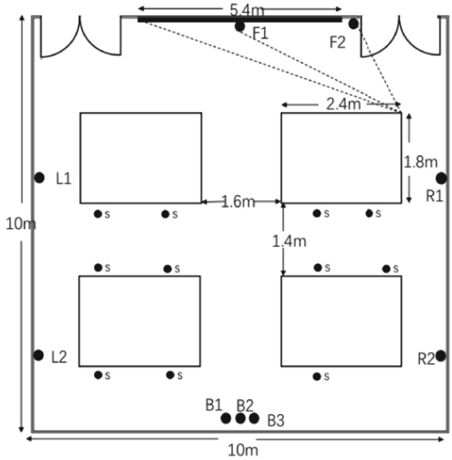

Fig. 2. The layout of a standard classroom and the installation positions of the two cameras in the front and rear

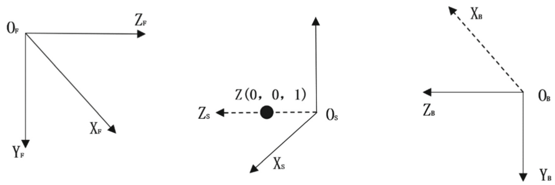

Fig. 3. 3 Local Coordinate Systems for Teacher Attention Visualization

The viewpoint visualization mentioned in this chapter refers to the projection of the face normal at the tip of the nose onto the teacher's teaching video shot by the rear camera, as shown in Fig. 3, the ray $O_S Z_S$ in the coordinate system $O_S X_S Y_S Z_S$ must be calculated first. The intersection points I on the plane $O_F X_F Y_F$, and then calculate the I' coordinates of the mapping point I in the coordinate system $O_B X_B Y_B Z_B$, it can be seen that the position information shown in Fig. 3 corresponds to the camera position.

The local coordinate system $O_F X_F Y_F Z_F$ of the front camera F (where the coordinate axis Z_F is perpendicular to the direction of the student is the positive direction), the local coordinate system $O_S X_S Y_S Z_S$ of the head of the subject S with the tip of the nose as the origin (by convention when the Y_S axis Vertical to the ground, the direction of the Z_S axis when it is vertical to the blackboard is the initial direction of the head, that is, the direction when the head is not rotated), and the local coordinate system $O_B X_B Y_B Z_B$ of camera B (the direction of Z_B perpendicular to the student is the positive direction).

When the teacher's head is rotated and translated, that is, the coordinate system $O_S X_S Y_S Z_S$ is rotated and translated relative to the $O_F X_F Y_F Z_F$ coordinate system, the following head rotation matrix R and translation matrix t can be obtained.

$$\mathrm{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{2}$$

$$\mathrm{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{3}$$

And suppose there is a point Z in the $O_S X_S Y_S Z_S$ coordinate system, and its coordinates are $(0, 0, 1)$, as shown in Fig. 3. According to the mapping relationship given by the above formula, the rotated Z point coordinates and the origin coordinates O_S are expressed in the $\mathrm{O_F X_F Y_F Z_F}$ coordinate system, respectively expressed as $O_S(t_1, t_2, t_3,)$ and $Z(R_{13} + t_1, R_{23} + t_2, R_{33} + t_3)$.

Finally, the representation of the teacher's viewpoint in the $O_F X_F Y_F Z_F$ coordinate system is obtained through the proportionality of the corresponding sides of the similar triangles. The specific calculation process is shown in various formulas.

$$\begin{bmatrix} X_F \\ Y_F \\ Z_F \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_S \\ Y_S \\ Z_S \\ 1 \end{bmatrix} \tag{4}$$

$$\begin{cases} x = t_1 + \Delta_x = t_2 - \frac{R_{13}}{R_{33}} t_3 \\ x = t_2 - \Delta_y = t_2 - \frac{R_{23}}{R_{33}} t_3 \end{cases} \tag{5}$$

The above transformation only represents the position of the teacher's viewpoint I in the coordinate system $O_F X_F Y_F Z_F$, but in order to visualize the viewpoint I, the image captured by the camera B needs to be displayed. As shown in Fig. 4, the figure can not only represent the position information of the viewpoint in the actual space of the coordinate system $O_B X_B Y_B Z_B$, but also the position information of the viewpoint in the coordinate system of the image captured by the camera B, and the viewpoint I is in the image The position information of is the value that needs to be solved.

The physical coordinates (x_K, y_K), (x_L, y_L), (x_M, y_M), $(x_N, y_{N)}$ of points K, L, M, and N in Fig. 4 can be obtained by measurement, and point K can be obtained by manual calibration. K, L, M, N mapping points K', L', M', N' coordinates in the image coordinate system captured by camera B $(x_{K'}, y_{K'})$, $(x_{L'}, y_{L'})$, $(x_{M'}, y_{M'})$, $(x_{N'}, y_{N'})$, the viewpoint $I(x_I, y_I)$ and it's in the image coordinate system There must be the following linear interpolation relationship between the mapping points $I'(x_{I'}, y_{I'})$.

$$\begin{cases} \frac{x_K - x_I}{x_K - x_N} = \frac{x'_I - x'_K}{x'_N - x'_K} \\ \frac{y_I - y_K}{y_L - y_K} = \frac{y'_I - y'_K}{y'_F - y'_K} \end{cases} \tag{6}$$

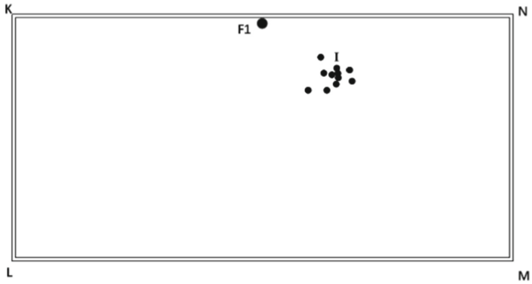

Fig. 4. Schematic diagram of the position information of viewpoint (Note: point F is the front camera, points K, L, M, N are the 4 corners of the blackboard plane respectively)

After transforming the above equation, the $I'\left(x_I', y_I'\right)$ solution equation was shown in the following equation can be obtained.

$$\begin{cases} x_I' = \frac{x_K - x_I}{x_K - x_N}(x_N' - x_K') + x_K' \\ y_I' = \frac{y_I - y_K}{y_L - y_K}(y_F' - y_K') + y_K' \end{cases} \tag{7}$$

4 Conclusions

Through the quantitative analysis of attention in the process of classroom teaching, some interesting preliminary conclusions are obtained as follows.

(1) Factors such as gender, major, family economic status, and growth environment (rural, township, or city) of the subject learners have no significant correlational influence on the length of time learners pay attention to the subject of education.
(2) After determining the learner's own role information, the improvement of teaching methods and the enrichment of teaching content can significantly improve the distribution of students' attention. That is to say, teaching methods and teaching content have obvious positive effects on learners' attention data. When the teaching method is to increase extracurricular knowledge, it is beneficial to increase the probability of learners by about 24%.
(3) At different stages of the learner's test, when the learner does not need to face the pressure of the test, the probability of the learner maintaining attention is significantly reduced, about 8%. Even if the teaching content can "cater" to the learners, using methods such as "introducing extracurricular knowledge" with relatively good attention data, the learners' attention is still about 10% lower than the pressure of exams. Therefore, in classroom teaching, appropriate introduction of examination pressure can better help the learners to maintain their attention, which can improve the teaching effect to a certain extent.
(4) Compared with the traditional teaching of book knowledge, the addition of extracurricular knowledge and classroom exercises has a certain effect on optimizing the

effect of classroom teaching. Among them, the teaching reform method of increasing extracurricular knowledge can better maintain the attention of the tested learners, and the learners' attention will disappear more slowly. Classroom drills can quickly improve the attention of groups of students and individuals, but it has no obvious optimization effect on the length of time to maintain a high degree of concentration, and the effect of delaying the disappearance of learners' attention is not obvious.

(5) Compared with the use of a single teaching method, a mixture of multiple teaching methods can improve the attention of the tested learners. Especially when teaching book knowledge, if multimedia means are used, the improvement of attention is about 33% compared with traditional oral teaching. When teaching extracurricular knowledge, traditional teaching should be used, because it can improve learners' attention data by about 13% than multimedia methods.

(6) Due to the small number of data samples, massive data analysis cannot be achieved. Therefore, the above conclusions are qualitative conclusions obtained through manual analysis of partial data. It is impossible to carry out correlation analysis on all-dimensional data such as teacher and student information, teaching methods, and teaching effects, and it is impossible to achieve a complete quantitative analysis. Therefore, it is necessary to further improve the collection methods and data mining of classroom teaching data.

In a word, based on computer vision, computer graphics, and high-performance parallel computing technology, this paper adopts a visual analysis method of student attention based on the recognition of head posture in a single image, and gradually develops the corresponding experimental system more satisfactory results. In view of the fact that the experimental system can only perform visual analysis of attention for a small number of people, in order to make this system practical, that is, to perform visual real-time analysis of the attention of 30 to 50 students in the classroom, we will further develop a large-scale computing cluster based on Research on parallel algorithms for head pose estimation with 4K high-definition video.

Acknowledgements. In this paper, the research was sponsored by the Social science fund of Shaanxi Province (Project No. 2019Q019), China Higher Education Association Special Project (Project No. 2020XXHYB13) and Shaanxi Provincial Department of Science and Technology Key R&D Program (Project No. 2022GY-317).

References

1. Dolecki M, Karczmarek P, Kiersztyn A, Pedrycz W. Face Recognition by Humans Performed on Basis of Linguistic Descriptors and Neural Networks [C]. New York: IEEE Press, 2016
2. Feng ZH, Huber P, Kittler J, et al. Random cascaded - regression copse for robust facial landmark detection [J]. IEEE Signal Processing Letters, 2015
3. Fu ZP, Zhang YN, Hou HY. Survey of Deep Learning in Face Recognition. New York: IEEE International Conference on Orange Technologies, 2014
4. Hang Z. A flexible new technique for camera calibration [J]. IEEE Transactions on pattern analysis and machine inteligence, 2000

5. Huber P, Hu G, Tena R, et al. A Multiresolution 3D Morphable Face Model and Fitting Framework. Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2016
6. Ren S, He K, Girshick R, et al. Faster R - CNN: Towards real - time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis &. Machine Intelligence, 2015
7. Mello SK. A selective Meta - analysis on the relative incidence of discrete affective states during learning with technology [J]. Journal of Educational Psychology, 2013
8. Xing Beibei, Yang Xianmin, Li Qinsheng. Education big data comes from collection technology [J]. Modern education technology 2016
9. Zhan Licai, Chen Fengyan, Meng Shimin. Technical support, dynamic structure and mechanism of online campus management under education big data research paradigm based on system dynamic theory [J]. Journal of distance education 2016