

Visualization Analysis of Learning Attention Based on Single-image PnP Head Pose Estimation

Li Dongxing^{1,a}, Liu Heng^{1,b}, Chang Wenquan^{1,c}, Xu Pengfei^{1,2,d}, Luo Zuying^{1,2,e,*}

¹ College of Information Science and Technology, Beijing Normal University, 19 Xijiekouwai Street, Haidian, Beijing, 100875, P. R. China

² Engineering Research Center of Virtual Reality and Applications, Ministry of Education, 19 Xijiekouwai Street, Haidian, Beijing, 100875, P. R. China

^a lidx@bnu.edu.cn, ^b 1041017843@qq.com, ^c changcwq@163.com, ^d

pengfei.bnu@qq.com, ^eluozy@bnu.edu.cn

*corresponding author

Keywords: Head Pose, Visualization, Attention Analysis, Classroom Evaluation.

Abstract: Learning attention analysis of students is the important indicator of classroom teaching/learning quantitative evaluation. Owing to the fact that the head-mounted eye tracker is expensive and unsuitable to be widely used in the large-scale classroom evaluation under expenditure limitation, in this paper, we use the PnP (Perspective-n-Point) method to estimate student's head pose for single-image. And then we achieve visualization of learning attention. Experiments demonstrate the following advantages of our method. (1) The method limits the average head-pose estimation errors under 4.88° with Biwi database. (2) This work has implemented student learning attention visualization analyses for three typical learning cases including engagement, attention, and disregard.

1. Introduction

Classroom teaching effect evaluation plays the key role in qualitative evaluation and promotion of elementary education [1]. Limited by the level of development of information technology, classroom teaching effect evaluation lacks the high-quantity evaluation approaches with high-sampling rate (1-2 times per second). Currently, the approaches with so-called low-sampling rate, mainly include CRS (Classroom Response System) using handheld devices, answering questions via touchscreens or surveying via e-questionnaire, so as to achieve the goals [1]. However, all the indicators whether classroom teaching arouses students' curiosity, inspires students to think, or develops students' concentration, have to be analysed by high-sampling rate and dynamic qualitative evaluation during classroom teaching. Therefore, the study on learning attention analysis of students with high-sampling rate and dynamic high-quality evaluation method has important theoretical research significance.

Head-mounted eye tracker [2] can be used to analyse students' learning attention visualization in class [3]. However, its cost is large and it is inconvenient. In order to achieve learning attention analysis more economically and practically in the large-scale classroom, sequentially this paper presents a method.

This study mainly has the following contributions:

(1) We analyse the demand and propose a relatively novel, economic and accurate method for students' learning attention visualization.

(2) Based on the visualization analysis results, this work can be used for three typical learning cases: engagement, attention, and disregard. Engagement represents the students focus on the teacher. Attention represents the students focus on the blackboard. Disregard represents the students focus on neither the teacher nor the blackboard.

2. Related Work and Method

In a standard classroom, the teacher faces with about 30 students for teaching in front of 4-meter-long blackboard. In order to monitor students' learning states and evaluate students' learning effect, one camera was usually installed at F point (the top of the front central blackboard). In order to analyse teacher's teaching skills, the other camera was usually installed at B point (in the middle of the back wall and opposite of the blackboard).

2.1. The Demand of Students' Learning Attention Analysis

Classroom teaching/learning process involves knowledge interaction and emotional interaction between teachers and students. The interactions often play an effective role in stimulating students' curiosity, inspiring students' deep thinking and making students focus on learning content. In theory, when the teacher is teaching, the students should be keeping sight on the teacher. When the teacher are writing on the blackboard, the students should be keeping mobile sight on some area of the blackboard. In contrast, when not interested in what the teacher teaches, students will frequently keep sight away from the teacher or the blackboard.

Therefore, it's an important indicator that we should analyse the states of students' attention during the process of classroom teaching, especially in an automatic way.

2.2. The Advantage and Disadvantage of Eye Tracker

Eye Tracker, an instrument specifically designed for attention analysis, is used to locate the direction of sight of one person. There are two types: Head-mounted eye tracker [2] and desktop eye tracker [3]. By detecting the eyeball pose and the head pose comprehensively, it can determine the screen area the person sees [4]. Both of them have high accuracy and sensitivity in attention analysis.

Desktop eye tracker is much cheaper. The weakness—it can only be used in fixed situation and the distance between the screen and the person is less than 1 meter—makes it not applicable. Comparably, Head-mounted eye tracker is more expensive. However, it is portable and can be far away from the object, which makes it meet the need in relative long-distance.

Moreover, we have to realize the problem clearly that one head-mounted eye tracker can be used for only one person and its cost is so large. Considering this, it is difficult to meet the need of in the large-scale classroom evaluation under expenditure limitation. Therefore, we have to find one more economical solution.

2.3. Students' Attention Analysis Characteristics in Classroom Teaching

During teaching, the fact that students are young, active and positive makes maintain order an important part in class. Although the teacher has the ability to allow the students not to leave the seats, to whisper to each other, he cannot be able to limit the students to look around. So the students can look everywhere without restraint in form of head rotation, instead of moving the eyeballs.

According to observations and the real situation of classroom teaching, there mainly has the

following characteristics of students' attention analysis in classroom evaluation.

(1) The equipments should be economical. For large-scale daily application, we must make full use of inexpensive equipments. This application is efficient and practical, and moreover can reduce the costs of use and maintenance.

(2) Don't distract students' attention. Attention analysis system must use non-contact test, so as not to distract students' attention.

(3) A large number of students in class. The number of students, to whom attention analysis system is used, is about 30-50 in a class.

(4) It can ignore student's eye rotation. Because of students' sight movement mainly through the head rotation, attention analysis only needs to recognize head pose of the students, without recognizing the eyeball pose.

2.4. Method

Based on the above discussion, this paper proposes a method for visualization analysis of learning attention based on Single-image and develops a corresponding system. The method has 6 steps sequentially.

(1) Video frames acquisition. With the help of LifeCam camera, we acquire classroom teaching videos and extract out of all frames.

(2) Calibration of camera parameters. To improve the head pose recognition accuracy, we use a convenient and accurate method in paper [5] to calibrate camera parameters.

(3) Face detection. Use the face detection algorithm invented by Professor Yu Shiqi to detect face from each video frame.

(4) Facial feature point detection. With the Stochastic cascade regression tree algorithm, we can get the coordinate information of 19 facial feature points, which is used to provide 2D information of each frame image.

(5) Head pose estimation. Using a standard facial statistical measurement model, we can get 3D coordinate information of each frame image, and then we can get the rotation and translation matrices, which mean the mapping relation between 2D and 3D models.

(6) Students' sight location. With the use of the information of rotation and translation matrices, and spatial coordinate conversion, we can project the gaze point on the frame image of the teaching video.

3. Experimental Results

3.1. The Head Pose Estimation Accuracy Test

In this paper, we use the Biwi Kinect database [6] to test the accuracy of the method. The Biwi Kinect database gives depth data and RGB images with head rotation in different directions of 20 individuals (14 men, 6 women), and also gives a total of 24 video frame sequences (a few people were recorded twice). The head position and the rotation angle of all the images are calibrated. The calibration errors for translation and rotation are 1 mm and 1 ° respectively and the range of the rotation angle is: pitch angle $\pm 60^\circ$, yaw $\pm 75^\circ$, roll $\pm 50^\circ$. Since our research focuses on whether students are paying attention to the blackboard area, we only concern about the student's head pose of the particular scope. As is shown in Table 1, the range of head rotation is reduced to: pitch angle $\pm 30^\circ$, yaw $\pm 45^\circ$, roll $\pm 50^\circ$ in this paper.

We use the head pose estimation method (R-CR-C+EPnP) proposed in this paper to calculate the head attitude rotation angle of all images in Biwi database that meet the need of attention analysis, so that we obtain an estimation of the pitch, yaw, roll and take the absolute value of the subtraction between the estimated value and the calibrated value as the error of the estimated value. We take the

average error of the estimated head rotation angle for all images as the mean error of the method. As is shown in Table 1, the average errors of the pitch, yaw and roll estimates obtained in this paper are 7.11° , 5.02° , and 2.51° . The fact that the total average error of the three rotation parameters (4.88°) is significantly less than the existing head pose estimation values [6-9] ($5.91^\circ - 25.21^\circ$), (R-CR-C + EPnP) proposed in this paper has a high accuracy of the test.

Table 1 Head pose estimation method comparison on average errors (unit: $^\circ$).

Method	Pitch	Yaw	Roll	Mean
random forests[16]	8.5	9.2	8.0	8.6
CLM[17]	18.30	28.30	28.49	25.21
CLM-Z[18]	12.03	14.80	23.26	16.69
CLM with GAVAM[18]	5.10	6.29	11.29	7.56
RF-TR-D[19]	5.15	7.8	4.8	5.91
R-CR-C + EPnP	7.11	5.02	2.51	4.88

3.2. Visualization Analysis of Typical Learning States

In order to apply the system to classroom teaching practice, we try to visualize the three states of students' learning states (engagement, attention, and disregard) in the classroom teaching process. In the experiment, the poses of teachers and students remain unchangeable as far as possible. In order to facilitate to display the visualization results, we take one of the teachers teaching image taken by the back camera as a student sight projection background, and project the student's sights (red) onto the teacher's teaching image. Specific effect is shown in Figure 1.

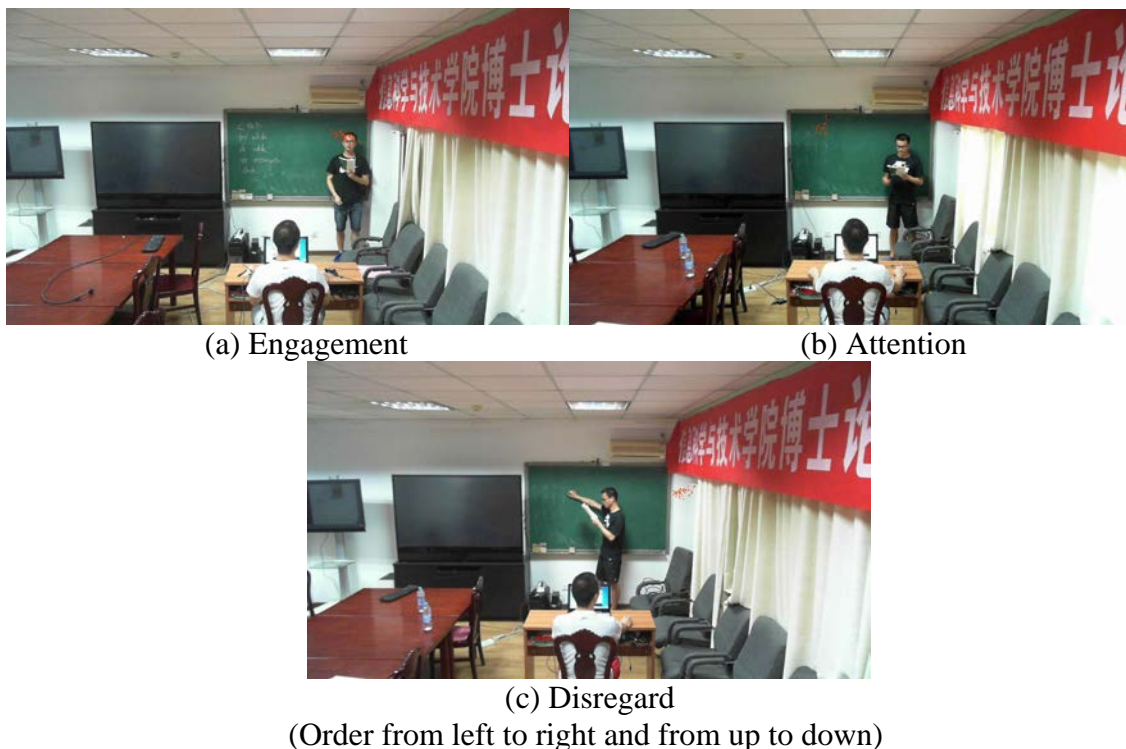


Figure 1 Attention visualization analyses for three typical learning cases.

As is shown in Figure 1 (a), the student is in a state of engagement. The student focuses on teachers, and he is eager to communicate with the teacher. The student's sights (red) based on the head pose are densely distributed near the head of the teacher. From the student's head pose, it also

can be seen that he is focusing on the teacher. As is shown in Figure 1 (b), From the student's head pose, it can be seen that he is focusing on the blackboard area where the student's sights (red) based on the head pose are densely distributed. As is shown in Figure 1 (c), the student is in a state of disregard. From the student's head pose, it can be seen that he is focusing on the area out of blackboard where the student's sights (red) based on the head pose are densely distributed.

It is proved that this method can be used to analyse the students' learning attention in classroom teaching to some extent.

4. Conclusion

Based on the technologies of computer vision and computer graphics in this paper, a method of learning attention analysis of students based on head pose recognition of single image is proposed, and a corresponding experiment system is developed. This paper validates not only the accuracy of head pose estimation, but also the results of the attention visualization analysis with this method. All have obtained the satisfactory results. The results show that this method can be used to analyse the students' attention in classroom teaching.

This experimental system can only focus on one person's visualization analysis and in order to put this system into practice, we do real-time analysis to 30-50 students in the classroom. For multi-person image, we will further improve head pose estimation algorithm, and meanwhile make good use of face recognition technology, parallel computing based on HPC, and 4K HD camera.

Acknowledgements

This research was sponsored by National Natural Science Foundation of China (61274033, 61271198).

References

- [1] Scott V. Clicking in the Classroom: Using a Student Response System in an Elementary Classroom [J]. *New Horizons for Learning*, 2014, 11(1).
- [2] <http://www.ergoneers.com/en/>.
- [3] Porta M, Ricotti S, Perez C J. Emotional e-learning through eye tracking[C]//Global Engineering Education Conference (EDUCON), 2012 IEEE. IEEE, 2012: 1-6.
- [4] Valenti R, Sebe N, Gevers T. Combining Head Pose and Eye Location Information for Gaze Estimation [J]. *Image Processing IEEE Transactions on*, 2012, 21(2):802-15.
- [5] Zhang Z. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [6] Fanelli G, Weise T, Gall J, et al. Real time head pose estimation from consumer depth cameras [M]//Pattern Recognition. Springer Berlin Heidelberg, 2011: 101-110.
- [7] Saragih J M, Lucey S, Cohn J F. Deformable model fitting by regularized landmark mean-shift [J]. *International Journal of Computer Vision*, 2011, 91(2): 200-215.
- [8] Baltrušaitis T, Robinson P, Morency L P. 3D constrained local model for rigid and non-rigid facial tracking[C] //Proceedings of Computer Vision and Pattern Recognition (CVPR), 2012: 2610-2617.
- [9] Kaymak S, Patras I. Multimodal random forest based tensor regression [J]. *IET Computer Vision*, 2014, 8(6): 650-657.