https://doi.org/10.20965/jaciii.2020.p0891

Paper:

A Students' Concentration Evaluation Algorithm Based on Facial Attitude Recognition via Classroom Surveillance Video

Simin Li, Yaping Dai, Kaoru Hirota, and Zhe Zuo[†]

School of Automation, Beijing Institute of Technology
No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, China
E-mail: zuzeus@bit.edu.cn
[†]Corresponding author

[Received October 20, 2020; accepted October 27, 2020]

To detect the students' concentration state in classroom, a DS (Dempster–Shafer theory)-based evaluation algorithm is proposed by measuring the students' Euler angles of their facial attitude. The detection of facial attitude angles can be implemented under the surveillance video with lower pixels. Therefore, compared with other methods for students' concentration evaluation, the proposed algorithm can be applied directly in most classrooms by the support of existing monitoring equipment. By using DS theory to fuse the concentration state of each student, the curve of students' overall concentration score changing with time can be obtained to describe the overall classroom concentration state. The design of the algorithm is proved to be feasible and effective under the dataset provided by computer front camera. The realization of the overall function effect of the algorithm is tested under the 35-person classroom video dataset. Compared with the average score from the questionnaire given by 20 reviewers, the accuracy of the proposed algorithm is about 85.3%.

Keywords: concentration evaluation, facial attitude recognition, Dempster–Shafer theory, classroom surveillance video

1. Introduction

Many methods based on computer vision have emerged to test students' classroom concentration with the development of the intelligent classroom. Among these methods, Mano et al. employed a model to identify and classify students' facial expressions and then assessed the emotions in learning activity [1]; L. B. and G. G. proposed a system to obtain the learner's concentration level in e-learning environment by detecting eyes and head movement [2]; Duan proposed a evaluation system based on machine vision to extract facial features to judge students' classroom concentration, which includes side face algorithm, head lifting (bowing) algorithm and eye closure algorithm [3]; By detecting and counting the faces in video, Sun took the sum of the effective head raising times and effective head lowering times as the concentration times [4]. Most of these methods evaluate the concentration state by extracting students' facial expressions or measuring their eyes' opening and closing degree, which can only be applied in detection environment with high pixels. However, the video provided by the monitoring devices in most classrooms cannot guarantee such high clarity. Besides, these methods mainly focus on single student's situation and pay little attention to the overall classroom concentration state.

Therefore, taking the hardware conditions of the monitoring facilities in most classrooms at present into consideration, a DS (Dempster–Shafer theory)-based concentration evaluation algorithm is proposed to detect the overall students' concentration state by measuring their facial attitude angles. Compared with the extraction of facial expressions, the detection of head attitude angles can be implemented under the surveillance video with lower pixels. Thus, the proposed algorithm can be applied directly in most classrooms by support of the existing monitoring equipment. For teachers, the overall concentration state in the classroom is more important than that of individual student. However, it is difficult for teachers to detect the students' overall learning state in class when there is a large proportion of teachers' and students' numbers. To make up for this deficiency, the proposed algorithm fuses the concentration state of each student by DS theory to obtain the overall concentration score in the classroom. The overall concentration score in class provided by the proposed algorithm can help teachers to review the teaching effect objectively and help improve teaching efficiency.

The proposed concentration evaluation algorithm mainly includes three modules: face detection, facial attitude angle measurement, and concentration detection. In the face detection module, MTCNN is used to extract face images from classroom surveillance video. In facial attitude angle measurement module, 2D key points in each face image is detected and matched with 3D face model, then the Euler angle is solved according to the rotation matrix. In concentration detection module, the angle values are compared with the objective standard values to score each student's posture separately, then the attitude scores of each student are fused with DS data theory to get the overall concentration score in the classroom. Af-

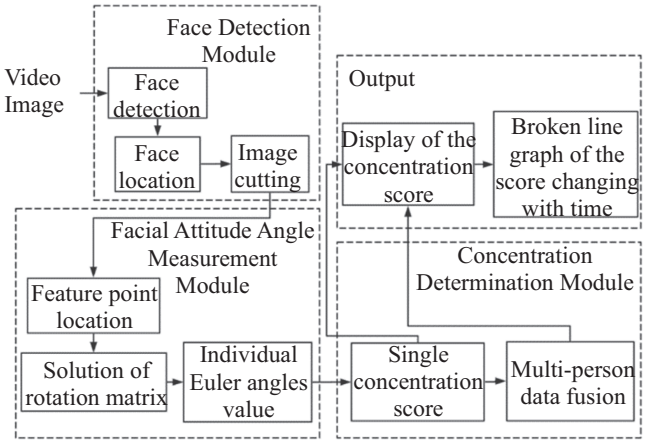


Fig. 1. The overall workflow of the algorithm.

ter the video reading is finished, the curve of the student's overall concentration state score about time is output.

The dataset used for experiment consists of two parts. The design of the algorithm is proved to be feasible and effective under the dataset provided by computer front camera. The realization of the algorithm's overall function effect in class is verified on the 35-person classroom video dataset. The obtained scores of students' concentration state are basically consistent with objective observation results.

Section 2 introduces the methods to realize the evaluation algorithm for concentration detection. The implementation and experimental verification of the algorithm are described in Section 3.

2. Concentration Evaluation Algorithm's Structure

In this part, the overall structure design of the algorithm is introduced and the theoretical methods of the three modules are elaborated.

2.1. The Algorithm's Structure

The overall workflow of the algorithm is designed according to the target function as shown in **Fig. 1**. The classroom video is input into the face detection module. The face detection is realized by using deep learning to cut the single-person facial images and sent them to the facial attitude angle detection module. The value of the facial postures' Euler angle is obtained through the feature points calibration and the solution of rotation matrix. Input the Euler angle value of single face posture into the concentration judgment module, and the single person concentration score is given. The multi-person concentration score is fused by DS theory to get the whole class concentration state score. The output consists of two parts: the quantitative display of the concentration score and the broken line graph of the score changing with time.

2.2. Face Detection Module

Face detection refers to locating faces from video images. In this algorithm, the deep learning method based on multi-task cascade convolution network MTCNN is used for face detection and extraction.

Based on cascade framework, MTCNN mainly consists of three sub-networks: P-Net (Proposal Network), R-Net (Refine Network) and O-Net (Output Network). The image processing is also divided into three steps from coarse to fine according to the three-layer structure [5].

Image pyramid is acquired by multi-scale transformation at first to adapt to face image detection of different sizes, and then the image pyramid is sent to P-Net for processing. P-Net mainly uses a full convolution network to obtain the bounding boxes of face regions and the bounding regression vector groups of these candidate frames from the image pyramid constructed in the previous step, and then evaluate and calibrate these candidate frames. Finally, non-maximal suppression (NMS) is used to remove a large number of repeated candidate regions. The output is the four coordinate information of the detected n candidate frames, the probability value of the candidate frame as a face, and the feature point positions of the face images. The input of R-Net layer is the output of P-Net layer. Compared with P-Net layer, there is a full connection layer (FC), which can be processed in more detail. O-Net adds a convolution layer on the basis of R-Net layer, and after more refined processing of candidate face areas, the output of this layer (as well as the MTCNN) includes three parts: bounding boxes' coordinate information and the probability that they are face images, and more accurate locations of 5 feature points of the face [6].

2.3. Face Attitude Angle Measurement Module

The video file captured and played by the camera is actually a 2D plane world, and the facial attitude angle is measured for the realistic 3D environment. Therefore, to determine the attitude angle, it is necessary to restore the true 3D pose of the object through the obtained 2D image information. That is to complete mapping transformation and calibration between four coordinate systems: 2D coordinates in pixel coordinate system, 2D coordinates in image coordinate system, 3D coordinates in camera coordinate system and 3D coordinates in world coordinate system. The mapping transformation relationship between the four coordinate systems is shown in Eq. (1) [7]. Where (u, v) is the feature points' 2D coordinate in the pixel coordinate system which can be detected directly, and (X_w, Y_w, Z_w) is the given 3D standard coordinate in the world coordinate system. (fx, fy) is the component of the focal length on the horizontal and vertical axes, and (u_0, v_0) are optical centers. The rotation translation matrix $(R\ T)$ is the target object attitude matrix. Make some approximate treatment to the image and camera parameters.

Considering that the focal length is equal to the image width and the optical center is close to the image center. Then the rotation matrix R which is required to determine

Table 1. 3D standard reference coordinates of key points.

Outer corner of the left eye	Outer corner of the right eye	Nose tip	Left corner of the mouth	Right corner of the mouth	Chin
−225, 170, −135	225, 170, −135	0, 0, 0	−150, −150, −125	150, −150, −125	0, −330, −65

the facial attitude angle can be solved [8].

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix}$$
$$= \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \quad \ldots\ldots (1)$$

In practice, it is only need to detect the position changes of key points on the face. Extracted outer corner of the left eye, outer corner of the right eye, nose tip, left corner of the mouth, right corner of the mouth and the chin as the six key points for observation. Using Dlib model to obtain the pixel coordinates of these six feature points, and referring to the biological frontal facial features, the 3D standard reference coordinates are shown in **Table 1**.

This algorithm uses the three Euler angles: Pitch, Yaw, and Row to describe facial posture [9], as shown in **Fig. 2**.

Therefore, the quaternion method is used to process the rotation matrix R to solve the Euler angles. The principle formula of the quaternion method is shown as Eqs. (2) and (3). Where α is the angle of rotation around the rotation axis, $\cos\beta_x$, $\cos\beta_y$, and $\cos\beta_z$ are the component of the rotation axis in the X, Y, and Z direction [10].

$$\begin{cases} w = \cos\dfrac{\alpha}{2} \\ x = \sin\dfrac{\alpha}{2}\cos\beta_x \\ y = \sin\dfrac{\alpha}{2}\cos\beta_y \\ z = \sin\dfrac{\alpha}{2}\cos\beta_z \end{cases}, \quad \ldots\ldots\ldots\ldots (2)$$

$$\begin{bmatrix} \varphi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \arctan\dfrac{2(wx+yz)}{1-2(x^2+y^2)} \\ \arcsin 2(wy-zx) \\ \arctan\dfrac{2(wz+xy)}{1-2(y^2+z^2)} \end{bmatrix}. \quad \ldots (3)$$

2.4. Concentration Determination Module

Combining the principle of facial attitude measurement module with experimental verification, the upper limit of the maximum angle of three Euler angles in practice is detected first. This is because in the theoretical calculation of facial attitude, the default maximum angle is taken as the theoretical maximum value of 180. However, in actual detection, the face image comes from the MTCNN as described in Section 2.2, when the deviation between the facial attitude angle and the standard position is too large, the image of the face area captured by camera will be too small to be detected by the MTCNN. Then the facial attitude angle of the object cannot be calculated. Therefore, in practical application, the three Euler angle values have a maximum detection limit. The standard value and scoring formula of Euler angle in single person concentration determination obtained from the test are shown in **Table 2**.

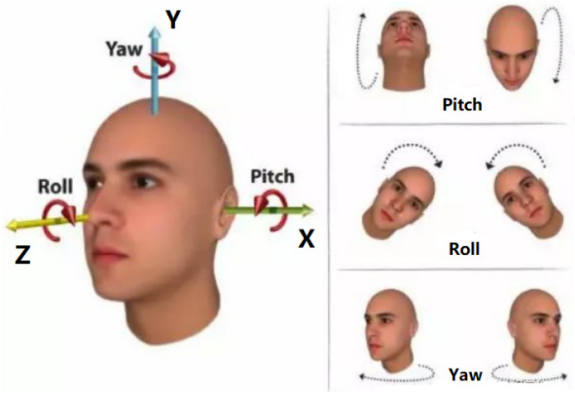


Fig. 2. Facial attitude Euler anglers.

It is considered that the three Euler angle values have the same influence factor on the concentration score result. Use DS evidence theory to fuse single-person score to get multi-player concentration score. The common data fusion methods include Kalman filter, Bayesian method, DS evidence theory, fuzzy logic, and so on. The reasons for choosing DS evidence theory can be summarized as follows:

i. The score of each student is synthesized by three Euler angles, and one of the advantages of DS synthesis formula is that it can integrate the knowledge or data from different experts or data sources.
ii. The data between the three Euler angles and the individual students are independent of each other, which meets the requirements of DS evidence theory for prior data.
iii. The prior data in the fusion process is the individual concentration score of each student, which is more intuitive than that in probability reasoning theory [11].

Therefore, compared with Bayesian method which is more inclined to classification [12] and Kalman filter method based on model prediction [13, 14], the DS evidence theory is chosen to fuse students' score.

After the processing of the previous function module, the Euler angle of each student's facial attitude in the

Table 2. Euler angles' scoring standard (degree).

Euler angle	Detection range	Standard value	Scoring formula
Pitch (front camera)	$[-180,-155]\cup[+155,+180]$	±180	$2-\lvert\text{Pitch}\rvert/90$
Pitch (classroom video)	/	170	$\lvert\lvert\text{Pitch}\rvert-170\rvert/180$
Yaw	$[-55,+55]$	0	$\lvert\text{Yaw}\rvert/180$
Roll	$[-46,+46]$	0	$\lvert\text{Roll}\rvert/150$

target picture has been detected. Regard each detected object's facial attitude angle as an information source, and each information source is independent of each other. Each information source includes a group of three Euler angles, and the score of each angle term can be regarded as an independent evidence body, so the multi-person concentration data fusion can be regarded as an information fusion problem of n independent information sources and each information source includes three evidence bodies. The video data is processed in the form of frames, so the information obtained from each frame can be regarded as the fusion in the spatial domain, that is, the fusion between three evidence bodies of each detection object. Using the DS theory formula shown in Eqs. (4) and (5) to merging data between information sources [15], the overall concentration score of students in the current classroom can be obtained.

$$m_i(h_i)=\frac{1}{K}\sum_{h_{i1}\cap h_{i2}\cap\cdots\cap h_{in}=h_i} m_i(h_{i1})\cdot m_i(h_{i2})\cdots m_i(h_{in}), \quad (4)$$

$$\begin{aligned} K &= \sum_{h_{i1}\cap h_{i2}\cap\cdots\cap h_{in}\neq\emptyset} m_i(h_{i1})\cdot m_i(h_{i2})\cdots m_i(h_{in}) \\ &= 1-\sum_{h_{i1}\cap h_{i2}\cap\cdots\cap h_{in}\neq\emptyset} m_i(h_{i1})\cdot m_i(h_{i2})\ldots m_i(h_{in}). \quad (5) \end{aligned}$$

Among the formula, $m_1(h_{1k})$, $m_2(h_{2k})$, and $m_3(h_{3k})$ represent three evidence bodies of facial attitude angel: score of Pitch, Yaw, and Roll respectively, and k represents the k-th student [16].

3. Experimental Verification and Result Analysis

The experimental verification of the overall algorithm is completed under two parts of dataset. In this section, the video provided by the computer front camera is firstly used to test the effectiveness and correct accuracy of the face detection module and face attitude angle measurement module. The realization of the overall function effect of the algorithm is tested under the 35-person classroom video dataset.

3.1. Experiment of Face Detection

In the experiment of face detection, the main parameters of MTCNN is set as shown in **Table 3**, which mainly includes three parts:

Table 3. Parameters set of MTCNN.

Parameters	Value
minsize	10
P-Net threshold	0.7
R-Net threshold	0.8
O-Net threshold	0.8
factor	0.709

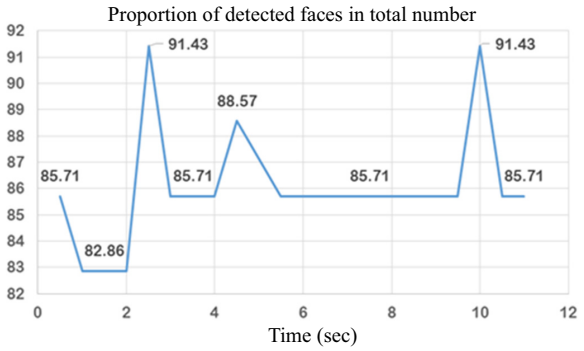


Fig. 3. Face detection rate versus time.

i. Minsize is the minimum image size for face detection. Images smaller than this size will be ignored for face detection.
ii. Factor is the iteration step value for reducing the pictures size, the images are transformed to meet the requirements according to this scale.
iii. Threshold of three network layers. The probability value of candidate boxes that contain face obtained by each layer network is compared with the threshold value of this layer network. Candidate boxes with larger probability value than this threshold can be sent to the next layer network for further screening.

Face detection effect is tested under the 35-person classroom video dataset. It is known that the frame rate of the video data used in the experiment is 24 frames per second. In the experiment, the video image is sampled every 12 frames (that is every 0.5 seconds), and a 10-second classroom video is randomly intercepted for detection. The detection rate is shown in **Fig. 3**.

The screen result of detection effect is shown in **Fig. 4**. From **Figs. 3** and **4**, it can be seen that for the real 35-person classroom with adverse conditions, including stu-

Fig. 4. The effect of face detection on MTCNN network under classroom.

Fig. 5. Part of the single face images obtained from the face detection module.

dents who climbing on the desk or be obscured and so on. The overall face detection rate of the used MTCNN network can reach for around 85%. For single person experiment with computer front camera, the detection rate of MTCNN network can be certainly reach 100%. Part of the single face images obtained from the face detection module are shown in **Fig. 5**.

3.2. Experiments for Face Attitude Angle Measurement

As the principle described in Section 2.3, the Dlib model is used to obtain the coordinates of the six key points firstly. Dlib model is the most widely used face feature point detect model. After a large number of image training by machine learning, the Dlib model can accurately mark 68 feature points on the face image. According to the number of 68 feature points obtained by Dlib model, the coordinates of six key points can be obtained as described in **Table 1**.

The marked effect of the six key points is shown in **Fig. 6**.

Take the video from computer front camera as input, after the step of rotation matrix calculation and Euler angle solution, the visualization effect of facial attitude measurement module is realized as shown in **Fig. 7**. The Euler angle of the currently detected facial attitude is displayed in the upper of the screen. The three numbers represent the values of yaw angle, pitch angle and roll angle in turn. After fine-tuning the coordinates under the world coordinate system of the key points on the face, the Euler angle of facial pose can be detected accurately as shown.

Fig. 6. Marked effect of the 6 key points based on Dlib model.

Fig. 7. Face attitude measurement result of front camera experiment.

Fig. 8. Face attitude measurement result of front camera experiment.

The facial attitude measurement experiment under the classroom surveillance video shows the result in **Fig. 8**. The Euler angle of facial attitude is calculated on the single face images cut by MTCNN as described Section 3.1 individually, and then summarized into the overall image of the classroom. Enlarge the images of single person, the results are shown in **Fig. 9**.

According to the principle of facial attitude measurement described in Eqs. (1)–(3), and the visualization mea-

Fig. 9. Facial attitude measure result of single person under classroom surveillance video.

surement results shown in **Figs. 7** and **8**, it can be seen that in the experiment of computer front camera with high pixel ratio, the Euler angle obtained by proposed facial attitude measurement method basically conforms to the objective situation. In the experiment under the classroom surveillance video dataset with more complex environment, the accuracy of face attitude angle measurement depends on the accuracy of facial key points marking. However, for some face images with large elevation angle, there are inevitably some errors in key points marking.

3.3. Experimental Verification Based on Front Camera

Under the computer front camera, the single-person concentration detection shows the result as in **Fig. 10**. The picture includes four parts of information:

i. The number of people detected in the current classroom is in the upper right corner of the screen.
ii. Personal concentration score is shown above the face image. When the score is greater than 75, it will be displayed in green font; if the score is less than 75, the font will be marked with red for warning.
iii. The corresponding warning signal is displayed above the individual score when the Euler Angle score is lower than the given threshold (Pitch, Roll, and Yaw corresponds to 'Look straight,' 'Look ahead,' and 'Sit up,' respectively).
iv. The overall score in the current classroom is displayed in the upper left corner of the screen. According to the score, the overall concentration atmosphere is divided into three grades: "focused," "relatively focused," and "unfocused." When the overall concentration score is higher than 70, it can be considered that the overall atmosphere of the classroom is in the state of "focused," and the score is displayed in green font. When the overall concentration score is between 30 and 70, the overall atmosphere is considered to be in a state of "relatively focused," and the score is shown in yellow font. When the overall concentration score is less than 30 points, the students in classroom are considered to be in the state of "unfocused," then the score will be shown in red.

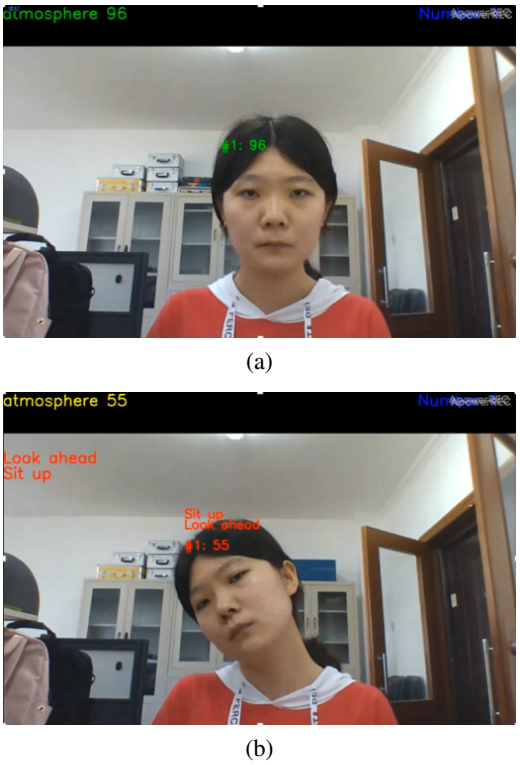


Fig. 10. Screen results of front camera experiment.

Table 4. Comparison of algorithm score result.

Artificial	Score score from algorithm	Accuracy
69.5	61	87.8%
43.5	37	85.1%
66.75	78	83.1%

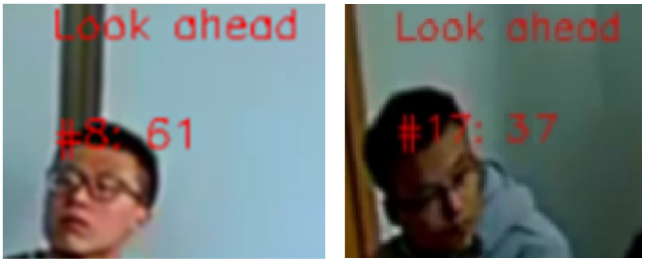
Select three images of student with different facial attitudes. Let 20 people grade the concentration state of the student in the picture separately, then take the average score of the artificial result as the objective standard of the student's concentration state. Compared with the score provided by the algorithm, the results are shown in **Table 4**.

As the data shown in **Table 4**, if regard the average result from the 20-people questionnaire as the objective score of the student concentration state, the accuracy of this estimation algorithm can achieve about 85.3%.

3.4. Experimental Verification Based on Classroom Surveillance Video

The concentration detection experiment under the classroom surveillance video shows the result as in **Fig. 11**. The picture includes four parts of information the same as introduced in Section 3.3.

(a)



(b) (c)

Fig. 11. Screen results of classroom video experiment.

Test the application effect of the algorithm in the traditional classroom on a 20-minute video data which is cut out randomly from the 35-person classroom surveillance video. The frame rate of video data used in the experiment is known to be 24 frames per second. In the experiment, the video image is sampled once every 720 frames (i.e., every 30 seconds). After the video is read out, the line graph of concentration score changing with time is obtained as shown in **Fig. 12**. The number of detected people at every time point is as shown in **Fig. 13**.

According to the time display on video monitor, among the 20 minutes of the classroom video used in this test, there is a 5-minute break between the 10th and 15th minutes. Combining **Figs. 12** and **13**, the graph line is shown in **Fig. 14**, where the interval of vertical dotted line is a 5-minutes of interclass time.

As can be observed from **Fig. 14**, in the first half of the video, although the concentration score and number of detected people fluctuated, the overall position remained in a relatively high level. Near the end of the class, students began to stir and the fluctuation of data increased. During the 5-minute break, the number of students facing the platform (camera) decreased due to their free activities, and the number of students detected showed a significant drop as well. By the 15th minute, the break was over, the students returned to their seats and faced the platform (camera) to back to the class state again. The overall concentration score and the number of people detected in the classroom showed a vibration rebound and gradually returned to the same situation as 10 minutes before the end of the class.

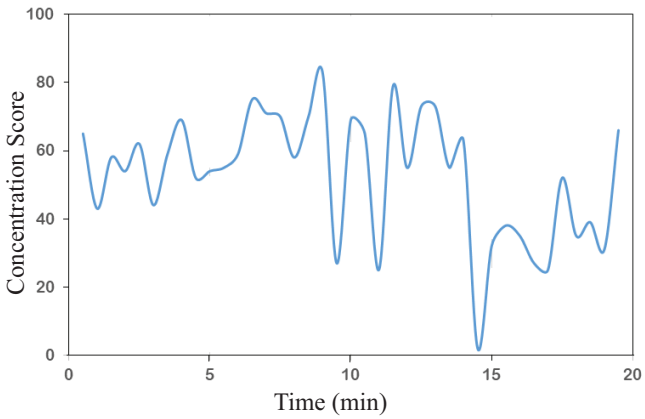


Fig. 12. Concentration score results of classroom video experiment.

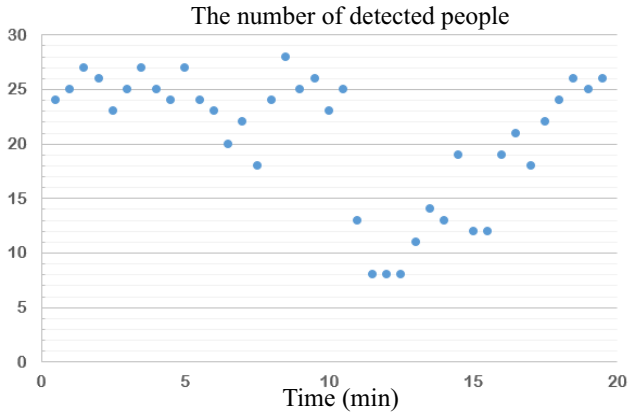


Fig. 13. The number of detected people at every time point.

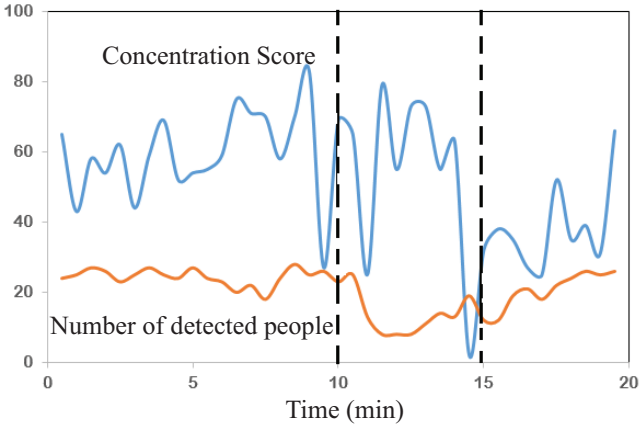


Fig. 14. Combination of concentration score and number of detected people.

4. Conclusion

To detect the students' concentration state in class, a DS-based evaluation algorithm is proposed by measuring the students' Euler angles of their facial attitude. By using the video data of classes provided by the existing

surveillance devices in classrooms, the objective description of the class is obtained. In the proposed algorithm, the MTCNN is used to extract face images from classroom surveillance video. The 2D key points in each face image is detected and matched with 3D face model, after that the Euler angle is solved according to the rotation matrix. The angle values are compared with the objective standard values to score each student's posture separately, then the attitude scores from each student are fused with DS data theory to estimate the overall concentration score in the classroom. After the video reading is finished, the curve of the student's overall concentration state score about time is output.

Compared with other methods for students' concentration measurement, the proposed estimation algorithm can be implemented under the surveillance video with lower pixels, so it can run under the support of the existing monitoring equipment in most classrooms. It also designed to give the overall concentration score of the class which helps teachers to review the teaching effect objectively and improve teaching efficiency.

The idea of using the facial attitude angle to detect student concentration is proved to be feasible under the data provided by the computer front camera. The overall function of the proposed algorithm is tested on the 35-student classroom video dataset. Compared with the average score from the questionnaire given by 20 reviewers, the obtained scores of students' concentration are proved to be consistent with objective observation with the accuracy for about 85.3%.

In future studies, we will try to add more criteria to the scoring rules. At present, the proposed algorithm scores the students' concentration state by detecting their facial attitude angles, so only the students facing the camera can be involved into the algorithm. However, many distracted students in reality (such as those who are playing mobile phones or sleeping) can not be extracted by face detection. To take those behavior into consideration, posture recognition need to be added into scoring. Then the influence of the students who can not be extracted by the face detection can be taken into account. Through refining these details, the algorithm will provide more comprehensive and reasonable estimation of concentration state.

References:

[1] L. Y. Mano, A. Mazzo, J. R. T. Neto, M. H. G. Meska, G. T. Giancristofaro, J. Ueyama, and G. A. P. Júnior, "Using emotion recognition to assess simulation-based learning," Nurse Education in Practice, Vol.36, pp. 13-19, 2019.

[2] K. L. B. and L. P. G. G., "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric," Procedia Computer Science, Vol.85, pp. 767-776, 2016.

[3] J. Duan, "Evaluation and Evaluation System of Students' Attentiveness Based on Machine Vision," M.S. Thesis, Zhejiang Gongshang University, 2018 (in Chinese).

[4] Y. Sun, "The Research of Pupil's Classroom Focus Based on Face Detection," M.S. Thesis, Hubei Normal University, 2016 (in Chinese).

[5] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From Facial Parts Responses to Face Detection: A Deep Learning Approach," IEEE Int. Conf. on Computer Vision (ICCV 2015), doi: 10.1109/ICCV.2015.419, 2015.

[6] J. Xiang and G. Zhu, "Joint Face Detection and Facial Expression Recognition with MTCNN," 4th Int. Conf. on Information Science and Control Engineering (ICISCE), pp. 424-427, 2017.

[7] J. Zhou, "Vision-based Human Pose Estimation in Smart Classroom," M.S. Thesis, Shanghai Jiao Tong University, 2011 (in Chinese).

[8] D. Li, H. Liu, W. Chang, P. Xu, and Z. Luo, "Visualization Analysis of Learning Attention Based on Single-image PnP Head Pose Estimation," Proc. of the 2017 2nd Int. Conf. on Education, Sports, Arts and Management Engineering (ICESAME 2017), pp. 1508-1512, doi: 10.2991/icesame-17.2017.324, 2017.

[9] K. Yu and J. Yin, "Design and Study on Hybrid Gestures Based on Head Gestures and Facial Features," Computer Applications and Software, Vol.35, No.12, pp. 209-215+258, 2018 (in Chinese and English abstract).

[10] T. Liu and Y. Ling, "Research on Relative Coordinate Detection System of 3-DOF Spherical Motor," J. of Shaanxi University of Technology (Natural Science Edition), Vol.35, No.2, pp. 43-49, 2019 (in Chinese).

[11] B. Chen, "DS Theory of Evidence-based Decision Closeness Fusion Algorithm," Presented at the 4th China Command and Control Conf., 2016.

[12] J. F. Liu, L. Liu, and J. He, "High resolution remote sensing image change detection based on Bayesian method," Technology Innovation and Application, No.11, pp. 1-5, 2019 (in Chinese).

[13] S. Cui, H. L. Jiang, H. Rong, and W. Y. Wang, "A Survey of Multisensor Information Fusion Technology," Auto Electric Parts, No.09, pp. 41-43, 2018.

[14] Z. H. Chen and L. Q. Huang, "Online Multi-target Tracking Algorithm Based on Kalman Filtering and Multiple Information Fusion," Information & Communications, No.03, pp. 35-38, 2019 (in Chinese).

[15] D. Han, Y. Yang, and C. Han, "Advances in DS evidence theory and related discussions," Control and Design, Vol.29, No.1, pp. 1-11, 2014 (in Chinese with English abstract).

[16] Y. Dai, F. Yang, H. Zhao, Z. Jia, and K. Hirota, "Auto Analysis System of Students Behavior in MOOC Teaching," Acta Automatica Sinica, Vol.46, No.4, pp. 681-694, 2020 (in Chinese with English abstract).

Name:
Simin Li

Affiliation:
School of Automation, Beijing Institute of Technology

Address:
5 Zhongguancun South Street, Haidian District, Beijing 100081, China

Brief Biographical History:
2019 Received B.S. degree from School of Automation, Beijing Institute of Technology
2019- M.S. Student, School of Automation, Beijing Institute of Technology

Main Works:
• Intelligent classroom, Deep learning

Name:
Yaping Dai

Affiliation:
School of Automation, Beijing Institute of Technology

Address:
5 Zhongguancun South Street, Haidian District, Beijing 100081, China

Brief Biographical History:
1990-1994 Ph.D., Beijing Institute of Technology
2002- Professor, Beijing Institute of Technology

Main Works:
• Networked control, Internet of things and e-experiment, Data fusion, Targets identification and tracking in video, Smart city relative
• Published 1 translation book and more than 80 papers, won "Quality Course Certification" from Beijing Municipal Education Commission

Membership in Academic Societies:
• Beijing Automation Association, Vice-Chief-Director

Name:
Kaoru Hirota

Affiliation:
School of Automation, Beijing Institute of Technology

Address:
5 Zhongguancun South Street, Haidian District, Beijing 100081, China

Brief Biographical History:
1982-1995 Professor, College of Engineering, Hosei University
1995-2015 Professor, Tokyo Institute of Technology
2015- Professor, Beijing Institute of Technology

Main Works:
• Image pattern recognition, Intelligent robotics, Fuzzy control, Artificial intelligence
• K. Chen, F. Yan, K. Hirota, and J. Zhao, "Quantum Implementation of Powell's Conjugate Direction Method," J. Adv. Comput. Intell. Intell. Inform., Vol.23, No.4, pp. 726-734, 2019.
• L. Chen, M. Zhou, M. Wu, J. She, Z. Liu, F. Dong, and K. Hirota, "Three-Layer Weighted Fuzzy Support Vector Regression for Emotional Intention Understanding in Human-Robot Interaction," IEEE Trans. on Fuzzy Systems, Vol.26, Issue 5, pp. 2524-2538, 2018.

Membership in Academic Societies:
• The Institute of Electrical and Electronics Engineers (IEEE), Life Member
• International Fuzzy Systems Association (IFSA), Fellow, Past-President
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT), Honorable Member, Past-President

Name:
Zhe Zuo

Affiliation:
School of Automation, Beijing Institute of Technology

Address:
5 Zhongguancun South Street, Haidian District, Beijing 100081, China

Brief Biographical History:
2004-2008 Ph.D., Beijing Institute of Technology
2009- Master Tutor, Beijing Institute of Technology

Main Works:
• Control theory and internal combustion engine control system application, high pressure fluid wave theory and fuel injection system
• Published more than 20 papers in SCI, EI, and other fields, and four authorized invention patents