Study on Visualization of Different Teacher Behavior Based on Teacher Experience during Trial Class

Sho Ooi, Shunyu Yao, and Haruo Noma
Faculty Information Science and Engineering, Ritsumeikan University, Shiga, Japan
Email: SHO.OOI@outlook.jp, syao@mxdlab.net, and hanoma@fc.ritsumei.ac.jp

***Abstract*—A new teacher at an elementary school, junior high school, or a high school has to teach classes from the first day. However, new teachers often find teaching difficult in a real-world teaching environment. In this paper, we analyze the gestures and behaviors of a beginner teacher and an expert teacher with the aim of increasing the teaching quality of beginner teachers. Image processing technology was used to automatically report visualized results of teaching behaviors. We devised the following approach: (1) a new teacher conducts a class in an environment replicating a real class, (2) the class behavior of the new teacher is systematically evaluated, and (3) new teachers objectively look back at their classes and derive insights for their development. The aim of this research is to evaluate and visualize the behavior of teachers during classes, focusing on (2) and (3). Specifically, we took videos of a trial class of expert and beginner teachers using a video camera (third-person view) and egocentric vision. The egocentric vision recognizes objects using YOLO algorithm, while the third-person view classifies the teacher behavior using Spatial-Temporal Graph Convolution Networks (ST-GCN) based on Open Pose. Then, we analyzed the differences between expert teacher behavior and beginner teacher behavior and visualized the results. In the behavioral analysis at STGCN, 'Writing on the board' constituted 74.5% of the approach of beginner teachers and 'Pointing to the board' was 36.9%, whereas 'Writing on the board' was 33.1% of the approach of expert teachers. Further, in the behavioral analysis by the YOLO algorithm, 'Writing on the board' was 41.1% and 'Talking at the front' 57.8% for beginner teachers, and 'Writing on the board' was 25.3% and 'Talking at the front' 74.0% for expert teachers. In other words, we confirmed that the experts were conscious of the whole classroom, and that beginners tended to do lessons only by writing on the board.**

***Index Terms*—teacher behavior analysis, visualization of teacher behavior, deep leaning, image processing, automatic analysis**

I. Introduction

A new teacher at an elementary school, junior high school, or high school has to teach classes from the first day. In teacher training courses, student teachers can experience classes, however, this is different from teaching at an actual school. In other words, it is difficult to adequately prepare student teachers and new teachers for the practical experience. Student teachers can witness the approach of an actual teacher in practical training at schools, but this is as short as one month at the longest, and the supervisor is also present during the class, which creates a difference from the real-world situation. Further, there is a cram school in Japan that provides experience in teaching prior to becoming a teacher, but this motivation differs from that of established school teachers. In other words, it is difficult for new teachers and student teachers to practice in an environment that accurately simulates the context of a real class. For this reason, new teachers without class experience are unable to respond to situations with quick judgments, and conspicuously conduct classes according to previously prepared teaching plans [1], [2]. Therefore, we think that a training system is necessary to improve the quality of classes. In order to achieve this, we have devised the following approach: (1) a new teacher conducts a class in an environment similar to a real class, (2) the class behavior of the new teacher is systematically evaluated, and (3) new teachers objectively look back at their classes and derive insights for their development. Fig. 1 illustrates the proposed system. The aim of this research is to evaluate and visualize the behavior of teachers during classes, focusing on (2) and (3). Specifically, we take videos of a trial class of expert and beginner teachers using a video camera (third-person view) and egocentric vision. The egocentric vision recognized objects using the You Only Look Once (YOLO) algorithm [3], and the third-person vision classified the teacher behavior using spatial-temporal graph convolution networks (ST-GCN) based on Open Pose [4], [5]. Then, we analyzed the differences between expert teacher behavior and beginner teacher behavior, visualizing the results.

II. Related Works

A. Teacher Nonverbal Communication

Studies on the analysis of the teacher's gaze using egocentric vision is abundant, with egocentric data obtained using a CCD camera and eye-tracker camera, and behavior patterns classified using a manual process [6]-[8]. It is necessary to manually classify a huge

Manuscript received March 1, 2020; revised July 10, 2020.

doi: 10.18178/ijlt.6.3.177-185

amount of the information from the teacher's gaze for a single class. In this study, we used the YOLO algorithm, which employs deep learning to automatically detect what is being looked at.

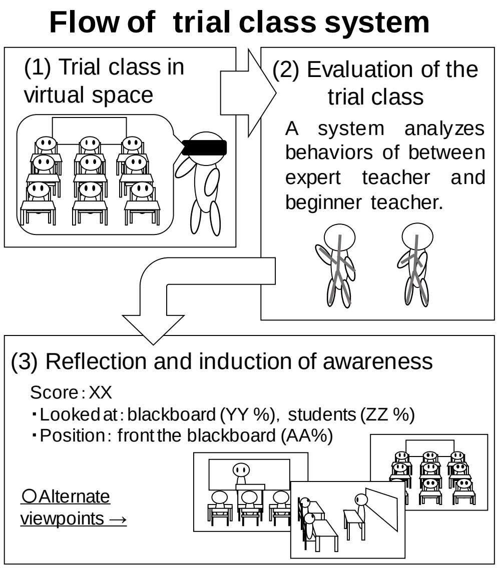


Figure 1. Note how the caption is centered in the column.

To understand the nonverbal communication of teachers, Kono *et al*. studied various teacher gestures [9], [10]. This was achieved by capturing the scenery of the lecture with a video camera, stopping the video every few seconds, and manually analyzing the teacher's attitude from the still images of teachers and graduate students. Nonaka *et al*. applied the cluster classification method to assess the behavior of teachers [11]. Zakaria *et al.* have examined teachers' nonverbal communication skills in online education, establishing the importance of nonverbal communication [12]. Barmaki focused on nonverbal communication between teachers and students, suggesting that there is an important role of gestures with crossed arms and open arms, and developed a simulation system to train teachers [13]. Bunglowala and colleagues suggest that nonverbal communication such as eye contact and gestures can provide students with a better understanding [14].

In addition, teacher behavior patterns have been classified manually. Kono *et al.* have classified 51 types of teacher behavior into 9 categories [15], and Nonaka has classified the attitudes taken by elementary school teachers into 51 types [16]. In each classification, numerous actions from the video images in the actual class are manually classified. In the trial class in this study, we did not anticipate that actions such as 'posture sitting on a chair or desk,' and 'lean on the desk' would be likely to be performed, and define 10 types of actions that are likely to appear in the trial class, classifying them automatically.

B. Action Recognition Methods

Research on behavior recognition is classified into that using wearable sensing systems and that using ubiquitous sensing systems. In wearable sensing, there are methods involving an acceleration sensor [17] and a wearable camera [18]. In recent years, acceleration sensor information has been identified by using deep learning methods such as a recurrent neural network (RNN) and long short-term memory (LSTM). On the other hand, in the wearable camera method, an action is estimated from the recognition of an object. In object recognition, SSD and YOLO use real-time processing as a deep learning method.

Ubiquitous sensing systems include RGB video-based and skeleton sequence-based methods, which recognize actions based on multiple features such as appearance and depth [19]-[24]. Skeleton sequence-based action recognition includes relative skeleton positions [25], [26], skeleton trajectory covariance matrix [27], and deep learning-based methods [4], [28]. Skeletal information detection methods include a method using a depth sensor such as Kinect and a deep learning base such as Open Pose [29]. The method using the depth sensor can process at high speed, but the accuracy is poor. On the other hand, the method using deep learning has a high accuracy of more than 90%, but the calculation volume is large. In recent years, a method for recognizing behaviors with a graph Convolution Neural Network (GCNN) from skeletal information has been developed.

As a GCNN-based action recognition, Wang *et al.* [30] were able to recognize actions by creating a graph with the detected object and skeleton positions as nodes. Yan *et al.* [4] created a graph with skeletal information in the time axis and spatial axis directions as nodes and attempted action recognition by convolution processing. Tang *et al*. [31] proposed a method to extract key frames by reinforcement learning and learn motions with GCNN.

In this study, we use deep learning-based YOLO for the egocentric vision data and the skeleton-based GCNN proposed by Yan *et al*. [4] for the third-person vision data, classify teacher behavior patterns using each method, then extract teacher behavior patterns.

III. Analysis of Teacher Behavior

A. Analysis of Egocentric Vision

In this study, we used the YOLO algorithm proposed by Redomon *et al*. The YOLO algorithm can perform real-time object detection and recognition. In addition to the YOLO algorithm, there are real-time object detection and recognition methods such as Regions with Convolution Neural Networks (R-CNN) [32], Fast R-CNN [33], Faster R-CNN [34], and Single Shot MultiBox Detector (SSD) [35]. The YOLO algorithm has a faster processing speed than R-CNN, Fast R-CNN, and Faster R-CNN, and has higher recognition accuracy than R-CNN, Fast R-CNN, and Faster R-CNN. In addition, the YOLO algorithm has slightly lower recognition accuracy than the SSD method, but this algorithm has a faster processing speed. Fig. 2 shows the YOLO detection system. This system has three processes; (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence. Fig. 3 shows a result retrieved from the YOLO algorithm using the egocentric video of a trial class.

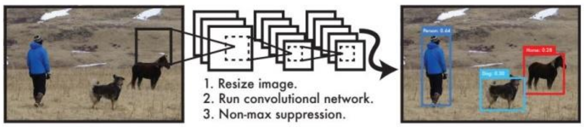


Figure 2. YOLO detection system [3].

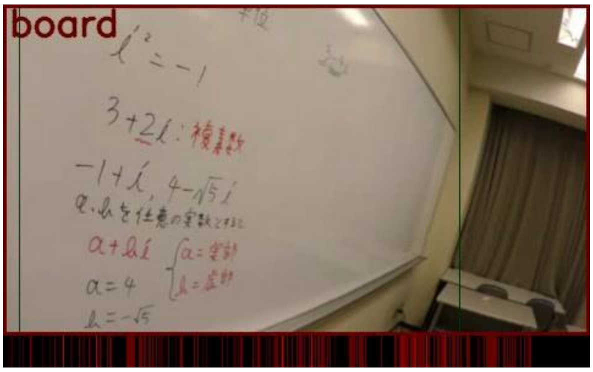


Figure 3. The result of object recognition using the YOLO algorithm during the trial class.

TABLE I. DATASET OF TRAINING FOR YOLO

Behavior	Total
Person	4,554
Face	1,854
Book (teaching materials)	1,858
Board	2,466
Arms	3,204

In this paper, the target objects of the egocentric vision are 'person,' 'face,' 'book,' 'board,' and 'arms.' Table I shows the numbers of data used for learning. At the time of learning, 10,000 epochs were learned, and the learning loss was 0.016. Moreover, we distinguished 5 behavior patterns from the combinations of the above target objects based on Arima's research [6]. The 5 behavior patterns were 'Writing on the board,' 'Talking at the front,' 'Talking at the side of the board,' 'Facing the class with teaching materials,' and 'Facing the board with teaching materials."

B. Analysis of Third-person Vision

As a number of classifications for teacher behaviors, Kono *et al.* proposed 51 kinds of behaviors for 9 categories: 'The teacher stands on the floor,' 'The teacher puts his hand on the desk,' 'The teacher has teaching materials,' 'The teacher makes gestures,' 'The teacher touches his body or his clothes,' 'The teacher holds my arm,' 'The teacher uses a board,' 'The teacher sits on the chair and the desk,' and 'The teacher walks in the classroom.' [15]. Nonaka categorized primary school teachers' behavior into 51 types [16]. The results in the previous research classify detailed behaviors from video in the actual class situation; however, in the trial class, there are no teaching desks or chairs, so there are no behaviors such as 'The teacher puts his hand on the desk,' or 'The teacher sits on the chair and the desk.' Therefore, in this study, we define 10 types of behaviors based on conventional research [8] as shown in Fig. 4. In addition, (A) to (J) in Fig. 3 are labels for each behavior.

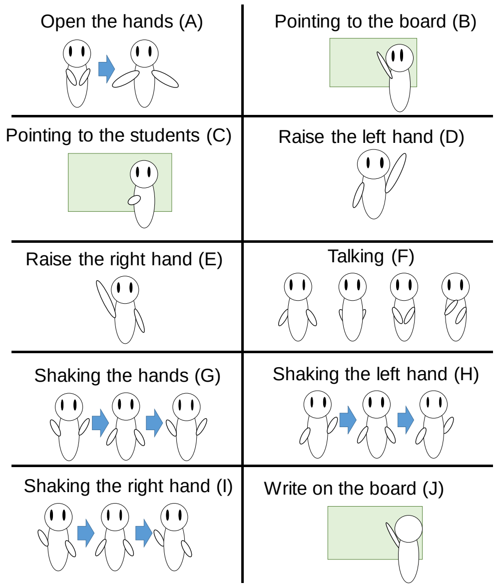


Figure 4. Ten types of teacher behaviors.

In this study, we analyzed the behavior of teachers using the spatial temporal graph convolutional networks (ST-GCN) proposed by Yan *et al.* ST-GCN [4] calculates the convolution between the skeleton data as well as the joints data using temporal information, and generates feature maps. The skeleton data was detected using the Open Pose algorithm. The Open Pose algorithm is a method that is able to detect the skeleton with a deep learning model from RGB video.

ST-GCN, proposed by Yan *et al.* [4], scans behavior data for 5 seconds and detects a behavior pattern; however, teacher behavior involves many variables. In this study, we define 3 seconds as one action unit, and assemble a model using action units to recognize the teachers' actions.

IV. EXPERIMENTS

A. Verify Accuracy of a Simple behavior Recognition

First, we needed to verify the recognition accuracy of a single behavior (of 10 types as shown in Table II) using the third-person vision described by section III. In other words, we verified the recognition accuracy the teacher behaviors using ST-GCN. Table II shows the numbers of training data and valid data. Equation 1 is a method for calculating recognition accuracy, where TP is True Positive, FP is False Positive, TN is True Negative, and FN is False Negative.

$$\mathrm{Accuracy}(\%) = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{1}$$

TABLE II. DATASET OF 10 TYPES OF TEACHER BEHAVIOR

Behavior	Total	Train	Valid
Open the hands (A)	240	211	29
Pointing to the board (B)	180	152	18
Pointing to the students (C)	126	116	10
Raise the left hand (D)	140	124	16
Raise the right hand (E)	175	133	42
Talking (F)	332	293	39
Shaking the hands (G)	222	187	35
Shaking the left hand (H)	214	187	27
Shaking the right hand (I)	214	184	30
Writing on the board (J)	200	175	25
Total	1992	1,749	243

B. Trial Class Experiment

Next, we analyzed the trial lecture videos with the egocentric vision of 'Student A' (male) and 'Teacher B' (male) using the model trained by the YOLO algorithm, and analyzed the trial class videos with the third-person vision of 'Student A' and 'Teacher B' using the model trained by ST-GCN. The system automatically recognized behaviors during the trial class and visualized a behavior state. Student A, who had a teacher license for mathematics, was a graduate student who wanted to be a teacher, and Teacher B, also with a teacher license for mathematics, is a part-time high school teacher. In this study, the duration of the trial class was 15 minutes, and the class was free for participants to attend. This time, two participants selected the high school mathematics range. Table III shows the details of two experimenters.

TABLE III. THE DETAILS OF TWO EXPERIMENTER

	Student A	Teacher B
Teacher license	Mathematics for junior high school and high school	Mathematics for junior high school and high school
Age	24	24
Sex	Male	Male
Job	Master course student	Part-time teacher (two years)
Preparation time for a trial class	5 minutes	5 minutes
Contents	Introduction of complex numbers	Introduction of quadratic function

In conducting this study, we explained the objectives and methods to all participants and obtained consent for their participation in the experiment. This study was conducted based on a review of the Research Ethics Review Committee at Ritsumeikan University (No. Kinugasa-Hito-2018-51).

V. RESULTS AND DISCUSSION

A. Results of a Simple Behavior Recognition

Fig. 5 shows the results of the simple behavior of the teacher. The average accuracy of the teacher behaviors is 88.9%. In terms of individual recognition accuracy, 'Writing on the board (J)' was 100%, 'Pointing to the students (C)' was 40%, and other behavior was 80-90%.

True Label \ Predection Label	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
(A) Open the hands	0.79	0.10	0.03	0.00	0.00	0.00	0.03	0.03	0.00	0.00
(B) Pointing to the board	0.00	0.78	0.00	0.00	0.00	0.06	0.00	0.00	0.17	0.00
(C) Pointing to the students	0.00	0.00	0.40	0.00	0.20	0.00	0.00	0.10	0.30	0.00
(D) Rise the left hand	0.00	0.07	0.00	0.79	0.07	0.00	0.00	0.07	0.00	0.00
(E) Rise the right hand	0.00	0.13	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00
(F) Talking	0.00	0.00	0.00	0.03	0.00	0.97	0.00	0.00	0.00	0.00
(G) Shaking the hands	0.00	0.03	0.00	0.00	0.00	0.00	0.94	0.00	0.03	0.00
(H) Shaking the left hand	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00
(I) Shaking the right hand	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.93	0.00
(J) Write on the board	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

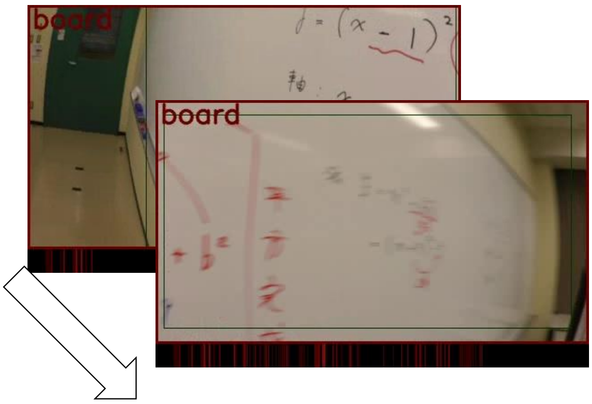
Figure 5. Accuracy of 10 types of teacher behaviors

The reason for the poor recognition accuracy for 'Pointing to the students (C)' in this study was that the method used ST-GCN based on Open Pose without depth information. In other words, when the teacher raised his hand, the system could not make a classification as to whether it was raised in front or behind, and the system failed in its behavior recognition. As a method of solving this problem, the 3D-based Open Pose could be used for the ST-GCN so that front and back movement in teacher behaviors can be detected, decreasing false recognition. Besides, we think that the system could be used to estimate the teacher's behavior by recognizing teacher movements in relation to objects.

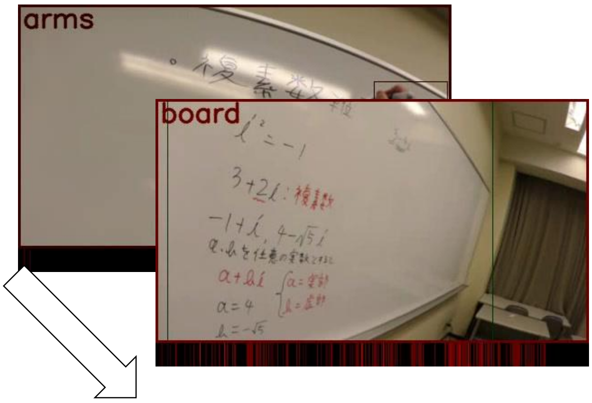
Based on the above results, we think that the model used with ST-GCN could perform classifications with 90% accuracy, and could be used for trial classes.

B. Results of the Egocentric Vision During Trial Class

Next, we describe the results from two trial classes using egocentric vision. Fig. 6 shows the results of recognized objects using egocentric vision, and Table 4 and Fig. 7 show the results of the object the teacher was looking at during each trial lecture.



(a) Student A



(b) Teacher B

Figure 6. Automatic classification of behavior during the trial class with egocentric vision.

TABLE IV. THE RATE OF GAZE IN A TRIAL CLASS WITH EGOCENTRIC VISION

Behaviors	Student A	Teacher B
Facing the class with teaching materials	0.007	0.007
Facing the board with teaching materials	0.000	0.000
Talking at the side of the board	0.000	0.000
Talking at the front	0.578	0.740
Writing on the board	0.414	0.253
Other	0.000	0.000

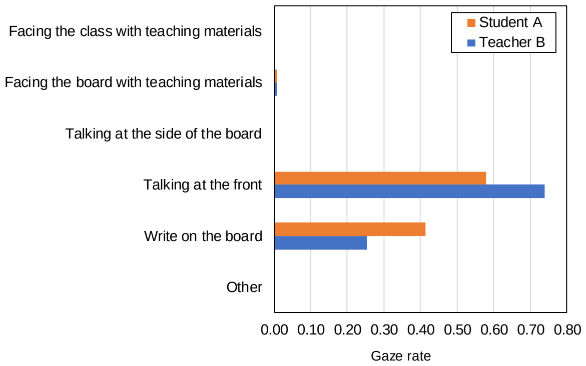


Figure 7. The rates of gaze in a trial lecture with egocentric vision.

In Fig. 6, it can be seen that Student A is looking at the board. As a detailed result, 'Writing on the board' is 41.1% and 'Talking at the front' is 57.8% in student A's case. 'Writing on the board' is 25.3% and 'Talking at the front' is 74.0% in Teacher B's case. In other behavior, there was no difference between Student A and Teacher B. In other words, teachers with practical experience teach the lesson with attention paid to students rather than 'Writing on the board.' This result is similar to the previous research by Arima *et al.* [6]. In addition, the reason why there are few actions related to teaching materials is that in this mock lesson, Student A and Teacher B engaged in the lessons without looking at the teaching materials (textbooks). Fig. 8 shows the rates of gaze in a conventional research with egocentric vision [6]. As a result, conventional research uses data from the actual class rather than a trial class. As a result, there is a slight difference in the behavior with the teaching materials compared to the trial classes. As a reason, the teachers were teaching in the classroom without a teacher's desk, we think that teachers had a textbook and were teaching.

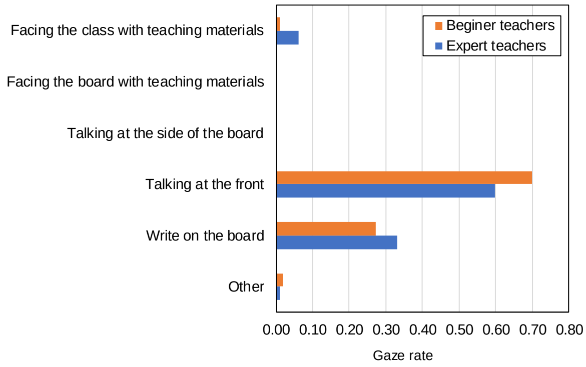
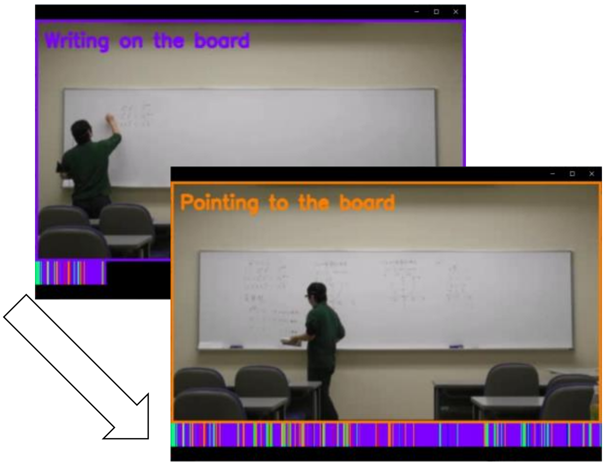
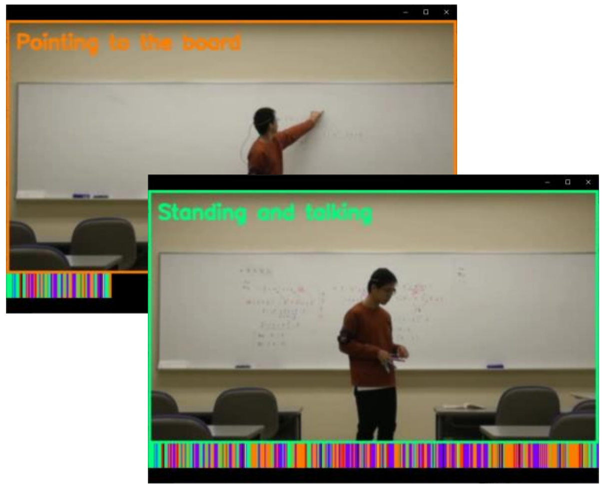


Figure 8. The rates of gaze in a conventional research with egocentric vision [6].

C. Results of the Third-person Vision During Trial Class



(a) Student A



(b) Teacher B

Figure 9. Automatic classification of behavior during the trial class with third-person vision.

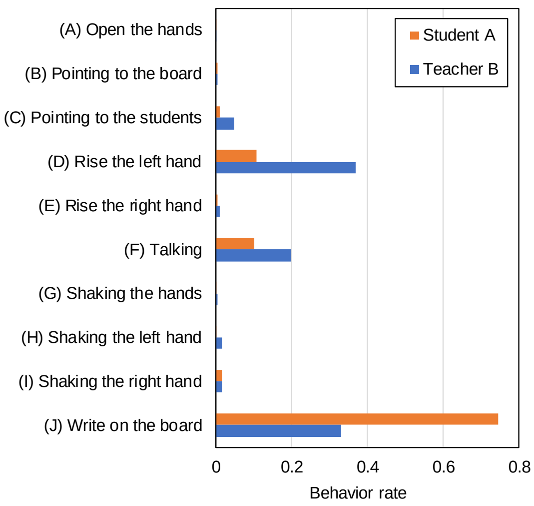


Figure 10. The rates of behavior in a trial lecture with third-person vision

Fig. 9 and Table V show the results of Student A and Teacher B's behavior during the trial class. For Student A, 'Writing on the board' is 74.5%. In other words, Student A was teaching with his body facing the board. On the other hand, Teacher B has the same ratio of 'pointing to the board' and 'writing on the board.' In other words, after Teacher B had written on the board, he pointed to the board and taught. Teacher B, unlike Student A, was found to be conducting the trial class with his body facing the student.

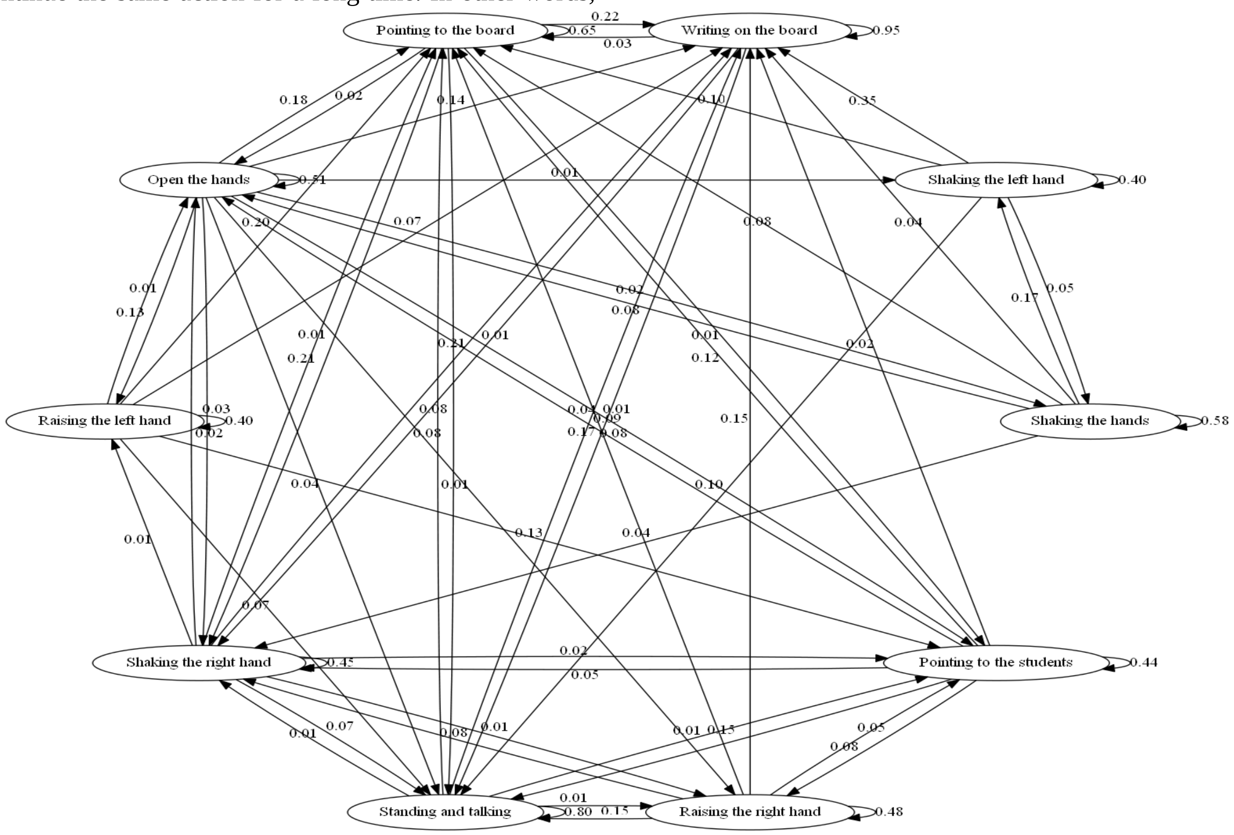
As seen in the results in Fig. 10, Student A performs the same action continuously, but Teacher B does not continue the same action for a long time. In other words, Teacher B uses nonverbal communication in the explanation during class.

TABLE V. THE RATE OF BEHAVIOR IN A TRIAL CLASS WITH THIRD-PERSON VISION

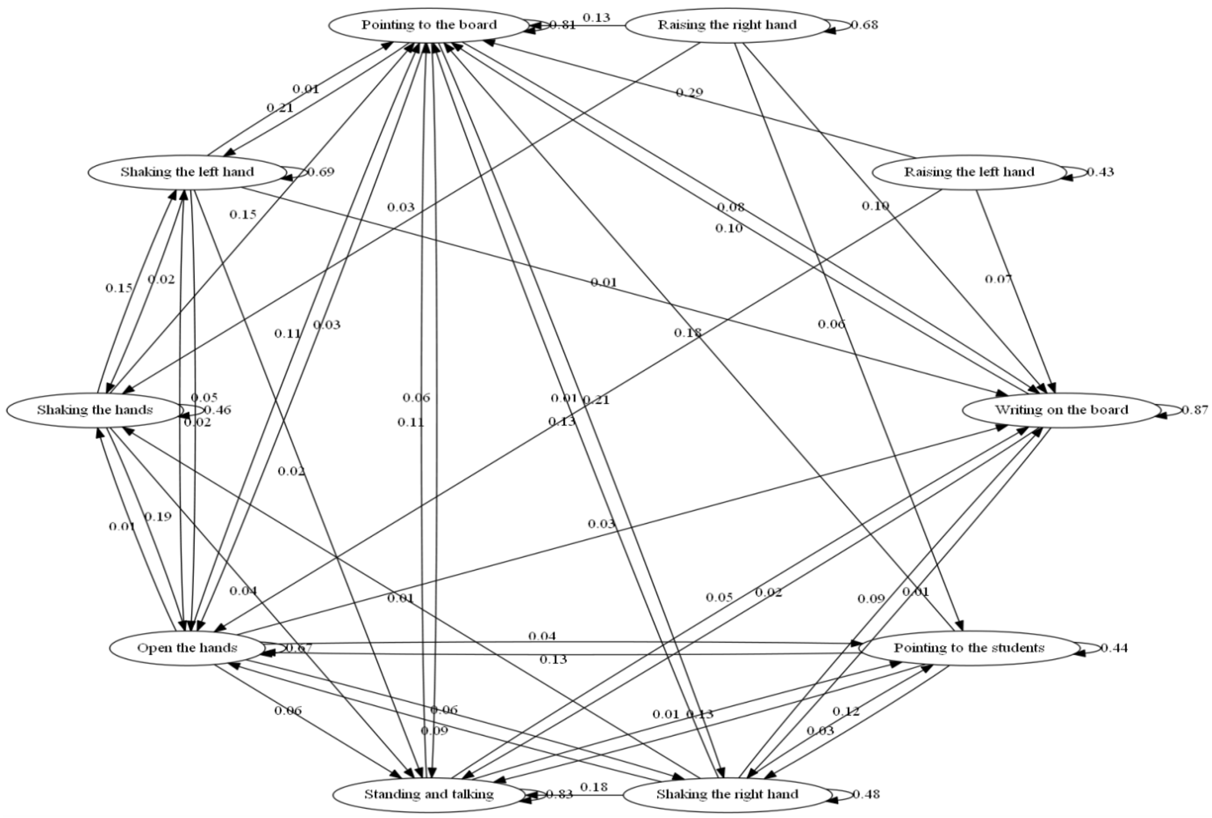
Behavior	Student A	Teacher B
Opening the hands (A)	0.011	0.049
Pointing to the board (B)	0.108	**0.369**
Pointing to the students (C)	0.005	0.01
Raising the left hand (D)	0.002	0.002
Raising the right hand (E)	0.005	0.005
Talking (F)	0.102	0.198
Shaking the hands (G)	0.003	0.004
Shaking the left hand (H)	0.002	0.016
Shaking the right hand (I)	0.017	0.017
Writing on the board (J)	**0.745**	**0.331**

D. *Stochastic Transition Model*

We generated an established transition model for the results visualized in Fig. 10. Fig. 11 shows the transition models of 10 types of behavior patterns in the trial classes of Student A and Teacher B. Referring to Fig. 11, Student A has many behavior pattern transitions, and Teacher B has a few behavior pattern transitions. This indicates that Student A is not patterned during the trial lesson, and the results for Teacher B suggest that some behavior is patterned during the trial lesson. For example, Teacher B shows a pattern of 'Writing on the board (J)' → 'Pointing to the board (B)' → 'Talking (F)' when explaining.



(a) Student A



(b) Teacher B

Figure 11. The transition models of 10 types of behavior patterns in the trial class.

VI. CONCLUSION

In this study, in order to conduct a quantitative evaluation of teachers in class, we automatically analyzed what they were looking at from egocentric video using a deep learning model. It was found that student teachers who have little experience teaching actual classes often turn their attention towards the board. For the purpose of quantifying teacher behavior, we attempted to establish a classification/recognition of teacher behavior as an initial step, incorporating Student A, who was new to teaching, and Teacher B, who had practical experience. We compared differences in behavior patterns between Student A and Teacher B from the egocentric vision and third-person vision.

Firstly, the difference between the first-person video images is illustrated. Student A, who had no practical experience, focused on the action of ‘Writing on the board.’ In contrast, Teacher B, who had practical experience, alternated between ‘Writing on the board’ and ‘Talking at the front,’ indicating that an experienced teacher considers there to be an appropriate balance between writing and speaking. We have shown that automatic classification is possible without significant differences compared to conventional classification [6].

Next, the results from the third-person viewpoint video were obtained. The classification accuracy by ST-GCN was obtained, and as a result, we roughly classified 10 types of motion and showed that they were associated with an average accuracy of 88.9%.

Moreover, when applied to an actual class, the difference in behavior during the class was obtained depending on the presence or absence of teaching experience. Student A, who had no teaching experience, focused on the action of ‘Writing on the board (J),’ whereas Teacher B, who had teaching experience, focused on ‘Writing on the board (J)’ and ‘Pointing to the board (B),’ as indicated by the egocentric vision.

In addition, we calculated the behavior pattern transition. The results showed that Student A did not have a fixed behavior pattern, and the behavior pattern transition was complicated. On the other hand, Teacher B treated a certain behavior pattern as a routine, for example, ‘Writing on the board (J)’ → ‘Pointing the board (B)’ → ‘Talking (F)’ when explaining.

Furthermore, in conventional methods to improve teachers' classes, they would take teachers' classes video and teachers watch their videos [12]-[14], [35]-[37]. However, it is difficult to watch back on the video because the teacher's job is busy. In this study, the system can automatically analyze and quantitatively evaluate teacher's behaviors by image processing using egocentric and third-person videos, therefore the teacher can easily review his/her video. By showing the results of visualizing the behavior to the teacher in this way, we think that the teacher can gain “awareness” about his / her behavior. In other words, teachers can improve their own behavior by looking back at their results of behavior quantitatively and improve the quality of their classes.

In the future, we will be able to visualize differences in behavior between experienced teachers and newly appointed teachers by analyzing the data of faculty teachers. In addition, we will increase the number of training data, deal with more detailed recognition of movement, evaluate the differences between experienced and newly appointed teachers quantitatively, and proceed with the development of a simulated class support system. In addition, by using 3D information, we can aim to improve the recognition rate and automatically analyze the behavior of teachers in class. In this study, only a single action recognition was performed for the simulated class, but by analyzing the timing in series of movements and using teachers' habits, voice data, and first-person viewpoint video, it will be possible to extract more details. In the future, I would like to further clarify the differences in behavior between newly appointed teachers and experienced teachers, and examine evaluation indices. Additionally, more detailed classification will be attempted by synchronizing movement analysis with speech.

Besides, we describe the analysis in the actual class. In the egocentric video, we think that by adding a function to detect the facial expression of the student, we can visualize the difference in fine movements by analyzing the degree of understanding of the student and whether or not he/she turned his or her gaze to the student.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Sho Ooi and Shunyu Yao conducted the research and developed the system, and wrote the paper; Sho Ooi, Shunyu Yao, and Haruo Noma discussed the evaluation method of the teacher behavior; all authors had approved the final version.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP 19K20750. We would like to thank Editage (www.editage.com) for English language editing.

References

[1] Ministry of Health, Labor and Welfare: Status of employee turnover of new graduates (2013). [Online]. Available: https://www.mhlw.go.jp/content/11650000/000369575.pdf

[2] Ministry of education. School teacher statistics survey (2018). [Online]. Available: http://www.mext.go.jp/component/bmenu/other/icsFiles/afieldfile/2018/03/28/139530301.pdf

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788.

[4] S. Yan, Y. Xiong, and D. Lin, "Spatial temporalgraph convolutional networks for skeleton-based action recognition," in *Proc. Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using part affinity fields," arXiv preprint arXiv:1812.08008, 2018.

[6] M. Arima, "An Examination of the teachers' gaze and self reflection during classroom instruction: Comparison of a veteran teacher and a novice teacher," in *Proc. 50th Annual Meeting of the Japanese Association of Educational Psychology*, 2008, p. 683.

[7] T. Sekiguchi and Y. Kawano, "The effect of class size on gaze patterns of teachers in class-examination by eye movement measurement," in *Proc. 50th Annual Meeting of the Japanese Association of Educational Psychology*, 2007, p. 682.

[8] T. Hirayama, A. Goto, M. Taniguchi, and H. Takeuchi, "Attempts to support the development of classroom practical skills, using the video record, based on the teacher's perspective," Meijo University Research Institute Bulletin, vol. 23, 2018.

[9] Y. Kono, "Nonverbal behavior of teachers in the classroom: Relationship between personality traits and attitudes of teachers," in *Proc. 29th Annual Meeting of the Japanese Association of Educational Psychology*, 1987, pp. 798-799.

[10] Y. Kono, "The relationship between the attitude of the instructor and the personality traits," Bulletin of Tokyo Gakugei University. Division 1, Educational Science, vol. 43, pp. 79-88, 1992.

[11] Y. Nonaka, "Exploratory study on attitude collection and classification taken by elementary school teachers during class," *Educational Psychology Research*, vol. 59, no. 4, pp. 450-461, 2011.

[12] N. S. Zakaria, N. A. Norul'Azmi, N. M. R. N. Yusoff, and H. Baharudin, "Cultivating Nonverbal Communication (NVC) skills among arabic language teachers in virtual classroom," *Creative Education*, vol. 10, no. 12, p. 2564, 2019.

[13] R. Barmaki, "Nonverbal communication and teaching performance," in *Educational Data Mining*, 2014.

[14] A. Bunglowala and A. Bunglowala, "Nonverbal communication: An integral part of teaching learning process," *International Journal of Research in Advent Technology Special*, pp. 371-375, 2015.

[15] Y. Kono and Y. Shinno, "Development of the coding system of teacher's posture (TP2002)," in *Proc. Japan Society for Educational Technology Conference*, 2002, pp. 627-628.

[16] Y. Nonaka, K. Numa, and W. Inoue, "The classification on the basis of improvement evaluation of nonverbal behavior skills through practice teaching contents," *Journal of Japan Society for Educational Technology*, vol. 35, pp. 5-8, 2011.

[17] K. Murao and T. Terada, "A motion recognition method by constancy decision," *Journal of Information Processing Society of Japan*, vol. 52, no. 6, pp. 1968-1979, 2011.

[18] S. Ooi, T. Ikegaya, and M. Sano, "Cooking behavior recognition using egocentric vision for cooking navigation," *Journal of Robotics and Mechatronics*, vol. 29, no. 4, pp. 728-736, 2017.

[19] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in Neural Information Processing Systems*, pp. 568-576, 2014.

[20] A. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 4489-4497.

[21] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4305-4314.

[22] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. V. Gool, "Temporal segment networks: Towards good practices for deep action recognition," in *Proc. European Conference on Computer Vision*, 2016, pp. 20-36.

[23] Y. Zhao, Y. Xiong, L. Wang, W. Zhirong, X. Tang, and D. Lin, "Temporal action detection with structured segment networks," in *Proc. IEEE International Conference on Computer Vision*, 2017, pp. 2914-2923.

[24] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1290-1297.

[25] H. Minamoto, M. Sano, and S. Ooi, "A human action recognition method for cooking rehabilitation," in *Proc. 78th National Convention of IPSJ*, 2016, pp. 263-264.

[26] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in *Proc. Twenty-third International Joint Conference on Artificial Intelligence*, 2013.

[27] A. Shahroudy, J. Liu, N. Tian-Tsong, and G. Wang, "NTU RGB+ D: A large scale dataset for 3D human activity analysis," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1010-1019.

[28] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part AffinityFields," arXiv preprint arXiv:1812.08008, 2018.

[29] X. Wang and A. Gupta, "Videos as space-time region graphs," in *Proc. European Conference on Computer Vision*, 2018, pp. 399-417.

[30] Y. Tang, Y. Tian, J. Lu, P. Li, and J. Zhou, "Deep progressive reinforcement learning for skeleton-based action recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5323-5332.

[31] R. Girshick, *et al*., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[32] R. Girshck, "Fast r-cnn," in *Proc. IEEE International Conference on Computer Vision*, 2015.

[33] S. Ren, *et al*., "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, 2015.

[34] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. European Conference on Computer Vision*, 2016, pp. 21-37.

[35] T. Kobayashi, "Comparison of the teacher's awareness by the difference in class analysis method," *Japan Society for Science Education*, vol. 33 no. 8, pp. 35-40, 2019.

[36] N. Hasegawa, *Study on Refraction Method and Trial Class*, Toho Publication, vol. 45, no. 2, pp. 99-107, 2017.

[37] K. Himeno, "A trial practice and evaluation of lesson reflection which focused on teacher's eye movement," in *Proc. Japan Society for Educational Technology Conference*, 2016, pp. 13-16.

Sho Ooi was born in Osaka, Japan. He received his Ph.D. degree in information science from Osaka Institute of Technology in 2018. Currently, he is an assistant professor with the Faculty of Information Science and Engineering in Ritsumeikan University, Japan. His research interests include computer vision, cognitive science, pattern recognition, and education technology. He is a member of the Information Processing Society of Japan (IPSJ), Institute of Electronics, Information, and Communication Engineers (IEICE), Institute of Image Electronics Engineers of Japan (IIEEJ), Robot Society of Japan (RSJ), and IEEE.

Shunyu Yao was born in Dalian, China. He received his B.S. degree in engineering from Ritsumeikan University and B.S. degree in information technology from Dalian University of Technology in 2019. Currently, he is a master course student with the Department of Digital Media Technology, Dalian University of Technology, Japan. His research interests include computer vision and machine learning. He is a member of the Information Processing Society of Japan (IPSJ).

Haruo Noma was born Ehime, Japan. He received his Ph.D. degree in engineering from Tsukuba University in 1994. Currently, he is a professor with the Faculty of Information Science and Engineering in Ritsumeikan University, Japan. His research interests include virtual reality, haptic interface, human interaction, cognitive science, pattern recognition, and education technology. He is a member of The Society of Instrument and Control Engineers (SICE), Human Interface Society (HIS), The Virtual Reality Society of Japan (VRSJ), Robot Society of Japan (RSJ), Institute of Electronics, Information, and Communication Engineers (IEICE), ACM, and IEEE.