Development of Scene Segmentation to Improve Work Efficiency of Learner Monitoring

Kaoru Sugita

Fukuoka Institute of Technology, Fukuoka, Japan

Email: sugita@fit.ac.jp

Received: June 01, 2024; Revised: August 03, 2024; Accepted: August 24, 2024; Published: December 27, 2024

Abstract

Since the outbreak of COVID-19, many universities have introduced video conference systems and e-learning systems, but these issues still make it difficult to obtain actual learning time. However, during operation of these systems, the participants or learner may not watch the video because they have also other tasks. For this reason, we have developed some prototype systems for monitoring learner behavior at watching learning content. In this paper, we describe the development of video scene segmentation based on a video correlation matrix, which reflects learner behavior, aiming to enhance the efficiency of learner monitoring. From our evaluation, we have been able to segment scenes from videos capturing learners based on the difference in correlation values between adjacent frames.

Keywords: Video processing, Scene Segmentation, Monitoring System.

1 Introduction

Nowadays, video conference systems are becoming popular for e-Learning and remote classes in universities and companies. However, some learners are lazy or busy, and such a system will cause these users to prioritize other tasks over watching videos. Since the outbreak of COVID-19, many universities have introduced video conference systems and e-learning systems, but these issues still make it difficult to obtain actual learning time.

Many learning support systems have been proposed to improve the effectiveness of lectures. These studies introduce a lot of educational approaches and systems aimed at enhancing self-directed learning (Mohamed, Mohamed, & Olfa, 2008), as well as effectively managing learning histories (Rapuano & Zoino, 2006) (Graf, Kinshuk, & Liu, 2008). In addition, some lecture videos have been made available on the Internet and utilized in various ways during classes (Deniz & Karaca, 2004) (Subbian, 2013). In these studies, the duration of usage and the rate of completion of e-learning content or lecture videos are often utilized as indicators for evaluating learning engagement. However, many authors don't focus on measuring the real learning time during e-Learning content and lecture video. Therefore, we have proposed a learning time monitoring system that evaluates the learning duration of e-learning content based on the time spent concentrating on watching the material (Sugita, Nakasone, Machidori, & Takayama, 2020) (Takegawa, Sugita, & Uchida, 2021). We have also introduced an averaged image to visualize viewer behavior during the learning content (Sugita, Implementation of an Average Image Composite Software for Viewer Visualizing Behavior During Learning Content, 2022).

In our previous research, we proposed a new concept called Universal Multimedia Access (UMA), which considers the digital divide causing from network environments, computer equipment, and user capabilities when accessing multimedia content (Maeda, Sugita, Oka, & Yokota, 2008). In addition, we implemented video content that supports text-based video search and voice reader functionality, utilizing text data used for generating subtitles (Sugita, A scene search method using subtitles appended to lecture video for supporting

self-learning, 2018). Furthermore, we refined an update version of user interface along with its functionalities (Sugita, Ito, Machidori, & Takayama, 2019). In these studies, we did not consider behaviors of learners utilizing the content during learning. Therefore, we developed a learning time monitoring system that measures gaze time while viewing a display device from the front by detecting faces in video frames (Sugita, Nakasone, Machidori, & Takayama, 2020). Through our system evaluation, we observed that the outcomes are influenced by learner behavior during the viewing of lecture content (Takegawa, Sugita, & Uchida, 2021). We have also visualized viewer behavior by creating an average image from recorded video frames of the viewer (Sugita, Implementation of an Average Image Composite Software for Viewer Visualizing Behavior During Learning Content, 2022). However, the system faced problems in distinguishing whether a learner was taking notes or working in other tasks while viewing the content. In this paper, we describe the development of video scene segmentation based on a video correlation matrix, which reflects learner behavior, aiming to enhance the efficiency of learner monitoring.

The paper is organized as follows. In Section 2, we introduce a video scene segmentation approach for monitoring learner. We present the implementation details in Section 3 and discuss the evaluation of video scene segmentation using the video correlation matrix in Section 4. Finally, conclusions and future work are given in Section 5.

2 Video Scene Segmentation Approach for Learner Monitoring

For improving the learning efficiency of live streaming and e-learning systems, it is essential to accommodate various types of learners. Learners can be categorized into diligent learners and lazy learners. Diligent learners concentrate when watching videos or taking notes during learning sessions. In contrast, lazy learners may engage in other activities or leave their seats during learning sessions.

In our previous work, we found an average image of recorded video have some visual differences (Sugita, Implementation of an Average Image Composite Software for Viewer Visualizing Behavior During Learning Content, 2022). However, it was difficult to distinguish actions such as note-taking and multitasking from average images.

In the video scene segmentation, individual frames are grouped into shots which are defined as a sequence taken by a single camera, and related shots are grouped into scenes which are defined as a single dramatic event (Kender & Yeo, 1998). In this study, we introduce a video scene segmentation for learner monitoring as shown in Fig. 1. Using a video scene segmentation to learner monitoring tasks, a lecturer can quickly find evidence of lazy behaviors during learning as shown in Segment2 (Scene2) and Segment3 (Scene).

We assume that learners are captured by a webcam while viewing educational content. In these videos, the webcam is stationary and only the subject moves within the video. Therefore, the video reflects only differences between frames corresponding to the learner’s actions and does not reflect the movement of the background. This fact suggests that the correlations between video frames contains valuable information about learner behavior.

A correlation matrix for video frames can be obtained by calculating the correlation among all recorded frames for a learner. The video frame correlation matrix can be derived by calculating the correlation among all frames recorded in the video of the learner. These correlation values reflect learner behavior as shown in Figure 2. These correlation values are expected to be high when the learner is diligently engaged in watching the learning content and low when the learner is disengaged or inactive.

Considering the characteristics of the video correlation matrix as mentioned above, the learner's video can be segmented into scenes based on correlation values. In particular, as the correlation value increases with smaller differences between frames and decreases with larger differences, the consistency of correlation values can serve as a criterion for segmenting the video.

3 Implementation

We have developed software for computing the video correlation matrix among video frames on a Windows PC using the software environment shown in Table 1. In the software, a video and a processing frame number are displayed on a window, while the processing time, processed frame, and the correlation value are output in a command prompt, as shown in Figure 3. The correlation value is calculated by ZNCC (Zero Mean Normalized Cross Correlation) supported by OpenCV 4.6.0. The software allows users to input such parameters as the video file name, the CSV file name (video correlation matrix), the scaling factor, the number of skipped frames, as well as the start and end frames.

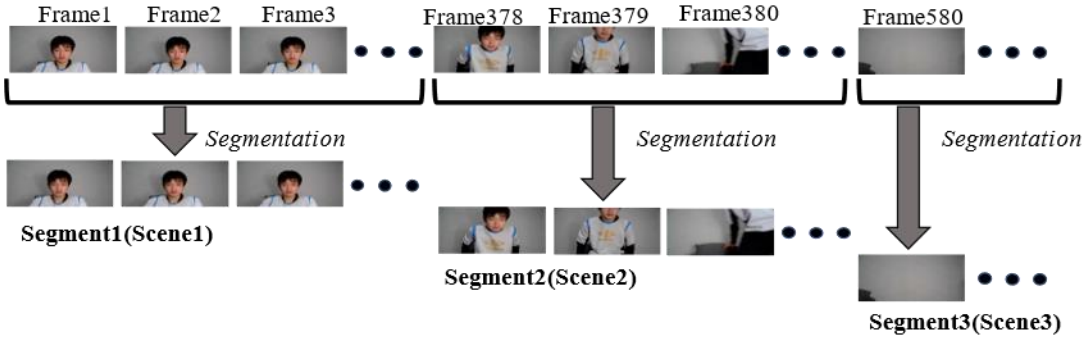

Figure 1: Video Scene Segmentation

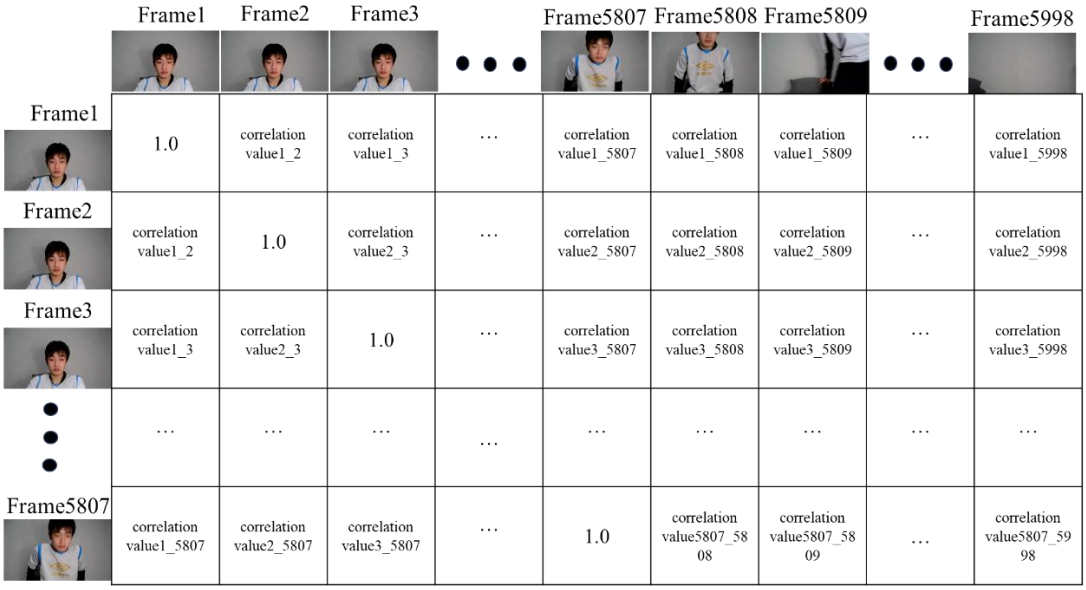

Figure 2: Video Correlation Matrix

Table 1: Software Development Environment

Software	Product
Integrated Development Environment	Visual Studio Professional 2021
Programming Language	Visual C++
Library	OpenCV 4.6.0

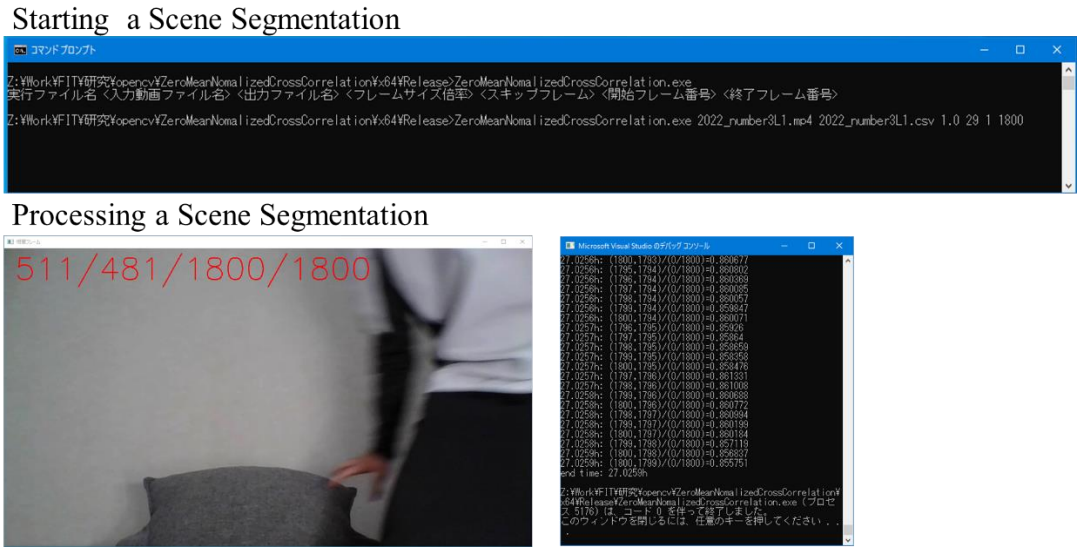

Figure 3: Implementation

4 Evaluation

We evaluated the video scene segmentation utilizing the video correlation matrix between video frames, employing the video depicted in Figure 4. The video features HD quality (1280 x 720 pixels) and runs at 30 frames per second (Total frames: 1786, with a duration of approximately 1 minute). Within this video, the learner stands up at approximately frame 406 and leaves the seat around frame 487. Our evaluation environment is detailed in Table 2.

4.1 Process Time for Video Correlation Matrix

When computing the video correlation matrix with a fixed frame size of 1280×720, a process time varied according to the number of frames, as shown in Figure 5. Both measured and approximate values are put on Figure 5.

The approximate values are obtained by fitting the measured values to a quadratic function for computing the video correlation matrix, and they are almost same values. The quadratic function y can be determined by the number of frames x according to equation (1):

$$y = 0.028747x^2 + 0.000989786x + 0.006862939 \quad (1)$$

In Figure 6 is shown the relation between frame rate and processing time by replacing the frame numbers with frame rates from Figure 5.

When computing the video correlation matrix with a fixed frame rate of 1 fps, a process time remained a constant value regardless of the frame size, as shown in Figure 7. From these results, we found that the processing time can be approximated by a quadratic function in relation to the frame rate, and it remains a constant value regardless of the frame size.

4.2 Visualization Results of Video Correlation Matrix

The visualization results of the video correlation matrix were conducted while varying the frame rate at 1280

x 720 pixel. In these results, the video correlation matrix remained largely unchanged for frame rates of 1 fps, 2 fps and 3 fps as shown in Figure 8 to Figure 10.

In the videos capturing learner, there are scenes with and without learner actions. These scenes can be segmented based on the continuity of correlation values in the video correlation matrix. Therefore, the frames are considered to belong to the same scene when the correlation values in the video correlation matrix exhibited the following continuity.

- The correlation values are consecutively close.
- The change in correlation value occurs continuously.

To verify the continuity of correlation values in the video correlation matrix, we visualized the difference in correlation values between adjacent frames, as shown in Figure 11. In the video correlation matrix, the diagonal elements of the video correlation matrix are 0. Consequently, in this visualization result, the difference in correlation values between adjacent frames and the diagonal elements equals the correlation value in adjacent frames. For this reason, the difference in diagonal elements becomes larger than the difference among neighboring values. Considering this fact, we can observe that the difference becomes particularly large around frame 208 and frame 406 in this visualization result.

Table 2: PC Environment of Performance Evaluation

Device	Product
CPU	Ryzen 9 5900X
Memory	80GB DDR4-3200
GPU	GeForce RTX 3080
SSD	2TB
OS	Windows 10 Pro 22H2

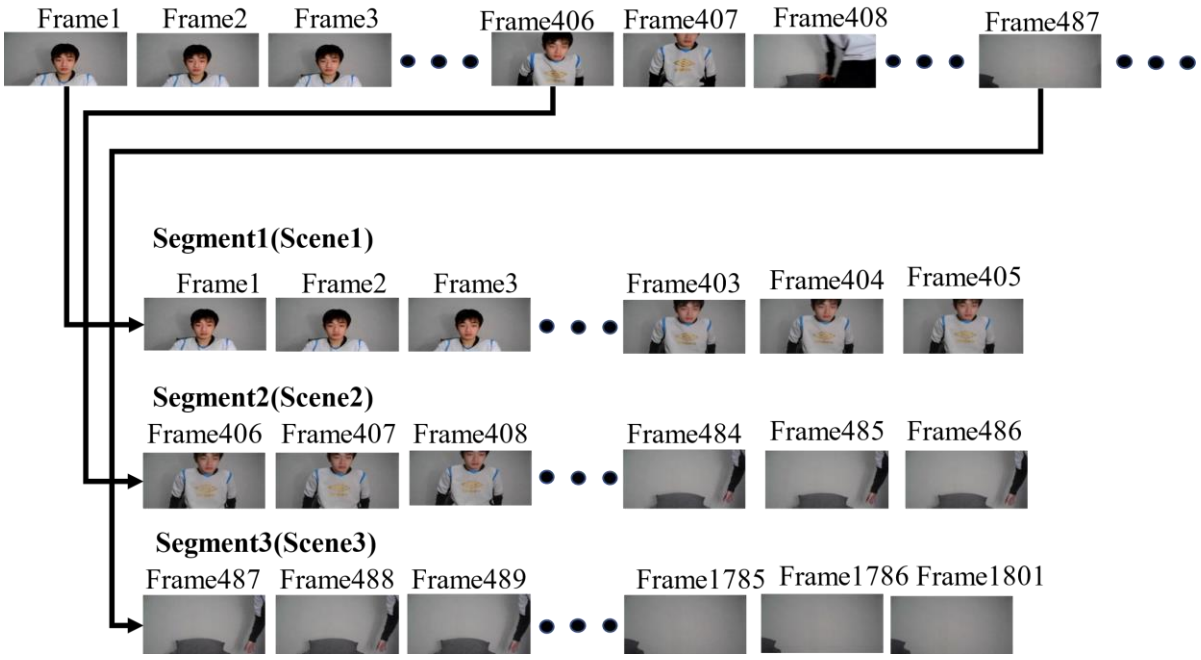

Figure 4: Organization of Video Frames

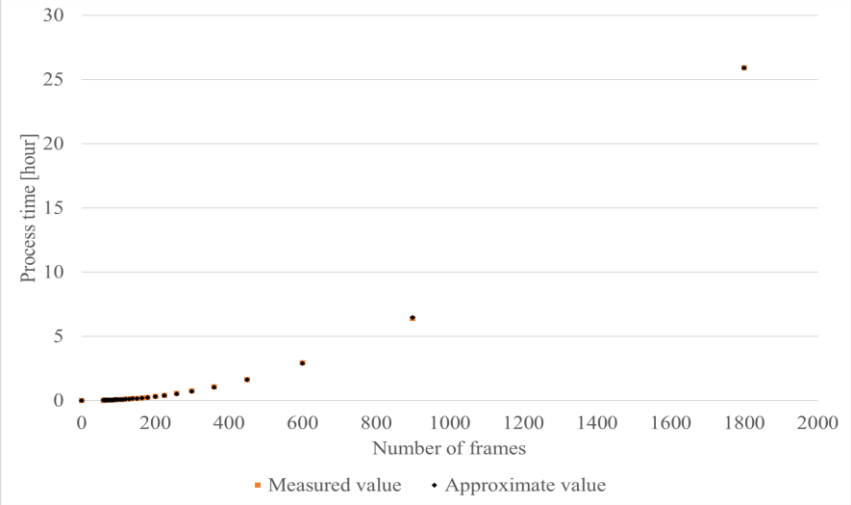

Figure 5: Process Time vs. Number of Frames

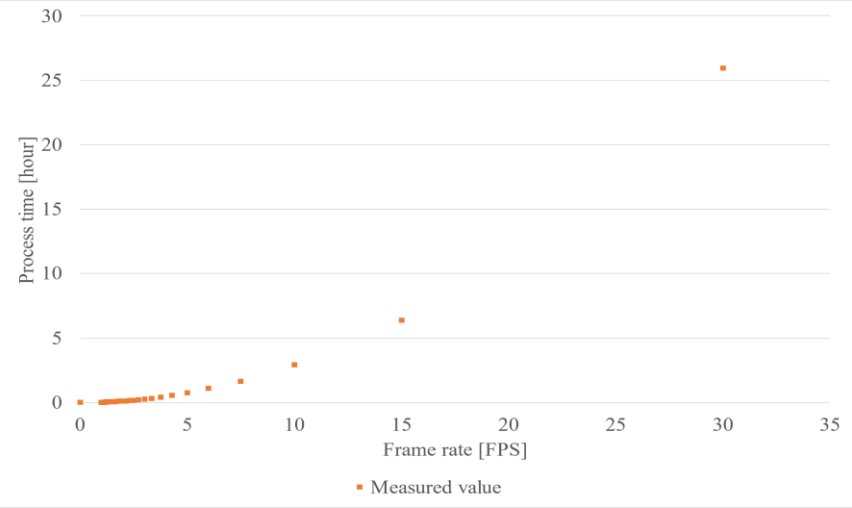

Figure 6: Process Time vs. Frame Rate

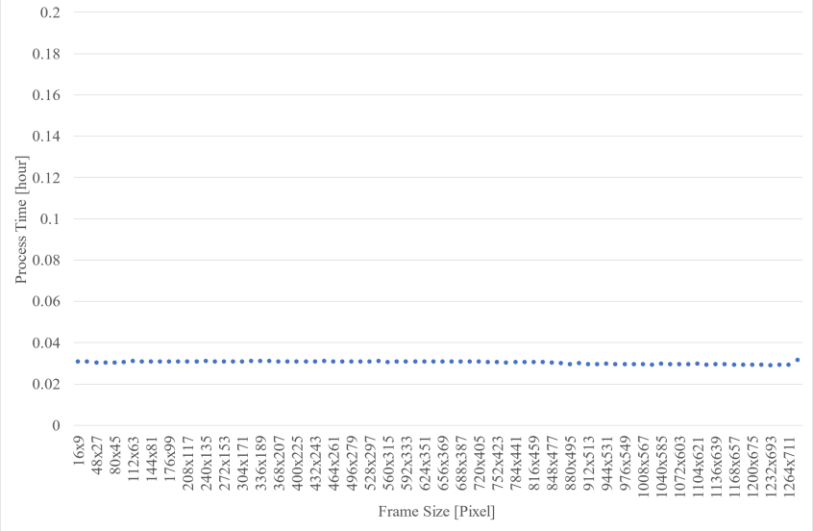

Figure 7: Process Time vs. Frame Size

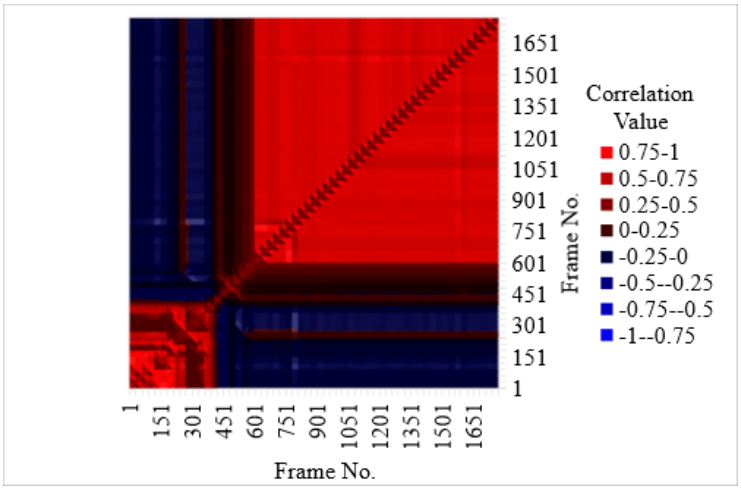

Figure 8: Correlation Matrix for Changing Frame Rates (1 [FPS])

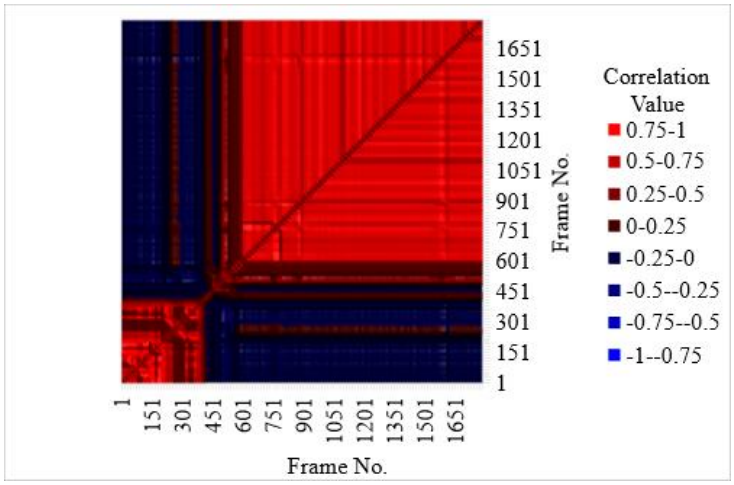

Figure 9: Correlation Matrix for Changing Frame Rates (2 [FPS])

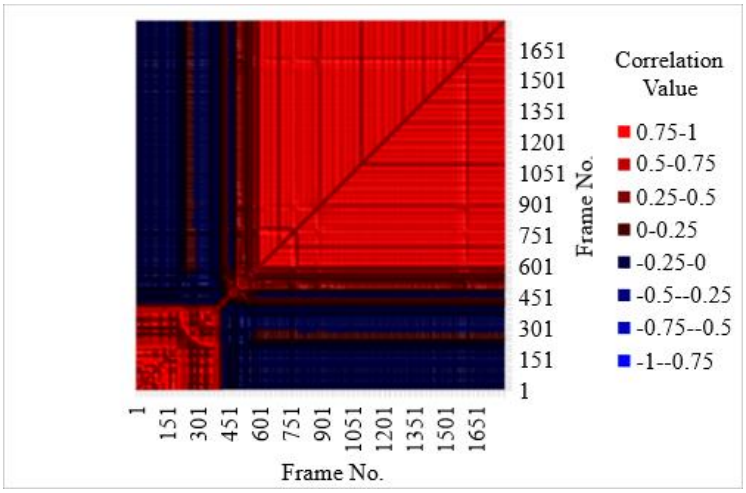

Figure 10: Correlation Matrix for Changing Frame Rates (3 [FPS])

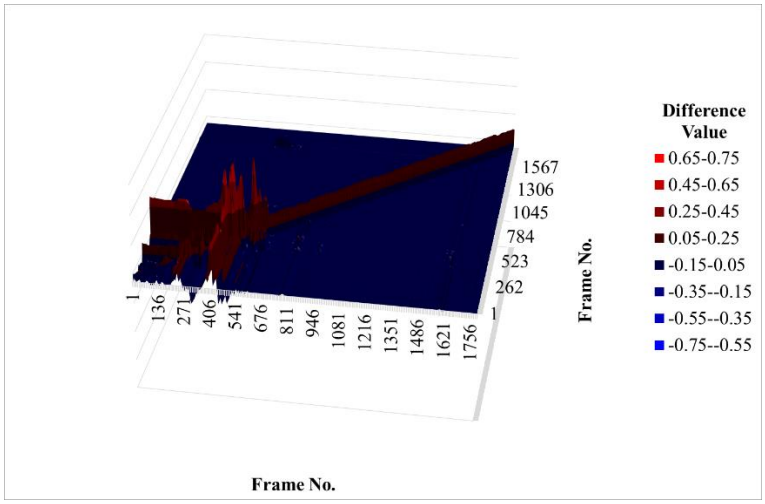

Figure 11: Difference between Adjacent Correlation Values (3[FPS])

4.3 Video Scene Segmentation

The results of video scene segmentation are shown in Figure 12 to 14. In these results, the video scene segmentation uses the aforementioned difference to divide scenes according to the following criteria:

- From the frame where the difference in correlation values exceeds a threshold to the frame where this difference becomes almost 0.
- From the frame where the difference in correlation values becomes almost 0 to the frame where it exceeds a threshold.

The threshold values are set to 0.1, 0.2, and 0.3 in Figure 12, Figure 13, and Figure 14, respectively. From these results, we confirmed that scenes can be segmented from videos capturing learners, based on the continuity of correlation values in the video correlation matrix.

Figure 12: Result of Video Scene Segmentation (Threshold > 0.1)

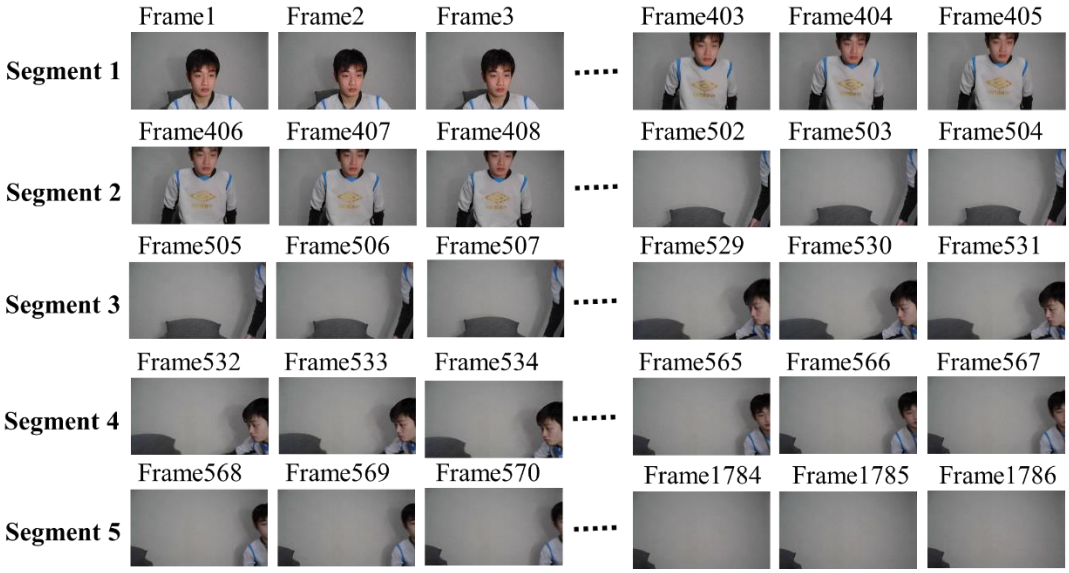

Figure 13: Result of Video Scene Segmentation (Threshold > 0.2)

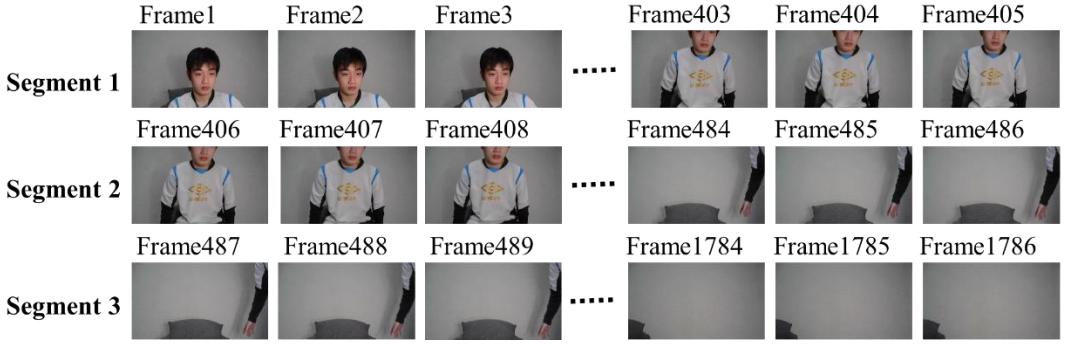

Figure 14: Result of Video Scene Segmentation (Threshold > 0.3)

5 Conclusion

In this paper, we described the development of video scene segmentation based on a video correlation matrix, which reflects learner behavior, aiming to enhance the efficiency of learner monitoring. Where, we calculated the video correlation matrix with different frame rates and frame sizes. Also, we visualized the video correlation matrix and the difference in correlation values between adjacent frames. Additionally, we segmented the video capturing learner into scenes using the difference in correlation values between adjacent frames of the video correlation matrix. From our evaluation, we found as the follows.

- The process time for calculating the video frame correlation matrix varies according to the number of frames and becomes an approximation of a quadratic function.
- The process time for calculating the video frame correlation matrix becomes a constant value regardless of the frame size.
- Even when the frame rate is changed, the video frame correlation matrix has variations only in the number of correlation values, with the values themselves remaining largely consistent.
- The difference in correlation values between adjacent frames becomes high with learner actions and low without learner actions.

- Scenes can be segmented from videos capturing learners based on the difference in correlation values between adjacent frames.

In the future, we would like to evaluate the accuracy of video scene segmentation affected by other computation of correlation values and by different videos capturing other learners.

References

[1] Deniz, D. Z., & Karaca, C. (2004). Pedagogically Enhanced Video-on-Demand Based Learning System. In Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training (pp. 415-420). IEEE.

[2] Graf, S., Kinshuk, & Liu, T.-C. (2008). Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour. In Proceedings of the Eighth IEEE International Conference on Advanced Learning Technologies (pp. 482-486). IEEE.

[3] Kender, J. R., & Yeo, B.-L. (1998). Video scene segmentation via continuous video coherence. In Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231) (pp. 367-373). IEEE.

[4] Maeda, Y., Sugita, K., Oka, T., & Yokota, M. (2008). Proposal of a New Concept of Universal Multimedia Access. In Proceedings of 13th International Symposium on Artificial Life and Robotics (pp. OS7-6). ISAROB.

[5] Mohamed, K. K., Mohamed, J., & Olfa, N. (2008). Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In Proceedings of the Eighth IEEE International Conference on Advanced Learning Technologies (pp. 241-245). IEEE. Retrieved from http://dl.acm.org/citation.cfm?id=2643085&dl=ACM&coll=DL

[6] Rapuano, S., & Zoino, F. (2006). A Learning Management System Including Laboratory Experiments on Measurement Instrumentation. In IEEE Transaction on Instrumentation and Measurement, Vol.55, No.5 (pp. 1757-1776).

[7] Subbian, V. (2013). Role of MOOCs in integrated STEM education: A learning perspective. In Role of MOOCs in integrated STEM education: A learning perspective (pp. 1-4). IEEE.

[8] Sugita, K. (2018). A scene search method using subtitles appended to lecture video for supporting self-learning. In Proceedings of 28th International Symposium on Artificial Life and Robotics (pp. OS7-7). ISAROB.

[9] Sugita, K. (2022). Implementation of an Average Image Composite Software for Viewer Visualizing Behavior During Learning Content. In Advances on P2P, Parallel, Grid, Cloud and Internet Computing. 3PGCIC 2022. Lecture Notes in Networks and Systems, vol 571. (pp. 346-352). Springer.

[10] Sugita, K., Ito, S., Machidori, Y., & Takayama, K. (2019). Some improvements of lecture video content for introducing a beginner programmer to Java programing language. In Proceedings of the 29th International Symposium on Artificial Life and Robotics (pp. 824-827). ISAROB.

[11] Sugita, K., Nakasone, Y., Machidori, Y., & Takayama, K. (2020). Development of a learning time monitoring system for supporting e-Learning. In Proceedings of the 25th International Symposium on Artificial Life and Robotics (pp. 737-740). ISAROB.

[12] Takegawa, C., Sugita, K., & Uchida, N. (2021). Evaluation of Re-watching Support Functions on Educational Content Using Learner Monitoring System. In Evaluation of Re-watching Support Functions on Educational Content Using Learner Monitoring System (pp. 143-144). IPSJ.

Author's Biography

Kaoru Sugita received the B. Eng., M. Eng. and Doctor of Software and Information Science degrees from Saitama Institute of Technology, Toyo University and Iwate Prefectural University in 1996, 1998 and 2003, respectively. From April 1998 to November 2002, he was working for ATA Company. From April 2003 to March 2004, he was a Post Doctor Researcher at Iwate Prefectural University. From September 2003 to March 2004, he was working with NetBridge Company. He joined Fukuoka Institute of Technology in 2004 and works currently as a professor in Department of Information and Communication Engineering, Fukuoka Institute of Technology, Japan. His research interests include multimedia communication systems. He is member of VRSJ, JSKE, IPSJ and IEICE.