Hindawi
Complexity
Volume 2022, Article ID 3053772, 8 pages
https://doi.org/10.1155/2022/3053772

Research Article

A New Approach to Estimate Concentration Levels with Filtered Neural Nets for Online Learning

Woodo Lee,[1] **Junhyoung Oh**,[2] **and Jaekwoun Shim**[3]

[1]*Department of Physics, Korea University, Seoul, Republic of Korea*
[2]*School of Cybersecurity, Korea University, Seoul, Republic of Korea*
[3]*Center for Gifted Education, Korea University, Seoul, Republic of Korea*

Correspondence should be addressed to Junhyoung Oh; ohjun02@korea.ac.kr and Jaekwoun Shim; silent99@korea.ac.kr

Received 25 October 2021; Accepted 8 April 2022; Published 21 April 2022

Academic Editor: Carlos Aguilar-Ibanez

The COVID-19 pandemic heavily influenced human life by constricting human social activity. Following the spread of the pandemic, humans did not have a choice but to change their lifestyles. There has been much change in the field of education, which has led to schools hosting online classes as an alternative to face-to-face classes. However, the concentration level is lowered in the online learning class, and the student's learning rate decreases. We devise a framework for recognizing and estimating students' concentration levels to help lecturers. Previous studies have a limitation in that they classified attention levels using only discrete states. Due to the partial information from discrete states, the concentration levels could not be recognized well. This research aims to estimate more subtle levels as specified states by using a minimum amount of body movement data. The deep neural network is used to continuously recognize the human concentration model, and the concentration levels can be predicted and estimated by the Kalman filter. Using our framework, we successfully extracted the concentration levels, which can aid lecturers and can be expanded to other areas. To implement the framework, we recruited participants to take online classes. Data were collected and preprocessed using pose points, and an accuracy of 90.62 % was calculated by predicting the concentration level using the framework. Furthermore, the concentration level was approximated based on the Kalman filter. We found that webcams can be used to quantitatively measure student concentration when conducting online classes. Our framework is a great help for instructors to measure concentration levels, which can increase the learning efficiency. As a future work of this study, if emotion data and skin thermal data are comprehensively considered, a student's concentration level can be measured more precisely.

1. Introduction

After the outbreak of the coronavirus in December 2019, it has spread worldwide and has caused much confusion in society [1]. The coronavirus is causing chaos in many parts of society and has a great impact on the daily life of mankind. Education is one of the most affected sectors as the coronavirus has persisted without any signs of improvement [2]. Most classes in elementary school, middle school, high school, and university have come to be conducted in the online learning method. In many schools that do not have sufficient preparation for online learning, the educational effectiveness is declining due to insufficient technical preparation and lack of operational experience [3].

Online learning is classified into synchronous distance education and unsynchronous distance education [4]. In synchronous distance education, lectures are conducted in real-time using useful tools such as Zoom or Google Meet [5]. In unsynchronous distance education, instructors upload the recorded video to the system, and students take the course at the desired time. Because synchronous distance education is a real-time lecture, if students turn on their cameras and show their faces, it is possible to determine the minimum level of participation in the class. For unsynchronous distance education, many universities develop and use various learning management systems such as Moodle and Blackboard [6]. By using these systems, it is possible to determine student participation by calculating the learning rate roughly.

Although the students participated in the class, they might not be focused on the content of the class. Xu and Yang found that when students learn through online learning, the dropout rate can go up to 95% because they desire to use their time for purposes other than educational purposes [7]. Even if students participate in synchronous distance education, the instructor cannot correctly determine the students' concentration due to actions such as taking other actions or turning the camera while attending class. Even if the learning rate is calculated using multiple learning systems, unsynchronous distance education has many weaknesses. Students turn on the class and engage in other activities, or they attack the system's vulnerabilities to adjust the speed of lectures and take them faster [8].

Therefore, appropriate measures should be taken by determining the students' concentration in class. Typically, lecturers have determined students' concentration levels based on their own experiences in online learning. For example, they would make inferences about whether students were concentrating on a lecture or not through various visual cues, such as the focus of students' eyes or their body movements during interactions. However, according to a study by Erol and Tekdal, when it comes to distance education, teachers currently do not have sufficient resources to supervise and evaluate students [9]. Therefore, an automated method for determining students' concentration levels is needed.

There have been many attempts to measure students' concentration levels using various methods, such as taking skin temperature [10], recognizing visual attention and students' emotions [11], and detecting electroencephalogram (EEG) signals [12–14]. However, these methods often do not work well in online classes because teachers cannot promptly interact with each student. In addition, these attempts lack detail because their concentration levels are classified as discrete states [15]. As students' concentration levels are simultaneously changing states, this information may aid lecturers.

Here, we develop a new framework that consists of a concentration level recognition network (CLRN) and Kalman filter (KF) [16] to overcome the limitations of existing methods. The CLRN is based on supervised learning, which is trained with the standard deviations of designated points in positioning a human being and classified labels. The CLRN provides the concentration levels as the probability of "high concentration." The concentration levels can be obtained by the CLRN simultaneously, and the KF identifies the patterns from the fluctuating levels. In addition, a future concentration level can be estimated by applying the KF. Ultimately, the concentration levels can be quantified by the CLRN with KF, which can aid the lecturers in better understanding the concentration levels of his/her students.

We implemented this framework in practice using videos of participants taking online lectures. First, the standard deviations of the pose points were extracted as a preprocessing step. Then, CLRN was constructed, and a loss function was grafted. Based on this, the concentration level of the participants was predicted, and performance of 90.62 % was derived. Moreover, the concentration level was completed by smoothing and approximating by applying KF to the result. In this paper, there are various abbreviations, and the list of abbreviations is summarized in Table 1.

2. Motivation and Related Work

The motive of our study is that a person's body movements can be a factor in recognizing his/her condition. Extracting the status of a human being, such as their emotional state, from body movements is an interesting research topic, which has recently become more important [17]. Several studies have extracted meaningful factors from the movements of individuals. They have examined whether students are concentrating on a lecture or not by checking various visual cues, such as the focus of the eyes or body movements of the students [18]. Generally, eye movement is a strong indicator for estimating the degree of concentration [19]. Research has demonstrated that concentration is amplified when the eye movements of participants maintain a central fixation [20]. In addition, body movement has been previously researched; for example, a model using the joints of the human body can estimate the pose of individuals [21]. Kinetic movement of an individual's body has been identified for the assessment of a patient's recovery process [22]. There have been previous studies that extract high-value features based on dynamic movements such as dance movement and aerobic [23, 24]. The pose of individuals has also been researched using video data, which can then be used to present a visual flow of poses [25]. Furthermore, emotion has been recognized from body movement via machine learning and is available in public data sets [26].

The studies mentioned above suggest that the relationship between the movement of individuals and effect is very close, and the relationship should be examined via a bidirectional rather than a unidirectional cause-effect approach [27]. Research to classify the various states of individuals by human body posture has been conducted; however, only binary states have been suggested as results [15]. Similar to previous research, we propose that the standard deviations of designated physical points comprise a core factor in measuring concentration levels. We use OpenPose as a backbone package; it is a well-known tool for analyzing body movements by detecting designated points of a human body. Several types of research have been used OpenPose; for example, sign languages were recognized by a transfer learning algorithm that utilized OpenPose [28]. When humans are focused on some subjects, the standard deviations of their movements will become lower because they engage in less wasted effort. Therefore, we designed the CLRN based on deep learning to find the subtle changes in the standard deviations. There has been similar research to recognize human states, such as emotion [29], via deep learning. However, the approach is limited in the sense that it does not provide simultaneous results. To address this problem, we apply a KF to deal with continuous and simultaneous data.

Table 1: Abbreviation list.

Abbreviation	Definition
EEG	Electroencephalogram
CLRN	Concentration level recognition network
KF	Kalman filter
ReLU	Rectified linear unit
ADAM	Adaptive moment
DNN	Deep neural network

3. Proposed Framework

Figure 1 shows the overview of our framework. In the first step of the framework, the student's video data recorded by a webcam is preprocessed. The preprocessed data are labeled by the two states based on the participants' self-reported intent. The labeled data are used for training the CLRN in the recognition step. The CLRN is devised with supervised learning for binary classification, and the data are prepared with a binary class (the data are labeled as zero or one).

The trained CLRN recognizes the continuous concentration levels, which are defined as recognition levels (S_r). The KF is used for smoothing and filtering the highly fluctuating S_r. In the estimation step, the KF provides an approximation of the concentration levels, which are called the estimation levels (S_e).

Our method to recognize and estimate human concentration levels consists of three steps: preprocessing, recognition, and estimation.

3.1. Step 1: Preprocessing. The first step of our framework is to extract the standard deviations of the pose points from the video data. The standard deviations (σ) of the X and Y coordinates are calculated for the top and middle parts, respectively. Note that we assume the standard deviations of the points are the core factor in measuring the concentration levels. Table 2 shows the notations of the results in the preprocessing step.

Algorithm 1 shows the process of the preprocessing step. The standard deviations are obtained through this algorithm and become the input data, the CLRN, which is discussed in the following section.

3.2. Step 2: Recognition. Algorithm shows the overall structure of the recognition step. Through the CLRN, the recognition levels (S_r) are obtained. The CLRN consists of four layers: two hidden, one input, and one output layer. The role of the hidden layers is to find the hidden features in the data. A network deeper than two layers does not improve the performance of the framework. The rectified linear unit (ReLU) is used as the activation function in both hidden layers. A sigmoid is used to make sure the probability is distributed relatively evenly from zero to one. ADAptive Moment (ADAM) estimation optimizer [30] is applied, and the initial learning rate is set as 0.1 %, which is the optimal value for the CLRN. The binary cross-entropy loss is chosen as the loss function (L) of the CLRN and is defined as

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right) log\left(1-\hat{y}_i\right)\right], \tag{1}$$

where the number of data items is N, the labels are y_i, and the prediction values from our deep neural network (DNN) are $\hat{y}_i$. Note that the values of y_i are obtained from the participants' self-reported intent. As the output value is a probability for binary classification, the binary cross-entropy is an appropriate value to determine continuous concentration levels.

3.3. Step 3: Estimation. The estimation step of the CLRN includes a KF to establish S_e. Algorithm 3 shows the overall process. There are three states in the algorithm: the prediction state, $\mathbf{s_p}\left(\mathbf{t}\right)$; estimation state, $\mathbf{s_e}\left(\mathbf{t}\right)$; and measurement state, $\mathbf{m_t}$. The error covariance matrix $(\mathbf{P_t})$ and the transition weight matrix $(\mathbf{A})$ are also defined. In the predicting step, $\mathbf{A}$ and an external noise matrix $(\mathbf{Q})$ are used, and lectures can modify those matrices. $\mathbf{A}$ is set to 1, and $\mathbf{Q}$ is set to 0 as an ideal case.

As part of the updating step, the Kalman gain $(\mathbf{K})$ is obtained at each update. $\mathbf{H}$ is a scale matrix, which is set to 1 by simplifying the problems. $\mathbf{s_e}\left(\mathbf{t}+1\right)$ and $\mathbf{P_{t+1}}$ are updated with $\mathbf{K}$. Finally, in the estimating step, the next estimated state $\mathbf{s_e}\left(\mathbf{t}+1\right)$ is recurrently updated.

We assume that each of the distributions of Ψ_{low} can be decomposed into two dominant levels with a certain function. The function is a bimodal distribution X, which is written as

$$\begin{aligned} N_k\left(\mu_k, \sigma_k\right) &= Ae^{-\left(x-\mu_k\right)^2/2\sigma_k^2}, \\ X &= \mathcal{N}_1\left(\mu_1, \sigma_1^2, A_1\right) + \mathcal{N}_2\left(\mu_2, \sigma_2^2, A_2\right), \end{aligned} \tag{2}$$

where σ_1 and σ_2 are the standard deviations, μ_1 and μ_2 are the mean values, and x is the input data.

4. Implementation

Three participants were recruited for this experiment, and each participant was recorded when they viewed an online lecture, and they were required to mark the times when they were concentrating on the lecture. This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Korea University Center for Gifted Education.

A webcam was used to record the 25-fps video data. For recording video data for the distraction (nonconcentration) case, the participants also marked the times when they were distracted.

The data from three participants are merged as a dataset because estimating the levels for each participant, respectively, could be biased per the characteristics of participants. Moreover, we expect to find general properties of concentration levels by using the merged data with our models. The merged dataset is labeled as two cases based on the markers of the participants. In total, 12 hours of video data

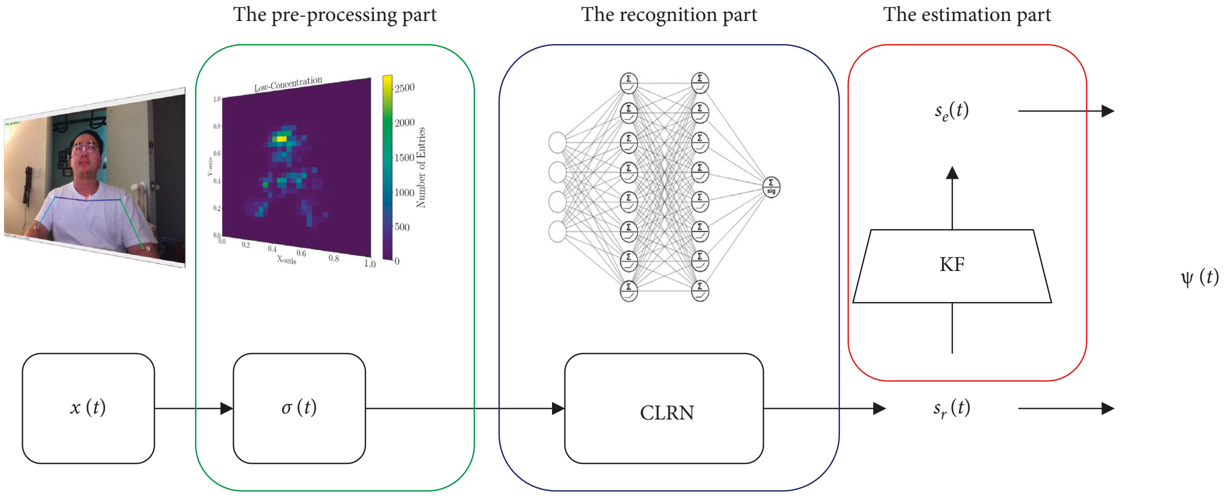

FIGURE 1: Overview of our framework. The framework consists of a CLRN to recognize the features and a KF to estimate the levels.

TABLE 2: Data symbols and descriptions.

Symbol	Description
σ^X_{Top}	σ of the top part's X coordinate
σ^Y_{Top}	σ of the top part's Y coordinate
σ^X_{Mid}	σ of the middle part's X coordinate
σ^Y_{Mid}	σ of the middle part's Y coordinate

```
Input: top.X, top.Y, mid.X, mid.Y
  for each Data ∈ {top.X, top.Y, mid.X, mid.Y} do
    for $D_i = \bigcup_{50i}^{50(i+1)} \text{Data}$ do
      $\sigma_{\text{Data}}$.append$(s.d.(D_i))$
    end for
  end for
Output: $\sigma^X_{\text{Top}}, \sigma^Y_{\text{Top}}, \sigma^X_{\text{Mid}}, \sigma^Y_{\text{Mid}}$
```

ALGORITHM 1: Data preprocessing.

```
Input: $\sigma^X_{\text{Top}}, \sigma^Y_{\text{Top}}, \sigma^X_{\text{Mid}}, \sigma^Y_{\text{Mid}}$
Input layer: $\in R^4$
  $1^{st}$ hidden layer: $\in R^8$ (activation function: Re$LU$)
  $2^{n\,d}$ hidden layer: $\in R^8$ (activation function: Re$LU$)
  Output layer: $\in R^1$ (activation function : Sigmoid)
Output: ConcentrationLevels $s_r(t) \in$ S
```

ALGORITHM 2: CLRN.

```
Input: $s_r(t) \in S$
for all $s_r(t) \in S$ do
  $\mathbf{s_e(t)} \leftarrow s_r$
  /* Predicting */
  $\mathbf{s_p(t) = A \cdot s_e(t)}$
  $\mathbf{P_t^{pre} = A \cdot P_t \cdot A^T + Q}$
  /* Updating */
  $\mathbf{K = P_t^{pre} \cdot H/(H \cdot P_t^{pre} \cdot H + R)}$
  /* Estimating */
  $\mathbf{s_e(t+1) = s_p(t) + K \cdot (m_t - H \cdot s_p(t))}$
  $\mathbf{P_{t+1} = P_t^{pre\ d^p} - K \cdot H \cdot P_t^{pre\ d}}$
end for
  /* Analyzing */
  Fitting the distribution of $\mathbf{s_e(t+1)}$
  user − defined function
Output: ConcentrationLevels $(\Psi(t))$
```

ALGORITHM 3: KF.

(consisting of 1M images) were recorded. A total of eight hours of data was marked for the concentration case; the other four hours of data were taken for the distraction case.

For step 1(preprocessing), certain pose points in the images are detected to measure the distribution of participant poses. Ten points of the human body are measured every 50 frames, which are classified as the top part (0–4) and the middle part (5–9). Figure 2 shows the points, and the coordinate data of the points range from zero to one. To detect the points, OpenPose [31–34] is used. OpenPose [31] is a very recent open-source package for detecting the keypoint of human poses. OpenPose is a real-time system for the body, foot, hand, and facial keypoint detection and is an

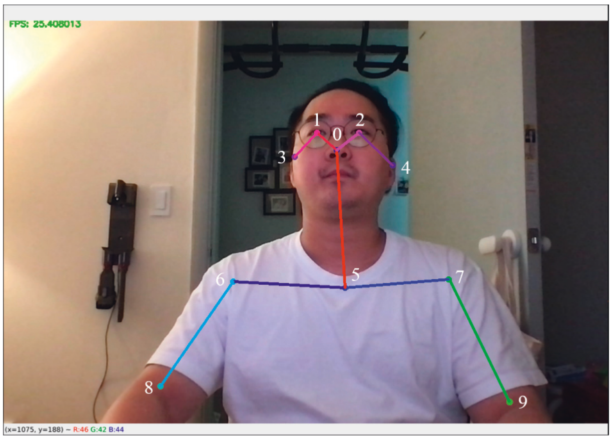

FIGURE 2: The points measured by OpenPose are shown. The middle and upper body are measured with ten points, respectively.

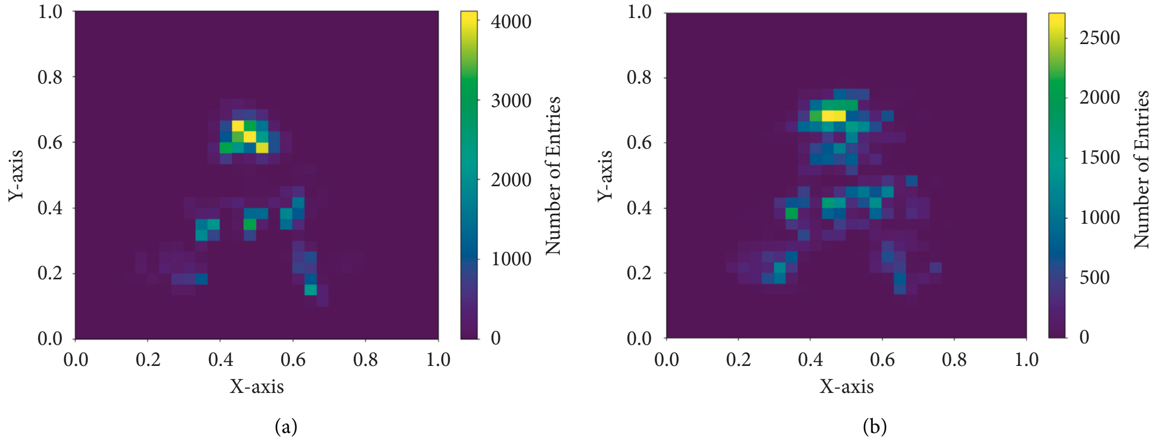

Figure 3: (a) The 2D histogram in the case of high-concentration; (b) the 2D histogram in the case of low-concentration.

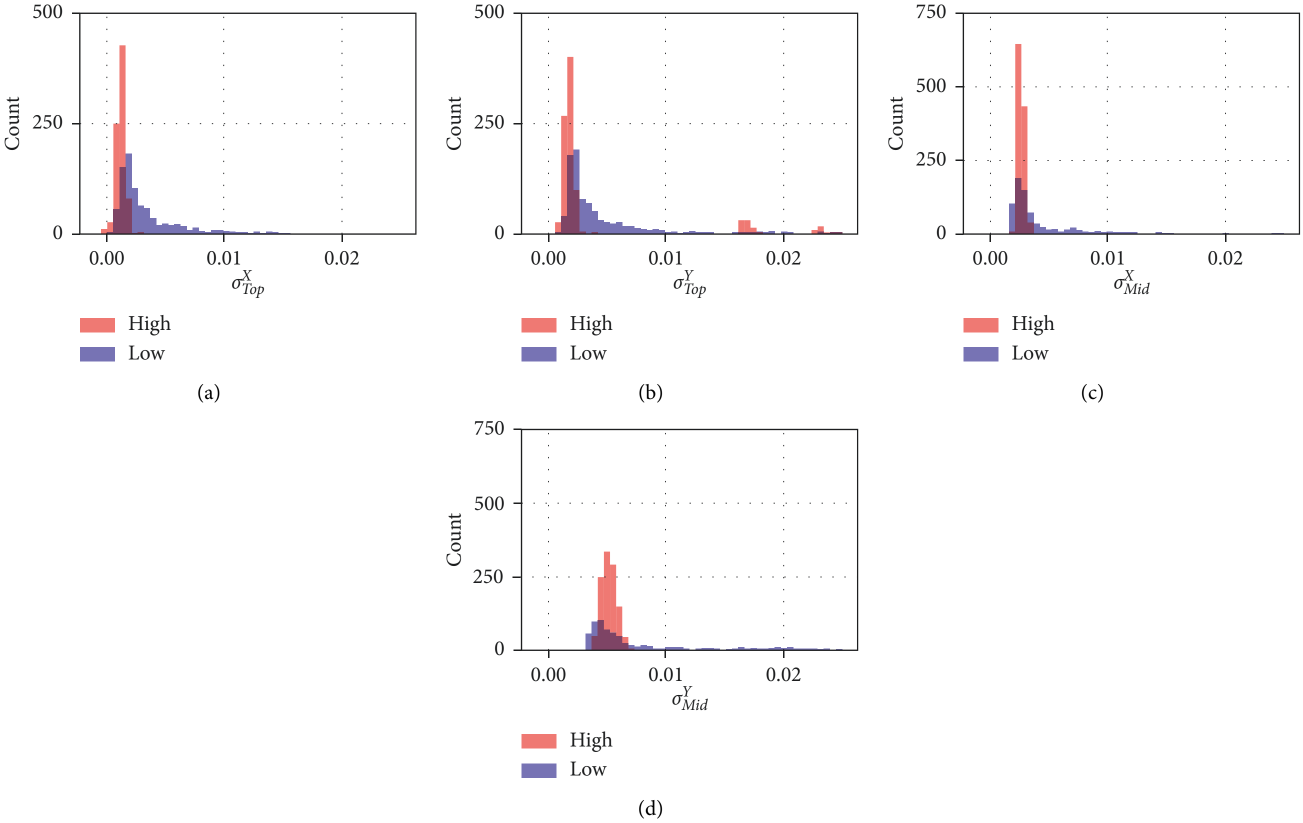

Figure 4: (a)–(d) show the standard distribution for each respective part. The red histogram indicates the high concentration case, and the blue histogram indicates the low concentration case.

appropriate package for continuously detecting these points. In our case, we only used the upper body of individuals as captured in the video data.

We then check the distributions of the pose points when the individuals were concentrating or not, as shown in Figure 3. The distribution in Figure 3(a) shows that the entries are gathered more closely around the body points, while those in Figure 3(b) are spread more widely. The difference is visually noticeable in this example, but it cannot be easily quantified to identify the concentration levels. In the preprocessing step, the input data are already divided into 50 frames, so the input data for the CLRN are not separated into minibatches.

For step 2(recognition), CLRN performs the task of predicting what participants marked while viewing the online lecture. K-fold is applied to cover insufficient data. The accuracy of 5-fold training ranged from 85% to 95% with a median of 90.62%.

Figure 4 shows the difference of the σs among each group. Nevertheless, there remain unexplained aspects, such

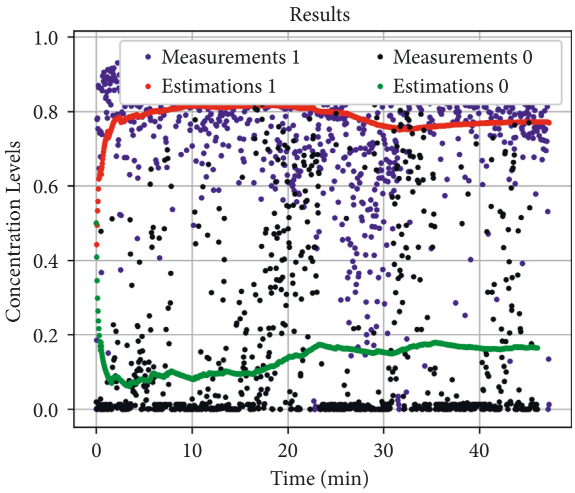

Figure 5: Estimation and measurement levels are shown. The measurement and estimation values are represented with the blue and red dots, respectively, when the students are concentrating highly. The black and green dots are the measurement and estimation values, respectively, when the students express low concentration levels.

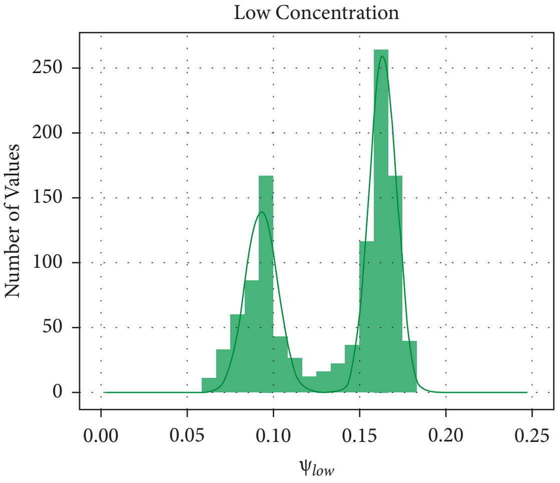

Figure 6: Ψ_{low} and the fit curve. The histogram shows the estimated concentration levels from our model. The histogram contains the data of all three participants when they are under low concentration levels.

as ambiguous patterns, whose correlation with the concentration levels is unclear. To this end, neural networks are applied to solve the problems as they are an appropriate method for obtaining nonlinear combinations from features. This allows us to identify hidden features that we cannot otherwise describe.

For step 3 (estimation), trained CLRN recognizes continuous concentration level, smoothing and filtering it using Kalman Filter, and finally approximates it. The students' state starts from $\mathbf{s}_{\mathbf{e}}(0) = 0.5$ because the students' concentration level is assumed to be 50 % at the beginning. $\mathbf{P}_0 = 0.9$ is the system error, which comes from the DNN, which was described in Section 3.

Figure 5 shows the estimation and measurement results for 2.5 second intervals. It indicates that the students maintained their concentration levels, and there were no external disturbances when they observed the lectures. Even though the measurements fluctuate widely every 2.5 seconds, the KF enables users to track the levels smoothly, which are shown as the green and the red dots, indicating the low (Ψ_{low}) and high (Ψ_{high}) concentration levels, respectively.

Figure 6 shows the distribution of Ψ_{low}. μ_1 and μ_2 are obtained as 0.09 and 0.16, respectively, which indicate that the students are entangled in two concentration levels.

5. Conclusion and Future Work

Many schools have been semicompulsory for distance education due to the coronavirus. However, distance education is economical in terms of price effect and can educate many students simultaneously. Furthermore, if distance education is carried out, cooperative learning can be performed in an interactive learning environment, and since home-based classes are possible, the time and effort of commuting to school are reduced. If the major disadvantage of distance education, the concentration level is low, can be overcome through this study, and more effective classes will be possible. We solve this problem by developing a novel framework consisting of a concentration level recognition network and a Kalman filter. We devised the model for aiding lecturers in estimating students' concentration levels using webcams as part of online classes. Our system presents the level every 2.5 seconds with 90.62% accuracy and estimates the next level of concentration by using the KF. In contrast to the previous research, such as VGG16 [35], our model takes a different approach to quantify the levels by capturing the variance of the detected pose points on individuals in the current state. Additionally, we estimate and track the level for the next time window. Our model offers a practical tool to monitor the level more precisely and aid lecturers in estimating the level. Academically, our model applies a novel approach to analyzing complex human states, in this specific case, concentration level. As future work, we plan to use not just body movement data but also emotion data [36] and skin thermal data [10, 37] to enhance the prediction of measuring human concentration levels. This paper will combine and process the measuring method used and the conventional techniques using deep learning. This work expects to provide helpful information on students' concentration levels and thus assist lecturers.

Data Availability

No data are available because of privacy issue.

Disclosure

An earlier version of this manuscript was preprinted in the arXiv [38], and several students participated to the earlier version.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to show their gratitude to Jakyung Koo, Nokyung Park, and Pilgu Kang at Korea University for their assistance in this research, which greatly improved the manuscript. Even though they are not included in authorship in this paper, their broader assistance lays the foundation of our research.

References

[1] W. McKibbin and R. Fernando, *The economic impact of covid-19. economics in the time of covid-19, Baldwin, B. Weder di Mauro (red.)*, pp. 45–51, Centre for Economic Policy Research (CEPR), London, UK, 2020.

[2] J. Daniel, "Education and the covid-19 pandemic," *Prospects*, vol. 49, no. 1, pp. 91–96, 2020.

[3] A. Pragholapati, *Covid-19 Impact on Students*, Universitas Pendidikan Indonesia, Bandung, Indonesia, 2020.

[4] R. F. Branon and C. Essex, "Synchronous and asynchronous communication tools in distance education," *TechTrends*, vol. 45, no. 1, p. 36, 2001.

[5] V. D. Soni, "Global impact of e-learning during covid 19," *SSRN 3630073*, 2020.

[6] P. Faisal and Z. Kisman, "Information and communication technology utilization effectiveness in distance education systems," *International Journal of Engineering Business Management*, vol. 12, 2020.

[7] B. Xu and D. Yang, "Motivation classification and grade prediction for moocs learners," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 2174613, 7 pages, 2016.

[8] J. A. Alokluk, "The effectiveness of blackboard system, uses and limitations in information management," *Intelligent Information Management*, vol. 10, no. 06, pp. 133–149, 2018.

[9] E. Koçoglu and D. Tekdal, "Analysis of distance education activities conducted during covid-19 pandemic," *Educational Research and Reviews*, vol. 15, no. 9, pp. 536–543, 2020.

[10] S. Nomura, M. Hasegawa-Ohira, Y. Kurosawa, Y. Hanasaka, K. Yajima, and Y. Fukumura, "Skin tempereture as a possible indicator of studentâ€™ s involvement in e-learning sessions," *International Journal of Electronic Commerce Studies*, vol. 3, no. 1, pp. 101–110, 2012.

[11] P. Sharma, M. Esengönül, S. R. Khanal, T. T. Khanal, V. Filipe, and M. J. Reis, "Student concentration evaluation index in an e-learning context using facial emotion analysis," in *Proceedings of the International Conference on Technology and Innovation in Learning, Teaching and Education*, pp. 529–538, Springer, Thessaloniki, Greece, June 2018.

[12] A. S. Al-Musawi, "Concentration level monitoring in education and healthcare," *Basic and Clinical Pharmacology and Toxicology*, vol. 124, no. s2, p. 36, 2018.

[13] S. Marouane, S. Najlaa, T. Abderrahim, and E. K. Eddine, "Towards measuring learner's concentration in e-learning systems," *International Journal of Computer Techniques*, vol. 2, no. 5, pp. 27–29, 2015.

[14] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using eeg signals from mobile sensors," *Sensors*, vol. 13, no. 8, pp. 10273–10286, 2013.

[15] R. Sacchetti, T. Teixeira, B. Barbosa, A. Neves, S. C. Soares, and I. D. Dimas, "Human body posture detection in context: the case of teaching and learning environments," *SIGNAL*, vol. 87, pp. 79–84, 2018.

[16] R. E. Kalman, "A new approach to linear filtering and prediction problems," *IEEE*, vol. 82, no. 1, pp. 35–45, 1960.

[17] H. Zacharatos, C. Gatzoulis, and Y. L. Chrysanthou, "Automatic emotion recognition based on body movement analysis: a survey," *IEEE computer graphics and applications*, vol. 34, no. 6, pp. 35–45, 2014.

[18] L. Lakshmi Priya Gg, "Student emotion recognition system (sers) for e-learning improvement based on learner concentration metric," *Procedia Computer Science*, vol. 85, pp. 767–776, 2016.

[19] M. Li, L. Cao, Q. Zhai et al., "Method of depression classification based on behavioral and physiological signals of eye movement," *Complexity*, vol. 2020, Article ID 4174857, 9 pages, 2020.

[20] M. M. Doran, J. E. Hoffman, and B. J. Scholl, "The role of eye fixations in concentration and amplification effects during multiple object tracking," *Visual Cognition*, vol. 17, no. 4, pp. 574–597, 2009.

[21] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *Advances in Neural Information Processing Systems*, vol. 27, pp. 1799–1807, 2014.

[22] L. M. Pedro and G. A. de Paula Caurin, "Kinect evaluation for human body movement analysis," in *Proceedings of the 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1856–1861, IEEE, Rome, Italy, June 2012.

[23] X. Zhai, "Dance movement recognition based on feature expression and attribute mining," *Complexity*, vol. 2021, Article ID 9935900, 12 pages, 2021.

[24] W. Fan and H. J. Min, "Accurate recognition and simulation of 3d visual image of aerobics movement," *Complexity*, vol. 2020, Article ID 8889008, 11 pages, 2020.

[25] T. Pfister, J. Charles, and A. Zisserman, "Flowing convnets for human pose estimation in videos," in *Proceedings of the IEEE international conference on computer vision*, pp. 1913–1921, Araucano Park, December 2015.

[26] F. Ahmed, A. H. Bari, and M. L. Gavrilova, "Emotion recognition from body movement," *IEEE Access*, vol. 8, pp. 11761–11781, 2019.

[27] I. Rossberg-Gempton and G. D. Poole, "The relationship between body movement and affect: from historical and current perspectives," *The Arts in Psychotherapy*, vol. 19, 1992.

[28] S.-K. Ko, C. J. Kim, H. Jung, and C. Cho, "Neural sign language translation based on human keypoint estimation," *Applied Sciences*, vol. 9, no. 13, p. 2683, 2019.

[29] R. Santhoshkumar and M. K. Geetha, "Deep learning approach for emotion recognition from human body movements with feedforward deep convolution neural networks," *Procedia Computer Science*, vol. 152, pp. 158–165, 2019.

[30] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014.

[31] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

[32] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1145–1153, Honolulu, HI, USA, July 2017.

[33] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, June 2017.

[34] M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi, and M. M. AlyanNezhadi, "Facial emotion recognition using deep

convolutional networks," in *Proceedings of the 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)*, p. 0017–0021, Tehran, Iran, December 2017.

[35] J. C Ruvinga and D. Malathi, "Human concentration level recognition based on vgg16 cnn architecture," *International Journal of Advanced Science and Technology*, vol. 29, 2020.

[36] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 4724–4732, Las Vegas, USA, June 2016.

[37] L. Nummenmaa, E. Glerean, R. Hari, and J. K. Hietanen, "Bodily maps of emotions," *Proceedings of the National Academy of Sciences*, vol. 111, no. 2, pp. 646–651, 2014.

[38] W. Lee, J. Koo, N. Park, P. Kang, and J. Shim, "A framework for recognizing and estimating human concentration levels,".