pre-submission

by Arief Setyanto

7
Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

CHEATING ACTIVITY DETECTION ON MOBILE SECURE ONLINE EXAM

22

Abstract

The online exam is one of the important tasks in online learning systems. Online exam proctoring, therefore, is important to ensure the credibility of the exam. The participants who take apart in the online exam are distributed in many locations. This work proposes an online exam system with an android application. We utilize the camera and audio recorder to capture human activity during the test. We collect the audio-video from 20 participants in seven cheating and one non-cheating scenario. To get the balance data between cheating -non-cheating activity, the participant, take seven cheating and seven non-cheating scenarios. The system record the face of a participant using the front camera and recording the sound. We propose a mid-level representation of the audio-video before the classification task. We perform data normalization into uniform units for each parameter between zero to ten. Finally, the MLP carries out the classification of a midlevel signal into cheating and no- cheating activity. We found that the accuracy of activity classification achieves 90.7%, while precision at 90.7% recalls 91 % and F1 score at 90.6%.

Keywords : Action Detection, Mobile Exam, Eye Detection, Face Detection, Multilayer Perceptron

1. Introduction

Electronics learning currently has become one of the games changing applications in the education sector. The traditional education in classical mode has gradually changed into online mode either in synchronous and asynchronous learning. The online meeting has been enabled by online video meeting applications, while an asynchronous model facilitated by online documents ranging from a static powerpoint, animation software, sharing a video to more complex interactive learning tools such as interactive programming tools. According to a survey in [1], 14.3% of American students taking online courses only without a traditional degree, while 15 % of students take a combination online and traditional degree. A significant improvement happened to compare to the same survey in 2013 [2]. The online education participants improved from 411.000 to 7.1 million, while the online education provider improved from 2.6% to 5% over the past year. Although the survey has been done in the USA, it indicates the general trend all over the world.

Examination plays an important role in ensuring the participants have understood the learning material. The examination can be either synchronous and asynchronous. Synchronous exam where the examiner and examinee work at the same time, it can be done by video call exam. The asynchronous online exam can be done by an interactive questionnaire done by the participants. Asynchronous mode permit examinee examines at any time and any place. Online proctoring would not be possible in this mode; therefore, if we need to ensure the test is properly done, an automatic proctoring is needed.

To date, several researchers have reported that online exam opens the higher possibility of cheating [3]–[5]. According to [4] and [5], 74% of students considered that cheating in the online exam is easy. While 29% of the respondents in their research committed cheating during an online exam. According to [6], honor code only gives the insignificant effect of reduction in online cheating, while warning double the reduction of cheating behavior.

Online proctoring is desirable due to the growth of massive open online courses (MOOC). Some online proctoring products available in the market such as, e.g., Kryterion, Web assessor and ProctorU, offer online proctoring services [7]. They usually put a surveillance camera on the user site; however, they still rely on human watching over through video streaming. [8], they propose a mobile exam system embedded in the Moodle system and identify various vulnerabilities that may violate exam security in m-learning environments and to design the appropriate security services and countermeasures that can be put in place to ensure exam security.

Automatics proctoring task has received attention from the researcher community. In [9], front face camera to track the participant eye movement. The exploitation of eye movement leads to an understanding of user focus. When the user focus leaves the testing sheet, it can be assumed that something happens. The multimedia analysis helps online proctoring automation [10], [11]. Gaze tracker, webcam and EEG installed to help online exam proctoring [10]. In [11], they set two cameras, which are one webcam, and a camera installed on special designed eyeglass (webcam). They record the user's face, capturing the participant's view area. Their design is working properly; however, they need a special installed camera to reduce the practicality.

Audiovisual has been exploited for activity/event recognition for other purposes such as social interaction, object tracking and sports analysis. Social cohesion among group members extracted from video and classified by support vector machine [12]. An out of the box way of identifying good behavior of job seeker utilized interviews video proposed by [13]. Identifying unwanted behavior in audiovisual data has also done by [14], [15]. Sport activity recognition such as attack defense, match report proposed by [16], while goals, saves, kick of recognized using Bayesian network [17].

This research makes an effort to avoid installing the special device to perform the online proctoring. We, therefore, design the exam on a smartphone with a requirement of double camera front and back, GPS, gyro sensors and voice recorder. Currently, those accessories are available on most of the smartphones. We will use those standard accessories to provide multimodal data captured during the exam. The camera, the voice recorder is responsible for user verification and test surveillance.

2. Related work

There are three possibilities regarding online exam model proctoring. Firstly, the exam leaves without human controlling [18], [19] [20]. Secondly, the human involved either online through a webcam or physically in a certain location [21]. The third option leaves the participant take the test without proctoring but records the exam process and let the multimedia analysis system work[11], [22].

In the Exam, we need to ensure user identity. There are many efforts in ensuring computer user identity, for example, utilizes keystroke behavior or biometrics. Keystroke is a major activity in using a computer; therefore, it attracts researchers' attention to identifying user activity. In [23], researchers use keystroke dynamics for online exam authentication. In [24], iris recognition for biometrics authentication for the same purpose. In [25], face recognition periodically has done during the exam in order to ensure the authentication of the examinee.

Beside participant identity, user action during an exam is an important clue of cheating behavior. There is much research dealing with action recognition such as [26]–[28]. Head movement is one important clue for activity recognition. [26] detect human activity by extracting head motion to analyze the socio-communicative and affective behavioral characteristics of interacting partners. There is much technique for head pose estimation. [27] using the multilevel structured hybrid forest to detect head and estimating the pose. Dornaika et. proposed 3D face pose estimation using 2D eigenfaces and DE algorithm [28]. Mutual information (MI) exploited to deal with pose estimation in an uncontrollable environment [29]. Another technique proposed by [30] exploited a mixture of linear regression with partially-latent output.

Eye-tracking plays an important role in understanding human attention. In the examination proctoring, exploitation of eye direction will be able to point out the user's point of view. Currently, most eye movement technology developed using near-infrared light. The human pupil will light up when they receive near-infrared light sources, and therefore, the currently available technology benefit from that pupil behavior. It is, however, costly and lacks practicability due to a need for the special device and even location for eye movement detection. In many medical purposes, this technology is viable due to the accuracy requirement and facilities availability.

However, this research would not be able to provide a special camera with near-infrared sources; therefore, we seek simple techniques to detect and track the pupil with a simple web camera. Some researchers made an effort to enable tracking the eye without near-infrared sources such as [31], [32] Cheung et. All proposed a webcam in desktop to recognize the pupil, eye corner and movements [31]. In 2016, [32] propose a normal camera for eye movement recognition. Meng et. All proposed CNN based eye movement detection and tested in a half-million frames and successfully recognize activities such as reading, browsing and watching video [33].

Sound has been identified as a strong clue for activity recognition in [11] sound considered to be a clue of cheating if a human voice exists. Therefore, speech extraction performed in modeling the cheating action. Sound event recognition has been attracted researcher attention, such as [34][35].

The contribution of this paper is:

Developing a mobile exam video cheating dataset consists of cheating and one non-cheating activity from twenty subjects, 280 actions with 36.120 frames in total.

Defining a midlevel data modeling for audio-video cheating dataset consists of head pose, eye tracking and audio.

Develop an artificial neural network to perform cheating activity detection based on the mid-level dataset.

The outline of the remaining paper. We explain our proposed methods in section 3, and part 4 is described result and discuss the achievement, in the last part we draw our conclusion

3. Multimodal Action Detection

3.1 Research Framework

This research is divided into four main tasks which are, defining the cheating actions, develops a dataset for pre-defined actions, mid-level feature extraction, and action recognition task. Figure 1 depicts the entire research framework.

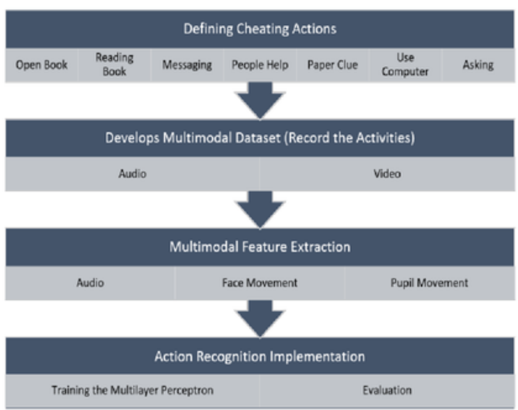


Figure 1 Research Framework

3.2 Multimodal Input

Multimodal cheating behavior in online exams involves audio and video data. We pre-process the video from the front camera of the smartphone. Before further processing, a standardization of the visual signal is carried out to reduce the volume of the data while maintaining the quality for feature extraction purposes. The size of the frames is under-sampled into 180 x 320 pixels size and reduces the frame rate into 12 per second. We consider that this data is enough considering that the video is quite low motion and the distance between the object under surveillance is closed enough. The audio is taken continuedly during the data collection for further processing. Figure 2 shows the condition of the normal position of a smartphone, a front camera and the face of the test taker.

Figure 2 Normal position of a mobile exam test taker

3.2.1. Face and Eye Recognition

The region of interest in the data transformation to the mid-level data representation is the face. Therefore, we need a number of task to localized the face and the eyes. Firstly, resize the frames into 180 x 320 pixels per frame with 12 frames per second (fps). Secondly, the face location has to be identified and the face landmark is calculated in order to identify key point of the face such as nose tip and eyes.

We extract the face from the whole image in each frame. use the histogram of gradients HOG [36]. We follow the implementation of [36] by divide the image into a 16x16 patch and calculate the orientation of the gradient in 9 bins of unsigned orientation between 0 to 360 degrees. The vector is then normalized to get L2-Norm, L1-Norm and L1-Sqrt in equations 1,2, and 3.

$$L2 - Norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad \textbf{(1)}$$

$$L1 - Norm : f = \frac{v}{\|v\|_1 + e} \quad \textbf{(2)}$$

$$L1 - Sqrt : f = \sqrt{\frac{v}{(\|v\|_1 + e)}} \quad \textbf{(3)}$$

Where v is the non-normalized vector containing all histograms in a given block and $\|v\|_k$ is the k-norm (k = 1,2 ..) and e is constant with a small value?

According to the Dalal et.all in [36], the histogram orientation with L2-Norm and L1-Sqrt gives a good performance. The result of this procedure is a localized face area, as figured out below.

After the face location is determined, the face landmark has to be identified. We implement the face landmark recognition proposed by [37]. Detect the face orientation in each frame and record the relative position to front face direction. The basic idea is finding 68 points within the face. Kazemi et.all has trained the detector and provide the model to be used in the detection task. They use a cascading ensemble regression tree (ERT) to estimate the position of the points in the face landmark. They use the formula below to define the face landmark as a shape S, where the points in the face is a member of the S vector.

In the training session, n face image (I) provided and the shape S. The training data consists of (I_i, S_i) where $i \in \{1 \ldots n\}$, n is the number of training images. The cascaded regression trained using equations 4,5 and 6.

$$\pi_i \in \{1, \ldots, n\} \quad (4)$$

$$\hat{S}_i^{(0)} \in \{S_1, \ldots, S_n\} \mid S_{\pi_i} \quad (5)$$

$$\Delta S_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)} \tag{6}$$

This is repeated for I = 1… N where N =nR where R is the initialization for each image I_i. Based on this data, regression 0 (r0) learn to yield the $S_i^{(1)}$. The regression is cascaded till from t = 0 to T. The formula for cascade regressing an estimating the shape can be seen in equations 7 and 8.

$$\hat{S}_i^{(t+1)} = \hat{S}_i^t + r_t(S_{\pi_i}, \hat{S}_i^t) \tag{7}$$

$$\Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \tag{8}$$

This process is repeated in a cascade of T regressors $r_0, r_1, \ldots, r_{T-1}$ to learn the best combination give a sufficient level of accuracy. Once the T regressor trained by n face images, a regression model for face landmark defined. This paper uses the pre-defined model to identify the face landmark.

Based of the face landmark, Identify the eye location Eye location is denoted by 37 – 42 of the left eye in the face landmark, and 43 – 48 for the right eye. The pupil is located in the middle of the eyes as the darkest object.

3.2.2. Face Tracking

Face movement responsible for detecting the pose of the participant relative to the camera. The face able to move in three-dimensional; however, we consider horizontal and vertical movement only.

Let the video is consist of n frames. The first frame would be frame 0, and the face tracking data start with i = 1 to n - 1. The x and y position is calculated by considering the displacement of the face position relative to the previous frame. The nose tips are considered as the center of the face. We consider the displacement of the position of the nose tips as face movement data. The value is set between 0 – 10 and whenever the value > 10, then it is recorded as 10.

$$f(x_i) = \begin{cases} x_{i-1} - x_i, & 0 < |x_{i-1} - x_i| < 10 \\ 10, & |x_{i-1} - x_i| \geq 10 \end{cases} \tag{9}$$

Where x_i is the current frame, x_{i-1} is the previous frame face coordinates.

Face tracking produces two output variable which is horizontal movement (x) and vertical movement y. The horizontal and vertical movement, then normalized in the range 0 to 10, is calculated respect to equation 9.

3.2.3. Eye-tracking

Eye-tracking indicates user focus to the user view. It will give us a clue when the user looks to another side of the test user interface focus. For example, when someone helps him with the paper outside of the camera view, the user's eyes will leave the test sheet focus.

The normal eye condition when the participant focus on the testing area (smartphone). Pupil position would be in two-dimension space x and y with +/- relative to the center of the eyes. For a single face, we assume that the left and right eye movements in the same direction.

The first task is to identify the face and the eyes in the face image. Secondly, estimate the center of the eye by identifying calculating $x_2 - x_1$ and $y_2 - y_1$, where x_2 and x_1 are the left and the right corner of the eye and y_1 and y_2 as the top and bottom of an eye. The job in every frame sequence is estimating the new position of the black node in the eye relative to the pupil in the previous frame. Figure 4 shows the face identification and pupil tracking during a particular frame in a cheating video. Once the position of the pupil detected, the relative position is calculated against the normal position of respect to equation 9.

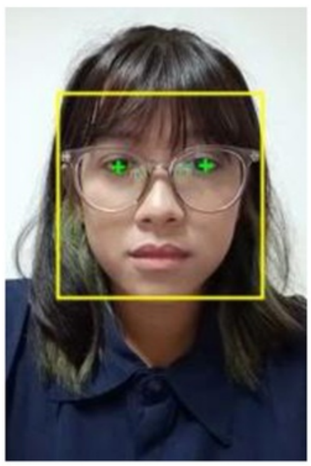


Figure 3 Face and pupil identification in a particular frame

3.2.4. Sound

During the simulation, the smartphone records the sound of the test environment. We transform the recorded sound and sample the amplitude of the sound in certain periods synchronous to the frame sampling. We normalized the sound into 0 to 10 and put on the time-series data.

3.3 Multilayer perceptron for activity recognition

Multilayer perceptron (MLP) is a class of artificial neural network (ANN). Multilayer perceptrons commonly referred as "vanilla" neural networks, especially it have a single hidden layer only [19]. The minimum architecture of MLP consists three layers which are an input, a hidden, and an output layer. Each node in the hidden and output layer is a neuron with a nonlinear activation function. Training the network to adjust the weight for each nodes is carried out by a supervised learning technique called backpropagation [20][21]. MLP capable to deal with non linear separation of classes due to the function of nonlinear activation function [22]. This capability is become the main differences between MLP and a linear perceptron.

MLP employs two common activation functions sigmoid in equation, 6 and rectifier linear unit (RELU) as described in equation 7:

$$S(x) = \frac{e^x}{1+e^{-x}} \quad (6)$$

$$R(x) = \max(0, x) \quad (7)$$

Where S(x) is the sigmoid activation function, x is the input variable, R(x) is the RELU function.

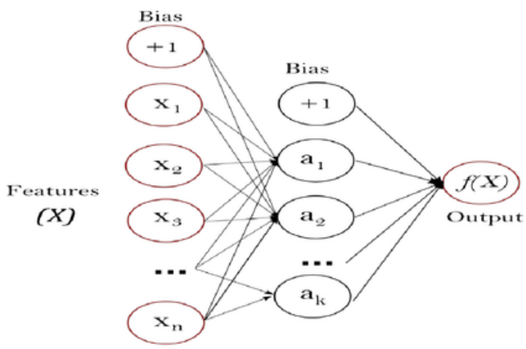


Figure 4 Multi-Layer Perceptron

In our experiments, we set the network with 20 hidden layers with 'RELU' activation function as described in equation 7 and trained using 'adam' optimizer. The input layer is set at 1290 nodes and the final output is a single node with binary class cheating and non-cheating.

3.4 Experiment Design

This research is designed as experimental research. A simulation of suspicious activity will be defined in seven possible cheating scenarios as follow

- Opening a book activity
- Reading a book on next to the smartphone
- Using a laptop to browse information
- Read some else clue outside the camera

- Send and receive a message from the second smartphone
- Someone else presence outside of the back camera
- Asking for someone else trough voice

Those activities are considered as cheating during an online exam. In this research, we make a real experiment using 20 subjects to test all the scenarios. We recorded 7 activities for each subject, and in total 140 activities recorded and another 7 times of each subject in normal (non-cheating) activity. Therefore, we have 280 activities in total, 140 cheating and 140 non-cheating.

3.5 Evaluation

We divide the data into 5 folds and carry out cross-validation. The proportion of training and testing data 80%, 20%, respectively. The evaluation is done for those five classifiers in each round. The quality metrics of the classification measured are accuracy, recall and precision and f measure.

4. Result and Discussion

The activity classification is carried out by a multilayer perceptron (MLP) on a standardized input parameter. In order to standardize the data, the duration of each activity in the listed scenarios are set in a fifteen-second duration. The shorter activity is padded with zero value in each variable.

We consider 12 frames per second is enough since the mobile exam scenario is controlled. In fact, when the examinee uses the smartphone the movement of the head will be limited by the camera position hold by hand would not be possible to move too far.

The original audio-video signal is transformed into face tracking, pupil tracking and sampled audio. The length of the output of the transformation is varied due to the duration of the video between 14 to 16 seconds. In order to ensure the equal length of the input vector to the network, we put zero paddings at the end of the signal. The expected length of the signal is 258 to fit the number of input nodes in the MLP architecture. Figure 5 shows the snippet of mid-level sound signal representation of a particular activity. Figure 6 and 7 shows the face tracking and eye tracking respectively. X and y-axis represent horizontal and vertical movement.

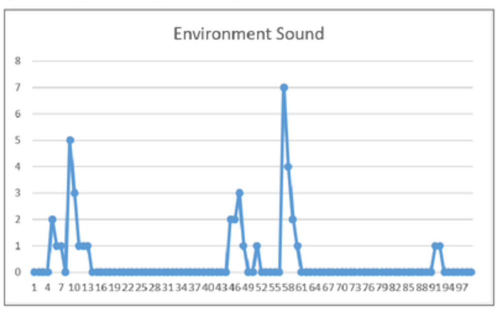


Figure 5 Sound mid-level representation

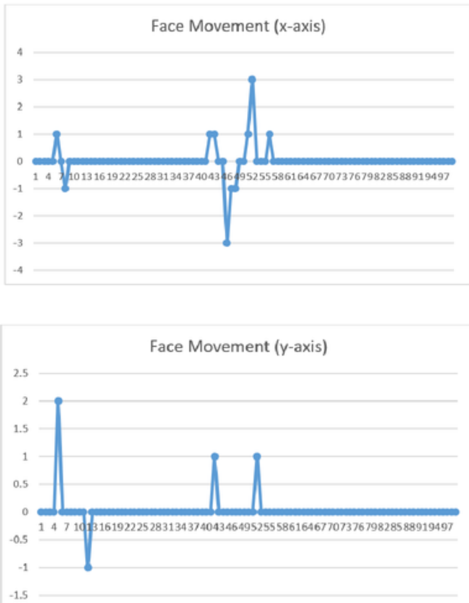
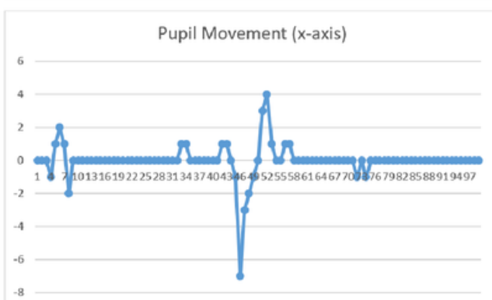


Figure 6 **Face Tracking movement mid-level signal**



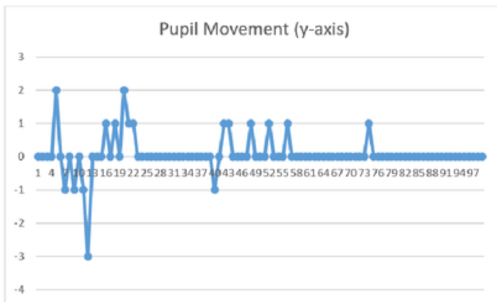


Figure 7 Output Signal

We concatenate the signal into a single vector for each activity. We transform the mid-level signal 5 x 258 into a vector with 1 x 1290 size.

The entire mid-level signal for all the dataset consists of 280 rows. We divide the data into fivefold randomly and validate the classifier using five-fold cross-validation. Table 1 presents the classification metrics of the experiments.

Table 1 Classification result for each fold.

Fold	Accuracy	Recall	Precision	F1 Score
1	0.892857	0.892857	0.902527	0.893131
2	0.892857	0.892857	0.912338	0.892033
3	**0.964286**	**0.964286**	**0.964286**	**0.964286**
4	0.892857	0.892857	0.894081	0.892299
5	0.890909	0.890909	0.899609	0.890475
Average	0.906753	0.906753	0.914568	0.906445
Std Dev	0.023013	0.023013	0.019887	0.023136

According to the experiments, we got the best model on the fold three, where the classifier successfully classifies 96% of the activity based on the available data. The rest of the experiments achieves 89% accuracy. We consider the balance weight between the recall and precision in calculating the F1 score. This is because of the balance portion of cheating and non-cheating class in the dataset.

The conversion of the video data into face tracking and pupil tracking has been done successfully. As shown in figure 5,6 and 7 The downsampling into 12 frames per second still produce a good result where the accuracy achieves well over 90% on average and in the best model (fold 3) 96% of accuracy is achieved. The simple mechanism to transform video and audio data into a simple representation still has enough discriminating power for a binary classifier.

Table 2 The performance of previous research [10]

Cheating Classifier	Recall	Precision	Accuracy
Webcam-only ACD	76.60%	84.60%	83.30%
Multi-modal ACD	82.10%	60.00%	81.10%
PCD	96.40%	94.10%	95.60%
Webcam-only MOOP	78.60%	90.40%	86.10%
Multi-modal MOOP	90.50%	93.80%	92.70%

Comparing to the previous research in [10] shows in table 2 — their research using webcam, EEG and gaze detector in their research. The accuracy of our result is slightly under their PCD and multimodal MOOP. It is reasonable because they use much more complicated tools. If we fairly compare to their results with a single webcam, our proposal overperforms compare [10]. We achieve 90.7% of accuracy while their web-cam only achieved 86.1% — another online proctoring with a double camera and sound recording proposed by [11]. Our result was 3.7% better than their accuracy at 87%.

5. Conclusion

This research presents an audio-video analysis for cheating action recognition to support a mobile exam proctoring. We conclude that the audio and video signal feasible to recognize the cheating and non-cheating activity. The average accuracy, precision, recall, and F1-score at $90.7 \pm 2.3\%$, $91.4 \pm 2.3\%$, $90.7 \pm 1.99\%$ and $90.6 \pm 2.3\%$ respectively. Compared to the state of the art in online cheating detection, we achieve a high recognition rate. Compare to webcam only proctoring in [10] and double camera plus audio in [11], we achieve better accuracy with almost a 4% margin. In addition, our research is designed for a smartphone exam environment with an implementation of standard smartphone accessories without any additional device. This research still needs more data for multiclass classification to identify which action happened. The possibility of using more sensors in the smartphone, such as accelerometer, the gyro will enrich the dataset and expected to improve the accuracy.

References

1. J. E. Seaman, D. Ph, J. Seaman, and D. Ph, “Distance Education State Almanac 2017,” 2017.
2. I. E. Allen and J. Seaman, “Grade Change Tracking Online Education in the United States,” 2014.
3. A. Fask, F. Englander, and Z. Wang, “Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment,” *J. Acad. Ethics*, vol. 12, no. 2, pp. 101–112, 2014.
4. C. G. King, R. W. Guyette, and C. Piotrowski, “Online exams and cheating: An empirical analysis of business students’ views,” *J. Educ. Online*, vol. 6, no. 1, pp. 1–11, 2009.

5. D. L. King and C. J. Case, "E-CHEATING: INCIDENCE AND TRENDS AMONG COLLEGE STUDENTS," *Issues Inf. Syst.*, vol. 15, no. 1, pp. 20–27, 2014.

6. H. Corrigan-Gibbs, N. Gupta, C. Northcutt, E. Cutrell, and W. Thies, "Deterring cheating in online environments," *ACM Trans. Comput. Interact.*, vol. 22, no. 6, 2015.

7. D. Foster and H. Layman, "Online proctoring system compared," *U.S. Pat. Appl. No. 12/723,666.*, pp. 1–12, 2013.

8. M. Kaiiali, A. Ozkaya, H. Altun, H. Haddad, and M. Alier, "Designing a Secure Exam Management System (SEMS) for M-Learning Environments," *IEEE Trans. Learn. Technol.*, vol. 9, no. 3, pp. 258–271, 2016.

9. R. Bawarith, D. Abdullah, D. Anas, and P. Dr., "E-exam Cheating Detection System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 176–181, 2017.

10 X. Li, K. M. Chang, Y. Yuan, and A. Hauptmann, "Massive open online proctor: Protecting the credibility of MOOCs Certificates," in *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 2015, pp. 1129–1137.

11 Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu, "Automated Online Exam Proctoring," *IEEE Trans. Multimed.*, vol. 19, no. 7, pp. 1609–1624, Jul. 2017.

12 H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," *IEEE Trans. Multimed.*, vol. 12, no. 6, pp. 563–575, 2010.

13 L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez, "Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior," *IEEE Trans. Multimed.*, vol. 16, no. 4, pp. 1018–1031, 2014.

14 W. Zajdel, J. D. Krijnders, T. Andringa, and D. M. Gavrila, "CASSANDRA: Audio-video sensor fusion for aggression detection," *2007 IEEE Conf. Adv. Video Signal Based Surveillance, AVSS 2007 Proc.*, pp. 200–205, 2007.

15 I. Lefter, G. J. Burghouts, and L. J. M. Rothkrantz, "Automatic audio-visual fusion for aggression detection using meta-information," *Proc. - 2012 IEEE 9th Int. Conf. Adv. Video Signal-Based Surveillance, AVSS 2012*, pp. 19–24, 2012.

16 Z. Wang, J. Yu, and Y. He, "Soccer Video Event Annotation by Synchronization of Attack-Defense Clips and Match Reports with Coarse-Grained Time Information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 1104–1117, 2017.

17 M. H. Kolekar and S. Sengupta, "Bayesian Network-Based Customized Highlight Generation for Broadcast Soccer Videos," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 195–209, 2015.

18 G. R. Cluskey, C. R. Ehlen, and M. H. Raiborn, "Thwarting online exam cheating without proctor supervision," *J. Acad. Bus. Ethics*, 2011.

19 A. Wahid, Y. Sengoku, and M. Mambo, "Toward constructing a secure online examination system," in *ACM IMCOM 2015 - Proceedings*, 2015.

20 N. L. Clarke, P. Dowland, and S. M. Furnell, "e-Invigilator: A Biometric-

Based Supervision System for e-Assessments," in *International Conference on nformation Society (i-Society), 2013*, 2013, p. 5.

21 S. Asha and C. Chellappan, "Authentication of e-learners using multimodal bometric technology," in *IEEE- International Symposium on Biometrics and Security Technologies, ISBAST'08*, 2008.

22 jALALI KHATEREH, "An Automatic method for cheating detection."

23 A.V. Senthil Kumar and M. Rathi, "Keystroke Dynamics: A Behavioral Biometric Model for User Authentication in Online Exams," in *Biometric Authentication in Online Learning Environments*, www.igi-global.com, 2019, pp. 183–207.

24 A. Bal and A. Acharya, "Biometric authentication and tracking system for online examination system," in *2011 International Conference on Recent Trends in Information Systems, ReTIS 2011 - Proceedings*, 2011, pp. 209–213.

25 A. Fayyoumi and A. Zarrad, "Novel Solution Based on Face Recognition to Address Identity Theft and Cheating in Online Examination Systems," *Adv. Internet Things*, vol. 04, no. 02, pp. 5–12, 2014.

26 B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Trans. Multimed.*, vol. 17, no. 7, pp. 1107–1119, Jul. 2015.

27 Y. Liu, Z. Xie, X. Yuan, J. Chen, and W. Song, "Multi-level structured hybrid forest for joint head detection and pose estimation," *Neurocomputing*, vol. 266, pp. 206–215, 2017.

28 F. Dornaika and B. Raducanu, "Three-Dimensional Face Pose Detection and Tracking Using Monocular Videos: Tool and Application," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 39, no. 4, pp. 935–944, 2009.

29 C. Qing, J. Jiang, and Z. Yang, "Normalized Co-Occurrence Mutual Information for Facial Pose Detection Inside Videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1898–1902, 2010.

30 V. Drouard, R. Horaud, A. Deleforge, S. Ba, and G. Evangelidis, "Robust Head-Pose Estimation Based on Partially-Latent Mixture of Linear Regressions," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1428–1440, 2017.

31 Y. M. Cheung and Q. Peng, "Eye Gaze Tracking with a Web Camera in a Desktop Environment," *IEEE Trans. Human-Machine Syst.*, vol. 45, no. 4, pp. 419–430, 2015.

32 K. Krafka *et al.*, "Eye Tracking for Everyone," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2176–2184, 2016.

33 C. Meng and X. Zhao, "Webcam-Based Eye Movement Analysis Using CNN," *IEEE Access*, vol. 5, pp. 19581–19587, 2017.

34 H. D. Tran and H. Li, "Sound event recognition with probabilistic distance SVMs," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 6, pp. 1556–1568, 2011.

35 C. Y. Wang, J. C. Wang, A. Santoso, C. C. Chiang, and C. H. Wu, "Sound Event Recognition Using Auditory-Receptive-Field Binary Pattern and Hierarchical-Diving Deep Belief Network," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 8, pp. 1336–1351, 2018.

36 N. Dalal and W. Triggs, “Histograms of Oriented Gradients for Human Detection,” *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

37 V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

pre-submission

ORIGINALITY REPORT

10%	7%	4%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Nazarbayev University Student Paper	1%
2	Submitted to Universiti Teknologi MARA Student Paper	1%
3	Submitted to Taylor’s Education Group Student Paper	1%
4	Xuanchong Li, Kai-min Chang, Yueran Yuan, Alexander Hauptmann. "Massive Open Online Proctor", Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15, 2015 Publication	1%
5	itechprosolutions.in Internet Source	1%
6	arxiv.org Internet Source	1%
7	jestec.taylors.edu.my Internet Source	<1%

8	manualzz.com Internet Source	<1%
9	Submitted to University of Westminster Student Paper	<1%
10	Submitted to ITM University Student Paper	<1%
11	Submitted to University of Stirling Student Paper	<1%
12	Submitted to De La Salle University - Manila Student Paper	<1%
13	Submitted to Auckland University of Technology Student Paper	<1%
14	Yousef Atoum, Liping Chen, Alex X. Liu, Stephen D. H. Hsu, Xiaoming Liu. "Automated Online Exam Proctoring", IEEE Transactions on Multimedia, 2017 Publication	<1%
15	Submitted to Asian Institute of Technology Student Paper	<1%
16	Submitted to University of Sheffield Student Paper	<1%
17	journalofbigdata.springeropen.com Internet Source	<1%
18	M.-Y. Chow, R.N. Sharpe, J.C. Hung. "On the	

application and design of artificial neural networks for motor fault detection. I", IEEE Transactions on Industrial Electronics, 1993
Publication | <1%

19	dspace.lboro.ac.uk Internet Source	<1%
20	Xiaolong Zhou, Haibin Cai, Youfu Li, Honghai Liu. "Two-eye model-based gaze estimation from a Kinect sensor", 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017 Publication	<1%
21	pure.qub.ac.uk Internet Source	<1%
22	Submitted to School of Business and Management ITB Student Paper	<1%
23	Abdourahman Houssein Ahmed, Kidiyo Kpalma, Abdoulkader Osman Guedi. "Human Detection Using HOG-SVM, Mixture of Gaussian and Background Contours Subtraction", 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2017 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		