AI Based Proctoring System

Shantanu Deorankar, Gaurang Pateriya, Aniket Itankar, Sakshi Bhise

Information Technology, GCOEA

Abstract**—** Massive open online courses (MOOCs) and other forms of remote education continue to increase in popularity and reach. The ability to efficiently proctor remote online exam- inations is an important limiting factor to the scalability of this next stage in education. Presently, human proctoring is the most common approach of evaluation, by either requiring the test taker to visit an examination center, or by monitoring them visually and acoustically during exams via a webcam. However, such methods are labor-intensive and costly. In this paper, we present a multimedia analytics system that performs automatic online exam proctoring. The system hardware includes one webcam and a microphone, for the purpose of monitoring the visual and acoustic environment of the testing location. The system some basic components that continuously estimate the key behavior cues: user verification using eye detection, text detection, voice detection, mouth opening detection, head pose estimation, no person detection, multiple face detection and phone detection. By combining the continuous estimation components, and applying a temporal sliding window, we design higher- level features to classify whether the test taker is cheating at any moment during the exam. To evaluate our proposed system, we collect multimedia (audio and visual) data from 24 subjects performing various types of cheating while taking online exams. Extensive experimental results demonstrate the accuracy, robustness, and efficiency of our online exam proctoring system.

Keywords— Online exam proctoring (OEP), user verification using eye detection, text detection, voice detection, mouth opening detection, head pose estimation, no person detection, multiple face detection and phone detection

I. INTRODUCTION

MASSIVE open online courses (MOOCs) offer the potential to significantly expand the reach of today's educational institutions, both by providing a wider range of educational resources to enrolled students and by making educational resources available to people who cannot access a campus due to location or schedule constraints. Instead of taking courses in a typical classroom on campus, now students can take courses anywhere in the world using a computer, where educators deliver knowledge via various types of multimedia content. According to a recent survey [1], more than 7.1 million students are taking, at least, one online course in 2013 in America. It also states that 70% of higher education institutions believe that online education is a critical component of their long-term strategy. Exams are a critical component of any educational program, and online educational programs are no exception. In any exam, there is a possibility of cheating, and therefore, its detection and prevention are important. Educational credentials must reflect actual learning in order to retain their value to society. When exams are administered in a conventional and proctored classroom environment, the students are monitored by a human proctor throughout the exam. In contrast, there is no convenient way to provide human proctors in online exams. As a consequence, there is no reliable way to ensure against cheating. Without the ability to proctor online exams in a convenient, inexpensive, and reliable manner, it is difficult for MOOC providers to offer reasonable assurance that the student has learned the material, which is one of the key outcomes of any educational program, including online education.

In this paper, we introduce an AI based proctoring system to perform automatic and continuous online exam proctoring (OEP). The overall goal of this system is to maintain academic integrity of exams, by providing real-time proctoring for detecting the majority of cheating behaviors of the test taker. To achieve such goals, audio-visual observations about the test takers are required to be able to detect any cheat behavior. Many existing multimedia systems [11], [18] have been utilizing features extracted from audio- visual data to study human behavior, which has motivated our technical approach. Our system monitors such cues in the room where the test taker resides, using a camera and a microphone. The camera is located above or integrated with the monitor facing the test taker. In this paper, the camera is referred to as the "webcam". The webcam also has a built-in microphone to capture any sound in the room. Using such sensors, we propose to detect the following cheat behaviors: (a) cheat from text books/notes/papers, (b) using a phone to call a friend, (c) using the Internet from the computer or smartphone, (d) asking a friend in the test room, and (e) having another person take the exam other than the test taker.

We propose a hybrid two-stage algorithm for our OEP system. The first stage focuses on extracting middle-level features from audio-visual streams that are indicative of cheating. These mainly consist of some basic components: user verification using eye detection, text detection, voice detection, mouth opening detection, head pose estimation, no person detection, multiple face detection and phone detection. Each component produces either a binary or probabilistic estimation of observing certain behavior cues. In the second stage, a joint decision across all components is carried out by extracting high-level temporal features from the OEP components at the first stage. These new features are utilized to train and test a classifier to provide real-time continuous detection of cheating behavior. To evaluate the OEP system, we collect multimedia (audio and visual) data from 24 subjects performing various types of cheating while taking a multiple choice and fill in the blank exam. Extensive experimental results demonstrate the accuracy, robustness, and efficiency of our online exam proctoring system in detecting cheating behavior. This paper makes the following contributions:

(A) Proposes a fully automated online exam proctoring system with visual and audio sensors for the purpose of maintaining academic integrity.

(B) Designs a hybrid two-stage multimedia analytics approach where an ensemble of classifiers extracts middle level features from the raw data, and transforming them into high-level features leads to the detection of cheating.

(C) Collects a multimedia dataset composed of two videos and one audio for each subject, along with label information of all cheating behaviors. This database is publicly available for future research.

II. Related Work

Over the years, the demand for online learning has increased significantly. Researchers have proposed various methods to proctor online exams in the most efficient and convenient way possible, yet still preserve academic integrity. These methods can be categorized into three categories: (a) no proctoring [2],[8] (b) online human monitoring[8],[5] and (c) semi automated machine proctoring[8],[12]. No proctoring does not mean that test takers have the freedom of cheating. Instead, cheating is minimized in various ways. In [2] authors believe they can prompt academic honesty by proposing eight control procedures that enable faculty to increase the difficulty and thus reduce the likelihood of cheating. In [8], the authors offers a secure web-based exam system along with network design which is expected to prevent cheating. Online human monitoring is one common approach for proctoring online exams. The main downside is that it's very costly in terms of requiring many employees to monitor the test takers. Researchers have also proposed different strategies in full monitoring, such as in [5], where they use snapshots to reduce the bandwidth cost of transmitting large video files. Authors in [12] attempt to do semi-automated machine proctoring, by building a desktop robot that contains a 360^{o} camera and motion sensors. This robot transmits videos to a monitoring center if any suspicious motion or video is captured. The main problem is that a single camera cannot see what the subject sees, and as a result even humans may have a hard time detecting many cheating strategies. For example, a partner who is outside the camera view, but who can see the test questions (e.g., on a second monitor), could supply answers to the test taker using silent signals, or writing on a piece of paper which is visible to the test taker.

Among all prior work, the most relevant work to ours is the Massive Open Online Proctoring framework, which combines both automatic and collaborative approaches to detect cheating behaviors in online exams. Their hardware includes four components: one webcams and a microphone. The camera is mounted above the monitor capturing the face. Motion is used for classification by extracting dense trajectory features. However, this work is limited to only one type of cheating (i.e., reading answers from a paper), with evaluation on a small set of 9 subjects with 84 cheat instances. Since many types of cheating do not contain high- level motion, it is not clear how this method can be extended to handle them

Beyond educational applications, in the multimedia community, there is prior work on audio-visual-based behavior recognition. Authors in [18] study audio- visual recordings of head motion in human interaction, to analyze socio-communicative and affective behavioral characteristics of interacting partners. [9] Automatically predicts the hireability in real job interviews, using applicant and interviewer nonverbal cues extracted from the audio-visual data. In [4], they automatically estimate high and low levels of group cohesion using audio video cues. In [7], the authors used audio-visual data to detect a wide variety of threats and aggression, such as unwanted behaviors in public areas. Their two-stage methodology decomposes low-level sensor features into high-level concepts to produce threat and aggression detection. The addition of audio to video was also proven to complement many visual analysis problems, such as object tracking [6], event detection retrieval in field sports [13], and vision-based HCI system [11].

III. Proposed Method

In this work, we aim to develop a AI based proctoring system to detect a wide variety of cheating behaviors during an online exam session. Our proposed online exam process includes two phases, the preparation phase and exam phase. In the preparation phase, the test taker has to authenticate himself before beginning the exam, by using a password and face authentication. This phase also includes calibration steps to ensure that all sensors are connected and functioning properly. Further, the test taker learns and verbally acknowledges the rules of using the OEP system, such as, no second person is allowed in the same room, the test taker should not leave the room during the exam phase, etc.

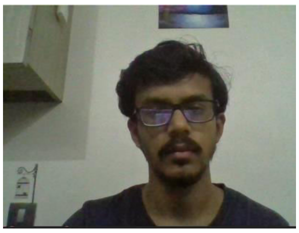
In the exam phase, the test taker takes the exam, under the continuous "monitoring" of our OEP system for real-time cheating behavior detection. We will be using two sensors (i.e., webcam and microphone) to capture audio-visual cues of the exam environment and the test taker. The sensed data is first processed using six components to extract middle-level features. These components are: user verification using eye detection, text detection, voice detection, include the component- dependent features, such as the mean and standard deviation within a window, and features based on the correlation among the components, such as the covariance features [16]. It is crucial to use a diverse and rich set of features to improve the overall detection performance of the OEP system, since the detection of some cheating behaviors relies on the ignition of multiple behavior cues. The remainder of this section describes the following topics: (A) the hardware components of the OEP system, (B) through (G) the six basic components of the system.

A. *Hardware component*

During an exam, the test taker may cheat by hearing or viewing forbidden information. Therefore, the OEP system hardware should be designed in a way to hear what the test taker hears and see what the test taker sees. So for vision purpose we will require Tensorflow version greater than 2.0 and below 2.4 and OpenCV library freely available in python as well as sklearn=0.19.1 for the sake of face spoofing. The model used was trained with these version and does not support recent ones.When it comes to audio part we will need Pyaudio and speech_recognition functionalities of python and natural language toolkit(nltk) The hardware requirement is Laptop with working camera

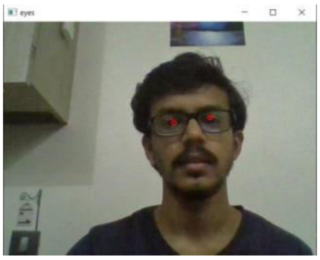
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

B. Face Detection



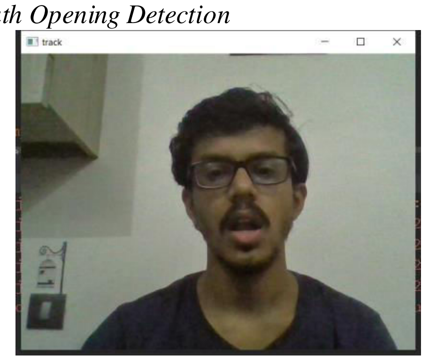
Earlier, Dlib's frontal face HOG detector was used to find faces. However, it did not give very good results. In face_detection different face detection models are compared and OpenCV's DNN module provides best result. It is implemented in face_detector.py and is used for tracking eyes, mouth opening detection, head pose estimation, and face spoofing. An additional quantized model is also added for face detector. This can be used by setting the parameter quantized as True when calling the get_face_detector(). On quick testing of face detector the normal version gave ~17.5 FPS while the quantized version gave ~19.5 FPS. This would be especially useful when deploying on edge devices due to it being uint8 quantized.

C. Eye tracking



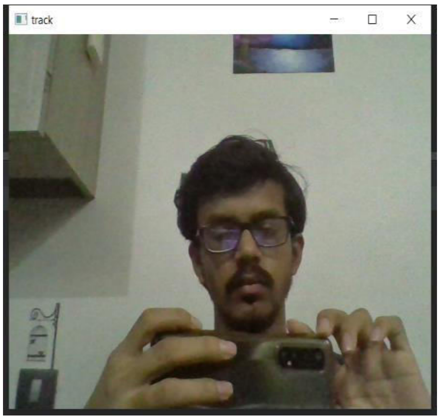
Takes 3D coordinates of facial landmarks and estimates creates a new black mask using NumPy of the same dimensions as our webcam frame. Store the (x, y) coordinates of the points of the left and right eyes from the keypoint array shape and draw them on the mask. it creates an image with the area between those points filled with that color. and later we can track the colored eyes in the eyeball tracking .eye_tracker.py is used to track eyes. However, it was written using dlib.

D. Mouth Opening Detection



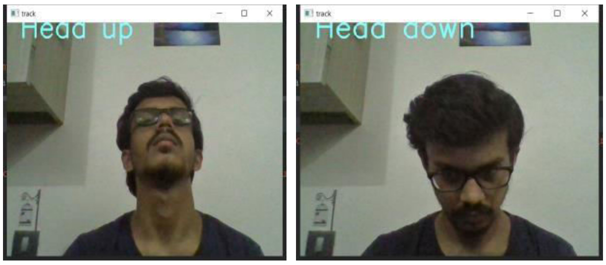
This is similar to the Eye tracking, It Takes 3D coordinates of facial landmarks and calculates the distance between the lips key points (5 outer pairs and 3 inner pairs) is noted for 100 frames and averaged. If the user opens his/her mouth the distances between the points increases and if the increase in distance is more than a certain value for at least three outer pairs and two inner pairs then infringement is reported. mouth_opening_detector.py is used to check if the candidate opens his/her mouth during the exam after recording it initially. It is using dlib which can be easily changed to the new models.

E. Person counting and mobile phone detection



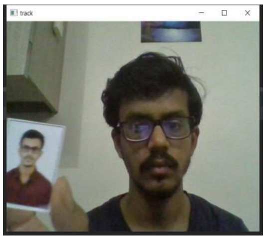
person_and_phone.py is for counting persons and detecting mobile phones. YOLOv3 is used in Tensorflow 2

F. Head pose estimation



Takes 3D coordinates of facial landmarks and estimates the rational and translational vectors at the nose tip head_pose_estimation.py is used for finding where the head is facing.

G. Face spoofing



Their approach is to use different color spaces like YCrCb and CIE L*u*v* and create histograms in these channels. Normal RGB color space is not used because the correlation between the red, green, and blue channels means that it obstructs separation between luminance and chrominance which is essential in spoofing attacks. In the paper they discussed the mathematics of converting RGB to YCrCb and CIE L*u*v*, however, we do not need to that to do that here as there are already predefined functions in OpenCV to do that. The histograms are concatenated into a feature vector FV = (Y, Cr, Cb, L, u, v) of size 1536 that is passed as input to an extra tree classifier for training purposes. They not only tested their method on faces but also on real and printed cork stoppers of wine bottles. face_spoofing.py is used for findingwhether the face is real or a photograph or image.

IV. EXPERIMENTAL RESULT

FPS Obtained:

Functionality	On Intel i5
Eye Tracking	7.1
Mouth Detection	7.2
Person and Phone Detection	1.3
Head Pose Estimation	8.5
Face Spoofing	6.9

If tested on a different processor a GPU will consider making a pull request to add the FPS obtained on that processor.

V. CONCLUSION

This paper presents a AI proctoring system for online exam, which aims to maintain academic integrity in e-learning. The system is affordable and convenient to use from the text taker's perspective, since it only requires having an inexpensive cameras and a microphone. With the captured videos and audio, we extract low-level features from basic components: user verification using eye detection, text detection, voice detection, mouth opening detection, head pose estimation, no person detection, multiple face detection and phone detection. These features are then processed in a temporal window to acquire high-level features, and then are used for cheat detection. Finally, with the collected database of 24 test takers representing real-world behaviors in online exam, we demonstrate the capabilities of the system, with nearly 87% segment-based detection rate across all types of cheating behaviors at a fixed FAR of 2%. These promising results warrant further research on this important behavior recognition problem and its educational application.

ACKNOWLEDGMENT

This project was sponsored in part by the Government College of Engineering, Amravati (GCOEA). The authors thank anonymous volunteers for participating in our OEP project. The authors also thank Sumit, Mohit, and Tushar for contributing to the study of human proctoring.

REFERENCES

[1] I. E. Allen and J. Seaman. Grade change: Tracking online education in the united states, 2013. Babson Survey Research Group and Quahog Research Group, LLC. Retrieved on, 3(5), 2014.

[2] G. Cluskey Jr, C. R. Ehlen, and M. H. Raiborn. Thwarting online exam cheating without proctor supervision. Journal of Academic and Business Ethics, 4:1–7, 2011.

[3] P. Guo, H. feng yu, and Q. Yao. The research and application of online examination and monitoring system. In IT in Medicine and Education, 2008. IEEE Int. Sym. on, Trans. on, 52(3):340–349, 2009.

[6] V. Kilic, M. Barnard, W. Wang, and J. Kittler. Audio assisted robust visual tracking with adaptive particle filtering. IEEE Trans. Multimedia, 17(2):186–200, 2015.

[7] I. Lefter, L. J. Rothkrantz, and G. J. Burghouts. A comparative study on automatic audio–visual fusion for aggression detection using metainformation. Pattern Recognition, 34(15):1953–1963, 2013.

[8] X. Li, K.-m. Chang, Y. Yuan, and A. Hauptmann. Massive open online proctor: Protecting the credibility of moocs certificates. In ACM CSCW, pages 1129–1137. ACM, 2015.

[9] L. Nguyen, D. Frauendorfer, M. Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. IEEE pages 497–502, 2008.

[4] H. Hung and D. Gatica-Perez. Estimating cohesion in small groups using audio-visual nonverbal behavior. IEEE Trans. Multimedia, 12(6):563–575, 2010.

[5] I. Jung and H. Yeom. Enhanced security for online exams using group cryptography. Education, IEEE Trans. Multimedia, 16(4):1018–1031, 2014.

[6] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vision, 42(3):145– 175, 2001.

[7] M. Reale, S. Canavan, L. Yin, K. Hu, and T. Hung. A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing. IEEE Trans. Multimedia, 13(3):474–486, 2011.

[8] W. Rosen and M. Carr. An autonomous articulating desktop robot for proctoring remote online examinations. In Frontiers in Education Conf., 2013 IEEE, pages 1935–1939, 2013.

[9] D. Sadlier and N. O'Connor. Event detection in field sports video using audio-visual features and a support vector machine. IEEE Trans. Circuits Sys. Video Technol., 15(10):1225–1233, 2005.

[10] E. Spriggs, F. De la Torre, and M. Hebert. Temporal segmentation and activity classification from first-person sensing. In Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), pages17–24, 2009.

[11] A. Tsukada, M. Shino, M. Devyver, and T. Kanade. Illumination-free gaze estimation method for first-person vision wearable device. In Proc. Int. Conf. Computer Vision (ICCV) Workshops, pages 2084–2091, 2011.

[12] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell., 30(10):1713–1727, 2008.

[13] A. Wahid, Y. Sengoku, and M. Mambo. Toward constructing a secure online examination system. In Proc. of the 9th Int. Conf. on Ubiquitous Information Management and Communication, page 95. ACM, 2015.

[14] B. Xiao, P. Georgiou, B. Baucom, and S. Narayanan. Head motion modeling for human behavior analysis in dyadic interaction. IEEE Trans. Multimedia, 17(7):1107–1119, 2015.

[15]