Video Interpretation for Cost-Effective Remote Proctoring to Prevent Cheating

Kiran P. Kamble and Vijay R. Ghorpade

Abstract In the rising era of globalization and digitization, remote education continues in gaining popularity and reach. Efficiently proctoring online remote examination is an important limiting factor to sustain the integrity of the exam as well as provide unprejudiced results. Currently human proctoring is the customer perspective to maintain integrity, either manually with the help of a test taker or by overseeing them visually through webcams. Online exams provide the examiner the choice, to choose the environment and the tools they wish to use during the exam. In response to this, our research proposes an application to detect fraudulent activities during online examination in real-time through the video recorded by the webcam of the examiner’s system. The application provides four features that continuously estimate the integrity of the exam: (1) User verification for checking impersonation by the examiner. (2) Multiple people together solving the exam. (3) Absence of examiner. (4) Detecting the use of mobile phones. The extensive experiment depicts accuracy of our cost-effective remote proctoring.

Keywords Integrity · Deep learning · Convolutional neural networks · Online proctoring and deep metric learning

1 Introduction

Online courses significantly expand the reach of institutions worldwide. It provides flexibility to the students who couldn’t reach the desired location. According to a survey [1] in America in 2013, more than 7.1 million students are enrolling at least
one online course. It also states that 70% of higher education institutions believe on online learning is a core part of their long-term planning. It is likely possible those students would perform cheating in this environment where no one is there to proctor them. Nearly 74% of online learners in 2013 concluded that it would be somewhat easy to cheat in online exams. Survey also concluded that in 2013, about 29% of the students admitted to cheating in online exams. During the COVID-19 pandemic, the count of organizing online teaching, conducting online exams, enrollment to online courses, and webinar-workshop has exponentially increased [2]. Statistics show that more than 30% of all test takers bring an unpermitted resource to their exam. Without proctor supervision, almost 600,000 items would have gone unnoticed before an exam in 2017 alone. Students have been issued penalties for performing malpractices. Hence, it is essential to sustain the integrity of online exams. In response to this, this research proposes an application to detect fraudulent activities [3] during online examination in real-time through the video recorded by the webcam of the examiner’s system.

2 Literature Survey

Automated Online Exam Proctoring [4] is a multimedia analytics system to perform automatic and continuous online exam proctoring. The examinee resides using two cameras and a microphone. The first camera is located above or integrated with the monitor facing the examinee (webcam). The other camera is to be worn or attached to eyeglasses, capturing the field of view of the examinee (wear cam). The sensed data is first processed using six components to extract middle-level features. These components are user verification, text detection, speech detection, active window detection, gaze estimation, and phone detection. After that, the middle-level features within a temporal window are fused to generate high-level features, which are then used for training and testing a cheat classifier. Their system is 6X slower-than-real-time speed based on the assumption that all six basic components process every frame in 25 FPS videos. In reality, it is very likely that they may process at a lower frame rate. Here the disadvantage is that it uses two cameras and eyeglasses thus increasing the cost of the system and has implemented a machine learning approach.

E-exam Cheating Detection System [5] has introduced the concepts of cheating and how it can be controlled in an online exam. It provides a technique for detecting and preventing students from cheating through continuous authentication and online proctor. E-exam management system is proposed to investigate cheating in the exam using Fingerprint Reader to authenticate the examinee, and Eye Tribe Tracker to continuously guarantee that the examinee is the one who is claiming to be. This system was developed in the visual C# and SQL server database. As a result, the system classified the examinee status as cheating or no cheating based on two parameters: the total an examinee time out the screen and the number of times, the examinee is out of the screen with an accuracy of 97.78%. The other applications built for online proctoring involve Proview [6] and ProctorU [7]. They are proctoring platforms with
AI-based machine-learning behavior analysis used to flag suspicious events. Exam sessions can be watched by test admins in real-time if desired; all are recorded end-to-end for later review. But the major disadvantage of these is they employ trained human proctors in their system. The motive of this proctoring system is to detect the majority of unfair activities that lead to cheating.

3 Proposed Method

3.1 Examinee Verification

Invigilator need to verify the identity of person who is going to appear for assessment. An IVOP system should be able to validate continuously whether examinee is actual who claiming to be in throughout the exam processes. Examinee expected to carry his or her identity card on which photo is present based on which invigilator supposed to manually verify the exam person in exam hall. While there is different types of verification for authentication of examinee, such as keystroke dynamics, fingerprint matching. We decide to authenticate with face as input due to cost-effective, automatic and robustness. There are various challenges for examinee authentication in IVOP. First the detection might be conducted in low lighting and different poses. Second, while capturing webcam partial occlusion occurred by the eyeglasses as show in Fig . The proposed work believes to be overcome these challenges. As the architecture show in Fig. 1 depicts the brief over of user verification. For the analysis we collected examinee recent photographs while registering. For each labeled name person more than 1000 sample are taken for further processing. The data set consist of few persons who are known to us and few persons are famous celebrity. Algorithm 1 show the detail step we followed, the $d_i$ represent path of data set for encoding. Result of encoding individual users are stored in $E(i)$. It initial data point of user 1. Loop over the entire dataset to read each face image. $Iname$ represent labeled name of individuals. Face extraction from each image using algorithm $Face Location$ input as $Im$. The result Face location is stored in $(x, y)$ coordinates. For the obtained coordinates we applied $FaceEncoding$ function which return encoding for each face. Face location detection is performed using HoG or Convolutional neural network mode. For each detected face we computed embedding of dimension 128 [8]. Subsequently we stored name of respective known person mapping with his or her 128d encoding in $N(E(i))$. As the examinee starts the exam, the webcam of his or her, Desktop or Laptop will be set on automatically. The system will be continuously monitoring the face of the examinee and displaying his or her name as label and if impersonation occurs then that new person will be detected as unknown thus resulting in declaring of fraudulent activity.
3.2 Checking Candidate Presence

This includes the scenario of the student getting disappeared that is moving from the place of exam and getting out of the field of the webcam which infers that examinee is undergoing some form of cheating by disappearing and will be detected as bounding boxes of examinee using face detection algorithm based on Convolution neural network and also with histogram of gradient. If the absence of face in the video. This will also result in a decrease in the integrity score and as stated before the reduction of integrity score below the threshold exam will be terminated. Figure 2 depicts the at the start of exam examine is present, however after specific interval of time it disappear which alert of absence. However, for some case IVOP system fail to identify presence of person in awkward poster of body, Such as face direction against the camera Fig. 3 indicate different face angle of IVOP system failure. To handle these case we proposed alternative as absence identification using captured

Fig. 1 Architecture of Interpreting video during online proctoring (IVOP)

Fig. 2 Time line of for verification of examine absence
background subtraction. The details of background subtraction is as follows. Though for few awkward poster IVOP system fail to detect person identity but actually these is not failure because if examiner is trying to observe or ask something to neighborhood, examiner will turn his eye or face which depicts this is as cheating. To handle against face condition we will capture different still background images with different lighting conditions before the start of exam or before examine appearing for exam notation as Bt. Face recognition IVOP model will start and for awkward poster as listed in Fig. 3 system will generate the flag as “examine is absence”, that will be first flag for final alarm for declaring as absence. Later on the examine is absence frame will subtract from the each Bt. If for each subtraction mean computation gives more than the some environment specific threshold value then second flag will set to 1. If both flags are ON then and then only system will trigger the final alarm as examine is absence.

### 3.3 Multiple People Attempting the Exam

Another case is the examinee seeking assistance from some other person during the exam will be detected as it will involve multiple people(faces) in the video frame will be displayed as multiple bounding boxes and thus again resulting in a reduction of integrity score. For the above three features, we are using face recognition and for face recognition, we are using Deep metric learning in dlib.
Algorithm 1: Examinee verification algorithm

**Data:** Collected data of registered examinee $d_i$

**Result:** $E(i), N(E(i))$

**Initialization:** $I_t = d_1, I_m = 0, e_i = 0, I_{name} = \"\"$, $I_f = 0$

**Loop** $I_t$ to $d_n$

\[
I_m = I_{image} \\
I_{name} = d_{name} \\
(x, y) = FaceLocation(I_m) \\
e_i = FaceEncoding(I_m, (x, y))
\]

**Loop** $i \leftarrow e_0$ to $e_n$

\[
E(i) = i \\
N(E(i)) = I_{name}
\]

**Loop** true

$\text{If } I_f = \text{read frame}(V_i)$

Preprocess $I_f$

\[
(x, y) = FaceLocation(I_f) \\
e_f = FaceEncoding(I_f, (x, y))
\]

**Loop** $i \leftarrow e_0$ to $e_n$

\[
C = MatchEncoding(e_f, i) \\
\text{If}(C == \text{True})\quad \quad \text{return } N(E(i)) \text{ and } (x, y)
\]
The model used is a ResNet [9] network with 29 convolutional layers. It is a version of the ResNet-34 network from the paper, Deep Residual Learning for Image Recognition by He, Zhang, Ren and Sun with few layers removed and the number of filters per layer reduced by half. The network has been already trained on a dataset of about 3 million faces. We are also using the face recognition library. This library uses HOG or CNN for face detection. Then it creates 128D encodings (A unique vector) for each person where each value in the vector represents a certain feature of a face. As stated before, each examinee has to register for the exam by providing his/her recent images. These images will be used to create the encoding of each student. Thus while running a video of the exam it compares the encodings of the examinee’s face in the video with the encodings of known students derived before and displays the result in the form of the bounding box and the name of the student.

3.4 Prohibited Object Detection

The examinee may use mobile either to seek answers directly or by contacting some other person. To avoid such cheating object detection is used. For object detection we are using faster Rcnnc inceptionv2 architecture pre-trained on the COCO dataset [10]. Faster R-CNN [11] is used for region proposal and inceptionv2 [12] is used for feature extractor. We have performed transfer learning for our class i.e person using a mobile phone. Object detection algorithm runs taking the webcam captured video as input and checks whether the examinee makes use of a mobile phone. If the mobile phone is detected integrity reduces. When the integrity goes below the threshold value exam stops and declares fraudulent. Transfer learning procedure: Step 1: Placing the jpg file and XML file in the training folder. Step 2: Placing the jpg file and XML file in the test folder. Step 3: Converting XML to .csv files. Step 4: Creating label map and config file. Step 5: Generating tfrecords as tfrecord is input to our architecture.

3.5 Difficult Level of Cheating

During online exam, examiner might be try some type of cheating which could not identified in simple observation. The type of cheating is described in bellow table indicated by spinning type of eye. Based on spinning type we have given remark. If any examiner wants to with this type of spinning in which category such as (1) Bending towards right side (2) Bending towards left side can be considered as suspect for cheating. Whereas other type wanting toward keyboard, spinning eye towards other than extreme left and right side is not belong to cheating suspect. Detail description of each type is given in Table 1 with this suspect we can restrict examiner to focus of his or her system only. One more possibility of cheating during online exam is unknowingly examiner make his or her lip movement and tries to ask something to their neighborhood. Table 2 show category of identification of suspect during lip movement.
Table 1  Difficult cheating with spinning eye

| Sr.No | Spinning Type         | Description               | Remark   |
|-------|-----------------------|---------------------------|----------|
| 1     |                       | Bending towards Right side| Suspect  |
| 2     |                       | Bending towards Down side | Keyboard |
| 3     |                       | Bending towards Left side | Suspect  |
| 4     |                       | Constant forward          | Attentive|
| 5     |                       | Bending towards North-East| Thinking |
| 6     |                       | Bending towards Up side   | Thinking |
| 7     |                       | Bending towards North-West side | Thinking |
Table 2  Difficult cheating with lip movement

| Sr. No | Frame 1 | Frame 2 | Frame 3 | Remark |
|--------|---------|---------|---------|--------|
| 1      | ![Image] | ![Image] | ![Image] | Suspect |
| 2      | ![Image] | ![Image] | ![Image] | Thinking |
| 3      | ![Image] | ![Image] | ![Image] | Thinking |

4  Dataset Used

Dataset for mobile detection is self-generated using the webcam of 2 megapixel and a certain part of the dataset is downloaded from Shutterstock. The total dataset consists of 2500 images. The dataset is divided into the ratio of 80:20 for training and testing purposes. The face recognition module was already trained one and we tested it on a dataset of 1412 images of 64 different people. One part of this dataset was images of celebrities that we collected from the Internet and another part is the images we got by capturing photos on webcam.

5  Experimental Result

This section highlights the endorsement of different component of IVOP system by answering the following questions. (1) How accurately can the system identify cheating? (2) How does various component set affect the result? (3) What is confidence score of detecting each prohibited activity type? (4) Is there any co-relation between various component? We started answering this question by providing performance evaluation for different components. The proposed work define evaluation with respect to the components of IVOP. How system correctly recognize known and unknown examiner, multiple people appearing, absences of examine and restricted object. Lastly the evaluation of whole model by maintaining Integrity score.
5.1 Evaluation of Examine

As we said in the data set we have created tow subset for the evaluation, First is 7 testing examine captured in live streaming labeled data set and second is 64 labeled data set contain face image of celebrity for which the accuracy is nearly 98%. In live streaming we captured image of each person individually and computed the following confusion matrix. Individually computing precision and recall. Based on individual precision and recall the average precision, recall and F1 score is computed. Tables 3 and 4 depicts the normalized confusion matrix and F1 score. This evaluation taken real time exam process with webcam mounted on monitor or laptop for which we obtained F1 score 81.76%.

5.2 Evaluation of Absence of Examine

IVOP system is designed with respect to the detecting and identification of face of examine. In real time scenario poster of examine face may vary such as, direction of

| Examine | Examine 1 | Examine 2 | Examine 3 | Examine 4 | Examine 5 | Examine 6 | Examine 7 |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Examine 1 | 0.82      | 0.13      | --        | 0.019     | 0.019     | --        | 0.0043    |
| Examine 2 | 0.0099    | 0.84      | 0.021     | 0.010     | 0.0056    | 0.0028    | 0.024     |
| Examine 3 | 0.040     | --        | 0.74      | --        | 0.17      | 0.011     | 0.026     |
| Examine 4 | --        | 0.089     | 0.057     | 0.84      | 0.005     | --        | 0.0076    |
| Examine 5 | 0.003     | 0.0085    | 0.057     | --        | 0.86      | 0.010     | 0.052     |
| Examine 6 | 0.013     | --        | 0.0045    | 0.0090    | --        | 0.92      | 0.045     |
| Examine 7 | 0.0012    | 0.011     | 0.042     | 0.027     | 0.20      | 0.021     | 0.69      |

| Examine | Precision (%) | Recall (%) | F1-score (%) |
|---------|---------------|------------|--------------|
| Examine 1 | 82.54         | 88.86      | 85.58        |
| Examine 2 | 84.06         | 83.35      | 83.70        |
| Examine 3 | 74.51         | 84.01      | 78.98        |
| Examine 4 | 84.03         | 80.65      | 82.30        |
| Examine 5 | 86.73         | 62.34      | 72.53        |
| Examine 6 | 92.76         | 92.34      | 92.54        |
| Examine 7 | 69.76         | 85.17      | 76.70        |
| Macro-F1/accuracy | –              | –          | 81.76        |
| Macro-precision | 82.05         | –          | –            |
| Macro-recall   | –              | 82.39      | –            |
face 90° towards monitor of system, 45° towards monitor or keeping half part of hand on face. For this type of cases face detection algorithm fails to detection presence of examine even if examine is there. Also if the distance between examine and monitor is too far. Such case we listed out in Fig. 3 using combine approach face recognition and background sub-straction we obtained 98% accuracy.

6 Conclusion

The proposed work presented cost-effective solution for remote proctoring. Examine verification module gives verification accuracy 81% in real time assessment. Absence of examine is handled by two approach first by absence of face and second by background subtraction accuracy obtained is 98%. Module of prohibited object detection gives mAP of 90%. Future scope for researcher is to design algorithm or model for difficult task, discussed in Sect. 3.5.

References

1. Allen E, Seaman J (2014) Grade change: tracking online education in the United States. Babson Survey Research Group Report
2. All India Council For Technical Education(AICTE), “Coursera’s Coronavirus Response Initiative” (2020)
3. Michael TB, Williams MA (2013) Student equity: discouraging cheating in online courses. Admin Issues J Educ Pract Res. Retrieved on 1 May 2015 from http://dx.doi.org/10.5929/2013.3.2.8
4. Atoum Y, Chen L, Liu AX, Hsu SDH, Liu X (2017) Automated online exam proctoring. IEEE Trans Multimedia 19(7):1609–1624. https://doi.org/10.1109/TMM.2017.2656064
5. Razan B, Basuhail A, Fattouh A, Gamalel-Din S (2017) E-exam cheating detection system. Int J Adv Comput Sci Appl (IJA-SCSA) 8(4). http://dx.doi.org/10.14569/IJACSA.2017.080425
6. Proview Homepage. https://info.talview.com/proview. Last accessed 05 Feb 2020
7. ProctorU Homepage. https://www.proctoru.com/. Last accessed 05 Feb 2020
8. Amos B, Ludwiczuk B, Satyanarayanan M (2016) Openface: a general-purpose face recognition library with mobile applications. CMU-CS-16-118. Tech. Rep, CMU School of Computer Science
9. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
10. Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Lawrence C, Dollár P (2014) Microsoft COCO: common objects in Context. CoRR
11. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
12. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. IEEE CVPR. Las Vegas, NV, pp 2818–2826. https://doi.org/10.1109/CVPR.2016.308