Using Face Recognition to Detect "Ghost Writer" Cheating in Examination

Huan He[1,4] Qinghua Zheng[1,2], Rui Li[4], Bo Dong[3,4]

[1] SPKLSTN Lab, Xi'an Jiaotong University.
[2] School of Electronic and Information Engineering, Xi'an Jiaotong University.
[3] National Engineering Lab of Big Data Analytics, Xi'an Jiaotong University.
[4] College of Distance Education, Xi'an Jiaotong University,
Xi'an 710049 China.
{hehuan, qhzheng, lrvberg, dong.bo}@mail.xjtu.edu.cn

Abstract. Cheating in examinations of the online distance education is a serious problem which may damage the fairness of exam and further undermine the credibility and reputation of certificates. In order to detect the "Ghost Writer" cheating strategy that existed in both online and offline exams, we propose the Student Identification by Face Recognition (SIFR) framework, a three layers architecture based on face recognition technique and micro-service principle, to detect the ghostwriter who takes the exam for others. In addition, we implement a prototype system based on open source projects and public cloud services. To evaluate the system, an experimental test was conducted with public data. The results indicated that the SIFR framework is feasible and the accuracy of detection is directly affected by the performance of face recognition service, which can be upgraded or replaced with better facial feature extraction module.

Keywords: Online distance education · Ghost writer · Cheating detection · Face recognition · Micro-service architecture

1 Introduction

Honesty is the cornerstone of all success, and there is no exception in education. With the rapid development of internet technologies, online distance education (ODE) plays an important role in promoting lifelong learning and providing foundation for long-term personal development [1, 2]. However, as the scale of enrollment increases, the problem of academic dishonesty becomes more apparent in online learning environment [3]. Especially, the problem of cheating in exams has been a major concern in ODE schools [4, 5], which not only damaged the fairness of examinations seriously, but also undermined the credibility and reputation of ODE certificates.

Present studies of cheating are focused on the following aspects: detection of cheating practices by analyzing multiply accounts' submission [6] or learning behavior and performance [7, 8]; motivations and environmental factors related to cheating [9-11]; and prevention methods [12, 13]. In order to detect and prevent cheating in online learning, technologies such as data mining and statistical methods were used to analyze students' learning behaviors or submissions [6-8]. In addition,

with the development of deep learning related technologies, face recognition has gradually become a mature technology provided as public cloud service on internet [14] and has been applied in educational environment to authenticate students [15] and to evaluate engagement by recognizing their facial expressions [16]. Since the literature has clearly highlighted the importance of the identification of cheating and feasibility of applying face recognition technology in online learning, we further add to this by proposing a technical framework to detect a typical cheating strategy in ODE examination context.

The major contributions of this paper are summarized as follows:

First, the "Ghost Writer" cheating strategy in ODE examinations is analyzed. And we summarized the challenges of anti-ghostwriter.

Second, the Student Identification by Face Recognition (SIFR) framework is proposed to address the challenges in detecting ghostwriters.

Third and last, a prototype system based on SIFR framework is implemented and experimental validated.

The rest of this paper is structured as follows: Section 2 describes the current examination system in ODE and analyzes the "Ghost Writer" cheating strategy. Section 3 proposes the SIFR framework for detecting the ghostwriter. Section 4 demonstrates the prototype system and discusses the experimental results. Section 5 concludes.

2 The "Ghost Writer" Cheating Strategy

Due to the wide geographical distribution of students, ODE schools usually commissioned learning centers located in various regions to recruit students and organize examinations. There are currently two types of examinations in ODE, entrance exams and course exams. Entrance exams are offline written examinations. Students must pass the entrance exams before starting online learning. Course exams are combinations of online and offline examinations. Some courses require students to take a written examination to test students' mastery of knowledge (e.g., engineering drawing, mechanical design, etc.) or include hands-on examinations (e.g., computer programming language, electrical and electronic technology, etc.), so only offline examinations can be used. Due to the huge size of enrollment and large number of course exams, it is a great challenge to maintain the fairness in the large-scale examinations.

In offline exams, due to the limitation in the number of examiners and their attention, it is difficult to monitor all activities in exam completely. Although cheatings such as "paging receiver" and "wireless earphone" have been prevented by metal detectors, some students still use the "Ghost Writer" cheating strategy in exams. They attempt to pass exams and earn credits by hiring ghostwriters to take the exam for them.

Before the exam begins, the examiner will inspect whether the photo on ID provided by the student is the same as that of himself/herself one by one. In order to pretend to be the student and pass the inspection, the ghostwriter merged student's photo and his/her own photo to make a fake photo indistinguishable from the student

and further counterfeit documents. They may also change their appearance (e.g., using glasses, fake beard, make-up, changing hair style, etc.) to make examiners difficult to judge immediately. In addition, it's a great pressure for examiners to check thousands of students in detail during exam season. Therefore, this cheating strategy has been implemented in some exams and has not been identified. The situation in online exams may be more serious. Since student's identity is validated only by username and password when login into exam system, which makes identifying ghostwriters more difficult.

By analyzing the above cheating practices of the "Ghost Writer" strategy and the difficulty in detecting and preventing this cheating strategy, we summarized three challenges of anti-ghostwriter as follows:

- Accurate detection in both online and offline exams. Although online examinations have become more prevalent in ODE, there are still many courses that have to adopt offline exams due to course content characteristics or technology limitations. Therefore, the solution of anti-ghostwriter should not only be able to support online exam, but also support the offline exams.
- Scalable for large-scale detecting. Since there are a large population of enrollment in ODE school every year and lots of courses provided to students, the solution of anti-ghostwriter should support horizontal expansion and contraction on demand.
- Affordable for learning centers and students. There are several technologies can provide reliable student identification such as handwriting matching, fingerprint recognition and iris recognition, etc. However, these technologies require the purchase of specialized equipment and software with trained personnel to operate, which would be a large investment and needs to be upgraded in hardware and software over time. It is unacceptable for ODE school to make additional large-scale investments in this respect. Especially for students, purchasing extra set of equipment only for online exams is impractical. As a result, the solution of anti-ghostwriter must be affordable for both ODE school and students.

3 The Proposed Framework

To detect and further prevent the "Ghost Writer" cheating strategy described above, we propose the Student Identification by Face Recognition (SIFR) framework for administrators and teachers to support anti-cheating in both online and offline ODE examinations. This framework depends on the following key technologies / services.

First, face recognition. As introduced in Section 1, face recognition has become a proven technology in many typical scenarios. With photo taken by a standard camera on mobile device or web camera on computer, this technology can provide facial feature detection and comparison at enough accuracy in typical scenarios without further investment in hardware. Not only does the industry have a large number of companies that offer related services, but there are also several open source projects published by companies, organizations or individuals, which provide solid foundation for establishing internal, private and customized services. *Second*, micro-services architecture. In the framework designed based on micro-services architecture, the key technologies or services can be packaged as web-based APIs for external system to

use. Without affecting external service access, the underlying implementation can be replaced, upgraded or expanded smoothly.

With the supports of above key technologies / services, the SIFR framework is presented in Figure 1. We will describe each layer of the SIFR framework in the following subsections.

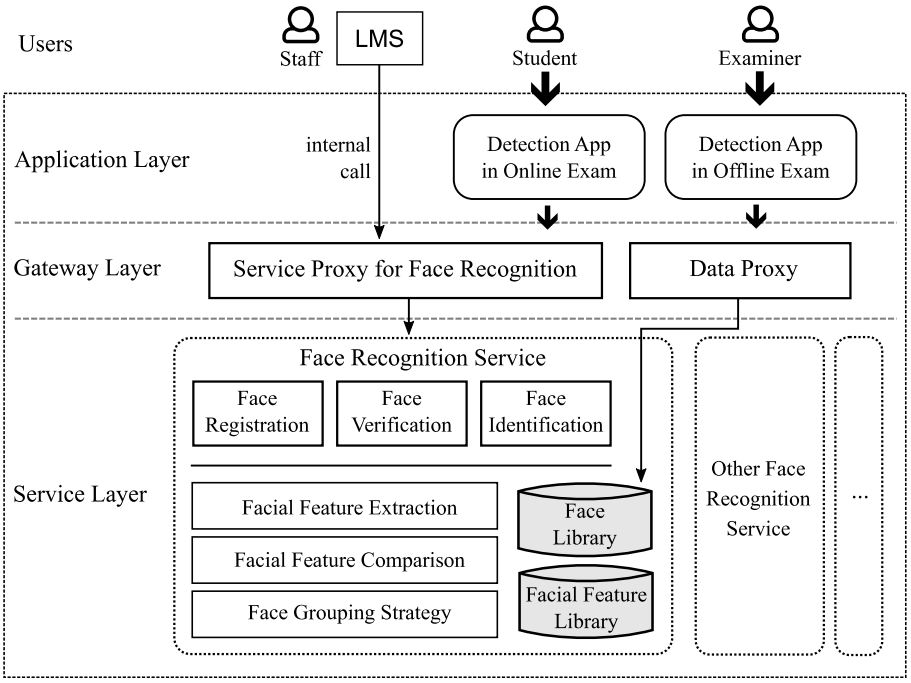

Fig. 1. The architecture of the SIFR framework.

3.1 The Service Layer

In this layer, there are two key technologies and one strategy to provide core support for face recognition service. *First*, the facial feature extraction is the most important function in this layer, it extracts a few basic measurements from the face in photo. *Second*, the facial feature comparison function calculates the difference of facial features between two faces. Small difference may indicate that two faces belong to the same person. *Third and last*, the grouping strategy is used to organize faces into groups to reduce computation load when identifying a specified face. Faces can be identified accurately even without grouping if the size of facial feature library is small. Nevertheless, as the number of faces increases, the possibility that several students have similar facial features will increase and the accuracy of face recognition may reduce. Therefore, it is necessary to design suitable grouping strategy in advance.

Based on the above three functions, we further summarize three core generic interfaces to meet the requirements of anti-ghostwriter.

- Face Registration. According to the principle of face recognition, a grouped facial feature library must be pre-built in order to provide base data for face comparison. As the basic interface at this layer, photos of all students will be imported through this interface into photo library, further processed and saved in facial features library by invoking key technology of facial feature extraction and specified grouping strategy. Because every student must submit a passport-style photo when they register in ODE school, the photo library and facial feature library can be built with this interface as soon as he/she submit it. The input of this interface includes a student ID, a group ID and a photo which belong to this student.
- Face Verification. This interface checks that weather the input photo belongs to a specified student by invoking the facial feature extraction and the facial feature comparison. If the photo belongs to the student, the verification is passed. The input of this interface includes a student ID and a photo to be verified.
- Face Identification. Similar to the function of face verification, by invoking the facial feature extraction and multi-times of facial feature comparison, this interface checks the input photo against a group of faces to find the student whom the photo may belong to. The input of this interface includes a group ID and a photo to be identified.

3.2 The Gateway Layer

In order to provide the scalable student identification service and related photo data to examiners, two types of proxies are required. As shown in Figure 1, the data proxy provides students' photo files to apps which help examiners to further identify students on their own. The service proxy connects to all services and exports the specified interfaces to users. In addition, the service proxy balances the work load from requests of apps to different workers. With the increasing requests of face recognition from users, single worker of face recognition service may not handle all of the computations. Therefore, more workers can be combined together, and the service proxy will distribute requests to them.

When the service provided in service layer is unavailable, upgrading or removed, the service proxy can switch to workable service in order to avoid interruption. On the other hand, heterogeneous services with same interfaces from different providers can be integrated into the service layer and provide face recognition service together. With this feature, any third-party AI service can be imported as a module in this framework to increase performance and accuracy of detection.

3.3 The Application Layer

Due to the differences in scenarios, the application of interfaces differs. The user in online exam scenario is student himself/herself, so the target student is clear for the app. Therefore, the app only need to use the interface of face verification. While in offline exam scenario, examiners may use both interfaces of face verification and face

identification to detect a student with student ID or an unknown student. Besides, the face registration interface will be invoked by LMS when staff in ODE school register student for exams.

4 The Implementation and Discussion

To verify the technical feasibility and workflow of the proposed SIFR framework, we implemented a prototype system based on open source projects and public cloud services. Details about each module used in each layer is listed in Table 1.

Table 1. The modules used in the prototype system of the SIFR framework.

Layer	Module	Implementation
Application layer	App in online exam	A plugin for exam system written in JavaScript
	App in offline exam	A HTML5 mobile app for Android smart phone
Gateway Layer	Service proxy	Nginx with a proxy server written in Python
	Data proxy	Nginx web server
Service Layer	Face recognition	A private service based on open source project: face_recognition and Flask in Python
		A public cloud service from Baidu Inc. [14]
	Data storage	Hadoop DFS as the photo library
		Redis database as the facial features library

In the application layer, we create a HTML5 mobile app for Android smart phone to help examiners in classroom and a plugin which is inserted in web page of exam system. The user interfaces of both apps are shown in Figure 2. With these apps the detection of "Ghost Writer" cheating strategy in both online and offline exams can be achieved without further investment in specialized devices. In addition, our private service and a public cloud service were both integrated in the prototype system. To further validate the performance of the SIFR framework, we tested the face identification interface by the mobile app of the prototype system with a small-scale face recognition dataset PubFig83 as students (made up of more than 100 images for each of 83 persons) [17], and other photos from internet as ghostwriters.

Table 2. The experimental results. (83 faces in one face group and 17 ghostwriters.)

	Student		Ghostwriter			
Service	Correct	Incorrect	Correct	Incorrect	Accuracy	Precision
Private service	75.9	7.1	17	0	92.90%	91.45%
Public cloud service	78.9	4.1	17	0	95.90%	95.06%

As listed in Table 2, all ghostwriters were correctly identified. On average, more students were wrongly identified as other students by our private service than public cloud service. The results indicate that the proposed SIFR framework is feasible and the performance of face recognition may be acceptable with public cloud service.

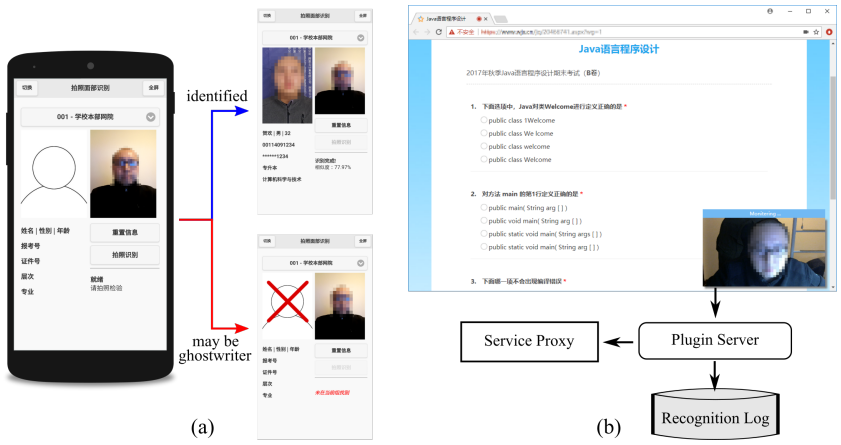

Fig. 2. The apps created for examiners and students. (a) The mobile app on Android. Examiners can use this app to identify students with two steps: 1) select the face group, 2) take photo. This app will automatically send photo and identify the person. (b) The plugin in exam system. After student login to the system, the plugin will request to authorize access to the camera and stay in background if permission is granted. During the exam, the plugin will randomly take photo and send back for detection. The recognition results will be saved for further analysis.

Although our private face recognition service is functional available, the accuracy is relative lower than public cloud service. This result may be due to small size of training sample and generic facial feature model. Besides, both the training sample size and hardware acceleration equipment used in public cloud service are far beyond ours, which also contributes to their accuracy. As a result, the facial feature extraction of our private service should be upgraded in order to achieve better performance. As mentioned in Section 2, ghostwriters may make their appearance difficult to identify by applying makeup or other means (e.g., pretending to be injured by wrapping gauze). In this situation, the facial features may not be detected properly, which indicates that there are still limitations in technical means, and further research is required not only in technologies but also in regulations and execution.

5 Conclusion

In order to detect the “Ghost Writer” cheating strategy in ODE examinations, we proposed a framework with face recognition to identify students in both online and offline exams. A prototype system was developed, which implemented two face recognition services. In addition, the system was tested on a small-scale public dataset, and the experiment results indicated that the proposed SIFR framework is feasible. Whereas the results also revealed that there were limitations in the accuracy of our private face recognition service. The future work will focus on improving the performance of our private service and conducting large-scale test in online exams.

Acknowledgments. This research was partially supported by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" under the National Key Research and Development Program of China with Grant No. 2016YFB1000903, the MOE Innovation Research Team No. IRT17R86, the National Science Foundation of China under Grant Nos. 61721002, 61502379, 61532015, and Project of China Knowledge Centre for Engineering Science and Technology.

References

1. Ding, X., Niu, J., Han, Y.: Research on distance education development in China. British Journal of Educational Technology. 41, 582–592 (2010).
2. Hu, F.: Return to Education for China's Return Migrant Entrepreneurs. World Development. 72, 296–307 (2015).
3. Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., Thies, W.: Deterring Cheating in Online Environments. ACM Transactions on Computer-Human Interaction. 22, 1–23 (2015).
4. Arnold, I.J.M.: Cheating at online formative tests: Does it pay off? The Internet and Higher Education. 29, 98–106 (2016).
5. Keresztury, B., Cser, L.: New Cheating Methods in the Electronic Teaching Era. Procedia - Social and Behavioral Sciences. 93, 1516–1520 (2013).
6. Alexandron, G., Ruipérez-Valiente, J.A., Chen, Z., Muñoz-Merino, P.J., Pritchard, D.E.: Copying@Scale: Using Harvesting Accounts for Collecting Correct Answers in a MOOC. Computers & Education. 108, 96–114 (2017).
7. Sabonchi, A.K.S., Görür, A.K.: Plagiarism detection in learning management system. In: 2017 8th International Conference on Information Technology (ICIT). pp. 495–500 (2017).
8. Salhofer, P.: Analyzing student behavior in CS courses: A case-study on detecting and preventing cheating. In: 2017 IEEE Global Engineering Education Conference (EDUCON). pp. 1426–1431 (2017).
9. Kalhori, Z.: The Relationship between Teacher-student Rapport and Student's Willingness to Cheat. Procedia - Social and Behavioral Sciences. 136, 153–158 (2014).
10. Toki, E.I., Tafiadis, D.C.: Identification of plagiarism by Greek higher education students. Do I cheat? In: 2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014). pp. 364–367 (2014).
11. Turner, S.W., Uludag, S.: Student perceptions of cheating in online and traditional classes. In: 2013 IEEE Frontiers in Education Conference (FIE). pp. 1131–1137 (2013).
12. Awad, M.K., Zogheib, B., Alazemi, H.M.K.: A penalty scheme for academic dishonesty. In: Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE). pp. 580–584 (2013).
13. Kaur, N., Prasad, P.W.C., Alsadoon, A., Pham, L., Elchouemi, A.: An enhanced model of biometric authentication in E-Learning: Using a combination of biometric features to access E-Learning environments. In: 2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES). pp. 138–143 (2016).
14. Baidu Face Recognition of AI Platform, http://ai.baidu.com/solution/faceprint.
15. Zhao, Q., Ye, M.: The application and implementation of face recognition in authentication system for distance education. In: 2010 International Conference on Networking and Digital Society. pp. 487–489 (2010).
16. Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A., Movellan, J.R.: The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. IEEE Transactions on Affective Computing. 5, 86–98 (2014).
17. Pinto, N., Stone, Z., Zickler, T., Cox, D.: Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In: CVPR 2011 WORKSHOPS. pp. 35–42 (2011).