Biometric Application for authentication and management of online exams

Amine Haytom, Christophe Rosenberger, Christophe Charrier, Charles Zhu, Clément Régnier

Biometric Application for authentication and management of online exams

Amine Haytom
GREYC lab
6, Rue Notre Dame de Nazareth, 75003 Paris
Haytom.amine@testwe.eu

Christophe Charrier
GYEYC lab
6, boulevard du Maréchal Juin, 14050 Caen
christophe.charrier@unicaen.fr

Clément Régnier
TestWe Company
6, Rue Notre Dame de Nazareth, 75003 Paris
clement.regnier@testwe.eu

Christophe Rosenberger
GREYC lab
6, boulevard du Maréchal Juin, 14050 Caen
christophe.rosenberger@ensicaen.fr

Charles Zhu
TestWe Company
6, Rue Notre Dame de Nazareth, 75003 Paris
charles.zhu@testwe.eu

Abstract

The main objective of this work is the design of a smart application to improve the management of remote exams, using new techniques to have a robust authentication by combining different proof of identity and the adaptation of anti-cheat tools and automatic monitoring with the detection of fraud.

In addition, conduct experiments under real-life conditions to identify inhabited behaviours with confidence indicators. Control the environment, and evaluate the system's ability to make a good decision.

Finally, respect for users' privacy requires the integration of reliable and secure tools to protect personal data and ensure the confidentiality of data exchanged.

Keywords
Personal data, Biometric signature, Learner behaviour.

1. Introduction

Today, with the rapid growth of online courses and exams, more and more students or professionals are applying for distance learning. Certainly, distance learning offers several advantages; the opportunity to take training in a distant establishment, arrange teaching for the personal and professional life of the learner, have great organizational flexibility and reduce accommodation and transportation costs.

It allows to have the same level obtained in classroom training. Secure tools must be used to safely access the exams. It is essential to control a student's access and environment during an exam. Online assessment and supervised tests address the issue of student verification and the environment in which access to unauthorized documents and resources is a major problem

Data processing and environmental control must respect the principles encouraged by the European regulation on the protection of personal information. The information often collected in a regulatory context can make the verification of a person a safe and secure step. Higher education that encourages schools to fight cheating more effectively, it is more interested in the quality and integrity of online proctoring.

This work can be divided into two parts; users authentication solution based on facial recognition and the integration of a security system to detect fraud during online exams.

A biometric model is proposed to maintain the integrity and quality of the online assessment; at beginning, the system integrates face recognition modality. Biometrics refers precisely to a computer technology that stimulates the physical presence of an individual, it conceives its identity and its movements as sources of dangers and risks, it represents a new era of modernity.

Generally, biometrics is divided into two categories; physical and behavioural. Physical analysis include the geometry of the hand, the retina, and the characteristics of the iris or the facial features. Behavioural analysis helps to verify a person's identity by examining measurable activity, such as gesture recognition and keystroke dynamics.

The integration of other modalities such as speaker recognition or detection allows to improve the efficiency of exam management system.

This paper represents information about using face recognition to obtain biometric authentication of a user.

1.1. Facial recognition

Many applications nowadays require personal data about the user, where the proof of identity is used to provide a secure service and to guarantee that a person is authorized to access an online multimedia service. In addition, it is necessary to convince users that the proposed solution is credible and that the information is processed reliably.

Without a doubt, facial recognition could play a necessary role to improve the identity verification of individuals. This biometric solution has a very high degree of acceptability when compared to other modalities such as iris or gesture recognition.

1.1.1 Generalities

We present in this chapter a solution to have a biometric signature based on the physical characteristics of the learner that will be used as proof of identity. Actually, facial recognition will be integrated into the remote examination management system as a basic modality to increase the security level of the application.

The biometric signatures are part of a lot of so-called sensitive data. So, we implement security measures and data protection processes by obtaining an encrypted template in the form of a secure code

The personal data are either material what the user possesses (QR-code, smart card, RFID tag), behavioural what the user knows how to do (typing dynamics and gesture recognition) or physiological data that which is related to the user (face, fingerprint, iris, voice).

1.1.2 Face recognition approaches

Nowadays, facial recognition systems have become an indispensable tool in many fields of application. It remains an acceptable solution for many people because it is less expensive and easier to use.

Facial recognition systems are well-known computer vision applications whose purpose is to verify or identify a person. There are two main approaches to facial recognition;

Facial recognition systems are well-known computer vision applications whose purpose is to verify or identify a person on the measurement of selected facial features. There are different approaches for facial recognition; Global approach: technically speaking, this approach is based on the global analysis of the face. The advantages of this approach the low time of calculates, the limitations are related to the size of the sample used as well as the difficulty to compute higher order statistics. Local approach: this approach is based on the analysis of face via different techniques, such as; Geometric measurements, Elastic Graph Matching and so on.

Facial recognition systems exploit digital images to provide contactless authentication. These systems have become widespread due to increasing demand in the areas of security and automatic identity verification. Nevertheless, it is possible to find approach with good performance that work despite the different types of constraints, so that the person to identify is in a controlled environment, using cameras with a specific orientation.

However, testing in a non-controlled environment (variable lighting, variable orientation of the person towards the camera, moving person, etc.) remain problematic, because the verification of the users is carried out with high uncertainty especially during the acquisition phase.

Certainly, there are other methods to verify the identity of a person such as Deep Learning based approach. To address the critical environmental issue, the research activity focused on the analysis of human perception in order to describe mathematically the processes that lead to the recognition of actions. Analysing human brain way to recognize faces and objects. Implementing algorithms operating in uncontrolled environments with a high level of accuracy. Today, there are several works and scientific publications based on artificial intelligence for user verification and identification.

Face recognition is a constantly evolving field, continually improving. Identity verification through deep learning has recently received a lot of attention from the research and industry community and is starting to be applied in a variety of areas, mainly for security.

We can find study that handle the authentication based on Convolution Neural Network [1], researchers compute VGG-Face CNN descriptors using CNN implementation based on the VGG-Very-Deep-16 CNN architecture and they evaluate it on a large databases (Labeled Faces in the Wild and the YouTube Faces).

Many open source library to develop ML models like Keras, Tensor Flow and Caffe. As well as several databases, for example CelebFaces, Labeled Faces in the Wild and Public Figures Face. Deep learning remains an effective technique and works much better than many facial recognition methods.

Today, surveillance systems are widely used for security purposes to monitor people in public places. A fully automated system is able to analyse information contained in the image or video through several image processing techniques

Face Detection is a technique for searching the entire face in an image or video. This phase represents a major step for most face recognition systems. The recognition stage serves to verify the identity of a person from an image or video by comparing the reference images with those acquired during the enrollment phase.

In this manuscript, we will present the techniques used for the detection of faces and the authentication. Next, we examine the learners' activities to detect frauds. While respecting the privacy of individuals.

2. Related work

In the literature, we can find studies that deal with remote exam management. (Pierre Beust, Valérie Cauchard, Isabelle Duchatelle) (First results of the experiment of remote monitoring tests) conducted a real-life remote monitoring experiment at the University of Caen Normandie.

This experiment allowed students in different countries to take this exam. The experiment was carried out as part of the training at (C.2.i.) Level 1 at the University of Caen Normandy. It is about a national certification to obtain the Computer and Internet Certificate.

The first results obtained during the experiment; 80 students volunteered on the University's platform, with more than 50% of candidates going so far as to set up an account with an organization that handles exam management. 31 participants completed the test as part of this experiment and most were under 25 years old, seven were absent and eight candidates could not pass the test due to technical problems.

3. Methodology

On the first hand, to authenticate users we start by capturing images using laptop camera, pre-processing tool have been used to enhance the input data. We use a Haar filter solution for face detection and then crop and normalize the faces. Next, we extract features using the convolutional neural network model to obtain a user model representing the signature of the user.

With the biometric model, learners can access based on the threshold used for the verification process. First, we compare the reference model with the signature acquired during authentication using a score distance. The idea is that if the score obtained is higher than the threshold the user is allowed to take the exam. However, if the score is below the threshold, the access will be refused.

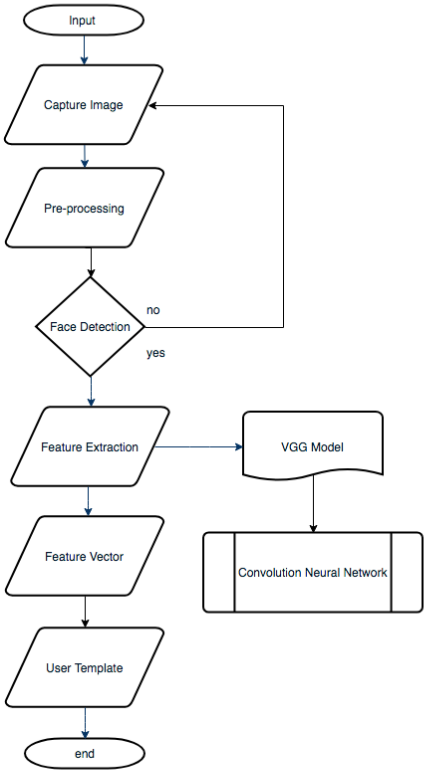

Figure.1: Feature extraction process for pattern recognition

3.1. Test Protocol (Preliminary test)

First of all, to obtain biometric data, the information is collected from the individual via an integrated camera of a "ThinkPad S440" computer. Next, pre-processing is applied to the model to improve the quality of data to be analysed.
Intel Core i7-4510U computer with a CPU of 2.00 and 2.60 [GHz] was used during the examination. The detection process is activated throughout the test by a user interface application developed by QT and C ++. The examination was performed under real conditions, the overall duration is 1 hour for each test, and we acquire an image every 3 seconds.

The first test consists in comparing a reference image with images acquired from a person authorized to take the examination (Figure.3), on the other hand the second test is to compare the reference image with a person is not authorized to pass examination (Figure.4).

3.2. Recognition process

Face recognition systems involve two important phases. The first step concerns the user enrollment: Enrollment means the acquisition of the biometric input data, the detection of region of interest and the extraction of features to define a template of each genuine use.

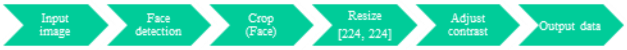

Figure.1: Pre-processing steps

The genuine model is stored as a reference after applying some quality requirements [7]. For face recognition based systems, the concerned biometric data is the face image, and the extracted features (Figure.1) are stored in the user device memory.

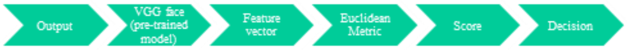

Figure.2: Steps for recognition

To perform a recognition test, we begin by applying the important VGG pre-training model to our reference image (Figure.2). Then, a feature vector of the input image is obtained, which represents the reference model or the unique signature of the learner.

In the second step called authentication, the system must decide if the user provide the expected pattern: the features extracted from the present pattern are compared to the stored features.

More specifically, the performance evaluation is based on the match score between the registered user and the test user, the learner can have access according to a decision threshold. Thus, in the verification system, two types of errors are defined. When a valid identity claim is rejected, a false miss error occurs. Conversely, when an imposter identity claim is accepted, a false acceptance error appears.

3.3. Data analysis

On the first hand, Euclidean distance was used to compute the similarity score between two images. The number of acquisition during the test is 720 images.

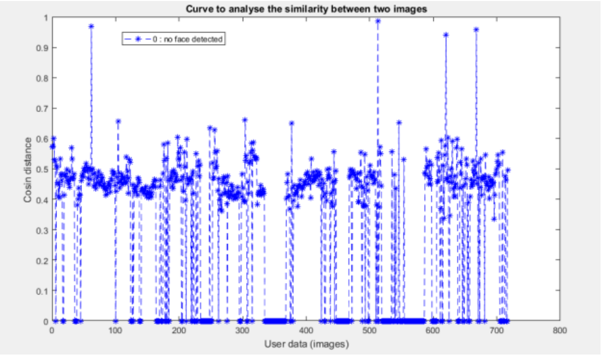

Figure.3 Curve represent user authorized to pass the exam

As we can see, each point represents the resemblance score after the comparison of two images, the first image being the reference image captured during the enrollment phase. This image is compared with images acquired during the exam session

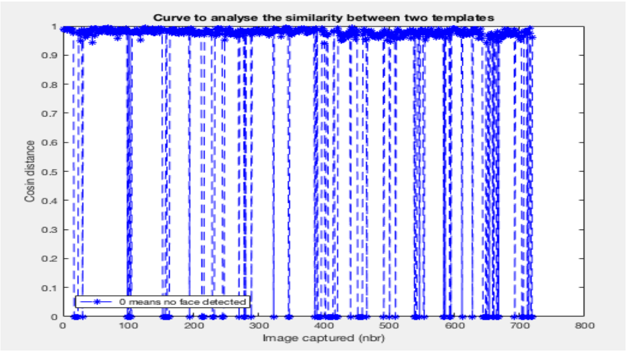

Figure.4 Curve represent user not authorized to pass the exam

The curve represents the cheat indicators of the preliminary results. Each value represents a point of resemblance between two images to detect cheating attempts.

The value zero means that the system does not detect any face, so if the distance is between 0.1 and 0.7 it is a person authorized to pass the test. If the distance is greater than 0.7 it means that the person is not allowed to take the exam.

We see that the value of the threshold is 0.7, when the amplitude of our signal is higher than the threshold at a given moment we consider that the user cheated.

To improve the threshold some machine learning tools have been used in order to create fraud detection model by setting the threshold value automatically.

4. Results

The idea here is to use and compare some learning method to detect the fraud automatically creating fraud detection model to manage online proctoring. The dataset for training represented mostly by normal cases. We begin the process by classifying input data by removing all zeros, then applying resampling to obtain user pattern.

In fact, 15 samples have been used to train the model, each sample is represented as one instance which is time series of continuous distance between reference pattern and new patterns acquired during the examination experiment.

Algorithm	AUC	CA	F1	Precision	Recall
Logistic Regression	0.889	0.5	0.444	0.792	0.5
Tree	0.944	0.929	0.929	0.940	0.929
SVM	1	1	1	1	1
kNN	1	0.714	0.645	0.802	0.714

Table.1: Table to evaluate the performance of some learning algorithm (cross-validation value = 3).

The idea here is to highlight several machine learning tools to create a model that will be used to detect fraud during an online exam. Here, some statistical tools such as "ROC curve" and "Area under curve" have been used to evaluate our biometric system. We can consider the system is ideal when the ratio value is equal to 1, more the value is lower more the performance decrease.

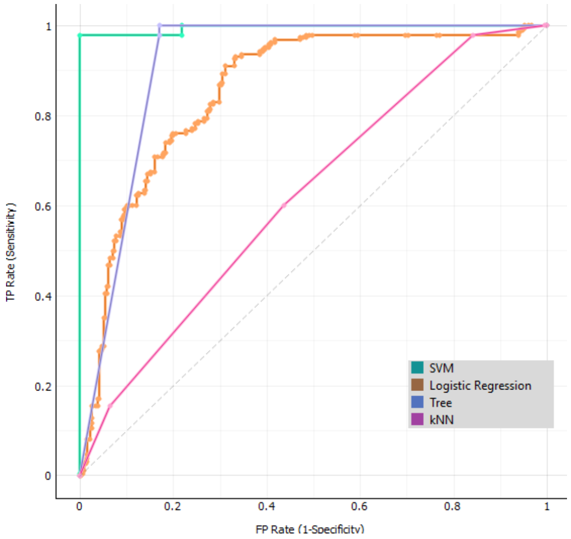

Figure.2: ROC curve to evaluate the performance of some learning classification tools.

To detect fraud, 15 samples were used to form our model. The target is a "normal case" or "fraud attempt". The first 10 vectors contain normal case values and the last 5 values of cheat attempts. After carrying out the experiment, as we can see SVM learning method represent the higher performance compared to the other methods

Many machine learning methods have been used to form our model. Logistic regression, kNN, decision of tree and SVM. It is obvious that the One-class SVM method is the best solution for classifying and detecting fraud in a remote exam. To conclude, we can integrate One-Class SVM model into our biometric system to detect an anomaly or create a fraud detection model to improve online assessment.

5. Future work

A biometric model is proposed to maintain the integrity and quality of the online assessment; at beginning, the system integrates face recognition modality. On the second hand, we integrate others modalities such as keystroke dynamics, speech detection and recognition, gaze tracking to improve system security and minimize cheating attempts. Also we will try to include other solutions to detect spoofing attack

6. Conclusion

The subject of this project is in the field of biometrics and the protection of the privacy of individuals and is particularly interested in the creation of an application with several features: authentication, monitoring of online learners, fraud detection and protection of personal data.

This work addresses the general problem of the protection of the privacy of a user during a remote examination. It seeks to define the different forms it can take fraud, as well as the expected properties of a secure biometric anti-cheat system that respects the privacy of learners.

Biometric technology is mainly used for security and smart applications. We should be careful about the drawbacks of ethical and social question raised by the use of biometrics. For one, we can imagine a situation where a person's biometric information would be stolen, this could have hazardous consequences with respect to the right to privacy.

Regarding privacy, it remains a major concern because we are collecting data not only about a person, but very important information that makes this person unique having unique behavioural or physical signature.

To conclude, the integration of biometrics into distance learning systems will help teachers to effectively control student authentication, course tracking, provide certificates in an automated manner, analyse student behaviour during exams, the validation of the certificates of success as well as the detection of fraud.

7. References

[1] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman. BMVC, cis.csuohio.edu, 2015.

[2] Jia Deng, Had Su, Jonathan Krause, Sanjeev Satheesh, - International Journal of computer vision, Springer – 2015.

[3] O Russakovsky, J Deng, H Su, J Krause “Imagenet large scale visual recognition challenge” - International journal of computer vision 2015 – Springer

[4] E Flior, K Kowalski “Continuous biometric user authentication in online examinations”- 2010 Seventh International Conference on Information Technology, 2010 - ieeexplore.ieee.org

[5] GB Huang, M Mattar, T Berg, E Learned-Miller - Workshop on faces in ‘Real-Life’ Images , 2008 - hal.inria.fr

[6] Y Benezeth, B Hemery, H Laurent, B Emile, C Rosenberger - Evaluation of Human Detection Algorithms in Image Sequences, 2010 – Springer

[7] Y Han, X Xu, Y Cai - Novel no-reference image blur metric based on block-based discrete cosine transform statistics - Optical Engineering, 2010 - spiedigitallibrary.org

[8] Y Benezeth, B Emile, H Laurent, C Rosenberger - Détection de la présence humaine et caractérisation de l'activité, 2009 - researchgate.net

[9] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), 2001, pp. 511–518

[10] Z Zhang, J Yan, S Liu, Z Lei, D Yi, SZ. Li - 2012 5th IAPR International Conference on Biometrics (ICB), 2012 - ieeexplore.ieee.org