# Online Invigilation: A Holistic Approach: Process for Automated Online Invigilation

Vaibhav Ahlawat
Computer Engineering Department
Indian Institute of Technology (BHU), Varanasi

Ahirnish Pareek
Computer Engineering Department
Indian Institute of Technology (BHU), Varanasi

S.K. Singh, Ph. D
Computer Engineering Department
Indian Institute of Technology (BHU), Varanasi

## ABSTRACT
Invigilation is an integral part of education and as education has evolved from conventional paper based methods to on-line ones, and so have the methods of invigilation. Major examinations are now online like TOEFL, GRE etc. But even with the assessment going online, invigilation still remains a manual affair; still officials have to be deployed on testing locations. Also in case of e-learning solutions the candidates are evaluated in their personal environment where there are no manual invigilators, thus a proper approach for online invigilation must be there. This paper aims to propose an invigilation model to automate the process and a tool for the same while taking into consideration the various constraints that come into picture for the specific scenario.

## Keywords
e-Invigilation, assessment, authentication, monitoring system, cheating.

## 1. INTRODUCTION
With the changes in technology, Education has changed as well, be it in terms of learning methods, teaching, assessment or invigilation. A major portion of educational content is going online which facilitates more penetration of the material i.e. easy availability, also freedom of choice for the learners as they can choose their time and material of study and can learn on their own pace.

Assessment is an important part of education. It encourages learning, provides feedback to the learner and the instructor, document competency and skill development, allows students to be graded, and allows benchmarks to be established for standards [1]. Assessments can also be done to relatively grade candidates for admission in universities or for jobs.

With the advent of e-learning, assessment has moved online too from conventional paper-pencil based methods but the process of invigilation is still mostly manual. In cases where manual invigilation is not possible, the use of unfair means is fairly easy thereby denting the very motive of assessment.

Invigilation involves both authentication and active or live proctoring. There has been work in both of these fields. Large portions of the work are done on the former i.e. monitoring the candidates for wrongful use of PCs which is a violation of the assessment policy (e.g. using web search to get results). Percival *et al.* proposed "The Virtual Invigilator", an approach that utilizes Intrusion Detection-type functionality to detect possible deviations away from standard procedure [2]. Other approaches, such as commercial offerings by Software Secure and Respondus have taken the approach of locking down what the browser and/or system is able to do during an assessment, thereby removing the opportunity for possible misuse [3, 4]. Yuan and Yang [5] have proposed a SIP based video surveillance system. These systems fail at identifying the authenticity of the candidate; hence do not result in automation of the system.

Software Secure has recognized the desire for remote proctoring of exams; however, their solution incorporates real time videoing of the candidate during the assessment. Whilst this does provide a level of authenticity, the real-time nature of the capture is storage and heavy bandwidth and the solution still requires a manual inspection by the academic to verify whether any problems exist. No level of automation exists within the process [6]. The solution given by N.L Clarke *et al.* aims at using biometrics for active authentication but does not include measures against spoofing like use of photograph to fool the facial recognition.

Among the various negative use-cases that exist for the scenario of online invigilation, our model tries to eliminate the following a) invalid authentication. b) Use of photograph to fool the authentication. c) Invalid use of web search and d) Use of external help to some extent. The authentication and the subsequent live proctoring uses facial recognition and techniques of liveness detection are applied so that the facial recognition system is not fooled by photographs of the candidate, judicious use of audio and video monitoring is used in case of suspicion so as to decrease the load on bandwidth (which can occur in case of continuous audio and video monitoring).

First a high level view of the model is discussed, then the procedure used for liveness detection and facial recognition are discussed in some more detail before arriving to results and conclusion.

## 2. MODEL FOR ONLINE INVIGILATION
In an online environment there can be various ways in which a candidate can try to use unfair means-

i. Imposter giving the examination on behalf of true applicant.

ii. Use of photograph to fool face verification system.

iii. Maintaining communication with other people that is unfair.

iv. Searching the web for right answers.

Proposed model tries to curb all of the above ways in which the candidates can fool the system. Fig. 1 shows the model along with the constituent stages.

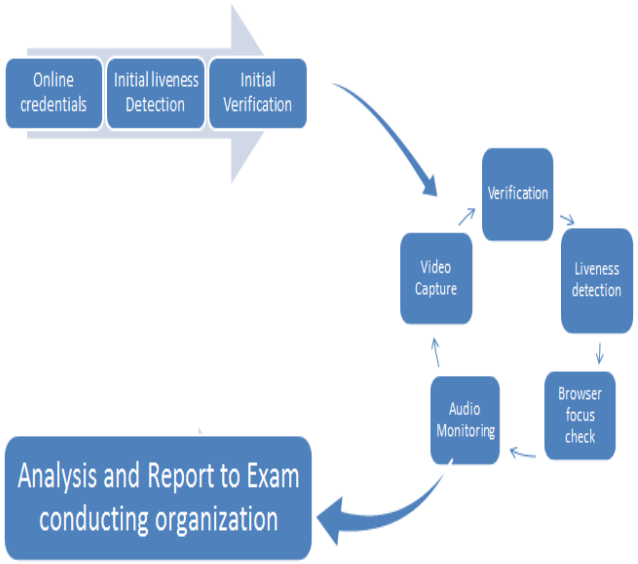


**Fig 1: Model for online invigilation**

## 2.1 Online Credential Verification

Username and password check is the most basic step of user authentication. This stage involves the username and password authentication of the candidate so as to establish the identity of the candidate and create a session for the same. In case the credentials are not correct, system prompts an error and does not move to further stages until the right credentials are entered.

## 2.2 Initial Liveness Detection

Photographs of the candidates can be used to pass the face verification step. So, liveness detection is performed to make sure the system is not compromised in this manner. Various approaches for liveness detection have been discussed in literature - [7] have used Eye blinking as a criterion to establish the liveness, [8] have used the fact that facial parts in real faces move differently than on photographs, Bao *et al.* have tried to detect the liveness using properties of skin [9]. But many of these methods don't work in the constrained scenario of online examination. Table 1 looks at the various methods for face liveness detection and evaluates them for our scenario of online examination.

Määttä *et al.* discussed an approach for liveness detection using Local binary patterns (LBP) [10]. We have tried to adopt this in our model as it is not computationally intensive and does not pose a bottleneck in network communication. The approach is discussed in detail in section 3.

## 2.3 Initial Face Verification

If the image of the candidate is already there in the database, then face recognition is done using the proposed hybrid approach discussed in section 3, otherwise the database is populated with the images of the candidate which are used for future references and for active authentication during rest of the examination. Also, if the camera detects more than one face then the system will raise an error, prompting for appropriate precautionary measures.

**Table 1. Comparison of liveness detection techniques**

| Approach | General Scenario | Our Scenario |
|---|---|---|
| Eye blinking | Good prediction rate | 1. Too many images will have to be sent.<br>2. Applicant might be solving question on paper. |
| Facial expression change | Not much optimized | -- do -- |
| Analysis of skin properties | work well for down-sampled photos | Likely to fail for higher-quality images. |
| Thermal Images | works well | No thermal camera available. |
| Depth Perception | works well | Hardware can be a constraint.(but possible) |

## 2.4 Active Authentication and Liveness Detection

Whenever the candidate changes a question, a photograph of him/her is clicked (as during question changing, he is most likely to be looking straight). Also photographs are taken at regular intervals of time and are processed for liveness detection and verification, failure of which alarms the system and puts it in suspicion mode. In suspicion mode audio and video verification, that are the 5$^{th}$ and 6$^{th}$ stages, are activated.

This stage also involves keeping a check on the usage of browser by the candidate. All modern browsers provide us with the functionality to check when a certain webpage goes out of focus, this is used to constraint the candidates to switch tabs only certain no. of times. It helps in reducing the use of web search.

## 2.5 Video Capture

When in suspicion mode, burst of photographs or a video is taken and stored in the database. If suspicion mode was triggered due to liveness detection failure then eye blinking is used to confirm the same, thus making the system more robust. The video can be analyzed after the exam manually in case of suspicion. Future work includes automating this process.

## 2.6 Audio Monitoring

The candidate can use the help of some other person during the exam, to curb this, audio monitoring is applied. First a sample of candidate's voice is made during registration and during test/exam, audio monitoring is activated after preset intervals of time or on lip movement of the candidate, which is then compared with the pre-recorded voice to check that the candidate is not using the help of some other person. Here we are concerned more about speaker recognition then speech recognition. This task of speaker recognition is done in two parts - Sampling and Verification. In Sampling, a voice sample is made of the person which we know is authentic and a number of features are extracted to form a voiceprint, template or model. Verification involves comparing the given voice with previously recorded voice prints and giving the closest match.

The speaker model based on Gaussian mixture model for text independent speaker identification by [11] works well for our purpose. This model attains 96.9% identification accuracy using 5 seconds clean speech utterances and 80.8% accuracy using 15 seconds telephone speech utterances with a 49 speaker population and is shown to outperform the other

speaker modeling techniques on an identical 16 speaker telephone speech task.

In case the voice does not match with that of the candidate, behavior is recorded in the system and an event is triggered that is pre-decided by the examiner - like a warning or ending of the session altogether.

# 3. DETAILS OF PROPOSED APPROACH

The proposed approach is a hybrid structure of three different algorithms for verification viz., Linear Discriminant Analysis, Linear Binary Pattern Histogram and Speeded-Up Robust Features (SURF), thereby taking into account spatial, texture and feature descriptor characteristics of query images. The reason for bringing three different approaches in a single module is to explore and analyse the query image for discriminative features in different domains so that our verification module would not depend on any single parameter for its operation and thereby limiting its capability and incurring accuracy loss in loosely constrained scenarios like remote invigilation. Verification step is preceded by liveness detection approach to verify whether it's a live image or an imposter image.

## 3.1 Liveness Detection

Määttä *et al.* explored some vital differences between face images and face prints in their work [10]. It says that face images and face prints reflect light in different ways due to difference in their surface texture properties. While a real face image is a 3d surface with unevenness, face print is a smooth 2d planar surface. This approach is based on texture analysis in the feature space. Local Binary Patterns, a powerful texture operator is used to describe the texture features and this feature set is then fed to an SVM classifier which determines whether the texture patterns characterize a live person or a fake image. More than one variant of LBP operator are used to sufficiently utilize differences in feature space.

## 3.2 Linear Discriminant Analysis

Keeping in mind the issues of variation in lighting condition and facial expressions, Belhumeur, Hespanha and Kriegmann proposed an approach that makes use of Fisher's Linear Discriminant, first developed by Robert Fisher in 1936 [12]. This approach assumes that all of the elements of a class i.e. face images of Lambertian surface lie in a 3D linear subspace (with some physical constraints) of the high dimensional image space. This observation becomes very useful in finding a linear projection of the faces from high dimensional image space to a considerably lower dimensional feature space. It produces a classification of images into distinct sets, called classes. The derivative of Fisher's Linear Discriminant then maximizes the ratio of between-class scatter to that of within-class scatter.

## 3.3 Linear Binary Pattern Histogram

Proposed by Ahonen *et al.*, this algorithm makes use of texture properties of face image to carry out the process of recognition [13]. The face image is first divided into small, local regions from which Local Binary Pattern features are extracted and concatenated into a single feature histogram. The motive behind using LBP features is that the whole face can be seen as if it is completely structured by micro-patterns which are invariant to grey-scale transformations.

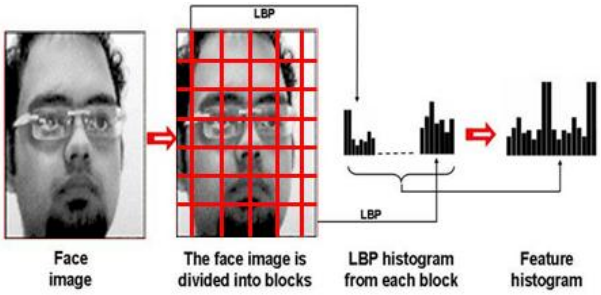


**Fig 2: LBP Overview**

The original LBP operator introduced by Ojala *et al.* labels the pixels of an image by thresholding the 3x3 neighbourhood of each pixel with the centre value [14]. If the intensity of the centre pixel is greater than or equal to its neighbour, then denote it with 1 and 0 if not. The resultant binary pattern is called Linear Binary Pattern.

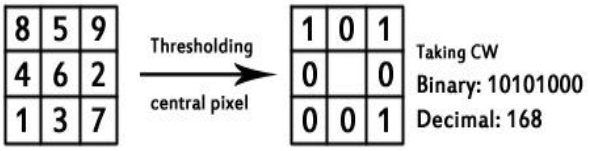


**Fig 3: LBP Encoding**

## 3.4 Speeded-Up Robust Feature

Suggested by Herbert Bay, SURF is a scale and in-plane rotation invariant detector and descriptor with better performance than SIFT [15]. Detectors in SURF are first used to find the interest points in an image, and then descriptors are used to extract the feature vectors at each interest point. Also, as compared to 128-dimensional SIFT; SURF has only 64 dimensions which reduces computational complexity to a great extent.

SURF uses the determinant of the approximate Hessian matrix and its local maxima applied to the scale-space are computed to select interest point candidates.

Figure 4.a shows an image and 4.b shows its interest points along with their strengths.

**Fig 4: a) and b)**

## 3.5 Hybrid Structure

Query image can be provided to the system in two ways – either clicking the image of the person right on the spot whose face is to be verified or providing a set of already stored query images as an input to the module. Face is detected in the query images with the help of Viola-Jones Face Detection Algorithm [16]. Ignoring the images containing multiple faces and no faces for the sake of simplicity, face is cropped from the query image.

### 3.5.1 Liveness Detection

Image is first normalized into a 64x64 pixel image after cropping. Then, $LBP_{8,1}^{u,2}$ is applied on the normalized face image and divide the resulting LBP face image into 3x3 overlapping region (with an overlapping size of 14 pixels). The local 59-bin histogram from each region are computed and collected into a single 531- bin histogram. Two other histograms from the whole face image using $LBP_{8,2}^{u,2}$ and $LBP_{16,2}^{u,2}$ are also computed, yielding 59-bin and 243-bin histogram that are added to the 531-bin histogram previously computed. A non-linear SVM classifier with radial basis function kernel is deployed to determine whether the query image is fake or not. More details can be found here [10]. If it's found fake, appropriate precautionary actions are taken. If found otherwise, verification step follows.

### 3.5.2 Verification

Firstly, Linear Discriminant Analysis is applied on the query image with the pre-computed threshold for identifying matching images and its predicted index is stored (= $index_{FF}$). Linear Binary Pattern Histogram is applied afterwards on the same query image and its predicted index is also stored (= $index_{LBPH}$). Saving predicted indexes from two different algorithms produces following observations based on their prediction capabilities: *a.)* If $index_{FF}$ = -1 and $index_{LBPH}$ = -1 (-1 signifies no matching), then we can stay assured that the query image is a non-matching image, *b.)* If $index_{FF}$ = X and $index_{LBPH}$ = -1, then SURF is applied on query image and image at $X^{th}$ index and result of SURF (=$flag_{SURF}$) is noted, and *c.)* If $index_{FF}$ = -1 and $index_{LBPH}$ = Y OR $index_{FF}$ = X and $index_{LBPH}$ = Y, then SURF is applied on query image and image at $Y^{th}$ index and $flag_{SURF}$ is noted. The reason for choosing $index_{LBP}$ in the case when both algorithms predicted their values is LBP's clearly better prediction capability as compared to LDA [12]. After applying SURF on the images, if $flag_{SURF}$ = 1, the indexed image is the correct match for the query image and it is not if $flag_{SURF}$ = -1. Flowchart of the verification module is described in Figure 5.

## 3.6 Database

For our approach, we used the publicly available NUAA Photograph Imposter Database [17] which contains images of real persons and face prints. The face images of real persons and their face prints were clicked in three different sessions with an interval of two weeks. Also, the physical and illumination conditions for each session are varying. In all, 15 subjects were invited to create the database. Traditional webcams with resolution 640x480 pixels are used to capture a series of subjects' face images. There are 500 images for each subject, clicked with the frame rate of 20fps. During the process, efforts were mainly directed towards making a live human look like a photo as much as possible. This was achieved to a great extent by asking each subject to look at the webcam frontally with minimum facial expression and movements such as eye-blink or head-movement.

The images of 15 subjects in the database are divided into two separate sets for training and test purposes. The training set consists of images from the first two sessions only. Remaining third session constitutes the testing set. The training set contains 1743 face images (889 and 854 from the first two sessions) and 1748 face prints (855 and 893). The test set has 3362 face images and 5761 face prints from the third session. All of the images are geometrically normalized into 64x64 pixels image.

**Fig. 6(a) Live Images**

**Fig. 6(b) Print Images**

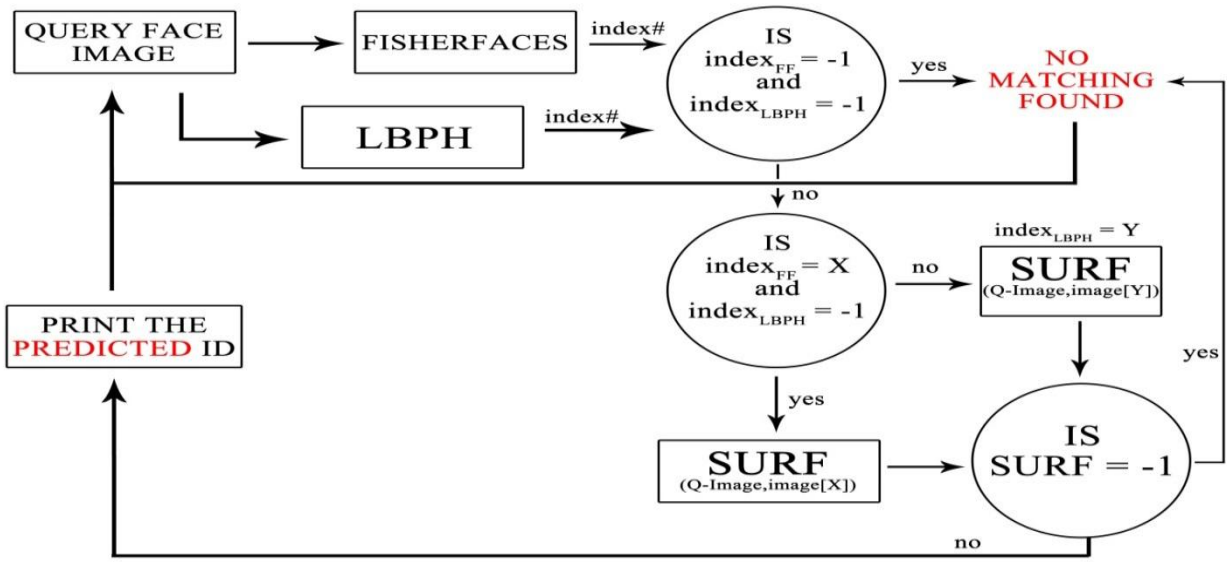


**Fig 5: Verification Module**

# 4. CONCLUSION AND FUTURE WORK

This paper has proposed a holistic approach to provide online invigilation of assessments and examinations. This aims at automating the process of invigilation while still maintaining the level of integrity that one would expect from a traditional assessment.

A simple prototype has been developed which takes into account liveness detection and facial recognition currently. Future work will include modifying the prototype fully in accordance to the proposed approach and evaluation of the software in a real-world assessment scenario.

For the sake of testing the proposed prototype, images of one of the authors' are integrated in the database so that full path of the prototype can be tested.

**Fig 7(a): Print images to test the prototype**

**Fig. 7(b): Snapshots of preliminary results**

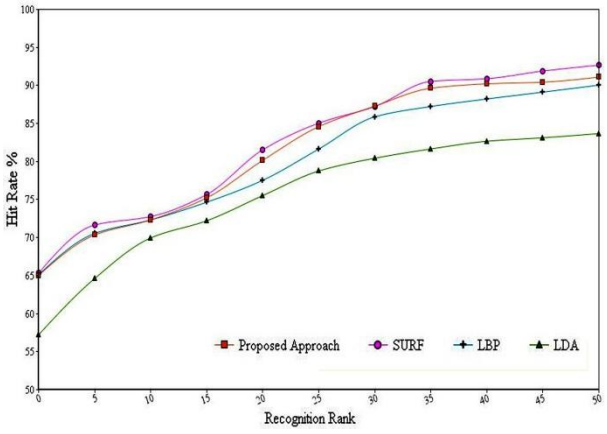


**Fig. 8: Cumulative Match Characteristics Curve**

As evident from the CMC curve in Fig.8, our proposed approach fares much better than LDA and comes close to SURF with an accuracy of 91.1% in the verification section. The liveness detection achieved around 1% false acceptance rate and 4.5% false rejection rate.

# 5. REFERENCES

[1] P. Broadfoot and P. Black. Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice*, Volume 11, Number 1, March 2004, pp. 7-26(20)

[2] N. Percival, J. Percival, C. Martins. The Virtual Invigilator: A Network-based Security System for Technology-Enhanced Assessments. In *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, October 22-24, 2008

[3] Software Secure. Remote Proctor. http://www.softwaresecure.com/solutions/remote-proctor.html

[4] Respondus. Respondus - Assessment Tools for Learning Systems. http://www.respondus.com/

[5] C. Yuan, Q. Yang. The Scheme of SIP-based Video Surveillance System. *Second International Workshop on Education Technology and Computer Science*, vol. 3, pp. 268-271, 2010.

[6] N.L Clarke, P. Dowland & S.M. Furnell. e-Invigilator: A Biometric-Based Supervision System for e-Assessments. *International Conference on Information Society (i-Society)*, 2013.

[7] G. Pan, Z.Wu, and L. Sun. Liveness detection for face recognition. In K. Delac, M. Grgic, and M. S. Bartlett, editors, *Recent Advances in Face Recognition*, page Chapter 9. INTECH, 2008.

[8] K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. *Image and Vision Computing*, 27:233–244, 2009.

[9] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In *2009 International Conference on Image Analysis and Signal Processing*, pages 233–236. IEEE, 2009.

[10] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *Proc. IJCB*, 2011, pp.1-7.

[11] Douglas A. Reynolds and Richard C. Rose. Robust Text-Independent Speaker Identification Using Gaussian Mixture speaker Models. *IEEE Transactions on Speech and Audio Processing* Vol-3, 1995

[12] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear Projection. In *IEEE transactions on pattern analysis and intelligence*, 19(7), (1997).

[13] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. In *IEEE Trans. Pattern Anal. Mach. Intell.*, 28:2037–2041, (2006).

[14] T. Ojala, M. Pietikäinen, D. Harwood, "A comparative study of texture measures with classification based on feature distributions." In *Pattern Recognition* 29 (1996) 51–59.

[15] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool. Speeded-up robust features (SURF). In *Comput. Vis. Image Underst.*, 110(3), 346-359 (2008).

[16] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*. (2001) 511 – 518.

[17] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Proceedings of the 11th European conference on Computer vision: Part VI*, ECCV'10, pages 504–517, Berlin, Heidelberg, 2010. Springer-Verlag.