International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

SEMI-AUTOMATED ONLINE EXAM PROCTORING

Akshat Jain[1], Divyam Sharma[2], Kunal Gupta[3], Harshita Mishra[4], Prof. Pushpendra Singh[5]

[1]Akshat Jain Information Technology & Inderprastha Engineering College

[2]Divyam Sharma Information Technology & Inderprastha Engineering College

[3]Kunal Gupta Information Technology & Inderprastha Engineering College

[4]Harshita Mishra Information Technology & Inderprastha Engineering College

[5]Prof. Pushpendra Singh Information Technology & Inderprastha Engineering College

---

Abstract - With the Advent of COVID- 19, remote learning has blossomed. Schools and Universities may have been shut down but they switched to applications like Microsoft Teams, Google Meet to finish their academic years. However, there has been no solution to examination proctoring. The capability to efficiently proctor remote online examinations is an important part to the measure the next stage in education. Presently, human proctoring is the most common approach of evaluation even in the online mode. However, such methods are labour-intensive and costly. In this project, we present a deep learning/AI- based system that performs automatic online exam proctoring. The system hardware includes a webcam and a microphone, to monitor the visual and acoustic environment of the testing location. The system includes six basic components that continuously estimate the key behaviour cues: face recognition, speech recognition, active window detection, gaze detection and phone detection. By combining the continuous estimation components, we generate a cheat score for the test-taker to classify whether the test taker is cheating at any moment during the exam.

Key Words : Object Detection, Deep Convolutional Neural Network

1.INTRODUCTION

Examination for students are a critical component of any educational program and online educational programs are no exception. In any competitive examination, there is a lots of chances of malicious activities and therefore its detection and prevention are very important. Monitoring every Student is very challenging task in terms of man power during Online Exam. Human face behavioral pattern play an important role in person identification. Video analytics can be used for a wide variety of applications to examine such behavior. Applications like Object detection, person identification, Gaze tracking, people counting and abnormal activity recognition etc.

- This system aims to identify the students who indulge in malpractice or suspicious activities during the online examination and alerts administration.
- Using such sensor , we propose to detect the following cheat behavior :-

1. Using phone call to friend.
2. Using the internet from the computer, smart-phone or switching tabs.
3. Asking a friend in the test room.
4. Having another person take exam other than test taker.

- A general framework for human behavior (suspicious and normal) analysis involves stages such as motion detection with the help of background modeling and foreground segmentation, object classification, motion tracking and activity recognition.
- We are using OpenCV for detection which based on DNN.

2.RELATED WORK

Facial detection based on LRF-ELM and CNN models of deep learning has been proposed on NUAA and CASIA spoof face detection databases. The author achieves higher accuracy and lower training time for LRF-ELM model on both the databases. They have suggested using various models for achieving higher accuracy of spoof face detection [1].

The paper [2] proposed a deep learning framework illustrating the representation ability of spatial and temporal information enabled deep network architecture for spoof face detection. The framework was used to further improve generalization efficiency by taking generalization as regularization via minimizing maximum mean discrepancy distance. The authors compare their framework on four different datasets and found good results. They have suggested using their framework to different problems in future for getting better results.

The problem of spoof detection as having same condition to capture the face image in training and testing give rise the phenomenon of lack of generalization. This problem was tried

to get a solution by proposing an unsupervised domain adaption framework by [3]. This framework uses handcrafted features as well as features learned through deep neural network. The authors claimed to create a new face spoofing dataset of 3000 images. The framework has the ability to transfer labelled face samples of source domain into unlabelled target domain. They have achieved 20% improvement over state of art frameworks in terms of accuracy.

The paper [4] proposes two-step CNN architecture for improving performance of spoof face detection. In the first step, the model is trained for local region of face image, learn deep features of it and in the second step, the whole model is trained on global level on real and fake face to improve generalization. The study was supported by the use of two standard datasets as Replay-Attack and CASIA FASD (Face Anti spoofing Dataset).Siamese Network based client information was used for liveness face detection model proposed by paper [5].The first step was used to perform face recognition and second step for liveness detection as reverse steps are used in most of research items. The authors suggested to use client identity information-based liveness detection in videos in the future.

Speech recognition systems are based on HMMs. These are models whose result is a sequence of symbols . Speech recognition uses HMMs because a speech signal can be viewed as a piecewise stationary signal.

In 1952, the Audrey system designed at Bell Laboratories was the first speech recognition system that recognized only digits spoken by a single person. After 10 years IBM produced the model in which recognized 16 English words. The first commercial speech recognition companies are Threshold Technology and Bell Laboratories that showed multiple person voice. A new statistical method called Hidden Markov Model (HMM) was introduced in 1980 which expanded to recognize a hundred words to several thousand words and to recognize an unlimited number of words.

Thiang, et al. (2011) presented speech recognition using Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) for controlling the movement of a mobile robot.

Input signals were taken for sampling from the microphone and then the extraction was done by LPC and ANN [6]. Ms Vimala. C and Dr V.Radha (2012) proposed a speaker independent isolated speech recognition system for the Tamil language. In this system Hidden Markov Model[HMM] was used for implementing acoustic model, feature extraction, language model and pronunciation dictionary which produced 88% of accuracy in 2500 words [7].

For large-vocabulary speech recognition systems ANN-Hidden Markov Model (ANNHMM) are used. Artificiel neural networks (ANNs) mathematical models of the low-level circuits in the human brain, to improve speech-recognition performance, through a model known as the ANN-Hidden Markov Model (ANNHMM).

For reaching higher Detection accuracy, developing speech compilation, low Word error rate, is depending upon the nature of language and addressing the problems of sources of variability through approaches like Convolute Non-Negative Matrix Factorization & Missing Data Techniques, are the major considerations for development of efficient speech recognition system.

In 2013 year Suma Swamy proposed an efficient speech recognition system which was experimented with Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), HMM which recognize the speech with 98% accuracy. For this the database consists of five words spoken by 4 speakers ten times [8].

In 2015, Google's speech recognition experimented with Connectionist Temporal Classification (CTC) trained Long Short-Term Memory (LSTM) approaches which are implemented in Google Voice. Google's English Voice Search system integrated 230 billion words from the actual user [9].

Object Detection [10][11] is modeled for classification problem where at all feasible locations we take slots of fixed sizes from the input object to feed these patches into an image classifier. Then it is fed to the classifier which determines the object's class in the window. Therefore, we know the category and location of the image objects [12].

Girshick were among the first to explore CNN for generic object detection and developed Region-based Convolutional Neural Networks(R-CNN) [11], which is inspired by the ground-breaking image classification results obtained by Convolutional Neural Networks (CNN) [13] and the success of selective search in the regional proposal for hand-crafted apps, which combines Alex Net with regional proposal system selective search. Since RCNN's proposal [11], many improved models have been proposed, including Fast R-CNN, which jointly optimizes classification and bounding box regression tasks, Faster RCNN [13], which allows an additional sub network to produce local proposals, and YOLO[10], which performs object detection by means of a fixed grid regression. All bring various degrees of improvements in recognition orderliness over the primary R-CNN and make object recognition more feasible accurate in real time [12].

Joseph Redmon gave a Unified, Real-Time Object Detection You Only Look Once[YOLO]. Their prior work is on detecting objects using a regression algorithm. Joseph Redmon has proposed a YOLO algorithm to get high accuracy and good predictions in this paper [14]. Juan Du gives the paper about Object Detection Based on CNN Family and YOLO. In this paper, they generally explained the object detection families like CNN, R-CNN and compared their efficiency and introduced the YOLO algorithm to increase the efficiency [15]. Matthew B. Blaschko gave the paper about Learning to Localize Objects with Structured Output

Regression. This paper is about Object Localization. In this, they used the Bounding box method for localization of the objects to overcome the drawbacks of the sliding window method [16].

Face recognition is one of the most important area in computer vision. It helps in security, to identify of someone, surveillance system etc. Eyes, mouth and ears are the main features of the face. So for detection of them face recognition is very important. Eye tracking is most used feature of face detection. There are two approaches for eye tracking. First is Geometrical model and second is machine learning model.

In context of Geometrical model approach Timm and Barth [18] proposed an approach in which they are using image gradients to locate eye pupil. They derive a function which consist squared dot product. The maximum of this function according to the location is responsible to find the eye pupil.

In context of machine learning model, the traditional feature extraction and cascaded classifier is used. Chen and Liu [19] proposed a paper in regarding the traditional feature extraction where they were using Discriminatory Haar Features(DHF) and support vector machine(SVM) for eye detection.

Sharma and Savakis [20] proposed a paper in which histogram of oriented gradients (HOG) based technique is used with the combination of support vector machine for eye tracking.

3. METHODOLOGY

Audio Data:-

- For Speech recognition we will use Google's speech recognition API to convert speech to text.
- We use the NLTK (Natural Language Toolkit) library to perform the natural language processing.

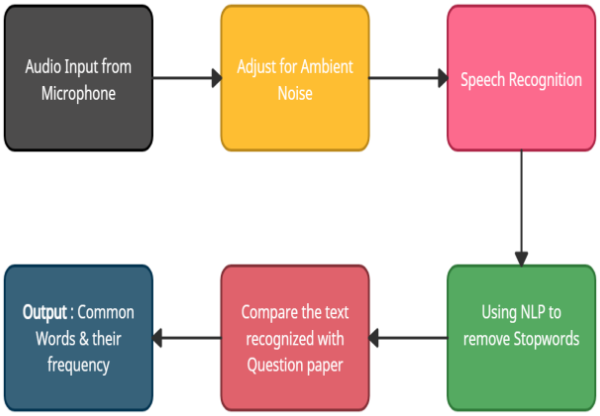
Visual Data:-

- Face Detection using YOLOv3 and OpenCV libraries in python to detect the number of faces present in the video.
- Mobile Phone Detection we will use the pre trained YOLOv3 model

Gaze Tracking:-

- Face Detection using OpenCV libraries in python to detect the faces from the input.
- Use pretrained 68 facial keypoint model to find the landmarks of the eyes, mouth and nose.

IMPLEMENTATION OF AUDIO MODULE :-



VIDEO PROCESSING MODULE :-

- In this section we are doing person counting and mobile phone detection that can be used formal practice or cheating during the exam.
- We are using YOLOv3 for this surrounding detection.
- Our system divides the input image into an S*S grid. If the center of an object comes into a S*S grid cell, then that grid cell is responsible for detecting that object.
- Each grid cell predicts Bounding boxes and confidence scores for those boxes.
- These confidence scores shows how confident the model is for the box that contains an object and also shows how accurate it thinks the box is that it predicts.

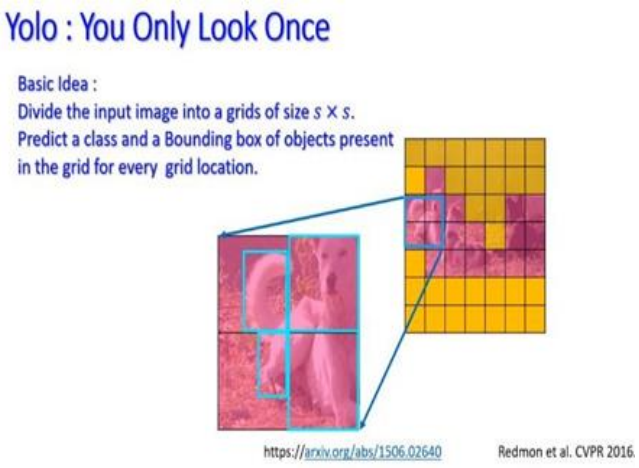


Fig 1 – Working of YOLO

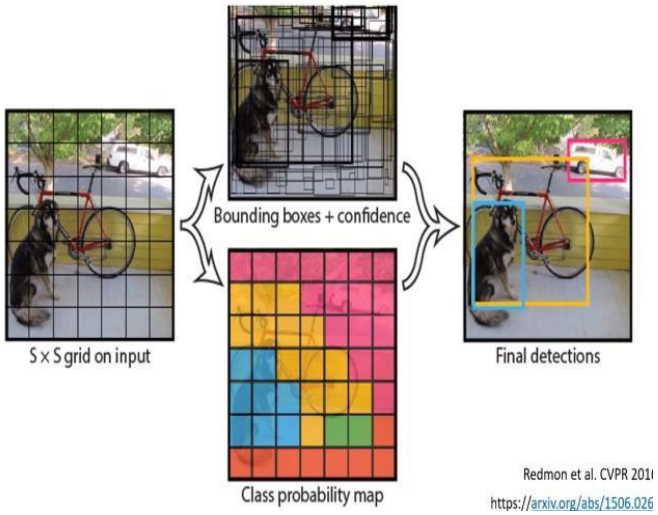
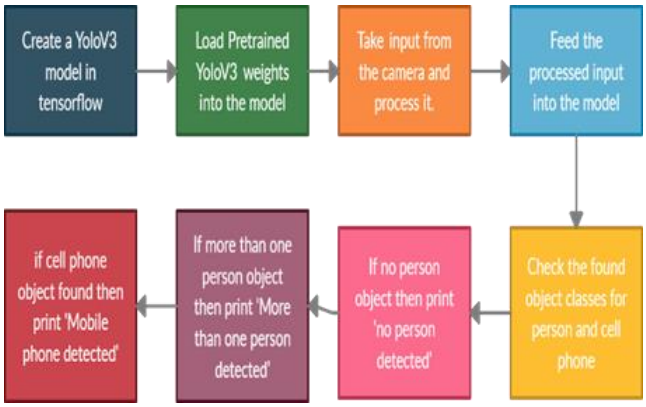


Fig 2 – Determine object belong to which class in YOLO

IMPLEMENTATION OF VIDEO MODULE :-

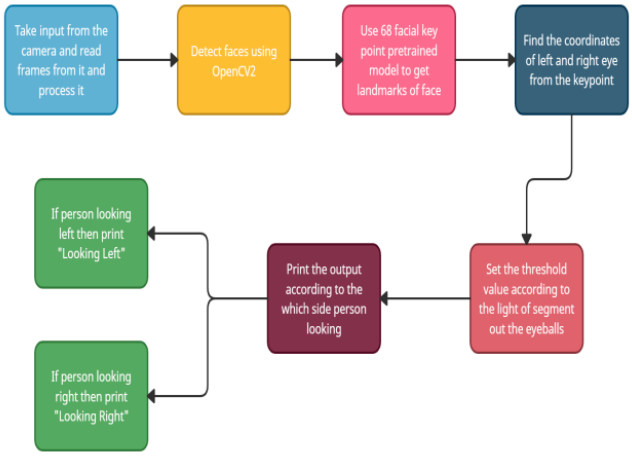


GAZE TRACKING MODULE :-

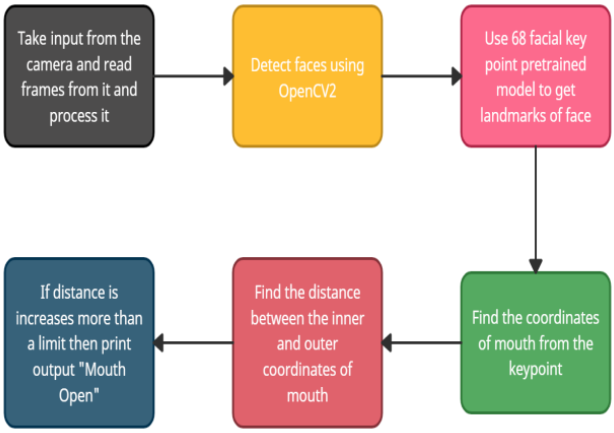
- In this module we are trying to track the eyes and mouth to prevent malpractices during the online exams.
- For this, First we use detect the face using OpenCV.
- Use Pretrained 68 facial keypoint model to find the landmarks of eyes and mouth from the face.
- After this set the threshold according to the light to segment out the eyeballs.
- After this module gives output according to the which side person is looking.
- For mouth, we find the distance between inner and outer part of the mouth.
- If distance is increases more than a limit then module return mouth is open.

IMPLEMENTATION OF GAZE TRACKING MODULE

For eye tracking:-



For mouth tracking :-



4.RESULT AND DISCUSSION

We developed a speech recognition module, mobile phone detection and person counting module and front-end in which teacher and student upload question paper and answers. In the speech recognition module, we used some new technologies like NLTK, PyAudio to take audio input from the user microphone and speech_recognition for converting audio into text. While in Mobile phone detection and person counting, we used YOLOv3, in this first we created the YOLOv3 model in TensorFlow then load YOLOv3 weights into the model, take input from the camera and the input then processed into the model and give the output. In the front-end we used javascript, bootstrap to create the page. Here some screenshots of our working modules.

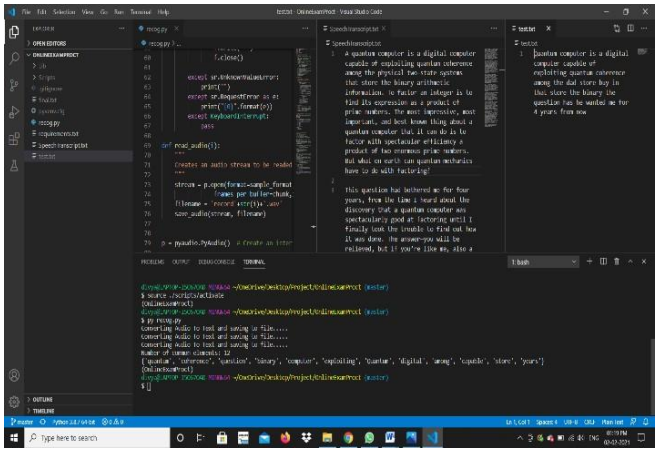


Fig 3 -Working of Speech Recognition

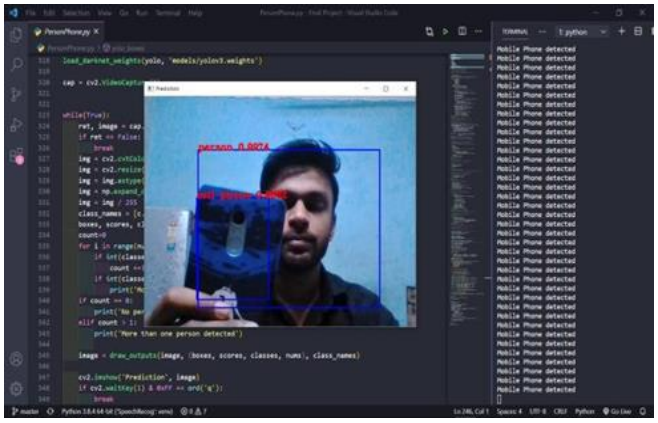


Fig 4 - When the person tries to cheat using a mobile phone

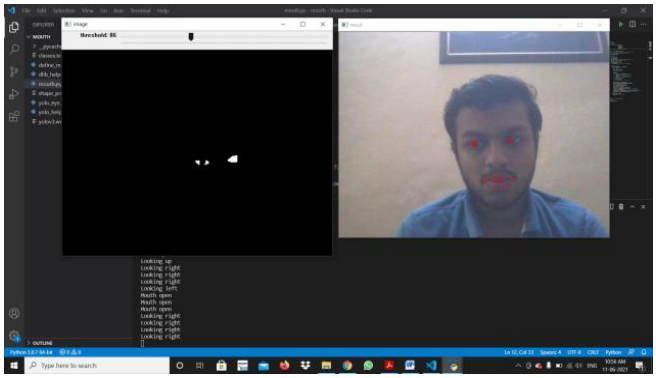


Fig 5 - Working of Gaze tracking

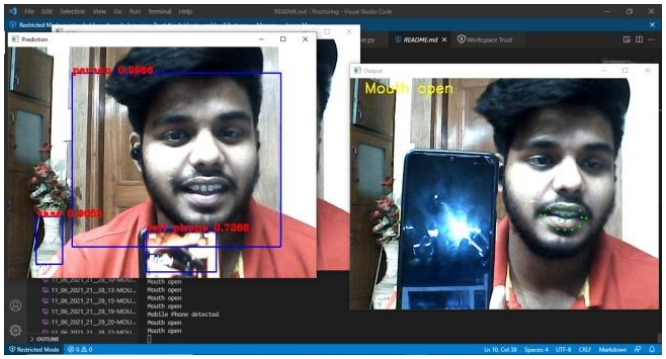


Fig 6 - Complete working of project

5.CONCLUSION

We completed our speech recognition module which responsible for detecting or recognizing the speech during exams, and also completed mobile phone detection and person counting module which is responsible to detect mobile phones and also for detection of more than one person during the exams. We also developed Gaze tracking module which contain eye tracking and mouth tracking. In eye tracking we find the eyeballs and from this we can find is student or person trying to do cheating by looking through eyes and in mouth tracking we can find that is person trying to cheat through speak or not during the examination. By this we can prevent cheating or malicious activity during examination.

REFERENCES:-

1. Akbulut, Y., Sengur, A., Budak, U., & Ekici, S. (2017). *Deep learning based face liveness detection in videos. 2017 International Artificial Intelligence and Data Processing Symposium (IDAP).* doi:10.1109/idap.2017.8090202
2. Li, H., He, P., Wang, S., Rocha, A., Jiang, X., & Kot, A. C. (2018). *Learning Generalized Deep Feature Representation for Face Anti-Spoofing. IEEE Transactions on Information Forensics and Security, 13(10),2639–2652.*doi:10.1109/tifs.2018.2825949.
3. Li, H., Li, W., Cao, H., Wang, S., Huang,F., & Kot, A. C. (2018). *Unsupervised Domain Adaptation for Face Anti-Spoofing. IEEE Transactions on Information Forensics and Security, 13(7), 1794–1809.* doi:10.1109/tifs.2018.2801312
4. Gustravo Botelho de Souza et.al., “On the learning of deep local features of robust face spoofing detection”,
5. Huiling Hao and Mingrao Pi, “Face liveness detectin based on client identity using Siamese Network”,
6. Thiang and Suryo Wijoyo, “Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot”, in Proceedings of International Conference on Information and Electronics Engineering (IPCSIT), Singapore, IACSIT Press, Vol.6, 2011, pp.179-183
7. Ms.Vimala.C and Dr.V.Radha, “Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM”, in Proceedings International Conference on Communication Technology and System Design 2011, Procedia Engineering 30 ISSN: 1877-7058, 13March2012, pp.1097 – 1102
8. Suma Swamy,K.V Ramakrishnan,“An Efficient Speech Recognition System”, Computer Science & Engineering:InternationalJournal(CSEIJ),V ol.3,No.4,DOI:10.512 1/cseij.2013.3403 August 2013, pp.21-27
9. Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays and Johan Schalkwyk (September 2015): Google voice search: faster and more accurate.
10. D J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.
11. A. Mekonnen and F. Lerasle, "Comparative Evaluations of Selected Tracking-byDetection Approaches," IEEE

Transactions on Circuits and Systems for Video Technology, vol. 29, no. 4, pp. 996-1010, 2019.
12. S. Shinde, A. Kothari and V. Gupta, "YOLO based Human Action Recognition and Localization," Procedia Computer Science, vol. 133, pp. 831-838, 2018.
13. L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," International Journal of Computer Vision, 2019
14. Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection",The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
15. OLO Juan Du1,"Understanding of Object Detection Based on CNN Family",New Research, and Development Center of Hisense, Qingdao 266071, China.
16. Matthew B. Blaschko Christoph H. Lampert, "Learning to Localize Objects with Structured Output Regression", Published in Computer Vision – ECCV 2008 pp 2- 15.
17. Deep learning approach to peripheral leukocyte recognition - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/YOLO v3-architecture-A-YOLOv3-pipeline-with- input-image-size-416416-and-3- typesof_fig4_334021766 [accessed 16 Mar, 2021]
18. F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," *Visapp11*, pp. 125–130, 2011.
19. S. Chen and C. Liu, "Eye detection using discriminatory Haar features and a new efficient SVM," *Image and Vision Computing*, vol. 33, pp. 68–77, 2015.
20. R. Sharma and A. Savakis, "Lean histogram of oriented gradients features for effective eye detection,"*Journal of Electronic Imaging*, vol. 24, no. 6, Article ID 063007, 2015.