Detection of Attentiveness from Periocular Information

MOULI LAHA

National Informatics Centre, Ministry of Electronics & IT, Government of India

---

ABSTRACT - *In this paper, an approach of predicting attentiveness of a person has been presented, using periocular region. One obvious approach of predicting the attentiveness is by training the system using a learning algorithm. This paper however presents an alternative novel approach in which a system is capable of segregating and detecting the periocular region with ability to analyze the attentiveness of the subject. The approach has also been developed with an intention of keeping the computational cost and storage requirement at a minimal level. Applications of the system include various fields where attentiveness becomes an important factor such as lecture hall or during driving. The proposed system along with its potential applications has been presented in this paper which also uses a number of commonly used vision-based tools and techniques.*

Keywords: Image processing; computer vision; attentiveness; periocular region; face recognition.

1. INTRODUCTION

Detection and recognition of objects have occupied an important place in the research activities of computer vision. Amongst all kinds of object detection, face recognition has taken the supremacy at a fast pace. Since facial region comprises of various unique and dominant features; performing detection, recognition and tracking based on facial features becomes a priority for various application purposes. As such, image of human face can be divided into many regions which are capable of providing various specific that will remain very close in most of the faces. Areas such as eyes, nose, lips constitute patterns and features as shown in figure 1.

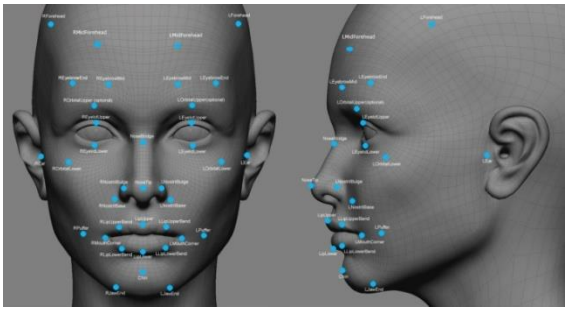


Fig -1: Features enclosed by the facial area.

Thus if a facial region is divided into blocks, features are extracted from them and a classification model may be formed based on an appropriate algorithm. In case of a learning algorithm such classification model will be capable of detecting faces in any captured frame with the aim to explore the region around the eyes called the periocular region in particular. Periocular region encloses various unique and distinguishable features and thus is capable of providing various important information regarding the face image under study, viz. gender, age, and ethnicity, for example. A number of such studies have been reported in the literature. A few notable amongst this are further studied in detail.

Periocular region can be considered as an alternative to the iris recognition system in many instances of biometrics [1]. There are several ways of detecting the periocular region. One such method of detection is by training a classifier by providing the periocular region as the ROI manually in the training dataset.

The study of the periocular region was reported to be successful in predicting age [2], gender [3] and ethnicity [4]. The approach provided in this research work reports a method of detecting the periocular region together with the capability of identification of attentiveness of the face under study.

In general attentiveness plays an important role in various fields of day to day activities. Though the study of attentiveness comes under the field of psychology, but in this research work a structure has been proposed along with the algorithm to grab a scene, detect a face and providing a warning tune if the person is not attentive. From a school classroom to driving in the highways, attentiveness is a key element to measure the mindfulness of the subject. Whether a student or an audience is paying attention towards the lecture being delivered or a driver in a moving car is feeling drowsy; examining the periocular region may give an approximate clue of the same. This paper presents an approach of predicting the attentiveness and also the efficiency as a metric for future application.

1.1 Why the face is detected first followed by periocular region?

The very common question which may arise is why the periocular region is not detected directly using a classifier model. The main scope of this research is to concentrate the

study of periocular region directly, with all of its variability. Creating a classifier model incorporating all kinds of variations in the periocular region and detecting them efficiently is a costly operation, computationally as there is a requirement of a huge number of training set to train the classifier. However developing a classifier model for detecting faces is comparatively easy since various notable features and patterns are enclosed by the facial area. Thus in the proposed approach of detecting and segregating out the periocular region uses firstly the detection of face then concentrate on an appropriate region of face suitable for periocular study.

1.2 How the features are extracted?

One of the important terms here is the feature. Features constitute the base of any image processing experiment. Any specific and well-defined entity within an image or any sensor data which can be detected and recognized repeatedly and consistently is known as feature. A feature can be a point, a line, a corner or an edge. Detailed study about features can be found in [5]. There are various feature descriptors available which helps in extracting the features from an image. Few of the standard feature descriptors are SIFT [6], SURF [7], HOG [8], and LBP [9] etc. For this work, HOG descriptor has been used.

1.3 Histogram of Oriented Gradients

The idea behind hog descriptor is that it works based on the distribution of intensity gradients or edge directions which affects the local appearance of a subject. HOG descriptor was first described by Navneet Dalal and Bill Triggs [8], researchers for the French National Institute for Research in Computer Science and Automation (INRIA) [ref]. They explored the HOG descriptor for detecting human using SVM classifier.

1.4 How the training takes place?

Next term to be explained is the learning algorithm. Learning can be supervised or unsupervised. In this case supervised learning has been incorporated. Supervised learning [10] is basically a process of inferring a function from a set of labeled training data. Supervised learning entails learning a mapping between a set of input variables X and an output variable Y and applying this mapping to predict the outputs for unseen data.

In the supervised learning paradigm, the goal is to infer a function f : X → Y , the classifier, from a sample data or training set $A_n$composed of pairs of (input, output) points, $x_i$belonging to some feature set X , and y_i∈ Y :A_n = $((x_1,y_1),...,(x_n,y_n)) \in (X \times Y)^n$

Typically X ⊂$\mathbb{R}^d$ and y_i∈ IR for regression problems, and y_i is discrete for classification problems. We will often use examples with y_i∈ {−1, +1} for binary classification.

There exist many learning algorithms among which SVM has been considered for this work.

1.5 Support Vector Machine [11]

SVMs are supervised learning model which are associated with classification and regression analysis of data. SVM classifies data by finding the best hyperplane, separating all data points of one class from those of the other class, shown in figure 2. It is assumed that both the classes are linearly separable.

Let x_i (i∈[1,N], $x_i \in \mathbb{R}^P$) be the feature vectors representing the training data and y_i (i∈[1,N], y_i∈ {−1,1}) be their respective class labels. We can define a hyperplane by <w, x> +b = 0 where w ∈$\mathbb{R}^P$ and b ∈ $\mathbb{R}$. Since the classes are linearly separable, we can find a function f, f(x) = +b with y_i f (x_i) = y_i (< w,x_i> +b) > 0, ∀i ∈ [1,N].

The decision function may be expressed as:

f_d (x) = sign(<w, x> +b) with f_d (x_i) = sign(y_i), ∀i ∈ [1,N].

Support vectors are training data points that are close to the separating hyperplane. To prove that they are enough to compute the separating hyperplane is a convex optimization problem. The dual formulation is favoured for its easy solution with standard techniques. Using Lagrange multipliers, the problem may be re-expressed in the equivalent maximization on α (dual form):

$$\alpha^* = \underset{\alpha}{argmax} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j <x_i, x_j>$$

Such that $\sum_{i=1}^{N} \alpha_i y_i$=0 and ∀i ∈ [1, N] $\alpha_i \geq 0$.

The hyperplane decision function can be written as

f_d (x) = sign ($\sum_{i=1}^{N} \alpha_i^* y_i$<x, x_i> +b)

The training takes place for 50 stages making the classifier more robust.

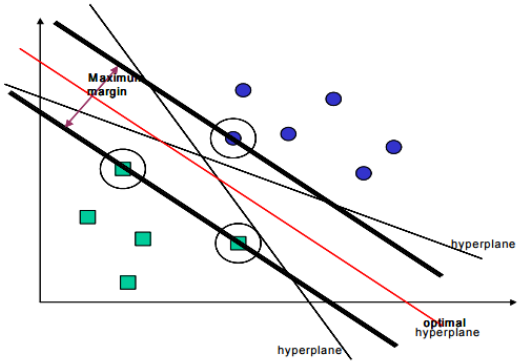


Fig -2: Graph depicting hyperplane in SVM.

The rest of the paper is arranged in the following manner: section 2 throws some light towards the related works which have already been performed in this field. Section 3 explains the approach followed by the validation and accuracy of the approach in section 4. Finally section 5 concludes the paper mentioning few of the future scopes of the process.

2. RELATED WORKS

Two major research areas that have been mentioned in this paper, viz. periocular region and predicting attentiveness, need some background survey.

Many works have already been done for identifying and detecting the periocular region [12] [13] [14] [15] [16]. Majority of such approaches include creating a classifier by providing a number of training samples consisting of the periocular region as the positive training images.

Various studies have been performed on predicting drowsiness from the eye contour [ref]. The yawning has also been detected from contouring the mouth area [ref]. But most of these studies incorporate the opening and closing pattern of the eyes and assume the state of drowsiness if the eyes are closed for more than some specified seconds, say 4-5seconds.

But the issue covered in this paper is not only drowsiness, but also attentiveness. A person may be inattentive even when he/she is active or not feeling sleepy. Few approaches to predict human attention are presented in this section though none of the works is based on the attention prediction with the help of the periocular region.

Kumar et al. [17] have analyzed how attention, identification and recognition of faces in a crowd can be carried out by segmentation and their relative visual saliency. They have carried out the experiment with gray scale images, considering intensity as the fundamental feature. They have obtained attentiveness of the faces through saliency score and measured the effect of segmentation on recognition and identification.

Zhong et al. [18] have proposed an attempt to develop an attention model for face recognition using deep learning technique. The attention model is based on bilinear deep belief network and the experiment is carried out on CMU PI and BioID face datasets.

Smith et al. [19] have also proposed a system to analyze human driver visual attention. The system operates with one camera and relies on the estimation of global motion along with color statistics to track the head and facial features of a person robustly. The system has made use of eye blinking and closure features in addition to the yawning which depicted large mouth motion.

3. PROPOSED APPROACH

Basically, the approach is to detect the periocular region from an initially detected faces.

3.1 Classification Procedure

3.1.1 How to detect faces?

In the initial step, a set of positive and negative images are collected or captured. Any image consisting of a visible face can be considered to be a positive image, whereas any image without a face acts as a negative image.

Second step is to specify the ROI within the positive images. For this paper, the ROI is the facial region around which a box is drawn. Basically, the starting coordinate of the box bounding the face and its height and width are stored while specifying the ROIs. These parameters are later used during the feature extraction process.

The notable features are extracted from the ROI region using HOG descriptor, which are fed to the training model. After feature extraction, SVM learning algorithm is applied to train the data for classification.

Upon completion of the training, a pattern gets generated which helps in the classification process.

3.1.2 How to detect periocular region from detected faces?

Since the face gets detected at the initial step, a box bounds the facial region. The periocular region exists in the upper portion of the face. If the height of the bounding box is reduced to half, the upper half will give the periocular region.

Let x, y, h, w be the four parameters which defines the bounding box. (x, y) denotes the starting corner coordinate. 'h' is the height of the box and 'w' is the width of the box.

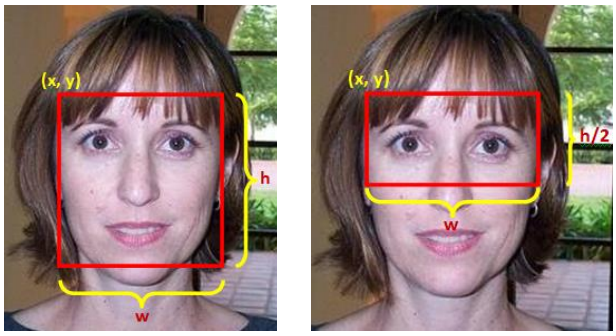
Keeping the starting position and width intact, the height of the box is reduced to half.

Box = f(x, y, h, w);

Box_half = Box.height/2;

where f is any specified function.

Thus, the upper part of the face consisting of the periocular region gets selected. The next set of operations will be carried out particularly in this selected periocular region. To make the process easy, the selected portion may be cropped out which however is not mandatory. It is assumed that the selected portion will be consisting of mainly two colors: black and skin color. The part of the visible hair, eyebrows and eyeball will contribute to the black color and the rest will be the skin color. No other color can possibly be involved as depicted in figure 3.



(a) **(b)**

Fig -3: (a) Box bounding the face. (b) Box bounding the periocular region which also shows possible colors and the validity which will be present in the image.

The basic operations that will be required are thresholding, segmentation and binarization using various morphological operators. Thresholding (shown in figure 4) refers to selecting a low and a high pixel intensity value or in some cases only a single value based on which segmentation can be done. The intensity of the selected threshold area is increased (or decreased) to a certain level, keeping rest of the intensities constant. The binarization operation is then carried out. This results in filling the selected region with 1's and the region outside the threshold area with 0's.

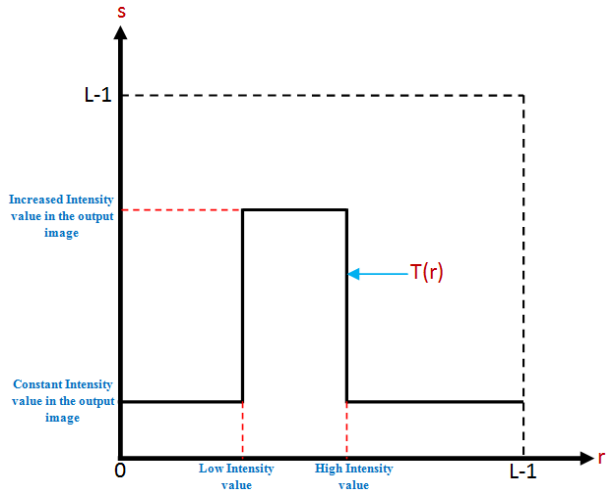


Fig -4: Thresholding.

In figure 4, r is the input intensity value whose range is [0, L-1] where $L = 2^n$, n being the number of bits each pixel represents. s is the output intensity value and T(r) is the transformation function where.

s=T(r)

The area of interest is the eyeballs. Thresholding is done in such a way that only the eyeballs are selected. In practice sole selection of eyeballs through thresholding is not possible. If the original image is converted to gray scale and a histogram is made, eyeballs along with the hair and eyebrows constitute the near black region. Thus, if a range is chosen such that the near black regions fall under the range and rest of the skin colored part remains outside the mentioned range. Now if this thresholded part is binarized, the near black region gets filled with 1's and the rest of the region becomes 0 i.e. black.

For input intensity values 0 to 255 in the histogram as shown in fugure 5, the range considered for this experiment is [0, 30] i.e. r_0 to r_{30} value is converted to 0 and r_{31} to r_{255} are converted to 1. Therefore s_0=s_1=…=s_{30}=0 and s_{31}=s_{32}=…=s_{255}=255.

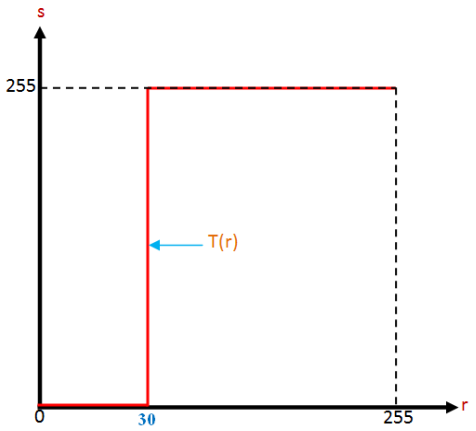


Fig -5: Binarization.

Next part is to detect the eyeballs from all other blackish regions. The segregation of the eyeballs from the frame is done using the following morphological operations:

- Firstly the boundaries of the eyeball regions are extracted by erosion followed by subtraction.
  $\beta(A) = A - (A \ominus B)$
- Second task is to fill the hole within the boundary, shown in figure 6. The equation for region filling is:
  $p_k = (p_{k-1} \oplus B) \cap A^C$ where p_0 is the starting point inside the boundary and k=1,2,3…the
  p_k contains all the filled holes and $p_k \cup A$ gives the filled holes along with their boundary.
- The final task is to extract all the connected components in an array form starting with p_0. The equation for the same is given by:
  $p_k = (p_{k-1} \oplus B) \cap A$

Polygonal approximation is done next to capture the essence of the boundary shape into a polygon using the least square error line fit merging technique.

Next a filtering operation is done to remove unwanted regions i.e. the eyebrows and hairs. A threshold value for number of pixels is provided. If the connected component consists of pixels more than the mentioned value, then those components are preserved, discarding the rest.

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	1	1	1	0	0
4	0	0	1	0	1	0	0
5	0	0	1	1	1	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

⇨

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	1	1	1	0	0
4	0	0	1	1	1	0	0
5	0	0	1	1	1	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

Fig -6: Hole filling of a connected component.

Next step involves computing the perimeter and area of the connected components. Perimeter is basically the sum of the boundary pixels of the connected components. After extracting the boundary of the existing connected components and storing the values in an array, the number of boundary pixels of each connected component are computed, which gives the perimeter. The total number of pixels constituting the boundary along with the pixels within it together gives the area of the polygon.

The main aim is to detect the eyeballs, which are near circular objects. To predict the level of circularity, a relation between the area and perimeter needs to be established such that a fully circular connected object will be denoted by 1.

Now perimeter of a circle = $2\pi r$ where r is the radius of the circle.

Area of a circle = πr^2

Circle = $\frac{4 \times \pi \times Area}{(Perimeter)^2}$

The range of the value of 'circle' will be within 0 and 1. In this work, a threshold of 0.60 has been chosen i.e. if 0.60<=circle<=1, the centroid of such circular objects are further computed. If in a single periocular region there exists 2 connected components with centroid, then the next step is to check whether the centroids are in the middle portion of each half of the periocular region or not. For that the periocular region needs to be divided into further two halves, width-wise this time.

Box_quad = Box_half.width/2;

The updated height and width of each box will be h/2 and w/2. There will be two boxes say Box_quad1 and Box_quad2. The midpoints of the two boxes are computed and a range from the midpoint is specified. For this paper, the range is [0, 0.25*(h/2)]. If the eyeballs are found to be located within the specified area, the subject is said to be attentive. From the figure 8, mid coordinates are (w/4, h/2) and ((w/4+w/2), h/2). The area under consideration is the area within the red circular part whose radius is in the range [0, 0.25*(h/2)]. For the person to be attentive, his/her centroids of the eyeballs must lie within the specified corresponding circular area.

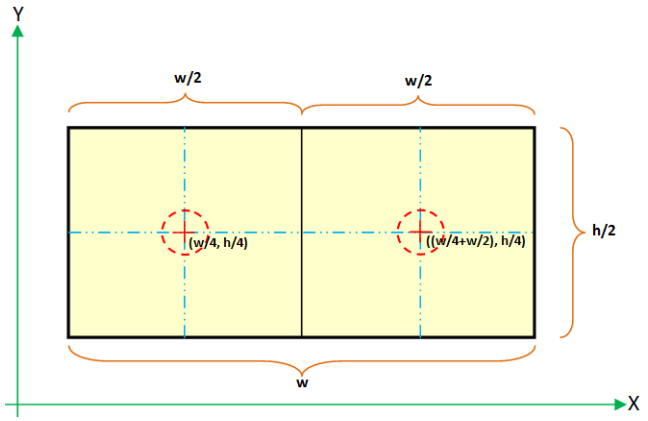


Fig -8: Within the periocular region, the acceptable location of the eyeballs (centroid) is within the red circular area.

For a real time data a video file is considered to be the input. The video is segregated into number of frames. The above mentioned algorithm is applied over each frame. Since a subject will blink, the value of circle will change accordingly for few frames alternatively in a specific pattern. If the value of circle does not get back to >=0.60 for say more than 5 seconds i.e. 125 frames, the subject is considered to be inattentive. Some special cases of inattentiveness have been shown in figure 13.

4. EXPERIMENTAL SETUP

The preliminary hardware setup that incorporated the above proposed approach consists of an external camera, a display unit and a motherboard. The camera grabs the live video of a scene which is fed as an input to the system. Video is nothing but a collection of frames. The above-mentioned algorithm is applied over each frame.

Various pre-recorded audio of warning messages are stored within the system. These messages are played according to the requirement. For instance, if a person is feeling sleepy then no centroids will get detected, hence the message is "The person is feeling drowsy". Similarly, if a person is looking elsewhere then the message is "The person is inattentive".

4.1 Few modifications to be incorporated in the setup as a part of the future work of this research

- Instead of playing a pre-recorded audio, a text-to-speech module may be incorporated within the system. In such a case the warning message in the text form will produce the corresponding speech form as the output.
- In case of implementing the system in a school classroom or a lecture hall, the algorithm is required to detect and label each face. Thus, if a student is inattentive, the system can warn or point to that particular labeled person.
- For employing the system in case of driver safety an alarm module may be incorporated so that it alerts the driver whenever he/she becomes inattentive.
- For employing the system in case of driver safety an alarm module may be incorporated so that it alerts the driver whenever he/she becomes inattentive.
- Also, an advanced version of the system is connecting the system with the brakes of the vehicle so that it can slow down in case the driver is inattentive for a long period of time.

5. EXPERIMENTAL RESULTS

The above approach is tested over few sample images and the overall accuracy rate is found to be 85% with 33.33% of false positive rate. Some results have been displayed in figure 11-14.

Also, to optimize the computation cost, few steps are followed:

- If a non-frontal face is detected, there is no need to run the algorithm for detecting the periocular region as it is clear that the subject is not attentive.
- If neither a frontal face nor a non-frontal face is detected, then it will be termed as 'not-a-face'.

6. CONCLUSIONS AND FUTURE SCOPE

This approach of detecting the periocular region and incorporating it to find out the attentiveness of any face in real time has been presented in this paper. The approach yielded 85% accuracy with 33.33% of false positive rate.

Employing this system inside a lecture hall or school classroom can be used as a metric to assess how many students are attentive in the class. Based on the metric of inattentiveness of students, arguably, the course structure or the topic may be modified. If more than half of the students are found to be inattentive, appropriate measures may be initiated to make the topic or the lecture more interesting.

Another important area of application for such a method will be in the transport industry. Whether a driver is attentively driving or is feeling drowsy or is looking at somewhere else can be found out using this tool. A number of expensive systems are available which require the whole face for proper assessment. The process of extracting features from the whole face is computationally expensive when compared with the periocular region for the same purpose.

Therefore, the proposed research possesses a huge scope of application and implementation in many areas.

Fig -9: Periocular region of attentive subject.

Fig -10: Result of thresholding and binarization.

Fig -11: Filling the holes within the connected components.

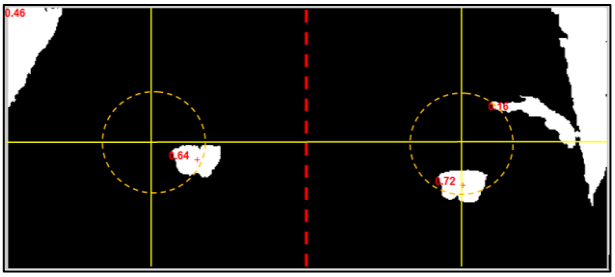


Fig -12: Detection of the eyeballs (centroid) which is situated within the specified boundary.

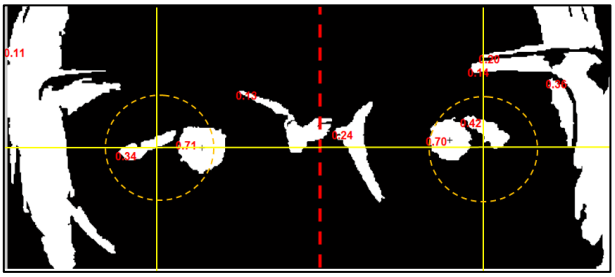
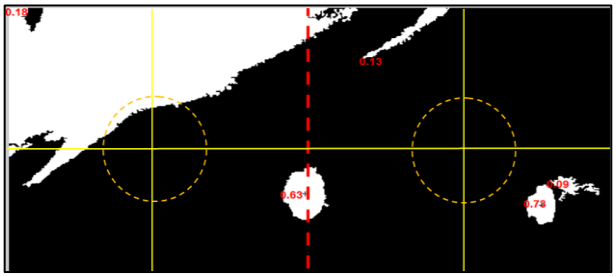
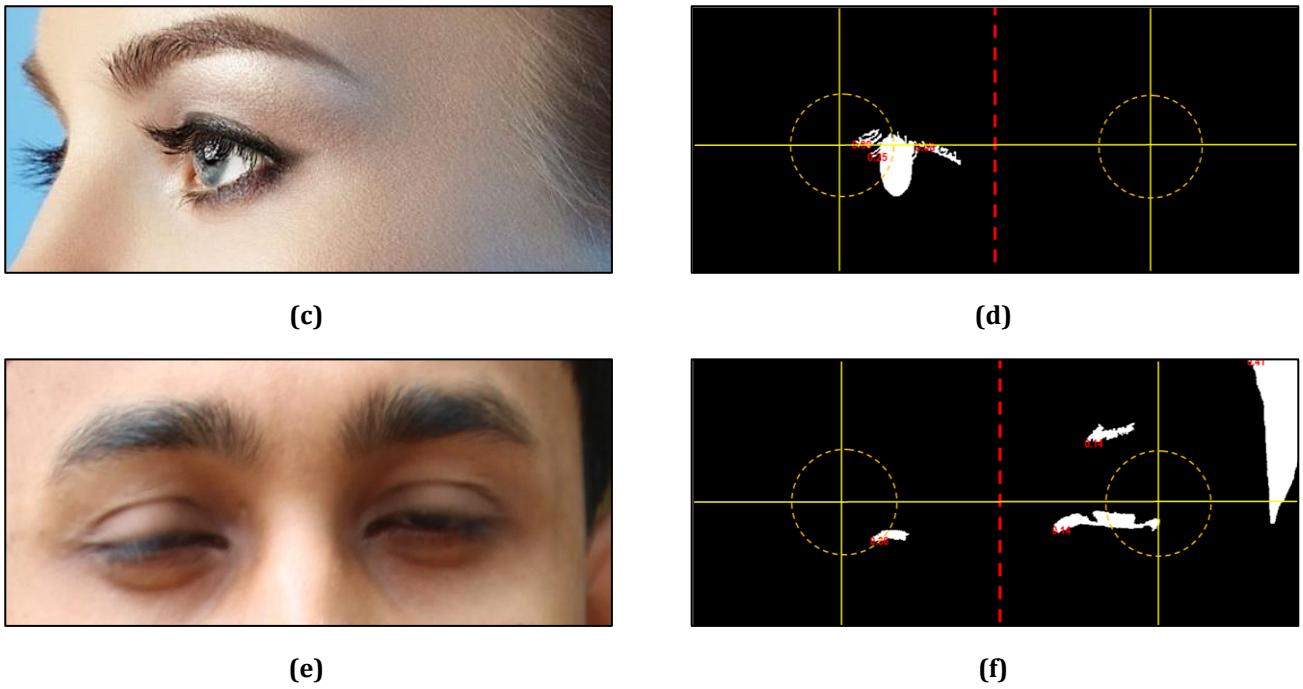


Fig -13: Periocular region of attentive subject with spectacles also satisfies the conditions of attentiveness.

(a)



(b)



(c) **(d)**

(e) **(f)**

Fig -14: (a), (c) & (e) show that the subject is inattentive. (b), (d) & (f) show their corresponding eyeball detection which fails the acceptance test.

REFERENCES

[1] Bharadwaj, Samarth, et al. "Periocular biometrics: When iris recognition fails." Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010.

[2] Juefei-Xu, Felix, et al. "Investigating age invariant face recognition based on periocular biometrics." Biometrics (IJCB), 2011 International Joint Conference on. IEEE, 2011.

[3] Merkow, Jameson, Brendan Jou, and Marios Savvides. "An exploration of gender identification using only the periocular region." Biometrics: Theory Applications and
Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010.

[4] Lyle, Jamie R., et al. "Soft biometric classification using periocular region features." Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010.

[5] Somajyoti Majumder, "Sensor Fusion and Feature Based Navigation for Subsea Robots", The University of Sydney, August 2001.

[6] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, January 5, 2004.

[7] Jacob Toft Pedersen, "Study group SURF: Feature detection & description", SURF: FEATURE DETECTION & DESCRIPTION, Q4 2011.

[8] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.

[9] Guo, Zhenhua, Lei Zhang, and David Zhang. "A completed modeling of local binary pattern operator for texture classification." IEEE Transactions on Image Processing 19.6 (2010): 1657-1663.

[10] Cunningham, Pádraig, Matthieu Cord, and Sarah Jane Delany. "Supervised learning." Machine learning techniques for multimedia. Springer Berlin Heidelberg, 2008. 21-49.

[11] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 3-24.

[12] Bakshi, Sambit, Pankaj K. Sa, and Banshidhar Majhi. "Optimized periocular template selection for human recognition." BioMed research international 2013 (2013).

[13] Ambika D R, Radhika K R, D Seshachalam, AN Exploration Of Periocular Region With Reduced

Region For Authentication : Realm Of Occult, Jan Zizka (Eds) : CCSIT, SIPP, AISC, PDCTA – 2013 pp. 241–248, 2013. © CS & IT-CSCP 2013.

[14] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," 2012 5th IAPR International Conference on Biometrics (ICB), New Delhi, 2012, pp. 439-445.

[15] Bharadwaj, Samarth, et al. "Periocular biometrics: When iris recognition fails." Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010.

[16] K. P. Hollingsworth, S. S. Darnell, P. E. Miller, D. L. Woodard, K. W. Bowyer and P. J. Flynn, "Human and Machine Performance on Periocular Biometrics Under Near-Infrared Light and Visible Light," in IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, pp. 588-601, April 2012.

[17] Kumar, Ravi Kant, et al. "Analysis of Attention Identification and Recognition of Faces through Segmentation and Relative Visual Saliency (SRVS)." Procedia Computer Science 54 (2015): 756-763.

[18] Zhong, Sheng-hua, et al. "Attention Modeling for Face Recognition via Deep Learning." CogSci. 2012.

[19] Smith, Paul, Mubarak Shah, and Niels da Vitoria Lobo. "Determining driver visual attention with one camera." IEEE transactions on intelligent transportation systems 4.4 (2003): 205-218.

BIOGRAPHIES

Mouli Laha has completed her B.Tech in CSE from WBUT (2014). She worked as a Research Fellow at CSIR-CMERI under the Ministry of Science, Government of India. She pursued her M.Tech in CSE from IIT (ISM), Dhanbad (2018). She has qualified GATE in 2014 and 2018 & UGC-NET in 2018. Currently, she is serving at National Informatics Centre under Ministry of Electronics and Information Technology, Government of India.