Intoxicated Person Identification Using Thermal Infrared Images and Gait

M. K. Bhuyan[1], Suchit Dhawle[2], Pradipta Sasmal[3], and Georgia Koukiou[4]

[1,2,3]Department of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati, Assam–781039, India
Emails: {mkb, s.dhawle, s.pradipta}@iitg.ernet.in
[4]Electronics Lab, Physics Department, University of Patras, Greece
[4]gkoukiou@upatras.gr

***Abstract*—In this paper, the activities of facial blood veins and temperature distribution and variations on the eye socket of drunk persons are studied using thermal images. Our proposed method considers walking pattern of a drunk person, which differentiate them from sober persons. For a sober person, vessels around the nose and eyes as well as nearer to the forehead region remain inactive and smooth, whereas these vessels become more active for a drunken person. Hence, thermal images can be used to detect drunkenness. The proposed method uses Curvelet Transform to capture the edges of a face to identify intoxicated candidate. The sclera and the iris are of the same temperature for the sober person. In contrary to this, the sclera temperature is high for intoxicated persons. SURF (Speeded up Robust Features) are used to accurately detect the temperature change in iris and sclera. The walking trajectory/pattern of a sober person and a drunk person can also be discriminated easily. Optical flow is employed to determine motion trajectories of drunk and sober persons. Finally, classification is done by using Random Forest and the Support Vector Machine. The experimental results show the efficacy of the proposed method.**

***Index Terms*—Intoxicated Person, Thermal Database, Curvelet, SURF, Optical Flow, SVM.**

I. Introduction

Security concerns include, to a great degree, bio-metrics and personal information [1]. Face identification [1] has been extensively examined in the visible spectrum. Research in face identification [1] has been prejudiced towards the visible spectrum for a number of reasons. One of those reasons is the accessibility of low cost cameras in the visible spectrum of the electromagnetic spectrum. Of late, importance has been given in obtaining information from faces in the thermal infrared spectrum [2,3]. The key reason behind this is that the temperature at various positions of the face depends on the physiological and the psychological circumstances of a particular individual, which in turn is strictly linked to the scattering of the vessel network on it [4]. With the help of these changes in the facial vessels, the distinction between drunk and sober person can be made.

Thermal properties of eyes are also being exploited now a days. Iris Identification is a central non-invasive biometric practice, which has been used till today for security reasons. Different properties of iris and sclera can be exploited with the help of thermal images. Drunkenness is a stimulating physiological condition, which can be inspected using infrared imagery. The physical characteristics of arteries and vessels of a drunken person change with the intake of alcohol. The thermal images of drunken and sober person are shown in Figs. 2–11. The sclera and the iris show different physical characteristics when a person consumes alcohol.

Drunk person identification using infrared images have been explored in the past [4–7]. But, most of these approaches are related to automotive anti-drunk driving systems. According to method presented in [2], the vessels network in the face is dense with large vessels, and a highly variable topology across individuals. A thermal gradient appears at the regions of the vessels, and intoxicated persons are prone to have more active blood vessels. In [6], the intoxication identification approach is based on extracting those blood vessels on the face which exhibit higher activity with alcohol consumption. But, no mathematical analysis is provided in this work to support intoxicated person discrimination capability. In [5] simple pixels on the thermal image of the face are used as features in the identification procedure, in which the concept of a “drunk feature space” is proposed. The approach in [5] is based on the fact that the facial vessels increase activity and some of them are brighter when the person consumes alcohol. For this purpose, in each pair of Infrared (IR) images, which corresponds to a specific sober and drunk person, filtering and morphological operations are carried out to reduce noise and enhance edges. Specifically, anisotropic diffusion is applied to enhance the vessels on the images, and after that, top-hat transformation is used for isolating the vessels from the face. Registration procedures are employed to reveal increased vessel activity on the face of the drunk person. Simple thresholding is applied to raise more active vessels. In drunk persons, vessels around nose and eyes become more active.

In [8] iris images captured in the near-infrared electromagnetic spectrum, and these images are compared before and after alcohol consumption. Due to alcohol consumption, the pupil dilates [8], which causes deformation in iris pattern, consequently affects iris recognition performance. Recently, emphasis has been given to acquire information from faces in thermal infrared spectrum [2]. The main reason is that the temperature of the face depends mainly on the physiological condition of the person. The human face has a mean temperature around 300 K, radiates according to the Wien's Law as a perfect black body, with maximum at 10 Lm wavelength. Thus, this region of electromagnetic spectrum (7–13 Lm) is the most appropriate for acquiring face information. In [4], the temperature distribution on the whole eye is examined before and after alcohol consumption. For the drunk person, iris appears darker as compared to sclera, which means that the sclera temperature increases. Thus, in a drunk screening procedure, the infrared images of the sober person are not needed. Histogram modification algorithms are employed when necessary, to show the gray level difference between the sclera and the iris of intoxicated persons. In this approach, for drunk person identification, a simple feature vector is formed by simply taking 20 different points on the face of a person. The drawback of this method is that the points are selected manually. If the training database is rotated then the interest points taken from the main vessel regions need to be estimated.

A significant number of research works on tracking of human motion have been reported. But, the work on motion trajectory and walking style of a drunk person are not explored sufficiently. It is well known that an intoxicated person has difficulty in walking in a straight line, while a sober person faces no such problems. This difference in the walking pattern can be utilized to distinguish an intoxicated person from a sober person. No previous works are reported in detecting the motion of an intoxicated person. Video surveillance involves acquiring and processing visual data from a scene to detect anomalies along time and space for purpose of recognizing interesting situations and perhaps generate alarms. The gait pattern of a drunk person is considerably different from that of a sober person. A major drawback of all the existing methods is that these methods does not simultaneously consider both eye and face information for distinction. Since alcohol brings changes in features of face and eyes, it is better to utilize both the information for better classification.

This paper tries to improve the accuracy of drunk person identification by using facial and eye thermal characteristics along with the gait pattern of intoxicated persons. We combine the output of these three classifiers to find out accurately whether a person is intoxicated or sober. The flow chart of our proposed method is shown in Fig. 1.

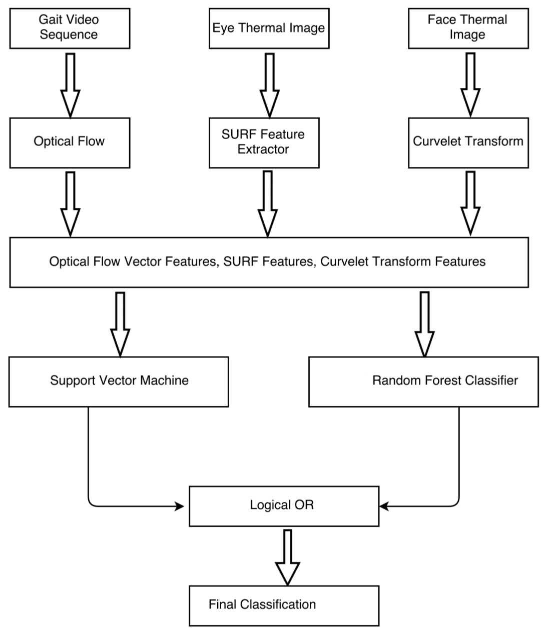


Fig. 1. Overall flow chart of the proposed method.

II. Methodology

As mentioned above, the main contribution of this research is the use of facial and eye thermal characteristics along with the gait pattern of intoxicated persons to distinguish them from the sober persons.

A. Discrete Curvelet Transform

It is found that the consumption of alcohol highlights the vessels in human face in the thermal images. These vessels look like edges on the face acting as singularities. Curvelet Transform is used here as edge detector. The curvelet transform is a multiscale directional transform that allows an almost optimal non adaptive sparse representation of objects with edges [9]. Frame elements in curvelet indexed by scale, location and orientation parameters in contrast to wavelets, where elements have only scale and location. The idea of Curvelets is to represent a curve as a superposition of functions of various lengths and widths obeying the scaling parabolic law. In this, angle polar wedges or angle trapezoid window is used in frequency domain. One essential advantage of the Curvelets is their ability to represent smooth edges [10]. Curvelet requires less decompositions to capture the edge information. Discrete curvelet transform is implemented using the wrapping based fast discrete curvelet transform.

B. SURF

We proposed to use Bag of Words approach for feature extraction from the thermal images of eyes. In this approach, the defect candidate region is represented as a histogram describing the frequency of certain patterns in the region. These patterns are Visual Words which together form a Visual Word Dictionary. Speeded up Robust Features (SURF) [11] is a widely used method to detect and represent interest points in an image. In our work, SURF descriptors are extracted from each of these key points by considering four different scales. First, an orientation is assigned to every key point which is computed using the responses of Haar wavelet in x and y direction in the circular neighborhood of the key points. This assignment of unique orientations to the key points allow the SURF descriptors to be invariant to rotations. Square regions are constructed around every key point, which are orientated in the directions of their corresponding key point orientation. The square regions are further divided into

4×4 sub-regions and inside each sub-region, Haar wavelet responses on horizontal direction dx and vertical direction dy are computed on regularly distributed sample points. These responses are summed up over the sub-region to give a vector represented by Eq. (1)

$$V = \left[\sum d_x,\ \sum d_y,\ \sum |d_x|,\ \sum |d_y|\right]. \tag{1}$$

The combination of the resulting descriptors for all the 4×4 sub regions is of length 64. Four square regions are constructed for every key point based on the four scales and corresponding orientations. Therefore, we get four SURF descriptors from every key point which are of length 64. The orientation and descriptor calculations are based on Haar wavelet responses and these responses can be calculated very quickly using integral images.

After the extraction of SURF features from the images, K-means clustering is performed on all the 64-dimensional features. This gives us K number of means which form the K-Visual Words of our Visual Word Dictionary. After the formation of the Visual Word Dictionary, the nearest neighbor amongst the Visual Words is found for each of the SURF features extracted from an image. For each extracted feature, a vote is given to its nearest Visual Word and a histogram of length K is created describing the frequency of each Visual Word in an image. Hence, every image is represented by a Histogram of Visual Words. This gives us a K-dimensional feature vector for every eye thermal image. In our work, the value of K is fixed at 500.

C. Optical Flow

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. Optical flow can arise from relative motion of objects and the viewer. Consequently. Optical flow can give important information about the spatial arrangement of the objects viewed and the rate of change of this arrangement. In our method, we only considered vertical velocity v. We simply added all the vertical velocities from all the frames and used image histogram on that matrix. Optical flow is used to find the motion information of a walking person - sober or drunken.

D. Classification

Random forest [12] is a powerful ensemble based classifier in which independent trees with low correlation are grown using different bootstrap samples of the dataset. Random Forest minimizes the high variance of individual trees and gives high accuracy in classification. For a test sample, each individual classification done on the basis of casting of votes by trees, and the final classification is done based on majority vote. Each tree uses around 67% of the available training data for its training because of Bootstrapping. The remaining 33% of the training data forms the Out-of-Bag data for the tree. Therefore, random forest classifier with N number of trees has N number of Out-of-Bag data subsets. These Out-of-Bag subsets provide an additional functionality of Out-of-Bag error to the random forest classifier where the performance of the trees on its Out-of-Bag subset is measured to find cumulative error. This can be used as the basis of tuning the parameters of the Random Forest, such as number of trees to be used and number of predictors to be considered at each node, and hence, this method eliminate the need for cross-validation and another separate test dataset. Random Forest classifier is employed on all the features from the face thermal database.

Fig. 2. Sober face thermal images.

Fig. 3. Drunk face thermal images.

SVM [13] is a binary classifier which determines a maximum margin hyperplane that separates two classes in the best possible way. We used the SVM classifier to effectively classify the eye thermal images and the gait based video frames. Drunk and non-drunk features obtained from eye database after applying SURF to it are fed to SVM classifier. Similarly, the walking features of two classes obtained with the help of optical flow are also given to SVM classifier.

III. Experimental Results

A. Dataset

The infrared images used in this work were acquired by means of the Thermal Vision Micron/A10 infrared camera [6]. The obtained information is in the thermal region of electromagnetic wavelengths, *i.e.* from 7000 to 13,000 nm, where the Wien curve has its maximum (at 9500 nm for 300 K). The temperature of the face depends mainly on the physiological condition of the person. The human face has a mean temperature around 300 K, radiates according to the Wien law as a perfect black body, with maximum at 10,000 nm wavelength. Thus, this region of the electromagnetic spectrum (7000 nm–13,000 nm) is the most appropriate to acquire information from the face. In this experimental procedure, 40 people were involved, composed of 30 males and 10 females. Each person consumed alcohol, four glasses of wine of 120 mL each, in the period of one hour (total of 62.4 mL alcohol). In each acquisition, a sequence of 50 frames were acquired from each person with a sampling period of 100 ms between the frames. The first acquisition of 50 frames for each specific person were obtained before alcohol consumption. A second acquisition of 50 frames were obtained 30 min after drinking the fourth glass of wine. The resolution of the infrared images is 128×160 pixels.

The experimental procedures require the availability of the thermal images of an intoxicated person, as well as the thermal

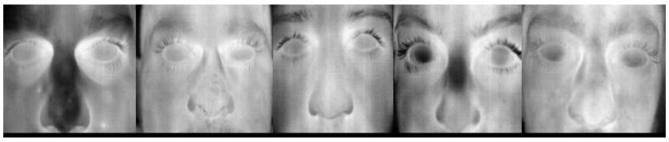


Fig. 4. Sober eye thermal images.

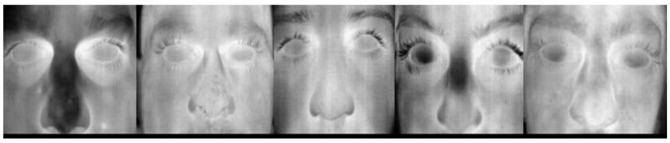


Fig. 5. Drunk eye thermal images.

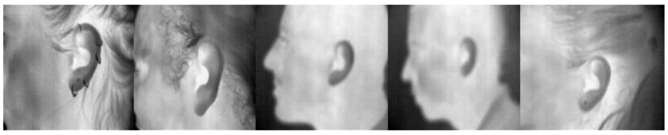


Fig. 6. Sober ear thermal images.

Fig. 7. Drunk ear thermal images.

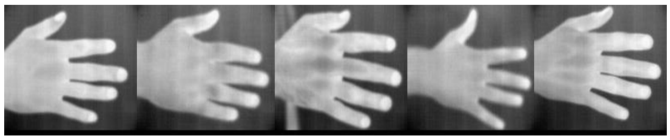


Fig. 8. Sober hand thermal images.

Fig. 9. Drunk hand thermal images.

images of the corresponding sober person so that comparisons can be carried out. Consequently, the people participating in the experiment had to be conscious about the requirements of the procedure. All participants were healthy and well aware of a possible risk they were undertaking. They also accepted their personal data such as age, weight and sex be available on the web for possible use by the scientific community. The term "intoxicated" in this project is attributed to a person who has consumed four glasses of wine, or a total of 62.4 mL of alcohol. This quantity of alcohol is considered the maximum that our researchers could drink and participate in the experiment. No blood tests were conducted in order to determine the alcohol quantity in their blood. According to various tests, only three glasses of wine are enough for any person to go beyond the limit of 0.5 gr/(L of blood), set for secure driving. This corresponds to 0.2 mg/L of exhaled air for a breathalyzer indication, which is the threshold for intoxication alarm. With the same quantity of alcohol, different persons are affected differently. This was realized by the measurements carried out by the police using a breathalyzer. With the quantity of 62.4 mL of alcohol given to all persons, the breath alcohol content ranges from 0.25 to 0.9 mg/L. It was found that this was the maximum concentration and was recorded 30 min after the consumption of the last glass of wine. The females were affected more than the males and the heaviest persons were affected less than the lighter. Specifically, for the males who participated in the experiment, the breathalyzer indication ranges from 0.22 to 0.37 mg/L of exhaled air. For the heaviest males the indication was the lowest. For the females, this indication is much higher, ranging from 0.49 to 0.89 mg/L. Finally, these measurements were normalized to the participant weight by obtaining the product weight and breathalyzer indication. This index ranges from 19 to 26 for the males, and from 30 to 50 for the females. Finally, it is worth mentioning that the people employed in this dataset were calm and in normal physical and psychological conditions during the experiment. No illness, no psychological stress, other pathological reason or any kind of body exercises were recorded for any one of the participants. They were asked to be present in the room of the experiment half an hour earlier and to keep calm till the first acquisition of frames. Actually, in this dataset experiment, the intoxicated situation is tested assuming that no other scenario happens, *i.e.* only the effect of alcohol consumption on the thermal signature of the face is measured.

During the acquisition procedure, the temperature and the dim light in the room were kept unchanged. A very dim light was available from a neighboring room for the researchers to be able to work. This light did not affect the operation of the infrared camera or the acquired infrared images. The distance of each face from the camera was around 30 cm and was kept constant from acquisition to acquisition. This results in a face that occupies the whole area of the frame and gives the images of the same person the capability of been easily compared.

For the walking trajectory based drunk and sober classification, one more database video was created. Four student's gait pattern were recorded, which resulted in a video of 240 frames long of MOV format. For the sober class, all the students walked in a straight trajectory with a normal hand movement. For the drunken class, students imitated the walking pattern of a drunk person. It is quite difficult for an alcoholic person to walk in a straight trajectory. This videos were recorded in a close passage with minimum lightening changes in a static simple background.

B. Results

For the intoxicated person identification, we first applied Curvelet to highlight the face vessels after the person consumes alcohol. Curvelet can efficiently capture facial features

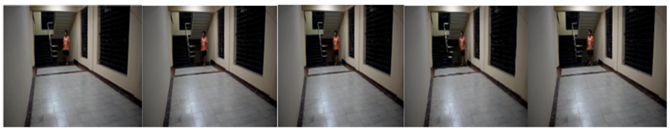


Fig. 10. Sober person walking sequence frames.

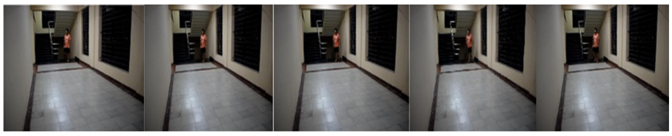


Fig. 11. Drunk person walking sequence frames.

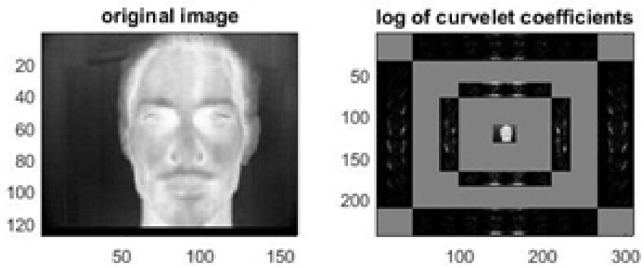


Fig. 12. Curvelet coefficients of drunk person.

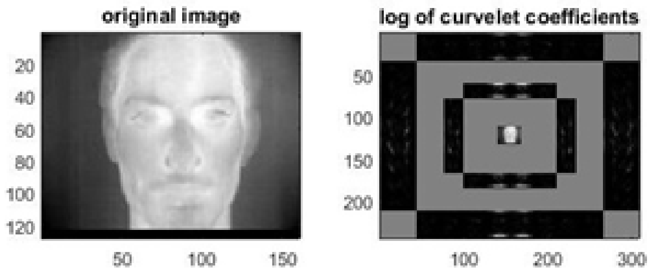
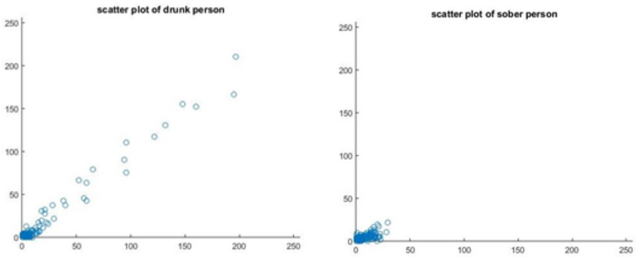
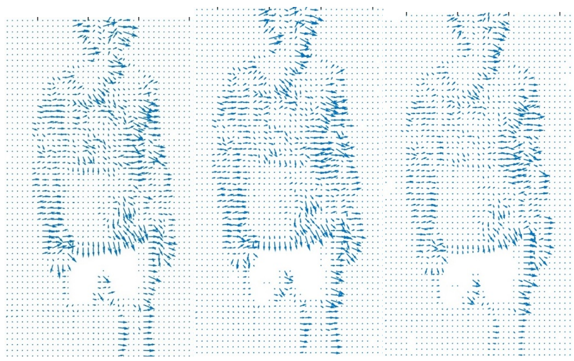


Fig. 13. Curvelet coefficients of sober person.



(a) (b)

Fig. 14. Plot showing more scattered values in (a) drunk person, (b) sober person. Horizontal and Vertical (U,V) velocity optical flow.



(a) (b) (c)

Fig. 15. Optical flow of three frame sequences as (a) no. 206, (b) no. 207, and (c) no. 208.

of a drunk and a non-drunk person. The curvelet coefficients for drunk and sober person are shown in Figs. 12–13. The vessels on the face of a drunk person are basically the edges, and the Curvelets are able to capture these features. All these feature vectors are classified by using random forest. We have altogether 1850 images for each of the drunk and sober classes. From the total of [3700 $\times$ 256] feature vectors, we used 65% of these features for training a random forest. With this percent of training, we obtained an 89.23% classification accuracy for a drunk person.

For the eye images, we used SURF features. The eye's thermal image of a drunk person exhibit different properties from that of a sober person. There is a temperature variation between the sclera and the iris part of the eye. For a drunk person, the sclera's brightness is more than the brightness of sclera of a sober person. And this difference is captured as a feature by SURF. In this case also, we have altogether 1850 images for each of the drunk and sober classes. From all the feature vectors, we only used 30% of feature vectors for training a SVM classifier. The accuracy given by the SVM classifier is 100%.

We also used gait of a person to identify whether he is drunk or not. Though the database for this is very less, we proved that this is an accurate way to identify an intoxicated person. The optical flow can effectively capture different walking patterns of a person. The optical flow vector for some frame sequences are given in Fig. 15. We have achieved almost 100% accuracy when the classifier was trained with 50% of features.

Combining the classified outputs: In this paper, we used three approaches to find out whether a person is drunk or not.

- Face Thermal Images: We obtained features from face images using Curvelet transform. These features are then passed through random forest classifier.
- Eyes Thermal Images: We obtained features from eye images using SURF. These features are then passed through SVM classifier.
- Gait Video Sequence: We obtained features from walking patterns using Optical Flow. These features are then passed through SVM classifier.

Test types	***Training data***	***Drunken***	***Sober***
Face thermal images	65%	89.23%	91.85%
Eyes thermal images	30%	100%	100%
Gait video sequence	50%	100%	100%

Finally we performed a logical or operation, *i.e.* if a person passes two tests out of the three then the person will be considered as a sober person. If the thermal image of face of a particular person and thermal image of eyes of the same person are classified as drunk, then the person will be considered as drunken.

IV. Conclusion

In the proposed approach, thermal infrared images are employed for intoxicated person discrimination along with the walking pattern. In intoxicated persons, blood vessels around the nose and eyes become more active. More bright and distinguishable blood vessels can be found in the forehead as well. The thermal images of 37 individuals were used for the feature formation and classification of a person from drunk to sober. Curvelet transform was used to find the features from the face thermal images. Random forest was used as a classifier for the differentiation of drunk and sober person.

Also, experimental results were presented which describe the temperature changes on the human eyes when somebody consumes alcohol. The basic evidence is that the iris remains in the same temperature while the sclera increases its temperature with alcohol consumption. Consequently, the iris appears darker in the thermal imagery. A physical explanation is that the sclera contains a denser blood vessels network than the iris, which increases the temperature of the sclera with alcohol consumption. So, thermal images of 37 individuals were used for the feature formation and classification.

Finally, gait patterns were used to detect whether an individual is drunk or not. The walking patterns of drunk and sober person were recorded. In this, 50% of the features were used for training and optical flow was employed for feature extraction. The database of hand and ear thermal images is also publicly available. But, the difference in physical characteristics of hand and ear which occurs after the consumption of alcohol could not be found out. In Optical flow, Histogram of Oriented Optical Flow can be used to reduce dimension.

References

[1] W. Zhao, R. Chellappa, P. Phillips and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] P. Buddharaju, J. Pavlidis and P. Tsiamyryzis, “Physiology-based face recognition in in the thermal infrared spectrum,” *IEEE Transactions on PAMI*, vol. 29, no. 4, pp. 613–626, 2007.

[3] D. A. Socolinsky and A. Selinger, “A comparative analysis of face recognition performance with visible and thermal infrared imagery,” in *Proc. 16th International Conference Pattern Recognition*, vol. 4, pp. 217–222, 2002.

[4] G. Koukiou and V. Anastassopoulos, “Drunk person screening using eye thermal signatures,” *Journal of Forensic Sciences*, vol. 61, no. 1, pp. 259–264, 2015

[5] G. Koukiou and V. Anastassopoulos, “Drunk person identification using thermal infrared images,” *International Journal of Electronic Security and Digital Forensics*, vol. 4, no. 4, pp. 229–243, 2012.

[6] G. Koukiou and V. Anastassopoulos, “Facial blood vessels activity in drunk persons using thermal infrared images,” in *4th International Conference on Imaging for Crime Detection and Prevention*, pp. 1–4, 2011.

[7] G. Koukiou and V. Anastassopoulos, “Drunk person screening using eye thermal signatures,” *Forensic Sciences*, vol. 61, no. 1, 2016.

[8] S. Arora and M. Vatsa, “Iris recognition under alcohol influence: a preliminary study,” in *Proceedings of the 5th IAPR International Conference on Biometrics (ICB)*, p. 33641, 2012.

[9] M. Alhanjor, “Curvelet and waveatom transforms based feature extraction for face detection,” *Al-Aqsa University Journal*, vol. 15, no. 1, pp. 41–66, 2011.

[10] E. J. Cand‘es and D. L. Donoho, “New tight frames of curvelets and optimal representations of objects with C2 singularities,” *Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2004.

[11] H. Bay, T. Tuytelaars and L. V. Gool, “SURF: Speeded up robust features,” in *Computer Vision ECCV 2006: 9th European Conference on Computer Vision*, pp. 404–417, 2006.

[12] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[13] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.