MIDSEM REPORT ON INTOXICATION DETECTION

UNDER DR.KAMLESH TIWARI AND DR.ASHUTOSH BHATIA

GARVIT SINGHAL(2021A7PS2226P)
SANJEEV MALLICK(2021A7PS2217P)

Intoxication Detection

Garvit Singhal
Sanjeev Mallick
Birla Institute of Technology and Science Pilani, Rajasthan 333031

Abstract

Detecting intoxication is very important in different situations like ensuring safety while driving, monitoring employees at work, and keeping public health in check. This report explores different ways to tackle this issue. Intoxication detection is necessary to keep people safe and reduce the risks that come with impaired decision-making and coordination caused by alcohol. Current methods, like watching behavior or using breathalyzer tests, have limitations—they can be subjective, invasive, and not always practical for real-time monitoring. Using deep learning frameworks, face detection algorithms, and thermal infrared images offers a non-invasive and continuous way to monitor intoxication. Our goal is to create a strong model for detecting intoxication using the above methods. This will improve accuracy, reliability, and the ability to scale up compared to current methods. By overcoming these challenges, our research aims to make public spaces safer and intervene promptly to prevent accidents and health problems related to alcohol.

1 Introduction

1. **What is the problem?**
   The problem is the detection of alcohol intoxication in individuals. This involves identifying and quantifying physiological or behavioral signals associated with alcohol consumption through various means, such as image analysis.

2. **Why it is important?**
   Alcohol consumption can impair judgment, coordination, and reaction time, leading to accidents, injuries and fatalities. Timely detection of intoxication is crucial for ensuring public safety in contexts such as transportation, workplace safety, and public events.

3. **What are the challenges? Why it is difficult?**
   Intoxication detection using deep learning, face detection, and thermal imaging faces challenges in diverse data representation, complex facial expressions analysis, noise interference, and real-time processing demands. Robust algorithms are needed to generalize across varied conditions, analyze facial expressions, minimize thermal imaging noise, and enable real-time processing, for the development of accurate and ethically responsible detection systems.

4. **What are current approaches?**
   Current approaches to intoxication detection primarily rely on behavioral observations, breathalyzers, or biochemical tests. Behavioral observations involve assessing physical symptoms such as slurred speech and impaired coordination. Breathalyzers measure blood alcohol concentration (BAC) through breath samples, while biochemical tests analyze bodily fluids such as blood or urine for alcohol metabolites.

5. **Where current approaches lack?**
   While current approaches are effective to some extent, they have limitations. Behavioral observations can be subjective and may not detect early stages of intoxication or individuals who exhibit high tolerance. Breathalyzers require direct cooperation

from individuals and may not be suitable for continuous monitoring. Biochemical tests are time-consuming, unhygienic and may not provide real-time results.

6. **What is your objective in this?**
Our goal is to use thermal infrared images along with deep learning and machine learning to detect intoxication more accurately and in real-time. By teaching computers to recognize patterns in these images associated with alcohol consumption, we aim to create a system that can continuously monitor for signs of intoxication without being invasive. This helps improve safety by allowing for early detection of intoxication to prevent accidents or health issues related to alcohol.

2 Intoxication Detection Using Two Stage Deep Neural Network

Abstract

Earlier to detect intoxicated people,we used breath alcohol testers as the primary method but its not hygienic , expensive and not scalable .In this paper a two stage deep neural network is used to overcome this issue.The first stage uses simplified VGG network to determine the age of the people and in second stage we use simplified dense net to identify facial features of drunk people. The age discrimination stage obtained an accuracy of 86.36%. Age classification is used in the first stage because it improves the accuracy significantly.

2.1 Summary

In the first stage we use simplified VGG to identify the age group whose results are used in second stage. For this stage we used IMDB-WIKI dataset to train the network. There is a image preprocessing stage and after this step the required training images are obtained and further images are obtained through data augmentation to build the age data set. Finally this simplified VGG is used to train the dataset and classify the age.

In the second stage the dataset specifically collected is used to obtain the facial image ROI through alcohol image preprocessing.Then the data augmentation is used to generate a more diverse dataset.Finally the the simplified Dense-Net was used for training and alcohol test identification.

2.1.1 Data Augmentation

The hardware used to collect the dataset was a webcam: Logitech C310(Resolution - 1280 x 720p/30fps ; Focus type:Fixed focal length ; Vision : 60 degree) Dataset was divided into 4 different stated. 1)Not drinking, 2) Drinking does not exceed the standard, 3)Excessive drinking, 4) Drinking seriously exceeds standards.The image preprocessing stage includes rotation ,finding region of interest , histogram equalization , median filtering and using grey scale.To increase the amount of original and overall data and avoid overfitting during training we perform data augmentation

2.1.2 Dataset

The age dataset of the first stage of this work is divided into 3 categories, namely 18-30 years, 31-50 years, and greater than 51 years having 322 ,313 and 541 pictures respectively. Using data augmentation the number of images were 19072 and a total of 1664 images were used for validation. In the second stage of intoxication recognition we used self recorded data, including images from 124 subjects: 77 people aged 18–30, 25 people aged 31–50, and 22 people aged above 51 years, of which 110 subjects are used for training and validation, and 14 subjects for testing

2.1.3 Deep Neural Networks

1. **Simplified VGG**
We are using a DNN called VGG16, but remove one part: a 6-layer section that involves convolutions (which are a way to process images) with specific settings, along with another section that involves fully connected layers. We also add batch normalization and ReLU after each convolutional layer. This helps make the network more efficient and accurate at recognizing patterns in images.

2. **Simplified Dense-Net**
   The simplified Dense Net architecture DNN is used in the second stage of the intoxication detection identification. The input pictures are 140 x 140 x 3; CI(Conv.) is the accumulation of the various functions such as padding ,convolution, batch normalization, ReLU and pooling. We used only 2 layers of DenseBlock with a bottleneck structure and 1 layer of transition layer.The two layers of DenseBlock have six and eight layers, and the original C1 block with 3×3 ZeroPadding, 7×7 Convolution, stride of 2, and 3×3 MaxPooling architecture is modified. First, three 3×3 convolutions are used, to replace one 7×7.After each convolutional layer,batch normalization and ReLU are added, and the original stride = 2 action is replaced with MaxPooling to achieve dimensionality reduction and to extract the significant feature.

2.1.4 Results

1. **Stage-1 result**
   A dataset about people's ages and a simplified VGG model were used to predict the age of people.After training the model using stochastic gradient descent with a momentum of 0.9,starting with a learning rate of 0.0001 and batch size of 128 for training and 32 for validation.At first learning rate was high to quickly reduce the loss,which started at 0.99 and went down to 0.61.As training continued,accuracy improved steadily reaching 89.06 and loss decreased to 0.3172.After training the model and testing it with 15,500 images the model achieved 89.36% accuracy.

2. **Stage-2 result**
   The result of alcohol test for age groups 18-30,31-50 and greater than 51 are 94%,83% and 81% respectively.

2.1.5 References

1. Republic of China Government Information Open Platform-Statistics on the Number of Injuries and Deaths Caused by Drunk Driving and the Number of People Over the Years. Accessed: Aug. 2019. [Online]. Available: https://data.nat.gov.tw/dataset/9018

2. G. Koukiou and V. Anastassopoulos, "Drunk person identification using thermal infrared images," Int. J. Electron. Secur. Digit. Forensics, vol. 4, no. 4, pp. 229–243, 2012.

3. K. Takahashi, K. Hiramatasu, and M. Tetsuishi, "Experiments on detection of drinking from face images using neural networks," in Proc. 2nd Int. Conf. Soft Comput. Mach. Intell. (ISCMI), Nov. 2015, pp. 97–101.

4. C. Wu, K. Tsang, H. Chi, and F. Hung, "A precise drunk driving detection using weighted kernel based on electrocardiogram," Sensors, vol. 16, no. 5, p. 659, May 2016.

3 Drunk Driving and Drowsiness Detection

Abstract

Driver drowsiness is based on real time detection of the driver's head,face and mouth where in Haar Cascade classifier for face and eye detection and template matching in the mouth region for yawning detection.

3.1 Summary

In this paper,driving behavior information-based approaches assess a driver's conduct by analyzing factors like speed, steering wheel angle, acceleration, lateral position, and braking patterns over time.It is trained and tested through Support Vector Machine (SVM) and Logistic Regression Models (LRM) on data. The feature-based approach looks at specific facial features like eye blinks, yawning, and head movements to analyze driver drowsiness.

3.2 Project Objective

The project aims to create a safety system for cars to prevent accidents caused by drunk or drowsy driving. It will use visual features like detecting eyes, face, and yawns for drowsiness detection.An alcohol sensor will be used for drunk state detection. If the system detects either condition, it will sound an alarm using a buzzer and vibrator to alert the driver. If the alerts are ignored three times within a set time, the engine will be automatically turned off, with an emergency LED indicator turning on.The proposed scheme uses facial features eye index and pupil activity to extract information of non alert behaviour of driver.The percentage of eye closure was used to indicate alertness level.Yawning detection was also performed by using mouth geometric features.

3.3 Methodology

Visual features consists of capturing eye state,head position and yawning.Drunk state detection would be done using an alcohol sensor to detect whether a driver is drunk or not.

3.3.1 Eye State

Visual analysis of eye state is done by a single camera for constantly monitoring alertness.The detection of face and eyes was done through AdaBoost algorithms and by adaptive template matching.Then visual features like pupil activity and eye index were computed.An SVM classifier was used to learn the driving patterns of the driver to classify people in alert or non alert situations.The paper also describes second approach which is using HAAR-like feature detection in which a web cam is used to capture real time facial images using HAAR-cascade classifier.An ROI is selected on the detected facial image and then eyes are detected using selected ROI.

3.3.2 Head Position

To track head position,Lucas-Kanade optical flow method is used. This method estimates feature points between two video frames to track changes in position. Unlike other methods relying on image intensity, which can be affected by lighting, this method is more reliable. Additionally,it used a Kalman-based tracking technique to reduce search areas for faster and accurate face detection. This approach is efficient and maintains accuracy.

3.3.3 Yawn detection

The paper proposes a computer vision to detect yawning, an important indicator of driver drowsiness. It used the Viola-Jones algorithm to detect the driver's face and AdaBoost algorithm to select critical features for classification. Once the face is detected, the system stops searching for other faces, saving time. Mouth detection is then performed using trained features, and the mouth region is converted to grayscale. A histogram of the mouth's position is saved as a reference. Yawning is determined by comparing pixel values of the open mouth region to a normal closed mouth position. If the pixel count for the open mouth or closed eyes exceeds normal thresholds, the system detects yawning, indicating potential drowsiness in the driver.

3.3.4 Drunk state detection

To prevent drunk driving, vehicle-based alcohol detection systems have been developed. These systems detect charged water clusters in exhaled breath, which are indicative of alcohol consumption. By applying an electric field and measuring the resulting currents, these systems can simultaneously detect breath and alcohol peaks. In India, where drunk driving is a major cause of road fatalities, such systems are crucial. One setup uses a low-cost alcohol sensor connected to a microcontroller unit (MCU) to control the car's motor based on the detected alcohol level. Another system employs a CCD camera to capture the driver's image and continuously monitors alcohol concentration. If the concentration exceeds a specified limit, the system activates alarms and prevents the engine from starting. It verifies the driver's identity by comparing images captured before and after alcohol detection, thereby preventing fraudulent attempts to bypass the system.

3.4 References

1. Sinan Kaplan, Mehmet Amac Guvensan, Member, IEEE, Ali Gokhan Yavuz, and Yasin Karalurt, Driver Behavior Analysis for Safe Driving: A Survey, IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 6, December 2015.
2. Ralph Oyini Mbouna, Seong G. Kong, Senior Member, IEEE, and Myung-Geun Chun, "Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring", IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 3, September 2013.

3. Mario I. Chacon-Murguia and Claudia Prieto-Resendiz, “Detecting Driver Drowsiness: A survey of system designs and technology”, IEEE Consumer Electronics Magazine, October 2015.

4. Mona Omidyeganeh, Member, IEEE, Shervin Shirmohammadi, Senior Member, IEEE, Shabnam Abtahi, Aasim Khurshid, Muhammad Farhan, Jacob Scharcanski, Senior Member, IEEE, Behnoosh Hariri, Daniel Laroche, and Luc Martel, “Yawning Detection Using Embedded Smart Cameras”, IEEE Transactions on Instrumentation and Measurement, Vol. 65, No. 3, March 2016.

5. T. Sivakumar, Dr. R. Krishnaraj, “Road Traffic Accidents (Rtas) Due to Drunken Driving in India Challenges In Prevention”, IRACST-International Journal of Research in Management Technology (IJRMT), Vol. 2, No. 4, August 2012.

6. Anirban Dasgupta, Anjith George, S. L. Happy, and Aurobinda Routray, Member, IEEE, “A Vision-Based System for Monitoring the Loss of Attention in Automotive Drivers”, IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 4, December 2013.

4 Detection of Intoxication in Automobile Drivers

Abstract

The paper discusses about technology to detect if someone is drunk, especially for situations like driving.The current methods are on facial recognition but need a lot of improvements. In this we mainly focused on a system using eye movements and CNN to predict intoxication. The study tested 5 methods, including different CNN architectures like VGG19 and MobileNet V2, as well as an LSTM + Attention mechanism approach.The results show that VGG16 performed the best for identifying intoxication.

4.1 Survey

The paper reviews facial recognition research, proposes age assessment without facial markers, evaluates facial actions for behavioral biometrics, and explores CNN efficiency for image identification. It investigates deep CNNs for classification, thermal imagery for intoxication detection, and automatic bimodal intoxication detection using CNNs and DNNs. Additionally, it examines ECG-based intoxication detection, suggests a two-stage neural network for drunk driving detection, and discusses real-time monitoring techniques and MCU chip-based alcohol detection. The paper also surveys driver drowsiness and drunkenness detection technologies, explores eye tracking technology for drowsiness detection, presents novel facial landmark optimization techniques, and evaluates DenseNets performance. It discusses the use of Inception modules in CNNs, technology for identifying passengers using central console controls, and considers marijuana intoxication alongside alcohol’s observable facial changes.

4.2 Proposed work

In studying drunk detection using neural networks, we noticed a unique method where two neural networks were used in succession, first to determine the user’s age and then to classify whether they were likely drunk or not. Our approach differs in that we aim to train the model in one go, exploring various algorithms and comparing their results. We focus on a dataset with 16 acquisitions from 41 images and plan to train a conventional CNN, along with testing more advanced architectures like ResNet, VGG, or MobileNet. Our goal is to enhance overall accuracy and provide a detailed analysis of which algorithms work best for this task. Additionally, we investigate the performance of LSTM and attention mechanisms, which haven’t been widely studied in this context, offering a new approach to the field.

4.2.1 Dataset

Initially we had 41 images,each has 4 quadrants having four levels of sobriety labeled by wine glasses. We need to pre-process this data into a form that could better be used as input for the classifiers. //The dataset, known as the Sober-Drunk database, is open-source and originates from a PHD study, containing 41 photos. Each photo consists of 16 different acquisitions, with each acquisition comprising 50 consecutive frames captured every 100 milliseconds, totaling five seconds. To preprocess the data, the initial step involves enlarging and segregating the .tif files into separate images, as each file contains four still photographs of the same individual. Data augmentation techniques, such as mirroring, rotating, and flipping, are applied to increase the variety of images. By augmenting the data, the number of images is expanded, aiding in achieving better results. These augmented images are then used to split the data into training and testing sets for further analysis. The initial cropping increases the dataset from 41 images to 164 images (Each individual image produces 24 new images including itself). We also flip and rotate the images so the number of permutations images in the dataset shoots up to 984. Hence, data augmentation provides a 2300% increase in the data.

4.3 Implementation

We tested five different CNN architectures to compare their effectiveness in detecting intoxication. Data augmentation involved cropping, rotating, flipping, and mirroring, increasing the dataset from 41 to 984 images. Our main goal was to determine which approach provided the most accurate detection of intoxication. We compared testing and validation accuracies across the different algorithms and architectures, including VGG19, VGG16, MobileNet V2, ResNet 50,and LSTM + Attention Mechanism.

4.4 VGG 19

It takes fixed-size RGB images (224x224x3) as input, with preprocessing involving computing the mean RGB value of each pixel across the training set. Using 3x3 kernels and a stride of 1, the model covers the entire image, maintaining spatial resolution with spatial padding. Max pooling with a 2x2 frame and stride 2 is applied for downsampling. Rectified linear unit (ReLU) activation functions are used for nonlinearity, improving classification accuracy. The model consists of three fully connected layers, with 4096 nodes in the first two layers, followed by 1000 channels for classification using the ImageNet dataset, and a softmax function for the final layer.

4.4.1 VGG 16

VGG16 is a well-established vision model architecture, with 16 of its 21 layers being weight layers. It comprises three dense layers, five layers with max pooling, and thirteen layers with convolutions. It requires an input tensor with three RGB channels and a size of 224x224. The arrangement of max pooling and convolution layers remains consistent throughout the architecture. The convolution layers include 64, 128, 256, and 512 filters respectively, and the model concludes with three fully connected layers, followed by a softmax layer for classification.

4.4.2 MobileNet V2

MobileNet V2, originally designed for mobile devices, offers strong performance with a low parameter count, making it suitable for classification tasks. It consists of two distinct blocks, each containing three levels. The first block includes a 1x1 convolution with ReLU6, followed by a depth-wise convolution and another 1x1 convolution. The second block follows a similar structure but lacks nonlinearity in the final convolution layer. MobileNet V2 utilizes ReLU activation functions to enhance the capabilities of deep networks within the nonzero volume portion of the output domain.

4.4.3 Resnet 50

We chose ResNet-50 for its large number of trainable parameters, making it suitable for our classification task. The model consists of identity and convolution blocks in each of its five phases, with three convolution layers in each block. ResNet-50 can be trained with around 23 million parameters and is commonly used in computer vision tasks due to its robustness. Skip connections are utilized to pass output between layers, mitigating the issue of vanishing gradients.

4.4.4 LSTM + Attention Mechanism

LSTMs (Long Short-Term Memory networks) employ gates to efficiently process information from both Long-Term Memory (LTM) and Short-Term Memory (STM). These gates include the forget gate, which discards unnecessary information from LTM. The learn gate integrates new input with STM, allowing recent information to influence the current input. The remember gate combines unforgotten STM and event data to update LTM. Lastly, the apply gate uses LTM, STM, and event data to predict the outcome of the current event, thereby updating STM.

4.5 Results

From our results, it appears that VGG 16 yielded the best performance for drunk detection. This outcome might be influenced by the dataset size, suggesting that with a larger dataset, different results could emerge. However, for smaller datasets, VGG 16 could be the most efficient algorithm. The validation accuracies for each approach are summarized in Table 1. VGG 19, VGG 16, MobileNet V2, ResNet 50, and LSTM achieved accuracies of 83.59%, 86.72%, 71.09%, 73.44%, and 75% respectively

4.6 References

1. A.S. Tolba, A.H. El-Baz, A.A. El-Harby, Face Recognition: A Literature Review, International Journal of Signal Processing, 2006, 2, (2), 88-103.

2. K. R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," Int. J. Comput. Vis., vol. 126, nos. 2–4, pp. 144–157, Apr.2018.

3. L. Benedikt, D. Cosker, P. L. Rosin, D. Marshall, Assessing the Uniqueness and Permanence of Facial Actions for Use in Biometric Applications, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2010, 40, (3), 449-460.

4. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

5. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 25. Stateline, NV, USA,Dec. 2012, pp. 1097–1105

6. G. Koukiou and V. Anastassopoulos, "Drunk person identification using thermal infrared images," Int. J. Electron. Secur.Digit. Forensics, vol. 4, no. 4, pp. 229–243, 2012.

5 Face Detection in Extreme Conditions: A Machine-learning Approach

Abstract

Detecting faces accurately in different situations has been difficult because of factors like facial expressions, lighting conditions, and obstructions. Recent studies show that using deep learning techniques can greatly improve face detection, even in challenging conditions. Identifying people through pictures has become popular, but methods like fingerprint or retina scanning are more reliable. Deep learning can also enhance performance in these areas. In this paper, a new approach is proposed called a deep cascaded multi-task framework. It uses deep convolutional networks in layers to detect faces and facial landmarks efficiently. Additionally, a new method for selecting difficult examples during learning is suggested, which helps improve performance automatically without manual intervention.

5.1 Introduction

Many algorithms have been developed for face detection, but reliably detecting faces in challenging conditions remains difficult. Deep learning techniques have shown promise in improving face detection accuracy, even outperforming humans. However, existing methods often fail to handle extreme variations in lighting, poses, and occlusions. This paper proposes a new approach using cascaded Convolutional Neural Networks (CNNs) to address these challenges. The framework consists of three stages: quickly identifying candidate face windows, refining these windows to reject non-face regions, and accurately detecting facial landmarks. By using multi-task learning, the algorithm's performance is significantly enhanced. The main contributions of the paper include introducing a novel cascaded CNN framework for face detection, designing lightweight CNN architectures for real-time processing, and implementing an effective method for online difficult sample mining to improve performance. Experimental results demonstrate the superiority of the proposed method over existing techniques in face detection tasks.

5.2 Face Detection

The Multi-Task Cascaded Convolutional Network (MTCNN) is a framework for detecting and aligning faces using deep learning. It consists of three stages of convolutional neural networks (CNNs) that identify faces and facial landmarks like eyes, nose, and mouth. MTCNN integrates both face detection and alignment tasks using multi-task learning. It quickly produces candidate face windows in the first stage using a shallow CNN, refines them in the second stage with a more complex CNN, and then further refines the results and outputs facial landmark positions using a third CNN. MTCNN performs better than traditional methods, accurately detecting faces in real-time. It uses loss functions to train each network, including cross-entropy loss for face detection and Euclidean distance for bounding box regression and landmark localization. These losses are weighted and combined to form the final total loss.

5.3 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a type of neural network used in deep learning. They're great for tasks like image recognition. CNNs work by processing data through layers of operations like convolution (extracting features), pooling (reducing data size), and fully connected layers (mapping features to output). The convolutional layer focuses on extracting specific features from input data, while the pooling layer reduces data size to speed up training. Fully connected layers map the extracted features to a one-dimensional vector for final analysis. CNNs are effective because they learn important patterns and relationships directly from the data, making them ideal for tasks like image classification and object detection.

5.4 Literature Review

Researchers are actively working to improve object detection accuracy in computer vision, particularly for faces and eyes. Kasinski introduced a three-stage hierarchical face and eye detection system using Haar cascade classifiers, achieving a 94% true positive rate for facial detection and an 88% detection rate for eyes with only 1% false positives. Zhang utilized deep convolutional networks to predict face positions and landmarks with high accuracy, outperforming traditional methods in multiple tests while maintaining real-time efficiency. Lang ye proposed a CNN framework for eye detection using raw color values of image pixels, achieving a 73% accuracy rate and 76% recall rate, surpassing previous methods. These advancements hold promise for various applications requiring precise object detection.

5.5 Result

The experimental results involved two model classes: MTCNN and Haar cascade algorithm, applied to the same dataset. MTCNN achieved a high accuracy rate of 99%, while Haar cascade reached 68%. A comparison of the results highlighted MTCNN's superiority in detecting faces. Further analysis using different datasets confirmed the effectiveness of MTCNN, particularly in deep learning techniques. The proposed MTCNN framework consistently outperformed other methods, demonstrating its potential for various applications such as face recognition and scene analysis. Future work may include expanding the framework to incorporate additional features like facial expressions or iris movement detection. Overall, the study concludes that MTCNN offers a promising approach for face detection tasks.

5.6 References

1. B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel eatures for multi-view face detection," in IEEE International Joint Conference on Biometrics, 2014, pp. 1-8.

2. P. Viola and M. J. Jones, "Robust real-time face detection. International journal of computer vision," vol. 57, no. 2, pp. 137-154, 2004.

3. M. T. Pham, Y. Gao, V. D. D. Hoang, and T. J. Cham, “Fast polygonal integration and its application in extending haar-like features to improve object detection,” in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 942-949.

4. Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, “Fast human detection using a cascade of histograms of oriented gradients,” in IEEE Computer Conference on Computer Vision and Pattern Recognition, 2006, pp. 1491-1498.

6 Driver Face Recognition and Sober Drunk Classification using Thermal Images

Abstract

The proposed system aims to address drunk driving accidents by capturing the driver's face using thermal imaging. Facial recognition, powered by Convolutional Neural Network (CNN), distinguishes between drunk and sober individuals. Drunk classification relies on Gaussian Mixture Model and Fischer Linear Discriminant for dimensionality reduction. Additionally, capillary junction points on the face are used to assess differences in blood temperature, aiding in the classification process. This approach offers a more robust and automated solution compared to manual checks and awareness programs.

6.1 Introduction

The paper introduces thermal face recognition as a solution to identify drunk drivers, addressing the rising concern of road accidents due to alcohol consumption. It proposes using Convolutional Neural Network (CNN) for face recognition and classification of drivers as sober or drunk based on temperature differences. The study reveals challenges like limited thermal face databases and data format issues. The paper details findings, the project model, face recognition and classification algorithms, dataset and experimental results, concluding with insights in respective sections.

6.2 Survey

The literature survey discusses various studies related to thermal face recognition for identifying drunk individuals. It highlights the effectiveness of Convolutional Neural Network (CNN) architecture in extracting features directly from raw thermal data without preprocessing. Studies have employed different algorithms for face recognition and classification, such as Weber Local Descriptor (WLD), Local Binary Pattern (LBP), Fisher linear discriminant, and Gaussian mixture model. These methods utilize specific facial points to detect changes in capillary and vein junctions due to alcohol consumption. CNN has shown superiority over other feature extraction methods in terms of accuracy and real-time recognition, especially in near-infrared (NIR) images. Furthermore, CNN algorithms have been tested for real-time recognition and tracking using wearable devices and low-cost portable cameras, demonstrating comparable results to conventional cameras but with faster processing times.

6.3 Proposed model

The proposed model consists of two stages: face recognition and drunk classification. In the face recognition stage, Convolutional Neural Network (CNN) is used to recognize faces from thermal images. Various layers of CNN, including convolution, ReLU, pooling, normalization, and softmax classifier, are employed for this purpose. In the drunk classification stage, features are extracted from the thermal images and reduced using Fisher Linear Discriminant (FLD). The classification is then performed using Gaussian Mixture Model (GMM). The process involves pre-processing, feature extraction, FLD for dimensionality reduction, and GMM for classification.The model aims to identify drunk drivers by detecting temperature differences near the eye socket and nose compared to sober individuals, using facial features extracted from thermal images.

6.4 Results

The dataset involves 31 males and 10 females who consumed four glasses of red wine each. Infrared images were captured using a FLIR camera. The proposed model achieved a face recognition accuracy of around 97% using CNN. After recognition, individuals were classified as drunk or sober using FLD and GMM, with an approximately 87% accuracy in classifying drunk drivers. The project aims to address the issue of drunk driving accidents by automatically identifying and classifying individuals, thereby improving road safety and aiding law enforcement. This approach can be applied in real-time scenarios to identify and prevent drunk driving incidents more effectively, potentially reducing the need for manual intervention and expediting justice.

6.5 Reference

1. Zhan, Wu Peng, Min Chen, T. (2016). Thermal face recognition usingconvolutional neural network. 6-9. 10.1109/OPTIP.2016.7528489.

2. Gabriel Hermosilla, Jose Luis Verdugo, Gonzalo Farias, Esteban Vera,´ Francisco Pizarro, and Margarita Machuca, "Face Recognition and Drunk Classification Using Infrared Face Images," Journal of Sensors, vol. 2018, Article ID 5813514, 8 pages, 2018.

3. 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.

4. S. Sarath and L. R. Deepthi, "Priority Based Real Time Smart Traffic Control System Using Dynamic Background," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0620-0622.