FLANN Based Matching with SIFT Descriptors for Drowsy Features Extraction

Vineetha Vijayan
School of Computer Sciences
Mahatma Gandhi University
Kottayam, India
vineetha.mphilcs5@iiitmk.ac.in

Pushpalatha KP
School of Computer Sciences
Mahatma Gandhi University
Kottayam, India
pushpalathakp@mgu.ac.in

***Abstract*—This paper presents an approach for extraction of drowsy features from face. Drowsiness during driving is one of the major issues of road accident. Driver drowsiness can happen due to fatigue resulting from physical or mental exertion, sedating effects of several medications, drug consumption, melancholy, or may be due to some disorders like obstructive sleep apnea. Drivers who do long distance driving mainly in tedious routes which do not require much driving input are also at a risk of getting sleepy. It is an indispensable task to detect the level of sleepiness in a person by monitoring various factors in intelligent vehicles. The current work combines Fast Library for Approximate Nearest Neighbours(FLANN) feature matching with Scale Invariant Feature Transform(SIFT) descriptors. SIFT has been widely used in face recognition and object detection tasks. SIFT algorithm is considered to be the most impervious to image deformations. The FLANN matcher matches the descriptors of features in a set with the features in the target set. The results show the superiority of FLANN-SIFT when compared with SIFT for drowsy driver detection.**

***Index Terms*—SIFT, FLANN, drowsiness, FLANN-SIFT**

I. Introduction

Drowsy driving is nowadays a very serious concern in the society. National Highway Traffic Safety administration(NHTSA), an agency of Department of Transportation in United States [1] has reported that around one lakh crashes happen in United States in a year due to the drowsy state of driver. In the year 2017, 795 deaths happened due to sleepiness of driver. Around 90,000 motor vehicle crashes happened in the year 2015 involving drowsy driving. As drowsy accidents have become a major contributor to mortality rates, systems that can detect driver drowsiness can help saving lives.

There is a wide range of development occurring in the field of automotive industry. One of the safety features introduced by Mercedes-Benz is Attention Assist. This system uses steering-wheel motions as the indicators of driver tiredness. It takes the blueprint of driving pattern of individual. When strays from these pattern, the system assists. A similar system has been introduced by Volvo which is a Driver Alert Control. The system alerts weary and inattentive drivers by using a vehicle mounted camera which is connected to the Lane Departure Warning System. Another drowsiness detection system has been introduced by Bosch, in which the vehicle steering occupies a sensor from which the system collects data and takes decisions after processing the data. But the usage of these systems is not ubiquitous as these are the systems embedded with luxury cars. Likewise several systems have been developed in the recent past to detect the drowsiness level in a driver.

The measures that can be used to detect driver drowsiness are (a) Physiological measures, (b) Vehicle based measures, (c) Behavioural measures, (d) Subjective measures and (e) Hybrid measures. This study concentrates on the extraction of behavioural measures of the driver, which focuses on eye features and yawning. The duration of eye blinks, the number of blinks, closed eyes and yawning are considered for detection. Frequent eye blinks, reduced duration of eye blinks, closed eyes for seconds, frequent yawning are signs of drowsiness. For capturing the behavioural measures of driver effectively, a FLANN based feature matching with SIFT descriptor is being used. The FLANN feature matcher is optimised for fast nearest neighbour search in huge datasets. This also works for features with high dimensions. It works more quickly than other matching techniques like Brute-Force Matcher for large datasets when fused with SIFT descriptor. The SIFT algorithm [2] proposed by David Lowe extracts distinctive invariant facial features of the driver and FLANN [19] provides feature matching with reference to the training images of the driver. SIFT when combined with FLANN would identify the drowsy facial features from the input images of driver more accurately.

II. Literature Review

Several studies have been done in the area of driver drowsiness detection based on various measures such as Subjective, Vehicle based, Physiological, Behavioural and Hybrid measures.

Hyung-Tak Choi, Moon-Ki Back and Kyu-Chul Lee [3] proposed a drowsiness detection system which is based on Multimodal Deep Learning. The system identifies both the visual facial feature which is a behavioural measure as well as the physiological measures. But the heterogeneity in data has been notified as a problem in the system. So in order to avoid the heterogeneity, a generative model has been used in representation. A Long Short Term memory(LSTM) has also been incorporated in the network to classify the driver's state. Juana, Anelisse, Minica and Juvenal [4] developed a real-time assistance system to examine the drowsiness level of driver.

978-1-7281-0899-5/19/$31.00 ©2019 IEEE

The system focuses on the physiological measures which comprises non-intrusive sensors and a heart rate monitor.

In order to reduce traffic accidents, Melissa, Brian and Natalia [5] in their work realises a system that can capture drowsiness. The work states the causes of driver drowsiness are sleeping hours less than 8, lack of sleep environment, lack of proper work schedule. The various facial changes notified in the study are frequent flicker, side to side head movements and yawning. A real time driver drowsiness detection system using deep neural networks has been developed by Vineetha Vijayan and Elizabeth Sherly [6]. The three deep neural networks namely ResNet50, VGG16 and Inception V3 are individually applied for the drowsy driver detection by capturing the behavioural measures of the driver. Also these three networks have been fused together to form an architecture that accurately measures the drowsiness. A wearable device is proposed by M. Choi et al. [7] which can monitor stress, fatigue and drowsiness conditions of a driver. The physiological and motional measurements are being tracked by the system.

A real-time system for drowsiness detection has been proposed by Isha Gupta, Novesh and Apoorva [8] which uses Histogram Oriented Gradient feature descriptor for face detection with an added feature of altering the threshold frames for face regions such as eyes and mouth. Inclusion of these makes the system more responsive to detection of drowsiness features.

Kyong Hee Lee and Whui Kim [9] has done a study on methods of feature extraction which is used to estimate the driver's level of drowsiness. They proposed a method using OpenCV and dlib for drowsy feature extraction. It considers head position, eye blinking and mouth features for the detection purpose. The correlation among these features have also been presented. But the factors such as refection of light, wearing of glasses and any other kinds of obstructions in front of face are considered as failing factors of the system.

Based on eye analysis, Ursulescu and Llies [10] proposed a drowsiness detection system. The system mainly focuses the eyes and tracks the open and closed states of eyes. While tracking the number of closed frames and when it crosses a threshold, a visual warning will be received by the driver which says that he/she is drowsy. A sensor fusion approach was proposed for drowsiness detection in wearable ultra-low-power systems by Kartsch et al [22]. Higher neural networks [23] can also be used for detection purposes.

Various measures can detect the drowsiness levels of driver. But Physiological measures [12][13] can irritate the driver because to get the physiological measures, some devices need to be weared or induced into the body of driver [11]. So this can make the driver feel uncomfortable during driving. Vehicle based measures are taken by mounting an external sensor or sensors connected to steering of vehicle and these measures are considered for drowsiness detection. But the variation in this vehicle based measures can also be a sign of rash driving. So system can get confused with rash driving and drowsy driving. As the driver have to provide survey regarding his own health to get the subjective measures, he may feel it uncomfortable if he is uneducated or fully unaware of the details or values of various measures to be given. So making a detection system with the subjective measures cannot reach common people. Another type of measure is behavioral measures in which the system tracks the facial expressions of driver and it is from these various facial expressions like eye blinking, yawning etc that the drowsiness level of the driver is calculated. Among all these measures, behavioral measures are considered to be more reliable than others.

III. Proposed Methodology

SIFT [2] is being proposed for extracting drowsy keypoints and descriptors. The drowsy features are extracted from all the faces in the database. When given a new driver face, the system matches and compares it with all the drowsy faces in the dataset using FLANN. The SIFT feature matching works good for scaled images but fails for faces with pose changes [14]. The matching is also scaled by a factor N. An increase in N value causes SIFT to fail. So this paper proposes a feature matching method called FLANN which overcomes these disadvantages of SIFT. The first step resamples the image using bilinear interpolation. This is shown in Figure 1. Gaussian smoothing blurs the succeeding images by reducing the noise and details. Convolutions will be applied to the left neighbours on further images by raising the standard deviation. The images are then downsampled and the convolution operation repeats until the images are too small to process. So each row in the figure is called an octave. A scale space is now constructed and the representation is normalized. Then Laplacian of Gaussian(LoG) has to be computed where SIFT takes the extrema of the Laplacian [20].

The input image description in varying scales can be defined by equation(1)

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where I(x,y) is the input image and G(x,y,σ) is the input image description. With the Gaussian Expression,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2}$$

The image obtained after convolution is the Gaussian image. (x,y) are the pixel coordinates of the input image and σ, the scale factor. LoG images are not scale invariant and dependent on the blur. If the scale, σ^2 in the denominator part of the LoG is removed, scale invariance can be achieved.

Laplacian of Gaussian can be represented as $\bigtriangledown^2$G, and thus the scale invariant LoG will be $\sigma^2\bigtriangledown^2$G, which becomes more complicated. But these complications can be avoided by Difference of Gaussian(DoG) operation as the results of DoG images are already come out with σ^2 multiplied. Any additional constituents coming in the algorithm can be avoided because the algorithm looks only for maximums and minimums in the images. Any further addition of constant

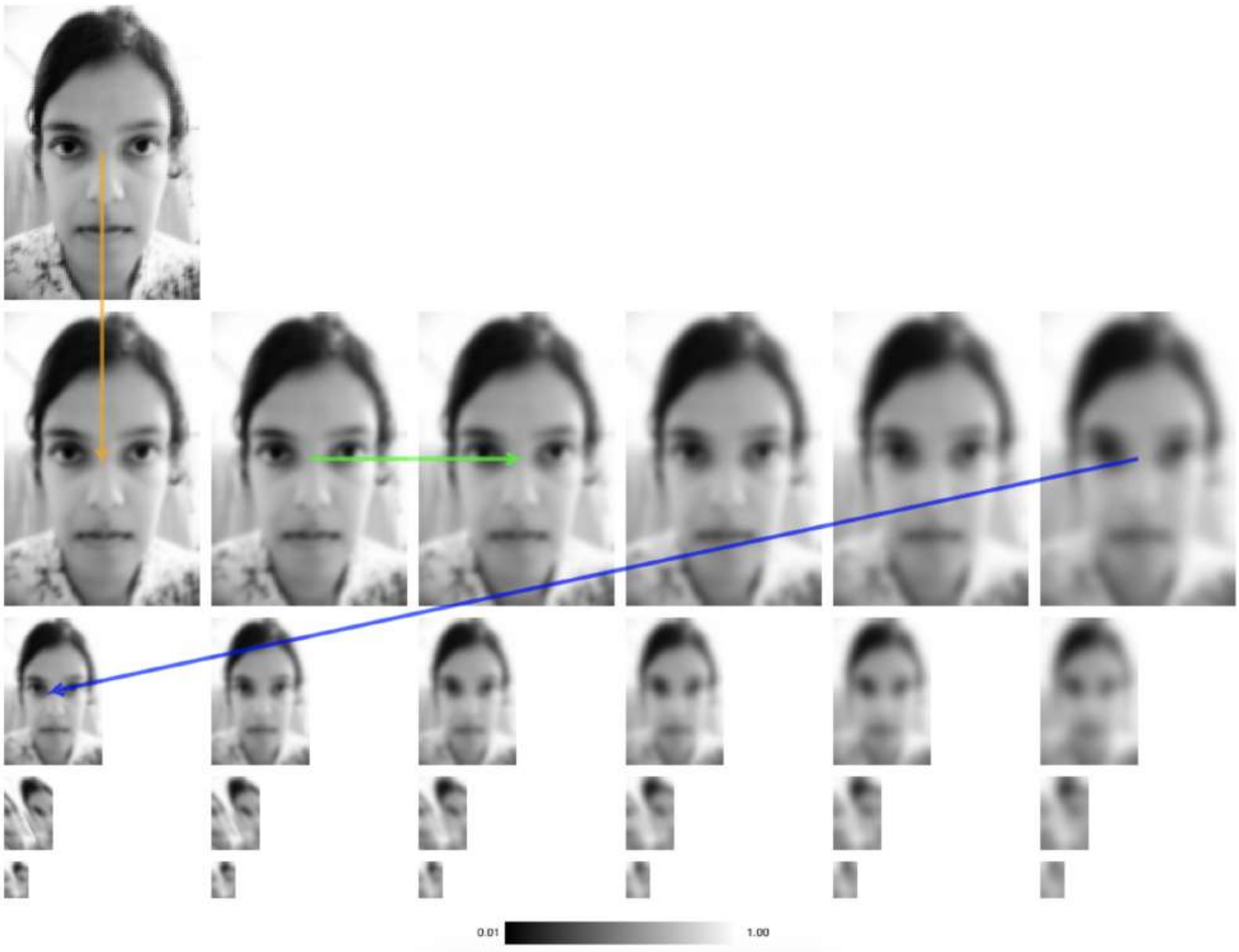


Fig. 1: Interpolation and Convolution Operations performed on Input image with dimension 200x300

values make the maxima and minima stay at the same location. The Difference of Gaussians have been found as in Figure 2 by subtracting two consecutive images and are shown in the successive rows. The extrema points now turned out are refined by Taylor expansion of the scale space function. It can be expressed as

$$D(x) = D + \frac{\delta D^T}{\delta x} x + \frac{1}{2} x^T \frac{\delta^2 D}{\delta x^2} x \tag{3}$$

The extrema points are found out from this equation which then forms the subpixel keypoints. These values matches the features more accurately and increases the stability of the algorithm.

FLANN matches a feature with another feature when the distance to that feature is less than a certain threshold value of the distance to the next nearest feature. This can reduce the false feature matchings and this is better done with FLANN as compared to SIFT.

The proposed system is trained on a subset of ImageNet database, which is used in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [16]. The dataset contains thousands of face images and can classify images into various categories. Physical attributes of the dataset includes variety in skin tone, fatigue, facial structure, clothes and hair styles. Hence, the model has learned rich feature representations for a wide range of images. It helps in this scenario for feature extraction which is robust to various backgrounds.

The proposed system also uses DrivFace database [14] that contains face images of various subjects while driving in real scenarios. The dataset contains 606 samples of 640x480 pixels each, obtained by capturing real time driving scenarios from 4 drivers (2 women and 2 men) with several facial features like glasses and beard. Each image has its gazes in three directions. First one is the ”looking-right” class. The looking right direction has the head angles between -45 and -30. The second one is the ”frontal” class which consists of head angles between -15 and 15. ”Looking-left” is the third kind of class that contains the head angles between 30 and 45. Thus the system for drowsiness detection considers robustness and efficiency in all circumstances.

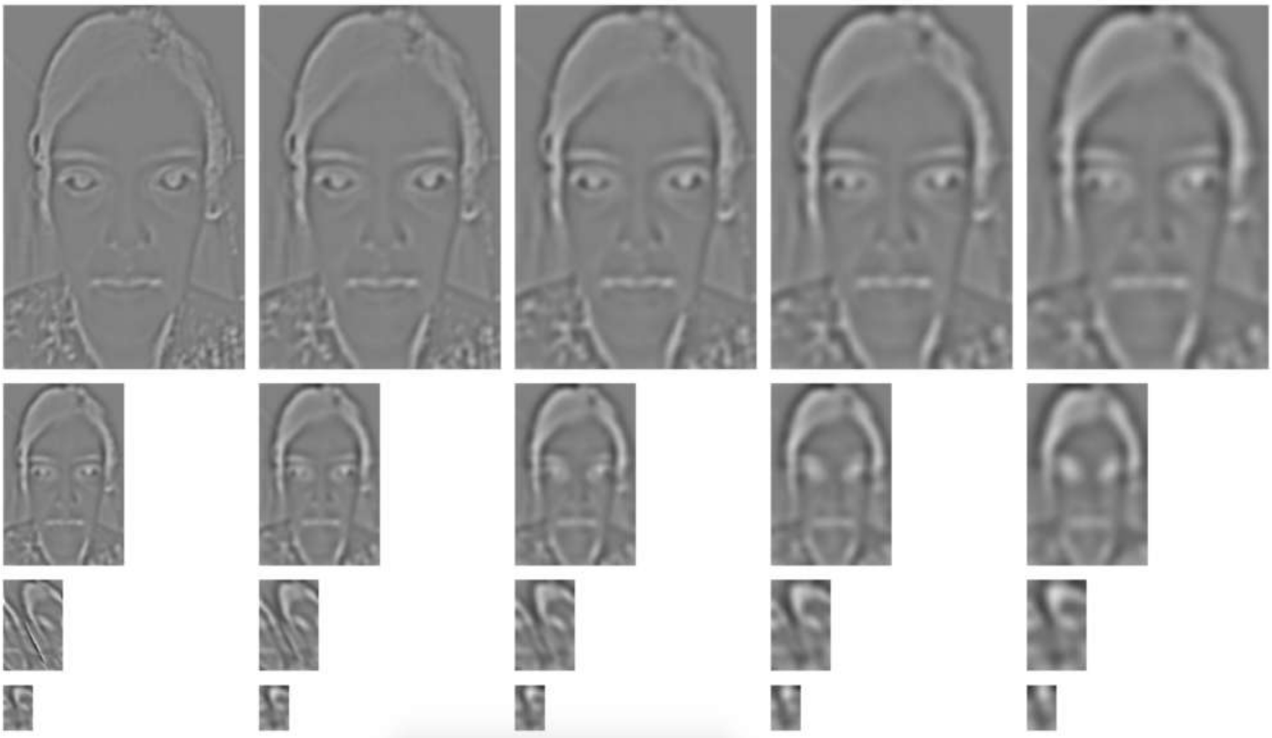
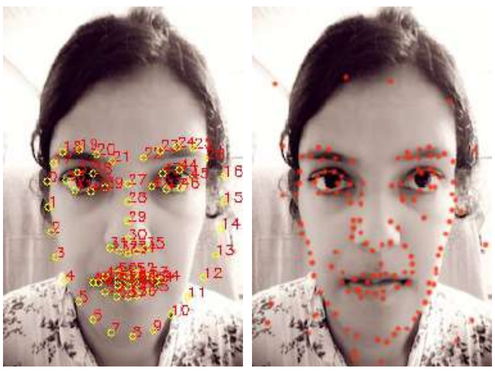


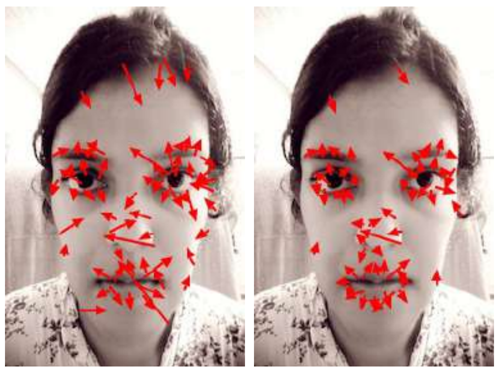
Fig. 2: DoG computation of individual pixel

Figure 3(a) shows the detected facial landmarks and 3(b) shows the features extracted by SIFT algorithm. 68 keypoints have been extracted that maps to specific spatial structures. Of that, 36-41 are the landmarks of right eye, 42-47 are the left eye points, 48-60 are the mouth outline points, 61-67 are the mouth inner points. These are the drowsy features considered for extraction. Figure 4(a) shows the extracted keypoints after applying a threshold value of Th=0.5 and Figure 4(b) illustrates the keypoints after applying Threshold value of Th=0.2. The number of keypoints reduces to more valid keypoints when the threshold value reduces from 0.5 to 0.2.



(a) Facial Landmarks (b) Feature detection

Fig. 3: Extracted facial features using SIFT



(a) Th=0.5 (b) Th=0.2

Fig. 4: Applying Threshold(Th) to SIFT features

There are certain mismatches found when the same person drives with different head orientation. Only a very less key-points are matched. A small change in the head orientation can match almost all SIFT features but when the orientation exceeds 45 degrees, matching becomes so poor [15]. Figure 5(a) shows the SIFT matching for the same person at the same pose and Figure 5(b) at varying poses. The pose in this current system refers to the head orientation.

Two dictionaries need to be passed for FLANN based matcher. The first dictionary states the required algorithm. The current system passes the kd-tree algorithm [21]. The second dictionary states the search parameters to be passed. This dictionary states the number of times the trees in the index have to be recursively traversed. Large values of parameters can provide better precision. But as the value of parameters increases, the time required also increases. The time consumption taken for processing is more in higher dimensional spaces for algorithms based on nearest neighbor approximations. No perfect methods are discovered to solve these kinds of higher dimensional problems. Still for the better performance, kd-tree algorithm has been used in the current system. The classical kd-tree algorithm proposed by Freidman [17] has some kind of inefficiencies such that the performance reduces with the increase in dimension. With lower dimensions, this algorithm works better. In order to speedup the search process, Silpa-Anan [18] proposed an improved version of the kd-tree algorithm in which it creates multiple randomised kd-trees. These randomised kd-trees have been used in the current system to increase the efficiency of the system by working with higher dimensions. As the number of randomised trees increases, performance increases. The nearest neighbour parameters, denoted as k, from kd-tree are passed for drowsy image detection. False matches are then reduced using the Lowe's distance ratio test. The distance ratio of extracted keypoints are calculated and when it goes below a threshold, it is considered as a good match.

Experiments and Results

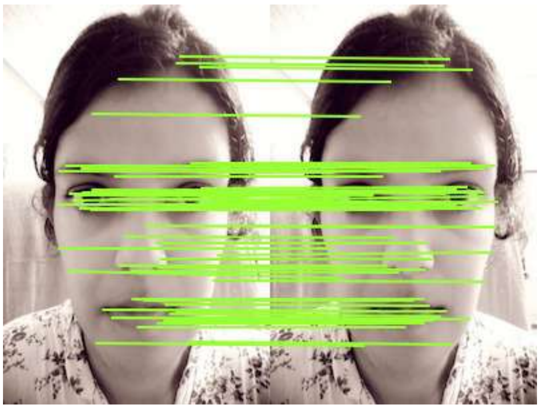
The proposed system is tested by providing images of drowsy drivers with different drowsy states and different facial situations on two different datasets. The different drowsy states include eye blinking and yawning. The different facial situations include bare faces, faces with glasses, barefaces at night, faces with glasses during night and faces with various head orientations. In order to evaluate the feature matching of FLANN in drowsy drivers, our work initially does the feature matching using SIFT. Matching has been done and the considered values are for the threshold 0.2, 0.3, 0.4, 0.5 and 0.6. As more valid keypoints are obtained with reduced threshold values, we have considered the threshold of 0.2 in feature matching. With Th=0.2, the SIFT matches 13 eye features(both eyes included) whereas FLANN matches exactly 12 keypoints with 6 keypoints for left eye and 6 keypoints for right eye. When the threshold value increases from 0.2 to 0.6, at Th=0.6, the number of features matched by SIFT algorithm for eye region is 24 and the number of features matched by FLANN is 15. The number of features matched by FLANN for eye region at Th=0.6 is almost nearest to the number of SIFT feature points of eyes when Th=0.2. Table I shows the number of SIFT matching points against various threshold values (0.2, 0.3, 0.4, 0.5 and 0.6) and Table II shows the number of FLANN matching points against the same threshold values. The tables provide a comparison between SIFT and FLANN feature matching and is better done with FLANN. For both the feature matchers, the threshold value that can be taken into account is 0.2 as this smaller threshold value provides more accuracy to feature point matching.

TABLE I. Number of SIFT matching points with various threshold(Th) values

Threshold(Th) Values	Th=0.2	Th=0.3	Th=0.4	Th=0.5	Th=0.6
Eyes	13	14	17	19	24
Nose	10	13	16	21	24
Mouth	23	26	28	32	35
Other Face Region	32	35	40	43	47

TABLE II. Number of FLANN matching points with various threshold(Th) values

Threshold(Th) Values	Th=0.2	Th=0.3	Th=0.4	Th=0.5	Th=0.6
Eyes	12	12	14	14	15
Nose	9	10	10	12	12
Mouth	20	20	24	24	25
Other Face Region	27	27	29	30	30



(a) Matching for the same person at the same pose

(b) Matching for same person at various poses

Fig. 5: Matching results of SIFT in a realtime scenario.

Figure 6 shows the matching results of FLANN matcher for same person at different pose. It can be found from this matching results that the false matches are reduced with FLANN as compared to SIFT.

The accuracy of the SIFT feature extractor on drowsy images is 85.714%. As the feature matching have been done with both SIFT and FLANN-SIFT, the accuracy measures obtained

TABLE III. Accuracy measurements of the methods of feature extraction and matching

Methods Used	Method of Extraction	Accuracy (Extraction)	Method of Matching	Accuracy (Overall)
SIFT-SIFT	SIFT	85.714%	SIFT	85.318%
SIFT-FLANN	SIFT	85.714%	FLANN	93.412%

for SIFT feature extraction and SIFT matching(SIFT-SIFT Method) is 85.318% where as the accuracy obtained for SIFT feature extraction and FLANN feature matching(FLANN-SIFT) is 93.412%. Table III illustrates the accuracy measure-ments of the extraction and matching methods. This accuracy is measured by considering the threshold value of 0.2.

Fig. 6: Matching results of FLANN for the same person at different poses in a realtime scenario.

Conclusion

This paper proposes a new method for drowsy feature extraction that have been developed by combining two previous approaches such as SIFT and FLANN. The drowsy features are extracted from face using SIFT feature extraction technique and the extracted features are matched using FLANN. The feature matching of SIFT has been compared with FLANN. Experimental results show that better features are extracted and matched with the FLANN-SIFT combination. It has also been demonstrated that the FLANN with SIFT approach reduces the false matches. One of the challenges that still exists in drowsy feature extraction is when the driver uses sunglasses. Addressing this issue can contribute as an extension to this work. In all other facial situations and for changes in illumination, the current system performs well. Thus the system for drowsy feature extraction exhibits robustness and efficiency in all circumstances.

Acknowledgement

Authors acknowledge the support extended by DST-PURSE Phase II, Govt of India

References

[1] Drowsy Driving NHTSA reports. Retrieved from https://www.nhtsa.gov/risky-driving/drowsy-driving, 08 January 2018

[2] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Gonput.Vis. 60(2), pp. 91-110, 2004.

[3] Choi, Hyung-Tak, Moon-Ki Back, and Kyu-Chul Lee. "Driver Drowsiness Detection based on Multimodal using Fusion of Visual-feature and Bio-signal." 2018 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2018.

[4] Pz, Juana Guadalupe Herrera, et al. "Electronic Device to Monitor Drowsiness." 2018 XIV International Engineering Congress (CONIIN). IEEE, 2018.

[5] Yauri-Machaca, Melissa, et al. "Design of a Vehicle Driver Drowsiness Detection System Through Image Processing using Matlab." 2018 IEEE 38th Central America and Panama Convention (CONCAPAN XXXVIII). IEEE, 2018.

[6] Vijayan, Vineetha, and Elizabeth Sherly. "Real time detection system of driver drowsiness based on representation learning using deep neural networks." Journal of Intelligent Fuzzy Systems Preprint, pp. 1-9, 2019.

[7] M. Choi, G. Koo, M. Seo and S. W. Kim, "Wearable Device-Based System to Monitor a Driver's Stress, Fatigue, and Drowsiness," in IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 3, pp. 634-645, March 2018.

[8] Gupta, Isha, et al. "Real-Time Driver's Drowsiness Monitoring Based on Dynamically Varying Threshold." 2018 Eleventh International Conference on Contemporary Computing (IC3). IEEE, 2018.

[9] Lee, Kyong Hee, Whui Kim, Hyun Kyun Choi, and Byung Tae Jang. "A Study on Feature Extraction Methods Used to Estimate a Driver?s Level of Drowsiness." In 2019 21st International Conference on Advanced Communication Technology (ICACT), pp. 710-713. IEEE, 2019.

[10] Ursulescu, Oana, Betina Ilie, and Georgiana Simion. "Driver Drowsiness Detection Based on Eye Analysis." 2018 International Symposium on Electronics and Telecommunications (ISETC). IEEE, 2018.

[11] Mahmoodi, Mohammad, and Ali Nahvi. "Driver drowsiness detection based on classification of surface electromyography features in a driving simulator." Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 233.4 : 395-406, 2019.

[12] Nadrag, Carmen, Vlad Poenaru, and George Suciu. "Heart Rate Measurement Using Face Detection in Video." 2018 International Conference on Communications (COMM). IEEE, 2018.

[13] Poorna, S. S., V. V. Arsha, P. T. A. Aparna, Parvathy Gopal, and G. J. Nair. "Drowsiness Detection for Safe Driving Using PCA EEG Signals." In Progress in Computing, Analytics and Networking, pp. 419-428. Springer, Singapore, 2018.

[14] Diaz-Chito, Katerine, Aura Hernez-Sabatnd Antonio M. L. "A reduced feature set for driver head pose estimation." Applied Soft Computing 45, pp. 98-107, 2016.

[15] Bhattacharya, Bhaskar, and Minhui Zhu. "Handling pose variation in face recognition using SIFT." 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2011.

[16] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) "ImageNet Large Scale Visual Recognition Challenge." IJCV, 2015.

[17] Friedman, Jerome. "An algorithm for finding best matches in logarithmic expected time." No. SLAC-PUB-1549. SLAC National Accelerator Lab., Menlo Park, CA (United States), 2018.

[18] Silpa-Anan, Chanop, and Richard Hartley. "Optimised KD-trees for fast image descriptor matching." 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008.

[19] Marius Muja and David G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", in International Conference on Computer Vision Theory and Applications (VISAPP'09), 2009

[20] Otero, Ives Rey. "Anatomy of the SIFT Method." Diss. 2015.

[21] Emiris, Ioannis Z., and Dimitri Nicolopoulos. "Randomized kd-trees for approximate nearest neighbor search." CGL-TR-78, NKUA, Tech. Rep. (2013).

[22] Kartsch, Victor Javier, et al. "A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems." Information Fusion 43 :pp. 66-76, 2018.

[23] Sarat Chandra Nayak and Mohd Dilshad Ansari *, "COA-HONN: cooperative optimization algorithm based higher order neural networks for stock fore casting", Recent Advances in Computer Science and Communications (2020) 13: 1. https://doi.org/10.2174/2213275912666190902121153