Integration of Driver Behavior into Emotion Recognition Systems: A Preliminary Study on Steering Wheel and Vehicle Acceleration

Sina Shafaei(✉), Tahir Hacizade(✉), and Alois Knoll(✉)

Technical University of Munich, Munich, Germany
{sina.shafaei,tahir.hacizade}@tum.de, knoll@in.tum.de

Abstract. The current status of the development for emotion recognition systems in cars is mostly focused on camera-based solutions which consider the face as the main input data source. Modeling behavior of the driver in automotive domain is also a challenging topic which has a great impact on developing intelligent and autonomous vehicles. In order to study the correlation between driving behavior and emotional status of the driver, we propose a multimodal system which is based on facial expressions and driver specific behavior including steering wheel usage and the change in vehicle acceleration. The aim of this work is to investigate the impact of integration of driver behavior into emotion recognition systems and to build a structure which continuously classifies the emotions in an efficient and non-intrusive manner. We consider driver behavior as the typical range of interactions with the vehicle which represents the responses to certain stimuli. To recognize facial emotions, we extract the histogram values from the key facial regions and combine them into a single vector which is then used to train a SVM classifier. Following that, using machine learning techniques and statistical methods two modules of abrupt car maneuvers counter, based on steering wheel rotation, and aggressive driver predictor, based on a variation of acceleration, are built. In the end, all three modules are combined into one final emotion classifier which is capable of predicting the emotional group of the driver with 94% of accuracy in sub-samples. For the evaluation we used a real car simulator with 8 different participants as the drivers.

Keywords: Emotion recognition · Multimodality · Driver behavior

1 Introduction

The recognition of emotions has wide implications in the different applications and recently got the attention of the researchers in the field of automotive for the purpose of driver fatigue detection [9,41,45,54], human-car interaction [9,29] and

G. Carneiro and S. You (Eds.): ACCV 2018 Workshops, LNCS 11367, pp. 386–401, 2019.
https://doi.org/10.1007/978-3-030-21074-8_32

respectively the highly and fully autonomous driving scenarios [21,37]. According to the *7-38-55* Rule, 93% of human communication is performed through nonverbal means, which consist of *facial expressions*, *body language* and *voice tone* [38]. Therefore, a system which aims at automatically analyzing the emotions of humans should mainly focus on these non-verbal channels. This research field respectively is called *Affective Computing* which is an emerging research field in enabling intelligent systems to recognize human emotions. The main challenges in automated affect recognition are *head-pose* variations, *illumination variations*, *registration errors*, *occlusions*, *identity bias* and (subject-independent affect recognition) [47]. The most common approach in the field of affective classification are multimodal methods. The general aim of multimodal fusion is to increase the accuracy and reliability of the estimates. Based on empirical studies and statistical measures, multimodal systems were consistently more accurate than their unimodal counterparts, with an average improvement of 9.83% (median of 6.60%) [14,43]. Fusion of multiple modalities into one single final output is a challenging task. The right fusion method highly depends on the underlying data. Common fusion techniques in the field of affective computing are *feature-level fusion*, *kernel-based fusion*, *model-level fusion*, *score-level fusion*, *decision-level fusion*, and *hybrid* approaches [14,43,44,53]. The most common fusion techniques are feature-level fusion and decision-level fusion. In feature-level fusion the data from separate modalities are first aggregated and then used as a single input into one model. In decision-level fusion each modality has its own trained model, and the predictions are then combined to a single output.

Moreover, most of the state-of-the-art systems have high complexities and are mostly benchmarked in ideal environments and on powerful computers with access to Graphics Processing Units (GPUs) [10,14,39,43]. This limits the applicability of using such systems for in-cabin environments due to the existing limitations regarding computation power in such environments. Considering this, no publicly accessible research, to this date, aims at enabling automated, robust affect recognition for car-drivers on embedded devices. In this work, we propose a system which is designed exclusively for in-cabin environments to tackle the important challenges of affect recognition in automotive and help to increase the emotional awareness of in-cabin environment by investigating the behavior-related modalities of the driver through monitoring the steering wheel for performed maneuvers and the change in acceleration of the car. We study **(RQ1)** the effects of emotions on behavior of the driver while driving, **(RQ2)** modeling the change in acceleration and steering wheel usage by the driver, according to the current emotional status and **(RQ3)** will try to demonstrate the benefits of multimodality in emotion recognition systems for robust predictions.

In the following remainder, we discuss in Sect. 2 relevant state-of-the-art methods in affective computing. Based on this knowledge, we tackle the aforementioned challenges of automatic emotion recognition for in-cabin environments with the help of our proposed approach at Sect. 3. In Sect. 4 we represent the integration of the considered modalities into our system beside the achieved results, and finally, in Sect. 5 we draw an outlook and list the open directions and challenges for future studies.

2 Related Works

Ekman and Friesen [17] proposed 6 universal human emotions (anger, disgust, fear, happiness, sadness, and surprise) and respectively developed the Facial Action Coding System (FACS). Their system interprets one or more facial muscle as Actions Units (AU) and the predefined combination of AUs represents one of the emotions. FACS acts as a basis for most of the automatic facial emotion recognition systems. Similarly, EMFACS (Emotional Facial Action Coding System) [18] which includes only emotion-related action units, is among the widely used systems. Alshamsi et al. [2] proposed a method for real-time emotion recognition on mobile phones. The system has two main parts as feature extraction and emotion classification. In order to extract features, authors label every pixel of a frame as 0 or 1 based on the intensity of pixels using BRIEF binary descriptor [7] a.k.a appearance-based approach. BRIEF differs from its competitors with higher recognition rates. Feature extraction step, can process 15 frames/second on mobile phones and as a result generates the 256-bit feature vector. Authors used the K-Nearest Neighbor (KNN) classification algorithm which uses Euclidean distance to compute the nearest point and requires minimal data distribution information. The trained system could achieve around 85% accuracy on a classification of 6 basic emotions using Cohn-Kanade dataset. Performance of BRIEF feature extractor outperforms previously used methods as *Similarity Normalized Shape Features (SPTS)* and *Canonical Normalized Appearance (CAPP)* [52]. Combination of SPTS and CAPP achieves 83% average detection rate which is still lower than BRIEF.

The majority of the researches in this field are focused on emotion recognition using a combination of several modules, but none of them considered classifying the emotions of driver based on his behavior during the ride. Behavior can be defined as the typical velocity and movements of body parts which then may be used to imply emotions as interest or boredom [4]. Some works with the help of biometric devices try to evaluate the stress, and the distraction level of the driver with the help of biometric devices like electroencephalogram (EEG) checkers [16,23]. Kamaruddin and Wahab [28] investigated the correlation between drivers speech and behavior in order to build a driver emotional indicator system. Their main goal was to identify the behavioral state of the driver through speech emotion recognition. For this purpose, they use the *Mel Frequency Cepstral Coefficient (MFCC)* [8] to extract features from the audio signals of a driver. Then these features were used to classify speech into 4 states of driver behavior, specifically as *talking*, *sleeping*, *laughing* and *neutral*. As a classifier, three different methods of *Multi-Layer Perceptron (MLP)* and a fuzzy neural network such as *Adaptive Neuro-Fuzzy Inference System (ANFIS)* and *Generic Self-organizing Fuzzy Neural Network (GenSoFNN)* were considered in their work. Dataset to train models was built from the mix of *Real-time Speech*

Driving Dataset (RtSD) [32], *Berlin Emotional Speech Database (Emo-DB)* [5] and *NAW* dataset [27]. Unlike the previous works, they not only consider words and speech but any vocal voice generated by the driver. A best example is laughing which is the main sign of being happy. There are various implementations of MFCC, but according to the comparison performed by Ganchev et al. [20] and Slaney's implementation [48] considerable improvement in performance was observed by integration of speech modality. However, this approach was not considered as an independent system in emotion recognition, but rather as an add-on to already existing functionality. This was due to the fact that speech is not a regular act while driving especially when the driver is alone in the car. Another complementary module to facial emotion recognition is the identification of head movements. The work of Samanta and Guha [46] shows an impact of head motions in affective recognition. In order to analyze head motions, the authors used *Acted Facial Expression in the Wild (AFEW)* dataset [13] which includes small video clips from 54 movies. All clips are labeled with seven basic emotion categories. After splitting these clips into frames, they consider the head as a 3-D rigid body which is defined by Euler angles (pitch, yaw, and roll) and use *incremental face alignment* method [3] to detect the head in each frame. As the first step of head motion inspection, *root-mean-square (RMS)* values of angular displacement, angular velocity, and angular acceleration are considered. For each emotion class, the nine-time series of RMS values are calculated. Their results show that *anger*, *joy*, and *neutral* states were more distinguishable from the rest. RMS measurements of *sadness* and *surprise* were very similar. In the next stage, they evaluated the impact of RMS measurements in emotion classification. As a classifier k-Nearest Neighbor method was used. After 10-fold cross-validation, overall accuracy of around 34% was achieved which was two times more than random guessing. It was also interesting for the authors to know whether head motions hold information complementary to facial expression. Experiments on AFEW dataset show that accuracy is increased 10% when RMS measurements of head motions are used jointly with facial expression. Findings of Hammal et al. [24] demonstrate that angular velocity and acceleration of head movements are considerably higher during negative affect relative to positive one. Behoora and Tucker [4] were able to find a link between body language and emotional state. By using non-wearable sensors, they tracked human skeletal joints and calculated velocity beside acceleration for each joint. Afterwards, machine learning techniques are applied to quantify body language states. From the perspective of methods, achieved results by previous experiments depict that during classification stage, decision tree based methods (like random forest and IBK) outperform others (like Naive Bayesian and C4.5) and can achieve up to 99% accuracy in recognition of 4 emotions. Vehicle acceleration and deceleration events also have been investigated in similar researches where the aim was to identify the driver and predict the age group [19,34].

The steering wheel is a part of the steering system which is manipulated by the driver. Zhenhai et al. [55] were able to detect driver drowsiness after analyzing time series of steering wheel's angular velocity. As a starting point, steering behavior under the fatigue state was inspected. Then after collecting some portion of data, the sliding window technique was applied and respectively the drowsiness state was recognized after examining high fluctuations in a temporal window. Moreover, the likelihood of eyes holding information about the behavior and emotional state was investigated by Cabrall et al. [6]. With the help of eye tracker apparatus, the authors tried to evaluate the distraction, drowsiness, and cognitive load of the driver during automated driving.

3 Proposed Approach

In order to identify the emotional state, we consider behavior-related data and facial expressions as the sources of input. The modality of facial expression is used as our main input to classify the emotions of driver and in order to increase the confidence and robustness of the predictions, we incorporate the behavior-modality. We model all modalities separately, meaning we use decision-level fusion in order to combine their outcomes. Therefore, each modality needs to be trained separately and then the outcomes need to be weighted and combined together. This architecture is visualized at Fig. 1. The facial expressions is represented by the node *Facial Expressions*. We evaluate this modality by an individual module of *Facial Modality*. The behavior-modality is composed of *Steering Wheel Usage*, and *Change in Acceleration* signals. The variables represent input data to the second module of *Behavior Modality*. Both modules analyze the emotional state individually and are fused on the decision-level to output the final classification of the emotional state.

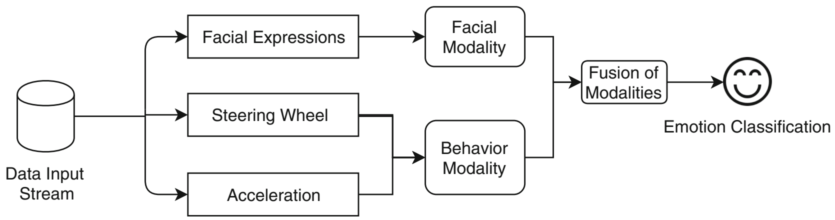


Fig. 1. Different modalities of the proposed system

3.1 Experimental Set Up

We set up a real car simulator for testing and evaluating our approach. On the simulator, driver has almost 180° of view and can easily watch over the cars in other roads at an intersection using the side monitors. Car controlling systems as the steering wheel, accelerator, brake, and transmission are almost identical in size and way of functioning with a real car. Sound system installed in the cabin also helps to convey a real sense of driving. During experiments, 15 drivers have participated, each of drove 36 min on average. Drivers got a chance to drive one route three times. Their attitude was different due to a prehistory condition narrated to them and the predefined situations on the road. Real emotional states of the drivers were collected using a questionnaire organized right after each ride. During the rides, drivers were being recorded using an external camera in a resolution of 960×720 and all other internal vehicle parameters as speed, acceleration and steering wheel position were collected by *virtual test drive* software of the simulator.

3.2 Facial Modality

To detect the *Region of Interests* (ROI) we used a facial landmark detector as designed by Kazem et al. [30]. After that, we used *Histogram of Oriented Gradients* (HOG) descriptors by applying a fixed sized sliding window over an image pyramid build upon them. The normalized HOG orientation features make this method capable of reducing false positive rates far better in comparison with *Haar Wavelet-based Detectors* [11]. Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbours (KNN) algorithms are most widely used supervised machine learning techniques for classification tasks. Our model was built based on linear kernel SVM with decision function of *One-vs-Rest*. In order to train our model, k-Fold Cross Validation method was used with a ‘k’ set to 10. The respective algorithm is represented at Algorithm 1. Facial data is represented by the aggregation of *extended Cohn-Kanade* (CK+) [35] and *Japanese Female Facial Expression* (JAFFE) [36] database.

Algorithm 1. The Facial Emotion Classifier

```
featureVector ← init list
SVMClassifier ← load model
while newFrame is exist do
    frame ← FetchVideoStream()
    grayFrame ← GrayscaleImage(frame)
    if faceTracker(grayFrame).Score < threshold then
        face ← detectFace(grayFrame)
    else
        face ← faceTracker(grayFrame).Position
    end if
    ROIarray ← FetchROI(face)
    for each ROI in ROIarray do
        hog ← HOGDescriptor(ROI)
        featureVector ← featureVector + hog
    end for
    result ← SVMClassifier(featureVector)
end while
```

3.3 Behavior Modality

Steering wheel angular velocity is one of the primary factors which was considered to be collected during all rides. Due to the fact that some situations like avoiding obstacles may require a sharp steering movement, we should observe the driver usage of steering wheel for some predefined period of time. Later, based on collected initial data, the average of angular velocity in steering wheel is calculated for small fraction of time. This averaged value will be used later as the threshold. Our aim is to detect angular velocity higher than dynamically defined threshold and mark this fraction of time as abnormal and when there is consecutively several abnormal values, the system will identify the emotional state of the driver accordingly. Another measurement factor which infers information about the behavior of driver during a ride, is the intensity of the changes in vehicle acceleration. In order to combine the modalities we use decision-level fusion as it was mentioned earlier. However, since not all of emotions extracted by facial expressions can be reflected in behavior modality, there is a need for high level representation of the emotional state of the driver. This is commonly performed by representing the emotional state with *valence and arousal* [31] as depicted at Fig. 2. Valence is positive or negative affective [22] and defines the description level on a scale from pleasantness/positive emotions to unpleasantness/negative emotions. Similarly arousal measures, is an indicator of how calm or excited the subject is and implies reactiveness of the subject to a stimuli [31].

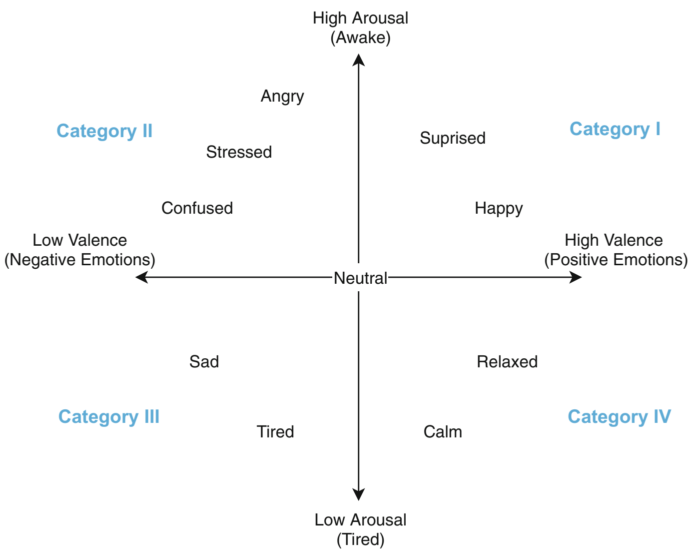
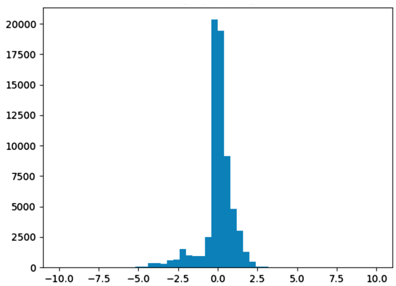
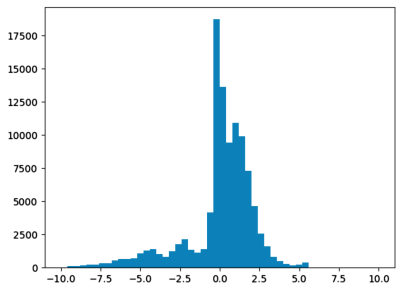


Fig. 2. Different categories of arousal-valence graph

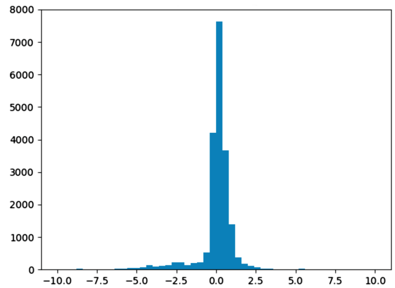
Generally speaking, drivers tend to drive more actively and make more often abrupt movements when they are *angry*, *happy* or *excited*. Respectively their driving behavior becomes more passive, and most likely don't make eye/body movements as actively as they do when they are *tired* or *sad*. These emotions can exactly match with the emotional categories depicted at Fig. 2. Here we achieve our main pattern based on the fact that active and aggressive driving skills are related to category I and II, whereas category III represents passive driving behavior and category IV depicts a neutral state of the driver.



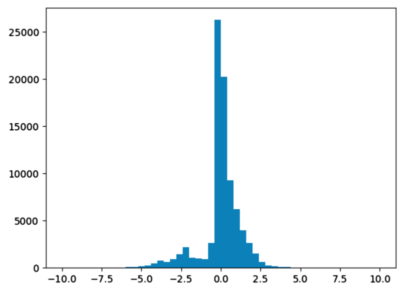
(a) Category I



(b) Category II



(c) Category III



(d) Category IV

Fig. 3. Frequency distribution of vehicle acceleration in 4 groups of emotional status

In order to compare the vehicle acceleration distribution during different emotional states, the histogram of frequency distribution is plotted at Fig. 3. The vehicle acceleration collection for category I, III and IV are very similar and do not hold much sign of distinction. On the other hand, Fig. 3b shows a wider range of values (–10 to +6) for category II which indicates the angry/stressed/confused drivers tend to accelerate and decelerate faster. We construct a single decision tree using 50 samples and export its flowchart tree at Fig. 4. The result reveals a considerable impact of the vehicle acceleration in prediction of emotions from category II. All 18 samples of category II are grouped using only one condition from vehicle acceleration (VA) which shows the important role of this module in prediction of *anger*, *nervousness* and *stress* of the drivers. The second condition in the decision tree uses the proportion of *sadness* emotion felt by the driver. This feature helps to group all 5 samples from category III. After this step, only samples from category I and IV are left un-grouped. For this purpose, conditions formulated by steering wheel (SW) rotation and *happy*, *neutral* and *sadness* features from the facial expressions module, are considered together. The feature vector is generated from the output of the three modules. The first and second values refer to VA and SW modules, and respectively last 7 values of the vector represent the output of facial expressions module which all are depicted at Table 1. The frequency for generating the feature vector is fixed to 2 min. For this period of time, SW parameter counts a number of abnormal steering wheel rotations, VA parameter makes a decision whether the driver is stressed or not and facial expression parameter counts the occurrences of the driver feeling each one of the seven basic emotions and gets normalized accordingly.

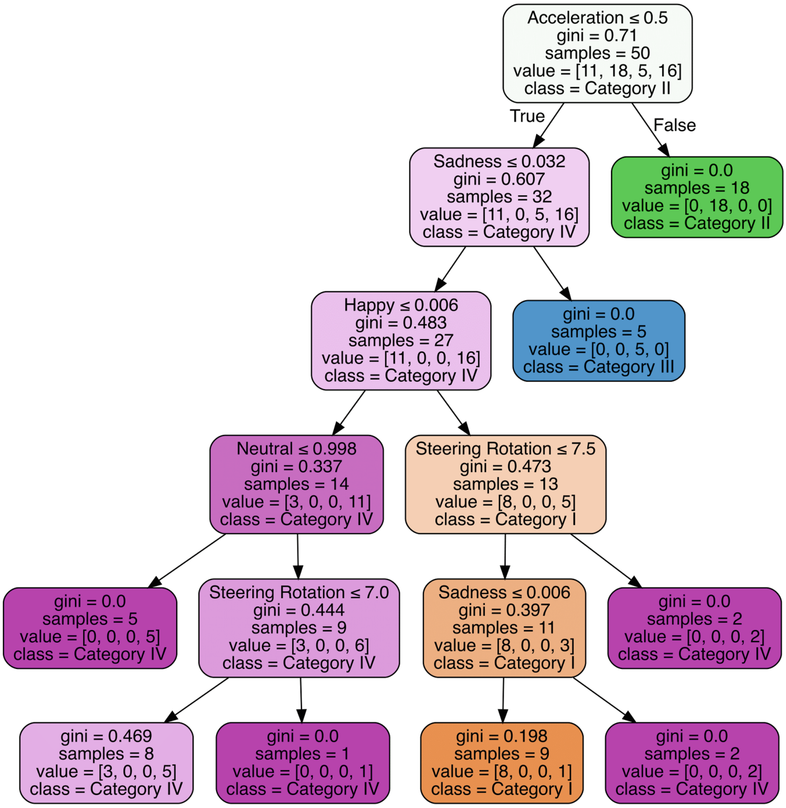


Fig. 4. Decision tree of combining the 3 different modules

After analyzing a single decision tree, we use the same feature vectors from 50 samples to train a random forest classifier (collection of decision trees), in order to achieve higher accuracy while increasing the robustness of the model. Hyper-parameters of the random forest are tuned using grid search and we set the *minimum sample lead* to 2, *maximum depth* to 2, *number of estimators* to 6 and set the *maximum features* variable to 'auto'.

Table 1. Feature vector of final emotion classifier

Module	Vector index	Parameter name	Value
VA	1	Acceleration	0 or 1
SW	2	Steering rotation	0 to ∞
Facial expression	3	Neutral	0 to 1
	4	Anger	0 to 1
	5	Disgust	0 to 1
	6	Fear	0 to 1
	7	Happy	0 to 1
	8	Sadness	0 to 1
	9	Surprise	0 to 1

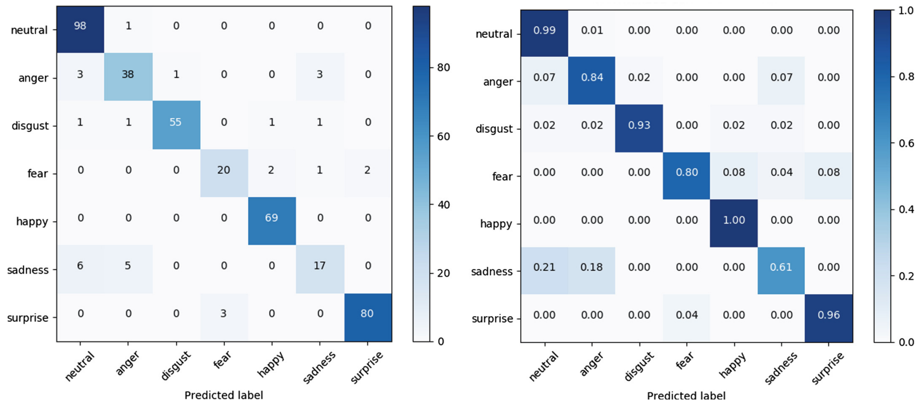
4 Evaluation and Results

Our proposed method for facial expression-based emotion recognition achieves 93% of accuracy after 10 fold cross-validation. Comparison of achieved result with the state-of-art methods which similarly tested on CK+ dataset is shown at Table 2. Obtained accuracy is higher than most of the previously proposed methods and only 2% less than work of Khan et al. [33], and Donia et al. [15].

Table 2. Comparison of different facial expression-based methods on CK+

Authors	Method	Accuracy
Cohn and Kanade et al. [35]	Active Appearance Models	83%
Alshamsi et al. [2]	BRIEF Feature Extractor	89%
Swinkels et al. [49]	Ensemble of Regression Trees	89.7%
Ouellet [42]	Convolutional Network	94.4%
Khan et al. [33]	HOG-based	95%
Donia et al. [15]	HOG-based	95%
Our Method	**HOG on ROI regions**	**93%**

Confusion matrix of our proposed approach at Fig. 5, demonstrates a nearly perfect performance in detection of *happiness* (100%), *surprise* (96%) and *disgust* (93%). A human face is mostly in a neutral state which is also true in driving situations, therefore it is important to detect the *neutral* state accurately, and in this case, our method achieves 99% of positive predictions. There is slightly



(a) Without normalization (b) With normalization

Fig. 5. Confusion matrix of facial-based emotion recognition classifier

a considerable margin of 20% in recognition of *sad* emotional status which was caused by the low number of samples for this emotion and high level of similarity shared with a neutral state in a human. In total, collected data of 10 rides from 8 different drivers was studied for evaluation of the system. Each ride was divided into sub-samples with the length of 2 min and gave us 79 samples for evaluation. Prediction of the driver emotion in one single ride is obtained from prediction results of its sub-samples. According to Table 3, final outcome of the tests on a single ride, represents that our proposed multi-modal emotion recognition system, achieved better results in comparison with each one of the modules alone. As it was initially considered, facial expression-based module plays the main role in final decision prediction, and SW along with VA are complementary modules.

Table 3. The results of each module in comparison with the fused one, in a single ride

Method	Accuracy	Precision	F1 score	Recall
Facial-based Module	54.54%	54.75%	50.45%	49.86%
SW-based Module	37.5%	10.3%	13.6%	25%
VA-based Module	68.18%	35.51%	37.76%	41.37%
Fusion of All Three Modules	**77.27%**	**73.39%**	**73.59%**	**75.89%**

The 77.27% of accuracy is obtained using multimodal emotion recognition system on data samples with 2 min of length. This condition is prone to errors and false predictions since in real-life situations the 2-min range could be easily falsified by situations like staying behind a red light. In order to cope with

such situations and increase the reliability of the results, we consider the decision taking step at the end of each ride by summarizing the emotion predictions performed for only sub-samples and choosing the most frequently felt emotion. In this way, our multimodal system achieves 94.4% of accuracy for classification into 4 emotional categories. Interestingly, the fusion of only behavioral-based modules (SW + VA) leads to 72.2% of accuracy in the same experiments. This demonstrates that in case of failures in camera(s) used for facial-based emotion recognition or highly illuminated situations, our proposed system is still capable of functioning in an acceptable level by relaying only on behavioral-related factors in order to predict the emotional status of the driver.

Table 4. Comparison of different unimodal and multimodal emotion recognition systems based on accuracy and different number of emotional classes

System	Type	Method	Classes	Accuracy
[40]	Unimodal	Electrodermal Activity (EDA)	3	70%
[51]	Unimodal	Facial Emotion Recognition	6	70.2%
[50]	Unimodal	Speech Emotion Recognition	3	88.1%
[26]	Unimodal	Speech Emotion Recognition	2	80%
[1]	Multimodal	EDA and Skin Temperature	4	92.42%
[12]	Multimodal	Speech & Facial Emotion Recognition	7	57%
[25]	Multimodal	Acoustic & Facial Emotion Recognition	3	90.7%
Our system	Multimodal	Facial and Vehicle Parameters	4	**94.4%**

A brief comparison of the state-of-the-art unimodal and multimodal works is presented at Table 4. Most of the multimodal approaches focus only on the fusion of speech and facial modules where the highest achieved accuracy is 90.7% by Hoch et al. [25]. However, they did consider only 3 classes as neutral, positive and negative emotion. Another notable method was proposed by Ali et al. [1] where they used the combination of EDA and skin temperature parameters of a driver as the input for a convolutional neural network and were able to get 92.4% of accuracy. Our proposed system by using the signals of vehicle controlling systems (steering wheel and acceleration/deceleration), along with the real-time facial expression-based approach achieved the highest accuracy rate of 94.4%.

5 Conclusion and Outlook

Many studies on emotion recognition are based on unimodal approaches where only visual or audio is examined and usually are aimed at classifying relatively basic emotions. There are multiple modalities that can be used to detect emotions. However, context in which the emotion is elicited and what modalities are most likely to be correlated should be taken into account. Various behavioral-based methods show promising results to such an approach. This can be more general in nature, for example head movement, gestures, and eye gaze, or can be specific to the environment of the vehicle. Some behaviors specific to the vehicle context can include acceleration, velocity, speeding, and steering wheel usage. Our main focus in this work was on developing a system with three different modalities of *steering wheel* and *vehicle acceleration* (as behavior modalities) beside the *facial expression* in order to study the correlation between emotional status and in-cabin behavior of the driver during a ride and more importantly, investigate the possibility of increasing the accuracy and robustness of emotion recognition systems by employing such related factors. We were able to represent the advantages of integration of behavior-based modalities into the emotion recognition systems and depict an structure for further developments on different other modalities. Our system was able to achieve relatively high accuracy rate after fusion of the modalities in comparison with each modality alone. Respectively we did demonstrate that our system stands on top of the similar multimodal works in this domain capable of classifying the emotions into 4 main categories with the accuracy of 94.4%. For the future, we plan to extend our system by integration of other different modalities and will try to study the shared models among all of the drivers according to their emotional states.

References

1. Ali, M., Machot, F.A., Mosa, A.H., Kyamakya, K.: CNN based subject-independent driver emotion recognition system involving physiological signals for ADAS. In: Schulze, T., Müller, B., Meyer, G. (eds.) Advanced Microsystems for Automotive Applications 2016. LNM, pp. 125–138. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44766-7_11
2. Alshamsi, H., Meng, H., Li, M.: Real time facial expression recognition app development on mobile phones. In: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 1750–1755. IEEE (2016)
3. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Incremental face alignment in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1859–1866 (2014)
4. Behoora, I., Tucker, C.S.: Machine learning classification of design team members' body language patterns for real time emotional state detection. Des. Stud. **39**, 100–127 (2015)
5. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B.: A database of German emotional speech. In: Ninth European Conference on Speech Communication and Technology (2005)
6. Cabrall, C., Janssen, N., Goncalves, J., Morando, A., Sassman, M., de Winter, J.: Eye-based driver state monitor of distraction, drowsiness, and cognitive load for transitions of control in automated driving. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 001981–001982. IEEE (2016)
7. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56

8. Chauhan, P.M., Desai, N.P.: Mel frequency cepstral coefficients (MFCC) based speaker identification in noisy environment using wiener filter. In: 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), pp. 1–5 (2014)
9. Chen, L.l., Zhao, Y., Ye, P.F., Zhang, J., Zou, J.Z.: Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst. Appl. **85**, 279–291 (2017)
10. Corneanu, C.A., Simón, M.O., Cohn, J.F., Guerrero, S.E.: Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans. Pattern Anal. Mach. Intell. **38**(8), 1548–1568 (2016)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
12. Datcu, D., Rothkrantz, L.: Multimodal recognition of emotions in car environments. DCI&I 2009 (2009)
13. Dhall, A., et al.: Collecting large, richly annotated facial-expression databases from movies. IEEE Multimedia **19**(3), 34–41 (2012)
14. D'mello, S.K., Kory, J.: A review and meta-analysis of multimodal affect detection systems. ACM Comput. Surv. (CSUR) **47**(3), 43 (2015)
15. Donia, M.M., Youssif, A.A., Hashad, A.: Spontaneous facial expression recognition based on histogram of oriented gradients descriptor. Comput. Inf. Sci. **7**(3), 31 (2014)
16. Fan, X.A., Bi, L.Z., Chen, Z.L.: Using EEG to detect drivers' emotion with Bayesian networks. In: 2010 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 3, pp. 1177–1181. IEEE (2010)
17. Friesen, E., Ekman, P.: Facial action coding system: a technique for the measurement of facial movement. Palo Alto (1978)
18. Friesen, W.V., Ekman, P., et al.: EMFACS-7: emotional facial action coding system. Univ. Calif. San Francisco **2**(36), 1 (1983)
19. Fung, N.C., et al.: Driver identification using vehicle acceleration and deceleration events from naturalistic driving of older drivers. In: 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 33–38. IEEE (2017)
20. Ganchev, T., Fakotakis, N., Kokkinakis, G.: Comparative evaluation of various MFCC implementations on the speaker verification task. In: Proceedings of the SPECOM, vol. 1, pp. 191–194 (2005)
21. Govindarajan, V., Bajcsy, R.: Human modeling for autonomous vehicles: Reachability analysis, online learning, and driver monitoring for behavior prediction (2017)
22. Guerrero Rázuri, J.F., Larsson, A., Sundgren, D., Bonet, I., Moran, A.: Recognition of emotions by the emotional feedback through behavioral human poses. Int. J. Comput. Sci. Issues **12**(1), 7–17 (2015)
23. Gutmann, M., Grausberg, P., Kyamakya, K.: Detecting human driver's physiological stress and emotions using sophisticated one-person cockpit vehicle simulator. In: Information Technologies in Innovation Business Conference (ITIB) 2015, pp. 15–18. IEEE (2015)
24. Hammal, Z., Cohn, J.F., Heike, C., Speltz, M.L.: What can head and facial movements convey about positive and negative affect? In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 281–287. IEEE (2015)

25. Hoch, S., Althoff, F., McGlaun, G., Rigoll, G.: Bimodal fusion of emotional data in an automotive environment. In: 2005 Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005), vol. 2, p. ii–1085. IEEE (2005)
26. Jones, C.M., Jonsson, I.M.: Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses. In: Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future, pp. 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia (2005)
27. Kamaruddin, N., Wahab, A.: Features extraction for speech emotion. J. Comput. Methods Sci. Eng. **9**(1, 2S1), 1–12 (2009)
28. Kamaruddin, N., Wahab, A.: Driver behavior analysis through speech emotion understanding. In: 2010 IEEE Intelligent Vehicles Symposium (IV), pp. 238–243. IEEE (2010)
29. Katsis, C.D., Rigas, G., Goletsis, Y., Fotiadis, D.I.: Emotion recognition in car industry. In: Emotion Recognition: A Pattern Analysis Approach, pp. 515–544 (2015)
30. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
31. Kensinger, E.A.: Remembering emotional experiences: the contribution of valence and arousal. Rev. Neurosci. **15**(4), 241–252 (2004)
32. Khalid, M., Wahab, A., Kamaruddin, N.: Real time driving data collection and driver verification using CMAC-MFCC. In: Proceeding of the 2008 International Conference on Artificial Intelligence (ICAI 2008), pp. 219–224 (2008)
33. Khan, R.A., Meyer, A., Konik, H., Bouakaz, S.: Human vision inspired framework for facial expressions recognition. In: 2012 19th IEEE International Conference on Image Processing (ICIP), pp. 2593–2596. IEEE (2012)
34. Krotak, T., Simlova, M.: The analysis of the acceleration of the vehicle for assessing the condition of the driver. In: 2012 IEEE Intelligent Vehicles Symposium (IV), pp. 571–576. IEEE (2012)
35. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94–101. IEEE (2010)
36. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. IEEE (1998)
37. Mcmanus, A.: Driver Emotion Recognition and Real Time Facial Analysis for the Automotive Industry (2017). http://blog.affectiva.com/driver-emotion-recognition-and-real-time-facial-analysis-for-the-automotive-industry
38. Mehrabian, A., et al.: Silent Messages, vol. 8. Wadsworth, Belmont (1971)
39. Mishra, B., et al.: Facial expression recognition using feature based techniques and model based techniques: a survey. In: 2015 2nd International Conference on Electronics and Communication Systems (ICECS), pp. 589–594. IEEE (2015)
40. Ooi, J.S.K., Ahmad, S.A., Chong, Y.Z., Ali, S.H.M., Ai, G., Wagatsuma, H.: Driver emotion recognition framework based on electrodermal activity measurements during simulated driving conditions. In: 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 365–369. IEEE (2016)

41. Ooi, J.S.K., Ahmad, S.A., Harun, H.R., Chong, Y.Z., Ali, S.H.M.: A conceptual emotion recognition framework: stress and anger analysis for car accidents. Int. J. Veh. Saf. **9**(3), 181–195 (2017)
42. Ouellet, S.: Real-time emotion recognition for gaming using deep convolutional network features. arXiv preprint arXiv:1408.3750 (2014)
43. Poria, S., Cambria, E., Bajpai, R., Hussain, A.: A review of affective computing: from unimodal analysis to multimodal fusion. Inf. Fusion **37**, 98–125 (2017)
44. Poria, S., Cambria, E., Hussain, A., Huang, G.B.: Towards an intelligent framework for multimodal affective data analysis. Neural Netw. **63**, 104–116 (2015)
45. Salih, H., Kulkarni, L.: Study of video based facial expression and emotions recognition methods. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 692–696. IEEE (2017)
46. Samanta, A., Guha, T.: On the role of head motion in affective expression. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2886–2890. IEEE (2017)
47. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(6), 1113–1133 (2015)
48. Slaney, M.: Auditory toolbox version 2. interval research corporation. Indiana: Purdue University 2010, 1998-010 (1998)
49. Swinkels, W., Claesen, L., Xiao, F., Shen, H.: SVM point-based real-time emotion detection. In: 2017 IEEE Conference on Dependable and Secure Computing, pp. 86–92. IEEE (2017)
50. Tawari, A., Trivedi, M.: Speech based emotion classification framework for driver assistance system. In: 2010 IEEE Intelligent Vehicles Symposium (IV), pp. 174–178. IEEE (2010)
51. Theagarajan, R., Bhanu, B., Cruz, A., Le, B., Tambo, A.: Novel representation for driver emotion recognition in motor vehicle videos. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 810–814. IEEE (2017)
52. De la Torre, F., Campoy, J., Ambadar, Z., Cohn, J.F.: Temporal segmentation of facial behavior. In: IEEE 11th International Conference on Computer Vision ICCV 2007, pp. 1–8. IEEE (2007)
53. Zadeh, A., Zellers, R., Pincus, E., Morency, L.P.: Multimodal sentiment intensity analysis in videos: facial gestures and verbal messages. IEEE Intell. Syst. **31**(6), 82–88 (2016)
54. Zhang, Y., Chen, M., Guizani, N., Wu, D., Leung, V.C.: SOVCAN: safety-oriented vehicular controller area network. IEEE Commun. Mag. **55**(8), 94–99 (2017)
55. Zhenhai, G., DinhDat, L., Hongyu, H., Ziwen, Y., Xinyu, W.: Driver drowsiness detection based on time series analysis of steering wheel angular velocity. In: 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), pp. 99–101. IEEE (2017)