# Automatic Inference of Driving Task Demand from Visual Cues of Emotion and Attention

*Alessandro Soro and Andry Rakotonirainy*

*Queensland University of Technology*
*Brisbane - Australia*

## ABSTRACT

Sensing the mental, physical and emotional demand of a driving task is of primary importance in road safety research and for effectively designing in-vehicle information systems (IVIS). Particularly, the need of cars capable of sensing and reacting to the emotional state of the driver has been repeatedly advocated in the literature. Algorithms and sensors to identify patterns of human behavior, such as gestures, speech, eye gaze and facial expression, are becoming available by using low cost hardware: This paper presents a new system which uses surrogate measures such as facial expression (emotion) and head pose and movements (intention) to infer task difficulty in a driving situation. 11 drivers were recruited and observed in a simulated driving task that involved several pre-programmed events aimed at eliciting emotive reactions, such as being stuck behind slower vehicles, intersections and roundabouts, and potentially dangerous situations. The resulting system, combining face expressions and head pose classification, is capable of recognizing dangerous events (such as crashes and near misses) and stressful situations (e.g. intersections and way giving) that occur during the simulated drive.

**Keywords**: Affective, Social, Computing, Automotive, in-Vehicle, Emotion, Facial Expression, Head Pose

## INTRODUCTION

Measuring or estimating the mental and physical demand of a driving task has a central role both in road safety research and for vehicle and in-vehicle information systems (IVIS) design. Previous research has shown exhaustively that the implementation of on-board electronics, often originally conceived as a safety aid, can actually turn into a safety threats, since it competes for the driver’s attention when providing safety related information.

Advanced driver assistance systems (ADAS) manage to reduce the driver’s cognitive workload by partially automating some driving tasks, as controlling the speed during a long drive, possibly adapting to small changes in the traffic flow. While such system has a clear advantage in terms of comfort and safety, as they relieve the driver from a series of barely operational duties, at the same time they can induce an unsafe adaptation on the drivers’ part, increasing the response delay in hazard detection tasks(Rudin-Brown & Parker, 2004).

From a different perspective, road congestion and time pressure, together with personal and situational factors, such as age and sex, previous anger or stress, competitiveness, sensation seeking, anonymity, are known to be related to driver aggression and anger (Soole, Lennon, Watson, & Bingham, 2011). Although intuitively emotion at the steering wheel can be regarded as a matter of pleasure and comfort, it must be remarked here that when driving, emotions such as anger and aggressive behavior are regarded as a major contributing factor to car crashes (AAAFoundation.org, 2009), comparable to alcohol impairment (Cook, Knight, & Olson, 2005)



https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2099-2

Algorithms and sensors to recognize natural human behavior, such as gestures, speech, eye gaze and facial expression, are becoming available, even on low cost hardware. Social cues, such eye gaze, body posture, facial expression and actions as gestures and touch, possibly combined with haptic feedback and powerful visualizations like augmented reality have proven effective in supporting communication and facilitating complex tasks, lightening the cognitive demand of computer applications and are now under the lens of road safety and car industry designers, with the aim of exploiting natural paradigms of interaction in the car, without competing for the driver's attention.

The present work focuses on the problem of sensing user behavior, specifically facial expression and head pose, with the aim of gathering emotional and attentional cues from which to infer the level of task difficulty that the driver is currently experiencing. After introducing the research problem both from the point of view of road safety, and from that of human computer interaction and machine learning we describe an original technology capable of recognizing several important phases of a drive: normal drive versus intersection negotiation, and the occurrence of dangerous events.

We discuss the strengths and limitations of our approach with the aid of a field exploratory test and a simulator experiment, respectively organized for making sense of the real world, complex problem, for collecting the necessary training data, and for testing the performances of the system.

## BACKGROUND AND RELATED WORKS

Emotion and Attention while driving have been long under the lens of road safety scientists. As said above, driving anger and aggressive behavior are a growing cause of concern: emotional arousal (as opposed to a neutral mood) when driving is known to decrease the drivers' performances in terms of lane adherence, steering wheel angle and sharpness of lane crossings (Cai, Lin, & Mourant, 2007). Within the car, several scenarios have been envisioned that rely on emotion recognition in order to, for example, improve drivers' productivity, wellbeing or pleasure while keeping in mind drivers' safety (Eyben et al., 2010). However, the automatic recognition of the emotional state of the driver, or more precisely, the classification of a suitable proxy for such emotional state, e.g., facial expressions, stress in the voice or a body postures, is still an open topic for research.

Hoch et al. (Hoch, Althoff, McGlaun, & Rigoll, 2005) exploit a fusion of audio and video modalities to classify the drivers' emotions according to three possible classes: neutral, positive, negative. They took audio and video recordings within a real car, but without motion (hence with little or no noise and relatively stable light). By doing so, they manage to achieve an average 90% recognition rate. The application domain they sketched is the improvement of human computer interaction in the car, for example, adapting the dialog strategies of the assistance and information systems, reducing mental workload and distraction.

MIT's SmartCar project (J. a. Healey & Picard, 2005; J. Healey & Picard, 2000) explored how an appropriate combination of sensors, capable of providing physiological data such as electromyogram, electrocardiogram, galvanic skin response and respiration can be fed to appropriate pattern recognition algorithm in order to predict driver's stress. In a naturalistic study, they compared the data gathered from sensors worn by the drivers to the level of stress self-reported by the participants, showing a fairly accurate (88.4%) rate of prediction for the measures above. The application domain traced by J. Healey & Picard (J. Healey & Picard, 2000) is the recognition of driver's stress, and hence comfort and wellbeing. The authors focus on the technology rather than depicting an application scenario, although in the concluding remarks they sketch the possible outcome of 'giving a quantified feedback to the individual' thus framing the possible interventions in the self-awareness domain and behavior change.

The 'Emotionally Responsive Car' (C. Jones & Jonsson, 2008; C. M. Jones & Jonsson, 2005; Jonsson, Nass, Harris, & Takayama, 2005; Nass et al., 2005) is envisioned to react to the driver's sensed emotion by changing the driver-vehicle interface. Example reactions include becoming 'less or more talkative depending on the mood of the driver' or changing 'the telematics, climate, music in the car in response the mood of the driver'. In a driving simulator, drivers are presented to challenging driving conditions to elicit a range of emotions, such as boredom, sadness/grief, frustration/anger, happiness or surprise. Emotional arousal is inferred from paralinguistic cues in speech recordings taken during the simulated driving task by means of an automated system and compare the results to evaluations performed by human experts.

Schroeter and colleagues (Schroeter, Soro, & Rakotonirainy, 2013) discuss a case study of emotion recognition from facial expressions in the car aimed at supporting the development of applications for enhancing self-awareness (thus influencing the driver behavior, both in real-time and over time), supporting social awareness while driving, e.g. to change drivers' attitude towards others, and finally improving urban awareness in and outside the car enhancing the understanding of the road infrastructure as a whole. In their case study Schroether and co-workers show qualitatively that facial expressions of the driver can be put in relation with significant events occurring during the drive, such as lane merging and intersection negotiation.

Discovering the focus of driver attention is not an easy task. Laboratory studies can exploit eye tracking to match eye-gaze and object in a controlled environment (Fletcher, Loy, Barnes, & Zelinsky, 2005; Horrey, Wickens, & Consalus, 2006; Pradhan, Pollatsek, Knodler, & Fisher, 2009). However, such solution have mainly been adopted for addressing the problem of driver distraction, e.g. bringing into focus the risks of in-vehicle information systems (IVIS) competing for driver's attention (Donmez, Boyle, & Lee, 2007, 2008; Jahn, Oehme, Krems, & Gelau, 2005; Roberts, Ghazizadeh, & Lee, 2012), rather than, as is our aim, for understanding the experienced task difficulty.

Previous research has often relied on self-reporting to assess the level of task difficulty and understand how drivers respond to on-road events.

Electrodermal activation, typically galvanic skin response, has been adopted to understand how drivers manage to respond to traffic and road demands. Seminal work on this subject evidenced that drivers manage to maintain a desired level of anxiety (Taylor, 1964). In more recent research, Wilde attributes such variations to the driver's subjective interpretations of the risk or probability of crash (Wilde, 1982), and Fuller to an attempt on the driver's part of maintaining a desired level of task difficulty (Fuller, 2005).

In both cases, drivers seem to aim at 'optimizing' the level of arousal, by adjusting their driving style (mainly the speed) in order to keep a desired, not too low, not too high, level of risk or difficulty, respectively.

However, in the view of a real world deployment, neither self-reporting, nor invasive measures such as electrodermal activation are suitable for measuring the task difficulty. Instead, we will describe how the combined adoption of head pose detection and facial expression recognition can provide an unobtrusive estimate of the task demand, support the implementation of the proposed scenario, and could potentially be implemented for personal mobile devices, and hence easily adopted on a large scale.

Algorithms and techniques for facial expression recognition from images or video involve the isolation and subsequent processing of a variable number of regions (features) of the face, and a comparison with corresponding regions from other expressions (typically the neutral expression) to determine changes in appearance or position of critical areas (see (Fasel & Luettin, 2003) for a comprehensive survey on this subject).

Changes in appearance, such as the onset of wrinkles on one's forehead when expressing surprise can be evidenced, for example using (combinations of) Gabor filters (Lyons, Akamatsu, Kamachi, & Gyoba, 1998) and feeding the resulting representation into an appropriate classifier, typically a multilayer perceptron (such as in (Zhang, Lyons, Schuster, & Akamatsu, 1998)) or a support vector machine (e.g in (Bartlett, Littlewort, Fasel, & Movellan, 2003)).

Starting from such background work we have developed a novel approach to driving task difficulty estimation, that combines emotional cues from facial expression to attentional indicators, such as head pose. In the following sections we describe in detail the collection of the training sets and hoe the raw classification can be processed to infer the occurrence of certain stressful events.

# DATA COLLECTION

Two sessions of data collection were performed to inform the present research. In Study 1 the researchers collected video footage in a naturalistic driving setting. This was aimed at making sense of the feasibility of applying traditional facial expression recognition techniques to the in-vehicle setting.

In Study 2, 11 participants were recruited and observed in a simulated driving task that involved several pre-programmed events aimed at eliciting emotive reactions, such as being stuck behind slower vehicles, intersections and roundabouts, and potentially dangerous situations, including a vehicle drifting against the traffic and a car that pulls out of a blind intersection failing to give way.

## Study 1

Four test drives were carried on at different times of day, capturing images of the driver at about 1 FPS from a camera installed on the dashboard of the vehicle. Sample captures are presented in figure 1. As expected, not all frames could be positively processed for face detection and emotion recognition; it is worth remarking here that this experiment is meant to be exploratory and to guide the design of the simulator study described further on, rather than to provide figures and statistics.

**Table 1: Frames captured during a real drive**

| | Face detected | No Face Detected |
|---|---|---|
| drive #1 | 624 (65.7%) | 326 (34.3%) |
| drive #2 | 2006 (68.7%) | 914 (31.3%) |
| drive #3 | 658 (54.2%) | 556 (45.8%) |
| drive #4 | 2933 (59.2%) | 2020 (40.8%) |

Table 1 presents the results of face detection for the frames captured during the test drives. Since the device and its pedestal were removed from the vehicle after each drive, the 4 series of captures were taken from slightly different perspectives. Also, the time of day varied between mid morning and late afternoon. Face detection failed for 35% to 45% of the frames, due in part to bad lighting, in part to wrong pose or more rarely to occlusions.

While this study does not provide useful insights into what facial expressions could be accurately recognized using state of the art algorithms in a naturalistic test drive, yet it shows that the real deployment scenario presents challenges than current research dataset don't capture. At the moment of writing, to the best of our knowledge, no dataset exists that provides annotated video footage from real drives suitable for automatic emotion recognition studies. The LISA-P Head Pose Database (Martin, Tawari, Murphy-Chutorian, Cheng, & Trivedi, 2012) has not been released for public use yet and will only provide facial landmarks for eye corners, nose tip and nose corners, i.e., prominent features specifically chosen for being the least sensitive to facial expressions.

To overcome this limitation, the study 2, described below, was designed to gather an initial dataset of facial expressions of drivers and explore the use of the driver's face/head to infer useful parameters about the drive.

(a) (b) (c) (d)

Figure 1: Challenges of in-vehicle facial expression recognition: (a) ideal case; (b) bad lighting condition; (c) bad pose; (d) occasional occlusions

## Study 2

Eleven participants (9 male, 2 female) were invited to take part in this study, aimed at recording realistic (if not real) facial expressions of drivers during a simulated drive from the city center of Brisbane to the BNE international airport.

Along the drive, the participants experience a number a pre-programmed events aimed at eliciting emotive reactions (see Table 4). Often the participant finds her/him-self stuck into traffic or obstructed by slow vehicles; in one case the vehicle preceding the participant is specifically programmed to proceed slowly and obstruct the road, accelerating if the participant attempts to overtake. Several intersections or roundabouts force the driver to stop and get into another traffic flow; the spacing between cars is initially small and increases gradually. Finally, a number of potentially dangerous situations are programmed, including a vehicle drifting against the traffic as well as a car that pulls out of a blind intersection failing to give way.

From inside the control room, the research team monitored the reactions of the driver, taking note of the facial expressions, especially in coincidence with these key events. Additionally, immediately after the driving session, the subjects were invited to reflect upon the experience, with emphasis on their reactions to the key events. The video recordings of the facial expressions of the drivers were later used for training the automatic classifiers described below.

The Advanced Driving Simulator used in the presented research consists of a complete and fully functional vehicle body installed on top of computer controlled mobile platform. All controls in the cabin, such as steering wheel, dashboard, pedals, electric windows, etc. are fully functional. All 5 seats are available for studies involving multiple occupants. A panoramic screen -composed from three 4x3m projected screens - provides 180 degrees of forward vision, while rear vision is simulated by means of small LCD screen that replace the 3 mirrors.

The cabin is mounted on top of a 6 Degrees of Freedom motion system that provides up to 700mm of motion in each direction, and up to 39 degrees of rotation in each direction. The motion system adds to the realism of the simulation providing shakes and a sense of acceleration consistent with the simulated drive. Such complex setup can provide an immersive and fully interactive environment, including traffic and roadway environmental characteristics, and provide the driver with high-fidelity motion, visual, auditory, and force feedback cues. A video camera mounted on the dashboard recorded a video of the face of the driver; in 5 of the 11 drives a second camera installed on the passenger seat recorded a video of the simulated drive and the hands of the driver.

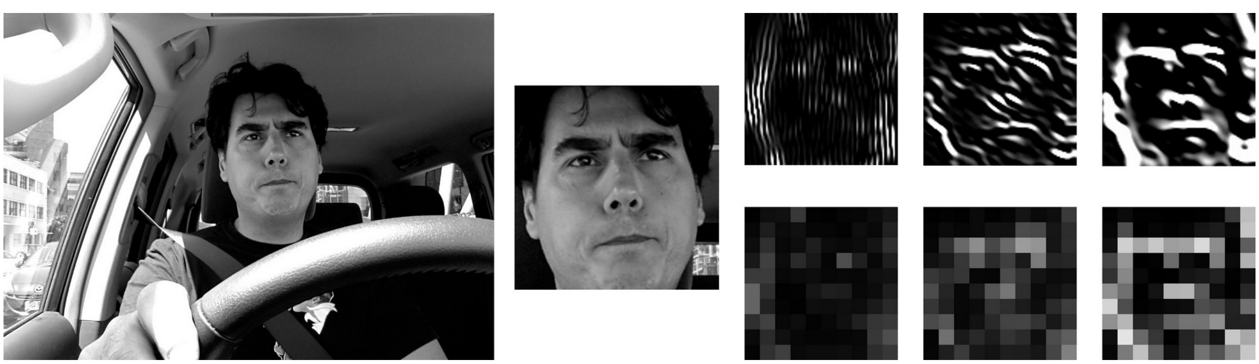

Figure 2: images captured are processed through a series of differently shaped/oriented Gabor filters reduced to a series of matrices of luminosity values, and finally linearized in a feature vector

## ANALISYS OF THE VIDEO COLLECTED

The video recordings of the driving sessions resulted in about 130 minutes of video footage, or 250.000 frames of frontal images of facial expressions. Of these, a smaller set was used to train an automatic facial expression classifier, namely those frames in which the face of the driver appeared clearly lighted, not covered by any occlusion (e.g., one hand kept in the line of sight of the camera) and from a frontal view, thus excluding those sequences in which the driver glances at the mirrors.

## Implementation

Our approach follows the method proposed in (Lyons et al., 1998), the implementation is based on open libraries: frames are first processed for face detection by means of an cascade classifier of Haar-like features (Viola & Jones, 2011); the resulting regions are normalized to position eyes/mouth at specific position, determined by means an active shape model (Cootes, Taylor, Cooper, & Graham, 1995) (ASM[1]).

Table 2: Confusion matrix of 7 class facial expression recognition: **N**eutral, **A**nger, **D**isgust, **F**ear, **H**appiness, **Sa**dness, **Su**rprise

| (%) | N | A | D | F | H | Sa | Su |
|---|---|---|---|---|---|---|---|
| N | **99.1** | 0.1 | 0.4 | 0.0 | 0.1 | 0.3 | 0.1 |
| A | 12.2 | **84.6** | 0.0 | 0.0 | 0.0 | 0.8 | 2.3 |
| D | 2.6 | 1.0 | **96.3** | 0.0 | 0.0 | 0.0 | 0.0 |
| F | 4.8 | 0.0 | 0.0 | **92.3** | 0.0 | 0.0 | 2.9 |
| H | 1.2 | 0.0 | 0.0 | 0.0 | **98.3** | 0.0 | 0.6 |
| Sa | 23.2 | 3.9 | 0.0 | 0.7 | 0.0 | **71.2** | 1.0 |
| Su | 2.9 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | **96.7** |

The image is then processed through a series of differently shaped/oriented Gabor filters (see Fig. 2) reduced to a series of matrices of luminosity values, and finally linearized in a feature vector. The training and evaluation of our base algorithm for facial expression classification was performed on the CK+ dataset. The CK+ database (Kanade, Cohn, & Tian, 2000; Lucey et al., 2010) includes 593 sequences from 123 subjects, each of which showing one of several possible emotions/expressions, and is constantly used as benchmark in related literature for the evaluation of facial expression recognition algorithms. Of the 593 sequences, 327 are provided with a (consistent and validated) emotional label based on the presence/absence of specific action units (e.g. dilated nostrils, inner/outer brow raise, etc.) according to the *Facial Action Coding System* (Ekman & Friesen, 1977).

Table 2 shows the accuracy of our classification algorithm, while Table 3 shows a comparison of our implementation to other results found in the literature that have been trained and evaluated on the CK+ dataset. It can be seen that the accuracy of our implementation is in line with state of the art approaches.

Table 3: Comparison of different approaches for facial expression classification

| Author(s) | Average accuracy |
|---|---|
| Our implementation | 91.2% |
| Shojaeilangari, et al. (Shojaeilangari, 2011) | 92.97% |

[1] Roughly, an ASM learns a statistical model of the shape of a face, starting from a training set of landmarked images, and constructing a map of acceptable deformations and of the expected appearance of the image close to each landmark. Then, the fitting algorithm iteratively tries to optimize the match between the texture at each landmark's current position and the expected texture for that landmark by moving each landmark within a certain range

| | |
|---|---|
| Naika, et al. (Naika C.L., Jha, Das, & Nair, 2012) | 83.09% |
| Mariappan, et al. (Mariappan, Suk, & Prabhakaran, 2012) | 87.4% |
| Jain, et al. (Jain & Aggarwal, 2011) | 85.8% |
| Chew, et al. (Chew et al., 2011) | 74.4% |
| Lucey, et al. (Lucey et al., 2010) | 83.3% |

However, when applied to images that come from a real or simulated drive, as in the case of the two studies detailed below, the algorithm hardly provides any meaningful result. We show below that the changes in head pose and lighting conditions, together with the peculiarities of the driving activity require further processing for selecting those frames that are more likely to result in a successful classification; this was done during the training phase by manually pruning the dataset, and in our automatic system by filtering the video sequence through a head pose classifier.

Hence, a second classifier was trained for automatically selecting the frontal frames from those frames in which the driver is looking left/right. Such head pose classifier was trained on the Color-FERET dataset (Phillips, Rizvi, & Rauss, 2000; Phillips, Wechsler, Huang, & Rauss, 1998). Additionally, the facial expressions recognized by the emotion classifier were grouped in *neutral*, *positive (*happiness, surprise), and *negative* (anger, disgust, fear, sadness).

Running the two classifiers on the collected images results the in a series of *tuples* $(h, neu, \neg, pos)_t$ where $h \in [-1.1]$ represents the head pose as it was recognized automatically by the system, with $h \approx 0$ if the driver is looking straight ahead, and $neu, neg, pos \in [0,1]$ with $neu \approx 1$ meaning that the system has recognized a neutral expression, $neg \approx 1$ meaning the system has recognized a negative (e.g. sad, angry frowned) expression and $pos \approx 1$ meaning the system has recognized a positive (e.g. smile, surprise) expression. Each tuple is representative of a single frame.

Table 4: The events programmed in the simulated drive in order to elicit emotional reactions on the drivers' part

| # | Description |
|---|---|
| 1 | Give way with gap acceptance test |
| 2 | Turn left with cyclist passing |
| 3 | Traffic light changes to amber |
| 4 | Right turn with gap acceptance test |
| 5 | Traffic light changes to amber |
| 6 | Two slow leading cars, blocking the way |
| 7 | Right turn with pedestrian crossing |
| 8 | Pedestrian with child crossing |
| 9 | Leading car going slow, accelerates when trying to overtake |
| 10 | Same car, breaks unexpectedly and without visible reason |
| 11 | Left turn at roundabout with gap acceptance test |
| 12 | Oncoming car going against the traffic |

| | |
|---|---|
| 13 | Slow moving car, partially blocks the road |
| 14 | Parked car pulling out, failing to signal or give way |
| 15 | Car from far right not giving way, almost unavoidable crash |

The goals of this study is then to identify possible algorithms $\varphi(h, neu, \neg, pos)_{[t=0 \ldots t=n]}$ for inferring the difficulty of the drive based on a (possibly short) observation of the drivers head and face appearance. 5 of the 11 test drives were then again annotated as to identify different phases and special events such as negotiating intersections (events 1, 4, 7 and 11) and shocking or dangerous events (events 10, 12 and 15).

## Dangerous events

We show empirically that dangerous or shocking events can be visually recognized from video footage of the head/face of the driver by showing the performances of a classifier, trained on the data collected in the simulated drive. 15 sequences were chosen, 10 taken during normal drive and 5 taken in correspondence or immediately following a collision or near miss. The 15 sequences resulted in 4709 frames; hence, in average, each sequence spans over 314 frames or 10.46 seconds.

A neural network was trained to classify each frame as belonging to a sequence of normal drive or dangerous event. The feature vector consists, for each frame, of the values of mean and variance of over a time span of 8 seconds preceding the frame of head pose and facial expression, resulting from the two classified described above, that is:

$$\varphi(h, neu, \neg, pos) = d \text{¿}).$$

A leave-one-out strategy was used for training: for all 15 sequences in turn, 1 was reserved as test set and the remaining 14 sequences were included in the training set. In this way we can positively argue that the network has learned to recognize the proposed behavior, either from different sequences of the same driver or from a corresponding sequence of another driver. The average results are summarized in table 5: a sequence is considered correctly classified if more than 50% of its frames (cumulatively reported in parentheses) have been correctly classified; an average accuracy of 80% was achieved using the above technique.

Table 5: Confusion matrix of 2 class normal/danger driving condition

| **(%)** | **normal drive** | **dangerous event** |
|---|---|---|
| Normal | **8** (2352 frames) | 2 (844 frames) |
| Danger | 1 (291 frames) | **4** (1096 frames) |

## Intersections

Giving way and merging into a flow of traffic has been identified as being among the most stressful tasks when driving, (see e.g. (J. Healey & Picard, 2000)), and hence it is useful to be able to distinguish such phases from normal drive, for example in order to adapt the behavior of mobile phones and other devices or suspend less important notifications. Sections of normal drive can be distinguished from intersections on the basis of head movements.

Again, a classifier was trained to classify each frame as belonging to a sequence of normal drive or to an intersection. The feature vector consists, for each frame, of the values of mean and variance of over a time span of 8 seconds preceding the frame of head pose alone, however in this case the neural network was trained for time series prediction using the technique of the sliding window: the feature vector consists of a series of 50 measures of mean and variance over the last 3 seconds of the head pose value. Single measures are taken at 10$Hz$, and hence span over a window of 5 seconds, facial expression is not exploited in this classifier; hence:

$$\varphi(h)=X\left(\left(\sigma^2,\mu\right)[h]_{t=-5sec}\ldots\left(\sigma^2,\mu\right)[h]_{t=0}\right)$$

As before, a leave-one-out strategy was used for training: from a total of 37 sequences (19 normal drive and 18 intersections) in turn, 1 was reserved as test set and the remaining 36 sequences were included in the training set. The average results are summarized in table 6: 95% was achieved using the above technique.

Table 6: Confusion matrix of 2 class normal/intersection driving condition

| **(%)** | **normal drive** | **Intersection** |
|---|---|---|
| Normal | 18 | 1 |
| Intersection | 1 | 17 |

# DISCUSSION AND CONCLUSION

We have presented a novel technique for in-vehicle vision based stress detection aimed at assessing the driving task demand on an *Intelligent Car*. Emotional arousal and fatigue are recognized threats, and considered to be involved in a large number of car crashes. The recognition and interpretation of the driving task demand is then the key to open the door to a vast panorama of intelligent applications aimed at improving the safety and comfort of the future cars.

However, existing techniques for emotion recognition draw from a context-neutral background, and have not been adequately assessed in a real automotive scenario. As a result, datasets for training and research comparison, as well as a thorough understanding of the issues, constraints and opportunities that arise during a real drive, are still missing. This work is meant to move a step in such direction.

We have provided a detailed description of the principles and algorithms required for training a vision based emotion recognition system. Our results show the feasibility of the approach proposed here, but more important, point out several constraints and challenges that any real-world implementation will have to face. To summarize, the contribution of our work is:

We propose a novel technique for automatic inference of driving task demand from visual cues of emotion and attention. We build on existing techniques adapting to a novel and challenging scenario; our approach combines facial expression recognition and head pose detection to overcome the limitations of the driving scenario: variable lighting conditions, occlusions, elusiveness of face expression.

We report on a test drive experience aimed both at assessing the validity of our approach in the real world and (more important) at uncovering issues and opportunities that are only faced out of the lab; we show that the real deployment scenario presents challenges than current research dataset don't capture; yet we show that our prototype implementation of the proposed approach is capable of achieving remarkable recognition rates, even in such challenging condition;

A number of issues remain open and need to be further explored. Further improvements in the classification accuracy can be reasonably achieved by gathering more and more samples from real test drives. A general trend in emotion and behavior detection is to take advantage of alternative, redundant channels and modalities. The in-car setting provides several such alternative modalities, many of which have been barely explored for emotion detection: acceleration and braking behavior, steering sharpness, gesture recognition, tailgating detection, are promising examples that allow non-invasive implementations even on inexpensive hardware.

Furthermore, while a driving simulator study as the one described here allows to safely and programmatically collect data on the emotional reactions of drivers, it is clear that the overall goal would be to achieve the necessary accuracy for a real world deployment. Hence, the natural next step will be to replicate the study described here in a naturalistic setting, so as to improve the proposed approach with *real* data.

## ACKNOLEDGEMENTS

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips et al., 2000, 1998).

Portions of the research in this paper use the CK+ Cohn-Kanade Facial Expression Database (Kanade et al., 2000; Lucey et al., 2010).

## REFERENCES

AAAFoundation.org. (2009). *Aggressive Driving: Research Update*. AAAFoundation.org.

Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on* (Vol. 5, p. 53). doi:10.1109/CVPRW.2003.10057

Cai, H., Lin, Y., & Mourant, R. R. (2007). Study on driver emotion in driver-vehicle-environment systems using multiple networked driving simulators. In *Proceedings of the Driving Simulation Conference North America (DSC'07)*.

Cook, L. J., Knight, S., & Olson, L. M. (2005). A comparison of aggressive and DUI crashes. *Journal of Safety Research*, *36*(5), 491.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, *61*(1), 38–59. doi:10.1006/cviu.1995.1004

Donmez, B., Boyle, L. N., & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident; Analysis and Prevention*, *39*(3), 581–90. doi:10.1016/j.aap.2006.10.003

Donmez, B., Boyle, L. N., & Lee, J. D. (2008). Mitigating driver distraction with retrospective and concurrent feedback. *Accident; Analysis and Prevention*, *40*(2), 776–86. doi:10.1016/j.aap.2007.09.023

Ekman, P., & Friesen, W. V. (1977). *Facial action coding system*. Consulting Psychologists Press, Stanford University, Palo Alto.

Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., & Nguyen-Thien, N. (2010). Emotion on the road - necessity, acceptance, and feasibility of affective computing in the car. *Advances in Human-Computer Interaction*, *2010*.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, *36*(1), 259–275. doi:10.1016/S0031-3203(02)00052-3

Fletcher, L., Loy, G., Barnes, N., & Zelinsky, A. (2005). Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems*, *52*(1), 71–84. doi:10.1016/j.robot.2005.03.010

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident; Analysis and Prevention*, *37*(3), 461–472. doi:10.1016/j.aap.2004.11.003

Healey, J. a., & Picard, R. W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 156–166. doi:10.1109/TITS.2005.848368

Healey, J., & Picard, R. (2000). SmartCar: detecting driver stress. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (Vol. 4, pp. 218 –221 vol.4). doi:10.1109/ICPR.2000.902898

Hoch, S., Althoff, F., McGlaun, G., & Rigoll, G. (2005). Bimodal fusion of emotional data in an automotive environment. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on* (Vol. 2, p. ii/1085 – ii/1088 Vol. 2). doi:10.1109/ICASSP.2005.1415597

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, *12*(2), 67–78.

Jahn, G., Oehme, A., Krems, J. F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(3), 255–275. doi:10.1016/j.trf.2005.04.009

Jones, C., & Jonsson, I.-M. (2008). Using Paralinguistic Cues in Speech to Recognise Emotions in Older Car Drivers. In C. Peter & R. Beale (Eds.), *Affect and Emotion in Human-Computer Interaction* (Vol. 4868, pp. 229–240). Springer Berlin / Heidelberg.

Jones, C. M., & Jonsson, I.-M. (2005). 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). Narrabundah, Australia, Australia: Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

Jonsson, I. M., Nass, C., Harris, H., & Takayama, L. (2005). Matching In-Car Voice with Driver State : Impact on Attitude and Driving Performance. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 173–181).

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 46–53). doi:10.1109/AFGR.2000.840611

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* (pp. 94–101). doi:10.1109/CVPRW.2010.5543262

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* (pp. 200–205). doi:10.1109/AFGR.1998.670949

Martin, S., Tawari, A., Murphy-Chutorian, E., Cheng, S. Y., & Trivedi, M. (2012). On the design and evaluation of robust head pose for visual user interfaces. In *4th ACM SIGCHI International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AUTO-UI)* (pp. 149–154). New York, New York, USA: ACM Press. doi:10.1145/2390256.2390281

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI '05 extended abstracts on Human factors in computing systems* (pp. 1973–1976). New York, NY, USA: ACM. doi:10.1145/1056808.1057070

Phillips, P. J., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(10), 1090–1104. doi:10.1109/34.879790

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, *16*(5), 295–306. doi:10.1016/S0262-8856(97)00070-X

Pradhan, a K., Pollatsek, a, Knodler, M., & Fisher, D. L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics*, *52*(6), 657–73. doi:10.1080/00140130802550232

Roberts, S. C., Ghazizadeh, M., & Lee, J. D. (2012). Warn me now or inform me later: Drivers' acceptance of real-time and post-drive distraction mitigation systems. *International Journal of Human-Computer Studies*, *70*(12), 967–979. doi:10.1016/j.ijhcs.2012.08.002

Rudin-Brown, C., & Parker, H. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. … *Research Part F: Traffic Psychology and …*.

Schroeter, R., Soro, A., & Rakotonirainy, A. (2013). Social Cars : Sensing , Gathering , Sharing and Conveying Social Cues to Road Users. In B. Guo, D. Riboni, & H. Peizhao (Eds.), *Creating Personal, Social and Urban Awareness through Pervasive Computing*. IGI Global.

Soole, D., Lennon, A., Watson, B., & Bingham, C. (2011). Towards a comprehensive model of driver aggression: a review of the literature and directions for the future. In C. N. Ferraro (Ed.), *Traffic Safety* (pp. 69–97). Nova Science Publishers, Inc.

Taylor, D. H. (1964). Drivers' Galvanic Skin Response and the Risk of Accident. *Ergonomics*, *7*(4), 439–451. doi:10.1080/00140136408930761

Viola, P., & Jones, M. (2011). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, pp. 511–518). IEEE Comput. Soc. doi:10.1109/CVPR.2001.990517

Wilde, G. J. S. (1982). The Theory of Risk Homeostasis: Implications for Safety and Health. *Risk Analysis*, *2*(4), 209–225. doi:10.1111/j.1539-6924.1982.tb01384.x

Zhang, Z., Lyons, M., Schuster, M., & Akamatsu, S. (1998). Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Automatic Face and Gesture*

*Recognition, 1998. Proceedings. Third IEEE International Conference on* (pp. 454–459). doi:10.1109/AFGR.1998.670990