Vision-based solutions for driver assistance

Adam Nowosielski
Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland
anowosielski@wi.zut.edu.pl

Abstract: The article presents a review on vision-based solutions for driver assistance. These solutions support the driver to keep safe travel conditions. They use diverse sensing modalities for the recognition of the environment around the vehicle. Upon detection a critical safety situation they supply the driver with the warning. Four assistance systems have been addressed: TSR - Traffic Sign Recognition, CAV - Collision Avoidance, LDW - Lane Departure Warning, and driver fatigue detection. Their structure and some existing approaches are presented. Furthermore, a solution for lane detection and another one for a driver fatigue detection are proposed in the article. They are prepared as the combination of existing image processing algorithms with the aim of presentation the ease of own limited solution creation. For the real-world and diverse working scenarios they would require a great deal of improvements.

Keywords: driver assistance, computer vision, TSR, Traffic Sign Recognition, CAV, Collision Avoidance, LDW, Lane Departure Warning, driver fatigue detection

1. Introduction

With the increasing development devoted to car safety it is the driver who becomes the weakest link in the security chain on the road. New solutions in the vehicle construction are introduced to prevent accidents. Special structures like crumple zones are designed to minimize the effects of accidents by absorbing the impact of the collision. The best solution, however, is to avoid risky situations. To help the deconcentrated, weary or sleepy driver many vehicles are now manufactured with the on-board driver assistant systems. Those systems are known by the ADAS acronym which stands for Advanced Driver Assistance Systems (although some authors use the DAS acronym). They use diverse sensing modalities and the computer vision approaches emerge here as one of the most important solutions.

Image-based techniques are currently widely used in many transportation related task. Depending on the purpose application computer vision solutions are developed for:

— Intelligent Transportation Systems,
— Autonomous car,
— Advanced Driver Assistance Systems.

The goal of Intelligent Transportation Systems (ITS) [1]: protect the safety, increase efficiency, improve the environment and conserve energy. The image processing techniques are used here mainly to help monitor the movement and flow of vehicles around the road network. The computer vision approaches offer the solutions for: vehicle counting, congestion calcu-
lation, traffic jam detection, lane occupancy readings, road accident detection, traffic light control, comprehensive statistics calculation.

Whereas the cameras in the ITS solutions are used as part of the road infrastructure, in the driver assistant solutions and autonomous cars cameras are the on-board equipment. They observe the road and provide the data for computer analysis. In the autonomous car the task is to substitute the driver with the altogether surrounding aware vehicle which is capable of independent movement in the traffic environment. The driver assistant solutions, on the other hand, are only created for the task of helping the driver with the driving. They offer the warning systems of different kind. The aim is to support the driver to keep safe travel conditions and to supply the driver with timely warning while a critical safety situation emerges. It is assumed that if the warned driver follows the automatic assistant recommendations the risk-burdened situation could be avoided and in effect the overall reduction in car accidents could be achieved. To inform or warn the driver the visual and auditory signals are generally used. The detection of certain emergency situations may lead to the automatic launch of vehicle braking systems taking an active part in the driving.

It is expected that vehicles should rely on the same perceptual cues that humans do [2]. Different infrastructure cuing vehicles would require huge investments and problems with the maintenance. The driver assistant systems interpret information from different sensors: radar, lidar, GPS and cameras. The paper focuses on vision-based solutions for the driver assistance. These solutions extend the capabilities of vision for the driver and allow a real-time response to occurring events. They base heavily on image processing. The following solutions are sequentially addressed in the consecutive chapters: TSR - Traffic Sign Recognition, CAV - Collision Avoidance, LDW - Lane Departure Warning, driver fatigue detection.

2. TSR – Traffic Sign Recognition

The road accidents are a serious problem for the society. One of the reasons is the omission of road signs by drivers. The traffic sign recognition has been an active field of research for several decades [3]. The achievements made in the past few years in the field allowed the implementation of solutions in real-world conditions. However, the contemporary commercial systems, installed in higher-class cars, are restricted to the limited number of sign types (usually round speed limit signs or overtaking restrictions signs) [4]. Those systems suffer from the low robustness to environmental conditions [4].

Traffic signs have been designed to be easily readable for a human. This makes even more surprising the fact that some methods outperform the human in the task. Such is the case with the Convolutional Neural Networks (CNNs) and the dataset with more than 50,000 images of German road signs gathered for the German Traffic Sign Recognition Benchmark held at IJCNN 2011 [3]. Despite this achievement, traffic sign recognition is still a challenging task. Many signs (e.g. speed limits signs) share the same general appearance. They compose a subset with elements very similar to each other. The problem becomes more difficult when various appearance of an individual sign are considered. The sign appearance changes under illumination, perspective, shadow presence, blur, and weather conditions (e.g. all kinds of precipitation). There is a problem also with partial occlusions [4].

Although humans are capable of road signs recognition in most situations achieving almost perfect accuracy, in the real-world scenario they are emerged in rich visual environment distracting from driving. A driving assistant capable of sign recognition is therefor a solution of high practical relevance.
The traffic sign recognition is a two stage problem. The first task is the sign detection followed by the classification. Since traffic signs have the form of a regular objects of specific colours, the detection process usually is based on low-level shape and color features \[4, 5\]. High cluttered background, in urban environment, may impede the task.

The classification task requires a compact representation of the image. Interestingly, some benchmark datasets, apart from the images, are provided with already computed features \[3\]: HOG features (Histograms of Oriented Gradient) \[6\], Haar-like features \[7\] and color histograms. Other low level features are also willingly used (e.g. low-frequency components of Discrete Cosine Transform \[5\]). The actual process of recognition/classification can be executed using a wide range of state-of-the-art machine learning methods \[3\]: neural networks, support vector machines, linear discriminant analysis, subspace analysis, ensemble classifiers, slow feature analysis, kd-trees, and random forests.

3. **LDW – Lane Departure Warning**

The Lane Departure Warning in a broader meaning is a road and lane understanding system. It includes the detection of \[2\]: the extent of the road, the number and position of lanes, merging, splitting and ending lanes and roads. Those tasks are expected to perform in an urban, rural and highway road conditions.

A human uses many perceptual cues for driving \[2\]: road color and texture, road boundaries, and lane markings. Those marking are used in automatic systems too. Solution such as lidar, laser, and global positioning are important complements here \[2, 8, 9\]. Although many manufactures equipped their vehicles with a lane assist solution some gaps and problems still needs to be solved \[2\]. A good survey of the approaches and the algorithmic techniques devised over the last years for different sensing techniques is presented in \[2\]. A bit older review but concentrated solely on vision-based system is presented in \[8\].

A generic system architecture for the lane understanding is provided in \[2\]. Authors claim that almost all literature lane assist algorithms can be mapped to subsystems of the proposed scheme. The scheme consist of image pre-processing and feature extraction at the beginning. The shadow removal or the image area truncation to the region below the horizon are performed at the first stage. Then, low level features are detected for the task of lane and road detection. Color, texture and lane marks are most frequently used. Those features constitute the base for the stage of road/lane model fitting. Next stage is the temporal integration that takes into account the new and previous frame information. The final step is the image to world mapping.

Here, a simple algorithm for the lane detection problem is proposed. It is prepared as the combination of existing image processing techniques with the aim of presentation the ease of own limited solution creation. Figure 1 presents two examples of lane detection. The algorithm used for the presentation purposes was created on the base of Hough transform. First, the image is truncated by removing the region above the horizon. Then, the image is binarized using the Otsu method \[10\]. Since the lane markings may take the form of a dashed line a temporal blurring is used (similarly as in \[9\]). The averaging should be adjusted according to the vehicle velocity and the exposure time. Here, the averaging of past five frames is employed. Line breaks are removed this way and the appearance of a long and continuous line is obtained. Then, the image is filtered using selected frequency domain features of block DCT components (the method proposed in \[11\] is employed). The final stage of the algorithm is line detection on the base of the Hough transform. It must be noticed that the curvature of
the road changes gradually. The lane markers are only slightly affected at the entry and exit points of the curve (see the right-hand side image at Fig. 1).

![Figure 1. Example of the lane detection](image)

The above presented solution is a simple combination of existing image processing algorithms suitable for the lane detection task. There are many solutions of that kind in the technical literature. It is said [9] that simple Hough transform-based algorithm solves the problem in roughly 90% of the highway cases (in addition without employing any tracking or image-to-world reasoning). It must be noticed, however, that real-driving scenario requires high reliability under miscellaneous case conditions.

4. CAV – Collision Avoidance

Collision Avoidance (CAV) systems form a group of the most advanced systems for driving assistance [12]. They are also known as anticollision systems. There also exists CM systems which goal is to mitigate the effects of collisions (Collision Mitigation) [12]. Both solutions warn the driver upon detection of the dangerous situation. Both are active safety systems and can use the braking system. The CAV system, however, is also able to swerve to avoid a potential collision [12].

The recognition of the environment around the vehicle is the crucial aspect of collision avoidance. The system should be able to detect obstacles of many kind: pedestrians, animals, other vehicles or static elements of road infrastructure. Prototypes of such systems are known in the literature for a long time (e.g. [13]). Figure 2 presents two examples from Volvo proprietary City Safety system. The system appeared in 2006 and car models for 2015 are equipped with the updated version. According to the manufacturer the new City Safety system is capable of recognition the following objects (also at night): pedestrians, vehicles, bicycles and big animals. The upper image in Fig. 2 depicts the detection of other vehicles. The bottom image presents the dangerous situation in which a cyclist rode from the side of the road.

Three type of environment sensors are used in todays implementations of CAV systems [12]: radar, lidar, and video camera. Existing solutions are usually hybrid and employ some kind of information integration [14]. Previous generation systems - ACC (Adaptive Cruise Control) systems - were mostly based on radar or lidar sensors [12]. Their task was to maintain safe distance to the primary target (by automatic acceleration or braking) and to resume the preset cruising speed when no such targets were detected by the system [15]. New generation systems (CAV) require the presence of video camera to perform the object recognition task.
Attempts to use cameras have already been undertaken in case of the previous generation systems (ACC). The exemplary solution is presented in [15].

The most important element of the CAV system is object detection. Vehicles and pedestrians are of great importance here. The scientific literature in that field is extensive. The detection procedure is typical to all image processing tasks - after the detection of candidates a classification step is required. The most commonly used general purpose segmentation methods for moving objects detection include [16]: optical flow method, frame difference method, background subtraction method. The classification usually bases on shape, motion and cooccurrence matrix [16].

Not all methods are applicable to a moving camera environment. Since the vehicle is in motion the internal camera is registering a moving scene which for example is a problem for background subtraction methods (due to the car movement static scene objects change their position in image plane). In such a case a sliding window approach (that considers an individual frame) may be a good solution. The procedure of sliding window considers all possible window scales and in all possible image locations. They are examined for the presence of the learned pattern (e.g. vehicle). A standard multipurpose Viola and Jones detector [7] can be successfully used for vehicle and pedestrian detection. Its real-time implementation
is possible. Furthermore, various solutions have been proposed in the scientific literature for vehicle (e.g. [17, 18]) or pedestrians (e.g. [6, 19]) detection.

Manufactures of premium class vehicles offer for their users as an optional equipment the thermographic camera. In poor environment condition (bad weather or night) such camera offers preview of the road much further than the range of headlights. Far-infrared or thermal night vision is a perfect technology for the detection of pedestrians at night [20]. A solution presented in [20] is based on infrared technology. Histogram of Oriented Gradients (HOG) [6] are used here for feature extraction. They are classified using Support Vector Machine [21]. In addition, after the successful detection human silhouettes are tracked using a Kalman filter [22].

5. Driver fatigue detection

The obvious and indisputable matter is the correlation of driver drowsiness with road accidents. The fatigue causes loss of concentration and prevents the driver from making quick and correct decisions [23, 24]. Many studies have been conducted to estimate the influence of weariness on driving behavior but the fact is that there is no simple, reliable way to determine whether the fatigue was a factor in an accident afterward and what level of fatigue the driver was suffering. It is estimated that around 20% of all road accidents are fatigue-related [24, 25]. The driver tiredness is then a considerable contributory factor in road accidents.

By monitoring the fatigue of the driver, the vehicle safety can be improved [23]. After the detection of drowsiness the driver assistant system is able to stimulate the driver with some auditory signals (e.g. play some lively music) or in critical situation may even stop the vehicle (after the sleep detection).

The driver fatigue/drowsiness detection system helps prevent accidents. It can recognize the driver steering patterns. Some inference can be conducted from the already described (in Sect. 3) lane monitoring system (attentiveness to the lane-keeping task [8]). Another possibility is the measurement of physiological driver parameters like heart rate or brain activity. These techniques however require some uncomfortable sensors to be placed on the body. The most convenient and noninvasive are vision-based driver monitoring systems. Drowsy drivers exhibit some observable behaviors in head and eyelid movement or eye gaze [23]. Computer vision techniques allow the constant observation of the driver face and eyes. It is possible to detect the state of the eyes (open, partially open or closed) and to discover the face or gaze direction. When the driver is drowsy the eyes start to close. Long lasting closure or downward diverted sight are factors indicating the increased risk situation. The most important issue here is the visibility of the eyes which in the case of dark sunglasses is a condition impossible to fulfill.

Face and eye detection are two important initial steps in vision-based driver fatigue detection systems. They are usually combined in a serial architecture, i.e. the localization of the eyes is limited to the already detected face region. Eye tracking is the next crucial step in the driver monitoring system. By comparing the temporal characteristics of the eyes region the appropriate conclusion of driver fatigue can be drawn.

The listed above tasks are classical face recognition problems. The face recognition technology has reached its maturity and under controlled conditions is regarded as solved [26]. It still remains, however, a number of issues for consideration. Nevertheless, for the face detection a well-functioning procedures exist. In last decades numerous methods of face detection were invented and a good classical review is presented in [27]. The following
categories for face detection have been proposed [27]: knowledge-based methods, feature invariant approaches, template matching methods, appearance-based methods. The survey [27] from the period of greatest interest in the field does not include, however, the Viola & Jones method which now is considered as universal, very good solution and the leading choice option for those opting for the ready to use solution. The method takes advantage form statistical boosting and uses Haar-like features proposed by Viola & Jones [7]. It is used to create state-of-the-art detectors (e.g. face detector [28]). The implementation is available for example in the common use OpenCV library (the Open Source Computer Vision library released under a BSD license, available at http://opencv.org/). The OpenCV implementation uses improved by Lienhart Haar-like features [29]. There is a possibility to train the cascade classifier on LBP (Local Binary Patterns) features [30]. The OpenCV library is also equipped with the ready-to-use eye detector constructed upon the same algorithm.

By employing existing algorithms and image processing techniques it is possible to create an individual solution for driver fatigue/drowsiness detection. The example depicted in Fig. 3 demonstrates the procedure of the closed eye detection. The OpenCV face and eye detectors are used sequentially to localize face and then the eyes. Simple feature extractor based on the two dimensional Discrete Fourier Transform (DFT) [31] is used to represent an eye region. The extractor selects low frequency coefficients from the spectrum (modulus values are considered). Those coefficients carry a substantial part of image energy. The modulus representation of the DFT is invariant to eye translation within the image plane which mitigates the impact of inaccuracies in eye localization. The details of the extraction method are provided in [31]. The eye image representation from the current frame is compared with five previous frames on the base of Euclidean distance. Averaged the resultant distance is presented on the chart (bottom part of Fig. 3). Human involuntary blinking are clearly visible in the chart and can be detected with a threshold value. Such procedure has a good toleration to light changes since only actual frames are considered (few previous frames). The function stabilizes in case of lack of changes. Since blinks are involuntary this is a good indicator of closed eye. This straightforward proposition presents the potential of face image processing methods but the practical implementation would require a great deal of improvements. Although driver’s face and eye detection are performed in moderate stable conditions (the distance and positions are almost constant) some unfavorable lighting conditions (night) may hinder the task. Usage of special equipment like infrared camera or lighting in such a case is indispensable.

There are many similar solutions in the scientific literature. One example is presented in [23]. It also employs the OpenCV implementation of face and eye detectors. The detection is supported by frame difference mechanism which aim is to localize the eyes by blinking. The template of the user’s open eye is created. Using this template the new eye position is localized on the base of template matching in a search region restricted to a small area from previous coordinates. The eye tracking is achieved in this manner. The best correlation coefficient indicates the new localization. The correlation coefficient indicates also the blinking. The correlation score is used solely for the detection of blinking and the analysis of blink duration. For very low values the system is assumed to lost the location of the eye and the procedure is restarted. High values indicate the open eye state and decreased values are identified as the blinking. The duration of the blink is the determinant of fatigue of the driver. It seems that in driving environment with dynamic high-contrast lighting (shaded and unshaded areas occurring one after the other) the presented algorithm would constantly reinitialize.

More sophisticated solution for fatigue monitoring is presented in [24]. The system operate in the visual and near infra-red (NIR) spectra. The infrared spectrum is invisible for the
driver and allow to use the additional infrared light source (NIR LED) for the night conditions and poor visibility. In each case (visible and infrared) two cascade of classifiers are used, one for the detection of the eye and the other for the verification stage. In the visual spectrum the color segmentation is used to detect skin regions. The adaptively window growing method (AWG) is utilized. Candidate regions are verified by the classifier operating with the higher order singular value decomposition (HOSVD) of the tensor of geometrically deformed images of real eye prototypes [24]. The eye detection in the infrared spectrum starts with the detection of pupil. Eye candidates are evaluated according to predefined geometrical relations and after that they are verified with the HOSVD classifier (trained with patterns specific to the NIR conditions). The driver drowsiness and fatigue is evaluated on the base of periods of closed eyes calculation. The percentage of eye closure mechanism for the detection of driver’s state is called PERCLOSE [24]. Such approach is simple and there is the need for a more complex analysis for the real life scenario.

6. Conclusions

The article reviewed vision-based solutions for the driver assistance. Those solution are designed for the improvement of safety and driver convenience. Driver assistance solutions received a significant research attention and now are successfully incorporated into vehicles.

Four assistance systems have been addressed in the article: TSR - Traffic Sign Recognition, CAV - Collision Avoidance, LDW - Lane Departure Warning, driver fatigue detection. Their structure and sample solutions have been presented. Most solutions share the same functional modules. They are, however, implemented differently in individual systems. The field of image processing is currently in such a stage of development that it is possible to propose for a given task a limited solution as combination of existing image processing algorithms. For the evidence two approaches have been proposed: for the lane detection and driver fatigue detection. These straightforward propositions present the potential of image processing methods. The practical implementation would require a great deal of improvements. The driver assistance system should reach low error rates. Otherwise, irritated drivers will reject any warning solution characterized by a large number of false alarms.
References


