A Monocular Vision Advance Warning System for the Automotive Aftermarket

Itay Gat
Mobileye Vision Technologies Ltd.
Jerusalem, Israel
itay.gat@mobileye.com

Meny Benady
Mobileye Vision Technologies Ltd.
Jerusalem, Israel
meny.benady@mobileye.com

Amnon Shashua
Hebrew University
Jerusalem, Israel
shashua@cs.huji.ac.il

September 29, 2004

Paper Offer #: 05AE-104
Session: AE7

Abstract

Driver inattention and poor judgment are the major causes of motor vehicle accidents (MVA). Extensive research has shown that intelligent driver assistance systems can significantly reduce the number and severity of these accidents. The driver's visual perception abilities are a key factor in the design of the driving environment. This makes image processing a natural candidate in any effort to impact MVAs. The vision system described here encompasses 3 major capabilities: (i) Lane Departure Warning (ii) Headway Monitoring and Warning (iii) Forward Collision Warning. This paper describes in detail the different warning features, the HMI (visual and acoustic) application design rules, and results of a study in which the system was installed in a commercial fleet and passenger vehicles.

1 Introduction

While the number of cars manufactured each year continues to grow - so do the figures of motor vehicle accidents. The alarming data show that around 10 million people around the world are injured in MVA each year. 20-30% of them are severally injured and around 400,000 are fatal injuries, resulting in death [1]. Research has shown that driver inattention and poor judgment are the major causes of MVAs. Extensive research has shown that intelligent driver assistance systems can significantly reduce the number and severity of these accidents. [2].

Controlled experiments have shown that when vehicles are equipped with crash warning systems, accidents are reduced by 78%. Providing the driver with 0.5 second alert to a rear-end collision can prevent as much as 60% of this type of accidents. The figures are even more impressive if the alarm is given 1 second in advance: a reduction of up to 90% is achieved (cited in [3]).

According to the National Highway Traffic Safety Administration (NHTSA), more than 43% of all fatal MVAs reported in 2001 involved a lane or road departure. This statistic increases every year, making it the single largest cause of automotive highway fatalities in the United States alone [4, 5].

NHTSA estimates that more than 1.5 million police-reported MVAs involve some form of driver inattention: the driver is distracted, asleep or fatigued, or otherwise "lost in thought". Driver distraction is one form of inattention and is a factor in more than half of these MVAs. The presence of a triggering event distinguishes a distracted driver from one who is simply inattentive or "lost in thought."

In most cases, failure to maintain safe headway can be attributed to driver inattention and/or misjudgment of distance. It has been shown that drivers tend to overestimate their headway and consequently drive with short and potentially dangerous headway [6]. An intelligent system may serve a double purpose in this case, both as an alert, and as an tool for "educating" the drivers. It was further shown that even imperfect systems are quite helpful in positively impacting drivers' habits [7].

Current solutions and alternative technologies

Most popular solutions today are based on Radar and Lidar technologies. The radar technology measures reflections from metal objects, and takes into account the Doppler effect in order to provide relative speed information. The Lidar systems use laser-beams to measure distance.

The high cost of radar systems limits their usefulness and the are on the whole restricted to high-end vehicles. Although Radar is unaffected by weather and lighting conditions, sensor data from the radar is extremely limited in the context of trying to interpret an extremely complex and dynamic environment. In most cases, the combination of smart processing with radar data works well for the constrained application of a distance control in highway driving, but there are situations where no matter how much processing is performed on the radar data, the data itself does not reflect the environment with a high enough fidelity to completely interpret the situation. Spatial resolution is relatively coarse for the detected field of view, such that detections can be improperly localized in the scene and object size is impossible to determine. The effects of this are that small objects can appear large, radically different objects appear similar, and position localization is only grossly possible. This leaves room for improvement, which becomes important as the sensing technologies are applied toward safety features.

Given that the the driving environment is designed around the human driver's ability for visual perception it may look natural to search for vision solutions. Therefore, another family of solutions is based on image systems with two sensors that can provide depth information in the image. Such systems are still rather expensive, require accurate calibration among the cameras. Moreover, the ability to provide depth information is only for the short range (up to 20m), whereas most of the vehicles on the road are much farther away from us.

Monocular vision systems are starting to emerge but they are usually focusing only on one aspect of the problem - e.g. lane departure warning [4]. It turns out that in many situations providing warning based on one modality may be too limited. For example lane departure system would gain a lot from insertion of information about vehicles on the road (blocking the view on the lanes). Furthermore, higher level of information about lanes can be of aid - for example unstable driving within a lane (indicated by lateral velocity) may be an important indication of intelligent systems.

This paper describes Mobileye's Advance Warning System (AWS) product which is based on technology enabling detection and accurate measurement of lanes, road geometry, surrounding vehicles and other information using monocular camera. The AWS description with its underlying. technology is given in the second section, whereas the third section provides thorough analysis of system performance. Finally the conclusion summaries the system capabilities.

2 System description

The technology of Mobileye enable the detection of lane marks and vehicles in complex environment together with various measurements on these objects. The analysis is carried out using a monocular vision system thus creating a reliable and cheap solution.

2.1 Building a detection scheme

Mobileye's detection system architecture loops through the following modules:

1. **Generate candidate regions of interest:** a systematic scan of the image for rectangular shaped regions at all positions and all sizes would be computationally unwieldy. An attention mechanism filters out windows based on lack of distinctive texture properties and in-compliance with perspective constraints on range and size of the candidate vehicle. On average, the attention mechanism generates 75 windows (out of the many thousands of candidates which could be generated otherwise) per frame which are fed to the classifier.

2. **Single frame classification:**

   The core of the detection process lies in the classification stage, where each region of interest is given a score that represent the likelihood of that region to be a vehicle. We are using several classification schemes throughout the system and they can be rather degenerate such as the nearest neighbor approach which employs relatively sophisticated local features such as those used by [12], or integration via a cascaded classifier such as the hierarchical SVM approach used by [11]. A particulary powerful scheme we employ borrows from the idea of the recognition-by-components using a 2-stage classifier algorithm. Namely, we breakdown the region of interest into sub-regions, create a local vector representation per sub-region, feed each of the local feature vectors to a discriminant function and integrate the local discriminant results by a second-stage classifier. The crucial difference from the conventional paradigm is the way we handle the training set. Since the number of local sub-regions are small we generate multiple local discriminants (one per local sub-region) by dividing the training set into mutually exclusive *training clusters*. The idea behind the subset division of the training set is to breakdown the overall variability of the class into manageable pieces which can be captured by relatively simple component classifiers. In other words, rather than seeking sophisticated component classifiers which cover the entire variability space (of the subregions) we apply prior knowledge in the form of clustering (manually) the training set. Each component classifier is trained multiple times —

once per training cluster — while the multiple discriminant values per subregion and across subregions are combined together via Adaboost [13].

3. **Multi-frame Approval Process:** candidates which survive the single frame classification thresholds are likely to correspond to vehicles. However, due to the high variability of the object class and the high levels of background clutter it is conceivable that coincidental arrangements of image texture may have a high detection score — an ambiguous situation which is likely to be unavoidable. Additional information collected over a number of frames are used in the system for further corroboration.

4. **Range measurement:** a more detailed description of our range and range-rate measurement are given in the next section.

The four basic steps above are also coupled with supporting functions such as host vehicle ego-motion (of Yaw and Pitch) [10], robust tracking — and of primary importance the classification scores of background sub-classes which include licensed vehicles, poles, guard-rails, repetitive texture, lane mark interpretation, bridges and other man-made horizontal structures. The sub-class scores play an important role in the final decision-tree multi-frame approval process.

2.2 Providing range and range-rate using monocular vision

Range to vehicles and range-rate are two important values required for any vision-based system. In this section only the essence of the algorithm is described where a full description can be found in [8].

As the data is collected from a single camera range must be estimated by using perspective. There are two cues which can be used: size of the vehicle in the image and position of the bottom of the vehicle in the image. Since the width of a vehicle of unknown type (car, van, truck etc) can vary anywhere between 1.5*m* and 3*m* a range estimate based on width will have only about 30% accuracy.

A much better estimate can be achieved using the road geometry and the point of contact of the vehicle with the road. We assume a planar road surface and a camera mounted so that the optical axis is parallel to the road surface. A point on the road at a distance Z in front of the camera will project to the image at a height y, where y is given by the equation:

$$y = \frac{fH}{Z} \tag{1}$$

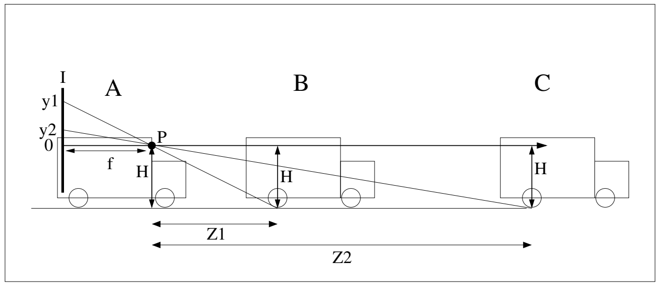


Figure 1: *Schematic diagram of the imaging geometry (see text).*

where H is the camera height, and f is the focal length of the camera (both given in meters).

Figure 1 shows a diagram of a schematic pinhole camera comprised of a pinhole (P) and an imaging plane (I) placed at a focal distance (f) from the pinhole. The camera is mounted on vehicle (A) at a height (H). The rear of vehicle (B) is at a distance (Z_1) from the camera. The point of contact between the vehicle and the road projects onto the image plane at a position (y_1). The focal distance (f) and the image coordinates (y) are typically in *mm* and are drawn here not to scale.

Equation 1 can be derived directly from the similarity of triangles: $\frac{y}{f} = \frac{H}{Z}$. The point of contact between a more distant vehicle (C) and the road projects onto the image plane at a position (y_2) which is smaller than (y_1).

To determine the distance to a vehicle we must first detect the point of contact between the vehicle and the road (i.e. the wheels). It is than possible to compute the distance to the vehicle:

$$Z = \frac{fH}{y}. \tag{2}$$

Figure 2 shows an example sequence of a truck at various distances. The distance from the horizon line to the bottom of the truck is smaller when the truck is more distant (a) than when it is close (b and c).

This outcome fits in with our daily experience, that objects that are closer to us are perceived as lower then objects that are further away. The relationship demonstrated here may be further used to deal with situations in which the assumption of planner road does not hold (e.g. starting to climb a hill or bumps on the road).

Beyond the basic range estimation, it is also important to provide information about the range-rate or the time it would take to cross the distance to current in path vehicle - *time to contact*. The human visual system is able to make very accurate *time to contact* assessments based on the retinal divergence (scale change of the target) and, therefore can be used in our monocular system. It can be shown that

(a)

(b)

(c)

Figure 2: *A typical sequence where the host vehicle decelerates so as to keep a safe headway distance from the detected vehicle. The detected target vehicle (the truck) is marked by a white rectangle. As the distance to the target vehicle decreases the size of the target vehicle in the image increases.*

if one assumes constant velocity, it is possible to obtain a simple relationship between the scale change and the time-to-contact:

$$TTC = \frac{\Delta t}{S-1}. \quad (3)$$

where TTC is the time to contact, S is the scale change between two consecutive images and Δt is the time difference between them. It is possible to add acceleration and deceleration into the computation [9].

2.3 Application

The Advance Warning System (AWS) provides a set of warnings for the driver based on the vision technology described above:

- **Lane Departure Warning (LDW)** The LDW module detects lane boundaries, finds the road curvature, measures position of the vehicle relative to the lanes, and provides indications of unintentional deviation from the roadway in the form of an audible rumble strip sound. The system can detect the various types of lane markings: solid, dashed, boxed and cat-eyes, and also make extensive use of vehicle detection in order to provide better lane detection. In the absence of lane markings the system can utilize road edges and curbs. It measures lateral vehicle motion to predict the time to lane crossing providing an early warning signal before the vehicle actually crosses the lane. Lane departure warnings are suppressed in cases of intentional lane departures (indicated by activation of turn signal), braking, no lane markings (e.g. within junctions) and inconsistent lane markings (e.g. road construction areas).

- **Headway indication and warning** The headway monitoring module provides constant measurement of the distance in time to the current position of the vehicles driving ahead in the same lane. The ability to indicate current in-path vehicle is dependent upon the information from the lanes detection module. While insufficient distance keeping is a major cause of MVAs, it is difficult for many drivers to judge this distance correctly while considering the traveling speed of the vehicle. The AWS headway display provides a visual indication when insufficient distance is being kept to the vehicle ahead, as well as a clear numeric display (in seconds) which provides an accurate cue for driving habits improvement for the driver.

- **Forward Collision Warning (FCW)** The FCW module continuously computes time-to-contact to the vehicle ahead, based on range and relative velocity measurements. An advanced image processing algorithm determines whether the vehicle ahead is in a collision path (even in the absence of lane markings) and provides audio warnings to the driver at predetermined time intervals prior to collision (e.g. 2.5, 1.6 and 0.7 seconds) alerting the driver to the danger and allowing appropriate action such as braking or steering away from the obstacle ahead [9]. The system uses information about driver actions (e.g. braking) to suppress warnings in situations that are under the driver's control.

2.4 AWS applicative overview

The AWS system consists of:

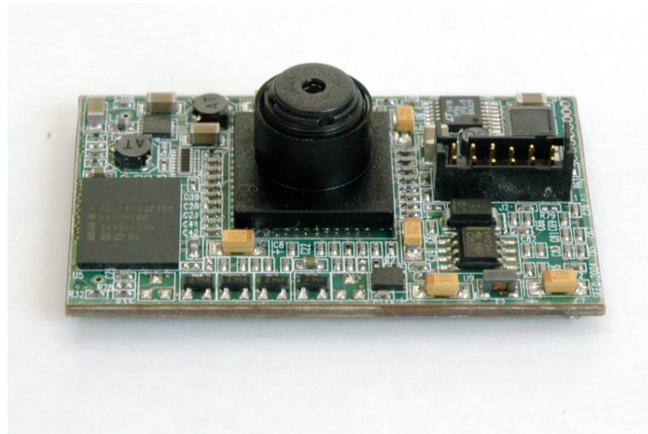
- **SeeQ** Real-time image processing unit running at 30 FPS on the *EyeQ* vision system on chip that include a Compact High Dynamic Range CMOS (HDRC) camera. The units' size is 3 over 5 cm and it is located on the windscreen (see figure3).
- Display and interface unit located in front of the driver
- A pair of loudspeakers for providing directional warnings

The interface unit is also connected to signals from the vehicle (vehicle speed signal - VSS, indicators and brake). The system is turned on shortly after ignition. An example of the video display is provided in figure 4. In this situation the system is active and has detected a vehicle in current path. The headway value for this vehicle is 1.2 seconds which is enough headway according to these settings. The driver sensitivity mode is indicated in the example as *NORMAL*.

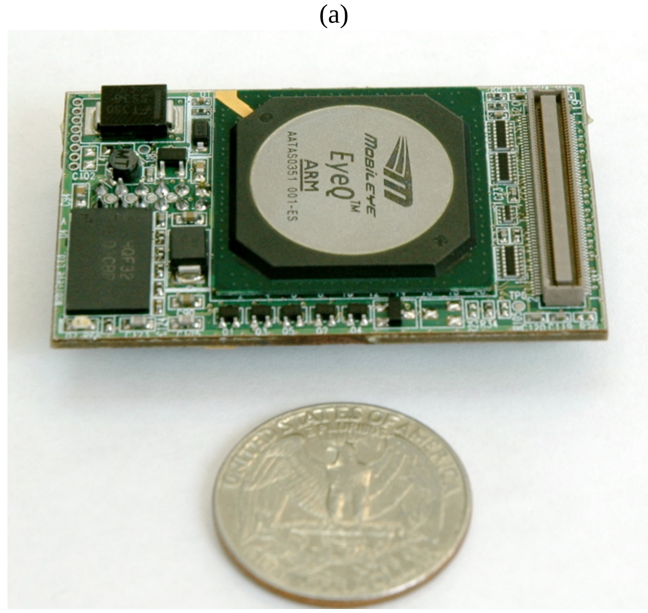
There is a small set of commands that the driver can pass to the AWS. Among them are a volume control and sensitivity control. Making the system more sensitive raise the frequency of alerts - for example, alerts when the driver is first getting close to crossing lanes are provided.

Apart from the warnings, the system also provides a visual indication of its availability. The system may be unavailable under two conditions:

1. **Failsafe** the system has low visibility and cannot provide alerts at this period. The problem may in some cases be a temporary one (e.g. low sun causing the image to be unusable, or dirty windscreen)
2. **Maintenance** The system is not receiving the required signals for computation (either the signal from the camera or speed signal from the vehicle).



(a)



(b)

Figure 3: *View of the SeeQ. (a) a road facing view of the camera. (b) Processing unit at the rear side of SeeQ*

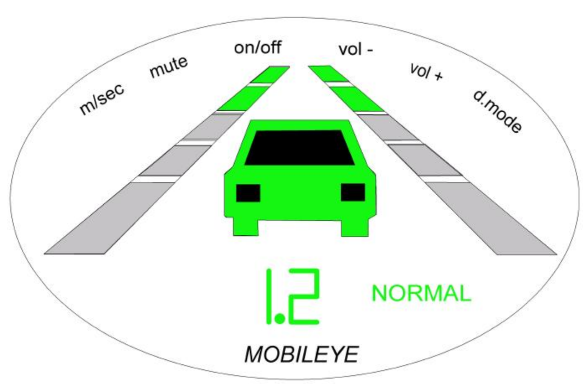
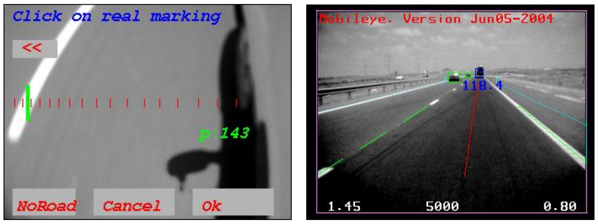


Figure 4: *Example of AWS display.*



(a) (b)

Figure 5: *Simultaneous capture of frontal and road images. (a.) Image of Road near the left wheel with 10cm markings from the wheel (shown in red). The distance from lane mark obtained by AWS (1.43m) is shown in green. (b). Frontal image showing vehicle detection (blue and green) and lanes (green). Note that the lane mark segment appearing in (a) is also visible at the lower part of (b).*

3 Performance and results

In order to evaluate the performance of the AWS system, two types of evaluations were carried out:

- Quantitative analysis: measured performance of each feature of the system.
- Qualitative analysis: effectiveness of the system as evaluated by professional and non-professional drivers.

3.1 Performance envelope

In order to facilitate the examination of accuracy of the AWS the following infrastructure was used:

Lane departure warning

A dual camera system was used in order to measure the accuracy of lane departure warning. The usual view of the driving scenario was accompanied with a camera placed above the wheel and looking downward on the road. The images coming from both cameras were synchronized by using an interchangeable sampling scheme. A view on the outcome of such a system is shown in figure 5.

headway control

The two issues that need to be examined in this situation are the availability of vehicle detection and the accuracy of distance measurement. The availability of the AWS is measured by manually inspecting video clips recorded while driving. The accuracy is measured by comparing the distance estimation of AWS with that obtained by Radar systems recorded simultaneously. A matching algorithm was used to synchronize the radar measurements and the ones provided by AWS.

(a) (b)

Figure 6: *The remote mounting structure.*

Forward collision warning

The performance of this capability is the most difficult to measure, as situations in which our vehicle is within less then 2 seconds from collision with the vehicle in path are rare and dangerous. The only solution is to simulate such conditions, either using balloon cars, or by placing the camera at a safe distance from the actual vehicle. We chose the second solution, and used a camera placed on a large Rod positioned above the vehicle. This camera displayed images of what a vehicle in crash situation could ”see”. The infrastructure used for this test is shown in figure 6.

3.2 Qualitative assessment

In field tests (clinic) the system was installed in several vehicles, drivers were instructed on the use of the system and were asked to drive normally and report the performance of the AWS. The most important question arising from this clinic was whether the system contributed to the safety feeling of the driver.
In order to asses this information we asked the drivers to answer several questionnaires:

- Pre driving questionnaire, in which we addressed the characteristics of the drivers (e.g. age, gender) and expectations from such system.
- After each drive: to report specific problems.
- After the whole process: we addressed more high level issues.

An offline kit that enables the recording (both video and log data) of alerts and other related events was installed in each vehicle. Throughout the drive, the following general information was also collected: as:

- Distance traveled
- Speed
- Time of day
- Eye gaze towards the AWS display

The information is currently being processed, and final results will be available soon.

3.3 Results

The performance envelope of the AWS was examined in various scenarios (e.g. highway, urban, day, night). The following statistics were based on a database of 30 hours of driving data randomly selected from over 300 hours of data from Europe, Japan, USA and Israel.

- **Lane Departure Warning** Availability values obtained were:
  - Well marked highway (day/night): 99.5%
  - Poorly marked highway (day/night): 97.5%
  - Country road (day/night): 98.6%
  - Bott's dots (day/dusk): 89.8%

  The average absolute error in measuring position of the lanes was 5cm and the false warning produced were less then 1 per 2 hours of average driving.

- **Headway control** The availability of vehicles detection was 99.5% The accuracy as measured in average absolute error in range is:
  - Vehicles up to 40m: 5.6%
  - Vehicles up to 60m: 6.0%
  - Vehicles up to 80m: 6.4%
  - All vehicles: 7.6%

  A typical performance is shown in Figure 7.

- **Forward Collision Warning** In all of the testing carried out using the collision simulation the warning was produced. The average time in advance was 1.5 sec. The false rate of the system was less then 1 false per 5 hours of driving and even those falses were short in nature (lasted less then 0.5 seconds). A typical performance is shown in Figure 8.

The subjective assessment in the clinic process is currently being evaluated.

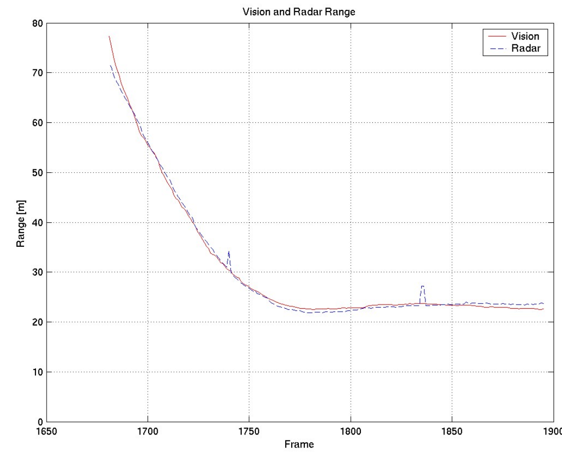


Figure 7: *Range for a typical sequence where the host vehicle decelerates so as to keep a safe headway distance of 24m from the detected vehicle. The lead vehicle was traveling at 80KPH.*

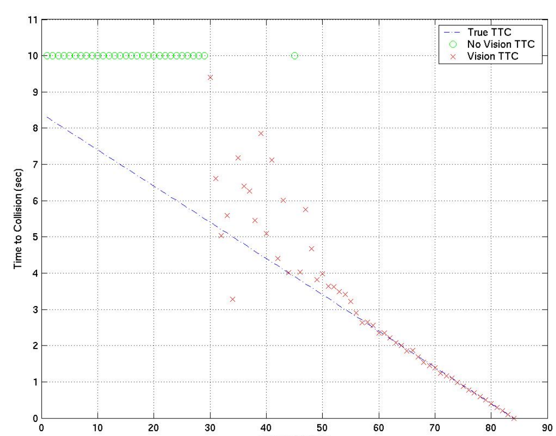


Figure 8: *Time to contact computed using vision divergence. Using the remote mounting infrastructure it is possible to compute the exact time to contact and to compare the computed results with it. The results show that reliable estimate is given up to 4 seconds before collision.*

4 Conclusion

The AWS is an advance system for the automotive aftermarket that offers a suite of active safety applications for accident reduction. The main purpose of the AWS system is to alert the driver and increase the awareness of drivers to dangerous situations that may be caused by weariness or by other distractions during driving.

Based on a single camera located on the front windscreen, the AWS detects and tracks vehicles on the road ahead providing range, relative speed, and lane position data. In addition, the system detects lane markings and measures and monitors distance to road boundaries.

A small display unit and a pair of left and right speakers inside the car provide timely audio and visual warnings, allowing the driver to react to various types of dangerous situations and to reduce the risk of accidents. It was demonstrated that a single low-cost camera can provide this information with satisfactory accuracy. The application built around the technology is such that it keeps a balance between supplying sufficient information and avoiding a surpass of alerts, that the driver would find annoying. Furthermore, the driver has the option of influencing the system's sensitivity and controlling the alert level provided. The performance of the system shows that both from the performance and the applicative points of view the system provides extra-value for the driver, and can reduce accidents and help in education of drivers for safer drive.

A second generation product that enhances the system capabilities and includes pedestrian protection is now being developed using advanced implementations of the principles used in the current AWS.

References

[1] Organization for Economic Cooperation & Development - Paris.

[2] The National Safe Driving Test & initiative Partners (CNN: http://edition.cnn.com)

[3] National Transportation Safety Board, Special Investigation Report - Highway Vehicle- and Infrastructure-based Technology For the Prevention of Rear-end Collisions. NTSB Number SIR-01/01, May 2001

[4] Iteris (http://www.iteris.com)

[5] P. Zador, S. Krawchuck and R. Voas Automotive Collision Avoidance System (ACAS) Program/First Annual Report. NHTSA - National Highway Traffic Safety Administration (http://www.nhtsa.dot.gov/) DOT HS 809 080, August 2000

[6] National Safety Council, Defensive driving course [Course Guide] 1992.

[7] A. Ben-Yaacov, M. Maltz and D. Shinaar , Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance. In *Hum Factors.*, pages 335 -342, 2002

[8] G. Stein and O. Mano and A. Shashua Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy In *IEEE Intelligent Vehicles Symposium (IV2003)*, June 2003

[9] O. Mano, G. Stein, E. Dagan and A. Shashua. Forward Collision Warning with a Single Camera., In *IEEE Intelligent Vehicles Symposium (IV2004)*, June. 2004, Parma, Italy.

[10] G. Stein, O. Mano and A. Shashua. A Robust Method for Computing Vehicle Ego-motion In *IEEE Intelligent Vehicles Symposium (IV2000)*, Oct. 2000, Dearborn, MI.

[11] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 23:349-361, April 2001.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004.

[13] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In Proceedings of International Conference on Machine Learning (ICML), pp. 148–156, 1996.