# Perception and data fusion on autonomous vehicles: the TerraMax™ experience

P. Grisleri[1], P. Cerri[1], S. Cattani[1]

1: VisLab - Università degli Studi di Parma, via G.P. Usberti 181/A, 43100 Parma - Italy

**Abstract**: This work proposes an implementation of Data Fusion for autonomous vehicles. In particular this work presents the fusion approaches developed for the TerraMax autonomous vehicle which competed in the DARPA Urban Challenge. The TerraMax Obstacle Detection system combines the positive qualities of cameras and LIDARs. LIDAR raw scan points are directly processed from an ECU to produce a list of tracked objects. The vision system performs a low level fusion between the disparity image obtained from a calibrated stereo camera pair and the raw scan data. This block still produces, after a low level tracking step performed using inertial data, a list of obstacles, but their consistency is augmented since vision can distinguish objects from noise, like dust or rain. The two independent lists of objects are used for a high level fusion step; tracking and other low cost techniques are used to increase obstacle persistence.

**Keywords**: Sensor fusion, autonomous vehicles, vision, LIDAR

## 1. Introduction

Unmanned ground vehicles (UGVs) need strong and robust sensing and perception capabilities to accomplish their tasks. Autonomous driving in urban environments is an extremely complex task due to moving obstacles usual of these scenarios. The highly dynamic nature of this problem (scenarios, background, illumination, obstacles) and the different features to be detected and tracked, such as pedestrians, vehicles, traffic signs and lights, suggest the combined use of multiple sensors to enhance the quality of perception.

Vision sensors passive and low cost show their strength when classification is needed while LIDAR sensors provide direct measurements for quantities such as obstacle distance and shape, or info for non-textured objects in the 3D space. On the other hand, vision can perform a better obstacle classification due to the higher amount of information available, and easily detect textured areas like lanes.

## 2. Sensor fusion

Each sensor has its own peculiarities due to the physical principles it is based on, and to the different results of the processing that can be exploited on its particular kind of data.

These peculiarities can be both strengths and limitations, depending on the kind of detection the sensor is employed for. The obvious extension to this basic scheme is to combine positive results from each sensor (lowering missed detections/false negatives) reducing the weakness of the overall system (like false positives).

Data fusion can also be exploited between detectors of different sorts. For example the output of a LIDAR system has been used to define the search range of a vision based lane detector.

Fusion can be implemented in different ways [1][2][3] and with different sensors [4]. In this work we present the two level fusion between vision systems and LIDARs that has been implemented on the TerraMax vehicle, depicted in Figure 1, when raced in the DARPA Urban Challenge. The truck was equipped with 11 cameras and 3 LIDARs. Two PCs were dedicated to the raw scan-data processing while 4 vision PCs were differently connected to both cameras and LIDARs performing image processing and low level fusion. Vision systems and LIDAR systems were connected to another module of the autonomous driving system named World Perception Server who was in charge of performing the high level fusion and tracking between the results of each sensor.

**Figure 1. The TerraMax autonomous vehicle.**

The truck successfully completed, together with other 10 top teams, the qualification process and participated to the DARPA Urban Challenge Event in Victorville, CA, in November 2007.

The experience done with TerraMax is being transferred to the new VisLab prototype Grandeur (Figure 2) built in collaboration with Mando Corp. and equipped with a LIDAR and a low cost camera for the detection of pedestrians in dangerous areas.

**Figure 2. The Hyundai Grandeur able to brake in front of dangerous pedestrians.**

## 3. Application on TerraMax

This section describes the implementation of the proposed sensor fusion on the TerraMax vehicle, starting from the hardware sensors used and going then through a subsection dedicated to the synchronization issues which are particularly critical during the realization of perception systems made up by different sensors.

The third subsection will be dedicated to the ego motion sensors, which are of basic importance for the autonomous guidance. Then a subsection is dedicated to the detailed description of the fusion implementation, while the last subsection describes the four perception systems based on vision together with their purposes.

### 3.1 Sensors

TerraMax has four Vision Systems on board: Trinocular, Stereo, RearView and Lateral. Each vision system is made of one COTS ruggedized computer. All the processing systems are boxed under the passengers seats where a direct air conditioning blow keeps them cooled.

Each system is differently connected, through the 800Mbps IEEE1394b bus or through the Giga-Ethernet port, to a subset of the sensors, depending on the application (see Table 1).

One or more detection algorithms, described in section [5][6][7], run seeing the same Hardware Abstraction Layer over a fine tuned Linux Fedora Core 6 with custom kernel.

Two different types of cameras have been used: Trinocular, Stereo and RearView use 9 PointGrey Flea2, featuring an XGA (1024x768), Bayer pattern, 1/3'' CCD sensor. The Lateral system, which has higher constraints from the measurement point of view, is based on 2 Allied Vision Technologies Pike2 cameras featuring a Full HD (1920x1080), Bayer pattern, 1'' CCD sensor.

Polarizing filters have been used on each optic to reduce reflections and artifacts.

All the cameras mounted externally were enclosed in a sealed box to prevent damage from rain and dust.

The cameras for the Trinocular system were mounted inside the cab, without any other protection than a thin cover to avoid windshield reflections.

| Vision System | TRINOCULAR | STEREO | LATERAL | REARVIEW |
|---|---|---|---|---|
| Cameras | 3x PtGrey Flea2 (XGA) | 4x PtGrey Flea2 (XGA) | 2x Allied Vision Technologies Pike 2 (HDTV) | 2x PtGrey Flea2 (XGA) |
| Cameras Position | Upper part of the windshield, inside the cab | 2 on the front camera-bar, two on the back of the truck, all looking downwards | On the sides of the front camera-bar | External, on top of the cab, looking backwards and downwards, rotated by 90° |
| Linked LIDAR | Front | Front, back | Not used | Back |
| Algorithms | Lane detection, stereo obstacle detection | Lane detection, stop line detection, curb detection, short-range stereo obstacle detection | Monocular obstacle detection | Monocular obstacle detection |
| Range | 7 to 40m | 0 to 10m | 10 to 130m | -4 to -50m |
| Notes | 3 stereo systems with different baselines | 2 stereo systems (front and rear) | Enabled when the truck stops at crossings | Overtaking vehicles detection |

**Table 1. Details about the vision systems on the TerraMax**

Trinocular cameras orientation was freely adjustable using 3-degrees-of-freedom screw lockable mounting.
The 3 IBEO Alasca XT LIDARs, like the one represented in Figure 3, provide a 240° scan on 4 separate planes with an overall vertical field of view of 3.2°. This laser is able to supply information up to four reflections against transparent objects. This is particularly useful in bad environmental conditions where the first reflection may be weak and caused, for example, by dust or rain, hiding the important reflection due to an obstacle. Two ECUs were processing separately the raw data from the front lasers and from the back laser.

**Figure 3. The IBEO A.S - ALASCA XT LIDAR.**

Additional information on the GPS location (latitude, longitude, heading), INS (speed, linear and angular accelerations) and Vehicle Status (motion direction, gear), are provided through the network by other computers directly connected to the physical sensors to the environment perception computers.
The sensors location and field of view is described in Figure 5.
Some vision systems share sensors with others: the front LIDAR are shared between 3 systems, the GPS, INS and Vehicle Status virtual sensors are seen from all the systems.

### 3.2 Synchronization

The synchronization between the two main sensor systems (Vision and LIDAR) is guaranteed by means of appropriate hardware and software.

One of the vision systems acts as Network Time Protocol (NTP) server maintaining the processing systems clocks tightly synchronized. As a matter of fact, different processing systems (like for example LIDAR ECU and Vision PCs) have clocks physically separated. For this reason it is necessary to guarantee that when a computer gathers or produces these new data, this data will have unique time reference. In other words we need a way to synchronize the clocks with enough precision that they can be considered a single one. Notice that for this purpose it is not mandatory to use a precise absolute reference, time such as the GPS time but it is sufficient to keep the clocks synchronized using NTP.

In addition, physical sensors, namely all the 11 cameras and the 3 LIDARs, have a unique sampling signal: a 12.5Hz square wave signal, generated by an IBEO SyncBox. Thus, in order to avoid that the high level of electromagnetic interferences due to the engine and other power systems onboard like switches and electro pneumatic valves can cause unwanted signal degradation, the strobe signal is amplified from 5 to 13V and transferred at this level close to of the sensors, here is translated to the level desired by the sensor and connected to the trigger signal of cameras and lasers.

### 3.3 GPS, INS and Vehicle Status

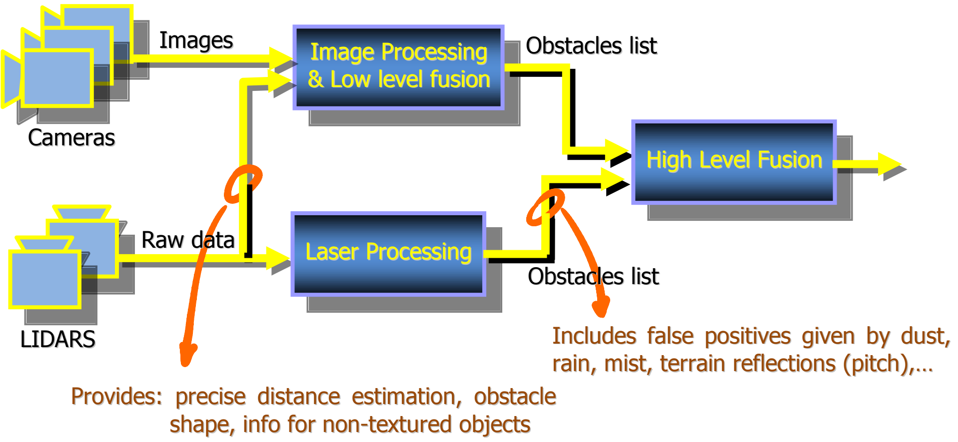
A Smiths Aerospace inertial reference unit was installed to provide the 6 Degree of Freedom measurements of the system inertial status integrated with the GPS receiver and wheel odometer data.

The accuracy of the positioning has been tested to be 10cm 95% of the time while the maximum error during the test period has been measured in 1.2m.

The vehicle status was available to the sensors, through the network interface. The vehicle status information like speed and gear were used by the vision systems to turn on and off the Lateral system when the truck was in a crossroad; speed was also used to decide the baseline to be used to detect obstacles using the front Trinocular system (see section 3.5).

### 3.4 Fusion implementation

In a schema without data fusion, images are captured from a calibrated stereo camera pair and processed inside each vision PCs depending on the application.
At the same time, data coming from the LIDARs pass through their own processing stage through their ECU. These two stages produce two different results, each one affected by the weaknesses typical of each sensor.



Includes false positives given by dust, rain, mist, terrain reflections (pitch),…

Provides: precise distance estimation, obstacle shape, info for non-textured objects

**Figure 4. Proposed data fusion scheme.**

The object list from a vision system is obtained from a dense 3D map with low accuracy measurements, thus objects may be hardly recognized or missing where the texture information is poor, and all the obstacle distances may be inaccurate, especially for far obstacles. On the other hand, LIDAR will supply a very accurate measurement on a sparse 3D representation of the world, concentrated on one or more planes. LIDAR classification is weak due to the sensor's physics which mainly rely on objects density and reflectivity

Introducing the two-level fusion in this scheme (Figure 4) means to couple together the two data paths in a convenient way such that the final result can take advantage of the strengths of both kinds of sensors. In this situation raw laser scanner data are taken in input by the vision processing step to help the vision system processing.

This list of obstacles produced by the vision block using the low level fusion will have LIDAR precision distance estimation.

On the other data path, the list produced by the LIDAR ECU will be affected by false positives due to dust, rain, etc. Another possible false positives source is the instantaneous pitch variation: whenever the vehicle changes its speed, the pitch variations due to the dampers action change the laser scanner orientation. If the beams hit the ground, following a sudden braking or a hole, fake objects are detected in the proximity of the truck, causing abrupt behaviors of the autonomous driving system. On the other hand if the beams are oriented toward the sky due to a strong acceleration, they can miss the detection of obstacles. Appropriate LIDAR point processing can avoid these unwanted results; however there are cases like non flat terrain or hard braking, where it is almost impossible to avoid this behavior.

The two obstacle lists are then taken as input by the high level fusion module, the World Perception Server (WPS). This module is responsible of performing the level tracking removing false positives due to the LIDAR list and producing a stable representation of the objects surrounding the vehicle.

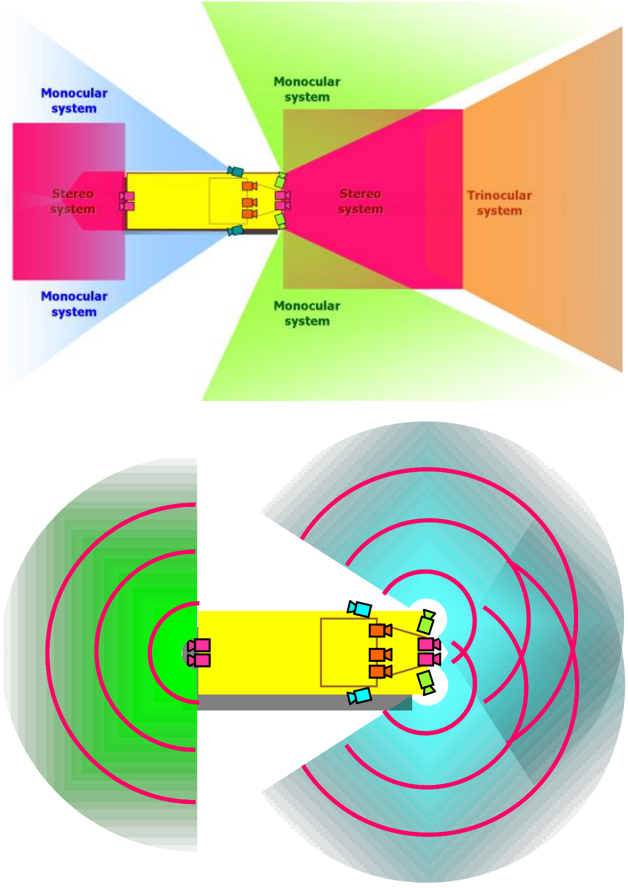


**Figure 5. Sensors position and field of view. From top: cameras and LIDARs.**

3.5 Perception systems

The TerraMax autonomous vehicle has been equipped with 4 vision systems as shown in Figure 5 on top. In this subsection each system is described more in detail, together with its specific implementation of the Low Level data fusion.

*Trinocular*

The Trinocular system cameras were employed to detect obstacles and lanes from 7 to 50m. In this area obstacles on the vehicle trajectory can be found, thus making this system probably the most critical for the sensing. It is formed by three stereo pairs with three different baselines, each specific baseline being more suitable for detecting obstacles in a certain range. The baseline size, detection ranges and correspondent speeds are detailed in the following table.

| Baseline | Size [m] | Range [m] |
|---|---|---|
| Short | 0.572 | 3 - 25 |
| Medium | 1.156 | 15 - 40 |
| Large | 1.728 | 30 - 50 |

This system has been successfully used for the first time during the 2005 DARPA Grand Challenge. Depending on the vehicle speed, a baseline is selected and the images from the stereo pair are processed to extract obstacles. The algorithm is based on the V-disparity approach [8].

The road slope is detected as first step, and then obstacles are found as everything that rises on the road surface over a certain threshold. The 3D reconstruction of the world is sent to the World Perception Server which collects and stows all the sensors data for the navigation suite.

This subsystem also captures the data coming from the laser scanner system. The detection rate in poorly texture areas takes advantage from the front LIDAR data. The main ground components are filtered and removed from the LIDAR output, then using the sensor calibration each LIDAR point is used to increase the density of the disparity image computed with the vision system. The result is used as input for the final flood fill step who merges pixels with similar disparity values. These data are used to perform a low-level fusion processing as described in section 3.4. The purpose of this phase is to obtain a more dense disparity map especially for the poor textured areas and in proximity of the beginning of the detection area, where obstacles are seen under different angles from the stereo pairs, producing a poor disparity map.

The tree cameras are mounted in the uppermost part of the windshield, inside the cabin. This high mounting position allows to see queued vehicles such as for example those parked on the road side or those approaching a crossroad. This feature simplifies the path-planner task, supplying a more accurate description of the world.

A lane detection algorithm is also running on this system. The use of the information from the stereo pair makes the result more reliable, reducing false positives that can result from the observation of poles or vehicle parts.

The lane detector processes a full resolution, color reconstructed image. The portions of images corresponding to the obstacles found in the previous step are removed from the search area. The grayscale and a yellow-enhanced image are processed with the same algorithm. The undistorted view-from-top is obtained using an inverse perspective mapping transformation and taking into account the pitch measured by the obstacle detector.

The next step the search for horizontal local luminance variations to detect lane markings is exploited. The algorithm is also able to classify different types of lane markings: white, single or dashed, and yellow, single or dashed.

A final tracking step is performed to get a more stable result, especially in cases where the line is not clearly visible.

In Figure 6 are depicted some typical results from the Trinocular system, showing different cases of lane markings and obstacles detection.

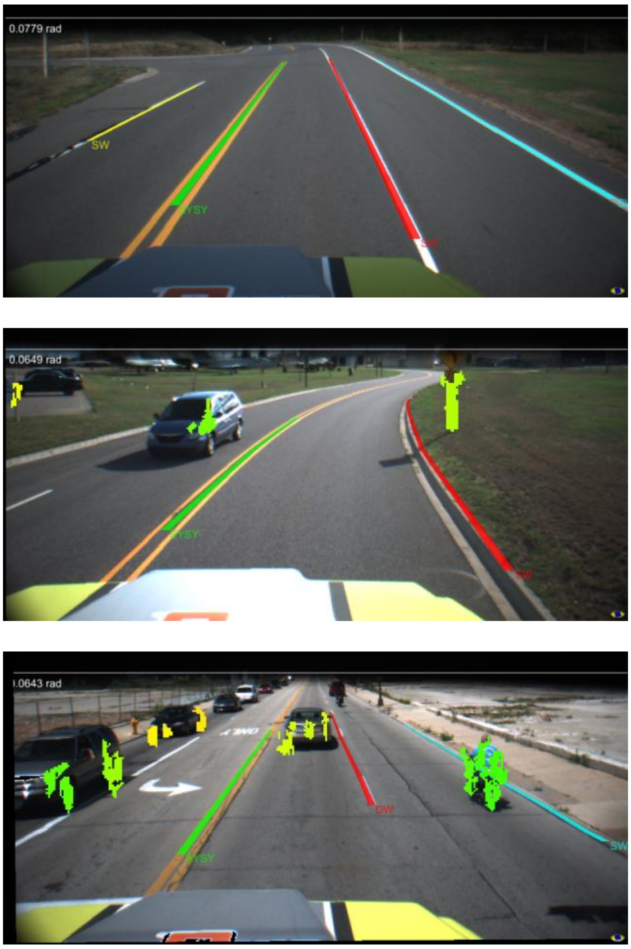


**Figure 6. Results from the Trinocular system: detection of obstacles and lane markings.**

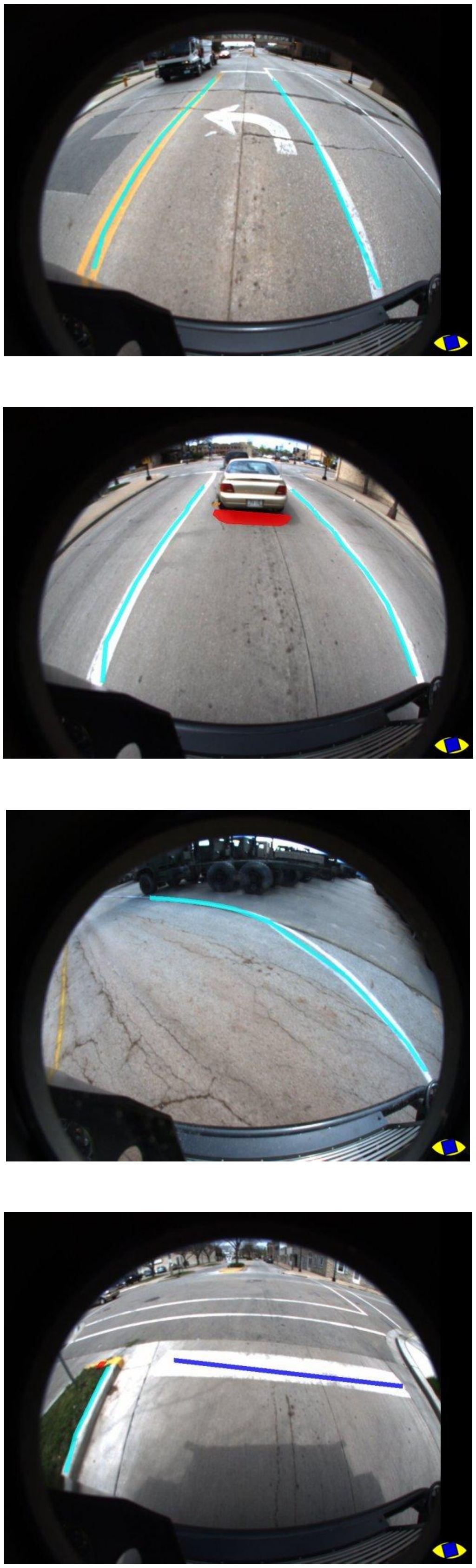


**Figure 7. Four results of the Stereo system: lanes, obstacle, sharp curves and stop lines.**

### *Stereo*

The Trinocular system has a coverage useful to drive at medium-to-high speed. Two Stereo systems for the narrow distance perception are provided in the front and in the back.

4 cameras are connected to the same PC which processes the images coming from the front or the back depending on the driving direction.

Thanks to the 160° aperture each system can monitor a 10x10m square detecting obstacles, stop lines and lane markings.

The fish-eye aperture allows detecting lane markings for the sharp curves.

This system uses the LIDAR information to refine the shape of the obstacles found. Distance is detected with good precision by the vision sensor while the shape might have some error due to the distortion of the IPM transformation. In fact the cameras setup selected for this system (fish-eye lenses) has the drawback of making it difficult to locate the exact boundaries of the detected items; to address this shortcoming, laser data are clustered and matched against the detected elements in the field of view, and when a correspondence is found the precise shape is associated to the obstacle. If no laser data are available, or the matching phase produces poor results, the system provides only the position of the obstacles, with no additional shape information.

The system uses low level data fusion with the laser scanner data to detect the precise shape of the object since the camera setup does not allow to obtain this data with reasonable precision.

Results of the Stereo system output showing the detection of obstacles, lanes for sharp curves, curbs and stop lines, are depicted in Figure 7.

### *Rear view*

The detection of incoming obstacles when driving on roads with multiple lanes is performed by the rear view system. Two cameras mounted on the top of the cabin and looking backward are used as input for this system. Cameras are rotated by 90° to take advantage and framing from the area close to the vehicle up to the horizon despite the mounting height. This system has the ability to overcome some LIDAR limitations, especially on dusty roads. This system uses the data fusion to refine the distances. Such correction is useful especially when the vehicle is pitching since the distance estimation is performed statically using a monocular camera. The detection is performed using an optical flow technique on the image after the perspective removal.

The high level processing is done searching and tracking the clusters of similar color. The road texture is removed by a comparison with the truck speed.

**Figure 8. Some results of the RearView system. When a car is in the lane at the back of the truck, it is shown with a red marker.**

Since the scene is framed from a high point of view, this system uses the LIDAR to correct the obstacle distance once the obstacles are found. Some output results are shown in Figure 8.

*Lateral*

When the truck stops at a crossroad, the Lateral system is turned on to monitor incoming traffic and see if the 10s gap required by the rules to merge into the traffic flow is available. The system performs this detection using 2 Full HD cameras, a standard background subtraction technique and a multi resolution approach to maintain a low processing resource consumption. No data fusion is used on this system since at a distance of 100m the LIDAR gives unreliable results due to the pitch and roll movements. Moreover, at that distance, the 4 planes and the horizontal resolution are not sufficient to give information on a single obstacle. The detection range goes up to 100m and the camera aperture is 90°: a wide area must be observed since crossroads might have different angles. Figure 9 shows some representative results of the Lateral system at different crossroads and in different situations.

**Figure 9. Result of the Lateral system: incoming vehicle are detected at crossroads.**

## 4. Conclusions

The TerraMax autonomous vehicle used active and passive sensors, taking advantage from the strengths of both approaches and overcoming their limitations through ad-hoc algorithms.

Some sensors have good performance but they do not cover every possible situation.
Vision based sensors work well in detection and classification but are weak in measurement.
LIDAR sensors work well in structure detection and measurement but they have problems with false objects due to ego motion and low reflective objects.
The novelty of this work lies in the fusion of data from the two sensors at two different levels. Results are promising since the truck performed very well during the test and the race. Low level fusion has been implemented and tested, and is now becoming a fundamental part of an innovative driving assistance system. High Level Fusion, although not thoroughly tested, proved to be effective during the Urban Challenge test bench.

Other vehicles adopted different solutions such as the replacement of Vision and LIDAR base sensors with High Definition LIDARs. This sensor seems to outperform Vision in dense 3D reconstruction of the surroundings. They have been the main data source for many top teams, definitely becoming a success factor in the Urban Challenge Final Event (UCFE). However, this kind of sensors do not seem a viable solution for integration on commercial vehicles given their mechanical characteristics (considerable weight, size, heavy rotating parts) and high production cost.

The expertise about sensor fusion acquired during the Urban Challenge has been capitalized in the development of an Advanced Driving Assistance System. The first project in which the know-how has been transferred is the Active Pedestrian Protection System (APPS), born from a cooperation between VisLab and Mando Corp.(Korea).

In this project a fusion between laser scanner and vision has been developed, with the aim of localizing potentially dangerous situations in specific urban scenarios. The first results obtained are described in [7].

The underlying idea of the project is the assessment of the scenario prior to the detection of pedestrians, in order to speed up the processing and, above all, focus the attention on specific dangerous situations.

The development is based on a specific urban scenario, in which a pedestrian is popping out behind a stopped vehicle, but the basic concept can be adapted to other dangerous situations.

**Figure 10. A dangerous pedestrian correctly localized from APPS system**

The laser scanner data are used to locate dangerous areas, particularly those behind stopped vehicles (or generic fixed obstacles); in these areas a pedestrian suddenly appearing is searched for. A typical situation of this kind of danger and the result of the processing is shown in Figure 10.

## 5. Acknowledgements

The authors would like to thank all Team Oshkosh for their precious support and for the endless effort put in the Urban Challenge project.
The authors would also like to thank Mando Corp. for founding the prosecution of this research.

## 6. References

[1] J.P. Hwang, S.E. Cho, K.J. Ryu, S. Park, E. Kim: "*Multi-Classifier Based LIDAR and Camera Fusion*", in Procs. IEEE Intl. Conf. on Intelligent Transportation Systems, Seattle WA, pp 467-472 Sept. 2007.

[2] C. Prembida, G. Monteiro, U. Nunes, P.Peixoto: "*A LIDAR and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking*", in Procs. IEEE Intl. Conf. on Intelligent Transportation Systems, Seattle WA, pp 1044-1049. Sept. 2007.

[3] M. Szarvas, U. Sakai and J. Ogata: "*Real-time Pedestrian Detection Using LIDAR and Convolutional Neural Networks*", in Procs. IEEE Intelligent Vehicle Symposium, Tokyo, Japan, pp. 213-218, June 2006.

[4] S.Milch, M. Behrns: "*Pedestrian Detection with radar and Computer Vision*", http://www.smart-microwave-sensors.de/Pedestrian_Detetcion.pdf

[5] Y.L. Chen, V. Sundareswaran, C. Anderson, A. Broggi, P. Grisleri*, P.P. Porta, P. Zani, J. Beck: "*TerraMax: team Oshkosh urban robot* ", Journal of Field Robotics, in press.

[6] A. Broggi, A. Cappalunga, C. Caraffi, S.Cattani, S. Ghidoni, P. Grisleri, P.P. Porta M. Posterli, P. Zani: "*TerraMax Vision at Urban Challenge 2007* ", IEEE Transactions on Intelligent Transportation Systems, in Press.

[7] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H.G. Jung, "*Localization and Analysis of Critical Areas in Urban Scenarios*", in Procs. IEEE Intelligent Vehicles Symposium 2008, pages 1074-1079, Eindhoven, Netherlands, June 2008.

[8] R. Labayrade and D. Aubert: "*A single framework for vehicle roll, pitch yaw estimation and obstacle detection by stereovision*", in Procs. IEEE Intelligent Vehicle Symposium, Columbus, USA, Location, Jun, 2003.