TerraMax Vision at the Urban Challenge 2007

Alberto Broggi, *Senior Member, IEEE*, Andrea Cappalunga, *Student Member, IEEE*, Claudio Caraffi, *Student Member, IEEE*, Stefano Cattani, *Student Member, IEEE*, Stefano Ghidoni, *Member, IEEE*, Paolo Grisleri, *Member, IEEE*, Pier Paolo Porta, *Student Member, IEEE*, Matteo Posterli, and Paolo Zani, *Student Member, IEEE*

***Abstract*—This paper presents the TerraMax vision systems used during the 2007 DARPA Urban Challenge. First, a description of the different vision systems is provided, focusing on their hardware configuration, calibration method, and tasks. Then, each component is described in detail, focusing on the algorithms and sensor fusion opportunities: obstacle detection, road marking detection, and vehicle detection. The conclusions summarize the lesson learned from the developing of the passive sensing suite and its successful fielding in the Urban Challenge.**

***Index Terms*—Autonomous vehicles, data fusion, lane detection, obstacle detection, Urban Challenge, vision systems.**

I. Introduction

THE CHALLENGES organized by DARPA are considered some of the most important milestones in the field of vehicular robotics. Not only did DARPA provide a vigorous help in the flourishing of new technological solutions, but the many researchers who accepted these challenges demonstrated that the current technology is indeed mature enough to drive vehicles in off-road and urbanlike environments.

The design and realization of autonomous vehicles capable of handling unknown and dynamic situations involve different aspects, spanning from new technologies to implementation and from concept to hardware issues, challenging many different research fields.

A completely autonomous vehicle needs to perceive the environment, assess the situation and react to it, and finally control the actuators to implement a correct behavior. In other words, the vehicle must have perception, decision, and control capabilities. The DARPA Grand Challenge (particularly the second one in 2005 when five vehicles were able to reach the end of the 130-mi course), the DARPA Urban Challenge, which witnessed the completion of the course by six vehicles, and other public demonstrations such as the PReVENT demonstration at the end of 2007 [1] showed that the control side has already reached an advanced development stage. Some advanced control systems (such as ABS, EBD, and ESP) have, in fact, already reached the market.

Manuscript received May 15, 2008; revised July 3, 2009. First published February 19, 2010; current version published March 3, 2010. The Associate Editor for this paper was F.-Y. Wang.

The authors are with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma, 43100 Parma, Italy (e-mail: broggi@vislab.it; kappa@vislab.it; caraffi@vislab.it; cattani@vislab.it; ghidoni@vislab.it; grisleri@vislab.it; portap@vislab.it; posterli@vislab.it; zani@vislab.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2010.2041231

Fig. 1. TerraMax vehicle.

Conversely, developing effective and reliable perception and decision systems is definitely more challenging. The perception problem is very complex and even not well defined: a perception system must deal with a high number of different situations and unknown scenarios. Objects in a scene have to be perceived and classified; a cluttered background—typical of an urban environment and, therefore, of a common situation—can lead to misdetections and introduce a high level of noise in the results. Similar considerations also apply to decision systems: to decide on the action to be taken, it is not sufficient to model the vehicle itself, like what happens on the control side, but it is mandatory to model the real world as well; therefore, a huge number of situations must be taken into account.

Different technologies have been tested and implemented on the robotic vehicles that participated in the DARPA Urban Challenge, ranging from conservative products up to completely innovative solutions. Based on the teams' capabilities, budget, and—most importantly—vision, the various vehicles have tackled the perception problem in different ways.

Team Oshkosh's vehicle, TerraMax (Fig. 1), was designed to operate with conservative technology so that the integration of the final sensing suite on new vehicles and/or its retrofit to existing vehicles could be straightforward. Furthermore, the choice of using passive technology (sensors not emitting any signal) is generally preferred in the military environment and is also dictated by the will of producing vehicles able to operate in close proximity, namely, in situations in which active sensors may interfere with each other.

The final solution was then to adopt cameras as the main sensing device and develop a 360° sensing suite based on artificial vision. Eleven cameras have been mounted on TerraMax to perceive all possible events around the vehicle that may

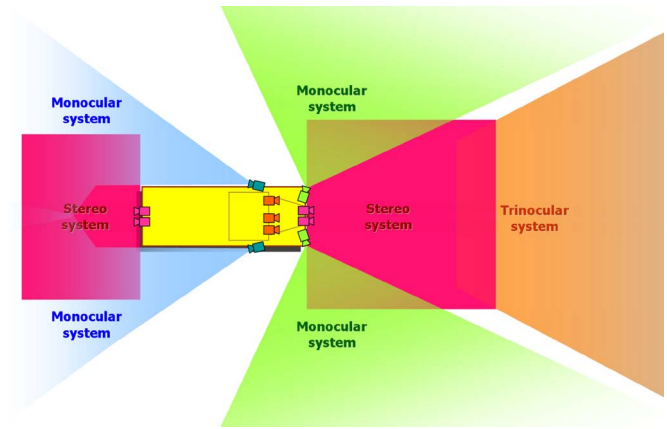

Fig. 2. TerraMax vision systems (TerraMax facing right) and their placement: orange—trinocular system; purple—stereo systems; green—lateral system; cyan—rearview system.

need to be considered by the decision system: the presence of obstacles, moving vehicles, lane markings, and horizontal road signs. To improve system robustness, three laser scanners have also been installed—two in the front and one on the back—to complement vision perception. In addition to providing their own information to the decision system, these three LIDARs supply raw data to the vision systems, which perform data fusion to fine tune their final results.

This paper is organized as follows. Section II presents the vision systems and provides details on the architectural and hardware solutions. Section III presents the obstacle detection functionality and shows how it is implemented using the different vision systems. Sections IV and V describe the solutions adopted for lane detection and vehicle detection, respectively. Section VI concludes the paper with a discussion on the lessons learned about the use of a vision system onboard an autonomous vehicle.

II. TerraMax Vision Systems

TerraMax perception is provided by 11 cameras grouped into four different systems, each supplying information about the environment in a specific area around the vehicle. Fig. 2 shows a sketch of the coverage of the vision systems.

A trinocular vision system (orange in Fig. 2) is used to perceive information about the road in front, including lane markings, road borders, and possible obstacles/vehicles on the driving path. Two stereovision systems (purple in Fig. 2) monitor the close proximity of the vehicle's front and back to locate obstacles and road signs. The precise localization of road signs (lane markings and stop lines) is essential due to the vehicle's abnormal size. Two monocular vision systems (green in Fig. 2) are used to surveil the sides of the vehicle when approaching intersections; normally off, these two high-resolution cameras start to process images as soon as the vehicle stops at an intersection and look for oncoming traffic in any possible direction (not only orthogonal to the vehicle direction). Finally, two other monocular vision systems (cyan in Fig. 2) keep track of overtaking traffic in adjacent lanes; their output is required when TerraMax needs to change lanes.

Fig. 3. Hardware arrangement inside the passenger seat for the various systems.

All the vision systems share a common hardware platform consisting of a rugged PC from SmallPC,[1] which is equipped with a mini-ITX board from Mini-Box.com,[2] an Intel[3] Core 2 Duo T2500 CPU, 2 GB of random access memory, and an Intel 945 GM chipset. Two FireWire 800 boards, one peripheral component interconnect (PCI), and one PCI Express provide four connectors on the back. All the PCs are placed under the passenger seat inside the cabin, together with LIDAR, INS/GPS, and vehicle control units, as shown in Fig. 3. The imaging hardware for the trinocular, stereo, and rear systems is based on Point Grey Research[4] Flea2 industrial cameras, which are equipped with a 1/3-in 1024 × 768 pixel Sony charge-coupled device sensor, while the lateral one features two full HD (1920 × 1080) FireWire B AVT[5] Pike2 industrial cameras; the exposure parameters like gain and shutter speed can be adjusted through the FireWire interface, while synchronization at 12.5 Hz with the three IBEO[6] AlascaXT LIDARs is provided by a dedicated SyncBox. Cameras are connected to the processing system using 7.5-m cables from AVT, which are specifically designed for industrial use in noisy environments. Cables, cameras, and boards are also equipped with locking screws to keep the connectors in a fixed position, even in critical environments with strong vibrations; all cameras placed outdoor are enclosed in sealed boxes to preserve them from dust and rain. Polarizers have been used to minimize the reflections.

The following sections present each single system in detail together with the solution adopted for their calibration, considering that calibration has to be performed on the fly,

[1]http://www.smallpc.com.
[2]http://www.mini-box.com.
[3]http://www.intel.com.
[4]http://www.ptgrey.com/.
[5]http://www.alliedvisiontec.com.
[6]http://www.ibeo-as.com.

Fig. 4. Trinocular portable calibration grid.

in many different places, and without any static calibration environment.

A. Trinocular

The trinocular system has been designed to provide robust obstacle and lane detection capabilities in the long range (7–50 m, with the lower bound determined by the presence of the hood in the cameras' field of view). Obstacle detection makes it possible to deal with certain kinds of situations that LIDARs can hardly (if at all) handle correctly, such as multiple vehicles queued at crossings, poles, and dust in unpaved areas (which can be perceived as a solid obstacle by these sensors, while stereovision tends to filter it out since it corresponds to a textureless region originating no reliable matches); it also provides additional information on the detected obstacles (such as height) and constitutes a reliable solution even in the case of complete LIDAR sensor failures. Finally, it improves the abilities of the lane detection system by masking image areas corresponding to objects. The lane detection module provides the path planner with a detailed description of the road geometry in front of the vehicle, along with a classification of each detected line (like its color and type).

Cameras are installed behind the upper part of the windshield, looking forward; Huber[7] mounts have been used to fix the cameras to the truck: these mounts allow adjusting the camera orientation with the necessary precision and also provide locking screws to avoid movements caused by vibrations. Myutron[8] 4.2-mm lenses have been chosen because they represent a good tradeoff between having a wide field of view and having a good resolution in the far range.

To provide accurate results, a sufficiently precise camera calibration is needed, and given the long baselines (up to 1.5 m) used in the trinocular system, this task turns out to be quite complex. A 5 × 25 m calibration grid, which is laid on a flat area, as shown in Fig. 4, is used to estimate cameras extrinsic parameters, which are then refined using the available laser data. Camera positioning has been eased by the use of software rectification, which can compensate for small misalignments; nonetheless, performing all the needed measurements and extracting the correct values from the collected data is a non-trivial time-consuming activity, which needs to be performed each time the cameras are serviced, and, as such, must be scheduled properly to avoid making the system unavailable for some time.

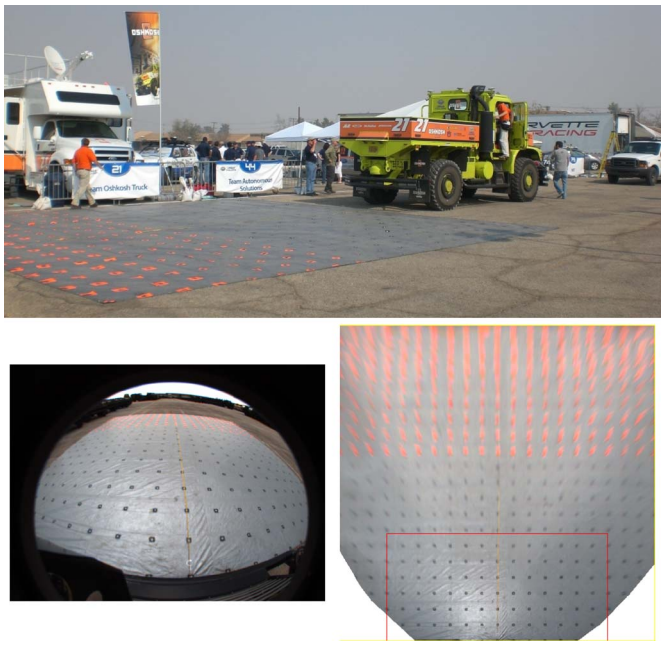

Fig. 5. Stereo calibration is performed by means of a grid painted on a portable tarp. The picture on the top shows TerraMax while calibrating the back stereo; shown on the bottom is the acquired image and the result after the pixel remapping using a lookup table.

B. Stereo

The trinocular system described in Section II-A can only be used for perception at medium to long distances; to extend its detection capabilities, TerraMax includes two stereo systems (one in the front and one in the back), which provide a precise sensing of obstacles and lane markings in the close proximity of the vehicle. The two stereo systems are identical from both the hardware and software standpoints: the four cameras are connected to the same PC, which selects the system to use based on driving direction and speed (forward, backward, or both when the speed is close to 0 km/h). Thanks to wide-angle (fisheye, about 160°) lenses from Fujinon,[9] the sensors gather information on an extended area of the immediate surroundings, which is about 10 × 10 m.

Calibration for the stereo system has to cope with highly distorting cameras used without any knowledge about the intrinsic and extrinsic camera parameters. An analytic approach to calibration would be computationally prohibitive: the equations that are normally used to model sferic lenses become too complex when wide-angle lenses are used. Therefore, an empirical strategy has been used: during an offline preprocessing, a lookup table that allows a fast pixel remapping is generated; namely, each pixel of the distorted image is associated to its corresponding pixel on the undistorted image. Images of a tarp with a grid of painted markers and placed in front of the truck

[7]http://www.hubermounts.com.
[8]http://www.myutron.com/index_e.asp.
[9]http://www.fujinon.com/.

are used to compute the lookup table (see Fig. 5). A manual system to pinpoint all the markers on the source image is used.

C. Lateral

Traffic merge is one of the most challenging situations the Urban Challenge vehicles were asked to deal with: after stopping at an intersection, vehicles are supposed to pull into traffic only when oncoming vehicles leave a gap of at least 10 s.

Assuming that the maximum speed of vehicles in an urban environment is 13 m/s (30 mi/h), it is possible to deduce the minimum detection range needed for traffic merge as $13\ \text{m/s} \cdot 10\ \text{s} = 130\ \text{m}$. Such a long distance is far beyond the typical LIDAR detection range; moreover, oncoming vehicles fall outside the field of view of the cameras used by the other vision systems, which cover the areas in front of and behind the vehicle. Hence, a specific vision system has been developed; the constraints to meet were the following:

1) long-distance detection capabilities;
2) different intersection angles;
3) detection of moving oncoming objects with speed estimation.

These constraints led to the choice of high-resolution and wide-field-of-view cameras to cover the several shapes an intersection can have. Cameras are mounted on a bar just in front of the vehicle at a 165-cm height from the ground, pointing 70° away from the truck longitudinal axis, and the field of view provided by the 8-mm optics from Kowa[10] is 85°.

To convert the position of detected vehicles from pixel to world coordinates, camera calibration is needed because of the mono approach to the problem. The constraint on long-distance detection makes it impossible to use a grid for calibrating the cameras, so the calibration process has been done by centering the TerraMax on an intersection and putting several cones along lateral roads at distances from 0 to 100 m. After acquiring some images of the intersection, the cones on the images have been marked to compute a homographic transformation that allows determining the correspondence between each point of the world with its representation on the image.

D. Rear

The rearview system has been designed to detect vehicles that are approaching the truck from the rear side: the system aim is to provide the data needed to determine if it is safe to change lanes.

The cameras are placed on the top of the cabin, at each side, pointing backward. The two cameras are rotated by 90°, acquiring images in portrait mode: this solution has been adopted because the field of view needs to extend quite far behind the truck, while, laterally, a narrow angle is sufficient to cover the adjacent lanes. Although the cameras are synchronized with the laser scanners, this is not an absolute constraint, since small errors can be tolerated in this phase. The optics used for this system have 4.2-mm focal length.

[10]http://cctv.kowa.com/frontend/landing_opto.asp.

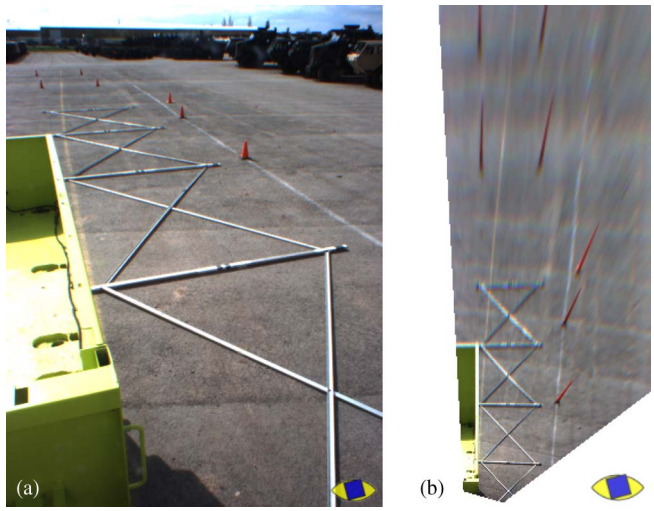

Fig. 6. (a) Rear calibration grid (the same used for the trinocular system). (b) Same image after homographic transform action.

The fields of view of the two cameras do not overlap, except for a small portion toward the horizon: the rearview system can therefore be considered as composed by two identical monocular subsystems performing the same task, which can be run at the same time on a dual-core machine. Calibration, which is used to evaluate the distance of the objects from the truck, is performed by acquiring images of a grid, as shown in Fig. 6.

III. Obstacle Detection

During navigation, TerraMax needs to sense the presence of obstacles in its surroundings soon enough to be able to adjust its trajectory to avoid them, whether they are moving or not. Long-range detection capabilities must be coupled with the ability of performing accurate maneuvers, which becomes essential inside crowded parking lots or traffic jams, with a lot of vehicles moving slowly along the lanes. Moreover, the lane detection systems can greatly benefit from a detailed knowledge of the presence and position of obstacles on and along the road.

Such requirements led to the development of the trinocular and stereo obstacle detection systems, which, once coupled together, can provide the needed environment reconstruction.

A. Trinocular

The cameras of the trinocular system are mounted on top of the cabin, with an asymmetrical configuration that generates three possible baselines (i.e., the distances between a pair of cameras). This solution, which was successfully tested during the DARPA Grand Challenge in 2005 [2], provides a substantial degree of flexibility, satisfying the different perception needs that arise during autonomous navigation.

The processing begins with the conversion of the raw Bayer patterned frames acquired by the cameras to subsampled gray-scale ones; these images are then rectified for stereo processing, correcting small hardware misalignment of the cameras, so that the corresponding epipolar lines become horizontal, and allowing more precise measurements. A special derivative filter [3]—useful to highlight ground features—is then applied to

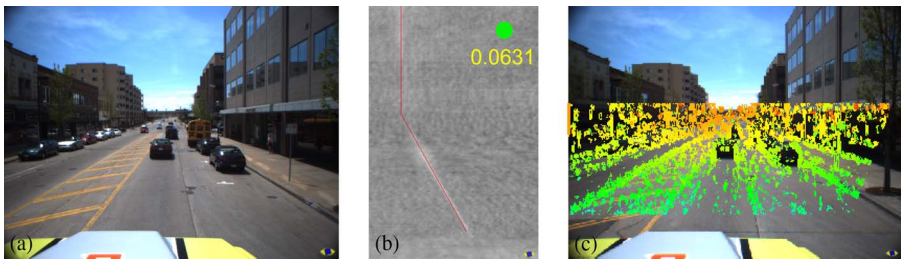

Fig. 7. First elaboration steps. (a) Sample frame acquired by the center camera. (b) V-disparity image with pitch estimation. (c) DSI.

the images coming from the largest baseline, and the resulting output is used to compute the so-called v-disparity image [4]. The v-disparity image is then analyzed to estimate the current vehicle pitch, as described in [2]. Fig. 7(b) shows a sample v-disparity image, with the detected ground correlation line highlighted in red, while the computed pitch value (in radiants) appears on the top-right corner.

The following step is the selection of one baseline out of the three available to be used for the generation of a disparity space image (DSI): shorter baselines provide less noisy results, which are useful in low-speed maneuvers, while longer ones allow a deeper perception range, thus permitting the detection of obstacles while they are still far away from the vehicle.

DSI computation is carried out using a standard correlation-based approach, with the sum of absolute differences as the associated cost function, which allows very efficient implementations; the matching phase is performed in an incremental fashion, thus becoming independent of the window size [5]. This algorithm has been improved by searching for matches along the epipolar line only for disparity values greater than the one expected for the ground (which varies along the image rows): this approach enhances the quality of the resulting map, with less spurious matches corresponding to disparities below the ground, but increases the overall complexity, since it generates a number of border cases that need to be carefully addressed. The resulting disparity map is shown in Fig. 7(c).

Both the v-disparity image and the DSI are computed by exploiting the parallelization opportunities offered by the target hardware: at a coarse level, input images are split into stripes and processed in parallel by independent threads on different CPU cores, while at a more fine level, the window matching algorithm exploits the available single-instruction–multiple-data capabilities through the use of the MMX and SSE instruction sets, operating on multiple pixels and rows (v-disparity image) or multiple disparity values (DSI) at the same time.

Once the disparity map has been computed, along with the corresponding 3-D world points, obstacle detection can take place. First, most of the elements belonging to the ground are removed, as contiguous pixels laying on the terrain surface have decreasing disparity values when analyzed from the bottom to the top of the image; the remaining pixels are grouped columnwise into layers with similar disparities. One of the limitations of the original obstacle detection algorithm [6] (originally conceived for the DARPA Grand Challenge in 2005 and used as a starting point for the development) was its inability to cope with multiple obstacles appearing in the same image column: in this case, the only detected obstacles was the

Fig. 8. Detection improvements using multiple layers. (a) Obstacles identified using a single layer. (b) Same frame processed using three layers, with some of the differences circled in red.

largest. The use of multiple layers during the grouping process removes this limitation, allowing it to cope with situations like the one depicted in Fig. 8, where a number of queued cars become stacked together by the perspective projection.

The compactness of each group is evaluated to discard elements that are too spread vertically to represent an obstacle; the surviving candidates receive a score proportional to the number of pixels in the group, and the score is increased if it is adjacent to groups with similar disparity. This value is then multiplied by the group average distance from the camera, so that the scores of objects placed in different positions become comparable. At the end of this process, a threshold on the score is applied to select only the elements that represent actual obstacles for the truck.

To improve the system detection rate in poorly textured areas, a data fusion approach has been devised, exploiting information coming from both the cameras and the front LIDARs. The first step consists in filtering LIDAR data to remove the main

ground components; next, for each LIDAR world point, the associated disparity value and image coordinates are computed using sensor calibration. The result is overlaid onto the previously computed DSI, generating a richer starting set for a final flood-fill stage, where these seed values are expanded, merging close pixels with similar disparity values. It should be noted that despite being originated by completely different sensors, seeds can be fused together since, after the respective filtering processes, they provide comparable information: a few points corresponding with a high confidence degree to obstacles. The flood-fill step is performed onto the original (not filtered) DSI using the computed seeds, and the resulting clusters are filtered to remove spurious ones (those too small to represent any obstacle). A final high-level tracker assigns to each detected element ID, age, and speed information.

B. Stereo

As previously described, the trinocular system can only monitor obstacles farther than 7 m, due to the presence of the hood. The vehicle proximity is therefore covered by the stereo system, which was derived from the one described in [7] and improved with LIDAR scan data fusion.

The use of an inverse perspective mapping (IPM)-based technique is justified by the assumption of an almost-flat terrain in the immediate proximity of the vehicle. Obstacle detection is performed in two steps: first, the two images, which were acquired simultaneously, are preprocessed to remove the very high lens distortion and the perspective effect, obtaining an IPM [8] image [depicted in Fig. 9(a) and (b)]; then, the two images are matched, computing the difference, so that any appreciable deviation from the flat-road assumption—which can be considered valid since the area of interest is limited to a maximum of 10 m—is labeled as an obstacle. Thanks to this approach, the system has the following characteristics.

1) It is able to detect any kind of obstacle without *a priori* defined obstacle classes.
2) It is not based on motion, which usually generates false positives when the vehicle moves.
3) It is robust with respect to shadows on the ground since it performs a comparison between images taken at the very same time and since shadows belong to the road texture.

After the difference between two IPM images has been computed, a Sobel filter is applied to avoid differences resulting from images with different brightness. After this step, data coming from LIDARs are clusterized and then overlaid onto the difference image, so that laser reflections in a particular region can boost the difference image and thus help the detection of obstacles. For each pixel above a given threshold, a morphological growing is applied, and then, connected areas appearing in the resulting image are localized and labeled: a progressive number is assigned to each label for further identification. A polar histogram is computed for each region; the reference point used to compute the polar histogram is the projection of the midpoint between the two cameras onto the road plane. These regions produce strong peaks on the polar histogram in correspondence to obstacles; thus, an additional filter can remove regions that do not correspond to obstacles. This filtering is performed considering the width of the histogram for the region of interest, which is computed in correspondence to a given threshold. When a polar histogram features several peaks, different values of width (w_1, w_2, etc.) are generated. Previously labeled regions are maintained only if the corresponding width is greater than a given threshold.

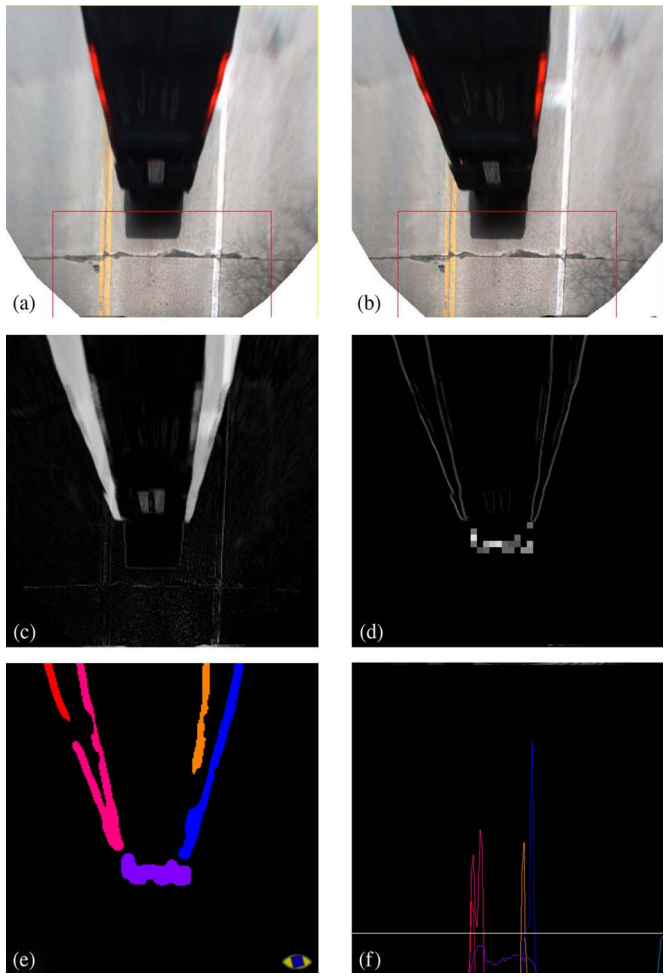

Fig. 9. Stereo obstacle detection processing steps. Starting from IPM images (a) and (b), a difference image is computed (c); then, Sobel filter is applied to avoid luminosity differences between the two images, and laser clustered data are overlaid (d). A morphological growing on each pixel above a given threshold is shown in (e), while (f) represents the polar histogram. The final output is shown in Fig. 10(c).

Data fusion with LIDAR can help provide a correct estimation of obstacle occupancy.

The processing steps are shown in Fig. 9, while some examples results are shown in Fig. 10.

IV. Lane Detection

Lane boundary knowledge has been essential to allow the autonomous vehicles to safely navigate the roads of the Urban Challenge. Knowledge about the road structure permits the following:

1) following roads whose *a priori* knowledge is incomplete (sparse waypoint following);
2) correctly choosing the lane to navigate in, without interfering with traffic in the opposite direction;

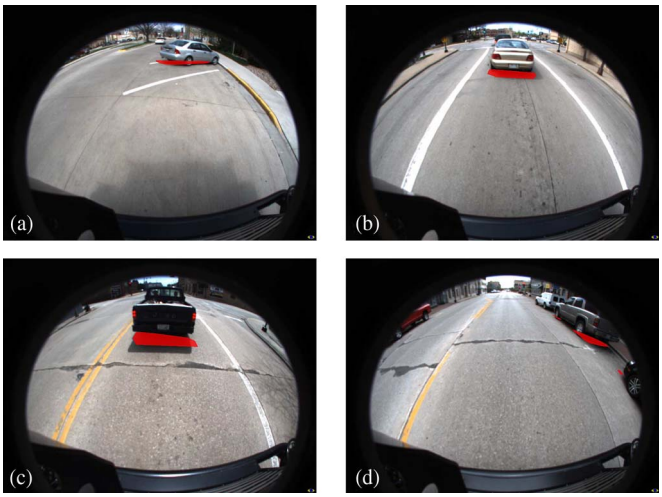

Fig. 10. Examples of stereo output. Obstacles are detected in the whole monitored area, and their shape is extracted by means of laser scan data.

3) switching to other lanes to avoid obstacles and during overtaking maneuvers.

The rules of the Urban Challenge [9] are described as follows.

1) There are three possible kinds of lane markings (lines):
   a) double solid yellow lines (coded as *SYSY* in the results), which are used to mark the edge between adjacent lanes with opposite traffic direction;
   b) (single) solid white lines (*SW*), which are used to mark road-side boundaries; this line marking was absent in most of the course, substituted by the presence of a curb or a barrier.
   c) dashed white lines (*DW*), which are used to divide adjacent lanes with same direction of traffic.
2) There are three possible structures for a road segment:
   a) one-way road with two lanes: SW at the sides and DW in the middle;
   b) two-way road with two lanes: SW at the sides and SYSY in the middle;
   c) two-way road with four lanes, two for each traffic direction: SW at the sides, SYSY in the middle, and DW to divide adjacent lanes with same direction of traffic.

A fourth line type, namely, single solid yellow (SY), to be used in replacement of and with the same meaning as double yellow lines, was also utilized to ease the construction of the site visit tracks of the teams involved [10].

A. Trinocular

In designing TerraMax, it was decided to obtain information about the road structure through vision-based lane detection. At the first stage, the task of lane detection was assigned only to the trinocular system, also exploiting color, given the fact that it carries important information. To obtain an extended perception range, during lane detection, the images were processed at full resolution (1024 × 768), obtaining color images through conversion of the Bayer input images. This conversion causes an additional computational cost, which cannot be delegated to the cameras because the system proved to suffer from the increased bandwidth request. Anyway, this problem is mitigated by the fact that lane detection is performed using just one of the three input images.

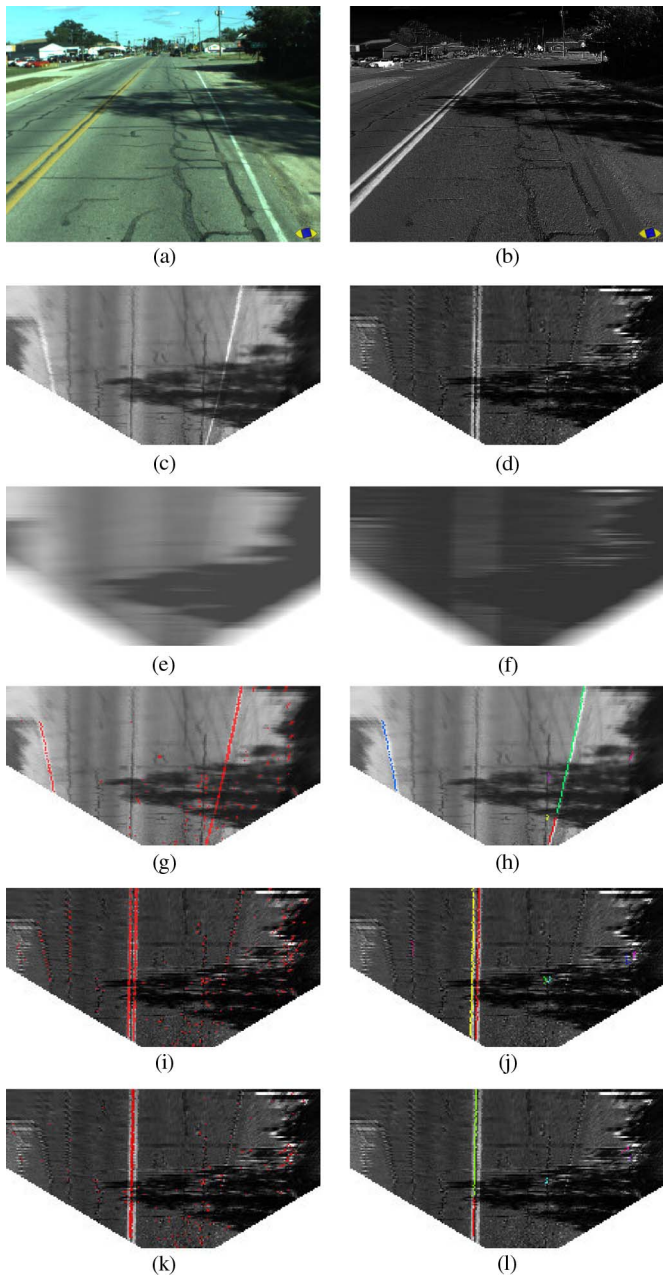

Fig. 11. Low-level processing of the lane detector for the trinocular system: a special case with diverging lane markings. (a) Right input image. (b) Yellow image. (c) IPM gray. (d) IPM yellow. (e) IPM gray averaged. (f) IPM yellow averaged. (g) DLD IPM gray. (h) Label DLD IPM gray. (i) DLD IPM yellow. (j) Label DLD IPM yellow. (k) DLDLD IPM yellow. (l) Label DLDLD IPM yellow.

From the input color image [Fig. 11(a)], two images are initially extracted.

1) A brightness image (*gray image*). This image is used for white line detection.
2) An image encoding higher values for pixel whose hue is closer to yellow [*yellow image*, Fig. 11(b)], for yellow

lines detection. Taking the three color components red, green, and blue, in the yellow image, the pixels values are set as

$$\text{pixValue} = \min\left(K\frac{\min(R,G)}{\max(B,B_{\text{th}})}, 255\right) \quad (1)$$

with K being a constant and B_{th} being a threshold used to avoid random values caused by noise in shadow regions.

The following low-level processing is then computed in parallel on both images (Fig. 11).

The algorithm looks for patterns eligible to be lane marking sections, i.e., for dark–light–dark [11] (DLD) transitions. To have constant DLD pattern widths, allowing simplifications and computation load benefit, an IPM is applied to the image [Fig. 11(c) and (d)]. To accomplish this task, the instantaneous pitch value obtained previously (Section III-A) is used; it should be noted that, to obtain the pitch value, no parallel lane marking assumption is made, so that cases of diverging lines, as in the example in Fig. 11, can be handled. The IPM in use maps a region that is 25 m wide, with a maximum distance of up to 40 m in a 333 × 120 pixel image; thus, pixels do not represent square regions of the real world. This is made to obtain a tradeoff between DLD pattern intelligibility and computational costs.

DLD transitions for each image row are then extracted[11] [Fig. 11(g), (i), and (k)]; to binarize DLD values, the algorithm uses a local threshold, which is proportional to the local brightness. This allows also detecting DLD patterns in shadow regions. The local brightness is estimated as the average value of pixels in a horizontal neighborhood. Note that computing the average of an IPM image corresponds to the average brightness over real-world regions of constant area. To ease the detection of double lines that are particularly close to each other, a dark–light–dark–light–dark (DLDLD) pattern extraction is also performed [Fig. 11(k)].

To delete false DLD patterns generated by objects, which, indeed, are not lying on the ground, a mask, which was created using the output from stereo obstacle detection (Section III-A), is applied at this stage. The resulting data are then clustered in labels [Fig. 11(h), (j), and (l)], discarding too short elements.

The length, position, and orientation of labels are examined to obtain the output road model, joining compatible segments to get candidate lines. Since pixels of the IPM image do not represent square real-world regions, at this stage, label points are converted into real-world coordinates, so that the lengths, curvatures, and orientations of labels become directly usable. The road model is then built in three steps:

1) identification of a dominant line, named *mother line* [Fig. 12(a), in purple], chosen considering:
   a) length;
   b) reliability (e.g., double lines are considered more reliable);
   c) persistence over time;

[11] On the analogy of Sobel naming conventions, these features are considered *vertical*.

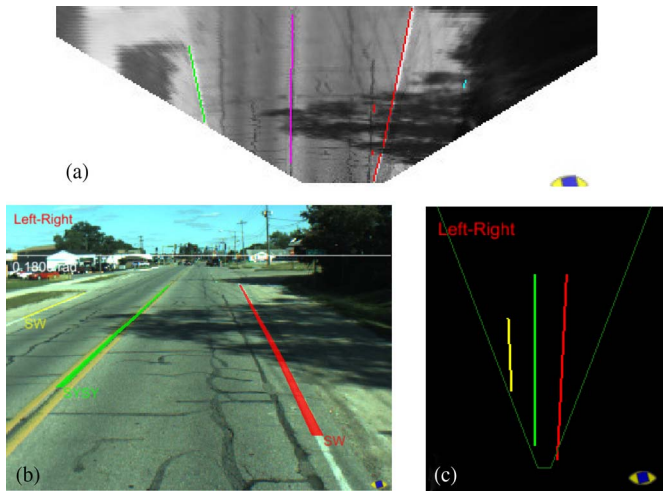

Fig. 12. (a) Categorization of labels based on compatibility with the mother line (plotted in purple): compatible labels at the left (green), right (red), far left (yellow), or far right (cyan) of the mother line. (b) and (c) Final results plotted on the input image and bird's eye view map. Green: close line at the left. Red: close line at the right. Yellow: far line at the left. SYSY: double solid yellow. SW: (single) solid white.

Fig. 13. Input images at the entrance of a sharp curve. (a) Trinocular input. (b) Front stereo input.

2) classification of the remaining segments as possible pieces of other (up to four) distinct lane boundaries, depending on the distance from the mother line;
3) creation of the lines at the side of the mother line and construction of the output model [Fig. 12(b) and (c)], categorizing left (in green) and right (in red) edges of the lane that the vehicle is currently navigating and boundaries of adjacent lanes.

Finally, the introduction of line tracking leads to increased stability and reliability of the results, particularly in the case of dashed lines.

B. Stereo

When Site Visit rules were disclosed, the introduction of 90° short-radius curves (*sharp curves*) obliged also introducing lane detection from the front stereo system: as can be seen in Fig. 13, only the stereo system allows effective lane detection while navigating sharp curves.

The algorithm used is almost identical to the one already described for the trinocular system; the strongest difference is that both vertical and horizontal DLD patterns are computed [Fig. 14(b)]. In the labeling process [Fig. 14(c)], two adjacent

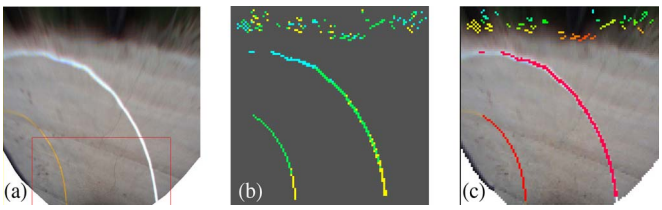

Fig. 14. Low-level processing of the lane detector for the stereo system [example in Fig. 13(b)]. (a) IPM image. (b) DLD image: yellow for vertical features, cyan for horizontal features, and green for both features simultaneously. (c) Obtained labels.

pixels share the same label when either of the following is true.

1) The two pixels are of the same kind (both vertical and horizontal DLD).
2) One of the two pixels corresponds to both a vertical and a horizontal DLD pattern.

Lane detection provided by the stereo system allows TerraMax to perform extremely precise maneuvering. In addition, using an analogous algorithm, the front stereo system performs detection of stop lines, making an accurate and safe approach to an intersection possible. It should be noticed that, given the limited perception range, the stereo system does not permit building a road model with the same reliability as the trinocular system, particularly in the case of sparse dashed lines, when just one short line segment is present in the considered field of view.

V. Vehicle Detection

A. Rear

Overtaking vehicle detection is a task that has widely been approached: see [12] for a recent study. In the rear system, vehicle detection is performed by evaluating the motion of objects in the image. In principle, every object moving on the road with a speed above a threshold and in a certain direction can be classified as an overtaking vehicle. The system monitors vehicles up to 40 m away from the truck and is capable of selecting only those that are approaching, discarding vehicles that are moving in the opposite direction.

The detection algorithm works on IPM images, which are obtained by applying a homographic transform to the original image; the transformed image is then divided into blobs of uniform color, which are used during the tracking phase; a tracker then computes associations between current and past clusters. These associations carry information about object motion of the object in that part of the image, thus providing optical flow [13]–[17] estimation. For each cluster, some parameters are evaluated, like color, aspect ratio, size, position, number of pixels, and center of mass. These data are used by the tracker to determine a similarity measure between clusters, as well as to filter out those that cannot belong to a vehicle: this dramatically reduces the number of elements to track and, therefore, the computational load.

The detection of shadows cast by vehicles on the road is used to validate the previous step results, as described in [18] and [19]. Shadows are easily detectable through the analysis of gray-level gradients and represent the most persistent clusters found by the algorithm: the introduction of shadow detection sensibly increased the algorithm performance.

Fig. 15. Example of vehicle detection. A red mark denotes a detected vehicle.

Obstacles found using the optical flow are then compared with those detected by the LIDAR: since the latter has a much higher precision, the position of obstacles found by vision is refined using the LIDAR data, if available. The fusion algorithm, in fact, does not create nor discard obstacles detected by the vision algorithm: otherwise, false positives or negatives caused by the laser scanner would impact on the vision system performance.

In Fig. 15, an example of the algorithm output is shown: a red mark indicates each detected vehicle. The computational load of the rear system is quite low, and the overall processing requires only 60 ms for each frame pair on the hardware used for the vision system.

B. Lateral

To deal with the large input image size, without overrunning the required time specifications, a hybrid multiresolution processing method has been designed. Two separate processings are applied on both captured images (right and left): the first one analyzes a horizontal slice of the full-resolution image, while the second one operates on a downsampled version of the whole image. Each process is executed independently, and the results are collected and fused together: this way, the computational load is considerably reduced if compared with a single full image processing. This approach provides high resolution on far-away zones, leaving enough time to also check near areas, and because of the independence of the four processes, it is possible to parallelize all of them, thus exploiting the multicore CPU.

The same detection algorithm is run for all the four processes and consists of a background subtraction approach with several additional features, which allow overcoming the problems due to camera vibrations and oscillating objects.

First, a difference image is created from the current frame and a reference image, which was computed as the time-weighted average of the last N frames; this way, vibrations that may affect the result are smoothed. The next step is to apply a threshold to the difference image to identify a set of zones in which movement was localized. All these zones are

used to generate a set of bounding boxes, through a simple histogram-based approach: a vertical histogram of the thresholded difference image is used to create several image slices, corresponding to high histogram values. The upper and lower bounds of the moving object are then searched in each vertical slice by computing a horizontal histogram and keeping the zones between the first and the last rows in which the histogram is high. It is important to notice that the adoption of this simple approach is mandated by the strict time constraints; this is why different approaches like labeling were discarded.

Once the list of moving objects is obtained, each bounding box is used to dynamically regenerate the scene background, an artifact that can be substituted to the reference image, to improve the difference image. The background generation algorithm starts from a black image, and at each iteration (one per acquired frame), the black zones of the background image are updated in the following way: if their corresponding pixels are not hidden by moving objects, they are filled with the pixels of the input image. This iterative process is incremental, but even if it generally takes a limited number of frames (depending on the speed of the vehicles) to complete, it is possible to use the partial generated background by filling the missing parts with the corresponding areas in the reference image. After the detection phase, the system tracks the moving objects and first infers information from the bounding boxes movements over time and then applies filtering policies to them.

The developed tracking system operates in two steps: the motion analysis phase, whose goal is to search for a mapping between the list of previously detected (and tracked) vehicles and the list of the new bounding boxes, and a filtering phase, whose goal is to discard all objects with a motion reckoned incompatible to a vehicle's one.

The last step of the algorithm is to convert the detected vehicle coordinates from the image to the world: this is done by applying an IPM to the points [20], so the output of the algorithm can be the real position, which is given in meters relative to the TerraMax, of each detected vehicle. However, this step is very critical, due to the high noise sensitivity of the IPM, particularly at long distances.

The algorithm showed a good detection rate during test sessions, particularly with respect to far vehicles; however, some false positives/negatives are present, even if the introduction of the tracking system has drastically reduced their number. As far as time is concerned, the overall system takes less than 100 ms on an Intel Core Duo 2.0-GHz (T2500) processor to process both right and left images in multiresolution mode. Some results are shown in Fig. 16.

Fig. 16. Detection results. The left column shows the complete result images, and the right column shows the same images zoomed in to highlight far-away detected objects.

VI. Conclusion

All the systems have been designed to provide results at a rate of 12.5 Hz, which has proven to be a good tradeoff between the processing time requirements of the various sensors and the planning and actuation constraints imposed by the urban scenario; the clock signal driving all the sensors was produced by the hardware SyncBox described in Section II.

Regarding the trinocular system, the Site Visit preparation helped to identify and solve a problem where highly textured road regions like lane marking borders were incorrectly detected as obstacles. After the Site Visit, LIDAR data integration was improved, and tracking was added to provide more information (namely, a unique ID and speed information for each obstacle) to the path planning system. During the National Qualification Event (NQE), the trinocular obstacle detection correctly identified road blocks such as the horizontal bar with stop signs and was also able to detect multiple vehicles queued before the truck at crossings. Data coming from this system integrated those coming from other sensors and provided a reliable alternative to the LIDAR. Tests also showed that the trinocular system, unlike LIDARs, could detect hanging and small obstacles.

The stereo obstacle detection proved to be robust in spite of the introduction of fisheye lenses, which are needed to cover a wide area in front of and behind the truck. Data fusion with LIDARs was added to the initially vision-only obstacle detector to enhance the algorithm and give to the vision system a higher precision in distance and shape estimation. Some problems were encountered in the case of little miscalibration; thus, an ad hoc filter for removing such kind of noise has been developed; similarly, another filter has been developed to remove some fake obstacles generated by LIDAR scan data, improving the overall performance. After the Site Visit, an ID was given to obstacles to ease the world perception server obstacle association process.

During the NQE and the Final Event (UCFE), the lane detection system allowed the TerraMax vehicle to stay correctly in the lane, also navigating sparse waypoint areas. It should be noted that TerraMax was the only vehicle to effectively use vision for lane boundary detection, while most of the teams used sensors (namely LIDAR) to perceive physical boundaries and map databases for GPS navigation. This has been possible because of the presence of curbs and barriers that physically defined the road boundaries. Furthermore, at the time of the final event, the course area was already well known by participants, easing navigation from maps and GPS. These facts belittled the efforts made by team TerraMax to develop a passive and more general sensor for road boundary detection. For close perception, a lane detection subsystem on the stereo module

was derived from the one implemented for the trinocular system with the difference of sharp-curve detection, which is a specifically developed module to localize very sharp curves (90° curves) that had to be driven during the Site Visit, and stop line detection was developed for the Site Visit and worked very reliably.

The lateral system has proven capable of accurately detecting obstacles up to 130 m and estimate their speed, but despite its performance, the lack of testing opportunities delayed its integration. The very high complexity of the overall sensing suite, as well as the interactions among the various systems, introduced many delays, with the team taking more time than expected to become familiar with the truck behavior. Given the tight schedule, those systems that were considered more critical and necessary to pass the qualifications, like lane detection and frontal obstacle detection, got higher priority, and this meant reaching the UCFE with a lateral system that is still not completely tested. At that point, where every single failure in any qualification test could end team TerraMax's effort, the system was deemed too risky to be kept running on the truck.

The approach chosen for the rearview system provided good performance since the early development stages, which was further improved by the introduction of shadow detection. The main advantages over LIDAR-based vehicle detection were the ability to handle queued cars and better off-road behavior: in such scenarios, dust clouds behind the truck were often incorrectly detected as obstacles by the LIDAR, while the rearview system did not generate any false positives. Unfortunately, like what happened for the lateral system, the rearview test and integration was delayed to the point that it became unsafe to use it during the NQE and UCFE.

ACKNOWLEDGMENT

The authors would like to thank all members of Team Oshkosh for their precious support and endless effort put on the Urban Challenge project.

REFERENCES

[1] M. Schulze, "Contribution of PReVENT to the safe cars of the future," in *Proc. 13th ITS World Congr.*, London, U.K., Oct. 2006.

[2] A. Broggi, C. Caraffi, P. P. Porta, and P. Zani, "The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA grand challenge on TerraMax," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Toronto, ON, Canada, Sep. 2006, pp. 745–752.

[3] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Procs. Int. IEEE Workshop Mach. Vis. Intell. Vehicles*, San Diego, CA, Jun. 2005, p. 65.

[4] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereo vision on non flat road geometry through 'V-disparity' representation," in *Proc. IEEE Intell. Vehicles Symp.*, Paris, France, Jun. 2002, pp. 646–651.

[5] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy, "Real-time correlation-based stereo: Algorithm, implementations and applications," INRIA, Sophia Antipolis, France, Tech. Rep. 2013, Aug. 1993.

[6] C. Caraffi, S. Cattani, and P. Grisleri, "Off-road path and obstacle detection using decision networks and stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 607–618, Dec. 2007.

[7] A. Broggi, P. Medici, and P. P. Porta, "StereoBox: A robust and efficient solution for automotive short range obstacle detection," *EURASIP J. Embed. Syst.—Special Issue on Embedded Systems for Intelligent Vehicles*, vol. 2007, no. 1, p. 6, Jun. 2007.

[8] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.

[9] DARPA Urban Challenge Rules, 2007. [Online]. Available: http://www.darpa.mil/grandchallenge/rules.asp

[10] DARPA Urban Challenge Technical FAQ, Jun. 2007. [Online]. Available: http://www.darpa.mil/grandchallenge/faq.asp

[11] S. Nedevschi, F. Oniga, R. Danescu, T. Graf, and R. Schmidt, "Increased accuracy stereo approach for 3D lane detection," in *Proc. IEEE Intell. Vehicles Symp.*, Tokyo, Japan, Jun. 2006, pp. 42–49.

[12] W. Liu, X. Wen, B. Duan, H. Yuan, and N. Wang, "Rear vehicle detection and tracking for lane change assist," in *Proc. IEEE Intell. Vehicles Symp.*, Istanbul, Turkey, Jun. 2007, pp. 252–257.

[13] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.

[14] Y. Zhang, S. J. Kiselewich, W. A. Bauson, and R. Hammoud, "Robust moving object detection at distance in the visible spectrum and beyond using a moving camera," in *Proc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, p. 131.

[15] C.-S. Fuh and P. Maragos, "Region-based optical flow estimation," in *Proc. Conf. Comput. Vis. Pattern Recog.*, Jun. 1989, pp. 130–135.

[16] B. Heisele and W. Ritter, "Obstacle detection based on color blob flow," in *Proc. IEEE Intell. Vehicles Symp.*, Detroit, MI, Sep. 1995, pp. 282–286.

[17] B. Heisele, U. Kreßel, and W. Ritter, "Tracking non-rigid, moving objects based on color cluster flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, San Juan, PR, Jun. 1997, pp. 257–260.

[18] C. Tzomakas and W. von Seelen, "Vehicle detection in traffic scenes using shadows," Institut für Neuroinformatik, Ruhr-Universität Bochum, Bochum, Germany, Tech. Rep. 98-06, Jun. 1998.

[19] M. B. V. Leeuwn and F. C. Goren, "Vehicle detection with a mobile camera," *IEEE Robot. Autom. Mag.*, vol. 12, no. 1, pp. 37–43, Mar. 2005.

[20] M. Bertozzi, A. Broggi, and A. Fascioli, "Stereo inverse perspective mapping: Theory and applications," *Image Vis. Comput.*, vol. 16, no. 8, pp. 585–590, May 1998.

Alberto Broggi (SM'89–S'93–A'96–SM'06) received the Dr.Eng. degree in electronic engineering and the Ph.D. degree in information technology from the Università degli Studi di Parma, Parma, Italy, in 1990 and 1994, respectively.

He is with the Università degli Studi di Parma, where he was an Associate Researcher with the Dipartimento di Ingegneria dell'Informazione from 1994 to 1998, was an Associate Professor of artificial intelligence with the Dipartimento di Informatica e Sistemistica from 1998 to 2001, and is currently a Full Professor of computer science with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione. He has authored more than 150 refereed publications in international journals, book chapters, and conference proceedings and delivered invited talks at many international conferences. His research interests include real-time computer vision approaches for the navigation of unmanned vehicles and the development of low-cost computer systems to be used in autonomous agents.

Dr. Broggi is the Editor-in-Chief of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and a member of the IEEE Intelligent Transportation Systems Society Council Executive Committee. He has been the Editor on the subject of intelligent transportation systems for the IEEE INTELLIGENT SYSTEMS MAGAZINE since 1999.

Andrea Cappalunga (S'08) received the M.Sc. degree in computer engineering in 2007 from the Università degli Studi di Parma, Parma, Italy, where he is currently working toward the Ph.D. degree with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione.

His research activity is focused on computer vision systems for intelligent and autonomous vehicles. He participated in the 2007 DARPA Urban Challenge as a member of team TerraMax.

Claudio Caraffi (S'05) received the M.Sc. degree in computer engineering and the Ph.D. degree in information technology from the Università degli Studi di Parma, Parma, Italy, in 2004 and 2008, respectively.

He is currently with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma. As member of team TerraMax, he participated in the DARPA Grand Challenge 2004 and 2005, and DARPA Urban Challenge 2007. He has been working on problems relating to system calibration, image stabilization, real-time obstacle detection, lane detection, and traffic sign recognition. His research interests are mainly in the area of computer vision applied to intelligent vehicles.

Stefano Cattani (S'05) received the M.Sc. degree in computer engineering and the Ph.D. degree in information technology from the Università degli Studi di Parma, Parma, Italy, in 2004 and 2008, respectively.

Since 2004, he has been with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma. His research interests are mainly focused on path and vehicle detection in the automotive field, based on monocular vision. He has been involved in DARPA Grand Challenge in 2004, 2005, and 2007 as a member of team TerraMax.

Stefano Ghidoni (S'04–M'09) received the M.Sc. degree in telecommunication engineering and the Ph.D. in information technology from the Università degli Studi di Parma, Parma, Italy, in 2004 and 2008, respectively.

He is currently with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma. He was part of team TerraMax during the 2007 DARPA Urban Challenge. His research focuses on obstacle detection and pedestrian recognition, applied to intelligent vehicles.

Paolo Grisleri (S'03–M'09) received the Dr.Eng. degree in computer engineering and the Ph.D. degree from the Università degli Studi di Parma, Parma, Italy, in 2002 and 2006, respectively.

Since 2002, he has been a Temporary Researcher with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma. His research interests are mainly focused on computer vision, data acquisition techniques, and system architectures for advanced driver assistance systems.

Pier Paolo Porta (S'06) received the M.Sc. degree in electronic engineering from the Università degli Studi di Parma, Parma, Italy, in July 2005, where he is currently working toward the Ph.D. degree with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione.

His research activity field is on autonomous vehicle and driver assistance system and is focused on short-range obstacle detection, camera calibration, and traffic sign recognition. In 2005, he was in charge of sensor fusion during DARPA Grand Challenge 2005. Since 2006, he has been responsible for a traffic sign recognition system developed in collaboration with Magneti Marelli and a close-range obstacle detector developed for the DARPA Urban Challenge 2007. He is the author of several publications on international scientific journals and conference proceedings.

Matteo Posterli received the M.Sc. degree in computer engineering from the Università degli Studi di Parma, Parma, Italy, in 2007.

From 2007 to 2008, he was a Researcher with the Department of Information Technology, Università degli Studi di Parma. He is currently with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma. His research activity was focused on computer vision systems for intelligent and autonomous vehicles. He participated in the 2007 DARPA Urban Challenge as a member of team TerraMax.

Paolo Zani (S'07) received the M.Sc. degree in computer engineering from the Università degli Studi di Parma, Parma, Italy, in 2005, where he is currently working toward the Ph.D. degree with the Artificial Vision and Intelligent Systems Laboratory (VisLab), Dipartimento di Ingegneria dell'Informazione.

He participated to the 2005 DARPA Grand Challenge and the 2007 DARPA Urban Challenge as a member of team TerraMax. His research activity is focused on efficient computer vision algorithms for intelligent and autonomous vehicles.