A Laserscanner-Vision Fusion System Implemented on the TerraMax Autonomous Vehicle

Alberto Broggi, Stefano Cattani, Pier Paolo Porta, Paolo Zani

VisLab - Dipartimento di Ingegneria dell'Informazione

Università di Parma

Parma, I-43100, Italy

{broggi,cattani,portap,zani}@ce.unipr.it

Abstract— **This paper presents a sensor fusion model developed for the 2005 Grand Challenge competition, an autonomous ground vehicle race across the Mojave desert organized by DARPA[1]. The two sensors used in this work are a stereo vision camera pair and an ALASCA laserscanner.**

An algorithm to filter laserscanner's raw scan data and to remove ground reflections is also presented.

Several tests were made to prove the reliability of this method, that has proved to be useful to extract the information required by the race.

Fusion was performed both at a low and medium level: terrain slope, detected with stereo vision, was used to correct pitch information of laserscanner raw data. Object segmentation is applied on a bird view bitmap where each pixel represents a square area of the world in front of the vehicle; this bitmap is obtained from the fusion of the ones generated by each sensor.

I. INTRODUCTION

When a human is driving a car, most of his senses are involved: vision, hearing, and even tact. Through the perceived information the driver can have a deeper description of the environment, and make suitable decisions. In the same way an autonomous navigation system can not rely upon a single type of sensor, but it is necessary to have different descriptions of the same scene, since each sensor has its own domain of applicability and may fail if used outsided of it.

For both a human being and an autonomous vehicle, the act of sensing means understanding the context in which he (or it) is involved, starting from the raw data sent from the sensors up to a symbolic description. Sensor fusion is therefore a process that develops on several abstraction layers[2].

Its main targets are:

- Increase sensor accuracy in a specified area
- Extend sensing coverage
- Increase results reliability
- Get more information form sensors' correlation
- Get an equivalent or even more robust sensor from the fusion of several cheap elements instead of a more expensive single one.

Obviously not all of these targets are always reached with fusion, but are meaningful to understand the power of this method.

---

[1]Defense Advanced Research Projects Agency.

[2]Generally also referred to as "levels".

Fig. 1. The TerraMax vehicle. Two ALASCA laserscanners are visible above and inside the front bumper, the three cameras and the two SICKS are placed on the bar over the hood.

To perform sensor fusion, it is also important to take into account how the sensors are going to work together, i.e. the sensor network. In the following, three of the most important kinds of sensor networks [1] are listed:

- **Redundant**: two or more sensors (usually of the same type) get data from the same area. This is useful in systems where is not possible to make a second measurement on the same scene (for example a moving vehicle) or when a backup is required. The network can work until there is at least one working sensor.
- **Complementary**: two or more sensors cover different areas (not only meant as a portion of space, but as sensing capabilities as well), and through fusion it becomes possible to have a wider environmental description. This is the case of different kinds of measurements of the same object at the same time (for example position and speed).
- **Cooperative**: it is possible to get additional information from fusion, in a way that can not be obtained from one sensor only. For example compute the position of one object starting from three distance measurements got from different sensors.

Usually sensor fusion for automotive purposes, both for road

safety and for autonomous vehicle projects, is performed at a high level. For example [2] presents a collision mitigation system, based on laserscanner and stereovision: data fusion in this case is executed after having extracted obstacles from both sensors' raw data. The information is used to evaluate the Time-To-Collision of each detected obstacle.

Tracking is another interesting feature that is usually related to sensor fusion. Many papers describe this characteristic related to the fusion process ([2], [3]). Tracking can be applied at different stages of the algorithms: in [2] it is placed inside each sensor's data computation, while in [3] it is located after data fusion and feature extraction.

Another completely different approach is described in [4] in which stereovision is used for validation of laser scan points. In particular tracked information and raw scan data are processed together to generate a new obstacles set. Due to the abundance of false detections, especially in case of strong pitch or non-flat terrain, stereo vision allows obstacles classification, increasing system reliability.

This paper presents a sensor fusion model developed for the 2005 Grand Challenge competition, the autonomous ground vehicle race across the desert organized by DARPA: it is based on both medium and low level fusion but no tracking is implemented in this version.

II. SYSTEM DESCRIPTION

The system described in this paper was developed in order to be used on the TerraMax vehicle (see figure 1) during the Grand Challenge competition 2005, the second edition of a challenge organized by DARPA to promote research on autonomous vehicles. This year the competition took place on October 8th 2005 in the Mojave desert in Nevada.

Team TerraMax was one of the five teams that finished the race. It is composed by Oshkosh Truck Corp., Rockwell Collins Inc., and University of Parma. The former provided an MTVR[3] truck modified with drive-by-wire and dealt with mechanical issues and logistics management. The second developed lidar processing, the vehicle control system, and path planning starting from sensor's high level data, and the last one realized a stereovision-based sensor for obstacle and path detection.

TerraMax is equipped with three front cameras placed to form three different baselines for stereo computation, one ALASCA laserscanners placed inside the front bumper and two SICKS mounted on the two sides of the truck, as shown in figure 1. A differential GPS system provides world coordinates to the vehicle.

A second laserscanner (also shown in figure 1) was installed as a backup above the bumper and it would have been used only in case of failure of the main one.

On board, embedded inside the passenger seat, there are seven PCs. One of them, called Vision PC, is used for acquiring images from the cameras and run the vision algorithms, while another one is an ECU, used to get laserscanner data. All the data transfer among them is performed via ethernet.

[3]Medium Tactical Vehicle Replacement.

A. Vision System

Vision System hardware is composed by three Micropix C-640 cameras placed on a rigid bar fixed to the truck chassis and a Pentium 4 PC. The distance between the left and right cameras is 1.5 m, while the central camera is placed 0.5 m far from the right camera and 1 m from the left one: in this way three baselines (1.5 m, 1.0 m, and 0.5 m) are available for stereo 3D reconstruction [5]. The vision system automatically selects the baseline –amongst the three available– depending on vehicle speed. When moving slowly (from 0 to 7 mph) the short baseline is used to focus the sensor's attention towards the region close to the vehicle, while at high speed (over 27 mph) the algorithm is tuned to detect far obstacles (up to 50 m away). In the other case, the medium baseline is used.
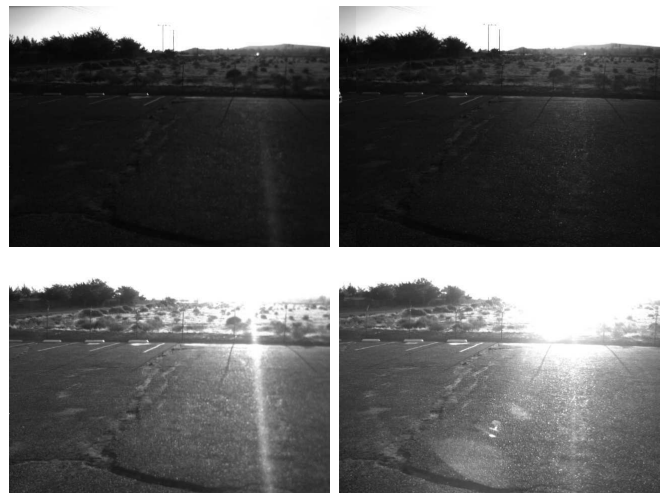

Fig. 2. Automatic Gain Control examples in direct sunlight. In the top images AGC is disabled while on the bottom ones it is enabled: the region of interest of the algorithm is the bottom part of these images.

An automatic gain control scheme is implemented into the image acquisition layer in order to improve the quality of the images and avoid image saturation or sub-exposure of the region of interest (i.e. the lower part of the image). See figure 2 as an example.

B. Laserscanner

The ALASCA laserscanner features a multi-layer technology and a multi-target support. The former allows this device to send four different light shots on different layers for each scan angle and the second makes the laserscanner wait up to two echoes for each shot. This approach lets the device perceive even below glass plate or metal fence or in rainy weather conditions.

The angle resolution is $0.25°$ at 10 Hz (the revolution frequency used on TerraMax) and the four layers are included between $+1.6°$ and $-1.6°$.

The ALASCA is connected via ARCnet to the ECU that sends scan data via ethernet to the vision PC. The ECU also features a CAN bus that carries high level data, i.e. a list of

objects generated by the ECU itself, but this feature is not used in our system.

Cameras and laserscanner are synchronized by means of a sync box, so that data from the two sensors refer exactly to the same scene. The two systems are shown in figure 3.

Fig. 3.   Sensors used for fusion on TerraMax.

III. ALGORITHMS

Two sensor fusion algorithms were developed [6]: one to remove ground reflections from laserscanner data and the other to fuse two bitmaps generated by the two sensors, each representing the same area in front of the vehicle.

A. Ground reflections removal

One of the main issues of using a laserscanner in an unstructured environment is the large amount of noisy data that is generated by ground reflections and need to be filtered before the obstacle detection step takes place. The two main causes are: uneven terrain and high vehicle pitch variations.

The algorithm that we have developed is based on the assumption that an obstacle is something in a vertical position while the ground is typically flat, quasi-horizontal or slightly bended. The ALASCA laserscanner, as described above, provides up to four scan points for each angle because of its multi-layer technology. Therefore these reflections (looking at a bird view map) in case of an obstacle will be clustered in a small area, otherwise in the second case –the ground– the scan points will be scattered in a large area.

The first step is to split the region of interest in front of the truck in 0.7 m side squares, then look inside each square to investigate how many echoes coming from different layers are present. If the scan points belong to a single layer, all the data that are inside the selected square are considered ground, and deleted. Otherwise they are maintained for further consideration. The value of 0.7 m was chosen via empiric tests.

The algorithm flowchart is shown in figure 4.

B. Pitch correction

The fusion between the stereo vision system and the ALASCA laserscanner was performed in two parts of the software. One inside the scan data processing and the other before the obstacle detection procedure.

Data received from laserscanner are related to the reference system fixed on the device itself, therefore a rototranslation is necessary to translate coordinates into the world reference system. In an unstructured environment real-time vehicle pitch suffers from abrupt and sudden variations. Since inertial sensors deliver pitch variations with a non negligible delay, the
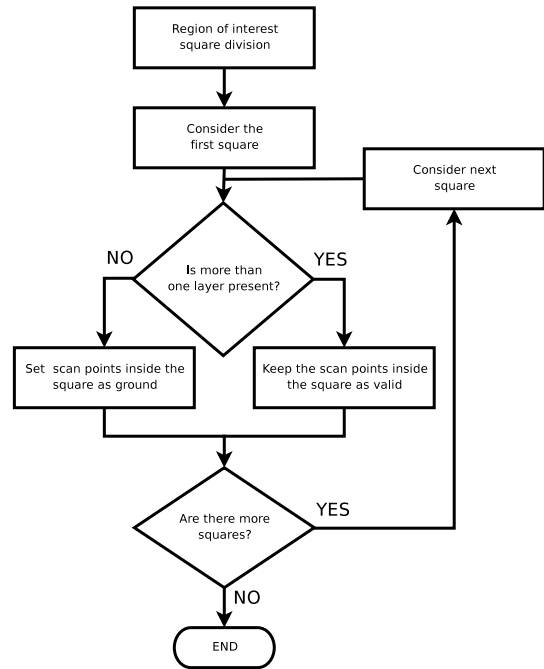

Fig. 4.   Ground reflections removal flowchart.

only feasible solution is to get real-time pitch values from stereo vision: starting from the V-disparity image [7], [5], it is possible to compute the cameras absolute pitch, and therefore correct laserscanner scan points world coordinates:

$$\theta_{ls,real-time} = \theta_{ls,static} + \theta_{cam,static} - \theta_{cam,real-time} \quad (1)$$

where $static$ values mean in rest mode, ls means laserscanner, and cam means camera. θ is the pitch symbol.

C. Sensor fusion

The main reason why fusion was developed is to improve obstacle detection and the results reliability. It can be classified between *redundant* and *complementary* types: the laserscanner covers the vision system field of view and is extended also on the sides. This system is independent of the number and type of sensors used, operating in a medium abstraction layer. Object segmentation is done on fused data.

Let's now analyze in detail the steps of the fusing process, summarized in figure 5:

- Raw data are acquired by pre-processing modules.
- Raw data are processed by each sensor module to filter noisy information.
- Each sensor module generates a bird view bitmap, encoding in each pixel a gray-scale color proportional to the chance that an obstacle is present in that area. This step is called "first weighing", and is performed on each sensor and related to the condition of that specific situation. It is a dynamic weighing.
- A second static weighing is applied to those bitmaps, based on the reliability of the sensor in the region of interest:
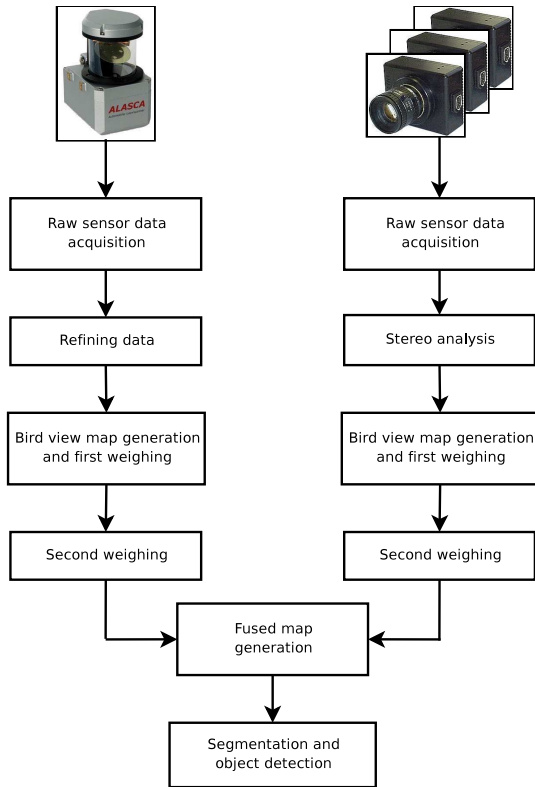
Fig. 5.   Sensor fusion flowchart as tested on TerraMax.

$$Sensor_j(i) = Sensor_j(i) \cdot Weight_j(i) \qquad (2)$$

where $Sensor_j$ is the j-th bitmap and $Weight_j$ is the weight map associated to the j-th sensor.

This map is filled offline referring to the capabilities of the sensor itself, maybe starting from empirical considerations on sensor reliability or from datasheet provided by the manufacturer.

- Now the bitmaps are ready to be fused together as a pixelwise average:

$$FusedMap(i) = \sum_{j=1}^{N} \frac{Sensor_j(i)}{N} \qquad (3)$$

where $FusedMap$ is the resulting bitmap and N is the number of bitmaps to fuse. The fused map can be considered as generated from a single sensor. On the bitmap generated we then execute segmentation in order to detect obstacles.

IV. RESULTS AND CONCLUSION

This section presents the results of the ground reflection removal algorithm and shows an example of a bitmap generated for sensor fusion. Several scenarios are presented to discuss the reliability of this kind of filtering. The tests described here took place off-road, in a typical Grand Challenge environment.

These figures show the filtered data remapped on a camera image and two bird view maps: the one on the left displays raw data and the other shows refined data after the ground reflection removal. Data remapping is possible since both positions of laserscanner and vision system are known: with a 3D to 2D coordinates transformation scan points are projected on images.

The system was tested in a dirt road and good results have been reached in cases of heavy pitch condition, or during a turn, where centrifugal force causes a significant raise of roll angle (as shown in figures 6.a, 6.b, 6.c); lateral bushes echoes are maintained while ground reflections are removed in both cases. In figure 6.a also the furthest bushes are deleted, this happens because of the high roll angle that makes laserscanner hit those obstacles only with one layer. Because of the distance of the obstacle (about 40 m), this is not considered a problem as the goal is mainly to have a stable result in the proximity of the vehicle, without false positives, even without perceiving some correct detections at greater distance.

In other situations, in correspondence to a softer turn as in figure 6.b, the bushes are maintained even if they are far, because hit by more than one layer.

In figures 6.d and 6.e the vehicle is running in a partially structured environment, i.e. near a building, parked cars or a fence. All the boundaries, related to vertical surfaces, are preserved and figure 6.e shows how the system can perceive even behind a metal fence.

However there are some situations that need to be examinated in detail. One of those is shown in figure 7.a: the vehicle is going downhill on a dike, due to the high pitch the hypothesis of horizontal ground is not valid anymore. In figure 7.b the slope is softer and no problem were noticed.

Another case that summarizes a limit situation is described in figures 7.c and 7.d: the vehicle is running in a wood; trees and bushes are detected as obstacles but while trees are insurmountables barriers, bushes are not. Therefore it seems to be necessary to introduce obstacle classification. This could be possible with an high level fusion: in fact the vision systems are typically used to improve laser-based classification capabilities [8].

The tests shown in figure 7.e are made while the vehicle is climbing a hill: on the very left of the images some fake echoes are still present; this is due to the high roll of the vehicle and a lateral slope, however this is not a big issue since placed in a peripheral area.

Another set of results, shown in figure 8, are examples of bitmaps generated from laserscanner module.

The system was tested in a number of scenarios and proved to generate reliable results. But was not used during the Grand Challenge due to the limited time for an exhaustive test and final integration.
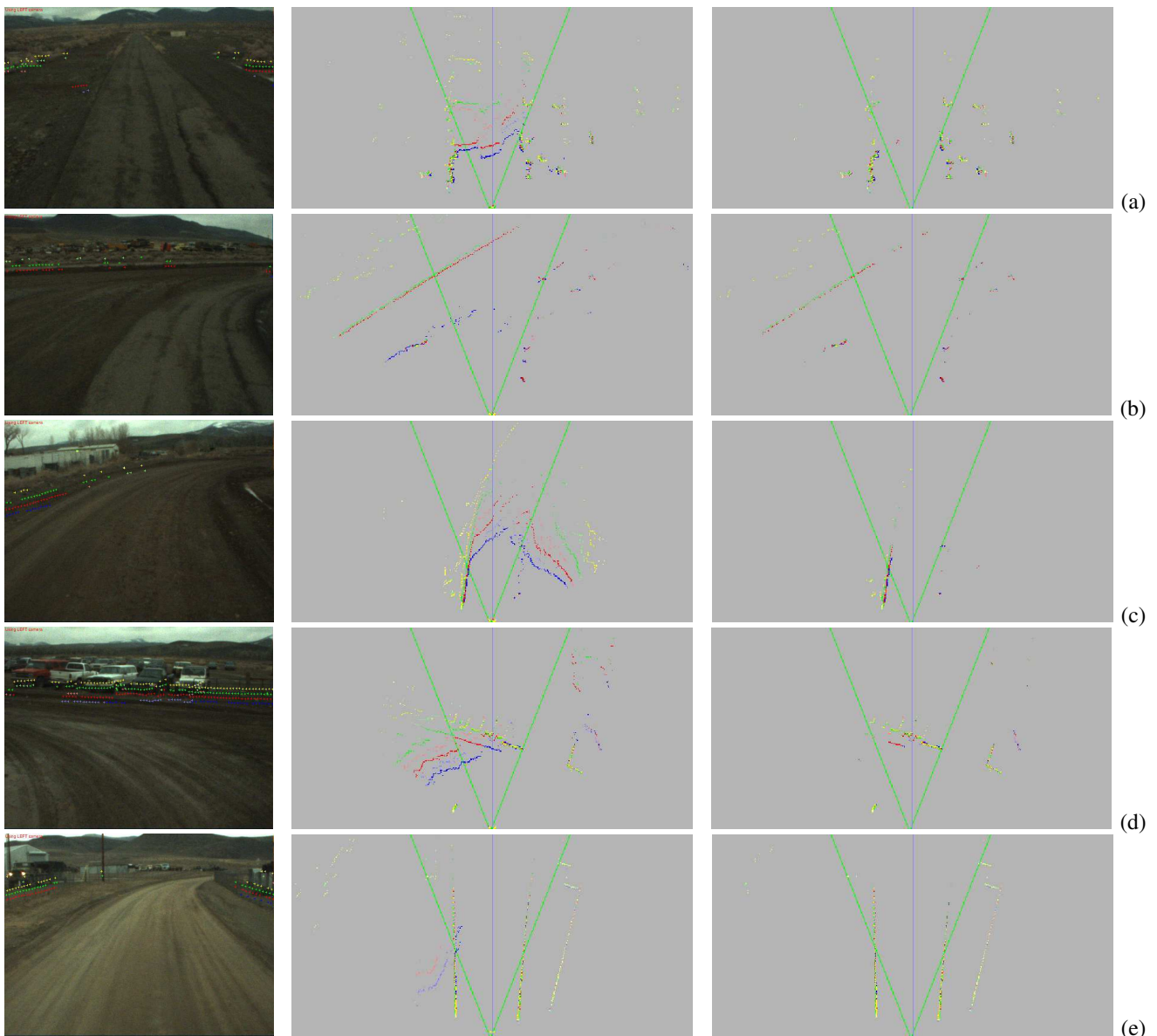
Fig. 6. From the left: filtered scanpoints remapped on the image, bird view of raw data, and bird view of data after ground reflections removal. The different colors refer to the four laserscanner layers. (a) The vehicle pitched down. (b) The vehicle is approaching a right turn. (c) Deep rolling during a right turn. (d) Partially structured environment: parked cars and a fence. (e) On the right both metal fence and the building behind it are detected.

REFERENCES

[1] A. Kirchner, H. Weisser, T. Sharnhorst, and D. Stüker, "Sensor Fusion - A research perspective," in *Vehicle Safety - New concepts for automotive safety*, Mar. 2001.

[2] R. Labayrade, C. Royere, and D. Aubert, "A Collision Mitigation system using Laser Scanner and Stereovision Fusion and its Assessment," in *Procs. IEEE Intelligent Vehicles Symposium 2005*, Las Vegas, USA, June 2005, pp. 441–446.

[3] N. Kaempchen, M. Buehler, and K. Dietmayer, "Feature-Level Fusion for Free-Form Object Tracking using Laserscanner and Video," in *Proceedings of 2005 IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, June 2005.

[4] M. Perrollaz, R. Labayrade, C. Royère, N. Hautière, and D. Aubert, "Long range obstacle detection using laser scanner and sterovision," in *Procs. IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan, June 2006, pp. 182–187.

[5] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle Detection with Stereo Vision for off-road Vehicle Navigation," in *Procs. Intl. IEEE Wks. on Machine Vision for Intelligent Vehicles*, San Diego, USA, June 2005.

[6] P. P. Porta, "Studio dell'integrazione di un sensore laser con un sistema di visione per la guida automatica in ambiente non strutturato," Master's thesis, Facoltà di Ingegneria, Università di Parma, July 2005.

[7] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection on non flat road geometry through V-disparity representation," in *Procs. IEEE Intelligent Vehicles Symposium*, Versailles, France, June 2002, pp. 646–651.

[8] K. Weiss, D. Stueker, and A. Kirchner, "Target modeling and dynamic classification for adaptive sensor data fusion," in *Procs. IEEE Intelligent Vehicles Symposium*. Columbus, USA: IEEE Computer Society, June 2003, pp. 132–137.
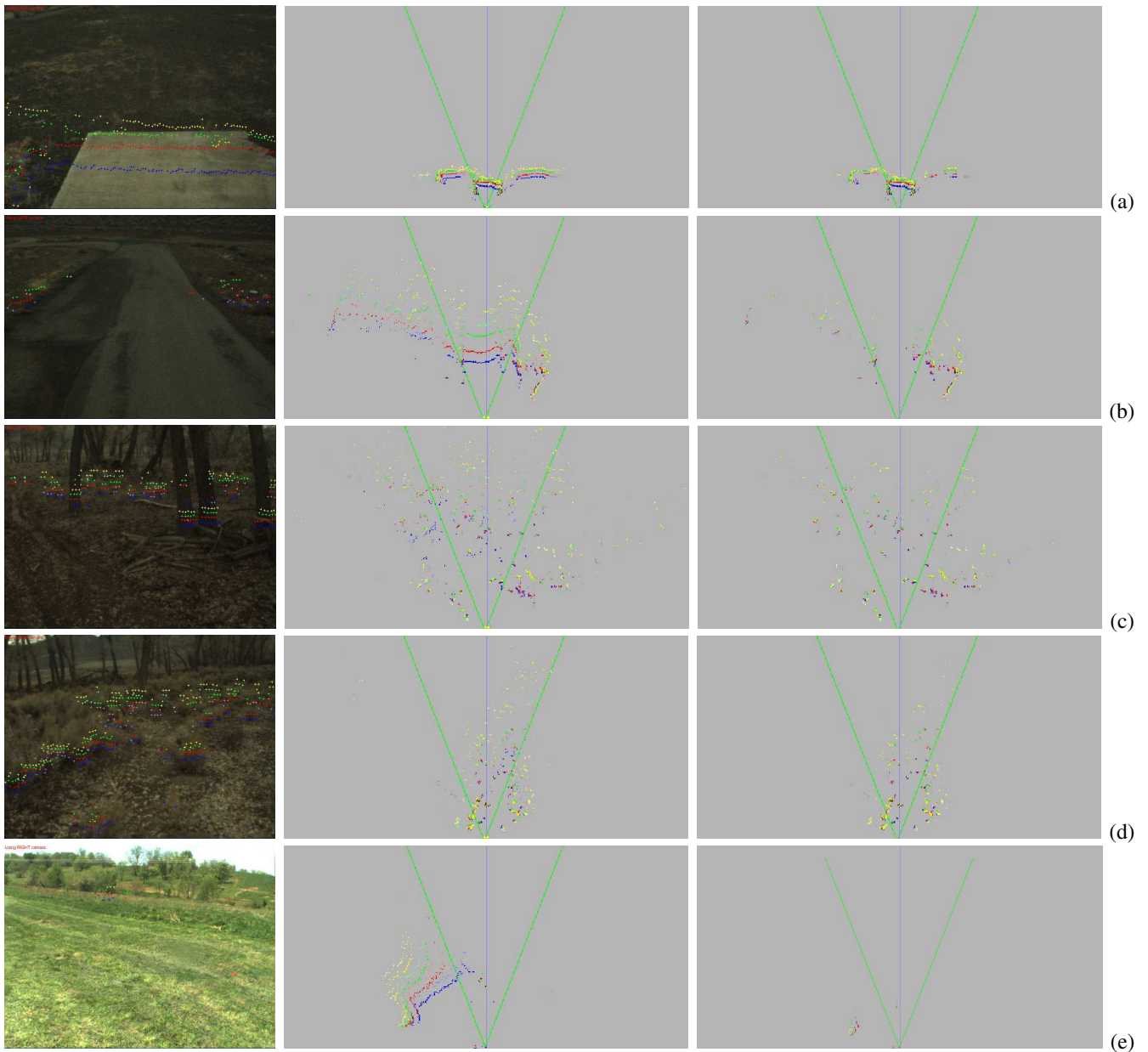
Fig. 7. From the left: filtered scanpoints remapped on a camera image, bird view of raw data, and bird view of data after ground reflections removal. The different colors refer to the four laserscanner layers. (a) Going downhill on a dike. (b) Vehicle is approaching the end of downhill. (c) Vehicle is approaching some trees. (d) The bushes in front of the vehicle are detected as obstacles. (e) High roll.
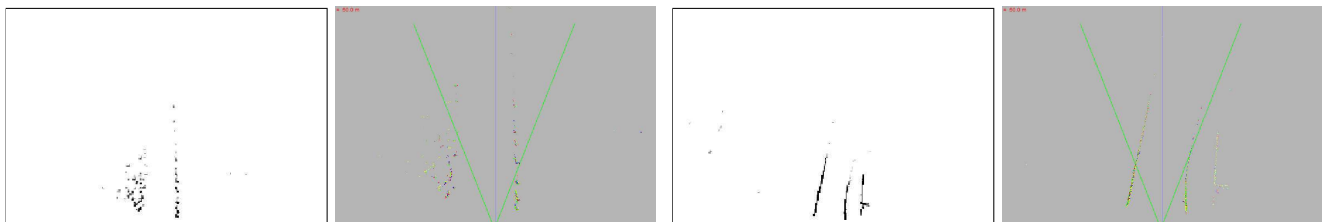

Fig. 8. Pairs of filtered data bird view map and bitmap generated for sensor fusion.