Excellabust
INNOVATION MANAGEMENT TRAINING 1
18 - 20 July 2016
Consiglio Nazionale delle Ricerche

Cooperative perception and environment characterisation

Dr. Annalisa Milella

E-mail: milella@ba.issia.cnr.it

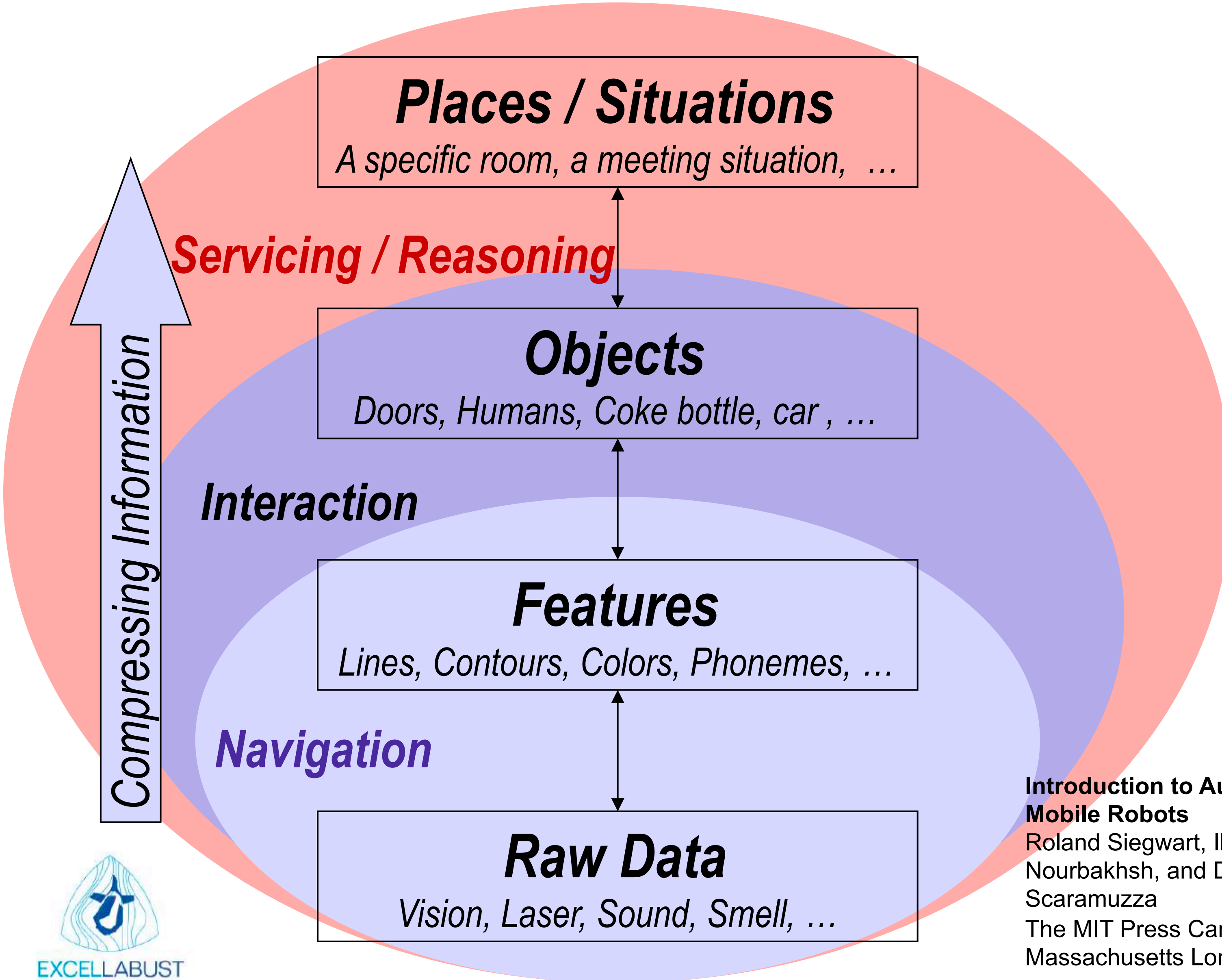
Outline of presentation

- Robotic perception: *main sensors and sensor classification*
- Vision sensors
  - *Case study*: a robotic visual inspection system for ship hull surfaces
- Perception in challenging environments
  - *Case study*: the QUAD-AV project – Ambient Awareness for Autonomous Agricultural Vehicles (multi-sensor ground/obstacle detection in unstructured outdoor environments)
- Cooperative perception
  - *Case study*: Air/ground vehicle cooperation for environment mapping: motivation and main challenges

Robot Perception

- **A robot has to perceive its environment** in order to interact (move, find and manipulate objects, etc.) with it.
- Perception allows making an **internal representation** (model) of the environment, which has to be used for moving, avoiding collision, finding its position and its way to the target, and finding objects to manipulate them.
- Without a sufficient environment perception, the robot simply **can't make any safe displacement or interaction**, even with extremely efficient motion planning systems.
- **The more unstructured an environment is, the most dependent the robot is on its sensorial system**. The success of industrial robotics relies on rigidly controlled and planned environments, and a total control over robot's position in every moment. But as the environment structure degree decreases, robot capacity gets limited.

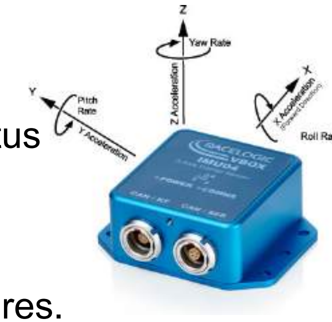
The perceptual pipeline: from sensing to ambient awareness

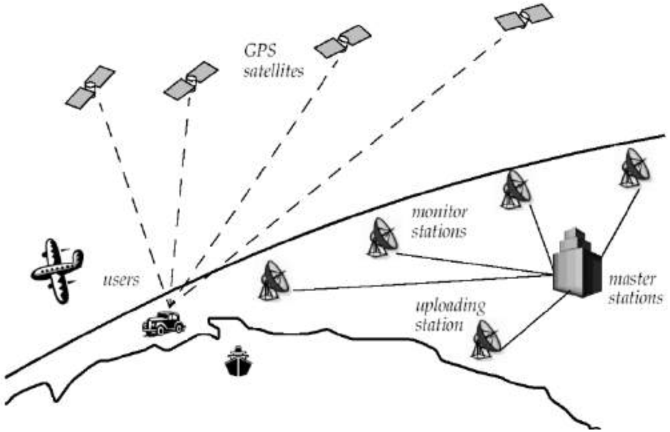


Introduction to Autonomous Mobile Robots
Roland Siegwart, Illah R. Nourbakhsh, and Davide Scaramuzza
The MIT Press Cambridge, Massachusetts London, England

Sensor classification

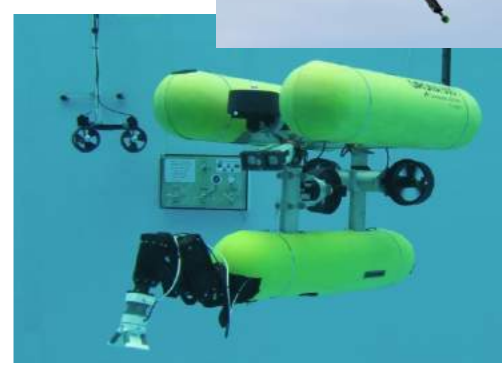
- What:
  - Proprioceptive sensors
    - measure values internally to the system (robot),
    - e.g. motor speed, wheel load, heading of the robot, battery status
  - Exteroceptive sensors
    - information from the robots environment
    - distances to objects, intensity of the ambient light, unique features.
- How:
  - Passive sensors
    - Measure energy coming from the environment
  - Active sensors
    - emit their proper energy and measure the reaction
    - better performance, but some influence on environment

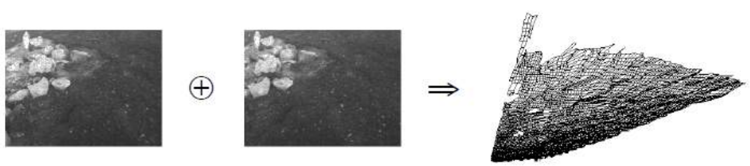




Vision sensors

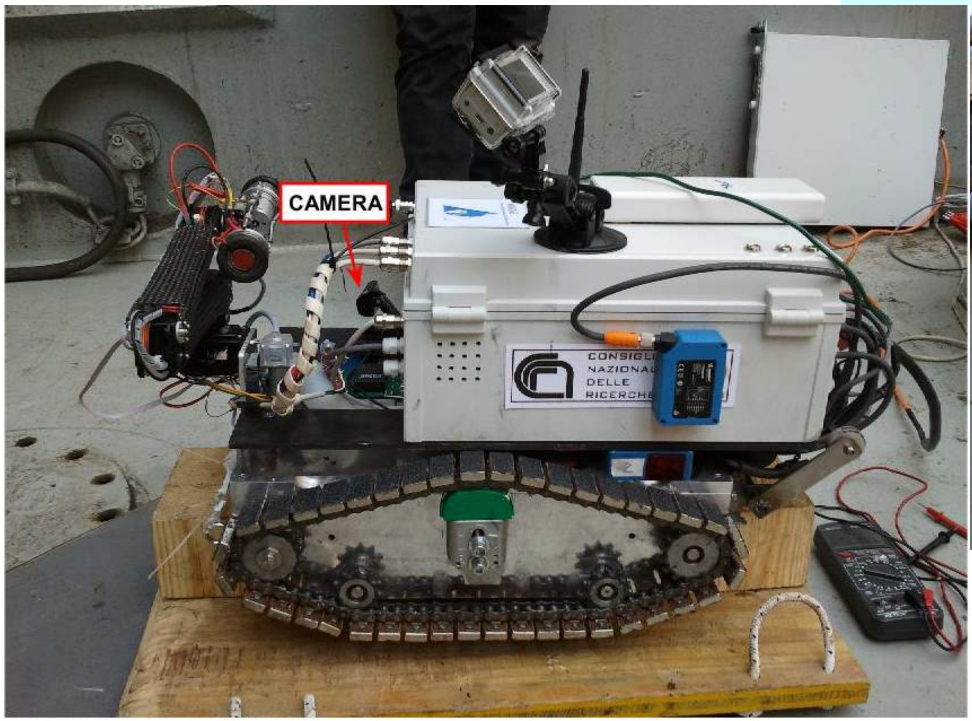
- Images carry a vast amount of information
- A vast know-how exists in the computer vision community
- Cameras: low cost, light and power saving
- Perceive data:
  - In a volume
  - Very far
  - Very precisely
- Easy to integrate on board
- Stereovision:
  - 2 cameras provide depth

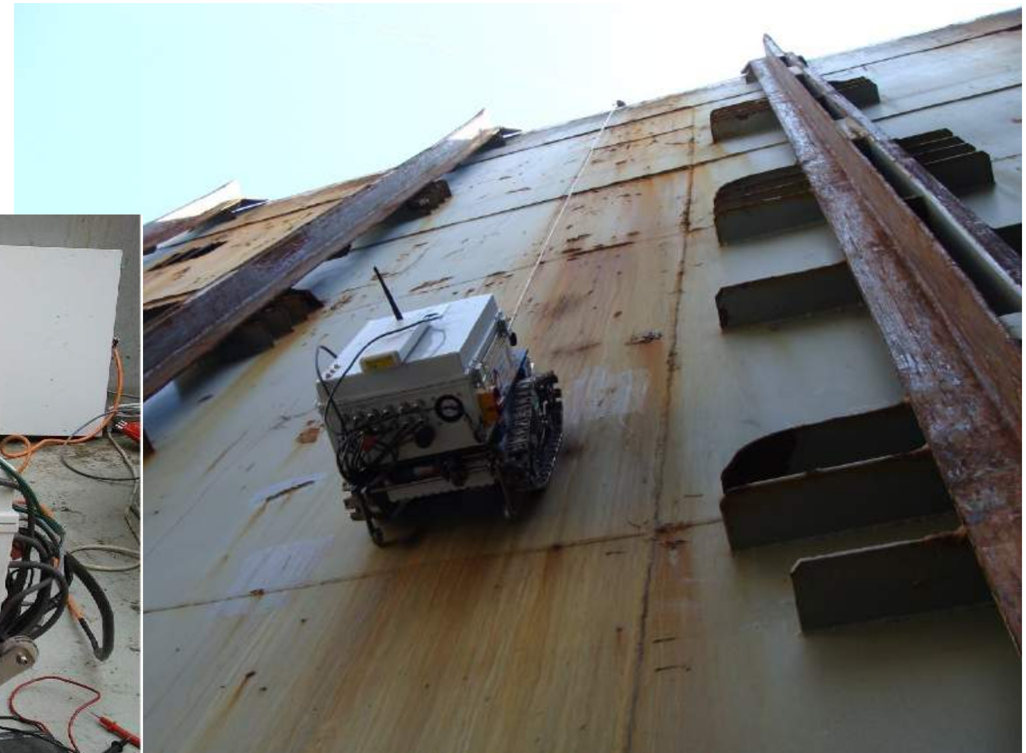




EXCELLABUST

Magnetic Autonomous Robotic Crawler (MARC)

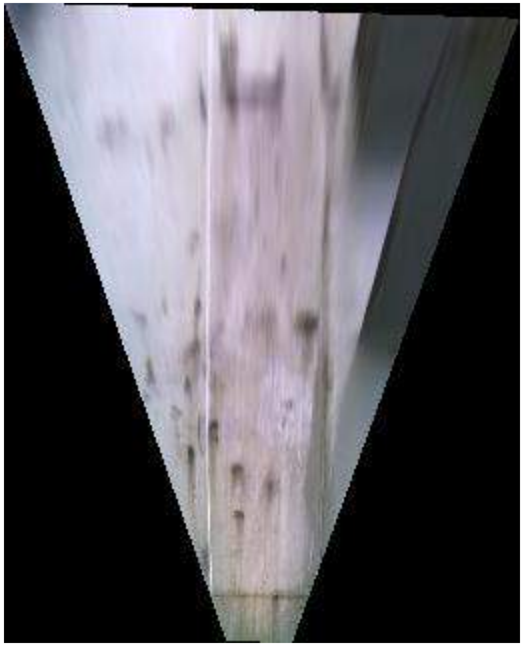




Scene mosaicking and reconstruction

SAMPLE IMAGE

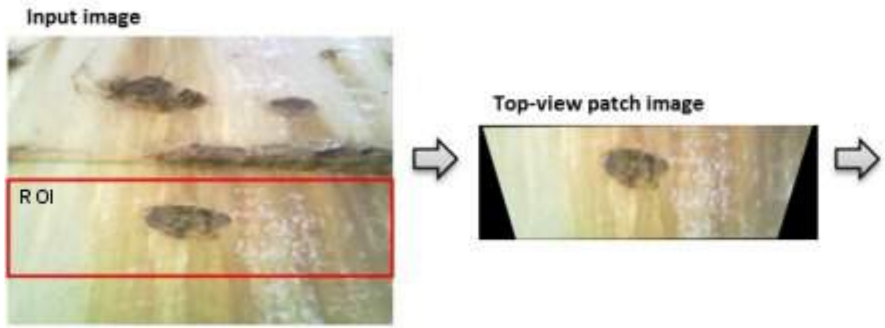
VIRTUAL TOP VIEW

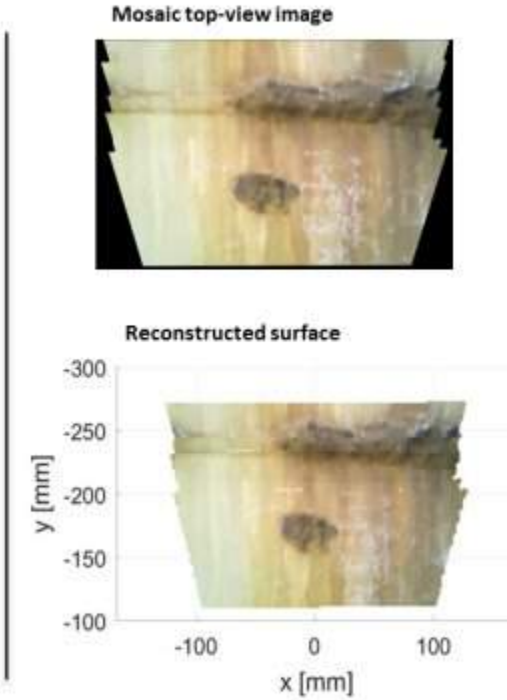


EXCELLABUST

Scene mosaicking and reconstruction

Near-to-far strategy





Mosaicking and reconstruction process.
From left to right:

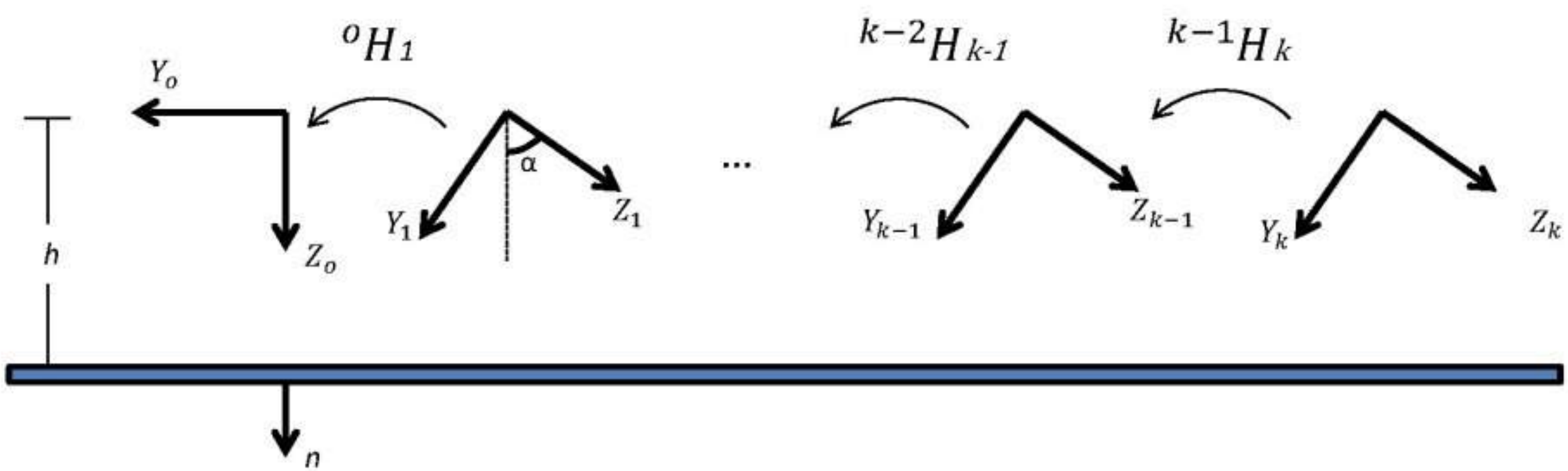
- a ROI is selected near the bottom of the image;
- the selected image patch is rectified to produce an overhead view;
- the overhead patch image is stitched together with subsequent overhead patch images to produce an overhead mosaic image;
- a metric reconstruction is obtained by converting pixel points into world points.

Mosaicking and motion estimation

$$\lambda \mathbf{x_i} = \mathbf{C}({}^{\mathbf{i}}\mathbf{R_j} + \frac{{}^{\mathbf{i}}\mathbf{T_j}\mathbf{n}^{\mathbf{T}}}{h})\mathbf{C}^{-1}\mathbf{x_j} = {}^{\mathbf{i}}\mathbf{H_j}\mathbf{x_j} \quad (1)$$

$${}^{0}\mathbf{H_k} = {}^{0}\mathbf{H_1}\,{}^{1}\mathbf{H_2} \ldots {}^{k-1}\mathbf{H_k} \quad (2)$$

Hypothesis:
- standard pin-hole camera model
- an estimate of the camera intrinsic (i.e., focal length, principal point, distortion coecients) and extrinsic (i.e., camera height h and tilt with respect to the hull surface) parameters is available from calibration;
- the inspected surface can be approximated by a plane, and contains enough texture for the registration process.

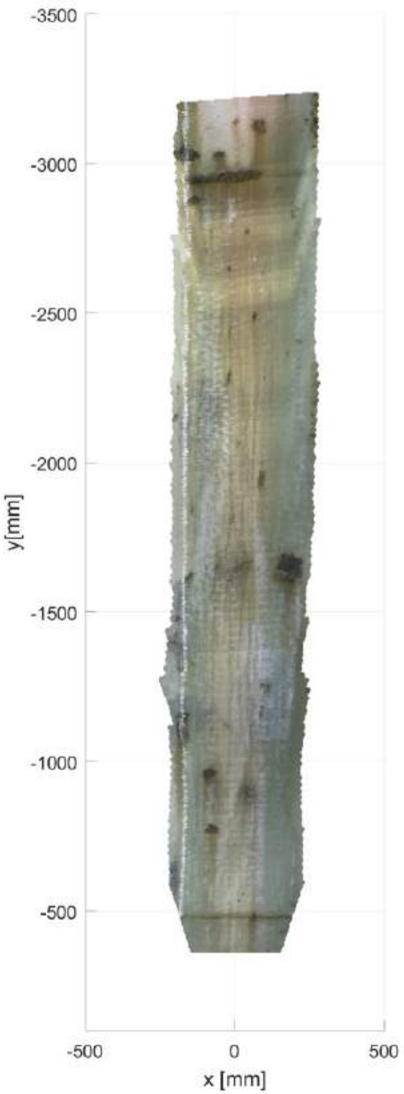


Experimental results

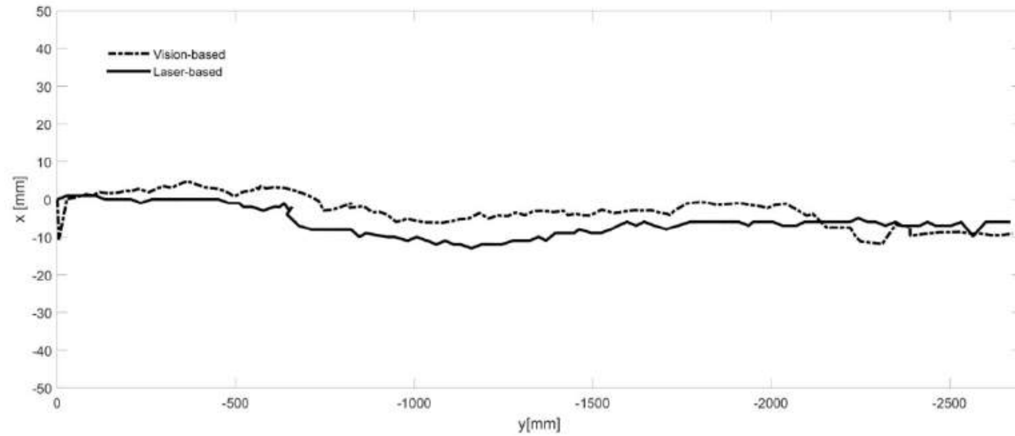
- The robotic visual inspection system proposed in this work was tested in field trials performed on a bulk carrier ship that was under repair in a shipyard in Varna, Bulgaria.
- The experiments consisted in MARC's following a vertical stiffener frame, while climbing a bulkhead at a maximum speed of 0.12 m/s.
- MARC was able to follow a linear surface on the basis of the measurements of vehicle heading and lateral displacement with respect to the stiffener provided by the lateral laser range sensors. The rear laser was also used to get an estimate of the robot translation in the vertical direction.
- A ramp was placed on the ground to allow the robot to autonomously climb on the ship hull.
- Images were acquired by the on-board camera at a resolution of 320 240 pixels and a frame rate of 10 Hz, and were processed offline. Specifically, one every fifth frame was processed for the mosaic reconstruction.

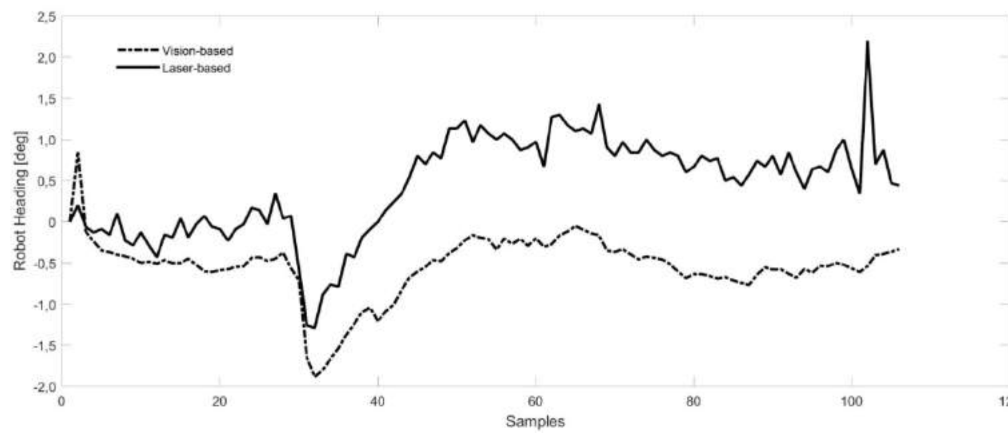
EXCELLABUST

Experimental results



106 frames



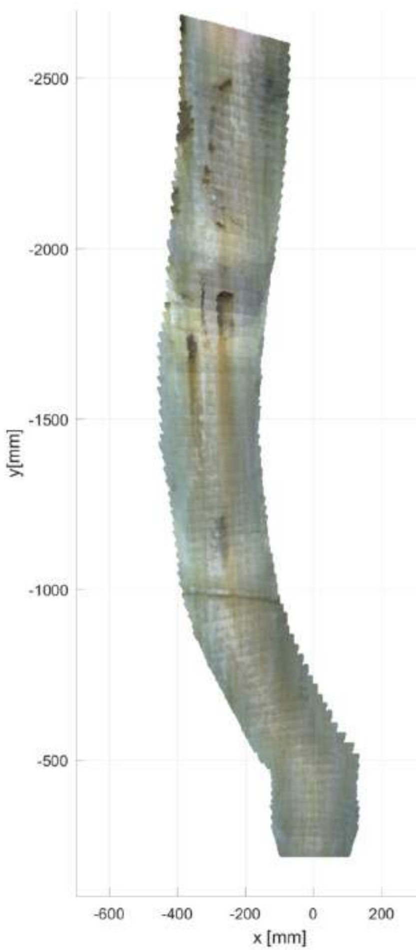


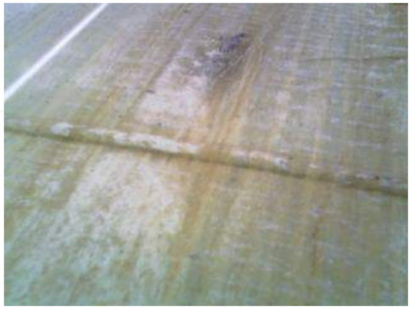
$E{\downarrow}p = 10.7 \pm 5.01$
$E{\downarrow}\theta = 0.99 \pm 0.47$

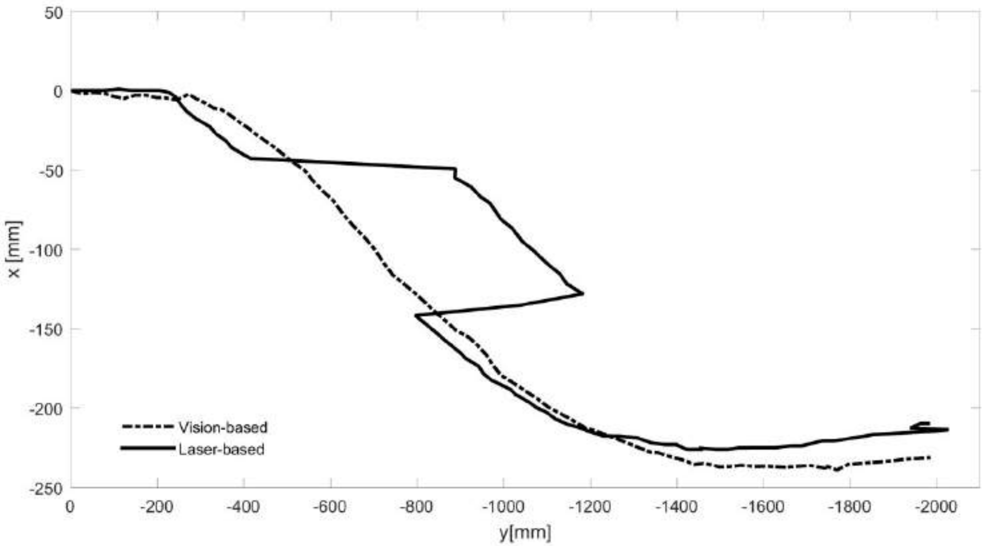
EXCELLABUST

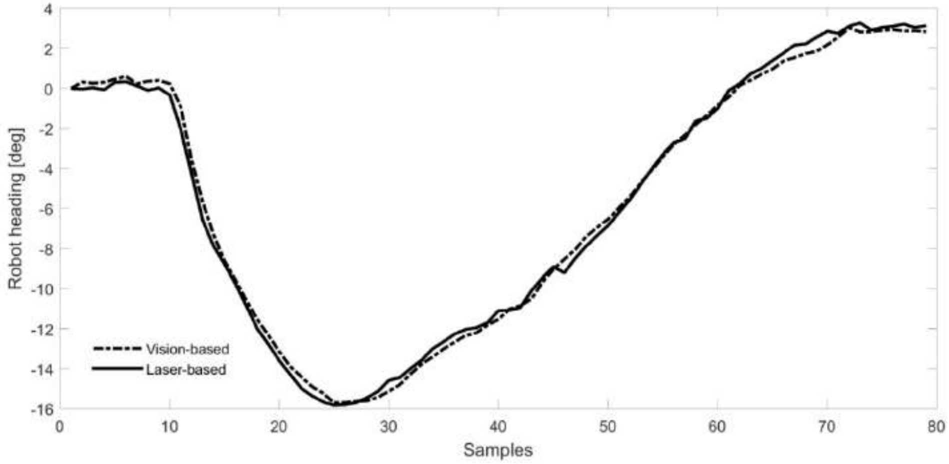
Experimental results

79 frames





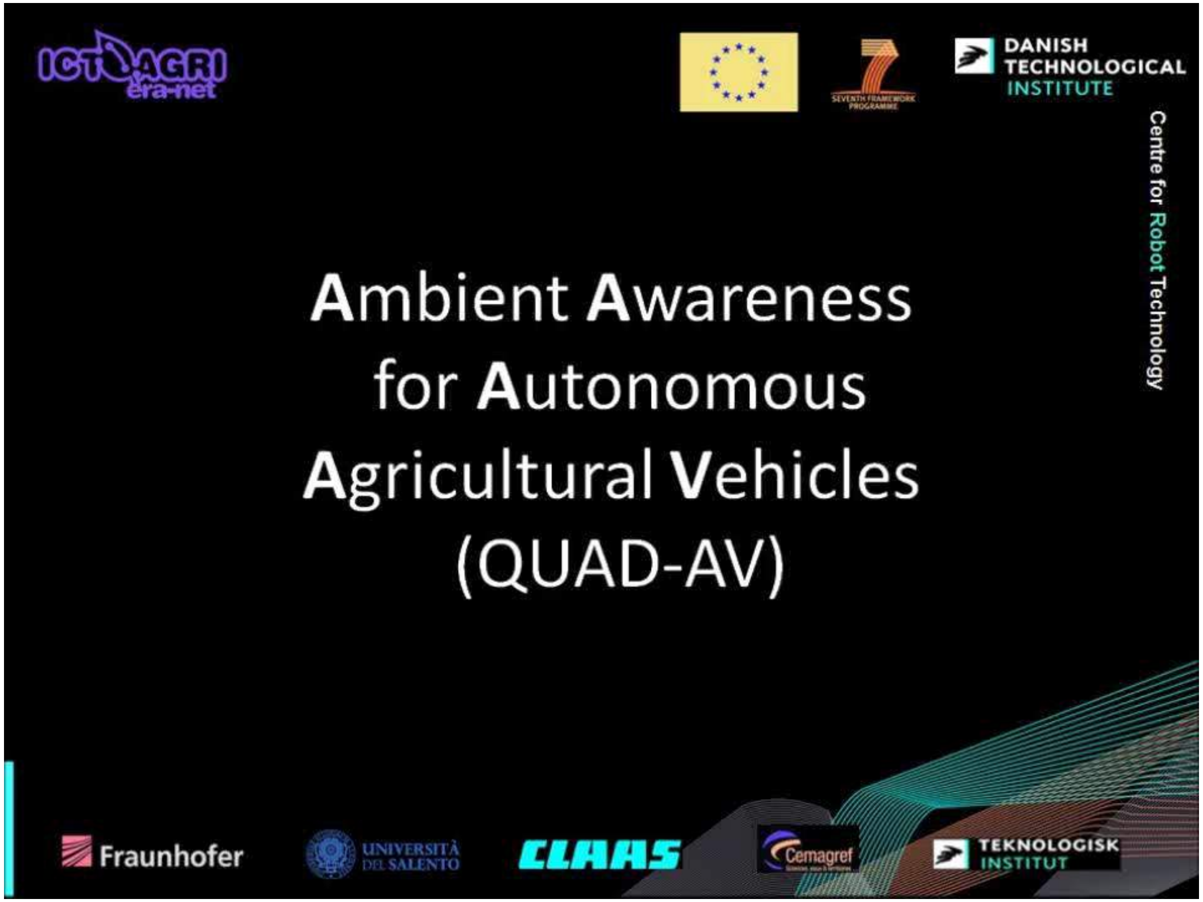




$E\downarrow\theta = 0.32 \pm 0.21$

Perception in challenging environmental conditions

Ambient awareness for agricultural robotic vehicles



G. Reina, A. Milella, R. Rouveure, M. Nielsen, R. Worst and M. R. Blas, "Ambient awareness for agricultural robotic vehicles," Biosystems Engineering, Volume 146, June 2016, Pages 114–132

Multi-sensory perception for obstacle detection

Positive obstacles

Negative obstacles

Moving obstacles / Live animals / People

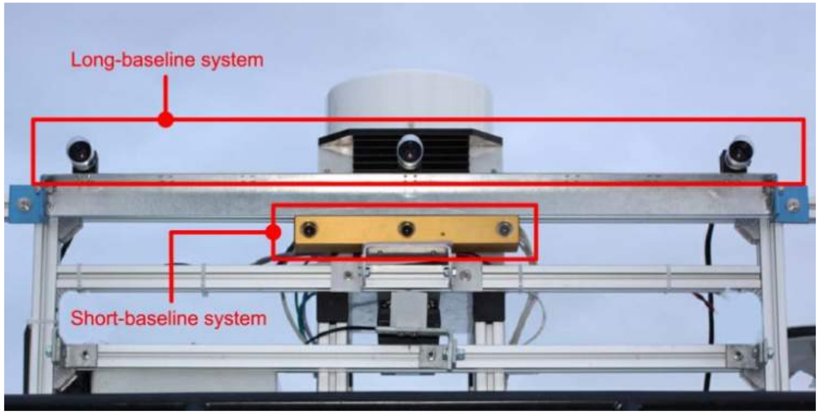
Difficult Terrain

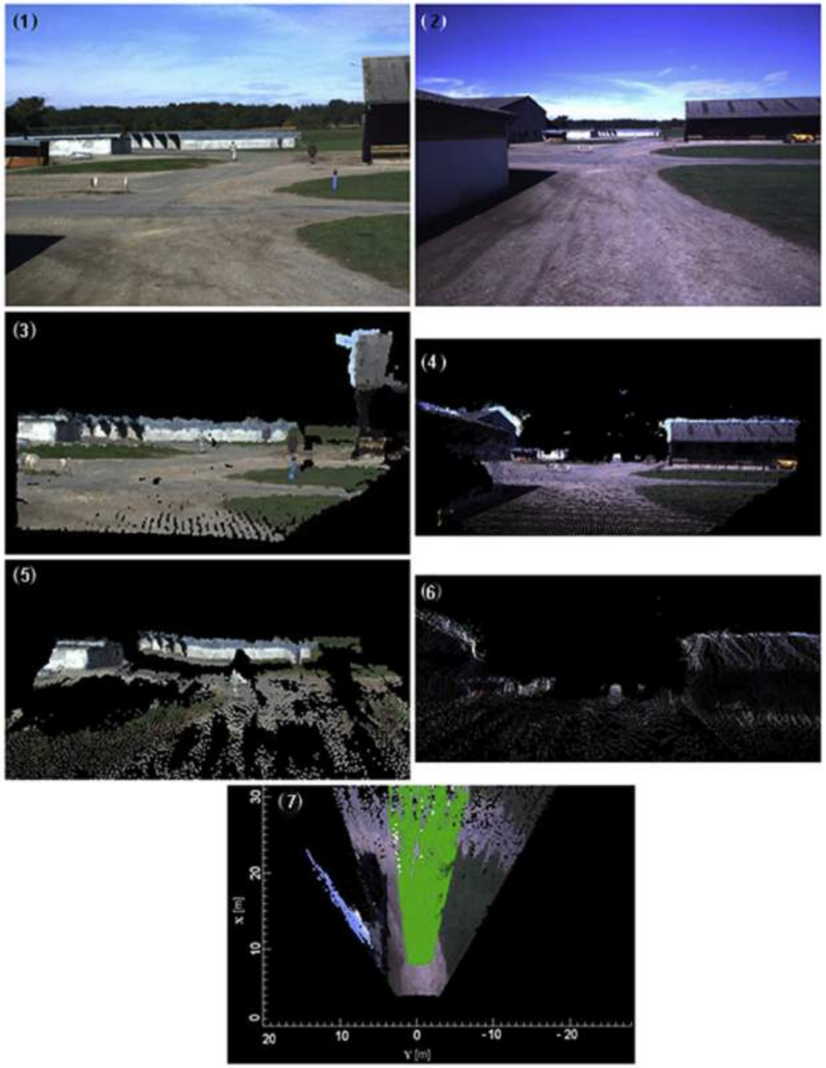
The potential of four technologies is investigated: stereovision, LIDAR, radar, and thermography. Advantages and disadvantages of each of these technologies are summarised in Table 1.

Table 1 – Advantages and disadvantages of different sensor modalities for outdoor perception.

Sensor modality	Advantages	Disadvantages
Mono/stereovision	Natural interpretation for humans	Risk of occlusions
	Relatively high resolution	Sensitive to lighting conditions
	Relatively high sampling rate	Poor performance in low visibility conditions (rain, fog, smoke, etc.)
	Rich content (colour and texture for mono and stereovision, range for stereovision)	Relatively short range (up to 15–30 m)
		Advanced processing algorithms and special hardware for data acquisition.
LIDAR	Accurate ranging over medium range (up to 30–40 m)	Costs related to accuracy and range
	Narrow beam spread	Sensitive to dust, fog and rain, and rounded surfaces
	Fast operation	Some risk of occlusion
	Fast lock on time	No colour or texture information
	Integration with rotating platforms for 3D acquisition	Relatively low sampling rate (except for high-end sensors)
Radar	Robustness to environmental conditions	Difficulty in signal processing and interpretation
	Long range (up to 100–150 m)	Low resolution
	Panoramic perception (360°)	3D map limited to pencil beam radar
	Multiple targets	
Thermography	Invariant to illumination	Relatively low resolution
	Robust against dust and rain	Some risk of occlusion
	Detect humans and animals	Difficulty in calibration

Multi-baseline stereo





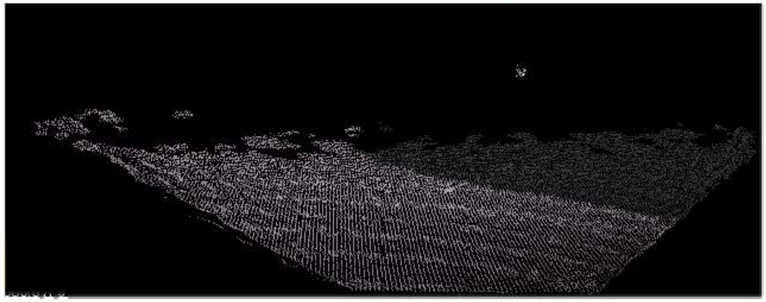
Short baseline system:
- Uses Bumblebee XB3
- 2 baselines: 12-24 cm
- 3.8mm focal length lenses

Long baseline system:
- Custom-built with 3 Flea3 cameras
- 2 baselines: 40-80 cm
- 12 mm focal length lenses

3D Reconstruction

XB3 image

Short range reconstruction by XB3



Flea3 image

Long range reconstruction by Flea3

Stereo-based Ground Detection

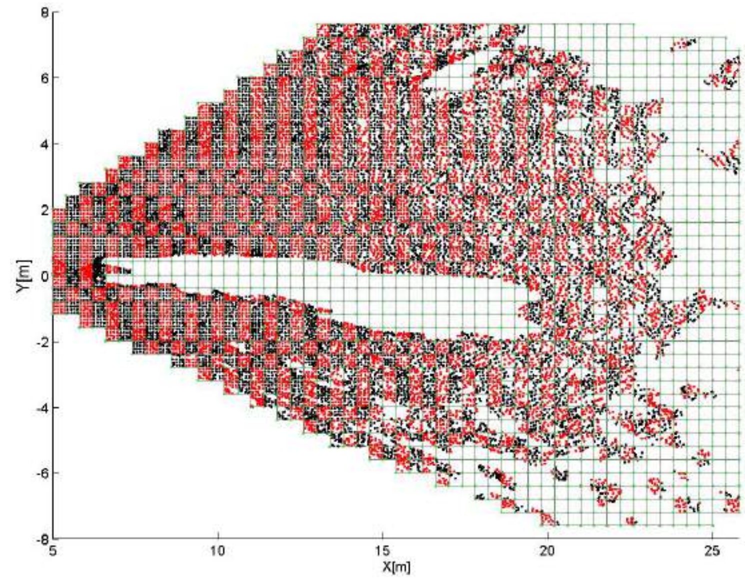
- A method for labeling observations based on their geometric properties, classifying the scene into **ground** and **non-ground** regions.
  - The ground model is constructed upon a set of **geometric features** that can be extracted from stereo reconstruction.
- **Self-learning scheme**: training instances are automatically produced, thus avoiding manual labeling.
- The model (i.e., the training set) is continuously updated using the most recent acquisitions to account for variations in ground characteristics during the vehicle travel.

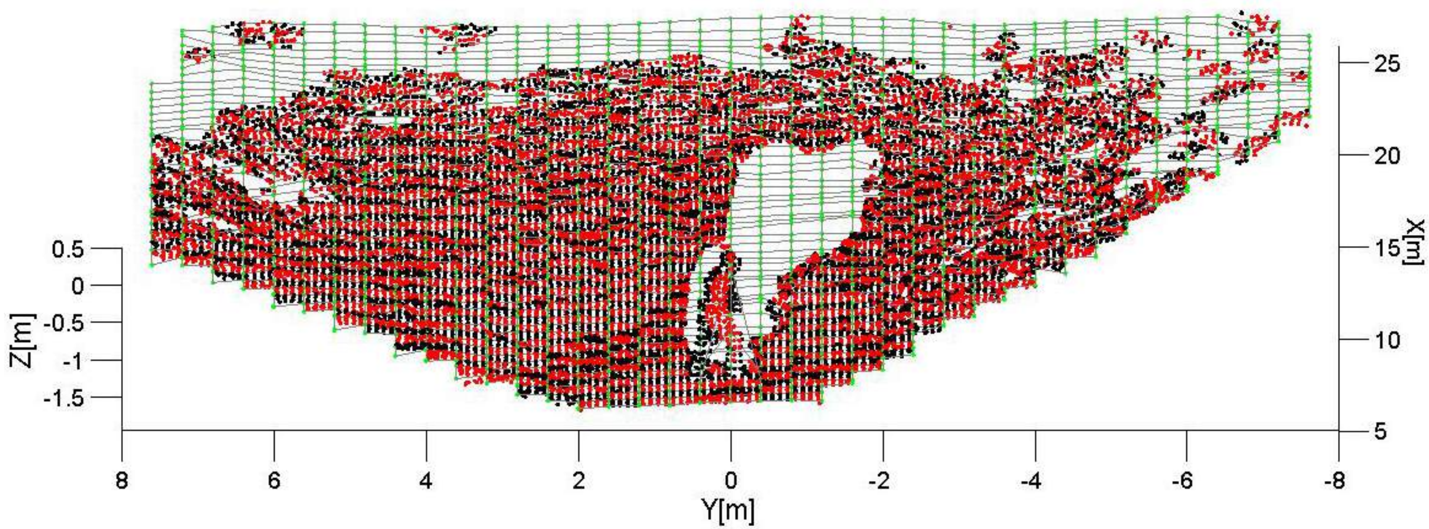
EXCELLABUST

Annalisa Milella · Giulio Reina
3D reconstruction and classification of natural environments by an autonomous vehicle using multi-baseline stereo, Intel Serv Robotics (2014) 7:79–92, DOI 10.1007/s11370-014-0146-x

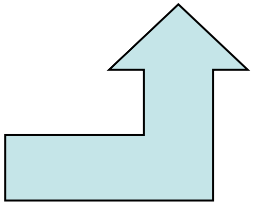
Geometry-based classifier

- The raw output is a cloud of range points.
- Points are divided into a grid of 40 cm x 40 cm terrain patches.





X-Y view

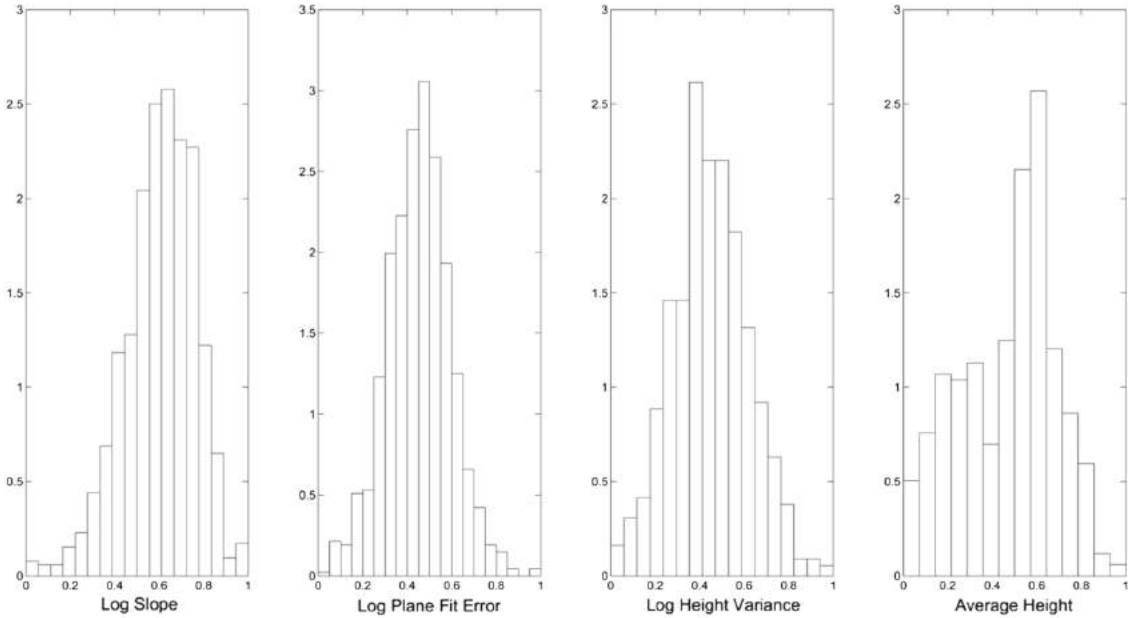


Geometric features

- Geometric features are statistics calculated from the elevation points associated with each terrain patch:
  1. *Average slope of terrain*: the angle between the least-squares-fit plane and the horizontal plane.
  2. *Goodness of fit*: the mean-squared deviation of the points from the fitted plane along its normal.
  3. *Variance* in the heights of the range points.
  4. *Mean* of the heights of the range points.

EXCELLABUST

Ground model



- All histograms exhibit an approximately **unimodal** distribution, which can be reasonably modeled with a Gaussian.

EXCELLABUST

Ground classification

- **One class-classification**
- **The pattern is an outlier, i.e., it is classified as a non-ground sample, if its squared Mahalanobis distance from the ground model is greater than a critical value:**

$$d^2 = (f - \mu)\Sigma^{-1}(f - \mu)^T$$

$$d^2 > \chi^2_{m;\alpha} \longrightarrow \text{Non-Ground}$$

$$d^2 \leq \chi^2_{m;\alpha} \longrightarrow \text{Ground}$$

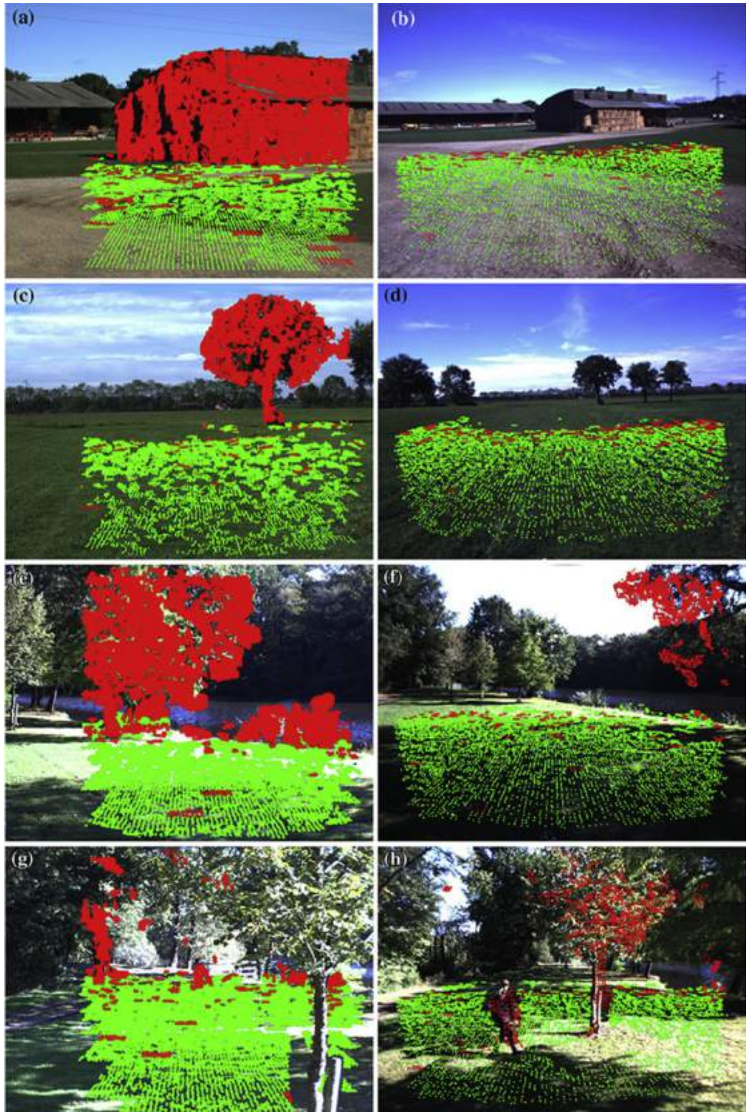
MULIVARIATE OUTLIER DETECTION APPROACH

NOTE: Assuming that the ground model vectors are independent and have a Gaussian distribution, it can be proved that the squared Mahalanobis distance is asymptotically distributed as the m degrees of freedom chi-square distribution.
Then, we can use the quantile α of the m degrees of freedom chi-square distribution as the delimiter (cutoff) for outlying observations.

Model Update

- **Problem**: A “static model”, built upon the initial properties of the ground, could soon fail or give poor results because of changes in ground properties during vehicle travel.
- **Proposed solution**: ***ROLLING WINDOW***
  - ***Initialization***: At the beginning of the robot’s operation, the training set is initialized under the assumption that the vehicle starts from an area free of obstacles, so that the stereo camera “looks” at ground only.
  - ***Update***: Then, the ground model is continuously updated as the vehicle moves: new ground feature vectors labeled in the most recent acquisitions are incorporated, replacing an equal number of the oldest ground instances.

Results



LIDAR-Stereo Combination

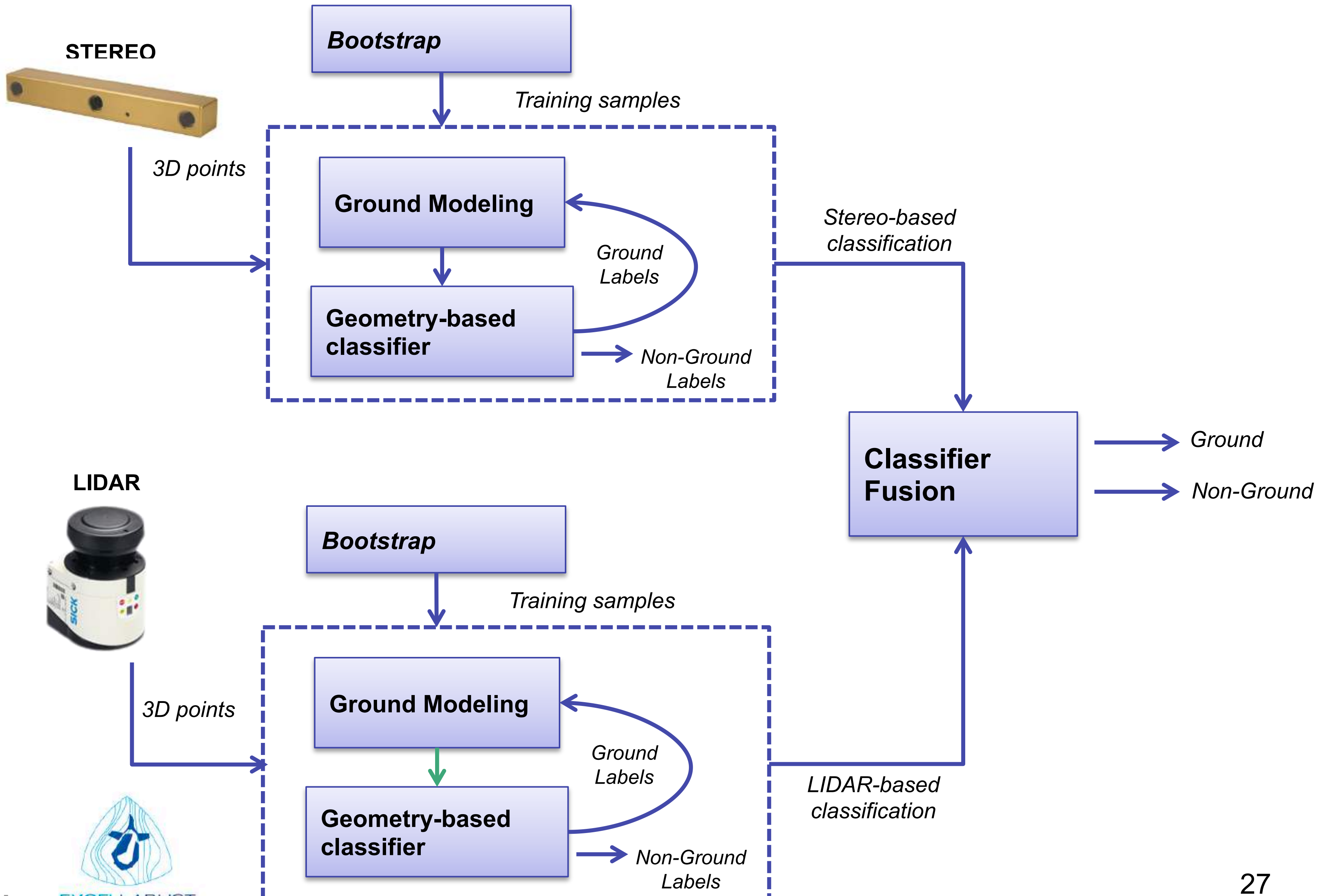
- LIDAR:
  - + is not affected by illumination conditions
  - + return dense and accurate 3D point clouds
  - – relatively low frequency (1 Hz or less) resulting in difficulties to capture dynamic obstacles
  - – fixed (non-scanning) types rely on accurate self-localization by the vehicle

- Stereo:
  - + high resolution in a suitable range of distances
  - + useful features for classification of different objects in the scene
  - – affected by lighting conditions
  - – the map can present holes due to low textures

Complementary characteristics – so combine them to get improved performance.

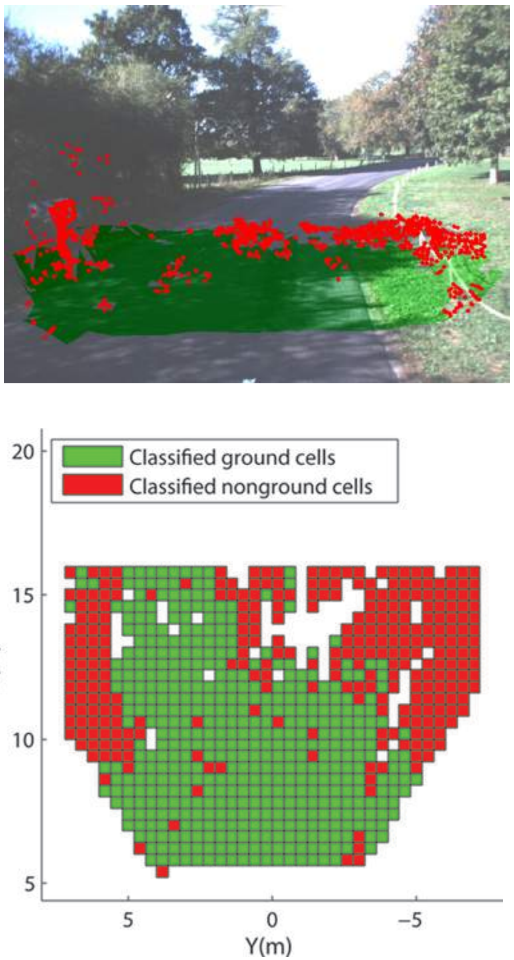
EXCELLABUST

Self-Learning Multi-Sensory Framework

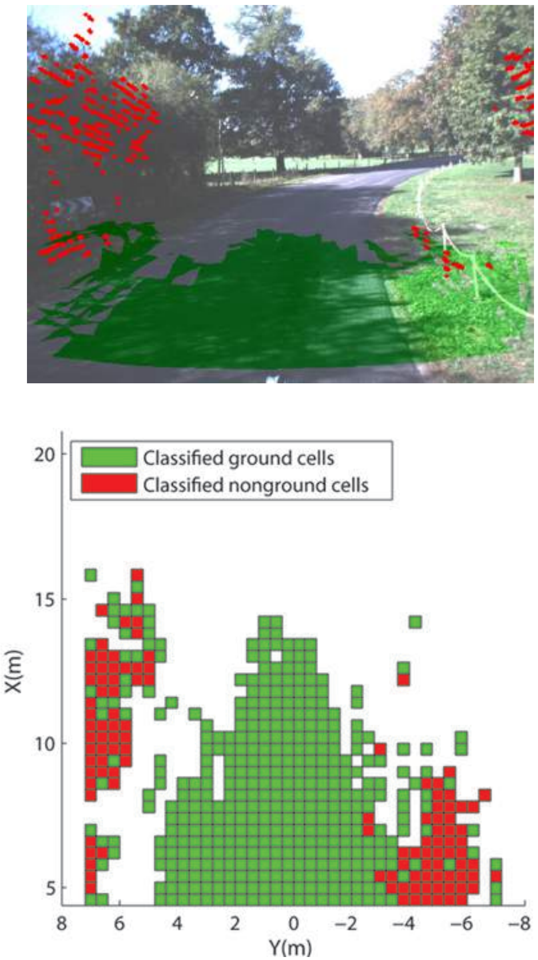


Results

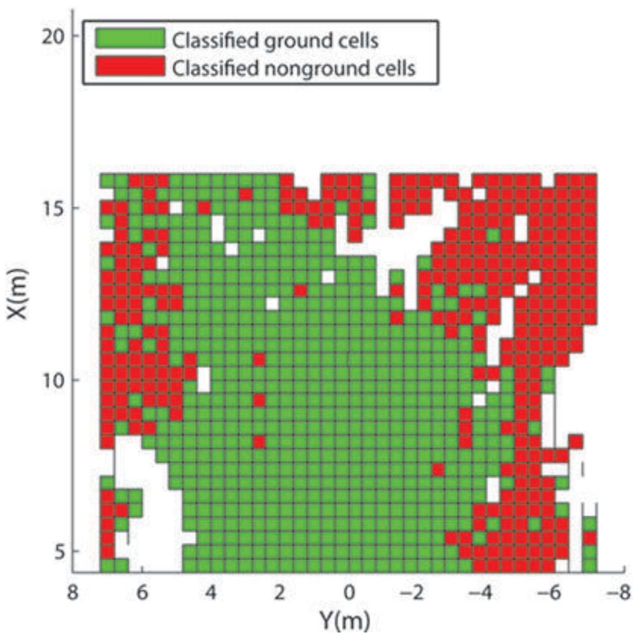
STEREO



LIDAR



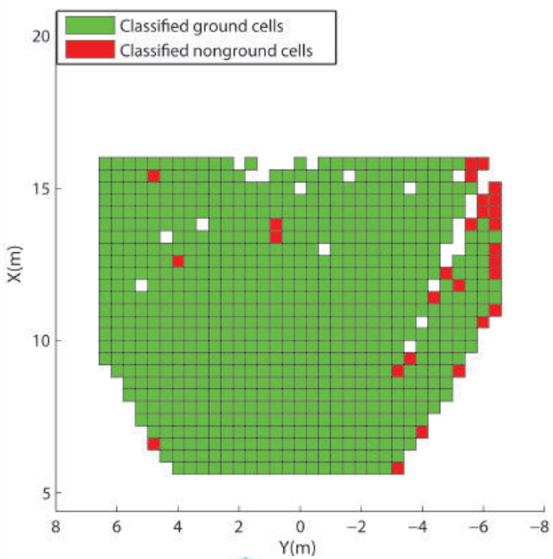
STEREO + LIDAR



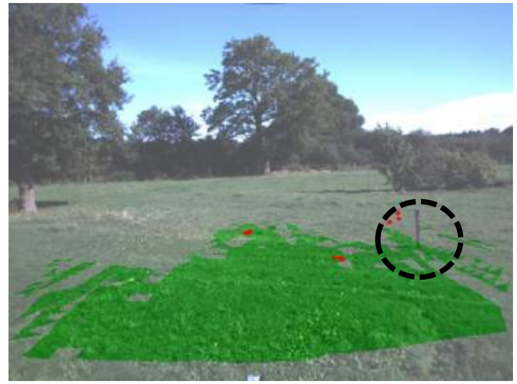
EXCELLABUST

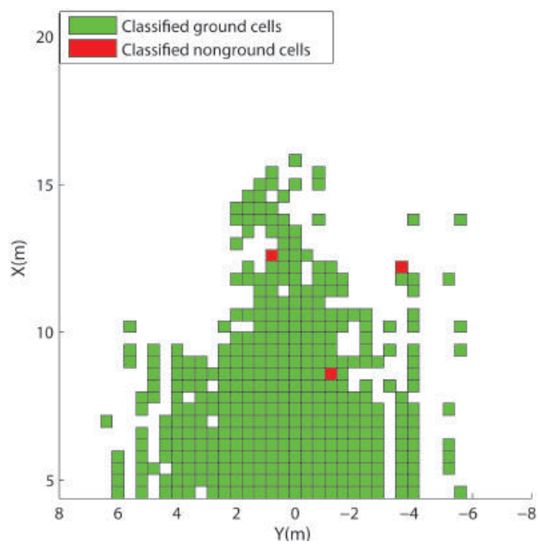
Results

STEREO

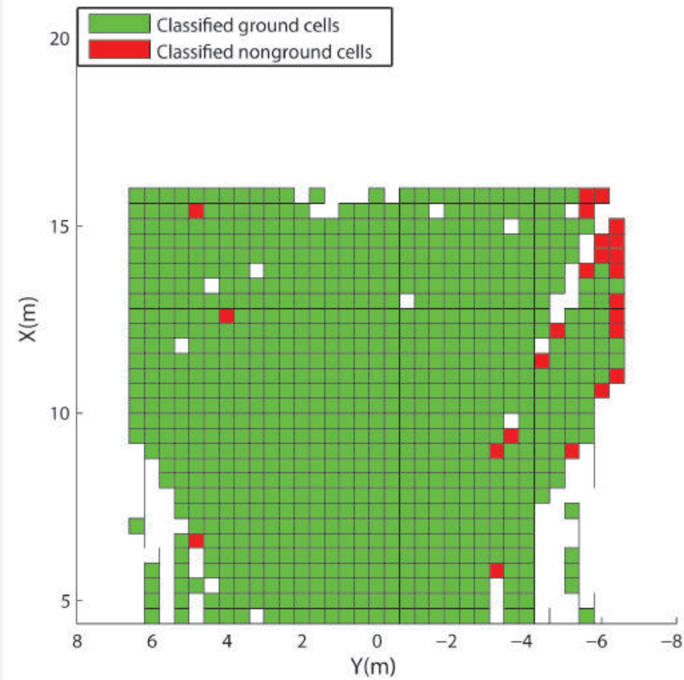


LIDAR





STEREO + LIDAR



EXCELLABUST

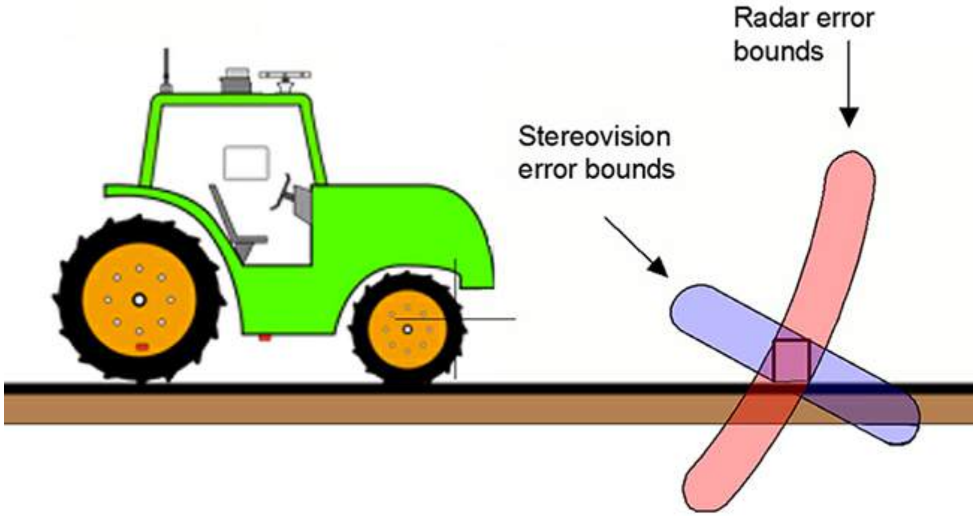
Results

Combining LIDAR and stereo imagery allows to:

1) ***widen the overall field of view*** of the perception system;
2) Improve the perception performance of the combined system due to complementarity of the two sensor modalities:
   - ***vision can help to overcome limitations of LIDAR***, such as sparseness of data and low acquisition frequency, by producing dense maps at relatively high frequency, being not affected by lighting conditions;
   - ***LIDAR can help to overcome limitations of vision***, such as reconstruction errors due to poor lighting conditions, shadows and lack of texture.

EXCELLABUST

Radar-stereo integration



OBJECTIVE:

combine the two sensors in order to improve 3D localization of obstacles up to about 30 m

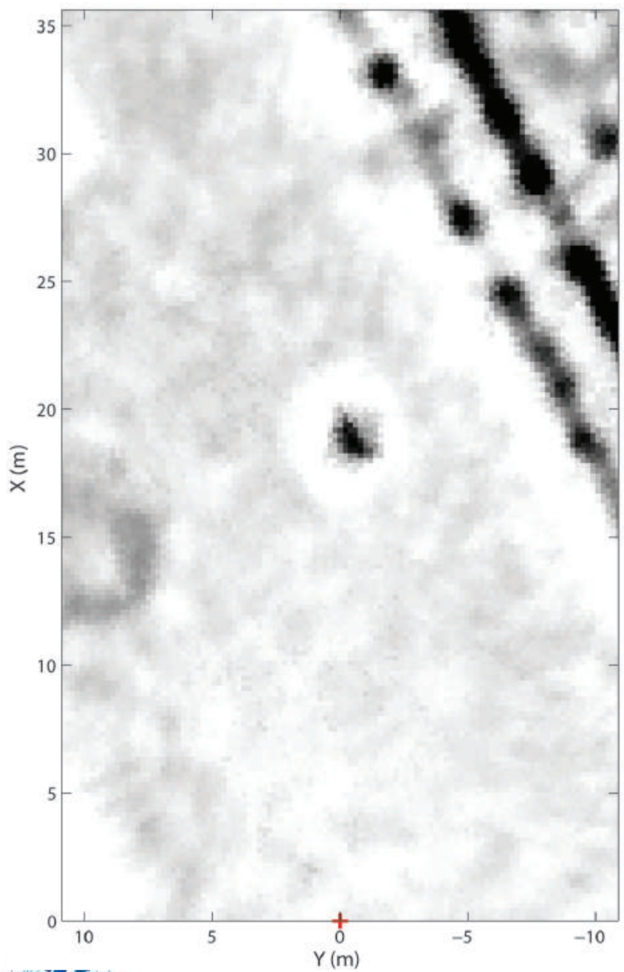
Radar:
- can accurately measure the distance of an object
- is robust to low visibility conditions (e.g., rain, fog, dust, and night-time)
- offers low vertical resolution

Stereovision:
- ensures good resolution to find the boundaries of an object
- is less accurate than radar in measuring range

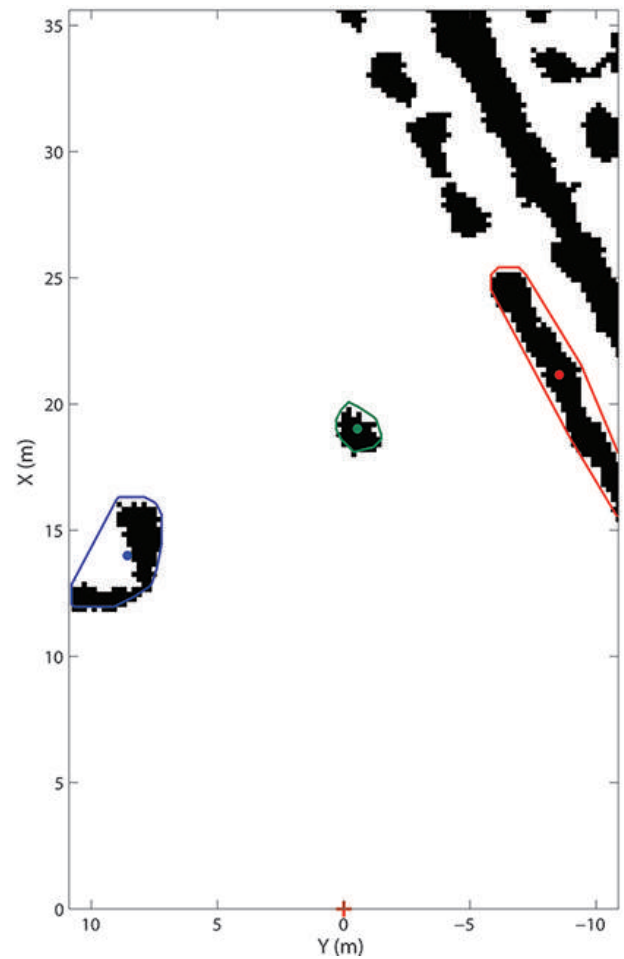
Radar-stereo based obstacle detection

- Two step- algorithm:
  - Radar-based multi-target 2D localization
  - Radar-stereovision combination for 3D target analysis
- Self-supervised scheme:
  - **2D radar range measurements** are used to define attention regions in the **stereo-generated point cloud** that can be then analyzed to augment the obstacle information content with **3D geometry** and **colour data**

Radar-based multi-target 2D localization



Raw radar image

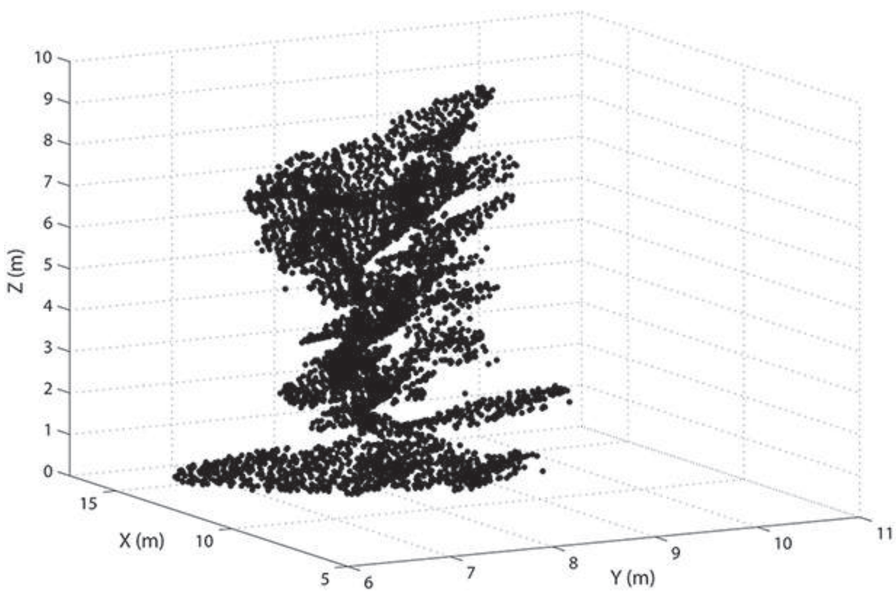


Obstacle localization by radar image CFAR binarization and centroid estimation

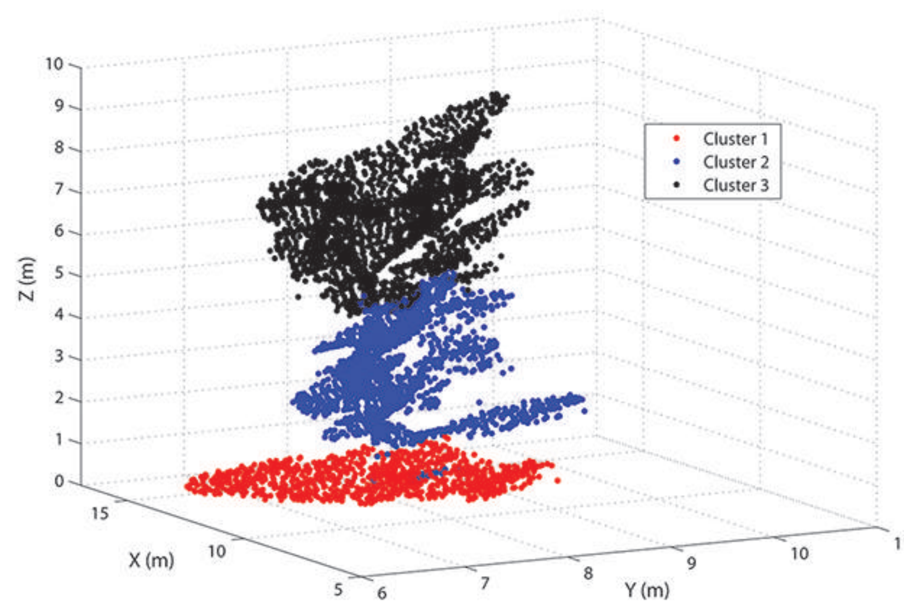
Radar-stereo combination for 3D target analysis

Step 1: Extraction of the sub-cloud

Radar-based obstacle detection drives extraction of 3D sub-cloud generated by stereovision



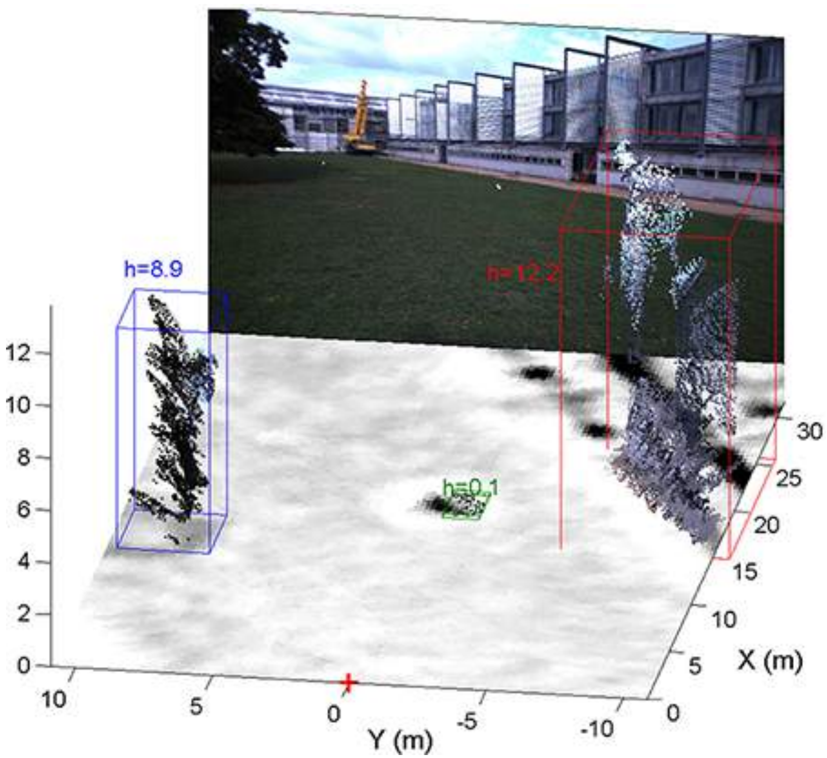
Gross sub-cloud



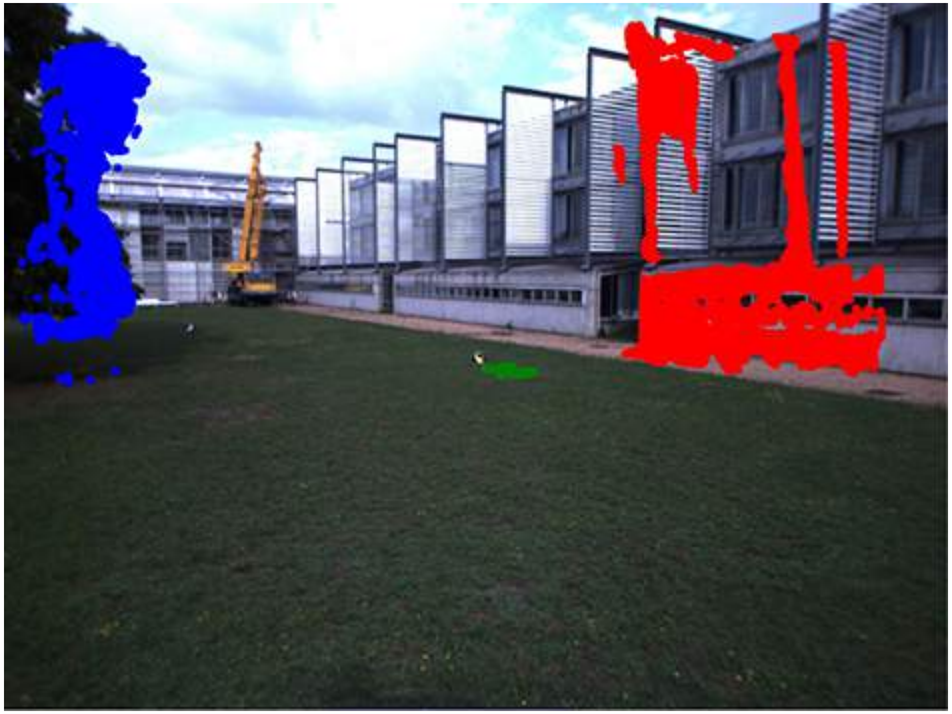
Sub-cloud segmentation based on MoG

Radar-stereo combination for 3D target analysis

Step 2: 3D obstacle characterization using stereo 3D and color information

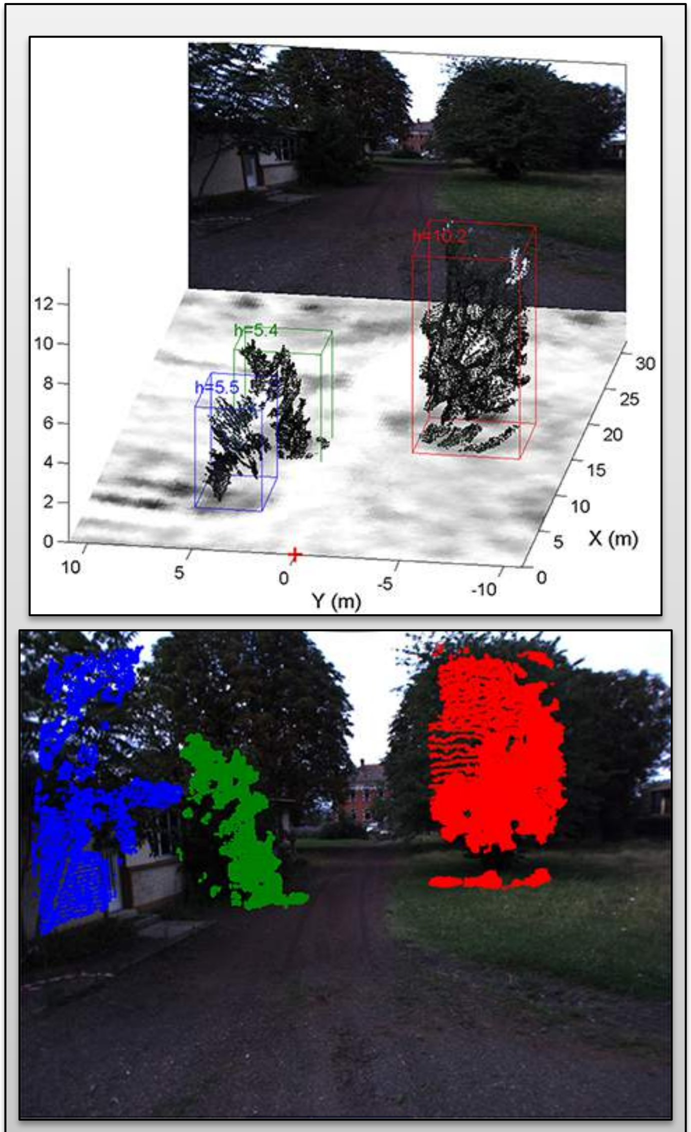


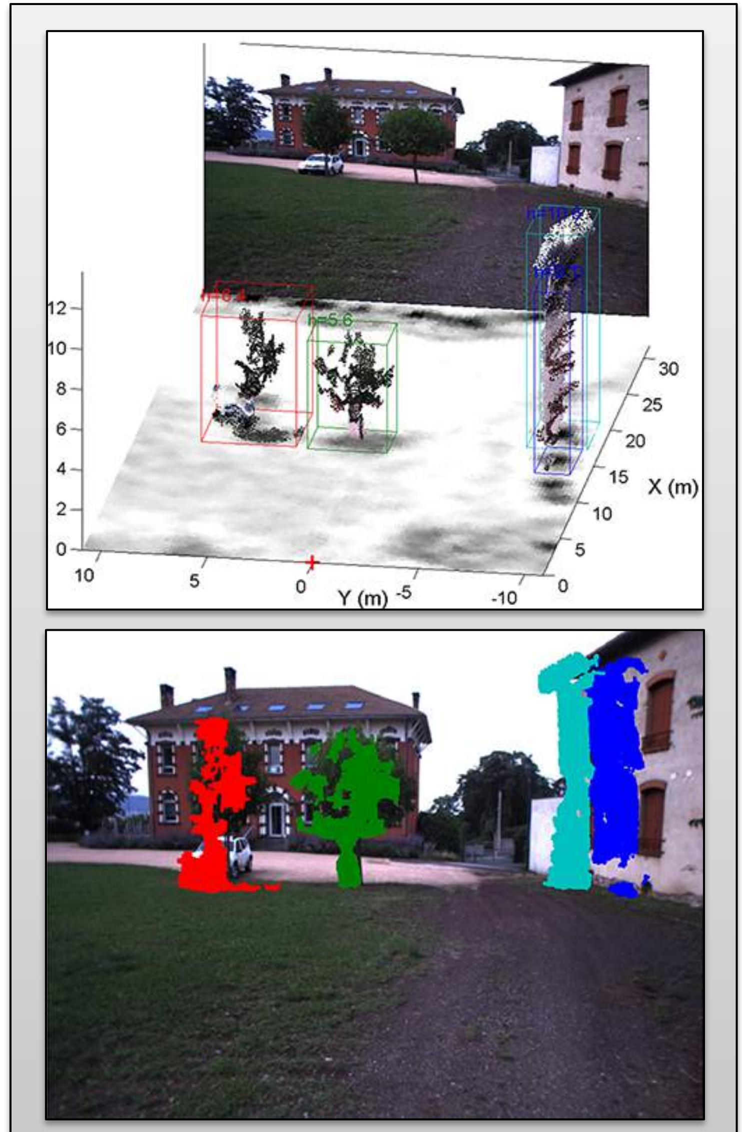
Stereovision reconstruction of radar-labeled obstacles



Sub-clouds associated to obstacles and projected over the original camera image

Results


h=5.4
h=5.5
h=10.2
12
10
8
6
4
2
0
10
5
0
-5
-10
Y (m)
30
25
20
15
10
5
0
X (m)


h=5.6
12
10
8
6
4
2
0
10
5
0
-5
-10
Y (m)
30
25
20
15
10
5
0
X (m)

From multi-sensor to cooperative perception

- The use of multiple sensors is mandatory to reach an adequate level of robustness
- Cooperation among multiple vehicles hold the potential to further improve the perception ability of each vehicle and reach an overall better knwoledge of the environment.
- **Cooperative perception can be defined as the collaboration between a fleet of robots for the estimation of the state of the environment by sharing information or developing cooperative actions.**

Cooperative perception

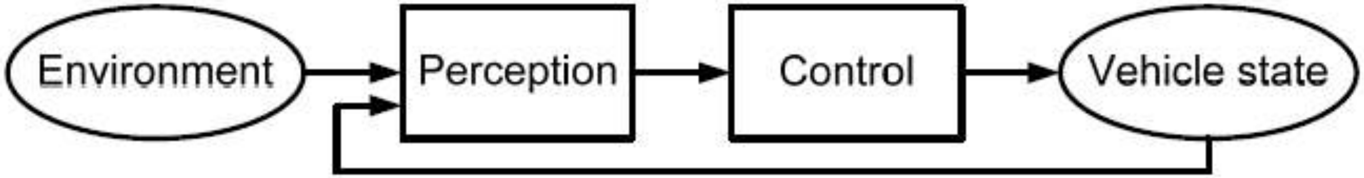


Figure 1.4 Intelligent vehicle operation paradigm

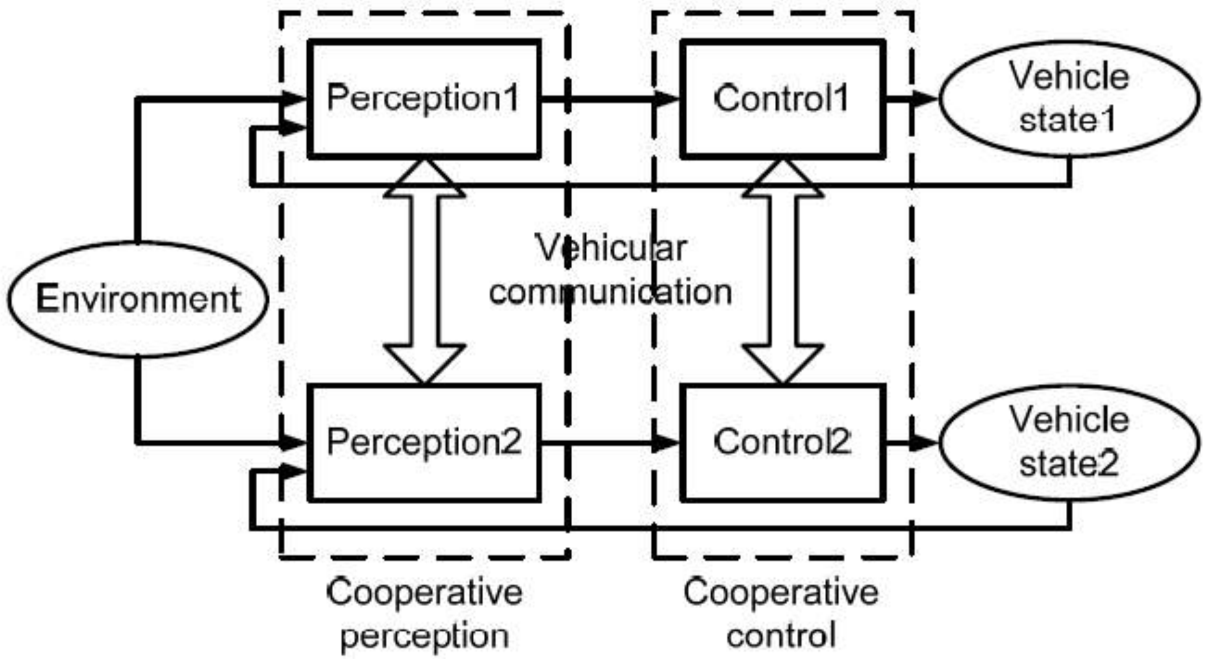


Figure 1.5 Vehicles cooperation paradigm

Hao Li. Cooperative perception : Application in the context of outdoor intelligent vehicle systems. Other. Ecole Nationale Superieure des Mines de Paris, 2012. English. <NNT : 2012ENMP0034>. <pastel-00766986>

EXCELLABUST

Why cooperative perception?

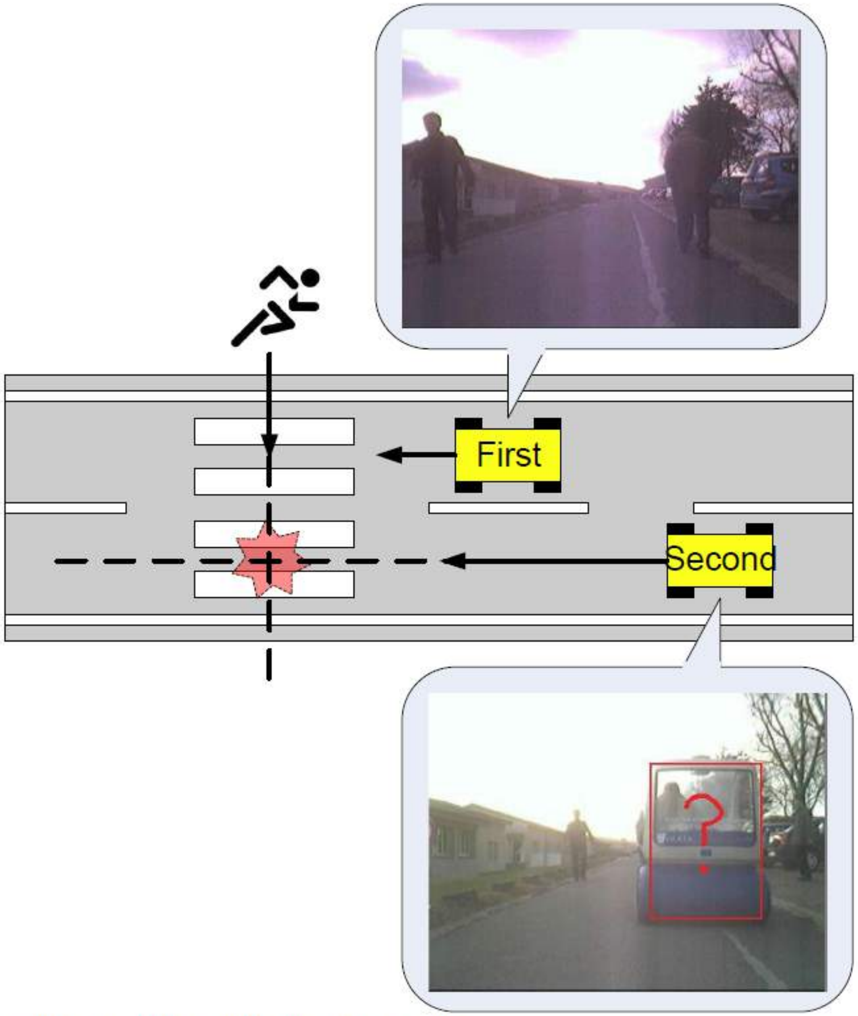


Figure 1.7 Overtaking scenario: potentially dangerous

Hao Li. Cooperative perception : Application in the context of outdoor intelligent vehicle systems. Other. Ecole Nationale Superieure des Mines de Paris, 2012. English. <NNT : 2012ENMP0034>. <pastel-00766986>

EXCELLABUST

Why cooperative perception?

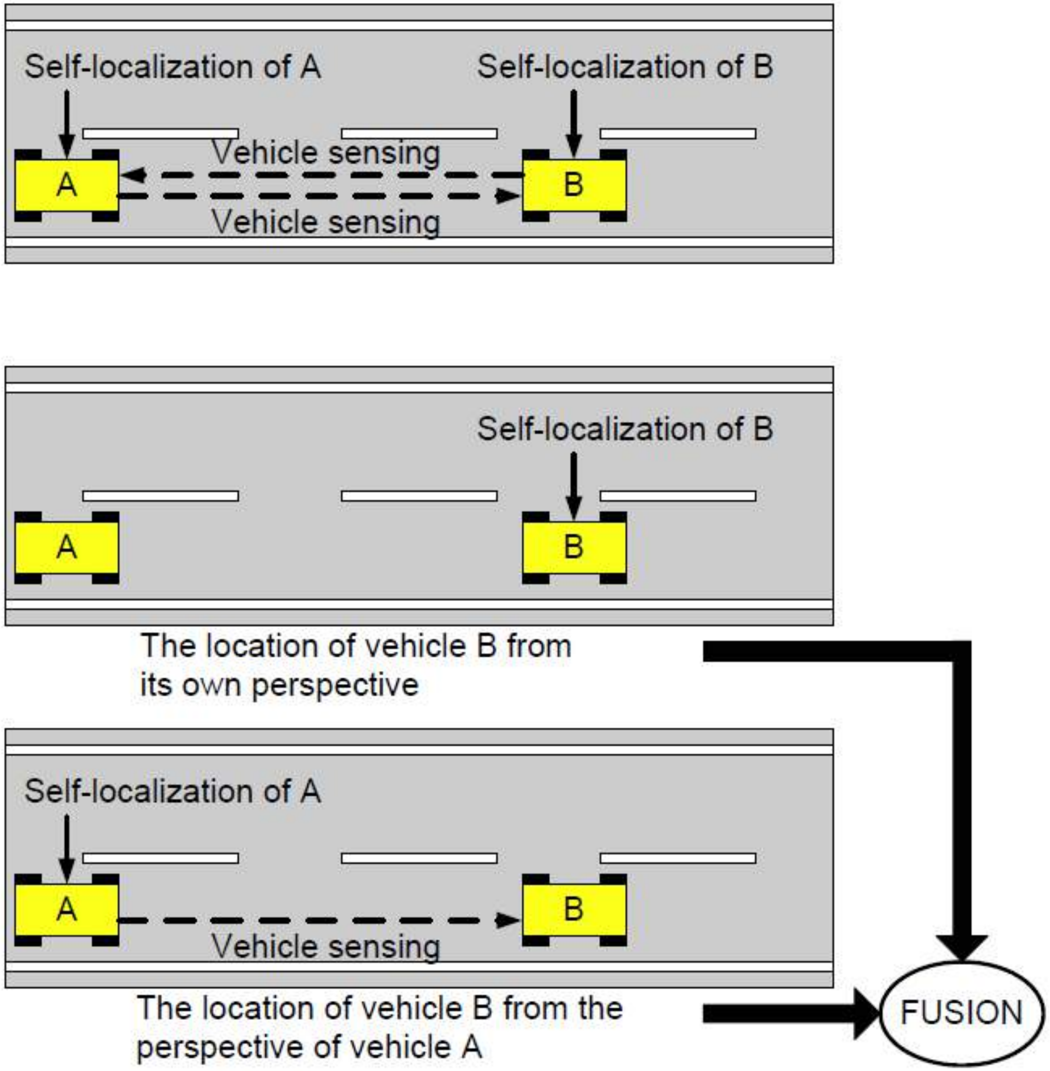


Figure 1.8 Multi-vehicles cooperative localization

Hao Li. Cooperative perception : Application in the context of outdoor intelligent vehicle systems. Other. Ecole Nationale Superieure des Mines de Paris, 2012. English. <NNT : 2012ENMP0034>. <pastel-00766986>

Air-ground cooperation

- Ground robots



Good at:
- ✓ Precise information gathering
- ✓ Physical intervention
- ✓ Long duration missions
- ✓ Heavy load carrying

Not so good at:
- ✓ Global information gathering
- ✓ Self localization
- ✓ High speed mobility
- ✓ Avoiding hazards

- Aerial robots


Good at:
- ✓ Global information gathering
- ✓ High speed mobility
- ✓ Avoiding hazards
- ✓ Communication relaying

Not so good at:
- ✓ Long duration missions
- ✓ Physical intervention
- ✓ Heavy load carrying

63

EXCELLABUST

Simon Lacroix, Guy Le Besnerais
“Issues in Cooperative Air/Ground Robotic Systems,” Chapter Robotics Research Volume 66 of the series Springer Tracts in Advanced Robotics pp 421-432

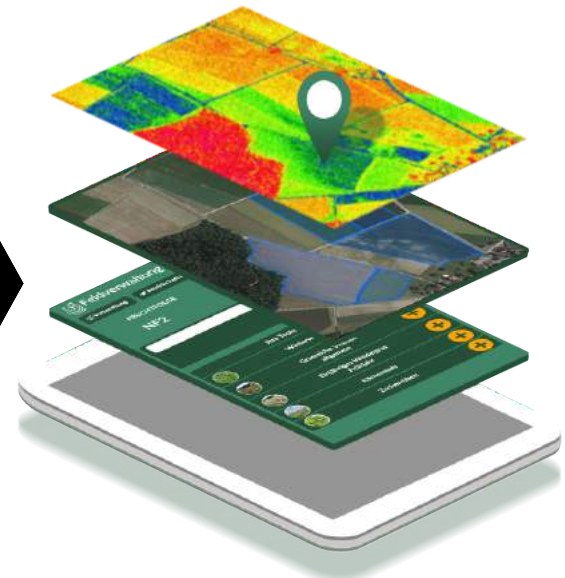
Cooperation schemes

- ***Aerial robots assist ground robots***. Aerial robots can provide the ground robots with information related to the environment (traversability maps, landmark maps). They can also straightforwardly localize the ground robots by perceiving them, or provide communication links with the remote operator station and between ground robots.
- ***Ground robots assist aerial robots***. Ground robots can assist the aerial robots by detecting cleared landing areas, or by providing them energy or the possibility to transport or recover them (as considered in [SKHR02, pro05] for instance).
- ***Ground an aerial robots cooperate to achieve a task***. Exploration and surveillance tasks can obviously be jointly achieved by teams of aerial and ground robots. Similarly, target detection and tracking tasks can benefit by the enhanced observation capacities brought by such teams – and more generally any combination of the two assistance schemes mentioned above can be foreseen.

Simon Lacroix, Guy Le Besnerais
“Issues in Cooperative Air/Ground Robotic Systems,” Chapter Robotics Research Volume 66 of the series Springer Tracts in Advanced Robotics pp 421-432

Air/ground cooperation for precision farming

Simultaneous Safety and Surveying for Collaborative Agricultural Vehicles (S3-CAV) FP7 ERA-NET ICT-AGRI-2



- The ground vehicle will carry **exteroceptive** sensors, which provide information about the geometry and the multichannel appearance of the environment. It will be also equipped with **proprioceptive** sensors to learn the local mechanical properties of the traversed terrain.
- The companion UAV will act as a “**flying eye**” for the ground vehicle and will be equipped with sensors allowing e.g. monitoring of crop growth, warning of potential danger e.g. prolific weed patches, or the identification of driving lanes within the field.

Air-Ground Localization and Map Augmentation

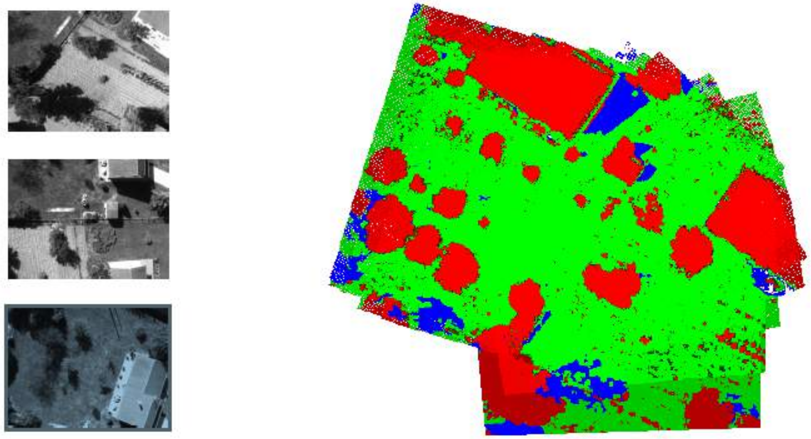


Fig. 1. Illustration of a traversability map built from a sequence of aerial monocular images [BLC06]. Left: some of the images processed, right: traversability map. Green, red and blue areas respectively correspond to traversable, obstacle and unknown ares.

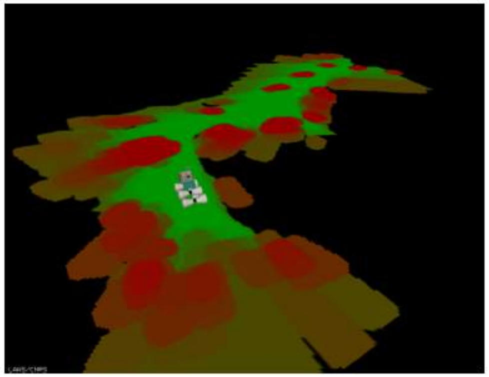


Fig. 2. A traversability map built by a ground rover as it navigates in the environment.

Simon Lacroix, Guy Le Besnerais
“Issues in Cooperative Air/Ground Robotic Systems,” Chapter Robotics Research Volume 66 of the series Springer Tracts in Advanced Robotics pp 421-432

- **Data types**: the acquired data can be images, either panchromatic or color, from which geometric 3D informations can be recovered, or directly 3D, as provided by a SAR or a Lidar for instance.
- **Data resolution**: the resolution of the gathered information can significantly change, depending on whether it has been acquired by a ground or an aerial sensor.
- **Uncertainties**: there can be several orders of magnitude of variation on data uncertainties between ground an aerial data.
- **Viewpoints changes**: beside the resolution and uncertainties properties of the sensors that can influence the detection of specific environment features, the difference of viewpoints between ground and aerial sensors generates occlusions that considerably change the effectively perceived area, and therefore the detectable features.

Air-Ground Localization and Map Augmentation

EXCELLABUST