# DR 1.2: Sensing, mapping and low-level memory II

Tomáš Svoboda*, Abel Gawel†, Dong-Uck Kong‡, Renaud Dubé†, Mario Gianni§, and the TRADR consortium

*CTU in Prague, Czech Republic

†ETH, Zürich, Switzerland

‡Fraunhofer IAIS, Sankt Augustin, Germany

§Alcor Laboratory, Department of Computer, Control, and Management Engineering "Antonio Ruberti" - Sapienza University of Rome.

⟨svobodat@fel.cvut.cz⟩



| | |
|---|---|
| *Project, project Id:* | EU FP7 TRADR / ICT-60963 |
| *Project start date:* | Nov 1 2013 (50 months) |
| *Due date of deliverable:* | Month 26 |
| *Actual submission date:* | February 15, 2016 |
| *Lead partner:* | CTU |
| *Revision:* | final |
| *Dissemination level:* | PU |



---

We report progress achieved in Year 2 of the TRADR project in WP1: *Persistent models for perception.* It describes the essential robot (UGV) perception functionalities and a new method for creating metrical maps.

---

## Executive Summary

The key objective of WP1 is to provide sensory data from all involved robots registered in space and time, to keep creating and updating robot centric representations, and ground them into the world coordinate frame. The obtained representations are furnished to other WPs, which maintain higher level situation awareness.

During Year 2 we concentrated on addressing challenges caused by dynamic elements of the scenes, e.g. varying presence of dense smoke which effectively blinds exteroceptive sensing (cameras and lidar). We focused on developing a new 3D mapping and map merging aiming at multi-sortie, multi-robot mapping persistence. We designed new algorithms for terrain perception in sensory deprived environments. We enrich the robot flippers by adding force sensors and design an algorithm for computing terrain profile from interoceptive measurements only. In collaboration with WP4, we also worked in a sensorless alternative for terrain contact sensing based on the residual dynamics of the flippers. For creating metric 3D maps we opted for the Continuous Trajectory Scan Matching (pose-graph SLAM variant) where the robot trajectory is the state of the map. Therewith exchanging updates on the map between robots, i.e. updating the nodes of the trajectories and co-registrations of nodes can be achieved efficiently with low communication bandwidth. The map merging is based on a flexible framework that permits to select which section of the trajectory one wishes to optimize and is capable of co-optimizing trajectories resulting from different sorties.

## Role of robot perception and metrical mapping in TRADR

The robot perception means the robot is able to analyze its neighborhood and act accordingly. Terrain recognition is essential for robot locomotion regardless whether the robot is teleoperated or moves autonomously. It is desirable the robot overcomes obstacles in a reasonable way - fast, safe, consuming less power and reducing cognitive load of a human operator. Automatic victim detection is important for many search and rescue scenarios. A human operator may provide final decision however, robots, when crawling through a disaster site should provide warning about possible victim locations.

The metrical mapping serves as the very basis for modeling the world. It is also the basis for sharing information between robots and also among several sorties and even missions.

# Contribution to the TRADR scenarios and prototypes

The new algorithms for sensing in sensory deprived environments together with related hardware advancements contribute to the models for acting (WP2), multi-robot collaboration (WP4) and also support the human-robot teaming (WP5). The advancements in algorithms that address the failure of exteroceptive sensing (Lidar, cameras) were significantly motivated by the end-user evaluations and we plan to deploy them in the Y3 of the project.

Based on the feedback from the end-user evaluations, we implemented several software packages and hardware upgrades that ease the deployment of the UGVs, e.g. warm restart of motor drivers.

# Persistence

Persistence in WP1 is addressed mainly by re-using the data in creating an enviroment model. The 3D metric map serves as the main basis for multi-modal (data), multi-source (robots), multi-level (abstraction, decisions) registration. In WP1, we are working on robust methods for merging partial 3D maps. The merging challenges include weak data overlap, dynamic changes in scenes, large displacement of local coordinate systems. The terrain perception and robot control algorithms use machine learning techniques in a quest of gaining experience from operator-robot interactions.

# 1 Tasks, objectives, results

## 1.1 Planned work

In Year2, WP1 planned to investigate "Sensing, mapping and low-level memory II Dynamic scene" (Milestone MS1.2). The work was divided into two tasks:

- Essential sensing and UGV control functionality (T1.3)
- Robot centric metrical maps and models storage (T1.4)

Both tasks emphasize a dynamic aspect of scenes. The plan for Task T1.3 was to improve low level sensing capabilities of UGV and enhance interconnection with WP4. The goal of Task T1.4 was to develop an algorithm for computing one single metrical map from possibly highly incomplete partial maps in dynamic environments. One of the main challenges expected was merging different modalities.

## 1.2 Addressing reviewers' comments

> The adaptive traversability for the UGV considers in the future also different environmental conditions, e.g. wet floor due to rain or snow, so as to adapt to different friction coefficients.

The adaptive traversability controls the robot morphology which is far more important for traversing obstacles then reacting to friction changes. The friction changes are currently not a major issue. At the moment a few rules are hand-crafted - like going up or down a staircase where stability and friction are of high importance.

> Traversability information can be integrated into the display and user controls, as the user needs to know how much time it will take a robot to move, and whether it is possible.

We will tackle this issue until the next TRADR Joint Excercise TJEx. We plan visualizing information helping the operator drive the robot in UGV Operator Control Unit (OCU) (robot roll/pitch, flipper torque visualization, visualization when the adaptive traversability needs more data or considers the situation unsafe). We still do not estimate the time-to-traverse as the actual speed/direction control is in operator's hands. Traversability analysis, performed on the 3D metric information coming from the robot laser sensor, can be exploited to inform the user about which regions of the environment can be traversable or not by the robot. The result of this analysis is provided to the planning algorithm for computing affordable paths towards the goal. On the other hand, this result can be provided to the user, in the form of colored point cloud, via the OCU. According to a predefined color range, the user can identify traversable areas and instruct the robot toward that directions. We are currently experimenting with this.

> Considerable thought should go into how to allow "just drive" on one hand, while still "pushing" the user towards easier terrains.

Pushing the user toward easier terrain is possible when the robot knows the terrain beyond its current view/measurement. This is more the role of the planning. The AT is applied in the situations when the robot "knows" only what it directly measures, i.e. it can work during the eploration when no complete 3D model is available. This point is closely connected to the visualisation plans, see above.

> In future demonstrations or reviews, it would be useful to see quantitative comparison to non-adaptive tests.

We performed some comparison experiments in [64] (Annex Overview 2.1). Generally, the traversal time is not always better with AT (and is not intended to be), but the cognitive load should be essentially lower. We think about setting up a task where the operator should drive the robot through a demanding terrain and look for objects of interest at the same time. With AT, he should be able to detect more objects than without AT.

> Role of persistence in the learned results (or vice versa, role of learning in promoting persistence) is still in early stages, and should come into play in the coming years of the project.

The main idea is that the AT is learns versatile models through extensive training in varied terrains and simulations. Learning AT during mission may be not the best idea because it is not sure the operator did drive the robot in the desired manner. Also, most situations are already covered by training data. What we could do is to detect situations far away from anything in the training data, remembering their location, and then manually traversing them to show what the ideal traversal should look like.

> Anomaly detection: clarify what is meant, put into context of related work.

The notion of anomaly has changed in Y2. In Y1, the anomaly was more about detection outliers in data. In Y2, the anomaly is understood as a failure of exteroceptive sensing, i.e. situations when Lidar and cameras produce outliers only, no usable measurement. At the moment, the AT starts arm/flipper active sensing when there is not enough probability of achieving safe state with any flipper configuration – typically because of missing data. We detail this issue in this deliverable.

> The victim detection classification learning methods should work well even with a subset of operational sensors.

We are now testing how far can we go with RGB data only - which is perhaps the prevailing case. On the other side of the problem, we are employing a thermal cam on a pan-tilt unit and developing active search algorithm.

> The victim detection module, because of the high false-result bias, may benefit from proposing candidates to the operator, which it can quickly rule in or out.

The 2D detections are fused in a 3D map. Collection of relevant snapshots may be displayed to an operator who can discard the relavant 3D area. However, we did not work out the human-in-the loop yet. This is still an open issue (from the implementation/integration) point of view.

> Globally persistent SLAM: key test would be in supporting, and being supported by, persistence: multiple runs, multiple robots.

The basic multi-sortie mapping is detailed in this deliverable.

> In WP1 and WP4, special attention should be given in the future to communication contention methods when scaling to larger teams of robots, e.g. in 3D mapping (CTSM), multi-robot collaboration, etc.

In Continuous Trajectory Scan Matching (CTSM) the robot trajectory is the state of the map. Therewith exchanging updates on the map between robots, i.e. updating the nodes of the trajectories and co-registrations of nodes can be achieved efficiently with low communication bandwidth. The higher bandwidth exchange of metrical map information is presently only done in offline batch processing of maps between sorties. Furthermore, the state trajectories represent valid paths of the robots that update over the course of time. Therewith they build a sparse connected traversability graph of the environment which can be facilitated for the planning.

## 1.3 Actual work performed

### 1.3.1 Task T1.3 (Essential sensing and UGV control functionality II – Scene part and object recognition, dynamic scene)

**Terrain perception in sensory deprived environments** We propose a combined hardware and software solution to predict the profile of terrain underneath and in front of the tracked robot. The algorithm exploits a prototype of a force sensor array installed inside a track of the robot, a robotic arm attached to the robot, proprioceptive measurements from joints and an inertial measurement unit (IMU), and information learned from a dataset of traversed terrains. The prototype of the force sensor (Fig. 2, is suitable for tracked robots and is installed between rubber track and its support, allowing it to serve as a tactile sensor. The arm is used to measure

Figure 1: From left: UGV robot approaches smoke area; Example of visual information that the operator sees inside a cloud of smoke: a crop out from the omni-directional camera (middle) and output of the laser range-finder (rainbow-colored point cloud in the right half of the image). Laser beams are randomly reflected by smoke particles. The resulting 3D point cloud is just noise close to the robot.

height of terrain outside the reach of the force sensor as contact between the arm end-effector and the terrain. To obtain well-defined contact points with the ground, we decided to take advantage of the flippers that can reach in front of the robot and are designed to operate on dirty surfaces or sharp edges. The original mechatronics of the robot allows to measure torque in flipper servos and thus detect physical contact between flippers and the environment. To be able to locate the contact point on the flipper exactly, we designed a thin force sensor between the rubber track and its plastic support (see Fig. 2). The sensor construction is a sandwich of two thin stripes of steel with *FSR 402* sensing elements between them which allows the rubber track to slide over it while measuring forces applied onto the track. There are six force sensing elements; the protecting sheet of steel distributes the force among them, the sensor is thus sensitive along its whole length. Figure 3 shows three examples of the sensor readings. The first case consists of a flipper touching flat floor. Although one would expect to see more or less equal distribution of the contact force along the flipper track, the torque generated by the flipper actually lifts the robot slightly and thus, most of the force concentrates at its tip (element n. 6). Compare this case with the third one (bottom), where the pose of the robot prohibits the lifting effect, and we therefore see the expected result. The second case (middle) shows an example of a touch in one isolated point.

We propose a novel active tactile exploration mode (ATEM), in which flippers and robotic arm autonomously explore the terrain shape in close vicinity of the robot. Estimated terrain shape and expected reconstruction accuracy are eventually displayed to the operator, at the moment in a form of simplified 2D graph. If ATEM is requested by the operator, the robot first

Figure 2: Prototype of the flipper force sensor: array of six sensing elements (FSR 402) is covered by a stripe of steel, forming a thin sensor that fits between the rubber track and the plastic track support. The stripe of steel protects the sensors from the moving rubber track and distributes measured force among them. The sensor mounted to the plastic track support (top). The sensing elements are passive sensors that exhibit decrease in resistance with applied force. For each sensing element, we use a reference resistor to form a voltage divider and an analog-to-digital converter expansion board for the Raspberry Pi computer to read the six voltages.

adjusts its flippers to press against the terrain and capture proprioceptive measurements. Then the initial probabilistic reconstruction of the underlying terrain shape is estimated from the captured data. If the reconstruction is ambiguous, the robotic arm explores the terrain height in the most inaccurate place. Eventually, the probabilistic reconstruction is repeated. As a result, reconstructed terrain shape with estimated variances is provided. More details can be found in [77] (Annex Overview 2.3).

**Controlling robot morphology from incomplete measurement** In Year 2 of the TRADR project we have extended and improved the adaptive traversability (AT) pipeline introduced in [83, 84] (see Figure 5 for an overview) in several ways. We introduce a safety measure which allows to invoke tactile exploration of non-visible terrain if needed. We develope several strategies for the tactile exploration with a body-mounted robotic arm. We suggest a regression forest based $Q$-function representation which allows easier marginalization over missing data.

The output of $Q$-learning is a $Q$-function $Q(c, \mathbf{x})$, which assigns the suitability value $q$ of a possible flipper configuration $c$ to the current state in terms of sums of discounted rewards, see [83] for detailed description. The state is described by a feature vector $\mathbf{x}$ computed from the proprioceptive (IMU, wheel odometry, ...) and exteroceptive data (Lidar, cameras ...). Instead of using only the expected value of $q$, we model the full $q$-value probability distribution function $p(q|c, \mathbf{x})$ (which will be further referred to as QPDF). Given this probability distribution and full feature vector $\mathbf{x}$, the

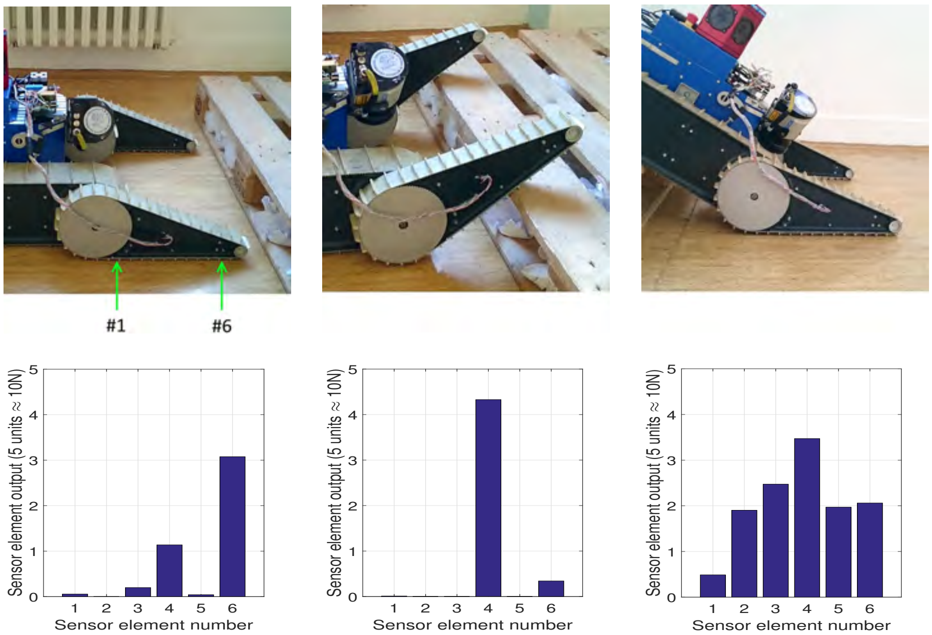


Figure 3: Examples of the force sensor readings. The top photos document the moments of the readings acquisition. The bottom line plots show raw readings of each sensing element, only corrected for bias.

optimal flipper configuration $c^*$ is:

$$\begin{aligned} c^* &= \operatorname*{argmax}_c Q(c, \mathbf{x}) = \operatorname*{argmax}_c E(Q \mid c, \mathbf{x}) = \\ &= \operatorname*{argmax}_c \int q \cdot p(q|c, \mathbf{x})\, dq \end{aligned} \tag{1}$$

There are two reasons for modeling the full QPDF: (i) measuring the safety of given flipper configurations and (ii) marginalization when only incomplete measurements $\mathbf{x}$ are available. Two different QPDF models: (i) Gaussian Processes and (ii) Regression Forests, are discussed in [84].

While proprioceptive data are usually fully available, the exteroceptive data are often incorrect or incomplete. This occurs in case of reflective surfaces such as water or in presence of smoke. We denote missing part of measurements as $\overline{\mathbf{x}}$, and available measurements as $\widetilde{\mathbf{x}}$, i.e. $\mathbf{x} = [\overline{\mathbf{x}}, \widetilde{\mathbf{x}}]$. In the case that $\overline{\mathbf{x}}$ is not empty, we marginalize $p(q|c, \mathbf{x})$ over the missing data $\overline{\mathbf{x}}$ to estimate $p(q|c, \widetilde{\mathbf{x}})$. Given the marginalized distribution and measurement $\widetilde{\mathbf{x}}$, the optimal flipper configuration $c^*$ is estimated as follows:

$$c^* = \operatorname*{argmax}_c \int q \cdot p(q \mid c, \widetilde{\mathbf{x}})\, dq. \tag{2}$$

However, the more features are missing, the higher the scatter of achievable $q$-values. Any state-action pair yielding a negative $q$-value, considering our

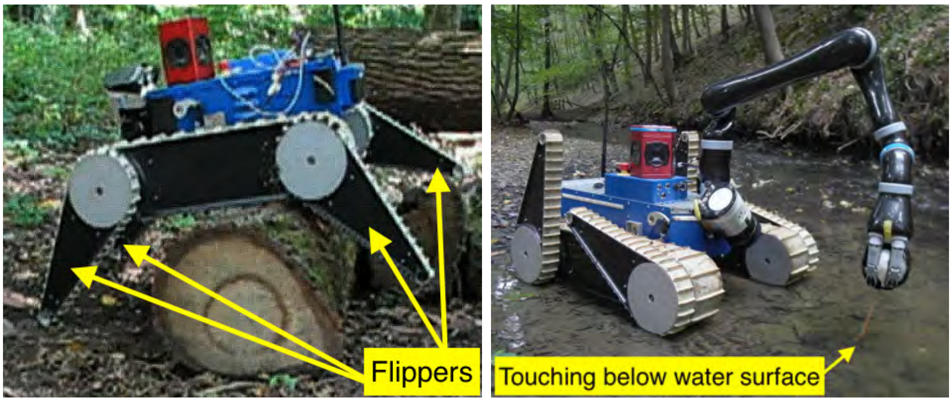


Figure 4: **Left:** Controlling robot morphology (flippers) allows for traversing obstacles. **Right:** Robotic arm inspects terrain below water surface compensating thus incomplete lidar measurement.


Figure 5: **Principle overview:** individual blocks in this scheme correspond to Sections IV-VII.

definition of the reward function.

$$S(c, \widetilde{\mathbf{x}}) = \int_{0}^{\infty} p(q \mid c, \widetilde{\mathbf{x}})\, dq, \tag{3}$$

that corresponds to the probability of achieving a safe state ($q \geq 0$) with the configuration $c$. Search for the optimal configuration $c^*$ (Equation 2) is restricted to the safe configurations only:

$$S(c, \widetilde{\mathbf{x}}) > \epsilon. \tag{4}$$

If none of the available configurations satisfies the safety condition (Equation 4), we use the robotic arm to evaluate selected missing terrain features; see Figure 5 for the pipeline overview. We propose several strategies that guide the active exploration of missing features in order to achieve a safe configuration as fast as possible. If all terrain features have already been

Table 1: Description of the states, actions and rewards

| | | |
|---|---|---|
| State | $\mathbf{x} \in \mathsf{R}^n$ | DEM, speed, roll, pitch, flipper angles, compliance, currents in flippers, actual flipper configuration |
| Actions | $c \in \mathbf{C} = \{1 \dots 5\}$ | 5 different, pre-set but self-adjustable flipper configurations [83]. |
| Reward | $r(c, \mathbf{x}) : \mathbf{C} \times \mathsf{R}^n \to \mathsf{R}$ | $\alpha \times$ user reward $+ \beta \times$ pitch penalty $+ \gamma \times$ roughness penalty |

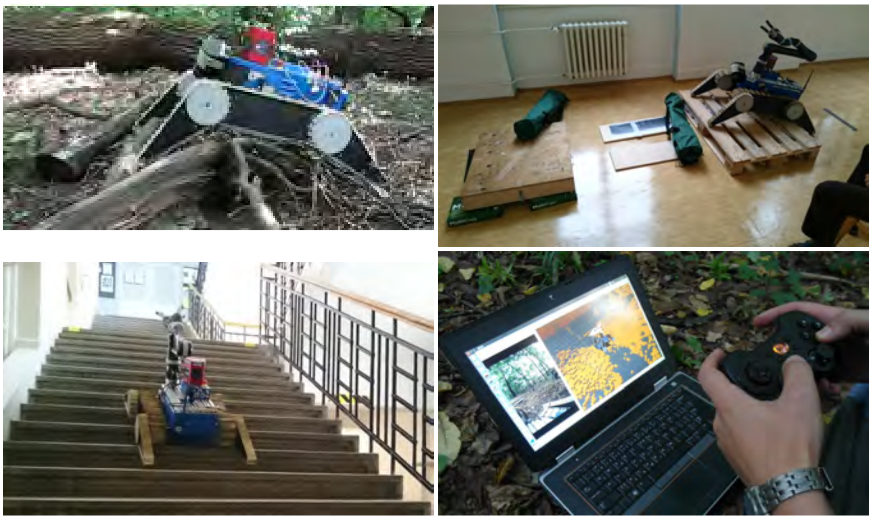

Figure 6: Top left: *Forest* obstacle. Top right: *Rubble* obstacle. Bottom left: *Stairs* obstacle. Bottom right: Operator controlling the robot using only sensor data.

measured and there is still no configuration satisfying the safety condition, then manual flipper control is requested from the operator. More details in [64] (Annex Overview 2.1).

**Victim detection** We have investigated Fully Convolutional Networks [52] for the task of victim segmentation in multimodal images, as deep networks can supposedly provide higher performance compared to boosted decision trees, with the trade-off of higher computational demands. A model from [52] was fine-tuned on our semisynthetic multimodal images of humans in arbitrary poses (i.e., victims), using an early-fusion scheme to add the depth and temperature modalities. We have observed that the information contained in these additional modalities cannot be fully utilized using the

Figure 7: Victim segmentation examples.

early-fusion scheme (without special encoding) and it provides only marginal performance gain despite the overall large performance gain achieved in fine-tuning. For depth information, this is consistent with previous observations reported in [34], where a larger performance gain is achieved by using a special depth encoding; a similar preprocessing may also be needed for the temperature. An example segmentation from a fine-tuned model using all the modalities is shown in Figure 7.

**Terrain classification using acoustic features** We explored sound as a complementary perception modality to classify terrain types with respect to the ground surfaces on which the UGV is moving. We collected audio data with the TRADR UGV platform driving on five basic human-classified terrain types: grass, pavement, gravel, carpet and sand, including different conditions, such as dry or wet gravel , grass with or without leaves, smooth or rough asphalt pavement or tiled pavement. We used several state-of-the-art acoustic feature extraction techniques and trained different types of classifiers. Average accuracy obtained on testing audio samples not used for training reached 93.5 % for samples of 1second duration, and 95.6% for 4 second samples. Most problematics was the classification of gravel, which was, not surprisingly confused with sand or pavement. These are promising results indicating that acoustic features are useful for terrain perception of a tracked robot. Details of the work done, the experiments and the results are presented in [73] (Annex Overview 2.7).

**Metric mapping** Next, we describe the advances within the mapping systems for the TRADR robots. The contributions address specific functionalities in the context of TRADR, such as map merging, efficient node sampling and multi-modality. A description of the general SLAM framework is given in [D 2.2]. Techniques for performing change detection are still under development and results cannot yet be reported at the time of this submission (January 2016). In this deliverable, we provide an overview of the current design phase of a change detection module aimed at providing this feature.

**Advancements in simultaneous robot localization and mapping: Non-uniform sampling of trajectory control points over a sliding window** In order to enable recognition of scene parts and objects of interest, a point cloud representation which faithfully represents the robot's environment is required. For achieving such quality representation in real-time, the mapping system needs to be able to process the incoming laser and odometry measurements efficiently. Developing techniques for computing accurate maps during real-time operation was a core focus in the second year of the TRADR project.

A common approach to the simultaneous localization and mapping problem is to consider, at each optimization step, a sliding window over the latest measurements. Within the current framework, this means that only the latest section of the robot trajectory is being optimized. The longer this sliding window is, the more measurements need to be processed, and therefore the longer the optimization requires. Nonetheless, longer sliding windows usually yields better results.

In order to determine an optimal sliding window size, we looked at its influence on the solution optimality, an analysis described in [20] (Annex Overview 2.10). We determined that, when operating the robot at normal driving speed, a sliding window of roughly 90 seconds (equivalent to 30 laser demi-rotations) contains most of the new information and yields a near to optimal solution. We have therefore chosen this sliding window size for the TRADR project.

As a second step to maximize the mapping efficiency, we asked ourselves whether the optimization correction power was uniformly distributed over that sliding window. Our intuition was that segments of the trajectory which have been optimized several times and have converged to a solution would need fewer correction than newer trajectory segments. The corresponding analysis, detailed in [20] (Annex Overview 2.10), shows that the most recent part of the estimated trajectory is very dynamic and indeed requires more correction to represent the real robot trajectory.

In TRADR, we built on this finding by developing an efficient algorithm for processing that sliding window. Compared to the commonly used uniform sampling of control points over the sliding window, this algorithm enables to produce a point cloud of same accuracy in less computational time, which is relevant for real-time operation. For more information on this algorithm, the reader is encouraged to consult section [20] (Annex Overview 2.10).

Fig. 8 shows a point cloud representation which was produced in real-time based on the previously mentioned algorithm. One can easily identify

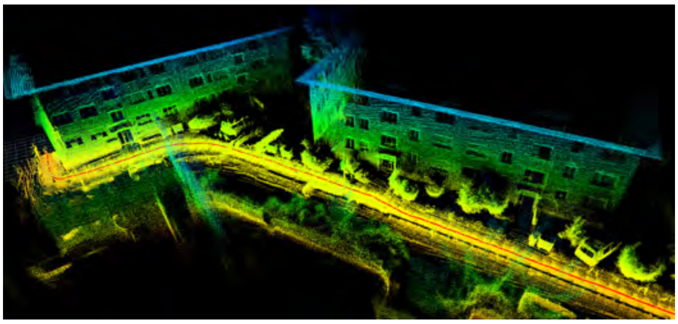

Figure 8: Point cloud generated in real time with non-uniform sampling of trajectory control points.

key elements such as buildings, vehicles, bushes and bins which represents a good basis for developing an object recognition system.

**Offline merging of metrical maps** We have developed the mapping framework so that it is both suitable for online operation and global offline optimization. The framework is flexible as it permits to select which section of the trajectory one wishes to optimize and is capable of co-optimizing trajectories resulting from different sorties. Good results are obtained as long as the robots are close in space and observe the same section of the environment. However, when different environment sections are explored during different sorties, the estimated trajectories experience a drift which is difficult to account for. When the estimated trajectories cross each other again, the robots might actually be in different sections of the environment. Without higher level input, it is not possible to determine that these two different parts should not be matched together. This justifies the need for detecting loop closures by looking closer into the map structural elements, which is an ongoing work.

**Change detection** We are in the process of developing change detection techniques to account for dynamics in maps over several sorties, potentially also within single sorties. This is important for both safe robot operation and map maintenance, two extremely relevant aspects for autonomous robots.

This module is being developed within a student project with a goal to identify and implement strategies for robots to detect changes in maps and update maps accordingly. The developed technologies should be general and as platforms independent as possible. However, to the date of the submission of this report we cannot report on results yet.

**Terrain contact modeling and classification** The contact between the sub-tracks of the TRADR UGV and the underlying surface is an important feedback for autonomous safe navigation [31, 58]. For example, during stair-climbing, ensuring the contact of the sub-tracks with the ridges of the stairs favours preventing tip-over situations thus improving the stability of the robot [32]. However, due to the lack of a proximity sensor on the sub-tracks of UGV platform, this feedback is not available for the control. In Year 1 we proposed an approach for estimating both the touch and the detach of the sub-tracks with the traversed surface [32]. This approach was based on a non-linear classifier [58]. Through this classifier we effectively learned the tolerance threshold of the sub-tracks torque undergoing the contact. In Year 2 we extended this work by developing a contact sensor that, given the measure of the residual at the sub-track motors, discriminates, on the basis of its trend, whether it has been determined by a disturbance due to the robot motion on the surrounding or it has been generated by a collision. The contact sensor extends the well-known Fault Detection and Isolation (FDI) schema [14, 18] to tracked vehicles and it relies on to a multi-resolution decomposition technique to extract from the residual local properties of the input signals, such as edges, spikes or transient [11]. Then, a sparse SVM classifier has been used to select those local properties, meaningful for the contact [71]. Experiments have revealed that the sensor is able to discriminate a contact on the basis of both the intensities and the frequency sub-bands occupied by the residual signal under consideration, reaching an accuracy of 84.31%. More details about this work can be found in Annexes, Section 2.12.

### 1.3.2 Task T1.4 (Robot centric metrical maps and models storage II – Dynamic scene)

**Layered visual SLAM for multi-density mapping** In TRADR, multiple aerial robots are to be used. An efficient visual SLAM architecture consisting of multiple independent layers is proposed [48] (Annex Overview 2.8). The fastest algorithm builds the basic layer and models the environment as a non-dense map to serve for urgent tasks such as localization and tracking. An even lower layer can be realized using sensor data as IMU measurements. Computationally more expensive algorithms build the middle layers, which provides semi-dense models of the environment for purposes, reaching from acute navigation planning to medium-urgent tasks as dynamics detection and graph optimization. The next higher layer builds full maps which are to be used for scene understanding of the entire sortie and for multi-map graph optimization. The highest layer is built by an algorithm which serves for long-term understanding of dynamic scene. This layer represents the persistency level and does graph optimization for maps gathered over time. In [47] (Annex Overview 2.9), three layers are implemented. A system overview is given by Fig. 9. The layers have different abstraction grades and different speed and benefit from each other and

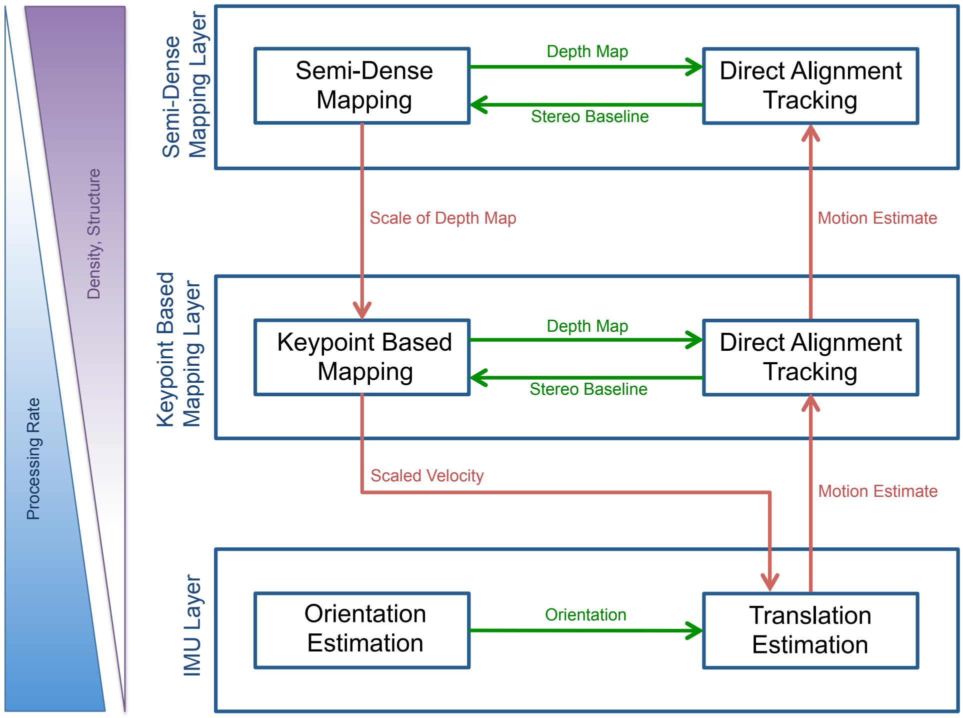


Figure 9: Overview of layered visual perception architecture

enhance the quality and the efficiency.

- An IMU provides 3D motion priors for the visual SLAM. We use velocity feedback from the vision-based next higher layer to close the open loop process of double integration and correct the velocity every time when a image frame is tracked. Estimating between two image frames, the position drift is sufficiently small.

- A slower keypoint based monocular method builds the intermediate layer. It uses IMU estimation as motion prior to better initialize the image-alignment-based direct tracking and reduce its computational cost for tracking. On the other side, this layer provides more accurately estimated velocity feedback to the IMU layer every time when a frame is tracked. This layer provides initialization for the tracking algorithm of the top layer.

- The top layer, the slowest one, provides semi-dense maps of the environment. Using two cameras and moving the camera pair so that the baseline of the pair lies orthogonal to the translational motion, we obtain stereo matches in two directions - one at a single time step and another over time. The map is then represented in metric scale and provides continuously a rescale factor to the intermediate layer, whose motion prior is thus also in metric scale.

With IMU-aided pre-estimation of motion, higher robustness against quick motions is expected. In our method, information flows not only from the faster to the slower levels, but also from the slower to the faster levels, what is novel to our best knowledge. The layers are individually complemented modules which can run on different computers.

Full description of the algorithm is given in the draft paper [47] (Annex Overview 2.9). We are now in the process of code optimization and experimental results are not available yet.

**Multi-modal map merging** One challenge in TRADR is the merging of mapping data gathered from different robots. For merging maps of same modalities, e.g. laser-laser or vision-vision, state-of-the-art modules were tested. Furthermore, a technique was developed to also merge vision and laser maps in the absence of further prior registration [29] (Annex Overview 2.11).

The researched technique is based on finding similar structural segments in vision and laser maps and reporting back the transformation between matching map elements. The full processing pipeline is depicted in Fig. 10 and described in full detail in [29] (Annex Overview 2.11). Therefore, individual maps are independently built from laser sensory data and vision data. The laser maps are constructed using the technique described in [Deliverable 2.2, ref to general description of mapper]. Vision maps are constructed facilitating a similar pose-graph-based pipeline with an additional densification step using Patch-based Multi-View Stereo algorithms (PMVS) [28].

Having two point-cloud candidates of equal densities, structural descriptors are calculated on a subset of the points. Structural descriptors capture the local neighbourhood around a chosen keypoint using a combination of a binning shape and a descriptor paradigm. Simple occupancy-based and binary density-comparing descriptors offer the best performances amongst all evaluated descriptors.

In a next step the descriptor dimensionality is drastically reduced by performing a projection to the most expressive dimensions. This can be achieved by assessing artificial noisy matches between nearby keypoints and observing the robust descriptor dimensions. This functionality was inspired by [1]. A reduced descriptor dimensionality significantly accelerates the following feature matching step.

A subsequent k-nearest-neighbour search between the two keypoint-feature maps results in the matching matrix. The densest regions of the matching matrix are then segmented into place matching candidates using the placeless approach by [55].

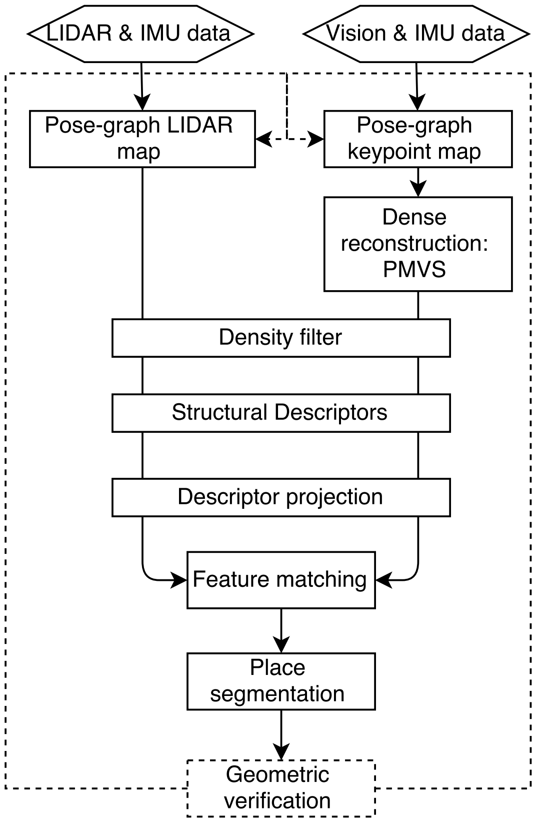


Figure 10: System diagram of the approach for matching and fusing vision-laser maps.

Before finally feeding the place matching candidates for map merging back into the map, a geometric verification is performed. The verification step assesses the individual feature matching candidates of a place match for a common transformation between the places.

Ultimately, the place matching candidates are added as constraints between the vision and laser map and can be optimized within the pose-graph framework.

## 1.4 Relation to the state-of-the-art

**Controlling robot morphology from incomplete measurements** To deal with incomplete data, the $Q$ function values have to be marginalized over the missing features. Such marginalization is often tackled by sampling [51, 72] or EM algorithm [30]. Especially for the Gaussian Processed (GPs) with Squared Exponential kernel, the moment matching marginalization method was proposed by Deisenroth et al. [16]. Different marginalization methods for GPs and Piecewise Constant functions were evaluated in [84].

We are not aware of any real mobile platform which would use arm as an active sensor for inspecting unknown terrain. Most of the efforts in active inference are directed towards active classification [19, 4, 42] or active 3D reconstruction. Doumanoglou et al. [19] use two robotic arms for folding an unknown piece of cloth whose type is recognized from RGB-D data (Kinect). One view is usually insufficient, therefore the cloth needs to be turned around to generate an alternative view. The turning action is implicitly learned with Decision Forests. Bjorkman et al. [4] also recognize objects from RGB-D data. In contrast to [19], Bjorkman et al. use the robotic arm as an active sensor, to touch the self-occluded part of the object in order to reconstruct the invisible 3D shape. While all these classification approaches actively evaluate features in order to discriminate the true object class from other possible classes as fast as possible, the $Q$-learning–based inference used in [83] evaluates the features in order to find some of the suitable flipper configurations that allow for a safe traversal.

**Terrain perception in sensory deprived environments** The problem of terrain characterization primarily using proprioceptive sensors, but also by sonar/infra-red range-finders and by a microphone is discussed in [62]. The authors exploit neural networks trained for each sensor and demonstrate that they are able to recognize different categories: gravel, grass, sand, pavement and dirt surface. Furthermore, they present a concept of terrain-characteristic curves that establish relationship between currents in motors driving the main wheels and resulting angular rate of the robot. In [66] we took a similar approach to train a regression function that maps from a space of features extracted from inertial sensors to parameters that compensate slippage in track odometry. In both cases the aim was to improve localization and control of the robot. Our current work focuses more on the actual terrain profile prediction, necessary for successful traversal.

Lack of sufficient visual information related to danger of collision with obstacles is addressed in [2]: decision whether it is safe to navigate through vegetation is based on wide-band radar measurements since it is impossible to detect solid obstacle behind vegetation from laser range-finder or visual data. Artificial whiskers offer an alternative solution; they mimic facial whiskers of animals and using them as a tactile sensor is a promising way to explore areas, which are prohibitive to standard exteroceptive sensors. Work of [69] presents a way to use array of actively actuated whiskers to discriminate various surface textures. In [63], similar sensor is used for a SLAM task. Two sensing modalities—the whisker sensor array and the wheel odometry are used to build a 2D occupancy map. Robot localization is then performed using particle filter with particles representing one second long "whisk periods". During these periods, the sensor actively builds local model of the obstacle it touches. Unfortunately, design of our platform does not allow using such whiskers due to rotating laser range-finder.

Relation between shape of terrain that we are interested in and configuration of the flippers is investigated in [61]. The authors exploit the knowledge about robot configuration and torques in joints to define a set of rules for climbing and descending obstacles not observed by exteroceptive sensors. We investigated this problem in [83, 84] by introducing the adaptive traversability algorithm based on machine learning. We collected features from both proprioceptive and exteroceptive sensors to learn a policy that ensures safe traversal over obstacles by adjusting robot morphology. Our motivation coincided with [61], aiming primarily to lower the cognitive load of the operator.

On contrary to the approaches exploiting only simple contact sensors, we extend our sensory suite with a robotic arm for further active perception for cases if necessary. Related to the active perception, relevant ideas and techniques come from the field of haptics. The work of [4] proposes to create models of objects in order to be able to grasp them. The idea is to complement visual measurements by tactile ones by strategically touching the object in areas with high shape uncertainty. For this purpose they use Gaussian processes (GP, [60, 81]) to express the shape of the object. We take a similar approach: we choose parts of terrain to be explored by the robotic arm based on uncertainty of the estimate resulting from the sampling process Probabilistic approach to express uncertainty in touched points is also described in [57], where only tactile sensors of a robotic hand are used to reconstruct the shape of an unknown object.

**Terrain classification using acoustic features** There are several approaches to terrain detection in the robotics literature based on a variety of sensors, such as motor slip measurement [67], velocity and acceleration features ([75], [12]), visual detection ([38], [65] and [54]), vibration ([7], [79], [78]), sound [50], or a mixture of several sensors ([40], [39]). Many of the aforementioned approaches are unsupervised and aim at automatically acquiring models to improve, e.g., the accurate navigation of the robot, or its autonomy. We use a supervised approach with human-labelled data for training the classifiers. Similarly to [50], we use acoustics data only to distinguish robot terrain interaction. However, the data collection methods, the robot, the feature extraction techniques and the classification schemes we used are different.

The choices we made for our approach were motivated by the primary purpose of the module for the generation of human-readable reports. While this may limit the use of the method for sensor fusion or the avoidance of hazardous terrain, the results suggest that the acoustic features may also be useful for other applications.

**Sensor-less contact detection for articulated tracked robots** Real-time detection of collisions has been widely studied in the literature, in particular in the context of robotic manipulators [70, 33, 45, 15, 35, 14]. Several approaches have been proposed, based either on comparison with nominal torques on desired motion of the robotic arm, or on the parallel simulation of robot dynamics, and on fault detection and isolation. These approaches are typically based on an accurate dynamic model of the manipulator, and in most systems this is complemented with a sensory apparatus measuring the presence of collision forces that produce work at the contact. Moreover, these approaches have three main limitations: (1) it is hard to add details about couplings, elasticity, friction and other nonlinear dynamics, which are required for high accuracy; (2) the performance of procedures for the identification of the parameters of the dynamic model strongly depends on both the experimental setting (e.g., with or without contacts) and the exciting trajectories; (3) they make strong assumptions to handle contacts. An alternative and appealing approach to detecting contacts is to use machine learning methods to learn a function assessing the occurrence of collisions on the robotic links. In this regard, Calandra *et al.* [9] proposed a mixtures-of-experts based on Gaussian Processes (GP) to learn the non-linear system dynamics of the humanoid robot *iCub* [41]. In this work, each of GP experts models a single contact type. Moreover, by using a gating network that activates and deactivates the individual GP experts the model can switch between contact types, thus generalizing to changing contact locations. However, although this approach does not require a spatially calibrated model of the skin, thus disregarding the information about the exact position of the contact, it still relies on the raw data of the tactile sensors placed on the robot skin for estimating the contact. Nevertheless, to the best of our knowledge there is no work that cope with contact detection, in particular without the support of any tactile sensor, for actively articulated tracked vehicles, as TRADR UGV. A first demonstration of how a machine learning technique can be applied for contact identification as been proposed by us in [58]. Our current work improves the statistical learning framework in [58]. Indeed, it also relies on a multi-resolution decomposition technique for extracting the features from the residual signal, computed according to the well-known Fault Detection and Isolation (FDI) schema [14, 18], in order to discriminate the contact on the basis of both the intensities and the frequency sub-bands occupied by this signal.

**SLAM using visual sensors** Known SLAM methods can be categorized by different density levels of the resulting maps. In point-based SLAM methods, features corners and blobs are used to find correspondences between images. Via triangulation, the distance of the points can be found, which is integrated into the map. Although a lot of such approaches such as [13, 46] are fast and efficient, the sparse representation of the environment is often not sufficient for navigation purpose. Here, sufficient existence of features is the essential condition. Semi-Direct Monocular Visual Odometry (SVO) [26] however works directly on image pixel intensities of interest areas, which allows high tracking accuracy without feature description. Despite higher density, the inherent sparsity of point-based maps remains.

To exploit full information of visual sensor data, dense SLAM methods models the environment by fully dense surfaces without relying on features. Also these type of methods operate directly on pixel intensities. Examples of dense methods are [56, 80]. A further development of dense methods proposes to use RGB-D cameras to directly obtain scene geometry. Although the environment structure is very well represented, this type of algorithms is computationally highly expensive.

A novel approach called Large-Scale Direct SLAM (LSD-SLAM) [23, 21] proposes an algorithm also working directly on pixel intensities without any need of interest points for mapping and tracking. LSD-SLAM however uses only image regions with sufficient large gradients and sufficient large angle to the stereo baseline. This result in a semi-dense maps, which are built more efficiently than dense methods and are more informative than point-based maps. Semi-dense method is very parameter-sensitive and suffers from inaccurate measurements in absence of specific camera motion ensuring sufficient translations with large angle to the baseline.

More and more researchers develop multimodal methods using visual sensors in combination with inertial sensors. Recent researches try to show that graph-based methods, which integrates redundant constraints and optimizes the graph, are more accurate than estimation fusion methods using filtering. A recent work, IMU Preintegraion on Manifold for Efficient visual Maximum-a- Posteriori Estimation [25] for example, proposes to preintegrate multiple inertial measurements between subsequent keyframes reducing the computational complexity. The preintegration theory addresses the manifold structure of the rotation group and suggests how to propagate the uncertainty through the measurements. The preintegrated IMU model is than integrated to a factor graph based visual-inertial framework and can be realized with any visual SLAM method.

Our recent work [47] (Annex Overview 2.9) and [48] (Annex Overview 2.8) aim at two main contributions:

- Development of a camera-based SLAM system which is divided into multiple mutually supported modeling levels that efficiently map the dynamic environment and provide different forms of environment representation for robots and humans
- Development of a multi-camera system for SLAM purpose that operates with baselines in multiple directions, thus increasing the robustness and the accuracy compared to the monocular solutions. While working on this concept, the authors of [23, 21] introduced a similar concept in IROS 2015 [22].

**Simultaneous Localization and Mapping (SLAM)** The SLAM problem has been addressed with several major approaches, e.g. Kalman-Filter, Particle-Filter, graph-based and their various variants [68] [74].

Graph-SLAM also pioneered by [53] and [24] treats the problem as an optimization task. It has become increasingly popular in recent years and an active community is performing research in this direction: [17] [43] [5] [6] [3] [10] [8].

In a classic graph-SLAM approach a map is represented as a graph. Nodes depict discrete states $\boldsymbol{x}_i$ of the system, i.e. robot poses and landmark locations, $i$ denoting the index of the node. Edges represent measurements $\boldsymbol{z}_{ij}$ between the nodes $i$ and $j$, i.e. the state variables. Graph-SLAM treats the SLAM-problem as an minimization of the error $\boldsymbol{e}_{ij}(\boldsymbol{x})$ between measurements $\boldsymbol{z}_{ij}$ and dynamic models $\boldsymbol{f}_{ij}(\boldsymbol{x})$ of the state $\boldsymbol{x}_i$ resulting in the updated states $\boldsymbol{x}_i^*$ facilitating the Gaussian assumption and information matrix $\boldsymbol{\Omega}_{ij}$. The predictions are often derived from wheel odometry or IMU integration, whereas measurement updates may come from scan-matching or key-point-tracking.

$$\boldsymbol{e}_{ij}(\boldsymbol{x}) = \boldsymbol{z}_{ij} - \boldsymbol{f}_{ij}(\boldsymbol{x}) \tag{5}$$

$$\boldsymbol{x}_i^* = argmin \sum_{ij} \boldsymbol{e}_{ij}(\boldsymbol{x}_i)^T \boldsymbol{\Omega}_{ij} \boldsymbol{e}_{ij}(\boldsymbol{x}_i) \tag{6}$$

The graph can be optimized by minimizing the error using non-linear least-squares optimization, e.g. the Gauss-Newton algorithm. However, in a naive implementation the optimization problem grows with the size of the map with the accumulation of states and measurements, increasing the computational time [49]. Several strategies have been proposed to limit the growing complexity, such as the work on *iSAM* by [44] [43] and the work on *Zebedee* [6].

Work [5] treats the SLAM problem in a continuous fashion, i.e. have a continuous state space $\boldsymbol{x}_{ij}(t)$ of a robot trajectory interpolating between state coefficients $\boldsymbol{x}_i$ and the continuous time index $t$. The technique was later formalized by [27]. Here, predictions $\boldsymbol{f}_{\boldsymbol{k},l}(\boldsymbol{x})$ and measurements $\boldsymbol{z}_{\boldsymbol{k},l}$, can be continuously registered on the interpolated continuous-time state, with $\boldsymbol{k}$ denoting a list of affected nodes and $l$ being an index.

The approach [6] extends the idea of continuous-time graph-SLAM to having a decoupled state-representation of a fixed-frequency correction trajectory composed with a constant high-frequency baseline trajectory generated from high-rate measurements. Therewith high fidelity of the measurements is preserved while keeping the state space size decoupled. The approach however has not yet investigated the concept of corrections in further detail, e.g. different correction curve representations or sampling strategies.

The recent work [3] uses the standard continuous-time approach with lidars facilitating insights from previous work with rolling shutter cameras. They treat lidar intensity imagery like slow rolling shutter cameras for trajectory estimation of a single robot. Continuous-time SLAM has the potential to limit state space size and enables advanced knot-placement strategies.

**Multi modal map merging** Multi-modality is closely related to heterogeneity addressing the heterogeneous robot sensor outfits. In mapping applications it is desirable to facilitate representations which express as many invariances as possible, i.e. scale, lighting, orientation. The fusion of lidar and vision data has received some attention in the recent years in the areas of place recognition.

The approach of [36, 37] for place recognition requires prior calibration between visual and lidar data. The authors define unique signatures for places consisting of both visual appearance and structural features.

[76] consecutively use visual and structural data. The authors consider a multi-sensor system with pre-registered lidar and vision. Applying appearance-based visual keypoint descriptors, a subset of place match candidates are found and consecutively checked via geometric verification.

[82] achieve visual localization within prerecorded high quality 3D lidar maps, augmented with reflectivity measures. Images from monocular cameras are matched against predicted views from the lidar reflectivity map.

[59] perform dense scene reconstruction by fusing lidar point-clouds with key-points from stereo-matching outside the lidar's field of view. They fa-

cilitate Conditional Random Fields (CRFs) and Delauney triangulation to infer the reconstruction assuming piecewise planarity of surfaces.

However, all existing approaches require either prior registration of sensor data or features which hold additional appearance-based dimensions, such as intensity values. Furthermore are most appearance-based features viewpoint- and lighting-dependant. Fusion of multi-modal features solely relying on structure has not yet been presented.

A means of fast fusion between the modalities is desirable for achieving an initial alignment for place recognition. Furthermore, the selection of structure based rather than appearance based features is desirable for robust fusion.

## References

[1] Keypoint design and evaluation for place recognition in 2D lidar maps. *Robotics and Autonomous Systems*, 57(12):1211–1224, 2009.

[2] J. Ahtiainen, T. Peynot, J. Saarinen, and S. Scheding. Augmenting traversability maps with ultra-wideband radar to enhance obstacle detection in vegetated environments. In *Proceedings of the IEEE/RSJ International Conference on Robots and Intelligent Systems 2013*, pages 5148–5155, 2013.

[3] S. Anderson, K. MacTavish, and T. D. Barfoot. Relative continuous-time SLAM. *The International Journal of Robotics Research*, 34(12):1453–1479, 2015.

[4] M. Bjorkman, Y. Bekiroglu, V. Hogman, and D. Kragic. Enhancing visual perception of shape through tactile glances. In *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, pages 3180–3186, 2013.

[5] M. Bosse and R. Zlot. Continuous 3D scan-matching with a spinning 2D laser. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4312–4319, 2009.

[6] M. Bosse, R. Zlot, and P. Flick. Zebedee: Design of a spring-mounted 3-D range sensor with application to mobile mapping. *IEEE Transactions on Robotics*, 28(5):1104–1119, 2012.

[7] C. Brooks, K. Iagnemma, et al. Vibration-based terrain classification for planetary exploration rovers. *Robotics, IEEE Transactions on*, 21(6):1185–1191, 2005.

[8] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kümmerle, C. Dornhege, M. Ruhnke, A. Kleiner, and J. D. Tardós. A comparison of SLAM algorithms based on a graph of relations. In *IEEE/RSJ*

*International Conference on Intelligent Robots and Systems (IROS)*, pages 2089–2095, 2009.

[9] R. Calandra, S. Ivaldi, M. Deisenroth, E. Rückert, and J. Peters. Learning inverse dynamics models with contacts. In *ICRA*, pages 3186–3191, 2015.

[10] N. Carlevaris-Bianco, M. Kaess, and R. Eustice. Generic Node Removal for Factor-Graph SLAM. *IEEE Transactions on Robotics*, 30(6):1371–1385, 2014.

[11] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser. Signal processing and compression with wavelet packets. In *Wavelets and their applications*, pages 363–379. Springer, 1994.

[12] E. Coyle, E. G. Collins Jr, and R. G. Roberts. Speed independent terrain classification using singular value decomposition interpolation. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4014–4019. IEEE, 2011.

[13] A. Davison, I. Reid, N. Molton, and O. Stasse. Monoslam: Real-time single camera slam. In *Transaction on Pattern Analysis and Machine Intelligence (TPAMI)*, 2007.

[14] A. De Luca and L. Ferrajoli. A modified newton-euler method for dynamic computations in robot fault detection and control. In *ICRA*, pages 3359–3364, 2009.

[15] A. De Luca and R. Mattone. Sensorless robot collision detection and hybrid force/motion control. In *ICRA*, pages 999–1004, 2005.

[16] M. Deisenroth, D. Fox, and C. Rasmussen. Gaussian Processes for data-efficient learning in robotics and control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):408–423, 2015.

[17] F. Dellaert and M. Kaess. Square Root SAM. *Robotics: Science and Systems*, pages 1–32, 2005.

[18] W. Dixon, I. Walker, D. Dawson, and J. Hartranft. Fault detection for robot manipulators with parametric uncertainty: a prediction error based approach. In *ICRA*, volume 4, pages 3628–3634, 2000.

[19] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis. Active random forests: An application to autonomous unfolding of clothes. In *Computer Vision–ECCV 2014*, pages 644–658. Springer, 2014.

[20] R. Dubé, H. Sommer, A. Gawel, M. Bosse, and R. Siegwart. Non-uniform sampling strategies for continuous correction based trajectory optimization. In *submitted*.

[21] J. Engel, T. Schoeps, and D. Cremers. Large-scale direct monocular slam. In *IEEE European Conference on Computer Vision (ECCV)*, 2014.

[22] J. Engel, J. Stueckler, and D. Cremers. Large-scale direct slam with stereo cameras. In *International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[23] J. Engel, J. Sturm, and D. Cremers. Semi-dense visual odometry for a monocular camera. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[24] J. Folkesson and H. Christensen. Graphical SLAM - a self-correcting map. In *IEEE International Conference on Robotics and Automation (ICRA)*, number April, pages 383–390, 2004.

[25] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. Imu preintegration on manifold for efficient visual- inertial maximum-a-posteriori estimation. In *Robotics Science and Systems*, 2015.

[26] C. Forster, M. Pizzoli, and D. Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[27] P. Furgale, T. D. Barfoot, and G. Sibley. Continuous-time batch estimation using temporal basis functions. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 2088–2095, 2012.

[28] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[29] A. Gawel, T. Cieslewski, R. Dubé, M. Bosse, J. Nieto, and R. Siegwart. Structure based vision-laser matching. In *submitted*.

[30] Z. Ghahramani and M. I. Jordan. Supervised learning from incomplete data via an EM approach. In *Neural Information Processing Systems (NIPS)*, pages 120–127. Morgan Kaufmann, 1994.

[31] F. Gianni, M. Pirri and M. A. Ruiz. Dynamic contact sensing for articulated tracked vehicles. In *RSS 2015 Workshop on Learning for Visual and Tactile Interaction*, pages 1–3, 2015.

[32] M. Gianni, F. Ferri, M. Menna, and F. Pirri. Adaptive robust three-dimensional trajectory tracking for actively articulated tracked vehicles. *Journal of Field Robotics*, pages n/a–n/a, 2015.

[33] G. Gilardi and I. Sharf. Literature survey of contact dynamics modelling. *Mechanism and Machine Theory*, 37:1213–1239, 2002.

[34] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik. Learning rich features from rgb-d images for object detection and segmentation. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8695 of *Lecture Notes in Computer Science*, pages 345–360. Springer International Publishing, 2014.

[35] S. Haddadin, A. Albu-Schäffer, A. De Luca, and G. Hirzinger. Collision detection and reaction: a contribution to safe physical human-robot interaction. In *IROS*, pages 3356–3363, 2008.

[36] K. L. Ho and P. Newman. Combining Visual and Spatial Appearance for loop closure detection.pdf. *European Conference on Mobile Robots (ECMR)*, 2005.

[37] K. L. Ho and P. Newman. Loop closure detection in SLAM by combining visual and spatial appearance. *Robotics and Autonomous Systems*, 54(9):740–749, 2006.

[38] A. Howard and H. Seraji. Vision-based terrain characterization and traversability assessment. *Journal of Robotic Systems*, 18(10):577–587, 2001.

[39] K. Iagnemma, H. Shibly, and S. Dubowsky. On-line terrain parameter estimation for planetary rovers. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, volume 3, pages 3142–3147. IEEE, 2002.

[40] K. D. Iagnemma and S. Dubowsky. Terrain estimation for high-speed rough-terrain autonomous vehicle navigation. In *AeroSense 2002*, pages 256–266. International Society for Optics and Photonics, 2002.

[41] S. Ivaldi, M. Fumagalli, M. Randazzo, F. Nori, G. Metta, and G. Sandini. Computing robot internal/external wrenches by means of inertial, tactile and f/t sensors: Theory and implementation on the icub. In *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots*, pages 521–528, 2011.

[42] Z. Jia, Y.-J. Chang, and T. Chen. A general boosting-based framework for active object recognition. In *British Machine Vision Conference*, pages 1–11, 2010.

[43] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 2012.

[44] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Fast incremental smoothing and mapping with efficient data association. In *IEEE In-*

*ternational Conference on Robotics and Automation (ICRA)*, number April, pages 1670–1677, 2007.

[45] S. Katsura, Y. Matsumoto, and K. Ohnishi. Analysis and experimental validation of force bandwidth for force control. In *Proceedings of the IEEE International Conference on Industrial Technology*, volume 2, pages 796–801, 2003.

[46] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality (ISMAR)*, 2007.

[47] D.-U. Kong. Layered visual perception architecture for efficient multi-density environment mapping and localization. Unpublished Draft Paper, December 2015.

[48] D.-U. Kong. Persistent mapping of dynamic environments by multiple aerial robots with multi-camera systems. Dissertation Proposal, January 2015.

[49] M. T. Lazaro, L. M. Paz, P. Pinies, J. Castellanos, G. Grisetti, and Others. Multi-robot SLAM using condensed measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1069–1076. IEEE, 2013.

[50] J. Libby and A. J. Stentz. Using sound to classify vehicle-terrain interactions in outdoor environments. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3559–3566. IEEE, 2012.

[51] D. J. Lizotte, L. Gunter, E. Laber, and S. A. Murphy. Missing data and uncertainty in batch Reinforcement Learning. In *Neural Information Processing Systems (NIPS)*, 2008.

[52] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[53] F. Lu and E. Milios. Globally Consistent Range Scan Alignment for Environment Mapping. *Autonomous Robots*, 4(4):333–349, 1997.

[54] L. Lu, C. Ordonez, E. G. Collins Jr, and E. M. DuPont. Terrain surface classification for autonomous ground vehicles using a 2d laser stripe-based structured light sensor. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 2174–2181. IEEE, 2009.

[55] S. Lynen, M. Bosse, P. Furgale, and R. Siegwart. Placeless place-recognition. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 303–310. IEEE, 2014.

[56] L. Matthies, R. Szeliski, and T. Kanade. Incremental estimation of dense depth maps from image sequences. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1988.

[57] M. Meier, M. Schöpfer, R. Haschke, and H. Ritter. A probabilistic approach to tactile shape reconstruction. *Robotics, IEEE Transactions on*, 27(3):630–635, 2011.

[58] M. Menna, M. Gianni, F. Ferri, and F. Pirri. Real-time autonomous 3d navigation for tracked vehicles in rescue environments. In *IROS 2014*, pages 696–702, 2014.

[59] O. Miksik, P. Patrick, and P. H. S. Torr. Incremental Dense Multimodal 3D Scene Reconstruction. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, 2015.

[60] A. O'Hagan and J. Kingman. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–42, 1978.

[61] K. Ohno, S. Morimura, S. Tadokoro, E. Koyanagi, and T. Yoshida. Semi-autonomous control system of rescue crawler robot having flippers for getting over unknown-steps. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3012–3018. IEEE, 2007.

[62] L. Ojeda, J. Borenstein, G. Witus, and R. Karlsen. Terrain characterization and classification with a mobile robot. *Journal of Field Robotics*, 23(2):103–122, 2006.

[63] M. Pearson, C. Fox, J. Sullivan, T. Prescott, T. Pipe, and B. Mitchinson. Simultaneous localisation and mapping on a multi-degree of freedom biomimetic whiskered robot. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 586–592, 2013.

[64] M. Pecka, K. Zimmermann, M. Reinstein, and T. Svoboda. Controlling robot morphology from incomplete measurements. *submitted*.

[65] J. Poppinga, A. Birk, and K. Pathak. Hough based terrain classification for realtime detection of drivable ground. *Journal of Field Robotics*, 25(1):67, 2008.

[66] M. Reinstein, V. Kubelka, and K. Zimmermann. Terrain adaptive odometer for mobile skid-steer robots. In *Proc. IEEE Int Robotics and Automation (ICRA) Conf*, pages 4691–4696, 2013.

[67] D. Sadhukhan, C. Moore, and E. Collins. Terrain estimation using internal sensors. In *Proceedings of the 10th IASTED International Conference on Robotics and Applications (RA)*, 2004.

[68] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza. *Introduction to autonomous mobile robots*. MIT press, 2nd editio edition, 2011.

[69] J. Sullivan, B. Mitchinson, M. Pearson, M. Evans, N. Lepora, C. Fox, C. Melhuish, and T. Prescott. Tactile discrimination using active whisker sensors. *IEEE Sensors Journal*, 12(2):350–362, 2012.

[70] S. Takakura, T. Murakami, and K. Ohnishi. An approach to collision detection and recovery motion in industrial robot. In *Proceedings of the 15th IEEE Annual Conference on Industrial Electronics Society*, pages 421–426 vol.2, 1989.

[71] M. Tan, L. Wang, and I. W. Tsang. Learning sparse svm for feature selection on very high dimensional datasets. In *ICML*, pages 1047–1054, 2010.

[72] M. A. Tanner and W. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540, 1987.

[73] A. Tesfay Gebru. Terrain type classification based on sound, July 2015.

[74] S. Thrun, W. Burgard, and D. Fox. *Probabilistic robotics*. MIT press, 2005.

[75] D. Tick, T. Rahman, C. Busso, and N. Gans. Indoor robotic terrain classification via angular velocity based hierarchical classifier selection. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3594–3600. IEEE, 2012.

[76] F. Tungadi, W. L. D. Lui, L. Kleeman, and R. Jarvis. Robust online map merging system using laser scan matching and omnidirectional vision. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7–14, 2010.

[77] V. Šalanský, V. Kubelka, K. Zimmermann, M. Reinstein, and T. Svoboda. Touching without vision: terrain perception in sensory deprived environments. In *submitted*.

[78] C. C. Ward and K. Iagnemma. Speed-independent vibration-based terrain classification for passenger vehicles. *Vehicle System Dynamics*, 47(9):1095–1113, 2009.

[79] C. Weiss, H. Frhlich, and A. Zell. Vibration-based terrain classification using support vector machines. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4429–4434. IEEE, 2006.

[80] A. Wendel, M. Maurer, G. Graber, T. Pock, and H. Bischof. Dense reconstruction on-the-fly. In *IEEE European Conference on Computer Vision (ECCV)*, 2012.

[81] C. K. Williams and C. E. Rasmussen. Gaussian processes for regression. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 514–520. The MIT Press, 1996.

[82] R. W. Wolcott and R. M. Eustice. Visual Localization within LIDAR Maps for Automated Urban Driving. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 176 – 183, Chicago, IL, USA, 2014.

[83] K. Zimmermann, P. Zuzánek, M. Reinstein, and V. Hlaváč. Adaptive traversability of unknown complex terrain with obstacles for mobile robots. In *ICRA2014: Proceedings of 2014 IEEE International Conference on Robotics and Automation*, pages 5177–5182, Piscataway, USA, June 2014. IEEE.

[84] K. Zimmermann, P. Zuzánek, M. Reinstein, T. Petříček, and V. Hlaváč. Adaptive traversability of partially occluded obstacles. In *2015 IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 3959–3964, Piscataway, USA, May 2015. Institute of Electrical and Electronics Engineers (IEEE).

# 2 Annexes

## 2.1 Pecka-TIE2016, "Controlling Robot Morphology from Incomplete Measurements"

**Bibliography** Martin Pecka, Karel Zimmermann, Michal Reinstein, and Tomáš Svoboda. Controlling Robot Morphology from Incomplete Measurements. Submitted to *IEEE Transactions on Industrial Electronics*, special section Motion Control for Novel Emerging Robotic Devices and Systems.

**Abstract** Mobile robots with complex morphology are essential for traversing rough terrains in Urban Search & Rescue (USAR) missions. Since teleoperation of the complex morphology causes high cognitive load of the operator, the morphology is controlled autonomously. The autonomous control measures the robot state and surrounding terrain which is usually only partially observable and thus the data are often incomplete. We marginalize the control over the missing measurements and explicitly evaluate a safety condition. If the safety condition is violated, a tactile terrain exploration by the body-mounted robotic arm gathers the missing data.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Restricted, paper under review. Not included in the public version of this deliverable.

## 2.2 Kubelka-RAS2016, "Improving Multimodal Data Fusion for Mobile Robots by Trajectory Smoothing"

**Bibliography** Vladimír Kubelka, Michal Reinstein, and Tomáš Svoboda. Improving Multimodal Data Fusion for Mobile Robots by Trajectory Smoothing. Submitted to *Robotics and Autonomous Systems*.

**Abstract** Localization of mobile robots is still an important topic, especially in case of dynamically changing, complex environments such as in Urban Search & Rescue (USAR). In this paper we aim for improving the reliability and precision of localization of our multimodal data fusion algorithm. Multimodal data fusion requires resolving several issues such as significantly different sampling frequencies of the individual modalities. We compare our proposed solution with the well-proven and popular Rauch-Tung-Striebel smoother for the Extended Kalman filter. Furthermore, we improve the precision of our data fusion by incorporating scale estimation for the visual modality.

**Relation to WP** The methods smoothes past trajectory estimate. It contributes to both tasks T1.3 – essential sensing and UGV control functionality and T1.4 Robot centric metrical maps and models storage.

**Availability** Restricted, paper under review. Not included in the public version of this deliverable.

## 2.3 Salansky-ICRA2016, "Touching without vision: terrain perception in sensory deprived environments"

**Bibliography** Vojtěch Šalanský, Vladimír Kubelka, Karel Zimmermann, Michal Reinstein, and Tomáš Svoboda. Touching without vision: terrain perception in sensory deprived environments. Submitted to IEEE International Conference on Robotics and Automation (ICRA) 2016.

**Abstract** In this paper we demonstrate a combined hardware and software solution that enhances sensor suite and perception capabilities of a mobile robot intended for real Urban Search & Rescue missions. A common fail-case, when exploring unknown environment of a disaster site, is the outage or deterioration of exteroceptive sensory measurements that the robot heavily relies on—especially for localization and navigation purposes. Deprivation of visual and laser modalities caused by dense smoke motivated us to develop a novel solution comprised of force sensor arrays embedded into tracks of our platform. Furthermore, we also exploit a robotic arm for active perception in cases when the prediction based on force sensors is too uncertain. Beside the integration of hardware, we also propose a framework exploiting Gaussian processes followed by Gibb's sampling to process raw sensor measurements and provide probabilistic interpretation of the underlying terrain profile. The profile is perceived by proprioceptive means only and successfully substitutes for the lack of exteroceptive measurements in the close vicinity of the robot, when traversing unknown and unseen obstacles. We evaluated our solution on real world terrains.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Restricted, paper under review. Not included in the public version of this deliverable.

## 2.4 Chmel-BT2015, "SW and HW Integration of an IP PTZ Camera onto a Mobile Outdoor Robot"

**Bibliography** Jakub Chmel. SW and HW Integration of an IP PTZ Camera onto a Mobile Outdoor Robot. Bachelor thesis, Czech Technical University in Prague, Faculty of Electrical Engineering. 2015.

**Abstract** The outdoor robot for urban search and rescue (USAR) missions lacks a high quality pan-tilt-zoom (PTZ) video stream. PTZ function can be simulated using an existing virtual camera, which uses data from LadyBug 3 (LB3). However, the LB3 camera may only use a digital zoom. This work is focused on SW and HW integration of a network (IP) PTZ camera onto the mobile outdoor robot. These types of cameras can add the optical zoom function to the robotic system. IP PTZ camera (Axis 214 PTZ) is the best choice for described problem in current situation. The goal of the thesis was to develop Robot Operating System (ROS) package that allows to use the camera on the TRADR robot. Most of source codes were written in Python and some in C++. In the thesis were performed several experiments that define the possibilities and limits of the developed solution.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Public. Available at `https://dspace.cvut.cz/handle/10467/61991`, pdf document at `https://dspace.cvut.cz/bitstream/handle/10467/61991/F3-BP-2015-Chmel-Jakub-SW_and_HW_Integration_of_an_IP_PTZ_Camera_onto_a_Mobile_Outdoor_Robot.pdf`

## 2.5 Mares-MT2015, "Safe Obstacle Traversal with Incomplete Data"

**Bibliography** Jakub Mareš. Safe Obstacle Traversal with Incomplete Data. Master thesis, Czech Technical University in Prague, Faculty of Electrical Engineering. 2015.

**Abstract** This thesis deals with terrain traversability for an unmanned ground vehicle (UGV) based on Niftibot platform. This mobile robot, which is dedicated for Urban Search and Rescue (USAR) missions, is equipped with auxiliary articulated tracks, so-called flippers. Flippers enhance robots ability to traverse complicated terrain, however they bring more degrees of freedom to control. Semi-autonomous control system which selects optimal flippers configuration with respect to traversed terrain is being developed at FEE, CTU. A system based on reinforcement learning and decision trees had been previously implemented. This system, however, required complete data from sensors. As model of environment was built using solely data from laser scanner, this condition was violated in some scenarios, e.g. in case of reflective surfaces. Therefore a partial reimplementation and an extension of the former system is introduced in this work. A new mode which utilizes JACO robotic arm for tactile exploration of terrain has been incorporated to the system. This helps to explore terrain invisible to laser scanner. The experiments with aluminium foil were performed to demonstrate that the arm helps the robot to complete information in robots map and furtherly use it to safely traverse terrain.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Public. Available at `https://dspace.cvut.cz/handle/10467/61738`, pdf document at `https://dspace.cvut.cz/bitstream/handle/10467/61738/F3-DP-2015-Mares-Jakub-Safe_Obstacle_Traversal_with_Incomplete_Data.pdf`

## 2.6 Salansky-MT2015, "Contact Terrain Exploration for Mobile Robot"

**Bibliography** Vojtěch Šalanský. Contact Terrain Exploration for Mobile Robot. Mater thesis, Czech Technical University in Prague, Faculty of Electrical Engineering. 2015.

**Abstract** For mobile robots, it is important to know the surrounding terrain. The goal of this diploma thesis is to provide a design and prototypical implementation of the method for terrain exploration via contact of robotic manipulator Kinova Jaco Arm with the terrain. There are more subgoals to be done within the implementation of contact exploration. This thesis deals with the contact detection of the robotic arm and the terrain, choosing the place from workspace for exploration, and the estimation of unknown places and mapping. The contact of the manipulator with the terrain is detected by using joint torques of actuators. The place that should be explored next is chosen to increase the expected usefulness for future estimation of the other unknown places. The places that are not explored are estimated by using proprioceptive sensors (tilt, currents in actuators). A method that saves and process the map from measured and estimated places is also provided in this thesis.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Public. Available at `https://dspace.cvut.cz/handle/10467/62094`, pdf document at `https://dspace.cvut.cz/bitstream/handle/10467/62094/F3-DP-2015-Salansky-Vojtech-Kontaktni%20pruzkum%20terenu%20pro%20mobilniho%20robota.pdf`

## 2.7 Tesfay (2015), "Terrain Type Classification Based on Sound"

**Bibliography** Tesfay Gebru, Abraham. "Terrain Type Classification Based on Sound." Master Thesis, Saarland University, July 2015.

**Abstract** The autonomous mobility of robots has great benefits to human exploring hazardous terrains. The motivation of this thesis is to detect different types of terrains traversed by a robot based on acoustic data from the robot-terrain interaction thereby helping to make the mobile robots more autonomous. The acoustic data was collected using a microphone mounted on our robot. Then, these recorded datasets were used to train classifiers so as to distinguish different terrain types from one another. Different acoustic features and classifiers were investigated, such as Mel-frequency cepstral coefficient and Gamma-tone frequency cepstral coefficient for the feature extraction, and Gaussian mixture model and Feed forward neural network for the classification. We analyse the system's performance by comparing our proposed techniques with some other features surveyed from distinct related works. Thus, we demonstrate the effectiveness of our approach using five different terrain classes which are trained using real data sets gathered from different ground surfaces. The experimental result indicates the average accuracy obtained is approximately 93.6% and it is enhanced to 95.2% with an increase in audio duration. In real applications, it is better to decrease the detection time and our system still has satisfactory performance using human-like terrain labelling even for smaller audio duration. These are very promising results which show that acoustics is an interesting domain that needs to be extensively explored to improve the autonomy of tracked robots.

**Relation to WP** This master thesis explores the use of acoustic features to characterise different terrain-types. Sound is a potentially interesting alternative environment perception modality and these features could provide useful complementary input to modelling the environment in WP1.

**Availability** Unrestricted. The PDF is downloadable at `http://www.dfki.de/web/forschung/publikationen?pubid=8232`.

## 2.8 Kong, Dong-Uck (2015), "Persistent Mapping of Dynamic Environments By Multiple Aerial Robots With Multi-Camera Systems"

**Bibliography** Dong-Uck Kong. "Persistent Mapping of Dynamic Environments By Multiple Aerial Robots With Multi-Camera Systems". Unpublished Dissertation Proposal, full document, January 2015.

**Abstract** Based on the requirements for the TRADR scenario, the central research question is formulated: How can particularly efficient mapping and localization methods be developed with respect to measurement quality, sensor data quantity and long-term data analysis for both robot and human action forces? This work proposes to use visual sensors, extensively reviews the state of the art methods in relevant domains, and conceptualizes an efficient system architecture for persistent map building. An overall system is suggested that provides multiple individual mapping levels which run with different speed on different platforms but use uniform methods to benefit from each other. Detailed concepts for basic mapping, map fusion in a single sortie, and map fusion for persistency over time are introduced.

**Relation to WP** This Proposal conceptualizes an efficient overall architecture for persistent mapping by visual sensors. It contributes to the tasks T1.4 Robot centric metrical maps and models storage.

**Availability** Restricted. Not included in the public version of this deliverable.

## 2.9 Kong, Dong-Uck (2015), "Layered Visual Perception Architecture for Efficient Multi-Density Environment Mapping and Localization"

**Bibliography** Dong-Uck Kong. "Layered Visual Perception Architecture for Efficient Multi-Density Environment Mapping and Localization". Unpublished Draft Paper, December 2015.

**Abstract** In this work, efficient multi-density environment mapping method is proposed that is realized by multiple visual perception methods hierarchically working together. We propose to use two rigidly connected visual sensors in combination with inertial sensors. A non-visual estimation layer based on IMU, a vision-based layer representing point-based world model, and a vision-based module representing edge-based world model are implemented. The suggested levels are individually complemented modules which are firstly algorithms independent from each other, but benefit from each other. Our main contributions are design of novel layered system architecture with bi-directional information flow, enhancement of map quality despite of low sensor rate, and robustness against fast camera motion.

**Relation to WP** This draft paper proposes multi-layer visual SLAM algorithm aided by inertial sensor. It contributes to the task T1.4 Robot centric metrical maps and models storage.

**Availability** Restricted. Not included in the public version of this deliverable.

## 2.10 Dube-ICRA2016, "Non-uniform sampling strategies for continuous correction based trajectory optimization"

**Bibliography** Renaud Dubé, Hannes Sommer, Abel Gawel, Michael Bosse and Roland Siegwart. Non-uniform sampling strategies for continuous correction based trajectory optimization. Submitted to IEEE International Conference on Robotics and Automation (ICRA) 2016.

**Abstract** Sliding window estimation is widely used for online simultaneous localization and mapping. While increasing the sliding window size generally yields improved accuracy, it also comes at an increase in computational cost. In order to reduce this cost, we propose smarter non-uniform sampling of the trajectory representation over the sliding window. This non-uniform distribution is possible with continuous-time representations that allow freely adjustable control vertex locations. We present two strategies for selecting the control vertices locations and evaluate them based on a real data laser-odometry SLAM problem. Our results clearly show that non-uniform distributions of control vertices can be superior to uniform distribution in terms of accuracy per computation time.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availability** Restricted, paper under review. Not included in the public version of this deliverable.

## 2.11 Gawel-ICRA2016, "Structure based Vision-Laser Matching"

**Bibliography** Abel Gawel, Titus Cieslewski, Renaud Dubé, Michael Bosse, Juan Nieto and Roland Siegwart. Structure based Vision-Laser Matching. Submitted to IEEE International Conference on Robotics and Automation (ICRA) 2016.

**Abstract** In multi-robot applications not every agent is equipped with the same sensor outfit. Registering maps created by different sensor modalities is a relevant step towards collaborative mapping. This work presents an approach for matching dense LIDAR point-cloud maps to densely reconstructed vision point-cloud maps using structural descriptors. The matching algorithm works independently of the sensors' viewpoint and lighting. We then present a novel approach for laser-vision place-recognition. We analyse a range of structural descriptors and present results of the method integrated in a larger pose-graph based mapping framework. Despite the fact that we match between the visual and laser domains, we can successfully identify place-matches using structural descriptors at a moderate precision of 72 % at 46 % recall at the best MCC scores and before geometric verification.

**Relation to WP** It directly contributes to T1.4 – Robotic centric metrical maps and models storage.

**Availability** Restricted, paper under review. Not included in the public version of this deliverable.

## 2.12 Gianni, Ruiz, Ferri, Pirri (2015), "Terrain contact modeling and classification for ATVs"

**Bibliography** M. Gianni, M. A. Ruiz Garcia, F. Ferri, F. Pirri. "Terrain contact modeling and classification for ATVs". Accepted to the IEEE International Conference on Robotics and Automation (ICRA'2016), January 2016.

**Abstract** We consider some specific ATV robot model in which the active parts of the tracks are flippers, and present a method that models the flippers terrain contact. The main idea is to extract the residual component, as formed by a disturbance due to the non modeled dynamics of the moving base link and an unexpected disturbance. We extend the FDI approach to comply with ATVs dynamics and, in particular, their flippers component. Under the hypothesis that the residual signal presents disturbance patterns that can be discriminated by those generated by unexpected high-frequency collisions of the flippers with the ground, we apply a classification method to recover the flipper contact. The wavelet packet transform is used to decompose the signal and generate from the different subbands a feature space. Finally, sparse SVM, based on feature selection discriminates the contact signal.

**Relation to WP** It directly contributes to T1.3 – essential sensing and UGV control functionality.

**Availablity** Restricted. Not included in the public version of this deliverable.