The EU Framework Programme for Research and Innovation H2020
Research and Innovation Action

Deliverable D5.1 CENTAURO Navigation Concept

Dissemination Level: Public

Project acronym:	CENTAURO
Project full title:	Robust Mobility and Dexterous Manipulation in Disaster Response by Fullbody Telepresence in a Centaur-like Robot
Grant agreement no.:	644839
Lead beneficiary:	KTH – Kungliga Tekniska Hoegskolan
Authors:	P. Jensfelt, J. Folkesson, G. Meneghetti, M. Felsberg, and Sven Behnke
Work package:	WP5 Navigation
Date of preparation:	2015-09-22
Type:	Report
Version number:	1.0

Document History

Version	Date	Author	Description
0.1	2015-08-21	P. Jensfelt and J. Folkesson	First draft of sections 1, 4 and 5
0.2	2015-08-27	G. Meneghetti	Implementation plan for SLAM
0.3	2015-09-01	P. Jensfelt	First almost complete draft
0.4	2015-09-02	P. Jensfelt and S. Behnke	Implementation plan for Task 5.3, more details to list of consumers of results in section 1.2 and more constraints on terrain map.
0.5	2015-09-22	J. Folkesson	Comments from IIT
1.0	2015-09-30	S. Behnke	Submitted version

Executive Summary

This deliverable describes the navigation concept for the CENTAURO system.

The core representation in the navigation concept is the spatial representation. It holds the knowledge that the system has about the world. We proposes a layered spatial representation consisting of the following three layers.

- The **global map** serves as the frame of reference for the system. It is a graphical model which is not necessarily globally consistent at a quantitative level but is consistent topologically and can therefore support global mission planning. It allows the operator to orient himself globally using image key frames or similar but it does not provide the means for a full 3D reconstruction of the world.
- The **local navigation map** represents the world in 3D or 2.5D if deemed enough. It gives the operator a 3D view of the environment and supports terrain classification and navigation planning. These local maps are anchored into the global map and can be used to derive a representation for the nodes in the global graph structure to support for example place recognition.
- The **local manipulation map** covers the space near the robot and is only created on demand. It has high enough resolution to support manipulation operation (planning, simulation, execution, etc).

The navigation concept also includes a novel navigation execution system. The unique feature is this is that it exploits the dual locomotion mode setup of the robot and allows both driving and walking.

Contents

1 Introduction

This deliverable reports on the conceptual design for the navigation in the CENTAURO project as specified in Grant Agreement [44]. This deliverable and the corresponding deliverables in the other work packages constitute the specification for how to design the system and its components. They will not be able to provide answers to everything since CENTAURO is a research project. However, these deliverables should ensure that the pieces will come together as smoothly as possible to form a working system. This requires, for example, that interfaces between the work packages / components can be identified and agreed upon. Assumptions made about the performance of certain parts of the system made in other parts of the system must be matched. Given the early stage in the project it is unlikely that everything stated in these deliverables will be accurate and that all assumptions will hold, but without making any decisions and assumptions no real progress on a project level can be made.

The objective of this particular deliverable is to ”Report on the concept for hybrid wheeled-legged navigation of the CENTAURO robot in rough terrain.”. It belongs to work package 5 Navigation (WP5). Contributions to this deliverable and the work to come, come primarily from KTH, LIU, UBO and IIT.

The overarching goal of the navigation work package (WP5) is to develop methods to support the operator when navigating in rough terrain. The developed methods should support and exploit the fact that the CENTAURO platform has both wheels and legs. That is, both driving and walking modes of locomotion are supported. Driving on wheels caters for speed when the ground is even and walking allows the CENTAURO robot to traverse rough terrain. As a start the operator would determine which mode to move the robot in manually, but later in the project the CENTAURO system should be able to handle more and more of the switching. The system should also adapt it posture to, for example, provide the largest possible stability when needed or reduce the overall width to pass through a arrow opening.

To support such rough terrain navigation in the context of a search and rescue mission, the CENTAURO system should construct a model of the environment as the robot moves. The mapping system must be able to map a large environment which has been completely or partly affected by a disaster like an earthquake. To assist the operator and perform some of the low-level tasks associated with navigation autonomously, the system must be able to assess the terrain for navigability. The methods for mapping and terrain classification must be efficient enough to support real-time execution. To reduce the cognitive load from the operator the system should also, ideally, learn models for terrain classification by fusing raw sensor data and operator input. This way, novel operating conditions can be dealt with and the system would be able to reduce the need for supervision as the mission progresses.

There are three important aspects to consider in these deliverables.

1. What information / inputs do we make use of and what do we require from these inputs. This gives constraints on other parts of the system, including sensors.

2. What outputs and actions are we expected to produce. These should be designed to fulfill requirements from other parts of the system and the system as a whole.

3. What methods and representations are used ”under the hood” to support the outputs and actions.

From a development perspective the two first items are needed to make sure that the pieces fit together and that the functional requirements on the system are met. The last one is important to consider to determine how to achieve this and to make sure that the assumptions are more firmly grounded.

1.1 Inputs

The inputs needed are listed in the table below. A more detailed description of this is provided in Section 4. In addition to what is below we require the specification for the work packages that this work OBpackage interacts with in order to design the interfaces.

What	When	From Whom	Nature
Existing sensor data similar to CENTAURO data	M6	UBO	RGBD
Existing sensor data via Central World Model (CWM)	M8	RWTH	RGBD
Local (as in the partner) data for local test	on demand	all	varies
Real data from CENTAURO sensor (not on final system)	M12	IIT	
Real data from CENTAURO sensor via CWM	M14	RWTH	
Real data from CENTAURO system	M30	IIT	
Real data from CENTAURO system via CWM	M30	RWTH	

Table 1: The inputs to the navigation workpackage.

1.2 Consumers

The users of the results are listed in the table below. Here we focus on consumers external to this work package. A more detailed description of this is provided in Section 4.

What	When	To Whom	Nature
WP4 Virtual Testbed	M12	RTWH	Local 3D models from CWM data
WP4 Virtual Testbed	M15	RTWH	Data structure for global map
WP4 Virtual Testbed	M18	RTWH	Global maps from simulated data
WP6 Manipulation	M12	UBO	Detailed local 3D model from local data
WP6 Manipulation	M18	UBO	Detailed local 3D model from CWM

Table 2: The external consumers of results from the navigation work package.

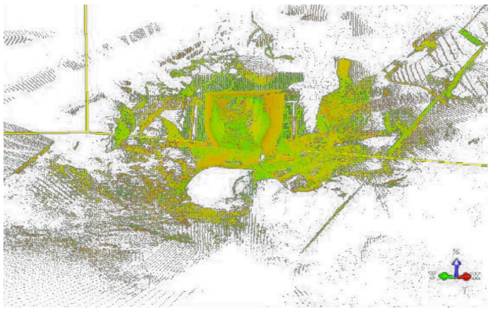
(a)  (b) 

Figure 1: (a) Rough terrain in DisasterCity, Texas, that is almost inaccessible by state-of-the-art systems. (b) 3D map acquired by range sensing.

2 Related Work

Disaster scenarios create complex 3D environments, as shown in Fig. 1. Many methods for motion planning in 3D are not applicable to platforms with multiple degrees of freedom due to limits on computation and payload. Urban Search and Rescue (USAR) represents a particularly hard case as real-time acting is necessary. As a result, many systems deployed within USAR scenarios and its testing arenas are either completely teleoperated or based on mapping and navigation in 2D [24, 25, 39]. Given a 2D map of the environment, one common approach is to plan the shortest trajectory in configuration space and to execute this trajectory via PD control. However, in harsh environments the safest path rather than the shortest path may be desired. Wirth *et al.* introduced an exploration strategy and path planner that utilizes occupancy grid maps for combining the distance transform (computing for each grid location the distance to the closest target) and the obstacle transform (computing for each grid location the distance to the closest obstacle) when planning a path to several exploration targets at the same time [64]. Consequently, the method selects the safest alternative consisting of target location and path to reach the target. Preliminary extensions towards exploration in 3D were introduced by Dornhege *et al.* [10].

When robots operate on rough terrain, it is important to consider the shape of the surrounding terrain and how the motion of the robot might influence this. Long-term motion planning in 3D is on most mobile platforms still computationally too expensive and only possible at low frequencies due to slow update rates of existing 3D mapping solutions. Therefore, several researchers introduced solutions with short-term look ahead that execute specific robot behaviors with respect to the current situation of the robot. Okada *et al.* introduced an autonomous controller for tracked vehicles that is based on continuous three-dimensional terrain scanning. The system obtains 3D terrain information around the robot by LIDAR sensors. The 3D map acquired around the robot serves as basis for a controller adjusting the position of sub-tracks on the robot (also sometimes referred to as "flippers") and thus enabling smooth navigation over rough terrain [35]. Magid *et al.* introduced a system for keeping the robot maximally stable at every step of its path while allowing the vehicle to loose balance in a controlled manner for facilitating safe climbing over debris [27]. Sheh *et al.* developed a method for behavioral cloning, a type of learning by imitation that produces control rules that clone the skills of an expert human operator. The method has been used for optimizing an autonomous controller that finally performed at a level comparable to that of a human expert when overcoming rough terrain [50]. Dornhege *et al.* introduced the concept of behavior maps, which is a framework to link certain robot behaviors on rough terrain such as climbing over stairs and obstacles, with structures detected from 3D point clouds. Both mapping and map classification were executed

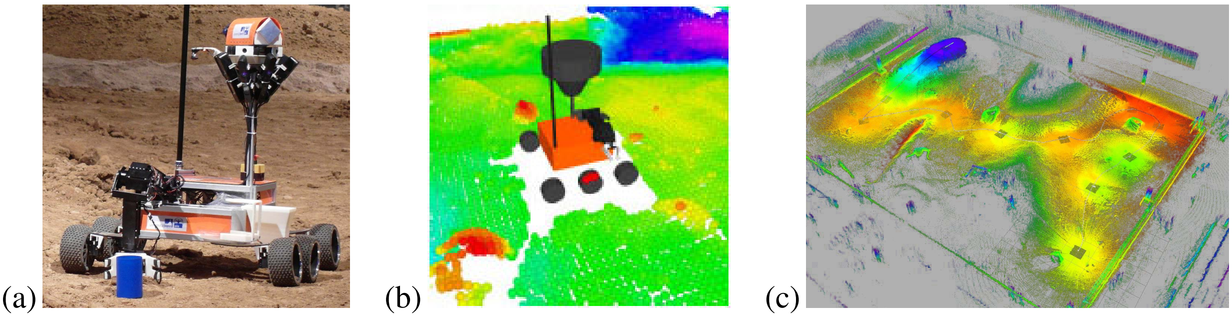


Figure 2: Explorer robot of University of Bonn at DLR SpaceBot Cup 2013 [57]. (a) grasping a water-filled cup; (b) omnidirectional terrain model from depth cameras; (c) 3D map of arena created with laser-based SLAM.

in real-time [9].

Murphy first demonstrated the deployment of shape-shifting robots in the USAR environment [31]. Shape-shifting robots can change configuration to adapt to surroundings when navigating through voids. Murphy showed in a case study at a collapsed house the need for robots adjusting their shape in order to access differently shaped voids [31]. Most shape-shifting robots that are designed nowadays are tracked and feature additionally four inclinable flippers mounted at the corners of the vehicles. For example, the robot *Quince*, developed by Japanese universities, was finally deployed at the Fukushima Daiichi nuclear power plant in the aftermath of the massive earthquake and tsunami that hit eastern Japan in 2011. The robot was used for inspection missions in highly contaminated areas [32]. There were several mechanisms of hybrid legged-wheeled platforms presented in the past [19, 59, 17]. For example, Takahashi *et al.* presented a mechanism for simultaneously executing wheel-mode and leg-mode [59], and Halme *et al.* introduce *rolking* (rolling-walking) to facilitate combined legged and wheeled locomotion for gaining effective natural terrain mobility [17]. While these mechanisms performed well under the mode of teleoperation, no integration with methods for situation awareness and motion planning were presented.

Algorithms for planning foothold selection and footsteps have been introduced in [3, 2]. They describe a system for real-time building of a local elevation map from 2D range measurements that supports a foothold selection algorithm employing unsupervised learning. In [2] a polynomial-based approximation method for creating decision surfaces is shown. The authors show that planned footholds enable the robot to walk more stably, avoiding slippages and falldowns. A geometric feature-based footholds identification system is described in [66], where candidate location for footholds are identified by support vector machine (SVM). In recent work, UBO developed hybrid driving-stepping locomotion for their mobile manipulation robot Momaro [47]. This robot was used with great success in the DARPA Robotics Challenge [45].

The techniques for Simultaneous Localization and Mapping (SLAM) are becoming quite mature and the problem of mapping simpler environments can be considered solved. Fig. 2 shows an example. In the last half decade there has been a lot of work on 3D mapping, using actuated laser scanners [33, 4, 11, 57] and cameras [38, 41, 5]. The recent development of affordable RGB-D sensors have created an even bigger effort in the direction of 3D mapping. Graphical models [13, 8, 21] and non-linear optimization [26] are now often used for mapping rather than traditional filtering techniques. One of the first methods making use of the new RGB-D sensor with great impact was the KinectFusion algorithm [20] which makes use of a limited TSDF (truncated signed distance field) volume for representing and fusing the information. Work such as Kintinuous [63] continue this development and extend it to larger volumes. Many recent methods also combine the geometric information with photometric information in the pose optimization [22]. In [62] the ICP algorithm from KinectFusion is combined with the approach of Steinbrücker et. al [55]. When working with dense 3D models the amount of data that needs to be stored becomes a problem when the models get bigger. Approaches to deal with this include multiresolution surfel maps [56] and discretization in the form of oct-trees [65] and the use of elevation maps [23, 40]. When robots move beyond simple navigation tasks geometry is not enough. In [34, 60, 58, 29, 37] the 3D model is extended with semantic information such as what parts of the model correspond to walls, floor, chairs, etc. In [16] RGB-D data is used to build large scale 3D maps augmented with recognized objects. The recognition is based on clustered planar regions and the objects are refined by replacing them with their corresponding CAD models. Recently, deep learning methods have been applied with great success for RGB-D object recognition [48].

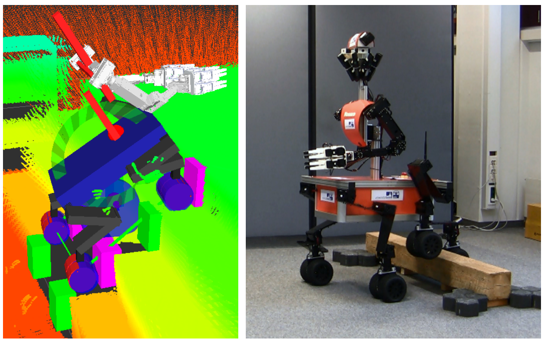


Figure 3: **Left**: Momaro climbing stairs in simulation. The green and purple boxes indicate detected obstacles which constrain the wheel motion.
Right: Momaro stepping over a wooden bar obstacle.

3 Progress in CENTAURO beyond the State-of-the-Art

In the CENTAURO project, we will make use of a hybrid mapping scheme where a dense 3D world model will be estimated on a local scale to support the online perception for mobility (terrain classification, assessment of traversability, etc) and a sparse graphical model will be used at a larger scale to ensure global consistency and for storing the explored part of the environment. On the local scale, our emphasis and contributions, will be on fast and efficient methods to register sensor data using a combination of geometric and appearance information allowing real-time mapping. On the larger scale, we will investigate how to use the dense local maps to generate powerful descriptors, including high level semantic information, to be used for robust data association in the graphical model. The starting point will be work like FabMap [7] and its follow ups and extensions [6, 36]. On both local and global scale, we will make use of an active perception strategy to allow the system to autonomously direct sensors to regions where more information is needed to improve the performance of the mobility system and to close loops [54, 14, 28].

We also contribute methods for semi-supervised learning of models for terrain classification where we will bring together the work on semantic mapping and more traditional terrain classification. For this we will extend our previous work on semantic mapping [42, 43] and spatial reasoning [53, 52, 1] and build on existing work [16] to match point cloud clusters to models of objects and structures such as beams and pipes to include in the map building on work such as [46, 51, 18, 12]. Probabilistic graphical models will be used to fuse the information from the low level terrain properties and the higher level semantic information. The semantic information will also help relieve some of the cognitive load from the operator. Finally, we will investigate how to incorporate information from CAD models of the environment (if available), as a prior for the mapping process, knowing that large parts of it may have undergone drastic changes.

The CENTAURO project will also advance the state of the art in autonomous and semi-autonomous navigation planning in rough terrain. In wheeled locomotion mode, the CENTAURO robot will be able to drive omni-directionally on flat surfaces. For legged locomotion, the system will execute body motions of maximal stability while moving into a target direction. The CENTAURO system will combine active perception of structures that can be utilized for stabilizing the posture of the platform such as footholds and supporting planes, with a physics simulation-based prediction of the outcomes of certain body motion commands.

Within the CENTAURO project, a new type of behavior system will be developed that, similar to the concept of affordances [15, 30], identifies stable body motion trajectories that are executable in the environment within a short time horizon. The system will be able to rank the terrain in terms of navigability and suggest paths for the operator to select from.

Special walking gaits will be developed for steps and stairs as well as behaviors for stabilizing the platform with detected structure elements such as foot holds. The CENTAURO system will allow autonomous navigation by planning obstacle-free paths in both wheeled and legged locomotion modes. It will maintain its body pose and adapt to the perceived height differences and slopes of the terrain.

4 Navigation Concept

This section describes the navigation concept for the CENTAURO system. Locomotion is based on omnidirectional driving and making steps when necessary. Initial experiences made with the mobile manipulation robot Momaro, shown in Fig. 4, were taken into account.

4.1 Requirements and Assumptions for Internal and External Consumers

We will start by analyzing the internal and external requirements posed on the navigation work package and discuss assumptions where applicable. The primarily external consumers of the results produced in the navigation work package are shown in Table 2. In addition to this there are internal requirements to realize autonomous execution of navigation in rough terrain making optimal use of wheeled and legged locomotion.

In the following sections the different consumers will be discussed and the requirements placed on the work in WP5 will be listed. These requirements will be of different types and at different levels. Especially for the internal requirements the hierarchical nature of the requirements will be clear.

4.1.1 WP5: Autonomous Execution of Navigation

The main functional output from the navigation work package is support for the operator to execute motions from A to B in an efficient way. Search and rescue missions are time critical. Therefore, when designing the representations and algorithms the resolution, accuracy, etc of these should be kept sufficient to solve the task, but not more.

We assume that the map does not have to be metrically consistent at a global level and that loop closure at the global level only needs to result in topological adjustments to the map. We will re-localize using denser information such as panoramic sensor data acquired at the corresponding node. We also need to be able to start a mission with an already existing map and register new map information to the existing map.
The requirements are that we can

- assess the terrain to tell where the robot can move with each locomotion mode,
- assist the execution of the motion from A to B, and
- start a mission with a map from a prior mission.

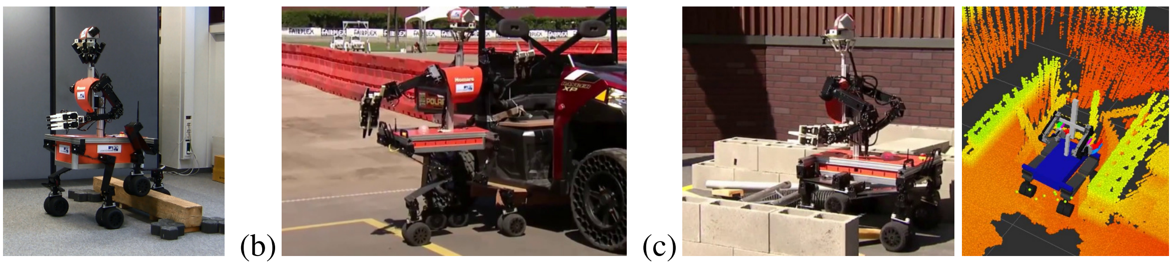
(a) (b) (c)


Figure 4: Mobile manipulation robot Momaro of University of Bonn [47]: (a) stepping over a bar; (b) climbing out of a car; (c) traversing debris.

4.1.2 WP3: Operator Interfaces

We make the assumptions that it is enough to be able to visualize the nearby surroundings of the robot in detail and with geometric information and that it is enough that the operator can get 2D projections of what the environment looks like. This way an operator can plan missions on a global scale by indicating the direction to move in there and give more precise commands at a local level. One way to think of this is that at a local level there should be geometry and texture but further away a textured ”view sphere” is enough.

When the robot is started in a new environment there will be no 3D at all as the robot has not acquired any information yet. The only available information for the operator is very local 3D information (sensor range) and long range 2D information (images). It will be necessary at this time to use a combination of 2D and 3D information in the operator interface. The local area will become larger and more complete with time. We then propose that we make use of this scheme throughout the mission to save bandwidth and computations.

Even though the information from WP5 will pass through the WP4's Virtual Testbed to WP3, the ultimate driver of the requirements below are WP3. To fulfill indirect requirements from WP3, WP5 has to support

- rendering accurate views of the environment from different view points at a local level
- rendering views that support commands about the direction to move at a global level

4.1.3 WP4: Modeling and Simulation

WP4 maintains a model of the world and provides a Virtual Testbed. WP4 acts as the central hub for information exchange in the system. Sensor data flows via it to the navigation system for example. The spatial part of the world model is based on both prior information and the map built by the robot from sensor data. The model supports simulations. These simulations can be used both to replace real hardware and to evaluate different plans before they are executed on the real robot. WP4 needs to be able to update the underlying model based on information from WP5. Our assumption is that for modeling real-world missions it is sufficient that the simulator only covers the local region around the robot in 3D and that the world further away is represented without geometry. The size of the local region and where the further away region starts will be task dependent. This assumption coincides with the assumption for the operator interface. To support WP4, WP5 needs to

- maintain a world model that supports (physics) simulation at the local scale
- maintain a representation at the global level which supports orienting oneself into the world

4.1.4 WP6: Manipulation

In the manipulation work package it is important to have accurate 3D models of the local environment to recognize objects and structures, to assess potential grasps, etc. In addition to this the following must be fulfilled.

- The model of space must be able to incorporate information from multiple sensors and types of sensors over time to improve perception of objects

- The robot base must be placed at appropriate locations in order to be able to perform object pick and place tasks[1].

4.2 Spatial Representation

The core of the navigation concept is a representation of the environment that supports the needs from inside the navigation work package, i.e., robust mobility and localization, and from the outside in the form of visualization, simulation and manipulation. The size of the environment under consideration and the requirements on the level of details posed by manipulation prohibits a single metric 3D map with uniform resolution. We will therefore consider the design of the spatial representation in terms of three layers

- Global map which covers the size of the environment
- Semi-local map supporting navigation planning and execution
- Local map supporting manipulation

4.2.1 Design

1. At the largest scale there is a global graphical model which is globally consistent in the sense that it supports navigation and decision making over the entire environment of interest. However, we put no constraints on it being metrically globally consistent. We assume that accurate geometric information is not needed for global mission planning but edge traversal information is needed. The global map supports
    - a global reference to which more local information can be registered
    - place recognition so that the robot is able to re-localize to previously visited places in the environment.
    - support global mission planning
    - the global map will be updated much slower than the local map. It will update as new information is available which is typically at the transitions between local maps.
    - the resolution of the global map will be about the size of the local maps.
2. At the second level in the spatial representation there is a semi-local 3D navigation map which is a sliding window supported by raw data saved for a short time. The scope of this map is large enough to support local / detailed navigation planning. When a new topological node is added to the global map, the sliding window data is used to generate a representation of the node. For efficiency reason, a representation in two 2.5-dimensional maps (elevation of ground and ceiling) might be considered.

    The semi-local navigation map

    - provides the operator station with the 3D model used to render a view of the robot's surrounding
    - supports terrain classification

[1] This can be addressed by including robot base motions in inverse kinematics and resolving redundancies according to prioritized control goals [49]. The assessment of potential robot base poses can be accelerated by precomputing an inverse reachability map [61].

3. The third level is a highly detailed 3D model catering to the needs of manipulation. Its extent is roughly defined by the workspace given by the reach of the robot's arms. The raw data from the second level's sliding window can be reprocessed with different parameters to get the manipulation map. Please refer to Deliverable D6.1 for more details. The manipulation map
   - is created on demand, i.e., when we want to perform manipulation operations
   - makes use of active sensing to acquire the views needed to build the model
   - The resolution/accuracy of the model will vary from millimeters to centimeters dependign on the needs and the sensing limits.

4.2.2 Terrain Map

The local map at the second level is used to generate a terrain map which in turn is used by the path planning (where to go but not how exactly) and the switching of locomotion mode. Information about navigability is maintained in a representation where each x,y position (cell) contains information about the traversability ("height constraints", "walk here", "stair steps", "need to step over", etc).

The terrain map must contain sufficient information to support full body navigation planning by being able to support

- assessment of the foot holds and
- assessment of if these foot holds can be connected by driving or if it is possible to make a step from one foot hold to the next.

1. On the scale of *individual wheels* (or wheel pairs), nodes correspond to possible foot placement locations, with costs indicating accessibility and' ' stability of foot holds, which are derived from geometric features, such as horizontal support surfaces, and surface material properties. Edges connect foot holds which can be directly reached by driving or by making a step, with costs indicating the effort of moving the foot from one node to the other.
2. On the scale of the *entire robot base*, nodes correspond to robot base poses and costs indicate the stability of the pose and the avoidance of obstacles. Each base node will correspond to four foot nodes (for the four robot legs) that are reachable from the base pose. Edges connect base nodes which can be reached by a combination of driving of the wheels and possible steps of a single foot, as indicated by the edges in the foot graph and verified using kinematic and balance constraints for the entire robot.

4.3 Navigation Execution

One of the main advantages that the CENTAURO platform offers, is that it supports both driving and walking. Driving allows the robot to move fast and efficiently when the ground allows it. When the terrain gets rougher it can use its legs to walk. Walking is slower and requires more power but in return much rougher terrain can be handled.

The operator is taking an active role in the operation of the robot at all times but the navigation system should relieve him from some of the cognitive load by helping with some of the low-level navigation capabilities. Depending on the properties of the terrain this results in different levels of support. On completely flat terrain the operator will be able to command the robot to move to a location. As the terrain gets rougher and driving is no longer possible the robot has to start walking. If the terrain contains only a few small and isolated obstacles, the long term aim is that the operator should be able to control the motion of the robot by specifying, for example, direction and speed. The system should then be able to plan where to put the feet down. As the terrain gets even tougher the operator will have to take control at a lower and lower level and the CENTAURO system will at this point assist with decision support functionality. The system should be able to recognize its limits and ask for help from the human operator, e.g. because terrain data is not available or because no feasible plan can be found.

It is the operator that decides what mode of locomotion to use with the help of suggestions from the system. The operator would also be the one that decides what foot print the robot should use, based on suggests from the system.

4.4 Sensing Requirements

We need to be able to perform accurate dead-reckoning while moving in rough terrain. This will require input from joint and wheel encoders, IMU(s) and exterioceptive sensors such as cameras to perform visual odometry. It is important that the sensors can operate both indoor and outdoor in the same mission and that different lighting conditions can be dealt with.

- IMU(s) for high bandwidth estimation of the relative body motion
- encoders on wheels for odometry
- encoders on joints to determine configuration and legged odometry
- joint torque sensors and if needed force-torque sensors on the hands and legs.
- omni-directional vision to support orientation invariant place recognition
- lidar sensor(s) for robust 3D perception[2]
- RGB camera(s) for terrain classification
- camera in the hand for mapping for manipulation and object perception

[2] such as the VLP-16 from Velodyne

5 Implementation Plan

This section provides a plan for the implementation of the concept. We present this in the form of the involved tasks. At a high level we foresee that the work on Task 5.1 and 5.2 will be highly integrated between KTH and LIU. KTH will focus mainly on the terrain classification to start with and the global map whereas LIU will focus on the local metric maps. IIT and UBO will work on the full body navigation capabilities which will result in functionality to support walking and driving with a relatively high level command interface which can then be used when executing the navigation and performing the locomotion mode switching.

5.1 Rough Terrain SLAM (Task 5.1)

A high-quality, intuitive visualization of the CENTAURO robot surroundings supports the operator to better understand the disaster area. The variety of disaster scenarios and manipulation tasks that the system will face suggest that it is best to follow a dynamic approach in the process of creating spatial maps. As described in Sec. 4, three layers (or levels) of representation will be investigated and proposed for the CENTAURO system to address the needs of different methods.

The development of multi-level environment maps will follow a three step approach. In the first step we will investigate the intermediate level of the spatial representation. The purpose of this map is to help the operator to perform local motion planning. The intended level of detail here could be realized with a visual SLAM algorithm executed on the CENTAURO robot. The quality, in terms of density, of the visualization will depend on the sensors mounted to the robots. It will be generated according to the motion of the robot in a "sliding window" procedure. The range of the area that will be captured has still to be discussed, since it will depend on different factors related to the mission. However, the evaluation process will give feedback for further decisions and iterative improvements.

For efficiency reasons, in the second step a 2.5-dimensional occupancy (or height) map representation will be considered for the intermediate level. It will consist of an elevation map of the scene, estimated from the computation of the height from the ceiling and from the ground.

Alongside, the development of the third level will be performed in the second step. A highly detailed map that will mostly be used for manipulation tasks in work package 6 [44]. However, it can also be used to navigate in dangerous areas in a more detailed scenario. The intended idea is to have a dense structure from motion pipeline at very low range. This will be evaluated in combination with other functionalities and we may deviate from this idea in case of bandwidth problems.

In a third step we will implement a global model of the disaster environment. This map will be created from the starting position of the CENTAURO robot and it will define a global reference frame for mission planning. In contrast to the previous maps, the level of detail is very low, but it will contain information on all areas already explored. For efficiency, key-frames with the CENTAURO robot poses as well as panoramic views will be kept to allow back tracking of the robot steps along the path. The implementation plan for the spatial representation is summarized in Table 3.

For all three steps we will have feedback on the performance of these maps with the other functionalities (Navigation, Manipulation, Terrain classification, ...). This is intended to guide further iterations on the map development.

5.2 Terrain Classification (Task 5.2)

We will develop the terrain classification to support autonomous execution of navigation and to provide semantic information for visualization to the operator. We will mainly focus on the navigation execution as the evaluation task. We will proceed with the development in steps.

1. In the first step we will make use of geometric information to classify the terrain. We envision three classes at this stage.
   (a) Drivable
   (b) Walkable
   (c) Non-navigable

   We start by operating on single sensor frames and use a point cloud as input format. Later we input the local 3D map from the spatial representation as it becomes available.

2. In the second step we will incorporate color and texture information, to refine the classification. At this stage we foresee the following decision mechanism:

Classes	Distinguishing features
Walkable and drivable	Flatness (geometric constraint)
Drivable and non-navigable	Material (e.g. mud vs tarmac)
Walkable and non-navigable	Stability (color, texture and geometry)

   We will evaluate a deep learning approach to the classification problem.

3. In the third step we will release assumptions about independent cells and incorporate spatial correlations using, for example, CRFs.

4. Later on we imagine working on some of the following issues
   - Incorporate object recognition / classification into the map. This extends the work in the second step by incorporating more high level information.
   - Supervised learning where an operator can provide input to the classification. The vision is that the operator can help with labeling when new categories / classes are encountered or the conditions have changed significantly so that the classification no longer works.
   - The are several strands of active sensing i) change sensor view to refine models, ii) use your feet to test the ground support, iii) carefully drive though stuff that might move away.

Table 3: Implementation plan for the spatial representation maps.

Step	Brief Description
Step 1	Intermediate level 3D map (navigation)
Step 2	Two tasks : third level highly detailed 3D map (manipulation) and intermediate level 2.5D map (navigation)
Step 3	First level 3D global map (mission planning)

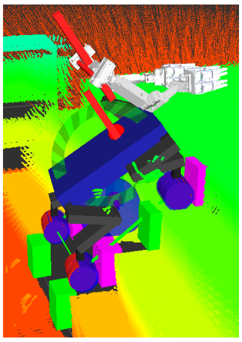
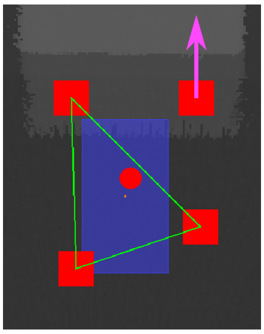
(a)  (b)  (c) 

Figure 5: Mobile manipulation robot Momaro of University of Bonn [47]: (a) Climbing stairs in simulation. The green and purple boxes indicate detected obstacles which constrain the wheel motion. (b) 2D height map of Momaro standing on the first of two steps. The robot is in stable configuration to lift the right front leg. Red rectangles: Wheel positions, red circle: COM, blue: robot base, green: support polygon. (c) The right front leg is lifted and placed on the next step.

5.3 Full-body Navigation Planning (Task 5.3)

We will develop in this task a motion planner for the robot base supporting driving as well as stepping locomotion modes, based on the initial experiences made with the mobile manipulation robot Momaro, shown in Fig. 5.

Input to the navigation planner will be a current robot pose (estimated by localization with respect to the created 3D environment map), the navigation goal (specified as desired robot pose in the map frame), and a navigation graph representation derived from the terrain map on two scales as described in Section 4.2.2.

Initially, we will realize omnidirectional driving on flat or sufficiently even surfaces, where no steps are required. Robot base poses will be planned in 3D (x, y, θ), for given height (z) and zero (upright) pitch and roll angles. The above navigation graph will be created from the terrain map and a cost-optimal sequence of nodes will be determined by A* search.

In a second step, we will extend the navigation planner by allowing stepping for individual feet while the robot is statically stable on three legs. For this, we will realize stepping primitives, which move the center of mass projection of the robot into the remaining support triangle, execute the step with the unloaded leg, and shift the weight back to the center of the four-foot support polygon.

In the third step, we will extend the planner by considering also the base height z and the angle ϕ between the vertical robot axis and the largest local terrain slope as parameters of the base pose that are considered for navigation planning. For the shift of the center of mass projection, we will include upper body and arm motions. In this way, we will extend the applicability of the navigation planning to sloped terrains, for which the base might need to be shifted to maintain balance.

If the planner has insufficient terrain information to find a robot trajectory to the navigation goal, boundary cells of the known map area will be indicated and paths to them can be planned in order to obtain more measurements of the unknown terrain from these views. An exploration heuristics will be devised that suggests a next best view pose with low navigation costs from the current pose and a low heuristic cost towards the navigation goal.

In order to incorporate the most recent sensory information about the terrain map, the estimated robot pose, and perceived obstacles, the navigation planner will regularly plan obstacle-free paths in both wheeled and legged locomotion modes while the robot is navigating.

We will realize autonomous and semi-autonomous operation modes through operator interfaces developed in T3.3, T3.4, and T3.5. In semi-autonomous operation, qualitatively different navigation plans will be presented for the operators to select among, similar to suggestions issued to car drivers by driver-assistance systems in modern cars. Planned paths can also be shown to the operator in direct control mode, if the navigation goal is known to the planner.

5.4 Autonomous Execution of Navigation (Task 5.4)

We will proceed in steps in the development of the execution of navigation.

1. We will start by treating the system as a wheeled robot and ensure that we can provide mobility in flat terrain. We will use the terrain classification and navigation map information to support planning of drivable paths and will execute these.

2. As a second step we will introduce the ability to switch between walking and driving where walking is controlled by the operator at a low level. This ensures that we have a system early on with the full complete mobility capabilities, although putting a lot of strain on the operator during walking.

3. Using the system developed in the second step as a benchmark, we will investigate how to best improve performance (as defined by WP8) of the system in terms of mobility by better support to the operator along the lines defined in Section 4.3.

6 Conclusions

We have presented a navigation concept. At the center of this is a spatial representation with three layers, global map, local navigation map and a local manipulation map. The latter is only created on demand.

We also presented a concept for using the spatial model to plan and execute navigation actions. The system will provide low-level support to the operator to perform navigation. The aim is that navigation is carried out in a way that is offers the best trade off between performance of the system and strain on the operator.

References

[1] Alper Aydemir, Andrzej Pronobis, Moritz Göbelbecker, and Patric Jensfelt. Active visual object search in unknown environments using uncertain semantics. *IEEE Transactions on Robotics*, 29(4):986–1002, August 2013.

[2] D. Belter, P. Łabęcki, and P. Skrzypczyński. Map-based adaptive foothold planning for unstructured terrain walking. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 5256–5261, May 2010.

[3] Dominik Belter and Piotr Skrzypczyński. Rough terrain mapping and classification for foothold selection in a walking robot. *Journal of Field Robotics*, 28(4):497–528, 2011.

[4] D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, and J. Hertzberg. Globally consistent 3d mapping with scan matching. *Robotics and Autonomous Systems*, 56(2):130–142, 2008.

[5] L.A. Clemente, A.J. Davison, I. Reid, J. Neira, and J.D. Tardós. Mapping large loops with a single hand-held camera. In *Robotics: Science and Systems*, 2007.

[6] M. Cummins and P. Newman. Highly scalable appearance-only SLAM–FAB-MAP 2.0. In *Proc. Robotics Science and Systems (RSS'09)*, 2009.

[7] Mark Cummins and Paul Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research (IJRR)*, 27(6), 2008.

[8] Frank Dellaert. Square root SAM. In *Proceedings of Robotics: Science and Systems*, Cambridge, USA, 2005.

[9] C. Dornhege and A. Kleiner. Behavior maps for online planning of obstacle negotiation and climbing on rough terrain. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3005–3011. IEEE, 2007.

[10] C. Dornhege and A. Kleiner. A frontier-void-based approach for autonomous exploration in 3d. In *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*, pages 351–356. IEEE, 2011.

[11] David Droeschel, Jörg Stückler, and Sven Behnke. Local multi-resolution representation for 6D motion estimation and mapping with a continuously rotating 3D laser scanner. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 5221–5226, 2014.

[12] Felix Endres, Christian Plagemann, Cyrill Stachniss, and Wolfram Burgard. Unsupervised discovery of object classes from range data using latent dirichlet allocation. In *Robotics: Science and Systems*, page 34, 2009.

[13] John Folkesson and Henrik Christensen. Graphical SLAM - a self-correcting map. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'04)*, New Orleans, LA, USA, 2004.

[14] Simone Frintrop and Patric Jensfelt. Attentional landmarks and active gaze control for visual SLAM. *IEEE Transactions on Robotics, special Issue on Visual SLAM*, 24(5), 2008.

[15] JJ Gibson. The concept of affordances. *Perceiving, acting, and knowing*, pages 67–82, 1977.

[16] Martin Gunther, Thomas Wiemann, Sven Albrecht, and Joachim Hertzberg. Building semantic object maps from sparse and noisy 3d data. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2228–2233. IEEE, 2013.

[17] A. Halme, I. Leppänen, S. Salmi, and S. Ylönen. Hybrid locomotion of a wheel-legged machine. In *3rd Int. Conference on Climbing and Walking Robots (CLAWAR'00)*, 2000.

[18] Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d object discovery via multi-scene analysis. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4850–4856. IEEE, 2011.

[19] S. Hirose and H. Takeuchi. Study on roller-walk (basic characteristics and its control). In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 4, pages 3265–3270. IEEE, 1996.

[20] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.

[21] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Incremental smoothing and mapping. *Robotics, IEEE Transactions on*, 24(6):1365–1378, 2008.

[22] C. Kerl, J. Sturm, and D. Cremers. Dense visual slam for rgb-d cameras. In *Proc. of the Int. Conf. on Intelligent Robot Systems (IROS)*, 2013.

[23] A. Kleiner and C. Dornhege. Real-time localization and elevation mapping within urban search and rescue scenarios. *Journal of Field Robotics*, 24(8–9):723–745, 2007.

[24] A. Kleiner and C. Dornhege. Operator-assistive mapping in harsh environments. In *2009 IEEE International Workshop on Safety, Security & Rescue Robotics*, pages 1–6, Denver, USA, 2009. (Best Paper Award Finalist).

[25] S. Kohlbrecher, J. Meyer, O. von Stryk, and U. Klingauf. A flexible and scalable SLAM system with full 3d motion estimation. In *2011 IEEE International Symposium on Safety, Security and Rescue Robotics*, Kyoto, Japan, 2011. IEEE.

[26] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. G2o: A general framework for graph optimization. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3607–3613. IEEE, 2011.

[27] E. Magid, T. Tsubouchi, E. Koyanagi, T. Yoshida, and S. Tadokoro. Controlled balance losing in random step environment for path planning of a teleoperated crawler-type vehicle. *Journal of Field Robotics*, 28(6):932–949, 2011.

[28] Christopher Mei, Gabe Sibley, and Paul Newman. Closing loops without places. In *International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, October 2010. 10.

[29] Andreas C. Müller and Sven Behnke. Learning depth-sensitive conditional random fields for semantic segmentation of RGB-D images. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 6232–6237, 2014.

[30] R.R. Murphy. Case studies of applying gibson's ecological approach to mobile robots. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 29(1):105–111, 1999.

[31] R.R. Murphy. Marsupial and shape-shifting robots for urban search and rescue. *Intelligent Systems and their Applications, IEEE*, 15(2):14–19, 2000.

[32] K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, and Y. Hada. Redesign of rescue mobile robot Quince - toward emergency response to the nuclear accident at Fukushima Daiichi nuclear power station on March 2011. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 13–18, 2011.

[33] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann. 6D SLAM – 3D mapping outdoor environments. *Journal of Field Robotics, Special Issue on Quantitative Performance Evaluation of Robotic and Intelligent Systems*, 24(8-9):699–722, 2007.

[34] Andreas Nüchter and Joachim Hertzberg. Towards semantic maps for mobile robots. *Journal of Robotics and Autonomous Systems (JRAS), Special Issue on Semantic Knowledge in Robotics*, 56(4):915–926.

[35] Y. Okada, K. Nagatani, K. Yoshida, S. Tadokoro, T. Yoshida, and E. Koyanagi. Shared autonomy system for tracked vehicles on rough terrain based on continuous three-dimensional terrain scanning. *Journal of Field Robotics*, 28(6):875–893, 2011.

[36] Rohan Paul and Paul Newman. Fab-map 3d: Topological mapping with spatial and visual appearance. In *Proc. IEEE International Conference on Robotics and Automation (ICRA'10)*, pages 2649–2656, Anchorage, Alaska, May 2010. 05.

[37] Mircea Serban Pavel, Hannes Schulz, and Sven Behnke. Recurrent convolutional neural networks for object-class segmentation of RGB-D video. In *Neural Networks (IJCNN), International Joint Conference on*, 2015.

[38] L.M. Paz, P. Piniés, J.D. Tardós, and J. Neira. Large-scale 6-dof slam with stereo-in-hand. *Robotics, IEEE Transactions on*, 24(5):946–957, 2008.

[39] Johannes Pellenz and Dietrich Paulus. Stable mapping using a hyper particle filter. In *RoboCup 2009: Robot Soccer World Cup XIII*, pages 252–263, 2009.

[40] P. Pfaff, R. Triebel, and W. Burgard. An efficient extension to elevation maps for outdoor terrain mapping and loop closing. *The International Journal of Robotics Research*, 26(2):217–230, 2007.

[41] P. Piniés and J.D. Tardós. Large-scale slam building conditionally independent local maps: Application to monocular vision. *Robotics, IEEE Transactions on*, 24(5):1094–1106, 2008.

[42] A. Pronobis, O. Martinez Mozos, B. Caputo, and P. Jensfelt. Multi-modal semantic place classification. *The International Journal of Robotics Research (IJRR)*, 29(2-3):298–320, 2010.

[43] Andrzej Pronobis and Patric Jensfelt. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'12)*, Saint Paul, MN, USA, 2012.

[44] European Commission Directorate General for Communications Networks, Content and Technology Components and Systems: Robotics. Grant Agreement 644839: CENTAURO—Robust Mobility and Dexterous Manipulation in Disaster Response by Fullbody Telepresence in a Centaur-like Robot, 2014.

[45] Tobias Rodehutskors, Max Schwarz, and Sven Behnke. Intuitive bimanual telemanipulation under communication restrictions by immersive 3D visualization and motion tracking. In *IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2015.

[46] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2):214–226, June 2007.

[47] Max Schwarz and Sven Behnke. Semi-autonomous concurrent driving-stepping locomotion. In *Late Breaking Result Abstract at IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[48] Max Schwarz, Hannes Schulz, and Sven Behnke. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 1329–1335, 2015.

[49] Luis Sentis and Oussama Khatib. Synthesis of whole-body behaviors through hierarchical control of behavioral primitives. *International Journal of Humanoid Robotics*, 2(04):505–518, 2005.

[50] R. Sheh, B. Hengst, and C. Sammut. Behavioural cloning for driving robots over rough terrain. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 732–737. IEEE, 2011.

[51] Jiwon Shin, Rudolph Triebel, and Roland Siegwart. Unsupervised discovery of repetitive objects. In *IEEE International Conference on Robotics and Automation, ICRA 2010, Anchorage, Alaska, USA, 3-7 May 2010*, pages 5041–5046. IEEE, 2010.

[52] Kristoffer Sjöö. Semantic map segmentation using function-based energy maximization. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'12)*, 2012.

[53] Kristoffer Sjöö and Patric Jensfelt. Learning spatial relations from functional simulation. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'11)*, 2011.

[54] C. Stachniss, D. Hähnel, and W. Burgard. Exploration with active loop-closing for Fast-SLAM. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04)*, pages 1505–1510, 2004.

[55] F. Steinbrücker, J. Sturm, and D. Cremers. Real-time visual odometry from dense rgb-d images. In *In Workshop on Live Dense Reconstruction with Moving Cameras (at ICCV)*, November 2011.

[56] Jörg Stückler and Sven Behnke. Multi-resolution surfel maps for efficient dense 3d modeling and tracking. *Journal of Visual Communication and Image Representation*, 25(1):137–147, 2014.

[57] Jörg Stückler, Max Schwarz, Mark Schadler, Angeliki Topalidou-Kyniazopoulou, and Sven Behnke. Nimbro explorer: Semi-autonomous exploration and mobile manipulation in rough terrain. *Journal of Field Robotics*, 2015.

[58] Jörg Stückler, Benedikt Waldvogel, Hannes Schulz, and Sven Behnke. Dense real-time mapping of object-class semantics from RGB-D video. *Journal of Real-Time Image Processing*, 2014.

[59] M. Takahashi, K. Yoneda, and S. Hirose. Rough terrain locomotion of a leg-wheel hybrid quadruped robot. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 1090–1095. IEEE, 2006.

[60] R Triebel, J Shin, and R Siegwart. Segmentation and unsupervised part-based discovery of repetitive objects. In *Proc. of the 6th Robotics: Science and Systems Conference (RSS)*, 2010.

[61] Nikolaus Vahrenkamp, Tamim Asfour, and Rudiger Dillmann. Robot placement based on reachability inversion. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1970–1975. IEEE, 2013.

[62] T. Whelan, H. Johannsson, M. Kaess, J. J. Leonard, and J. McDonald. Robust real-time visual odometry for dense rgb-d mapping. In *In Proc. of the International Conference on Robotics and Automation (ICRA13)*, May 2013.

[63] T. Whelan, J. B. McDonald, M. Kaess, M. F. Fallon, H. Johannsson, and J. J. Leonard. Kintinuous: Spatially extended kinectfusion. In *In RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, Sydney, Australia, July 2012.

[64] Stephan Wirth and Johannes Pellenz. Exploration transform: A stable exploring algorithm for robots in rescue environments. In *2007 IEEE International Workshop on Safety, Security, and Rescue Robotics*, pages 1–5, 2007.

[65] K.M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Burgard. OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems. In *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation*, 2010.

[66] Jie Zhao, Jie Chen, Yubin Liu, He Zhang, and Aixiang Cheng. Geometric feature-based footholds identification for an insect-like walker. In *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*, pages 2345–2350, Dec 2013.