Optimizations on Semantic Environment Management: an Application for Humanoid Robot Home Assistance

Eduardo Munera[1], Juan-Luis Posadas-Yagüe[1], Jose-Luis Poza-Luján[1], J. Francisco Blanes[1], Jose E. Simó[1],

Abstract— This article introduces some optimization mechanisms focused on environment management, object recognition, and environment interaction. Although the generality of the presented system, this work will be focused on its application on home assistance humanoid robots. For this purpose, a generic environment formalization procedure for semantic scenery description is introduced. As the main contribution of this work, some techniques for a more efficient use of the environment knowledge are proposed. That way, the application of an area-based discrimination mechanism will avoid to process large amounts of data, useless in the current context, improving the object recognition, and characterizing the available interactions in the current area. Finally, the formalized description, and the optimization procedure, will be tested and verified on a specific home scenario using a humanoid robot.

Index Terms— environment modelling, humanoid robot, home robot, robot navigation.

I. INTRODUCTION

In contrast with other robot applications, in which environment are defined as a wide open space, home assistance tasks requires a precise and constant knowledge of its surrounding and the objects within. It is also critical to perform an accurate recognition of scene objects, and evaluate its interaction capabilities to properly execute its assigned tasks.

Due of this, robots which aims to serve as domestic assistant, usually have to perform some type of localization mechanism. This allows the robot to be aware of its position and navigate between rooms. In most cases these localization systems make use of a given environment map to avoid robot to explore when executing its assigned tasks. These maps usually defines the environment as an static scenario with dynamic objects. Furthermore, when performing its tasks, robot must evaluate and characterize perceived elements in order to establish a match between them and the possible objects, whose definitions are usually stored in large data collections. This action implies a high computational load, especially when must deal with big houses or highly interactive environment.

To achieve an efficient use of the available resources on environment management tasks and object matching the following objectives are proposed:

- Obtain a formalized definition of the home environment which can be suited for any possible scene. Definition will be focused on the semantic information of the domestic environment objects and its characteristics.

- Optimize the object management process and apply an area based restriction, such as the areas of interest, for bounding computational costs.
- Analyze the environment interaction capabilities according to the active area restrictions in each case.

The work presented in this article is organized as follows: in Section 2 some definitions for environment characterization are reviewed. Along Section 3 a formalized semantic based description of home environment is detailed. In Section 4 the proposed optimization is introduced. In addition, object management enhancements are obtained by the application of area based discrimination mechanisms. Some test are analyzed in Section 5 in order to validate the proposal and quantify the benefits of the introduced optimization. Finally in Section 6 conclusions and the future work is detailed.

II. ENVIRONMENT DESCRIPTION TECHNIQUES

A service robot which aims to efficiently perform some domestics tasks, in the same way that any kind of robot which plans to perform autonomously, usually requires some kind of environment knowledge [14]. Environment representation techniques are related to robotics since the developments of early navigation methods [2]. These have evolved from basic occupancy grid models [3] to complex 3D ones [21], offering a better recognition and characterization of the surrounding objects [22].

Several works, like [26] or [8], have been developed in order to model a 3D environment by using cuboids. In the first case an Octree [24] based solution is proposed, while in the second a Rtree [6] topology formed by a set of rectangular cuboids is established. But despite of improvements, 3D definition of the whole environment will always be more complex to manage than 2D. Although this 2D representation some times can not provide the required information, but in many of this cases a 2.5D representation can bring enough information for basic localization and interaction tasks. In [13] the advantage to generate an enriched map for indoor environments in a 2.5D representation is introduced. The benefits of gathering 2D, 2.5D, and 3D models in the same definition are also detailed in [10]. In [11] this techniques are applied to generate an environment model using a combination between configuration space maps (c-space) [12] and dynamic Voronoi Diagrams [4].

Scene objects are a fundamental part in a home environment. Due of this a reliable object recognition and management is a must for achieve any proposed goal. In [25] a hierarchical system for managing 3D object models is presented. This approach aims to define different level

[1] Institute of Control Systems and Industrial Computing, Polytechnic University of Valencia. Building 8G - Access D - Polytechnic City of Innovation, Valencia, Spain. emunera@ai2.upv.es; {jposadas; jopolu; pblanes; jsimo}@disca.upv.es

of objects for each type of elements (background, furniture, manipulable objects, etc) in order to assign a proper Octree resolution according to its requirements. Consequently a highly adaptable solution is obtained. Other works like [15] also propose the definition of a Global Structure Histogram (GSH) to establish a generalized representation of 3D objects. Furthermore, some contributions have proposed to simultaneously model and recognize 3D objects [16] [9] by managing Object Databases.

Although the importance of the geometrical information of the objects, other kind of characteristics must also be analyzed. Between them, the semantic classification of the objects offers information about its meaning. As an example, [27] proposes a table of object definitions which have included semantic information for object recognition. That way, a semantic-based recognition offers a more human alike approach. Despite of this, introduced solutions requires a semantic classification of the sensor information[23] [5]. As a result, semantic map definitions encourages human interaction and promotes its application on service tasks and housework [18].

Therefore, environmental modelling and object classification can be set as active topics in autonomous robotic. Multiple approaches have been offered, each one introducing a new point of view that is suitable for successfully resolving the proposed issues. For that reason, several of these contributions will be combined for offering a flexible and efficient description model of the environment which would be suitable for the geometric and semantic definition of any possible scenery. Furthermore, semantic information in the domestic environment will ease the definition of the services tasks, and the way that human users interact with the robots, encouraging the establishment of more natural interface. Nevertheless, since this contribution is focused on map information management and its optimization it won't be addressed mapping techniques or semantic data classification mechanisms.

III. HOME ENVIRONMENT FORMALIZATION

As has been introduced, a proper domestic environment description, and a detailed object characterization, are basic requirements for robot home assistance. Along this section a hierarchy of elements for environment definition is presented as is show in Fig. 1. This representation aims to characterize any environment in the most general way.

A. Home Environment Hierarchy

A generic description of the environment guarantees its application in any possible scenario, including a home environment. Also, it can be managed by any robot despite of its perception capabilities. As has been detailed in Fig. 1, the hierarchy is set by the following elements:

1) Atomic Elements: This is the simplest structure in the environment and offers a basic geometric and color (or texture) definition . In this case, elements aim not to define a whole object but bounding a basic feature of the scene. Atomic elements will be defined as a part of an environment object. Geometric information are defined as a 2D rectangular section just as the used on Rtrees. Also can be defined in 2.5D assigning a height value to the rectangular section.

2) Environment Object: As one of the most important factors its definition must provide a detailed and completed information and specify its interaction capabilities.

Object geometric definition is ready to work with 2D, 2.5D and 3D representations. The main difference between them is the type and the number of elements that compose the object. The 2D objects can be formed by a single 2D atomic element or a combination of many. The 2.5 only can be define as a 2.5 atomic element, which bounds the whole real object. Finally the 3D object is composed by one or many 2.5D atomic elements. Semantic definition of each object is specified by the Semantic Tag. This tag refers to an entry to the semantic definition dictionary which will be introduced next. That way, object can be associated with a full description of the object semantics, and its interaction capabilities, without increasing the load of the data structure.

Object must also include a time mark to provide information about current time. This is specially critical for the management of dynamic objects, that may modify its position along time, and for information sharing between other robots.

3) Avoided Areas: Characterized as those regions bounded in the map which can not be physically accessed, such as columns or walls. Avoided areas are defined as a list of vertex that are sequentially connected between them. This helps to frame the environment in the most simple way, and also encourages the use of point-in-polygon [7] algorithms in order to check if an element is within an avoided area.

4) Semantic Definition Dictionary: This semantic tag refers to an entry in the semantic meanings dictionary. In this structure all the relevant information about the semantic of each object is gathered and identified by a tag. That way, information about object characteristics can be obtained beyond the sensor capabilities. The information compiled into the semantic dictionary will also depend on the characteristics of the described scenario.

5) Global Map: Global Map representation is the highest layer in the hierarchy of the environment definition. Consequently this is strictly dependent on the characteristics of the environment objects, defined by the atomic elements, and avoided areas. Global map also introduces information about the geometrical characteristics of the map. That way, bounds are defined by a list of vertex $(v_1, v_2, \ldots, v_n)$ in the same way as the avoided areas.

Thus, map is set by its bounds, a list of avoided areas (that are characterized by their owns vertex lists), and finally a list of objects enclosed in the scene. The objects in the list can be stored either in its 2D, 2.5D, or 3D representation, according to the definition of that elements. That way simple 2D objects can be defined in the most light way without disabling the management of complex 3D representation of other objects. By allowing the coexistence of different levels of representation, and its adaptation to the simplest geometry
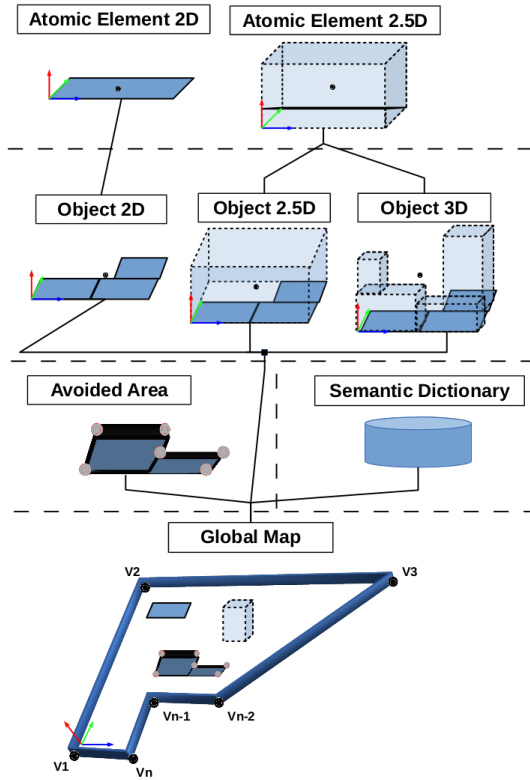
in each case, the efficiency of the system is improved.

Fig. 1. Environment hierarchy.

B. Object Semantics

In a domestic environment robots usually have to perform a certain degree of interaction with its environment and with humans. These interaction requires a knowledge of the environment beyond the measured capabilities, due to being related with their meaning. That way, robots have to be able to identify the semantic characteristics of each object which aims to establish an advanced interaction. Semantic information is gathered in three main categories:

1) Interaction capabilities: Here is described the information about how the robot can interact with this object, as well as the possible restriction or the conditions that have to been satisfied in order to establish this interaction. This information can also be extended by relating the object to a certain mission.

2) Dynamic properties: Provide information about the dynamism of the object. This property points if a certain object is static or if is a mobile element. In this last case it must also characterize the displacement capabilities of the object like the speed or its movement axis.

3) Localization: Some static objects can offer useful information for the navigation system, therefore these kind of objects can be considered as landmarks. Other objects, defined as restricted-area objects, can be located in different places but always inside of the same room or area, regardless of whether it is a static or a dynamic object. Finally, free objects can modify their position along all the environment

and consequently do not provide any kind of localization information.

IV. SYTEM ENHANCEMENTS

The presented formalization offers a common frame to describe the home environment. Furthermore, the offered flexibility for characterizing the scene objects allows to choose the most optimal representation of each one. Despite of this, large scenarios and a big number of objects could increase costs for map management. That way, the performance of a robot which performs a specific mission could be compromised depending on the environment in which it operates. In order to reduce or avoid those situations, a Zoom Map structure implemented as an attention mechanism [1] is presented for narrowing the computational costs. The Zoom Map data structure represents a subsection of the full map definition and collects all the information within the area. By analyzing the density of objects in the environment it can be specified an area size that guarantees a maximum number of managed objects, and consequently the worst case execution time. Nevertheless other parameters could be taken into account in order to optimize the optimum size of the interest area, and on future works can be considered the application of a dynamic size according to several conditions.
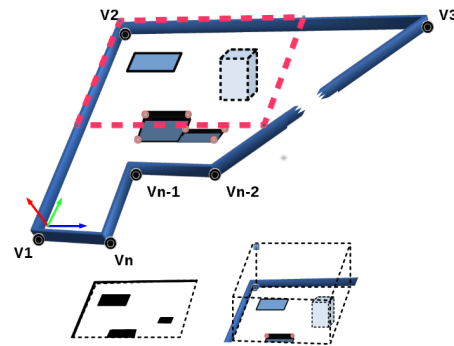

Fig. 2. Zoom Map: 2D and 2.5D areas of interest.

To improve the flexibility of the system, the Zoom Map can be specified either in 2D or in 2.5D. On one way, a 2D area of interest is focused only on a 2D projection of all the objects into the ground level, giving basic information for an optimum performance of those tasks which only requires to avoid obstacles. On the other way, 2.5D deals with all the objects (in 2D, 2.5D or 3D) enclosed into the bounding box defined as the area of interest. The Zoom Map have to update the location of the area of interest in two different situations. The first one takes place when the robot has reached a threshold distance from the center of the area. The second one is triggered by a time out that indicates the need to update the area. In both cases the conditions can be parameterized according with factors like the dynamic of the robot, the dynamic of the environment, etc.

In eq. 1 and eq. 2 are characterized the computation time of the whole and the zoomed map respectively. In these equations t_i and $t_{i'}$ represents the respective times

for map management along n iterations, performed before map actualization, while $t_{m'}$ represents the cost for zoom actualization. Management times are related with the number np of perceptions p and the number ne, or ne', of map elements m or m' (in the full or the zoomed representation respectively). The map actualization time $t_{m'}$ is defined by the comparison between the robot position pos and the map elements m. As is specified in eq. 3, the computational costs are determined by the number n of executions of the algorithm with t_i and $t_{i'}$ costs in each case. Without lack of generality t_i will always be equal or greater than $t_{i'}$, and according to the omega notation T_{map} is bounded below by $T_{map'}$ asymptotically.

$$T_{Map} = \sum_{i=0}^{i=n-1} t_i \rightarrow t_i = \sum_{j=0}^{j=np} \sum_{k=0}^{k=noe} f(p_j, m_k) \quad (1)$$

$$T_{Map'} = t_{m'} + \sum_{i=0}^{i=n-1} t_i' \rightarrow \left. \begin{array}{l} t_{m'} = \sum_{k=0}^{k=ne} f(pos, m_k) \\ t_i' = \sum_{j=0}^{j=np} \sum_{k=0}^{k=ne'} f(p_j, m_k') \end{array} \right\} \quad (2)$$

$$\left. \begin{array}{l} T_{Map} \rightarrow O(n) \\ T_{Map'} = \rightarrow O(n)' \end{array} \right\} \begin{array}{l} \forall np \in \mathbb{N}, \forall ne \in N \cdot O(n) \geq O(n)' \\ \rightarrow T_{Map'} \in \Omega(T_{Map}) \end{array} \quad (3)$$

V. USE CASE AND TESTS

Developed system has been defined as a suitable tool to describe and manage the environment information. In order to check its performance in a real case, a group of tests focused on the analysis of this implementation, have been proposed. To provide a testing platform, the V-Rep simulator [20] will be used for modelling the home environment and all the objects.
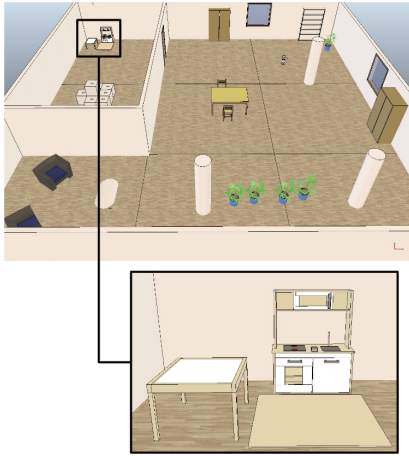

Fig. 3.   Home environment model in addition with the HumaBot competition scenery.

As the main scenario for the experiments, it has been modeled a home environment with two rooms just as can be observed in Fig. 3. In this figure can also be noticed that inside one of these rooms is located a small table and kitchen which corresponds with the normative scenario of the Humabot Challenge [17]. The Humabot Challenge aims to present the robot as an integral part of the house by helping its occupants. The selected challenges are held in the kitchen, where three different tests must be carried out: put off the fire, check the item stock and generate a shopping list, and prepare the meal. For performing these tasks the humanoid robot Nao [19] has been chosen as the official platform in this competition. Because of this, and due to its versatility, the Nao robot will also be used during the following experiments.
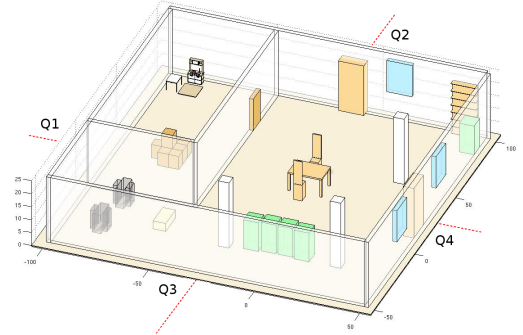

Fig. 4.   Graphic representation of the formalized environment.

Proposed tests are designed in order to evaluate the advantages of the developed environment hierarchy, the semantic information and the area based optimization. The first step is to establish the description file which contains the formalized description of the environment and the semantic dictionary. In Fig. 4 the graphical representation of the formalized environment is depicted.

One of the more critical steps for offering an optimal management of the information is the selection of the size of the area of interest. This size will determine the amount of objects, and consequently the amount of data, to deal with. For that reason, a set of test to select the proper size of the interest area, and to narrow the quantity of managed information, are developed. In each one will be analyzed the increment of managed objects as the size of the area of interest grows, starting from the size of the robot perimeter until the whole map size is reached. Up to five tests will be performed by firstly placing the robot in the center of the map, and then in the center of each quart of the map (Q1, Q2, Q3 and Q4), just as has been showed on Fig. 4. Taking a look to the Fig. 5 and establishing 20 atomic elements as the average number of managed entities, it can be noticed how the most suitable size for the area of interest is of 5,5 meters approx.

Once the size of the area of interest is set, main experiments can be performed. These experiments are designed in order to study two main factors: how the zoom map can define the environment interaction capabilities in each case, and how the system will be optimized thanks to the proposed techniques. Nao robot has been configured to perform a predefined trajectory all around the scenario. During this displacement the area of interest will be actualized according to the triggering conditions. That way, system is parameterized
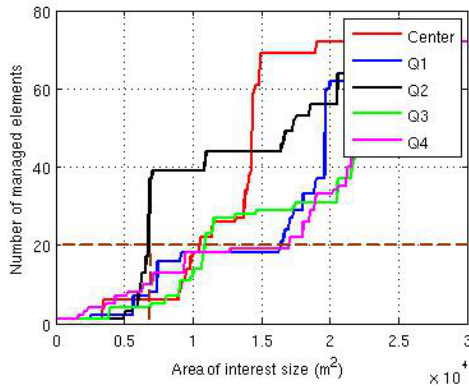
Fig. 5. Number of managed objects according to the size of the area of interest.

Fig. 7. Interaction capabilities offered by the zoom bounded objects.

to actualize the Zoom Map when the robot has traveled a 40 % of the area size. The evolution of the areas of interest along the performed trajectory can be checked in the Fig. 6.

The set of objects gathered in each area of interest, and its interaction capabilities, will define the range of actions that the robot can perform. The interaction capabilities of each object can be listed by analyzing the semantic definition of each object. These definitions are stored in the semantic dictionary, and each object is marked with a semantic tag which points to its own definition. Thus, the available actions in each case can be related with the behaviour selection and the execution of the mission-oriented tasks. Furthermore, missions can be assigned according to different areas. In this concrete test, when the area of interest is focused on the small kitchen, defined as the HumaBot scenario, the robot can analyze all the semantic information in order to find out the possible actions to perform during the challenge. An illustrative diagram can be found on Fig. 7.
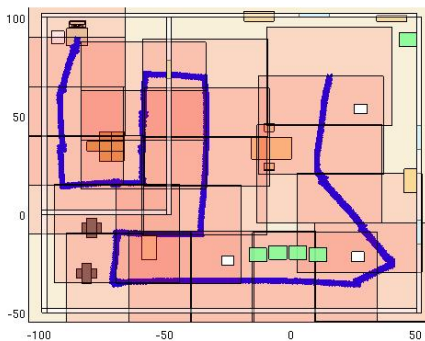

Fig. 6. Evolution of zoom map areas during the performed trajectory.

In order to verify the optimization obtained, several simulations have been performed. Some experiments have been configured to manage the full representation of the map, while some others are designed to deal with the zoomed one. Management times (characterized by the equations 1 and 2)will be stored in order to compare each representation and quantify the improvements. Although in every case the
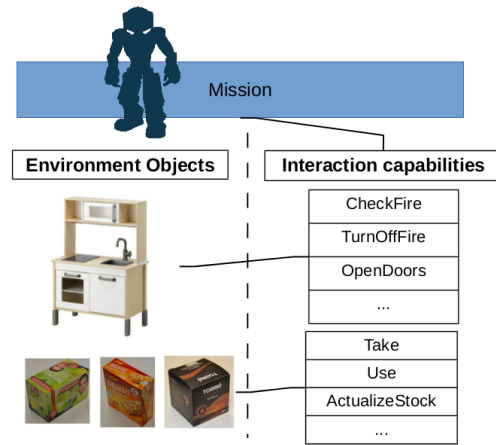
number of managed objects will be at least the same as in the whole map, and usually lower, time costs reduction have to be carefully analyzed. Furthermore, the costs derived from the evaluation of the actualization trigger conditions and from the actualization have to be also taken into account. As can be observed in the Fig. 8, during these tests the application of the Zoom Map, even with the actualization costs, offers an average time reduction around the 90%, and is up to 87% in the worst case.
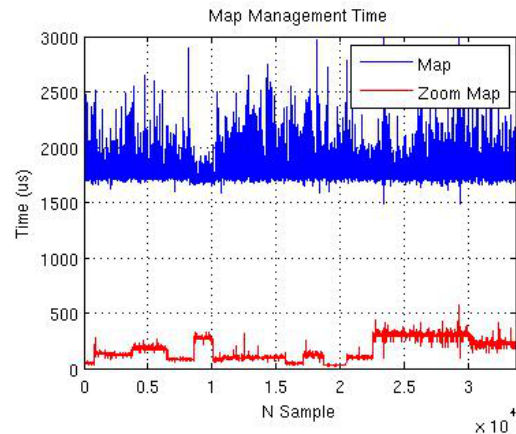

Fig. 8. Management time for map and zoom map during the proposed test.

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

As main conclusion must be pointed that the introduced formalization process offers the capability to describe and characterize any possible environment for mobile robots. That way, due to the flexibility of the system, the number of possible applications is increased. The introduction of some area based optimization techniques, defined by the Zoom Map mechanisms, provides a significant improvement of the system by selecting a subsection of the environment according to the requirements of the mission. The semantic definition of the environment objects, in combination to the

Zoom Map restrictions, defines the range of actions that the robot can perform in each section of the scenario. That way, mission system can be enhanced by evaluating the possible interactions and relating them to the proposed mission goals.

The generalized formalization and the system improvements have been tested and verified in the result section. In the use case is showed a real application for a small wheeled robot, which has been configured to manage the formalized representation of the environment. Tests have been also designed in order to quantify the computational improvements obtained through the proposed optimization mechanisms. During several simulations, the application of a Zoom map region has proved to perform in a more efficient way then the full map, even that a sporadic actualization phase is required.

B. Future Work

The presented work opens several lines of future work. Next developments will be focused on the application of the presented mechanisms to enhance the system, to improve the contribution itself, and to analyze its performance on physical systems.

The semantic definition of elements, and its interaction capabilities, have been related with the execution of a mission. In order to improve the mission system, future works will analyze how the mission tasks, and behaviours, can be assigned to a specific area of interest, or related with a kind of object in the environment. Semantic information will be also used to discriminate objects which acts as landmarks. As future work it will be developed a navigation system designed to work with this definition as an input to match sensor information in order to recognize landmarks and estimate the robot position.

Finally, Zoom map will be improved to work with areas of interest which will be dynamically resized according to the requirements of the system. In order to select a proper size will be need to perform a deep study of the parameters involved in the actualization triggering process and the characteristics of the active task, such as the dynamic of the environment or the average speed of displacement.

ACKNOWLEDGMENT

REFERENCES

[1] Carlos E Agüero, Francisco Martín, Luis Rubio, and José María Cañas. Comparison of smart visual attention mechanisms for humanoid robots. *Int J Adv Robotic Sy*, 9(233), 2012.

[2] James L Crowley. Navigation for an intelligent mobile robot. *Robotics and Automation, IEEE Journal of*, 1(1):31–41, 1985.

[3] A Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, 1989.

[4] M Gahegan and I Lee. Data structures and algorithms to support interactive spatial analysis using dynamic Voronoi diagrams. *Computers, Environment and Urban Systems*, 24(6):509–537, 2000.

[5] C Gurau and A Nüchter. Challenges in Using Semantic Knowledge for 3D Object Classification. In *Proceedings of the KI 2013 Workshop on Visual and Spatial Cognition*, pages 29–35, 2013.

[6] A Guttman. *R-trees: A dynamic index structure for spatial searching*. ACM, 1984.

[7] Eric Haines. Point in polygon strategies. *Graphics gems IV*, 994:24–26, 1994.

[8] S Khan, A Dometios, and C Verginis. RMAP: a rectangular cuboid approximation framework for 3D environment mapping. *Autonomous Robots*, pages 1–17, 2014.

[9] S Kriegel, M Brucker, ZC Marton, T Bodenmuller, and M Suppa. Combining object modeling and recognition for active scene exploration. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2384–2391. IEEE, 2013.

[10] B Lau, C Sprunk, and W Burgard. Incremental updates of configuration space representations for non-circular mobile robots with 2D, 2.5 D, or 3D obstacle models. In *European Conference on Mobile Robots (ECMR)*, pages 49–54, Orebro, Sweden, 2011.

[11] B Lau, C Sprunk, and W Burgard. Efficient grid-based spatial representations for robot navigation in dynamic environments. *Robotics and Autonomous Systems*, 61(10):1116–1130, 2013.

[12] T Lozano-Perez. Spatial planning: A configuration space approach. *Computers, IEEE Transactions on*, 100(2):108–120, 1983.

[13] RC Luo and CC Lai. Enriched indoor map construction based on multisensor fusion approach for intelligent service robot. *IEEE Transactions on Industrial Electronics*, 59(8):3135–3145, 2012.

[14] Ren C Luo, M-H Lin, and Ralph S Scherp. Dynamic multi-sensor data fusion system for intelligent robots. *Robotics and Automation, IEEE Journal of*, 4(4):386–396, 1988.

[15] M Madry, CH Ek, R Detry, K Hang, and D Kragic. Improving generalization for 3D object categorization with Global Structure Histograms. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 1379–1386, Vilamoura, Algarve, Portugal, 2012. IEEE.

[16] ZC Marton and D Pangercic. Combined 2D3D categorization and classification for multimodal perception systems. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 30(11):1688–1695, 2011.

[17] IEEE-RAS International Conference on Humanoid Robots. Humabot challenge. URL http://www.irs.uji.es/humabot/humabot-challenge, 2014.

[18] D Pangercic, B Pitzer, M Tenorth, and M Beetz. Semantic object maps for robotic housework-representation, acquisition and use. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 4644–4651. IEEE, 2012.

[19] Aldebaran Robotics. Nao software 1.14.5 documentation. URL http://www. aldebaran-robotics. com, 2014.

[20] Eric Rohmer, Surya PN Singh, and Marc Freese. V-rep: A versatile and scalable robot simulation framework. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1321–1326. IEEE, 2013.

[21] S Thrun, W Burgard, and D Fox. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*. IEEE, 2000.

[22] A Torralba, KP Murphy, WT Freeman, and MA Rubin. Context-based vision system for place and object recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 273–280. IEEE, 2003.

[23] R Vieux, J Benois-Pineau, JP Domenger, and A Braquelaire. Segmentation-based multi-class semantic object detection. *Multimedia Tools and Applications*, 60(2):305–326, 2012.

[24] J Wilhelms and A Van Gelder. Octrees for faster isosurface generation. *ACM Transactions on Graphics (TOG)*, 11(3):201–227, 1992.

[25] KM Wurm, D Hennes, D Holz, RB Rusu, C Stachniss, K Konolige, and W Burgard. Hierarchies of octrees for efficient 3d mapping. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4249–4255. IEEE, 2011.

[26] KM 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*, page Vol. 2, 2010.

[27] H Zhao, Y Liu, and X Zhu. Scene understanding in a large dynamic environment through a laser-based sensing. In *Robotics and Automationand Automation (ICRA), 2010 IEEE International Conference*, pages 127–133. IEEE, 2010.