KSEM

STARTPAGE

PEOPLE
MARIE CURIE ACTIONS

International Outgoing Fellowships (IOF)
Call: FP7-PEOPLE-2011-IOF

PART B

"KSEM"

Contents

1 RESEARCH AND TECHNOLOGICAL QUALITY

1.1 Research and technological quality, including any interdisciplinary and multidisciplinary aspects of the proposal

Environment models are resources for enabling robots to perform their tasks more reliably, efficiently, and competently by using information about the environments. The proposed project KSEM (Knowledge-enabled Semantic Maps) investigates a new generation of environment models that are to enable robots to perform everyday manipulation tasks in human environments effectively and efficiently. To this end, KSEM proposal extends state-of-the-art models such as Semantic Object Models with probabilistic spatio-temporal information: they probabilistically represent where objects typically are, where objects belong, the state of objects depending on their location (perishable objects are typically to be found in refrigerators), the dependence of activity contexts on object locations (if pots are on the oven then the activity context is probably meal preparation). Using these representational mechanisms KSEM projects provides semantic object maps that can characterize objects and places in the environment with respect to the role they play in activities (affordances). These are necessary for autonomous robots to do the right actions to the right objects in the right way: when setting the table, the robot should retrieve the cups from the cupboard because they are clean. When cleaning the table the robot should put the dirty cups into the dishwasher to clean them. Semantic maps in KSEM project enable the robot to infer this kind of commonsense knowledge. Technologically, KSEM will be realized as semantic object models combined with relational probabilistic models that relate objects to activities, places, time, roles, etc. KSEM semantic maps are (semi-)automatically acquired through the combination of semantic object model acquisition, the observation of everyday activities, the use of web instructions on everyday activities and statistical relational learning. Usefulness of the KSEMproject will be demonstrated in the context of setting the table, loading the dishwasher, and cleaning up as the primary task domain. The joint probability distributions of semantic maps in KSEMare to represent the dynamic aspects of the environment in the context of everyday activities, which include:

- *the typical locations of objects:* cups and plates can typically be found on the table, in the dishwasher, and in the cupboards;
- *the home/storage location of objects:* cups and plates are stored in a particular cupboard before they are used and after they are cleaned;
- *typical arrangements of objects:* some objects are typically used together in certain configurations, for example at lunch;

- *action-related places*, the places where actions and activities are performed;
- *role/affordance-based object models:* the roles that objects play in activities;
- *time-space-state representations:* the changes of objects' state and position over time.

SHORTEN According to the research topics of the KSEM project, we will now briefly review relevant research directions in the areas of semantic object maps, object perception, perception of everyday activities, knowledge representation for robots, statistical relational models, spatio-temporal learning, and entity resolution.

1.1.1 Semantic Object Maps: Acquisition and Use

As semantic maps include models of the environment structure and the static components of the environment including the furniture, appliances, etc. we will first consider the state of the art in the acquisition and use of 3D semantic object maps for everyday manipulation.

In Wuenstel and Moratz (2004), Wünstel and Moratz use a graph representation to detect chairs, but the relation descriptions are manually estimated, and thus it is unclear whether the proposed method scales. Posner et al. Posner et al. (2008) use probabilistic graphical models such as Markov random fields to label planar patches in outdoor urban datasets. Their work is based on that of Anguelov et al. Anguelov et al. (2005) and Triebel et al. Triebel et al. (2007b), which define point-based 3D descriptors and classify them with respect to object classes such as chairs, tables, screens, fans, and trash cans in the former, respectively wires, poles, ground, and scatter in the latter.

An EM-based algorithm for learning 3D models of indoor environments is presented by Liu et al. Liu et al. (2001). The maps are created using mobile robots equipped with laser range finders, but they do not include any semantic information. Grau Grau (1997) uses a stereoscopic camera system and a knowledge base in the form of a semantic net to form 3D models of outdoor environments. A hybrid representation including spatial and semantic aspects is proposed by Galindo et al. Galindo et al. (2005) for an indoor environment, but their approach needs further investigation.

Iocchi and Pellegrini Iocchi and Pellegrini (2007), Mozos et al. Mozos et al. (2007), and Vasudevan et al. Vasudevan et al. (2007) use 2D laser sensors to create a map used for navigation and additional semantics are acquired through the use of vision. An object-based approach for cognitive maps is used to recognize objects and classify rooms in different categories in Vasudevan et al. (2007), while in Mozos et al. (2006, 2007), places are semantically labeled into doorways, kitchens, corridors, rooms. The advantages of these representations are straightforward: they keep the computational cost low enough and base their localization and pose estimation on the well-known 2D SLAM (Simultaneous Localization and Mapping) problem, while the problem of place labeling is solved through the usage of feature descriptors and machine learning. However, by reducing the dimensionality of the mapping to 2D, most of the world geometry needed for manipulation is lost. Also, the label categories need to be learned a priori through supervised learning and this makes it unclear whether these representations scale well.

With few exceptions, in particular in the area of cognitive mapping as Vasudevan et al. (2007) and the work of Modayil and Kuipers Modayil and Kuipers (2004), but also including the more recent work by Triebel et al. Triebel et al. (2007a), maps do not represent objects relevant to any robot tasks apart from navigation.

For modeling the static part of the environment Nüchter and Hetzberg Nüchter and Hertzberg (2008) classify 3D sensed data from a laser sensor into walls, floor, ceiling, and doors based on angular thresholds. Static objects including standing humans are detected in two steps, first a hypothesis is extracted from a depth image using an appearance based method, then 3D matching to a model is performed to evaluate the classification. This however requires 2D models for all poses of each object to be detected, as well as the complete 3D models. The same holds true for postures in the case of humans.

1.1.2 Object Detection, Recognition, and Reconstruction for Pick-and-Place Tasks

Semantic maps in KSEMmust also include models of the objects of daily use, which are to be manipulated by the robot. Thus, we now consider recent work in object detection, categorization, localization, and reconstruction for robotic object perception.

A vision-based grasping system which segments objects on a table and constructs triangular meshes for them is presented by Richtsfeld and Vincze Richtsfeld and Vincze (2008). While the presented method is general and works for many objects, it creates complex models for certain objects, which could be simplified through the usage of geometric primitives. A simplification of the modeling problem is used in Grasp-It Miller and Allen (2004), where geometric shape primitives are used to model each object as a sphere, cylinder, cone or box.

Geons (geometric icons) Bley et al. (2006) can be used to develop generic category descriptions using geometric primitives, but also additional category knowledge, for the purpose of building libraries of grasp models for a variety of objects.

A computer vision- and machine learning-based method is used by Saxena et al. Saxena et al. (2008) to train classifiers that can predict the grasping points in an image. This is then applied to images of unseen objects. To obtain 3D positions of grasping points, the authors use stereo cameras, but their approach works reliably only to the extent provided by the training data. Another issue is the segmentation of objects, since grasp points are provided with no information about what objects are in the scene and to which of them do the identified points correspond. Bone et al. Bone et al. (2008) use an accurate line laser and a camera to build models and identify grasping points for novel objects with very encouraging results. However the system was tested only on two objects, thus its scalability is not clear.

In purely computer vision based approaches, features like the ones described by Lowe Lowe (2004) or Lepetit and Fua Lepetit and Fua (2006) are used to find matches between parts of a scene and a database of object images. The problem with these kinds of approaches is that they only work for objects that are in the database, and since no knowledge about the 3D information is known, the system can easily make mistakes and return false positives (e.g., a cereal box containing a picture of a beer bottle printed on it will get recognized as a bottle of beer). Some of the solutions adopted involve the offline creation of complete 3D models for the targeted objects and finding feature spaces to match the partial views with models in the database, as presented by Collet et al. Collet et al. (2009). Another approach to obtain 3D information directly from camera images is to project CAD models from a database to the image and search for good matches in the edges domain, as in Ulrich et al. Ulrich et al. (2009) for example. While this is a more direct method, it is still dependent on a database of CAD models.

Available models of complex objects are decomposed into superquadric parts by Biegelbauer and Vincze Biegelbauer and Vincze (2007) and Zhang et al. Zhang et al. (2004), and these models are matched to a point cloud. This, however, needs a database of models, and moreover, their decomposition into superquadric components, which is often difficult to obtain. A sample consensus based approach for model decomposition is presented by Schnabel et al. Schnabel et al. (2007), where a set of 3D geometric primitives (planes, spheres, cylinders, cones and tori) are fit to noisy point clouds. Since the point clouds presented there are complete, the authors do not need to reconstruct the missing parts.

Thrun and Wegbreit Thrun and Wegbreit (2005) describe a method for detecting and verifying symmetries in point clouds obtained from a single viewpoint which works very well for nicely segmented objects, however the problem of under- or over-segmented objects remains.

Object categorization goes hand in hand with segmentation and is usually performed using a single sensing device. Given a large set of training values containing all possible views, most approaches try to abstract the problem by using features like the ones proposed by Quack et al. Quack et al. (2007), Yan et al. Yan et al. (2007), and Savarese and Fei-Fei Savarese and Fei-Fei (2007), which work best on low scale and texture variance. The scaling variance can be reduced significantly by a previous segmentation. Saxena et al. Saxena et al. (2007) try to extract the 3D world out of only one view in order to improve the segmentation of objects. Another approach is to actively explore the environment and segment objects using the motion to generate 3D shape information, as done by Welke et al. Welke et al. (2008) and Feldman and Weinshall Feldman and Weinshall (2008) for example.

To improve the segmentation, Lai and Fox Lai and Fox (2009) take a randomized approach by classifying a "soup of segments" generated by different combinations of clusters, and use the results to get a final segmentation. They also explore different domain adaptation techniques in order to incorporate synthetic data in training their classifier. Though they are working with outdoor data, the same techniques can be applied to segment and classify indoor objects using low resolution 3D data, e.g. from time-of flight cameras.

1.1.3 Symbolic Knowledge for KSEM

For performing high-level tasks, robots need large amounts of semantic information from their environment model, like the types, locations and properties of objects. Very few systems exist that offer this kind of deep semantic environment information.

At the highest level of abstraction, one considers the recognized objects, whose semantic meaning is only implicitly represented: Humans immediately associate various properties with something called a “cupboard”, while robots usually do not have this kind of knowledge. Without an explicit knowledge representation, different robots or even different parts of the same robot may have a very different notion of an object.

Deeper semantic representations, which also describe object properties, such as the point at which to grasp an object or location of the opening of a container like a bottle, are used by Okada et al. Okada et al. (2007) but are mainly hand-coded and do not leverage the power of hierarchical, abstract knowledge representations. Galindo et al. Galindo et al. (2008) present a system for automatically building maps that combine a *spatial hierarchy* of local metric and global topological maps with a *conceptual hierarchy* that describes some semantic properties of rooms and objects. In this respect, their approach is similar to ours, but the conceptual hierarchy is much simpler and the spatial description much coarser.

1.1.4 Statistical Relational Models

In order to declaratively represent highly complex domains, in which there are relations between a variable number of relevant entities and which are furthermore governed by uncertainty, one requires a representation formalism that combines statistical with relational components, abstracting away from concrete entities to compactly represent general principles about the relevant aspects of the real world. In statistical relational learning, a number of such formalisms have been proposed, as presented by Getoor and Taskar Getoor and Taskar (2007) and De Raedt De Raedt (2008). Statistical relational learning methods have countless applications, including collective classification Neville and Jensen (2003), link prediction Taskar et al. (2003) and object identification Singla and Domingos (2006). In particular, the methods of statistical relational learning allow to represent full-joint distributions over logical propositions about a changing set of entities in a concise, declarative manner, and they provide a fully integrated framework for learning and inference. Therefore, they are ideally suited to the representation of the probabilistic components of semantic maps in KSEMas we envision them, since we need to consider, for example, spatial relations between objects in the environment, their attributes changing over time, their relevance to activities taking place and the effect these activities may have upon them.

1.1.5 Spatio-Temporal Learning

Our intended research in learning spatio-temporal structures for semantic maps in KSEMis inspired by research in visual analytics for the analysis of movement data. Andrienko et al. learn concepts like the working place, the living place, typical navigation routes and other spatio-temporal behavior patterns from GPS-data from their car Andrienko et al. (2007). While in their case, the data mining tasks are performed by human experts using visual analytics methods, other researchers such as Liao et al. Liao et al. (2007b,a) perform some of these learning tasks using probabilistic learning methods (hierarchical conditional random fields). Acquiring such models is also investigated in the pervasive systems community, where Philipose and his colleagues Pentney et al. (2007); Landwehr et al. (2007) and Intille and his research group Intille et al. (2006) learn models of daily activities from ubiquitous sensor networks.

If maps are to be used for more than navigation and mere obstacle avoidance, a semantic interpretation of the observed scenes is required, which necessitates a meaningful labeling of objects appearing in mapped scenes. Posner et al. Posner et al. (2008) propose a two-stage process to solve this problem, where, at the local level, classification is based on appearance descriptors, and at the global scene level, Markov random fields (MRFs) are applied to model relationships. Triebel et al. Triebel et al. (2006) draw upon statistical relational learning methods, specifically associative Markov networks, to solve a similar problem — again, however, on laser data only and at a low level of abstraction.

1.1.6 Entity Resolution

A key problem that has not yet received substantial attention in autonomous robotics is the inference about whether two different observations resulted from the same real world object. This type of inference is needed to recognize previously observed objects again, which in turn must be solved to build spatio-temporal environment models.

This problem can be viewed as an entity resolution problem that arises in many information integration scenarios: Given two or more sources containing records on the same set of real-world entities (e.g. mapped objects), the problem is to determine which of the records actually refer to the same object, as we typically have no unique identifiers that tell us what records from one source correspond to those in the other sources. Furthermore, the records representing the same entity may have differing or even partially contradictory information. For example, one record may have an attribute falsely assigned (as a result of sensor inaccuracy), another may be missing some fields. An entity resolution algorithm attempts to identify the matching records from multiple sources (i.e. those corresponding to the same real-world entity), and merges the matching records as best as it can. Entity resolution algorithms typically rely on user-defined functions that (a) compare fields or records to determine if they match (are likely to represent the same real world entity), and (b) merge matching records into one, and in the process perhaps combine fields (e.g., creating a new name based on two slightly different versions of the name).

While early work has often phrased the problem as a classification problem where a pair of records would be independently classified as either “matching” or “non-matching” Fellegi and Sunter (1969), more recent approaches have phrased the problem as link prediction in statistical relational models, e.g. Singla and Domingos Singla and Domingos (2006): One defines an equivalence relation over entities and considers, in a probabilistic setting, any predicates that state the same things about two entities as evidence for the entities referring to one and the same real-world entity; the importance of certain predicates can be learned from statistical data. Most importantly, this relational approach considers the objects collectively and does not assume that the equivalence of a pair of objects is independent of other equivalences.

1.2 Appropriateness of research methodology and approach

In order to clarify the motivation from above, we identified the following four subgoals of the project:
Subgoal 1: Representations in KSEM will be hybrid by integrating geometric descriptions (such as primitives, meshes, voxels), first-order description-logic-based symbolic representations (e.g. symbolic object properties, relations between objects and object classes) and probablistic first-order representations (including Markiv logic and Bayesian logic networks). Together with a representation language based on OWL (Web Ontology Language) and a concept ontology based on researchCyc (an encyclopedic knowledge base), these representations can be stored, extended and they can be queried by a Prolog-based inference mechanism.
Subgoal 2: Perceptual mechanisms for detection, categorization, recognition, localization and reconstruction of objects of daily use. Additionally, the robot needs capabilities to percieve and interpret arrangements of objects (spatially and/or temporally related). This includes methods for identity resolution and estimation of object state based on partial information and context to allow affordance-based manipulation.
Subgoal 3: Learning of semantic maps for autonomous world state interpretation. Some of the identified learning problems are:

- Learning typical locations of objects such as storage or usage locations, and interpretation of containers based on a generalization of their contents (e.g. fridge);
- Learning action-related places are places where actions are performed or objects take on certain roles in actions. This allows a robot to learn e.g. the concept of a chair by its role to support people as opposed to purely geometric or appearance based recognition;
- Learning arrangements of objects is important to abstract from a continuous coordinate system of object locations to a symbolic, relational model of a scene which allows e.g. for context interpretation of a table arrangement;

- Knowledge-intensive learning from selected web sites such as researchCyc for encyclopedic knowledge, ehow.com or wikihow.com for task instructions and the OpenMind Indoor Common Sense (OMICS) knowledge base. This constitutes important background information for KSEM project.

These learning problems require autonomous and on-demand (partial) processing of sensor data to enable life-long learning.
Subgoal 4: Applications for robot control to enable real-world problem-solving within a household assistance system such as execution of everyday manipulation actions (e.g. emptying a shopping bag, set the table for a given meal, or cleaning the table, while stowing items in their respective places, such as the fridge or dish washer) and task-dependent partial environment mapping.

All of the subgoals named above are to be implemented in a system that is to run on real robot platforms in real household environments.

1.3 Originality and Innovative nature of the project, and relationship to the state-of-the-art of research in the field

The ultimate goal of the project KSEM is the investigation of environment models that can be inferred from a combination of environment exploration and activity pbservation. To this end, we will design, implement and empirically analyze semantic model acquisition mechanisms for objects and everyday activities in order to generate semantic maps of human living environment.

As an example, consider the following query that might arise within a household scenario: A human asks the robot to retrieve a glass he had been drinking from earlier. This simple request can spawn several queries to KSEM system, such as: Where is a glass that this person had contact with earlier? There might be multiple glasses, so the system has to select the one which was used for drinking and disregard other glasses that the human might have cleaned earlier. Also, for picking up the glass, grasp analysis might need a surface mesh of the object or a precomputed list of applicable grasps, and the motion planner needs an occupancy voxel map or surface mesh of the whole environment for collision avoidance or possibly additional manipulation constraints linked to the glass, such as maintaining a vertical orientation over the whole action. A system that is able to tackle above described challenges, to the best of our knowledge, does not exist.

1.4 Timeliness and relevance of the project

With all the recent advances in the field of personal robotics and with first personal robots projected to be deployed in households within next 5 to 10 years, KSEM's relevance is enormous for robots to perform in highly dynamic and unpredicted situations. It will also enable researchers from other fields such as artificial intelligence, psychology, ergonomy, etc to get a grasp at the data and study human living patterns, social acceptance of robots and other high-level concepts. The KSEM project's relevance under the Marie Curie fellowship within a European setting is connected to the innovative ideas it introduces by combining in a novel way approaches to build semantic environment models to be used in robots' everyday operations. The applicant will need to be additionally trained within the host group, as will be discussed in Section 2, to pursue the proposed research, which is also a significant part of the Marie Curie program. Since the KSEM project is setup such to exploit the mutual leverage with the open source ROS community [1] and the PR2 Beta community [2] in terms of code reusability, feedback from other experienced researchers and critical reviews, it will make a significant contribution in enhancing the continuously advancing scientific excellence within Europe. Both the applicant and the environment in which the project will be realized will benefit significantly from the realization of the proposed project, since also members, senior and junior ones, from the host group will have their opportunity to assist the KSEMresearch.
ERIC: Check if this is enough.

[1] www.ros.org
[2] http://www.willowgarage.com/pages/pr2/pr2-community

1.5 Host research expertise in the field and quality of the group/supervisors

Michael Beetz from Intelligent Autonomous Systems group, Technische Universität München (TUM) will represent the host organization in the project and Pieter Abbeel from EECS department, University of California at Berkeley will represent the partner organization. Prof. Beetz is a vice coordinator of the German national cluster of excellence COTESYS (Cognition for Technical Systems) where he is also co-coordinator of the research area "Knowledge and Learning". Prof Beetz was a member of the steering committee of the European network of excellence in AI planning (PLANET) and coordinating the research area "robot planning". He is associate editor of the AI Journal. His research interests include plan-based control of robotic agents, knowledge processing and representation for robots, integrated robot learning, and cognitive perception. Prof. Beetz has published over 200 articles in the first-class conferences and journals in the fields of robotics, computer vision and artificial intelligence [3]. Currently he is leading a group of 2 postdoctoral fellows, 25 grad students and a large number of undergrad students and visiting researchers. He is also serving as a principal investigator for TUM's PR2 Beta Program. Prof. Beetz has successfully coordinated and trained more than 15 postdoctoral fellows and PhD students, in addition to a larger number of master's and diploma students. Most of his group alumni currently hold faculty positions in Universities and Research Institutions in Europe and United States. The equipment that is required for the realization of the project at the host organization is already available. In particular, three robot manipulation platforms – a PR2[4], an iCub[5] and a custom-built robot with KUKA leightweight arms and DLR/HIT hands – are available for real-world experiments on a daily basis. So is the testbed kitchen environment with an ample set of ubiquitous devices such as Kinects sensors, ceiling cameras, RFID readers, etc.

Pieter: Above like description for you and Ken

References

Andrienko, G., Andrienko, N., and Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations*, 9(2):38–46.

Anguelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., and Ng, A. (2005). Discriminative learning of Markov random fields for segmentation of 3d scan data. In *In Proc. of the Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 169–176.

Biegelbauer, G. and Vincze, M. (2007). Efficient 3D Object Detection by Fitting Superquadrics to Range Image Data for Robot's Object Manipulation. In *IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy*.

Bley, F., Schmirgel, V., and Kraiss, K.-F. (2006). Mobile Manipulation Based on Generic Object Knowledge. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*.

Bone, G., Lambert, A., and Edwards, M. (2008). Automated Modeling and Robotic Grasping of Unknown Three-Dimensional Objects. In *Proceedings of the 2008 IEEE International Conference on Robotics and Automation, Pasadena, USA*.

Collet, A., Berenson, D., Srinivasa, S. S., and Ferguson, D. (2009). Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation. In *IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan*.

De Raedt, L. (2008). *Logical and Relational Learning*. Cognitive Technologies. Springer.

Feldman, D. and Weinshall, D. (2008). Motion Segmentation and Depth Ordering Using an Occlusion Detector. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(7):1171–1185.

Fellegi, I. P. and Sunter, A. B. (1969). Ad//dx.doi.org/10.2307/2286061. *Journal of the American Statistical Association*, 64(328):1183–1210.

Galindo, C., Fernández-Madrigal, J.-A., González, J., and Saffiotti, A. (2008). Robot task planning using semantic maps. *Robot. Auton. Syst.*, 56(11):955–966.

Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., and et al. (2005). Multi-Hierarchical Semantic Maps for Mobile Robotics.

Getoor, L. and Taskar, B. (2007). *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

[3]http://ias.cs.tum.edu/publications
[4]http://www.willowgarage.com/pages/pr2
[5]http://www.robotcub.org

Grau, O. (1997). A scene analysis system for the generation of 3-D models. In *NRC '97: Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, page 221.

Intille, S., Larson, K., Tapia, E. M., Beaudin, J., Kaushik, P., Nawyn, J., and Rockinson, R. (2006). Using a live-in laboratory for ubiquitous computing research. In Fishkin, K., Schiele, B., Nixon, P., and Quigley, A., editors, *Proceedings of PERVASIVE 2006*, volume LNCS 3968, Berlin Heidelberg. Springer-Verlag.

Iocchi, L. and Pellegrini, S. (2007). Building 3D maps with semantic elements integrating 2D laser, stereo vision and INS on a mobile robot. In *2nd ISPRS International Workshop 3D-ARCH*.

Lai, K. and Fox, D. (2009). 3D laser scan classification using web data and domain adaptation. In *Proceedings of Robotics: Science and Systems*, Seattle, USA.

Landwehr, N., Gutmann, B., Thon, I., Philipose, M., and Raedt, L. D. (2007). Relational transformation-based tagging for human activity recognition. In *Proceedings of the 6th Workshop on Multi-Relational Data Mining (MRDM)*, Warsaw, Poland.

Lepetit, V. and Fua, P. (2006). Keypoint recognition using randomized trees. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9):1465–1479.

Liao, L., Fox, D., and Kautz, H. (2007a). Extracting places and activities from gps traces using hierarchical conditional random fields. *International Journal of Robotics Research*.

Liao, L., Patterson, D., Fox, D., and Kautz, H. (2007b). Learning and inferring transportation routines. *Artificial Intelligence*.

Liu, Y., Emery, R., Chakrabarti, D., Burgard, W., and Thrun, S. (2001). Using EM to Learn 3D Models of Indoor Environments with Mobile Robots. In *ICML*, pages 329–336.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Miller, A. and Allen, P. K. (2004). Graspit!: A Versatile Simulator for Robotic Grasping. *IEEE Robotics and Automation Magazine*, 11(4):110–122.

Modayil, J. and Kuipers, B. (2004). Bootstrap learning for object discovery. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-04)*, pages 742–747.

Mozos, O. M., Rottmann, A., Triebel, R., Jensfelt, P., and Burgard, W. (2006). Semantic Labeling of Places using Information Extracted from Laser and Vision Sensor Data. In *In Proceedings of the IEEE/RSJ IROS Workshop: From sensors to human spatial concepts*, Beijing, China.

Mozos, O. M., Triebel, R., Jensfelt, P., Rottmann, A., and Burgard, W. (2007). Supervised Semantic Labeling of Places using Information Extracted from Laser and Vision Sensor Data. *Robotics and Autonomous Systems Journal*, 55(5):391–402.

Neville, J. and Jensen, D. (2003). Collective classification with relational dependency networks. *Journal of Machine Learning Research*, 8:2007.

Nüchter, A. and Hertzberg, J. (2008). Towards semantic maps for mobile robots. *Journal of Robotics and Autonomous Systems (JRAS), Special Issue on Semantic Knowledge in Robotics*, 56(11):915–926.

Okada, K., Kojima, M., Tokutsu, S., Maki, T., Mori, Y., and Inaba, M. (2007). Multi-cue 3D object recognition in knowledge-based vision-guided humanoid robot system. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2007.*, pages 3217–3222.

Pentney, W., Philipose, M., Bilmes, J., and Kautz, H. (2007). Learning large scale common sense models of everyday life. In *Proceedings of AAAI 2007*, Vancouver BC.

Posner, I., Cummins, M., and Newman, P. (2008). Fast Probabilistic Labeling of City Maps. In *Proceedings of Robotics: Science and Systems*, Zurich.

Quack, T., Ferrari, V., Leibe, B., Van Gool, L., and Zurich, E. (2007). Efficient mining of frequent and distinctive feature configurations. In *IEEE 11th International Conference on Computer Vision (ICCV)*, pages 1–8.

Richtsfeld, M. and Vincze, M. (2008). Grasping of Unknown Objects from a Table Top. In *Workshop on Vision in Action: Efficient strategies for cognitive agents in complex environments*.

Savarese, S. and Fei-Fei, L. (2007). 3D generic object categorization, localization and pose estimation. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8.

Saxena, A., Driemeyer, J., and Ng, A. Y. (2008). Robotic Grasping of Novel Objects using Vision. *The International Journal of Robotics Research*, 27(2):157–173.

Saxena, A., Sun, M., and Ng, A. (2007). Learning 3-D Scene Structure from a Single Still Image. *IEEE 11th International Conference on Computer Vision (ICCV), 2007.*, pages 1–8.

Schnabel, R., Wahl, R., and Klein, R. (2007). Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*, 26(2):214–226.

Singla, P. and Domingos, P. (2006). Entity Resolution with Markov Logic. In *In ICDM*, pages 572–582. IEEE Computer Society Press.

Taskar, B., Wong, M. F., Abbeel, P., and Koller, D. (2003). Link prediction in relational data. In *in Neural Information Processing Systems*.

Thrun, S. and Wegbreit, B. (2005). Shape from symmetry. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Bejing, China. IEEE.

Triebel, R., Kersting, K., and Burgard, W. (2006). Robust 3D scan point classification using associative markov networks. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 2603–2608.

Triebel, R., Óscar Martínez Mozos, and Burgard, W. (2007a). Relational Learning in Mobile Robotics: An Application to Semantic Labeling of Objects in 2D and 3D Environment Maps. In *Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications (GfKl)*, Freiburg, Germany.

Triebel, R., Schmidt, R., Mozos, O. M., , and Burgard, W. (2007b). Instace-based AMN classification for improved object recognition in 2d and 3d laser range data. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), (Hyderabad, India)*.

Ulrich, M., Wiedemann, C., and Steger, C. (2009). Cad-based recognition of 3d objects in monocular images. In *International Conference on Robotics and Automation*, pages 1191–1198.

Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. (2007). Cognitive maps for mobile robots-an object based approach. *Robot. Auton. Syst.*, 55(5):359–371.

Welke, K., Asfour, T., and Dillmann, R. (2008). Object separation using active methods and multi-view representations. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 949–955.

Wuenstel, M. and Moratz, R. (2004). Automatic Object Recognition within an Office Environment. In *CRV '04: Proceedings of the 1st Canadian Conference on Computer and Robot Vision*, pages 104–109.

Yan, P., Khan, S., and Shah, M. (2007). 3D Model based Object Class Detection in An Arbitrary View. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–6.

Zhang, Y., Koschan, A., and Abidi, M. (2004). Superquadric Representation of Automotive Parts Applying Part Decomposition. *Journal of Electronic Imaging, Special Issue on Quality Control by Artificial Vision, Vol. 13, No. 3*, pages 411–417.

2 TRAINING

Clarity and quality of the research training objectives for the researcher
How have I integrated and performed so far, how will that continue **Relevance and quality of additional research training as well as of transferable skills offered**
How will I get better? **Host expertise in training experienced researchers in the field and capacity to provide**
mentoring/tutoring (outgoing and return host) Continuation of tutoring, assistance through ROS, teaching

3 RESEARCHER

Research experience
Research results including patents, publications, teaching etc., taking into account the level of experience
Independent thinking and leadership qualities
Match between the fellow's profile and project
Potential for reaching a position of professional maturity
Potential to acquire new knowledge

4 IMPLEMENTATION

Quality of infrastructure / facilities and international collaboration of host (outgoing and return host)
Practical arrangements for the implementation and management of the research project (outgoing and return host)
Feasibility and credibility of the project, including work plan
Practical and administrative arrangements and support for the hosting of the fellow (outgoing and return host)

5 IMPACT

Potential for acquiring competencies during the fellowship to improve the prospects of reaching and/or reinforcing a position of professional maturity, diversity and independence, in particular through exposure to transferable skills training
Contribution to career development or re-establishment, where relevant
Potential for creating long term collaborations and mutually beneficial cooperation between Europe and the Other Third Country.
Contribution to European excellence and European competitiveness
Benefit of the mobility to the European Research Area
Impact of the proposed outreach activities

6 ETHICS ISSUES