

Maps, Places, and Worlds for Robots

James J. Little
Computer Science
University of British Columbia
Vancouver, BC, Canada

Abstract—Vision is a powerful sense that permits a robot to look around itself and gather information both about the immediate present and the near future. The future arrives through the more distant physical space through which the robot can move, and the possible actions and events that may arise. To know the future a robot needs to parse the world with the aid of its models and experience.

Lasers and other active sensors have proven their ability to provide accurate geometric information. But context and the meaning of the space surrounding the robot, the objects and the actions they permit are only accessible with the more complete sensory input of vision. Vision as a sensor is computationally demanding, but resources have improved to make it practical.

Moreover there has been a convergence of the interests of roboticists and vision scientists – both want to explore and act in the world. Many of us have accepted that we must learn the patterns of data using machine learning, but we must also integrate categorical descriptions of our world, prototypical information that no individual robot is yet capable of learning. Vision provides the anchoring for concepts. I will discuss recent advances and trends linking vision and robotics through spatial descriptions and the connections with objects, actions, and meaning.

I. INTRODUCTION

When you have brought your new Apple iRobot¹ home for the first time, you are faced with a challenging task – introducing the robot to its new home/workspace. Of course the robot knows about homes and typical tasks. That’s why you bought the sleek, stylish robot, in addition to the fact that it promised a simple interface. It prompts you to take it on a tour of the house, naming the rooms, pointing out the appliances, and identifying the occupants of the house.

The promise of the now discontinued Aibo – communication and simple behaviours – shows that even simple visual sensors using strong features (SIFT[18]) can enable interesting visual tracking and recognition. Built in to the home robot will be the necessary concepts – tasks, objects, locations. Your home vacuum robot “knows” only about stairs, objects, infrared walls, and random search. Your iRobot knows about kitchens, doors, stairs, bedrooms, beer (you have the party version of the iRobot that can bring you beer in the entertainment room). How does it tie the sensory flow it receives to its plans, names, and goals in its repertoire? One could imagine that the home robot could explore the home unaided, and then upload its sensory data to a human assistant who could point out the locations of

¹Apologies to two major corporations that may be clashing over naming rights in the future.



Fig. 1. SIFT features found by our stereo-equipped robot. SIFT features in the three stereo images are matched; horizontal and vertical lines indicate the horizontal and vertical disparities respectively (from [17]).

object relevant to the robot’s behaviours, knowing the types of situations modeled in the robot.

This fanciful thought experiment is not so far in the future, but some applications such as assistive technologies[22], [1] operate in contexts where rich visual sensing is deployed, but the range of objects may be limited, say, for example, in a care facility where patients’ rooms are more constrained in their contents. Here it may be effective to learn the connection between features of the visual stream and their connections to world states and sequences of actions.

One missing element is the structure of the environment: the objects in the world, their functional relations, and their spatial layout.

Visual sensing offers some solutions. There are many ways to compute features or keypoints useful for localization and detection/recognition. Reference Lowe[18] (see Mikolajczyk and Schmid[23] for a discussion of the performance of feature detectors). Many solutions, in recognition, mapping and localization, and learning object appearance begin with these interest points.

A. Mapping the world

Laser and other active sensors are not only becoming more compact, less power hungry, and less expensive, but also our techniques for solving mapping and localization (SLAM) have become increasingly powerful[37]. Solutions based on passive visual sensors, both monocular[2] and stereo[28],

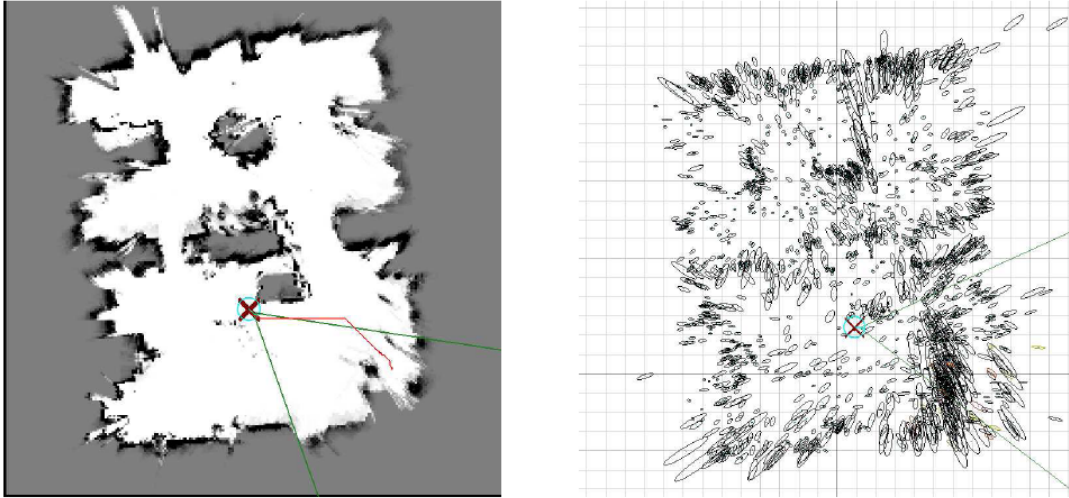


Fig. 2. (a) Occupancy grid constructed for the maximum-likelihood sample at the end of exploration. (b) Landmark map constructed for the maximum-likelihood sample at the end of exploration (from [30]).

[29], [30], are capable also of delivering both excellent localization in and high-quality geometry descriptions of unstructured environments. Figure 1 shows a set of features during a robot’s tour of our lab, with their disparities. As the robot moves in the world, it aggregates local stereo information into an occupancy grid[4] where the robot plans paths and represents the structure of solids and voids. stereo sensor delivers real-time data about the presence of obstacles, the full arsenal of probabilistic methods for localization, and estimation of actions are available[29], [30]. Figure 2) shows the occupancy grid constructed for the maximum-likelihood sample at the end of exploration, and the landmark map constructed for the maximum-likelihood sample at the end of a traversal of the lab; error ellipses show the spatial uncertainty of the SIFT feature points (projected from 3d). Note the density of the features.

During map-building the process of localizing the landmarks eliminates the dynamic scene elements. We intend with long-term use of the maps will enable us to track configurations of features that move rigidly and so label some scene elements as movable – we should then be able to enhance our map with parts of the environment. Section III builds on this insight and close observation of moving objects to isolate them and build 3d model shape and appearance models.

B. Visual context

Visual sensing can also situate the operation of a robot in a larger context: providing image and scene contexts, object categories, object recognition, actions, and intentions. Connecting the robot to the task and the elements of the scene simplifies the sensing requirements, by narrowing down the focus. We will see later how *José* used this aspect of vision.

How do we plan and operate in a higher-level description? In pioneering work, Rimey and Brown[27] showed a

selective vision system that used Bayesian estimation and decision-theoretic control to decide where and how to gather evidence. We will see later (Section IV) how POMDPs have become the model of choice for sensing and action.

Context tells the robot where to look and links with the task through the types of locations and objects involved in the task. Context directs visual processing by informing the robot “where to look”. The task combines “where” and “when to look”. All depend on models.

“When to look” depends on the sequence of the task and exogenous events that may not be consistent with the normal staging of a robot’s operations. Control of attention helps the robot cope with them. Most models of visual attention[13], [40] depend on a model of the visual saliency or novelty of a visual feature or phenomenon, whatever its modality, and model the spatial structure of the visual attention path and its temporal course. These are essentially bottom-up. The active realization of this concept is shown in Elder [3] where the high-resolution camera is directed by low-resolution cues to salient regions.

The innovation in Itti and Baldi’s new model[12] is to draw the attention to spots where the visual event is unlikely given a model of the image. This introduces a component of the scene and its content via the prior model.

Context can direct attention in a top-down fashion, moving from a global estimation of the category of the scene to relative spatial location of interesting objects in the scene. [38] introduces the concept of the “gist” of the scene, the visual characteristics (cues) that enable us without close inspection to determine the category of a location. They demonstrated that one could learn the gist, described in terms of responses to banks of filters, and then discriminate broad categories of scenes such as corridors, offices, and streets. Whereas localization and recognition using feature points employs specific discriminative features, gist identification targets descriptors that correspond more to spatial organiza-

tion and structure, such as parallel, rectangularly-orientation, or clutter.

In the spirit of [5] we wish to connect semantic and spatial information. They use a representation of the local appearance of an object and tag the world map with that appearance representation so that it can be indexed later. Likewise Vasudevan et al.[41] take the first steps toward building world models that connect spatial representations with the labeled content of the scene.

Relative spatial location – configuration of objects – depends on their relation to tasks – unless the environment is in disorder. It is relatively unlikely that a desk, for example, will be arranged so that the objects next to the keyboard and beside the monitor are automotive parts. There is a need for stronger perception of objects, stability, continuity, identity before identification. Low level processes can tease out identity and coherence, which precedes the process of categorization, but stronger spatial information can direct processing to appropriate locations.

By teaching a robot the categories of the objects in its environs we can avoid the difficult problem of categorization. But will RFID technology obviate this work entirely, by making object self-identifying? Likely in some not too near future, when all legacy appliances have been abandoned, but the task will remain problematic for some while.

In [10] Bill Gates predicts that robotics will be an exciting field in the coming years, while promoting Microsoft Robotics Studio as the software substrate for solving the difficult control software problems such as concurrency and coordination.

As mentioned in Gates' article, it can be assumed that the robot agent often works in an environment where networks of cameras observe the activity of the people and the robot. Many projects have constructed test homes rich with sensors and context aware components. It has become apparent that many of the techniques that have been developed with a view toward applications in surveillance, which often use networks of cameras, also apply to the world of service robots[21]. The robot can also be instrumental in recovering the relative geometry of the cameras as it explores its world[26].

Recent workshops[43], [44], [45] have highlighted the growing importance of the cognitive dimension of robotics and the necessary to import into robotics the tools of reasoning and planning.

II. STRUCTURING SPACE

Others have described the connection between metric conceptions and representations and topological representations that connect regions or places. In [15] Kuipers describes the Spatial Semantic Hierarchy that builds representations of the spatial domain at multiple levels: sensorimotor, control, procedures, topology, geometry. The base level present in many sensed representations is geometry, where metric quantified information about the position and orientation of objects is maintained. The more tractable topological representation captures paths and connections between distinguished locations.

Our robotic systems are very well localized in a map whose structure is either a single large map, or an atlas (by analogy to manifolds) of charts, each of which is a local spatial representation.

We can look to research in ontological description of objects[36] to assist us. Researchers have recently explored two important aspects of increasing usability of systems – using external knowledge and the structured information about the world in the form of ontologies[36].

We were motivated to pursue this by the work of Kautz and his colleagues[25] on assistive technologies where they track people in their homes and observe their everyday activities using RFID tags to sense the people as they use tagged objects. In a sensed home with cameras the limitations of RFID tagging can be overcome[21]. The objects are labeled with their description, for example, paring knife, spoon, cup, jug, and so forth. But classification of activities is hampered by the level of detail of the labels. To improve classification, they use an ontology, a description of the objects in the world, and their properties. Ontologies are hierarchical, and provide a variety of levels of abstraction of labels (properties), so that paring knives and butter knives are knives, while knives and spoons are both utensils, and whisks are utensils also, and so forth. By choosing the right level of abstraction, action classification improves.

Surprisingly it is more and more possible to find information supplying ontological description for informal situations and objects, amazingly available for many categories of objects[36]. Ontologies for more formal and technical subjects, such as geographical entities[19], have been extensively studied.

We propose that using such descriptions is a central problem for home robotics. The home description will be a useful tool, describing the objects in the home: rooms, appliances, furniture, and so forth.

How will the robot acquire this? By pointing, demonstration, download, exploration? If exploration, we may face several obstacles in sharing ontologies with the robot. Once the robot has learned the structure of the environment, will it partition the world into events homeomorphic to some of our concepts? There certainly will be hidden correlations between the objects and situations in our world and the actions they foster, and an exploring robot will uncover some novel dependencies.

We do not yet have hierarchical knowledge of categories of objects, nor their appearance, but extensive current work on representing the appearance and shape of objects[16] allows us to use these appearance models in recognition. These models need not be acquired from images, but their shape can be gotten from CAD models in the design process. Our work on object discovery (Section III) goes part way to this goal in an unstructured environment. Object discovery is not recognition, as in [5], but creation of models of scene elements by identifying independently moving parts of the scene and then growing a 3d model over several observations.

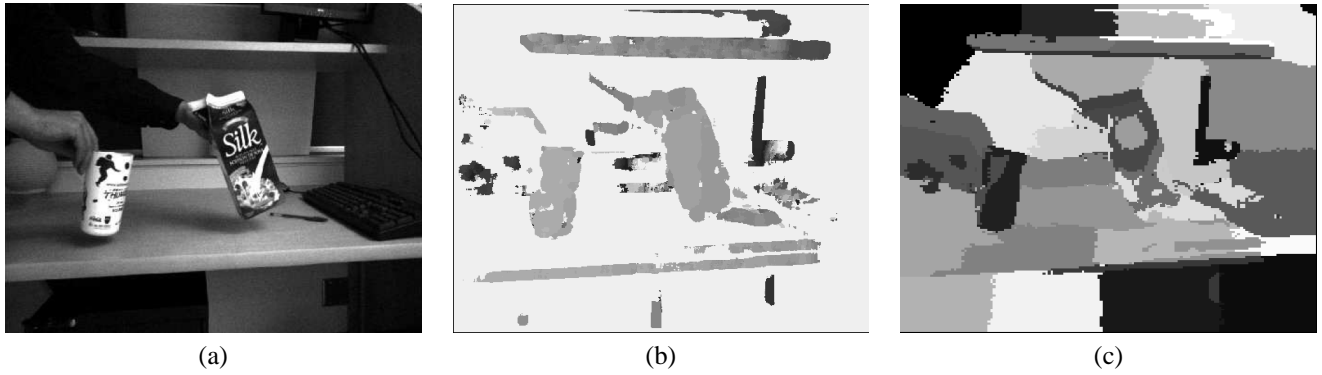


Fig. 3. (a) Input image for object discovery by motion, appearance, and shape (b) depth image from stereo (c) segmentation of the image and depth map using normalized cuts based on both the pixel intensity and depth.

III. FEATURE-BASED OBJECT DISCOVERY

We recovering object descriptions[35] from coherent structures isolated from their surroundings by differences: depth or, more generally, parallax and appearance, that provide cues to the identity over time of an object. Our work is similar in spirit to [32] where they build models from two types of features: interest point similar to SIFT, and MSER[20], tracking the features over long sequences, and then using the appearance models aggregated over the frames to match objects in other frames of the movie.

We examine the problem of *object discovery*, autonomous acquisition of object models, using a combination of shape, appearance and motion. We propose a novel multi-stage technique for detecting rigidly moving objects and modeling their appearance for recognition. First, a stereo camera is used to find a sequence of images and depth maps of a given scene (Figure 3).

Then the scene is oversegmented using normalized cuts[31], We use Normalized Cuts (NCuts), a technique for segmenting a graph based on a weight function between nodes. NCuts can be applied to an image by treating the pixels as nodes. Our weight function is based on the difference in intensity, depth and 2D image position between pixels (Figure 3(c)). SIFT features[18] are matched between sequential pairs of images to identify groups of rigidly moving features; the 3d movement of these features determines which regions in the segmentation of the scene correspond to rigid objects, grouping oversegmented regions as necessary (see Figure 4).

Segmentation links with the rigidly moving points (core points) so that we can tie regions in the image, with appearance models, to the point sets. A voting process adds new feature points (adjunct points) to the aggregated point sets. These additional features are extracted from segmented regions and combined with the rigidly moving image features to create snapshots of the object’s appearance (see Figure 4). Over time, even when objects have ceased moving, these snapshot are combined to produce models that are effective for recognition. The models contain shape information, since all feature points contain depth information, and appearance, from the aggregation of image regions.

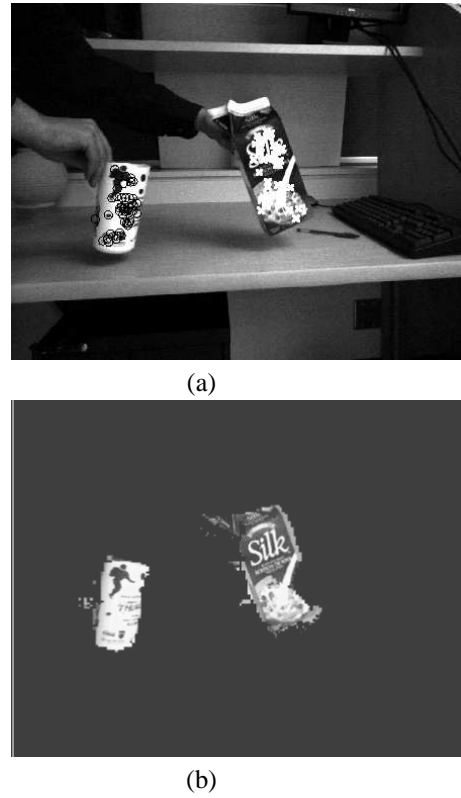


Fig. 4. (a) This image shows the position of every feature within the regions corresponding to the rigidly moving features in the input image, both core features and adjunct features. Features on cup on the left are indicated with Os while those on the milk box on the right are indicated with Xs. (b) Object snapshots of the cup and the milk box acquired through region voting.

Our object discovery method envisions a robot system seeing moving objects – we have ourselves moved the objects in our experiments, but in the future the robot will continually revisit the scene. Objects that have moved can be indexed through their SIFT features and our process applied over pairs of images where the robot has moved instead. Other methods (Ferrari et al.[8], for example) extend local groups of features to connect them into an appearance model that can be used in recognition.

Our current research plans entertain the possibility of learning the spatial relations between objects in the scene. Constellation models[9] encode varying spatial configurations of model parts for recognition by Gaussian distributions of relative part location learned from examples. Conditional random fields now can describe spatial relations between regions[39], [42]. Both of these representations are 2d, where full object spatial relations would have to be expressed in 3d. Where do we get the information for learning spatial configurations of objects in a robot’s world? Several possibilities arise: housing plans, virtual worlds as in electronic games, online contests such as the ESP game, and finally the world itself, via exploration, and object recognition and discovery.

IV. ACTING IN SPACE

The role of *José* [6] uses our visual processes, localization, mapping, navigation and human-robot interaction, in the context of a particular robotic task: serving food to a gathering of people. To accomplish this task, a robot must reliably navigate around a room populated by groups of people, politely serving appetizers to humans. The robot must also monitor the food it has available to serve, and return to a home base location to refill when the food is depleted. Problems specific to the serving task were also solved using vision, including finding people to serve and monitoring food. As well, the robot’s success depends significantly on its interaction with users, its “persona”, and its ability to generate appropriate actions and responses. In *José’s* planner, we identified specific points in the world with the tasks such as refilling the food. Similarly a service robot will have in its behaviours defined locations, where activities occur: the kitchen, the laundry, for example. Tying the location and its appearance to the sensory process is the problem of *grounding*, of determining the connection between beliefs and physical objects and situations, through the sensed data.

We argued in [17] that robots can be modeled as a set of capabilities, processes that enable behaviours. A collection of tasks can be performed in a role. For example, our robot enacted the *José* role of waiting as well as the Homer (Human Oriented MEssenger Robot)[7] role. It is tedious if not infeasible to engineer each role for the robot out of behaviours. It is more sensible to model the environment, the robot, and its actions by a POMDP[17], which has proven its worth in complex assistive applications[1].

A. Selecting features for actions

In many cases tasks[17] organize space into sequences of locations, where the linkage between spaces. When designing planners, whether based on traditional, though reactive planners, as in [24], or POMDPs, there is a classification or recognition step in which the fluent in the planner or the state variable in the POMDP estimates its discrete state.

In [11] we demonstrated how to select features that communicate in a game, essentially finding the correct aspect of the sensory stream to correspond with the action. However, it is feasible to link the actual elements of the visual stream

that are relevant to a particular task, by representing the connections between the utility of the action and the features within the confines of a task that exposes the effect of the visual signal. Reinforcement learning is the usual solution in such cases, but it is infeasible to test thoroughly state space, which requires an enormous number of experiments.

The solution we suggest is to “teach” the robot, rather than have it learn by trial and error. That is, the robot must import the cultural descriptions that adhere to locations, which take their meaning from the actions that occur at those places. The form of these description will be ontologies of objects and models of their dynamics under actions, expressed as Bayes nets, for example. For visually guided robots, we will need to ground the entities in appearances, and connect them with their spatial organization. Then we can sense, reason, decide, and act with them. Already we have some of the tools for connecting hierarchical spatial descriptions with probabilistic reasoning (Smyth and Poole, [34]). There, a system is developed for reasoning about general hierarchical models combined with qualitative information about the distribution of properties; the running example is rooms and types of rooms in a house.

V. SUMMARY

The challenge of visually guided robotics, particularly as partners with humans, is to progress from the laboratory into the home. Within the context of assistive technologies, much more limited success can be rewarding, but it is combined with the obstacle of greater risks and liabilities.

By endowing locations with semantic tags, cultural information situates the embedded system within a richer context. Would the robotic system be able to share context with other confederates throughout the world?

Because vision is capable of gathering information not only near but in the middle distance and far away it supports short, middle and long-term prediction of actions. The accuracy of these predictions depends on the ability to assess the spatial priors over contexts and objects. In any particular application, there has to be a balance between narrow focus (feature detection for localization, for example) and broad support (context identification, categorization), and the tradeoff between these two will be driven by the reliability of visual sensing.

The challenge at hand is to acquire the structured information about the world, the ontology of objects and their spatial organization and appearance, connect it to maps, and formulate robotic controllers to accomplish complex tasks. Exploration, recording of layout and appearance, and mining ontologies will deliver valuable maps, contexts, and theories of the world.

VI. ACKNOWLEDGMENTS

The author gratefully acknowledges the contribution of the National Science and Engineering Research Council of Canada and GEOIDE, a Canadian Network of Centres of Excellence. As well, many past and present members of our lab have contributed to the development of the ideas explored here.

REFERENCES

- [1] Jen Boger, Jesse Hoey, Pascal Poupart, Craig Boutilier, Geoff Fernie, and Alex Mihailidis, A Planning System Based on Markov Decision Processes to Guide People with Dementia Through Activities of Daily Living, *IEEE Transactions on Information Technology in BioMedicine*, 10(2), 323-333, 2006.
- [2] Andrew Davison, Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the IEEE Int. Conf. on Computer Vision*, pages 1403–1410. Nice, France, 2003.
- [3] James H. Elder, Simon J.D. Prince, Y. Hou, M. Sizintsev, and E. Oleviskiy, Pre-attentive and attentive detection of humans in wide-field scenes, *International Journal of Computer Vision*, in press.
- [4] Alberto Elfes, "Using occupancy grids for mobile robot perception and navigation", *IEEE Computer*, 22(6), pages 46–67, 1989.
- [5] Staffan Ekvall, Patric Jensfelt, and Danica Kragic, Integrating Active Mobile Robot Object Recognition and SLAM in Natural Environments, *IROS 06*, 2006.
- [6] Pantelis Elinas, Jesse Hoey, Darrell Lahey, Jefferson D. Montgomery, Don Murray, Stephen Se, and James J. Little, Waiting with José, a vision-based mobile robot, *IEEE International Conference on Robotics and Automation*, May 2002, pages 3698–3705.
- [7] Pantelis Elinas, Jesse Hoey and James J. Little, HOMER: Human Oriented Messenger Robot, *AAAI Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*, March 2003, pp. 45–51.
- [8] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool, Simultaneous Object Recognition and Segmentation from Single or Multiple Model Views, *Int. J. Comput. Vision*, Vol. 67(2), 2006, pages 159–188.
- [9] Rob Fergus, Pietro Perona, and Andrew Zisserman, Object Class Recognition by Unsupervised Scale-Invariant Learning, *CVPR 03*, Vol. 2, pages 264–271.
- [10] Bill Gates, A Robot in Every Home, *Scientific American*, Vol. 296, No. 1, January 2007, pages 58–65.
- [11] Jesse Hoey and James J. Little, Value-Directed Human Behavior Analysis from Video using Partially Observable Markov Decision Processes, in press, *IEEE PAMI*.
- [12] Laurent Itti and Pierre Baldi, A Principled Approach to Detecting Surprising Events in Video, *Proc. IEEE CVPR*, 2005, pages 631–637.
- [13] Laurent Itti, Christof Koch, and Ernst Niebur, A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans PAMI*, 20(11):1254–1259, 1998.
- [14] Danica Kragic, Márten Björkman, Henrik I. Christensen, Jan-Olof Eklundh, Vision for robotic object manipulation in domestic settings, *Robotics and Autonomous Systems*, Vol. 52(1), pages 85–100, 2005.
- [15] Benjamin Kuipers, The Spatial Semantic Hierarchy. *Artificial Intelligence* 119: 191-233, 2000.
- [16] Hendrik P.A. Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel, Image-based reconstruction of spatial appearance and geometric detail, *ACM Trans. on Graphics (TOG)*, Vol 22(2), pages 234–257, 2003.
- [17] James J. Little, Jesse Hoey, and Pantelis Elinas, Visual Capabilities in an Interactive Autonomous Robot, in *Cognitive vision systems : sampling the spectrum of approaches*, eds.: Christensen, Henrik I.; Nagel, Hans-Hellmut. – Berlin : Springer, 2006, VIII, 365 S. – (Lecture notes in computer science)
- [18] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pages 91-110.
- [19] David M. Mark, Barry Smith, and Barbara Tversky, Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. In *Freksa, C., and Mark, D. M., Editors, Spatial Information Theory: A Theoretical Basis for GIS*, Berlin: Springer-Verlag, Lecture Notes in Computer Science No. 1661, pp. 283-298, 1999.
- [20] Jiri Matas, Ondrej Chum, Martin Urban, and Tomas Pajdla, Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC.*, pages 384–393, 2002.
- [21] Gérard G. Medioni, Alexandre R.J. François, Matheen Siddiqui, Kwangsu Kim and Hosub Yoon, Robust Real-Time Vision for a Personal Service Robot, to appear, *Computer Vision and Image Understanding*, special issue on Human-Computer Interaction.
- [22] Alex Mihailidis, Pantelis Elinas, Jen N. Boger, and Jesse Hoey, Pervasive Computing to Enable the Mobility of Older Adults with Cognitive Impairments: An Anti-Collision System for a Powered Wheelchair, *IEEE Transaction on Neural Systems and Rehabilitation Engineering* (to appear).
- [23] K. Mikolajczyk and C. Schmid, A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27(10), pages 1615–1630, 2005.
- [24] Armin Mueller and Michael Beetz, Designing and Implementing a Plan Library for a Simulated Household Robot, *Cognitive Robotics Workshop*, AAAI 2006.
- [25] William Pentney, Henry Kautz, Matthai Philipose, Ana-Maria Popescu, and Shiao-kai Wang, Sensor-Based Understanding of Daily Life via Large-Scale Use of Common Sense, *Proceedings of the 26th Annual Conference of AAAI*, 2006.
- [26] Ioannis Rekleitis, David Meger, and Gregory Dudek, Simultaneous Planning, Localization, and Mapping in a Camera Sensor Network. *Robotics and Autonomous Systems*, 54(11), pages 921–932, 2006.
- [27] Ray D. Rimey and Chris M. Brown, Control of Selective Perception using Bayes Nets and Decision Theory, *IJCV*, 12(2-3), 1994, pp173–207.
- [28] Stephen Se, David Lowe and James J. Little, Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Landmarks, *Intl. Journal of Robotics Research*, Vol. 21, No. 8, August 2002, pages 735–758.
- [29] Robert Sim, Pantelis Elinas and James J. Little, A study of the Rao-Blackwellised particle filter for efficient and accurate vision-based SLAM, in press, *International Journal of Computer Vision/International Journal of Robotics Research Special Joint Issue on Vision in Robotics*
- [30] Robert Sim and James J. Little, Autonomous vision-based exploration and mapping using hybrid maps and Rao-Blackwellised particle filters, *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, Beijing, 2006.
- [31] Jianbo Shi and Jitendra Malik, Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8):888–905, 2000.
- [32] Josef Sivic, Frederik Schaffalitzky, and Andrew Zisserman, Object level grouping for video shots. In *Proceedings of the European Conference on Computer Vision*, 2004, pages 85–98.
- [33] Aaron Sloman, Jeremy Wyatt, Nick Hawes, Jackie Chappell, and Geert-Jan M. Kruijff, Long Term Requirements for Cognitive Robotics, *Cognitive Robotics Workshop*, AAAI 2006.
- [34] Clinton Smyth and David Poole, Qualitative Probabilistic Matching with Hierarchical Descriptions. *KR 2004*: 479-487
- [35] Tristram Southery and James J. Little, Object Discovery using Motion, Appearance and Shape, *Cognitive Robotics Workshop*, AAAI 2006.
- [36] E. Munguia Tapia, T. Choudhury, and M. Philipose, "Building Reliable Activity Models Using Hierarchical Shrinkage and Mined Ontology," in *Proceedings of PERVASIVE 2006*, B. Heidelberg, Ed. Dublin, Ireland: Springer-Verlag, 2006.
- [37] Sebastian Thrun, Dieter Fox, and Wolfram Burgard, Monte Carlo localization with mixture proposal distribution. In *Proceedings of the 2000 National Conference of the American Association for Artificial Intelligence (AAAI)*, pages 859–865, 2000.
- [38] Antonio Torralba, Kevin P. Murphy, William T. Freeman and Mark Rubin, Context-based vision system for place and object recognition, *ICCV*, 2003, pages 273–280.
- [39] Antonio Torralba, Kevin P. Murphy and William Freeman, Contextual Models for Object Detection using Boosted Random Fields, *NIPS'04*.
- [40] John K. Tsotsos, Sean M. Culhane, Winky Y.K. Wai, Yuzhong H.Lai, Neal Davis, and Fernando Nuflo, Modeling visual-attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–545, 1995.
- [41] Shrihari Vasudevan, Stefan Gächter, Marc Berger, Roland Siegwart, *Cognitive Maps for Mobile Robots – An Object based Approach*, From sensors to human spatial concepts (geometric approaches and appearance-based approaches), Workshop at IROS 2006, Beijing.
- [42] John Winn and Jamie Shotton. The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects, *CVPR06*, Vol. 1, pages 37–44.
- [43] From sensors to human spatial concepts (geometric approaches and appearance-based approaches), Workshop at IROS 2006, Beijing.
- [44] *Cognitive Robotics Workshop*, AAAI 2006, Boston.
- [45] *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006)*, Workshop Toward Cognitive Humanoid Robots, 4 December 2006, Genoa, Italy.