Indoor Mobile Robotics at Grima, PUC

L. Caro · J. Correa · P. Espinace · D.
Maturana · R. Mitnik · S. Montabone · S.
Pszczółkowski · D. Langdon · A. Araneda ·
D. Mery · M. Torres · A. Soto

Abstract This paper describes the main activities and achievements of our research group on Machine Intelligence and Robotics (Grima) at the Computer Science Department, Pontificia Universidad Catolica de Chile (PUC). Since 2002, we have been developing an active research in the area of indoor autonomous social robots. Our main focus has been the cognitive side of Robotics, where we have developed algorithms for autonomous navigation using wheeled robots, scene recognition using vision and 3D range sensors, and social behaviors using Markov Decision Processes, among others. As a distinguishing feature, in our research we have followed a probabilistic approach, deeply rooted in machine learning and Bayesian statistical techniques. Among our main achievements are an increasing list of publications in main Robotics conference and journals, and the consolidation of a research group with more than 25 people among full-time professors, visiting researchers, and graduate students.

1 Introduction

Robotics is reaching a level of maturity that is starting to allow robots to move out of research labs. During the last years, diverse companies have emerged offering a first generation of robots aimed to operate in partially structured applications such as lawn mowing [60], building surveillance [32][43], and floor cleaning [33][60], among others. This new technological scenario illustrates some of the notable progress in the field, however, the creation of truly adaptive and robust robots is still an open research issue. In particular, research issues such as robust operation in dynamic enviroments, high level semantic understanding, human-computer interaction, reasoning about other

Computer Science Department, School of Engineering
Pontificia Universidad Católica de Chile (PUC)
Vicuña Mackenna 4860
Santiago, Chile

agents, knowledge transfer, real time inference schemes, and integration of behaviors and tasks, appear as some of the research challenges in the cognitive side of Robotics [10].

At Grima [30], our research group on Machine Intelligence, we are facing some of the previous challenges by conducting theoretical and applied research on 3 main areas: Robotics, Computer Vision, and Machine Learning. In particular, in terms of Robotics our focus has been the creation of new techniques for the development of autonomous social mobile robots for the case of indoor environments. Since its inception in 2002, our group has been actively working in this task. At present, our group is composed by 4 full-time faculty professors and more than 20 graduate students, including PhDs and MSc.

Autonomous robots able to interact with people and to perform useful tasks in indoor environments have been a long standing goal for the Robotics community. Leaving aside hardware and locomotion issues, this goal requires the development and interconnection of a large number of robot behaviors such as mapping, localization, path planning, people detection and recognition, and scene undertanding, among many others. Furthermore, the social aspects of a robot behavior requires the development of algorithms able to follow social rules, detect and eventually recognize humans, express emotions, etc. Most of these issues have marked our research agenda, where our main efforts have been focused on the cognitive aspects of mobile robotics.

In this paper, we describe our main research activities and contributions to the area of indoor mobile robotics. Most of our activities are closely connected, however, to improve the clarity of this document, we classify them around 3 main lines of research: autonomous indoor navigation, semantic scene understanding, and social robots. Accordingly, this paper is organized as follows. Section 2 describes our progress in autonomous indoor navigation. Section 3 presents our algorithms for scene understanding. Section 4 presents our social robot, as well as activities using robots in the area of education. Finally, Section 5 presents the main conclusions of this work and future avenues of research.

2 Autonomous Indoor Navigation

Starting from the restrictions of the block world of Shakey [52] and the static world limitations of the Stanford Cart [50], the mobile robotics community has made lots of progress during the last 50 years. The research emphasis has been on providing robots with the ability of autonomously navigate natural environments using information collected by their sensors. In terms of indoor environments, one of the main peculiarities is that globally accurate positioning systems, such as GPS, are not available. As a result, the problems of automatic construction of maps of the environment and accurate estimation of the position of the robot within a map, tasks known as mapping and localization, have been highly relevant. Most practical activities of mobile robots require that the robot knows its location within the environment, and this in turn requires that the robot knows the environment where it is moving. This has led to research considering the problem of mapping and localization at the same time, which has been called the Simultaneous Localization and Mapping (SLAM) problem, or the Concurrent Mapping and Localization (CML) problem [40].

There has been an extensive research literature dealing with mapping and localization for mobile robots, e.g. see [69][19][5] [72]. The pioneering development in this

area was the paper by Smith et al. [63] who proposed a SLAM technique based on the Kalman filter to the problem of estimating topological maps. For the non-Gaussian case, Thrun et al. [71] presented a general approach that can be used with general distribution functions. A more recent successful approach to solve the SLAM problem is the FastSLAM algorithm [49]. This approach applies to topological maps, and is based on a factorization of the posterior distribution of maps and locations. During the last couple years new developments in the area of computer vision have led to intensive research to the so-called area of visual SLAM, where illustrative papers are [14] and [37]. In our case, our focus has been on developing new techniques for the problem of localization and mapping for indoor wheeled robots. Similarly to most works in this area [62] [72], our research evolution has closely followed advances in the perceptual side, starting from early work using sonar and basic visual perception routines [66], then using 2D laser range finders [3][2], and, more recently, incorporating 3D range sensors and more robust semantic vision based solutions [22]. As a distinguishing feature, in all these cases we have followed a probabilistic approach, deeply rooted in machine learning and Bayesian statistical techniques. We describe next some of our contributions to this area.

2.1 Localization and mapping using range sensors

In terms of SLAM, in [3] and [2] we present a complete probabilistic representation to this problem. We consider the case of a robot equipped with an odometer and range sensors, such as sonars or 2D laser range finders. We formalize the problem of mapping as the problem of learning the posterior distribution of the map given sensing information, where the map is represented by an occupancy grid [20]. Our key idea is based on noting that the posterior distribution of the map is determined by the posterior joint distribution of the locations visited by the robot and the distances to the obstacles from those locations. The advantage of this method is that it does not provide a single estimate of the map, as the EM-based solution [71], but it produces multiple maps showing the notion of variability from the expected posterior map.

Figure 1 shows a graphical representation of our probabilistic model, where non-observable variables have been circled for clarity. This representation forms the basis of our model and has some similarities to the work in [51], as well as to probabilistic graphs of the sort we could formulate related to Kalman filter approaches. Our representation was developed in earlier unpublished work and has been implicitly adopted in [72].

According to Figure 1, the distribution of the process is determined by three models. A *motion model* [68] describes the dependency of the current location, Z_t, on the previous one and the current odometer reading, Z_{t-1} and U_t, respectively. We adopt a Gaussian motion model. A *perception model* [68] describes the dependency of laser readings V_t on the true distances to obstacles, θ_t. We adopt a truncated Gaussian distribution, with standard deviation σ. The limits of the distribution correspond to 0 and d_{max}, where d_{max} corresponds to the maximum range of the laser device. Finally, we need a prior distribution for the map, $\mathbf{M}$. We assume that cells in the map are independent, having the same probability of being occupied, p. These models are discussed in greater detail in [2].

Using these models we derive expressions for the posterior joint distribution of locations Z_t and distances θ_t given odometer readings U_t and range readings V_t. We show that there is no closed form solution to the posterior, but by exploiting different
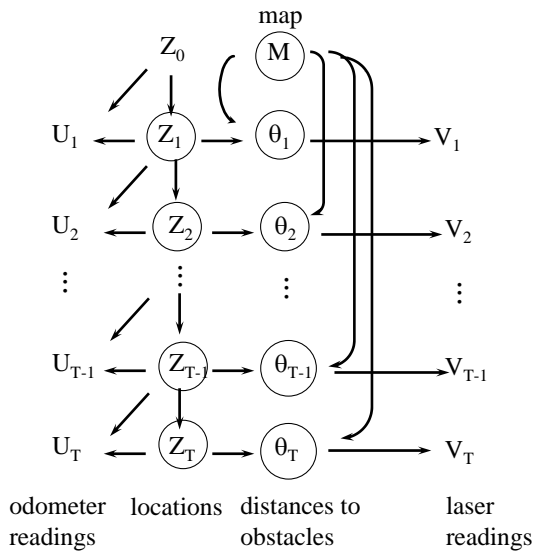
Fig. 1 Graphical representation of our approach to the SLAM problem.

factorizations of this distribution, we derive three sampling algorithms based on importance sampling [2]. We test these algorithms using real data obtained from two indoor office environments at Carnegie Mellon University and Pontificia Universidad Catolica de Chile. In both cases, we obtain maps that closely match ground true information. Figure 2 shows an example of the map obtained using raw odometry and laser data, as well as, the maps obtained using the proposed approach.

After developing tools to build maps from the environment, in [24] we propose a method to achieve real time global localization as well as position tracking. This method takes advantage of structural regularities of indoor environments to speed-up the likelihood estimation step in a Monte Carlo localization scheme [16]. More precisely, the proposed Monte Carlo localization strategy works over a map of line segments corresponding to the main walls of the environment. The likelihood function that computes each particle weight is based on a Modified Hausdorff Distance [18]. This provides a score for the matching between line segments extracted from the map and current laser readings. Additionally, the number of particles of the Monte Carlo localization method is adapted, allowing the system to have a large number of particles during the global localization phase, where the position of the robot is unknown, and a reduced set of particles during the state tracking phase, where the position of the robot is limited to a small area. We use data from a real office environment to compare the performances of this method and a classic localization scheme based on a cell to cell comparison using an occupancy grid representation [72]. Our results indicate that the proposed method converges faster from global to state tracking localization. Furthermore, the proposed method uses less particles that in average have higher likelihood values. A summary of the comparison results is shown in Table 1.

In further research related to robot navigation, in [31] we analyze and compare popular path planning methods for omni-directional robots in a dynamic environment with the presence of multiple agents. Also, in [56] we analyze and compare different
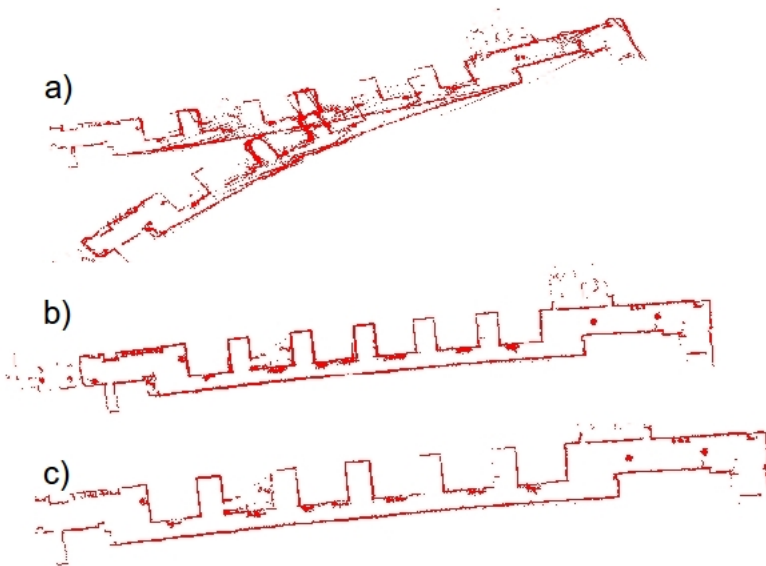
**Fig. 2** a) Map obtained from raw odometry and laser data. b-c) Samples from the posterior distribution.

Table 1 Comparison of the proposed robot localization method with respect to a point to point traditional approach.

	Number of particles. (same speed)	Iterations to converge.	Average likelihood (best sample).
Traditional MC localization	2700	14	0.91
Proposed method	5000	9	0.94

estimation and prediction techniques applied to the problem of tracking of multiple robots.

2.2 Topological maps using visual perception

The previous methods are based on 2D metric representations of the environment. This type of approaches has shown a high degree of success when operating in real time in natural environments [72] [62], however, they suffer from some limitations. For example, the usual structural symmetries of indoor building produce data association problems that are hard to solve with the 2D view of a laser range finder. Furthermore, problems such as modifications of the environment due to changes in the position of furniture, uncertainties due to the state of doors, or partial occlusions due to people walking around, also diminish the robustness of solutions based on 2D laser range finders. As a complementary solution to SLAM methods based on metric maps, we also develop topological representations of the environment using visual perception. In particular, in [23] and [39], we present a method for unsupervised selection and posterior recognition
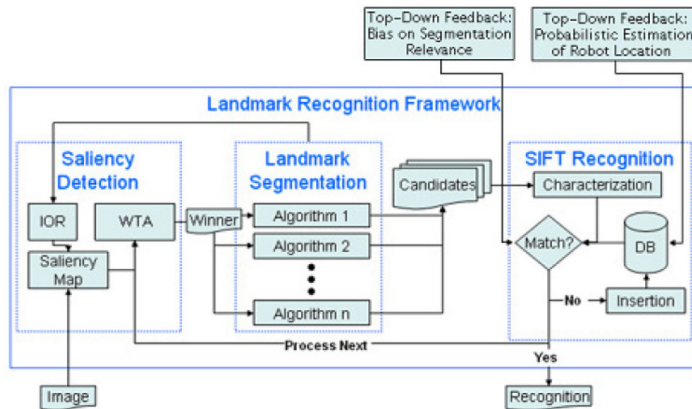
Fig. 3 Schematic operation of our approach for unsupervised selection and recognition of relevant visual landmarks. This approach integrates attention, segmentation, and landmark characterization mechanisms, in addition to top-down feedback mechanisms.

of visual landmarks in images acquired by a mobile robot. Figure 3 shows an overall view of this approach.

Our approach is based on a mixture of bottom-up visual features based on color, intensity, and depth cues, with top-down feedback given by spatial relations and memories of the most successful predicting features of previously recognized landmarks. Our overall goal is to select interesting, meaningful, and useful landmarks that can be used by a mobile robot to achieve indoor autonomous navigation.

The bottom-up approach for the selection of candidate landmarks is based on the integration of three main mechanisms: visual attention, area segmentation, and landmark characterization. Visual attention provides the ability to focus the processing resources on the most salient parts of the input image. This eliminates the detection of irrelevant landmarks and significantly reduces the computational cost. Area segmentation provides the ability to delimit the spanning area of each potential landmark. As there is no segmentation method that works properly in every situation, we adaptively integrate multiple segmentation algorithms in order to increase the robustness of the approach. The idea is to learn, for each landmark, the segmentation that provides the best recognition. Finally, landmark characterization provides a fingerprint for each landmark given by a set of specific features. These features allow the system to recognize and distinguish each landmark in subsequent images.

We complement the bottom-up landmark selection approach introducing two modalities of top-down feedback. First, to increase efficiency, we use an estimation of the robot position to reduce the searching scope for a potential match with a previously selected landmark. The estimation of the robot position is based on a Sequential Monte Carlo Localization method [64] using a metric map representation augmented with topological information. Second, to increase robustness, we keep a record of the previous successful recognitions of each landmark. We use this information to bias the influence that each bottom-up segmentation plays in the recognition. We achieve this goal by adaptively updating a set of weights that control the relevance of each segmentation in the recognition of each landmark.

Figure 4 shows images that highlight the typical landmarks detected in the sequences. Rectangles with identification labels are superimposed around new landmarks
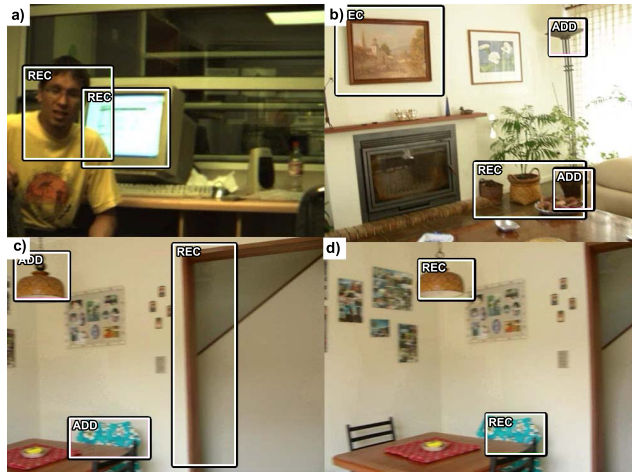
Fig. 4 a-b) Images from the testing sets. Examples of typical landmarks detected by the system are highlighted. c) Two selected landmarks are added (ADD) to a database. d) The two selected landmarks are recognized (REC) in a posterior frame taken from a different point of view.

(ADD) and recognized landmarks (REC). As an example, Figure 4c) shows two landmarks that are aggregated to the database, while Figure 4d) shows the recognition of these landmarks in a posterior frame captured from a different position.

We perform a more intensive evaluation using 10 video sequences captured by a stereo vision system in an indoor office environment (see Figure 4a). In this case, the images are automatically obtained by a mobile robot navigating inside an office building. Each sequence corresponds to the trip of the robot crossing the long corridor displayed in the map of Figure 2. The images of this dataset do not feature relevant illumination changes but they include moving objects. Given that our final goal is to achieve autonomous robot navigation, recall is measured as the number of times that a selected landmark is correctly recognized in posterior times that the robot visits a given place of the environment. To run this experiment, we use the 10 sequences of the office dataset. We obtain an average recall performance of 5.8.

2.3 Active localization using visual perception

Most state-of-the-art techniques for visual perception are mainly based on passively extracting knowledge from sets of images. In contrast, a mobile robot can actively explore the visual world looking for relevant views and features that can facilitate the recognition of relevant objects, scenes, or situations. This type of active visual perception scheme is highly pervasive in the biological world where, from human to bees, perceptual actions drive visual inference [11][6].

In [12], we use concepts from information theory and planning under uncertainty to develop a sound strategy to actively control the robot sensors. As a testbed, we use a differential drive wheeled robot provided with an odometer on each wheel and a color camera. This camera has a limited field of view and is mounted on a pan-tilt mechanism. By using the odometer, the robot is able to track its motions. By using the

video camera and suitable computer vision algorithms, the robot is able to distinguish a set of visual landmarks that represent the map of the environment. We assume that the map of the environment is known in advance.

The basic idea of our localization approach resides on actively using the pan-tilt mechanism to direct the visual sensor to areas of the environment where the robot expects to obtain the most useful information to achieve an accurate localization. We accomplish this goal by adding a new step between the prediction and observation steps of the Monte Carlo localization. In this new step, the robot uses the prediction of its current position and the map of the environment to assign a score to each possible perceptual action. This score considers the expected utility and the cost of executing each action.

We test the approach using a Pioneer-3 AT robot, equipped with a Directed Perception PTU-C46 pan-tilt unit and a Point Grey Dragonfly2 camera. We facilitate the landmark detection task by using a set of artificial visual landmarks. These landmarks consist of cylinders covered by two small patches of distinctive and easily detectable colors. Figure 5 shows pictures displaying the robot, the testing environment, and some of the artificial visual landmarks used in our tests. We test two strategies: i) A fixed camera always oriented according to the robot direction of motion, and ii) The proposed strategy to select perceptual actions. Figure 5-c shows the results of one of our tests. In the figure the robot moves from left to right. We observe that the active perceptual scheme achieves a better position estimation. This is explained by the fact that the robot is able to sense landmarks that are not available to the fixed camera case. Towards the end of the robot trajectory, the landmark sensing algorithm misses the two last landmarks closer to the robot path, and as a result, both approaches decrease the accuracy of the estimation. Figure 5-d shows that the active sensing scheme also provides an estimation with less variance than the fixed camera case.

3 Semantic Scene Understanding

Clearly, a robot with autonomous navigation capabilites is a valuable technology, however, to perform useful tasks in an indoor environments, such as fetching objects or delivering messages, one also needs to provide the robot with a higher semantic understanding of the environment. As an example, a robot operating in an indoor environment must recognize when a busy space does not correspond to an obstacle that needs to be avoided, but it corresponds to a door or an elevator, which in turn might have a handle or a push button that can be manipulated to successfully complete a task. Current efforts under these lines are still limited to highly specific situations, therefore, there is a need for more general solutions.

Our research agenda in the area of high level semantic understanding has been closely related to recent advances in the area of computer vision [74] [42] [25] [26]. These advances have been based on the successful application of machine learning techniques to the computer vision area. Following this trend, we have developed new techniques for people detection and tracking using a mobile platform [48][13][38], people recognition [45] and, recently, scene recognition by a mobile robot [22][21]. Next we provide further details of our application on scene recognition.
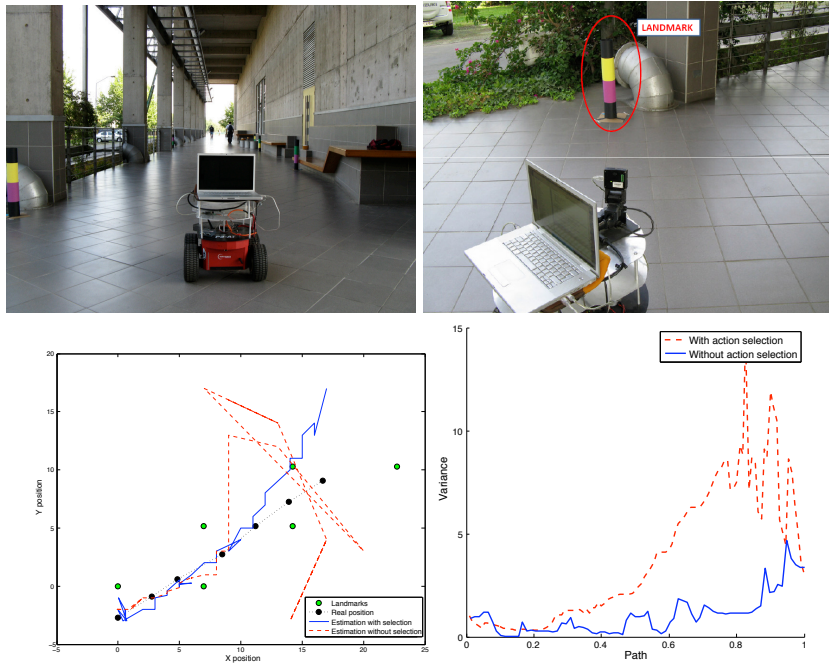
Fig. 5 a-b) Robot, testing environment, and some of the artificial visual landmarks. c) Estimation of robot position. d) Variance of estimation of robot position.

3.1 Indoor scene recognition

Current approaches for scene recognition present a significant drop in performance for the case of indoor scenes [61]. This can be explained by the high appearance variability of indoor environments. To face this limitation, in [22] we propose a method that includes high-level semantic information in the recognition process. Specifically, we propose to use common indoor objects as a key intermediate semantic representation. We frame our method as a generative probabilistic hierarchical model, where we use object category classifiers to associate low-level visual features to objects, and contextual relations to associate objects to scenes. As an important advantage, the inherent semantic interpretation of common objects facilitates the acquisition of training data from public websites [36] and also provides a straighforward interpretation of results, for example, by identifying failure cases where some relevant objects are not detected. This constrasts to current state-of-the-art methods based on generative unsupervised latent models, such as PLSA and DLA [25][7], where the intermediate representation does not provide a direct interpretation.

We base our category-level object detectors on Adaboost classifiers operating on visual features based on integral histograms using grayscale [17], fine grained saliency [48], Gabor [73], Local Binary Patterns (LBPs) [55], and HOG visual channels [15]. Additionally, we enhance our pure visual based classifiers using geometrical information obtained from a 3D range sensor on-board our robot. This facilitates the implementation of a focus of attention mechanism that allows us to reduce the computational load.

A comparison to several state-of-the-art methods using images acquired by mobile robots in two different office environments indicates that our approach achieves a significant increase in recognition performance. Specifically, when training the different models with similar amounts of generic data taken from the web and testing afterwards each model using images where relevant objects are present, our results indicate that the proposed method achieves an average recognition rate of 90%, while the accuracy of alternative methods ranges around 65%. In a second experiment, when we train the alternative models with images coming from the same office environment used for testing, we observe more competitive results. In this case, the best performing alternative model reaches an average accuracy of 82% that is closer but still lower than the 90% average accuracy of the proposed method trained with generic data. These results demonstrate suitable generalization capabilities of the proposed method and also support previous claims indicating that current state-of-the-art methods for scene recognition present low generalization capabilities for the case of indoor scenes [61].

Figure 6-a shows examples of scene recognition using the proposed method. In this case, the method is able to recognize a conference room through the correct detection of a projection screen and a wall clock. Figure 6-b shows a map of part of one of our office environments where we test the approach. This figure displays the trajectory followed by the robot during one of its runs and the labels assigned by the robot to the visited places (ground truth /estimation). The only situation when one place is not labeled correctly occurs during the second time the robot crosses the central hall, as no objects are detected in that part of the sequence. We believe that a more extensive implementation, based on a larger set of object detectors, can help to solve these situations.
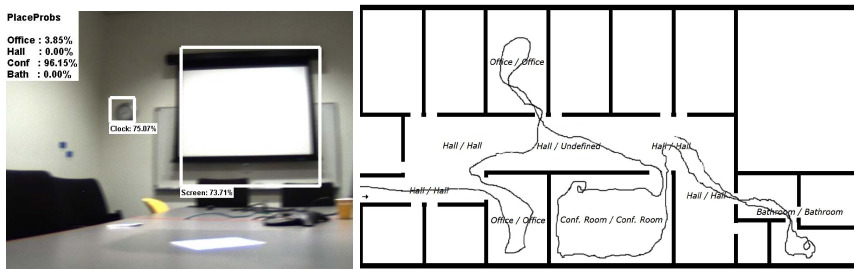
Fig. 6 a) Correct detection of a conference room. b) Map of an office environment indicating the robot trayectory and the label assigned to the different scenes (ground truth /estimation).

One inconvenient with the previous approach is that it needs to run all the object classifiers available at each inference step. This is highly inefficient because it is usually enough to detect a subset of the available objects to recognize a scene with enough confidence. To solve this limitation, in [21] we use concepts from information theory to propose an adaptive scheme to plan a suitable strategy for object search. The key idea is to execute only the most informative object classifiers. Under the proposed scheme, we use the current scene estimate to sequentially execute the classifier that maximizes the expected information gain. We select classifiers until it is possible to identify the current scene with enough confidence.

Figure 7 shows executions of our method in an office environment using a total of 7 object detectors. Figure 7-a shows a case without using adaptive object search,

(a) Without adaptive object search   (b) With adaptive object search
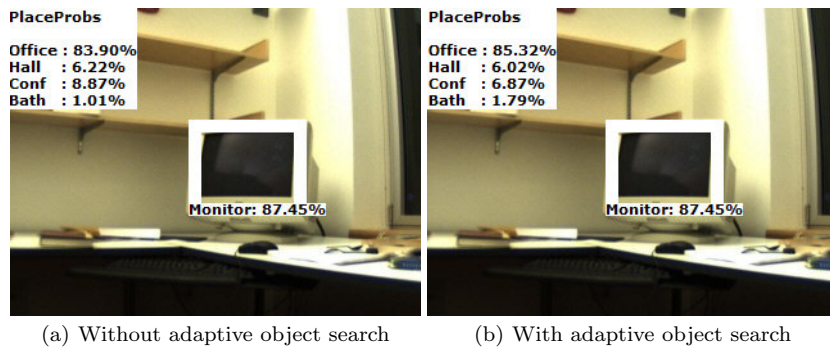
Fig. 7 Executions of the performance of our scene recognition approach in an office environment.

while Figure 7-b shows a case where we include adaptive object search. In the adaptive case, we add classifiers until the value of the respective information gain is not significant. We can see that, in both cases, detections are almost identical, and results differ slightly due to a sampling effect. The main difference between both executions is that in the first case all the available object detectors are executed, while in the second case the method runs only 5 object detectors: Screen, Urinal, Railing, Soap Dispenser, and Monitor. The reason of this behavior is that at the beginning of the inference process, when no objects are detected, the adaptive object search scheme chooses to run classifiers associated with objects that are highly discriminative with respect to a specific scene type, such as a Screen or an Urinal, because the eventual detection of those objects maximizes information gain. This is an expected result because we initially use a flat prior for the scene distribution and therefore the detection of informative objects produce peaked posteriors. By avoiding to run two object classifiers, the computational time for the object recognition task is speeded-up by a factor of ≈ 1.41. While this gain in performance might be considered marginal, the advantage of using an adaptive object search grows with the number of object classifiers available, where in a large scale case an efficient object search can avoid the execution of hundreds of object detectors. Also, it is important to notice that using adaptive object search also produces an overhead but, according to our experiments, this overhead is not significant.

4 Social Robots

A particular feature of a robot that operates in a human inhabited environment, such as an office building or a hospital, is that it needs to interact with people on a daily basis. In this case, it is highly useful, and in some cases a requirement, to provide the robot with suitable social behaviors. For example, in the case of boarding an elevator the robot must wait its turn.

Robots, such as Pearl [59] and Valerie [29], are examples of a new generation of social robots able to interact with people and to play the role of mediators to specific knowledge sources. Pearl is intended to assist elderly individuals with mild cognitive and physical impairments, as well as support nurses in their daily activities. Valerie operates as a robot receptionist for Newell-Simon Hall at Carnegie Mellon University, helping people by providing information about university members, campus directions,
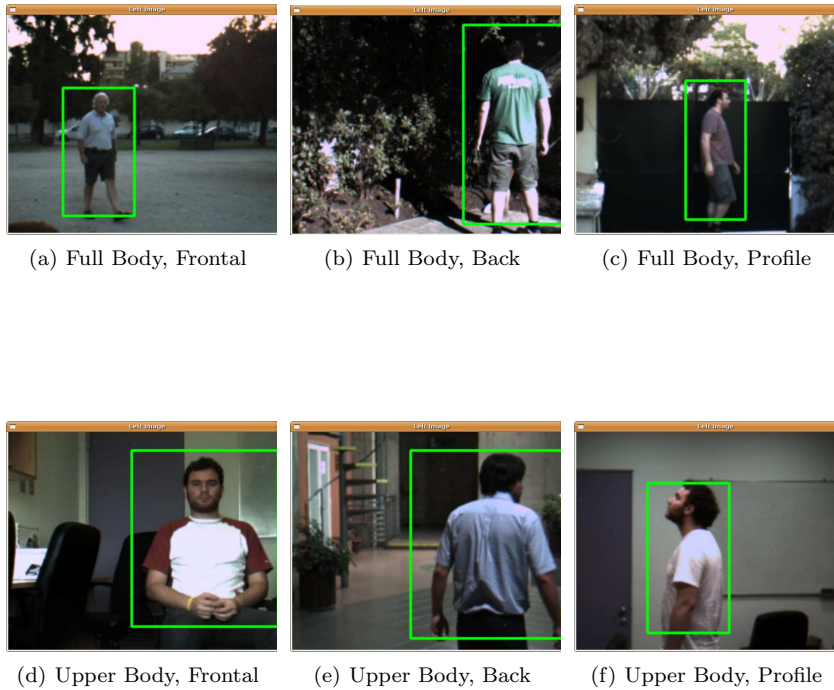
or retrieving data from the web. Although these robots perform specific tasks, they are designed to operate in highly general environments.

In particular, an intensive research activity on social robots has focused on museum robots that guide their human counterparts through the museum, explaining them the relevant aspects of the different expositions and halls. This is the case of robots such as Rhino [9], Minerva [70], Sage [53], and Joe Historybot [54]. As an example, Rhino was the first museum tour-guide robot, installed in mid-1997 at a museum in Germany. Rhino was responsible for greeting visitors and guiding them through a fixed set of museum attractions. As another example, Sage, later renamed as Chips, was a robot that operated in the Dinosaur Hall at Carnegie Museum of Natural History, USA, providing tours and presenting audiovisual information regarding bone collections. See [27] and [8] for reviews about social robots.

In the context of social robots, the detection of humans plays a key role. In [38] and [48] we target this problem using vision and range sensors mounted on a mobile platform. In [38], we present a method for human detection fusing information from differente visual cues. In [48], we extend the approach to the case of detecting people under several poses and partial views of the body. The main novelty of this approach is the feature extraction step, where we propose novel features derived from a visual saliency mechanism. In contrast to previous works, we do not use a pyramidal decomposition to run the saliency algorithm, but we implement the algorithm at the original image resolution using the so-called integral image. Examples of the different poses detected by this system can be seen in Figure 8.

Besides human detection, we also work on human tracking [13][65] and recognition [45][44]. For example, recently in [45] we propose a novel approach to face recognition using a generalization of the popular visual features known as Local Binary Patterns (LBPs) [55]. While many variations of LBP exist, so far none of them can automatically adapt to the training data [1] [67] [41] [77]. Our algorithm is based on the observation that the operation of a LBP over a given neighborhood is equivalent to the application of a fixed binary decision tree. Therefore, we use standard decision tree induction algorithms, in place of a fixed tree, to learn discriminative LBP-like descriptors from training data. As a major advantage, by using training data to learn the structure of the tree, we can effectively build an adaptive tree, whose main branches are specially tuned to encode discriminative patterns for the relevant target class. In particular, in the case of face recognition, our comparative results indicate that our approach outperforms several state-of-the-art solutions when it is tested on the FERET [58] and the CAS-PEAL-R1 [28] benchmark datasets [45][46].

In terms of robots interacting with people, in [47] we present an application of mobile robots in education. So far, most of the approaches that use robotics technologies in education aim to teach subjects closely related to the Robotics field [57][35][34]. Examples of these subjects are robot programming, robot construction, artificial intelligence, algorithm development, and mechatronics [75][76][4][34]. In our application, we use a robot to help students in the creation of abstract models of relevant concepts and properties of the real world by physically illustrating them. For example, a mobile robot can use its rotational capabilities to illustrate angular relations, or it can use its acceleration and velocity capabilities to illustrate relevant kinematic principles. Furthermore, by acting as a situated mediator of the educational activity, and using a collaborative and constructivist learning approach, the robot can guide the educational activity, playing a key role to increase the motivation and social bonds among the stu-

(a) Full Body, Frontal   (b) Full Body, Back   (c) Full Body, Profile

(d) Upper Body, Frontal   (e) Upper Body, Back   (f) Upper Body, Profile

Fig. 8 Figure a-f present results of our human detection system able to detect people under different poses. The system uses a new fast feature descriptor based on visual saliency and the integral image.

dents. As far as we know, this is the first time that robotics technologies are used in such educational settings, opening a new paradigm to apply robots in education.

Figure 9 shows the general setting of our educational framework. As shown, we extend the local capabilities of our mobile robot by providing it with remote wireless interfaces. These remote interfaces correspond to handheld devices distributed to each of the students. Using these devices, the robot can either, individually interact with each student by sending an exclusive message to the corresponding handheld device, or it can communicate with the group as a whole, by sending a shared message to all the students. In [47] we describe details of the implementation of this educational framework on two real cases used to teach concepts related to geometry and physics.

In a parallel research track, we recently developed a social robot that has the ability to express emotions. The interface of this robot is the result of a collaboration with people from the School of Art. The body of the robot corresponds to a sculpture (1.6 mts. tall) that exhibits an eye in its upper part. It has an eye-ball with a pan-tilt system, a video camera, and a 3D range sensor. By moving its eyelid and activating different sets of lights, the robot can express emotions, such as happiness, tiredness,
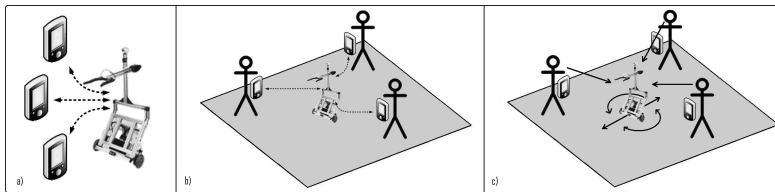
Fig. 9 General setting of the proposed educational framework. a) The robot is augmented with remote interfaces. b) The robot can communicate independently with each student by handheld devices. c) The robot can communicate visually with the group of students.

and so on. Figure 10-a shows the complete body of the robot, while Figure 10-b shows its main emotional interface. Our goal is to use this robot to autonomously guide the visit of groups of high school students who often visit our Department of Computer Science to learn about our facilities and main research activities. We believe that this is a challenging problem that will provide us with a suitable scenario to integrate and test several of the technologies developed in our group.

Fig. 10 a) Our social robot where its body correspond to a sculpture. b) The robot eye is its main emotional interface .

5 Conclusions

In this paper we present some of the main research activities that we have been developing since 2002 at our research group on Robotics and Machine Intelligence. Our research focus has been on the cognitive side of Robotics, mainly using machine learning and Bayesian statistical techniques. As a major achievement we have a working robot able to navigate in indoor buildings using visual and range sensors. Furthermore, our robot is able to perform high level perceptual behaviors such as scene recognition, and human detection, tracking, and recognition.

In terms of human and funding resources, during the last years there has been an increasing interest in Chile and our university about robotics technologies. To illustrate this we can consider that at PUC in the year 2002 the only course related to Machine Intelligence was the traditional AI deductive class. Today, our research group counts with more than 25 people among full-time professors, visiting researchers, and graduate students, and more than 15 courses in the area. Furthermore, during 2010 the first 3

graduate students obtained a doctoral degree as part of the group. We believe that the development of a critical mass of researchers and the need for a major association are some of the priorities for the development of Robotics in Latin America. In this sense, currently our group has collaborations with robotics groups at the Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) in Mexico, and the Robust Robotics Group at Massachusetts Institute of Technology (MIT) in USA.

As future work, we believe that currently high level scene understanding is one of the major challenges to deploy robust robots in natural indoor environments. Motivated by current successful combination of computer vision and machine learning techniques, we believe that the learning scheme is the right path to cope with the high ambiguity and complexity of the visual world. Furthermore, we foresee that the embodied and decision making nature of a mobile robot can be of great help by facilitating the development of active vision schemes, currently not deeply explored by main approaches to recognize relevant parts of the visual world. In particular, we believe that a synergistic ensemble of robust feature and region detectors from computer vision combined with suitable machine learning based classifiers and feature selectors, all guided by active vision modules, is a fruitful approach to improve the current semantic perceptual capabilities of mobile robots.

6 Acknowledgments

References

1. T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.
2. A. Araneda, S. Fienberg, and A. Soto. A statistical approach to simultaneous mapping and localization for mobile robots. *The Annals of Applied Statistics*, 1(1):66–84, 2007.
3. A. Araneda and A. Soto. Statistical inference in mapping and localization for mobile robots. In *Advances in Artificial Intelligence, Iberamia-04, LNAI 3315*, pages 545–554, 2004.
4. R. Avanzato. Mobile robot navigation contest for undergraduate design and k-12 outreach. In *Proc. of Conference of American Society for Engineering Education (ASEE)*, 2002.
5. T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM), part II: State of the art. *Robotics and Automation Magazine*, 13:108–117, 2006.
6. D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
7. A. Bosch, A. Zisserman, and X. Muñoz. Scene classification via PLSA. In *European Conference on Computer Vision (ECCV)*, 2006.
8. C. Breazeal. Towards sociable robots. *Robotics and Autonomous Systems*, 42(34), 2003.
9. W. Burgard, A. B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.
10. A. Clark and R. Grush. Towards a cognitive robotics. *Adaptive Behavior*, 7:5–16, 1999.
11. T. Collett. Landmark learning and guidance in insects. *Proc. of the Royal Society of London, series B*, pages 295–303, 1992.
12. J. Correa and A. Soto. Active visual perception for mobile robot localization. *Journal of Intelligent and Robotic Systems*, 58(3-4):339–354, 2010.
13. P. Cortez, D. Mery, and E. Sucar. Object tracking based on covariance descriptors and an on-line naive Bayes nearest neighbor classifier. In *Proc. of Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, pages 139–144, 2010.

14. M. Cummins and P. Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.
15. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of European Conference on Computer Vision (ECCV)*, 2005.
16. F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 1999.
17. P. Dollar, Z. Tu, P. Perona, and S. Belongie. Integral channel features. In *British Machine Vision Conference*, 2009.
18. F. Donoso-Aguirre, J.-P. Bustos-Salas, M. Torres-Torriti, and A. Guesalaga. Mobile robot localization using the Hausdorff distance. *Robotica*, 26(2):129–141, 2008.
19. H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping (SLAM), part I: The essential algorithms. *Robotics and Automation Magazine*, 13:99–110, 2006.
20. A. Elfes. *Occupancy Grids: A probabilistic framework for robot perception and navigation*. PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.
21. P. Espinace, T. Kollar, N. Roy, and A. Soto. Indoor scene recognition through object detection using adaptive objects search. In *European Conference on Computer Vision, Workshop on Vision for Cognitive Tasks*, 2010.
22. P. Espinace, T. Kollar, A. Soto, and N. Roy. Indoor scene recognition through object detection. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
23. P. Espinace, D. Langdon, and A. Soto. Unsupervised identification of useful visual landmarks using multiple segmentations and top-down feedback. *Robotics and Autonomous Systems*, 56(6):538–548, 2008.
24. P. Espinace, A. Soto, and M. Torres-Torriti. Real-time robot localization in indoor environments using structural information. In *IEEE Latin American Robotics Symposium (LARS)*, 2008.
25. L. Fei-Fei and P. Perona. A Bayesian hierarchical model for learning natural scene categories. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 524–531, 2005.
26. P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 2010.
27. T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166, 2003.
28. W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Transactions on System, Man, and Cybernetics (Part A)*, 38(1):149–161, 2008.
29. R. Gockley, A. Bruce, J. Forlizzi, M. P. Michalowski, A. Mundell, S. Rosenthal, B. P. Sellner, R. Simmons, K. Snipes, A. Schultz, and J. Wang. Designing robots for long-term social interaction. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2199–2204, 2005.
30. Machine Intelligence Group at Pontificia Universidad Catolica de Chile Grima. http://grima.ing.puc.cl.
31. F. Haro and M. Torres. A comparison of path planning algorithms for omni-directional robots in dynamic environments. In *IEEE Latin American Robotics Symposium, LARS*, 2006.
32. Mobile Robots Inc. *http://www.mobilerobots.com/*. 2008.
33. iRobot. *http://www.irobot.com/*. 2008.
34. M. Jansen, M. Oelinger, K. Hoeksema, and U. Hoppe. An interactive maze scenario with physical robots and other smart devices. In *Proc. of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE)*, 2004.
35. F. Klassner and S. Andreson. Lego mindstorms: Not just for k-12 anymore. *IEEE Robotics and Automation Magazine*, 10(2):12–18, 2003.
36. T. Kollar and N. Roy. Utilizing object-object and object-scene context when planning to find things. In *International Conference on Robotics and Automation (ICRA)*, 2009.
37. K. Konolige, J. Bowman, J.D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua. View-based maps. In *Proceedings of Robotics: Science and Systems*, 2009.
38. S. Pszczółkowski and A. Soto. Human detection in indoor environments using multiple visual cues and a mobile robot. In *Iberoamerican Congress on Pattern Recognition (CIARP), LNCS 4756*, pages 350–359, 2008.

39. D. Langdon, A. Soto, and D. Mery. Automatic selection and detection of visual landmarks using multiple segmentations. In *IEEE Pacific-Rim Symposium on Image and Video Technology (PSIVT), LNCS 4319*, pages 601–610, 2006.

40. J. Leonard and H. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *Proc. of IEEE International Workshop on Intelligent Robots and Systems*, volume 3, pages 1442–1447, 1991.

41. S. Liao and A. Chung. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In *Proc. of Asian Conference on Computer Vision (ACCV)*, 2007.

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

43. Robowatch Technologies Ltd. *http://www.robowatch.de.* 2008.

44. D. Maturana, D. Mery, and A. Soto. Face recognition with local binary patterns, spatial pyramid histograms and naive Bayes nearest neighbor classification. In *Proc. of International Conference of the Chilean Computer Science Society/IEEE CS Press*, 2009.

45. D. Maturana, D. Mery, and A. Soto. Face recognition with decision tree-based local binary patterns. In *Proc. of Asian Conference on Computer Vision (ACCV)*, 2010.

46. D. Maturana, D. Mery, and A. Soto. Learning discriminative local binary patterns for face recognition. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, 2011.

47. R. Mitnik, M. Nussbaum, and A. Soto. An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4):367–382, 2008.

48. S. Montabone and A. Soto. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing*, 28(3):391–402, 2010.

49. M. Montemerlo and S. Thrun. Simultaneous localization and mapping with unknown data association using FastSLAM. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2003.

50. Hans Moravec. The Stanford cart and the CMU rover. In I. J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 407–441. Springer-Verlag, 1990.

51. K. Murphy. Bayesian map learning in dynamic environments. In *Advances in Neural Information Processing Systems 12, NIPS*. 1999.

52. N. J. Nilsson. Shakey the robot. Technical Report 323, AI Center, SRI International, 1984.

53. I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. An affective mobile educator with a full-time job. *Artificial Intelligence*, 114(1-2):95–124, 1999.

54. I. Nourbakhsh, C. Kunz, and T. Willeke. The mobot museum robot installations: A five year experiment. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3636–3641, 2003.

55. T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

56. J. L. Peralta-Cabezas, M. Torres-Torriti, and M. Guarini. A comparison of Bayesian prediction techniques for mobile robot trajectory tracking. *Robotica*, 26(5):571–585, 2008.

57. M. Petre and B. Price. Using robotics to motivate back door learning. *Education and Information Technologies*, 9(2):147–158, 2004.

58. P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for Face-Recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

59. J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 42(3-4):271–281, 2003.

60. Probotics. *http://www.probotics.com/.* 2008.

61. A. Quattoni and A. Torralba. Recognizing indoor scenes. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

62. R. Siegwart and I. Nourbakhsh. *Introduction to Autonomous Mobile Robots*. MIT press, 2004.

63. R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I.J. Cox and G.T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 167–193. Springer-Verlag, 1990.

64. A. Soto. Self adaptive particle filter. In *Proc. of International Join Conference on Artificial Intelligence (IJCAI)*, pages 1398–1406, 2005.

65. A. Soto and P. Khosla. Probabilistic adaptive agent based system for dynamic state estimation using multiple visual cues. *Springer Tracts in Advanced Robotics (STAR)*, 6:559–572, 2003.

66. A. Soto, M. Saptharishi, J. Dolan, A. Trebi-Ollennu, and P. Khosla. CyberATVs: dynamic and distributed reconnaissance and surveillance using all terrain UGVs. In *Proc. of the International Conference on Field and Service Robotics*, pages 329–334, 1999.

67. X.Y. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Proc. of International Conference on Analysis and Modeling of Faces and Gestures*, pages 168–182, 2007.

68. S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99:21–71, 1998.

69. S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002.

70. S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F.Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Minerva: A second generation mobile tour-guide robot. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pages 1999–2005, 1999.

71. S. Thrun, W. Burgard, and D. Fox. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning and Autonomous Robots*, 31/5:1–25, 1998. Joint Issue.

72. S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

73. M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1–2):61–81, 2005.

74. P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

75. E. Wang and R. Wang. Using legos and robolab (LabVIEW) with elementary school children. In *Proc. of Conference on Frontiers in Education*, volume 1, pages T2E–T11, 2001.

76. J. Weinberg, G. Engel, K. Gu, C. Karacal, S. Smith, W. White, and X. Yu. A multidisciplinary model for using robotics in engineering education. In *Proc. of Conference of American Society for Engineering Education (ASEE)*, 2001.

77. L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Proc. of European Conference on Computer Vision (ECCV), Workshop on Real-Life Images*, 2008.