



UNIVERSIDAD
DE MÁLAGA

Doctoral Dissertation

Probabilistic Techniques in Semantic Mapping for Mobile Robotics


José Raúl Ruiz Sarmiento
2016

Tesis doctoral
Ingeniería Mecatrónica
Dpt. de Ingeniería de Sistemas y Automática
Universidad de Málaga



UNIVERSIDAD
DE MÁLAGA

AUTOR: José Raúl Ruiz Sarmiento

 <http://orcid.org/0000-0002-9929-5309>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es

UNIVERSIDAD DE MÁLAGA
DEPARTAMENTO DE
INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

El Dr. D. Javier González Jiménez y el Dr. D. Cipriano Galindo Andrades, directores de la tesis titulada "Probabilistic Techniques in Semantic Mapping for Mobile Robotics" realizada por D. José Raúl Ruiz Sarmiento, certifican su idoneidad para la obtención del título de Doctor en Ingeniería Mecatrónica.

Málaga, 3 de octubre de 2016



Dr. D. Javier González Jiménez



Dr. D. Cipriano Galindo Andrades

Dept. of System Engineering and Automation
University of Málaga
Studies in Mechatronics



Probabilistic Techniques in Semantic Mapping for Mobile Robotics

AUTHOR: José Raúl Ruiz Sarmiento

SUPERVISORS: Javier González Jiménez
Cipriano Galindo Andrades

Thesis defended on 25th November 2016

JURY:

Antonio Jesús Bandera Rubio (Málaga University, Spain)
Luís Filipe de Seabra Lopes (Aveiro University, Portugal)
José Luis Blanco Claraco (Almería University, Spain)

*To my family,
in heaven and on earth.*

*A mi familia,
en el cielo y en la tierra.*

Table of Contents

Table of Contents	i
Abstract	v
Acknowledgments	vii
Resumen de la Tesis	ix
I Thesis description	1
1 Introduction	3
1.1 Motivation	4
1.2 Contributions	6
1.2.1 Contributions to contextual scene understanding	6
1.2.2 Contributions to semantic mapping	7
1.2.3 Publications	7
1.3 Thesis framework	8
1.4 Thesis outline	11
2 Theoretical background	13
2.1 Probabilistic Graphical Models	13
2.1.1 The happiness example	14
2.1.2 Learning the models	16
2.1.3 Probabilistic inference	16
2.2 Knowledge bases	17



2.2.1	Ontologies	17
2.2.2	Happiness from an Ontological stance	18
3	Contextual scene understanding	21
3.1	Introduction	21
3.2	Related work	22
3.3	Testbed	25
3.4	Contributions	26
3.4.1	UMA-Offices dataset	26
3.4.2	The UPGMpp library	27
3.4.3	Testing CRF learning approaches	29
3.4.4	Exploiting Semantic Knowledge for CRF learning	30
3.4.5	Including rooms into the equation	33
3.4.6	Further enhancing CRFs performance: coherence and efficiency	35
3.4.7	Learning from experience	36
3.5	Discussion	37
4	Semantic Mapping	39
4.1	Introduction	39
4.2	Related work	40
4.3	Contributions	44
4.3.1	The Object Labeling Toolkit	44
4.3.2	Robot@Home dataset	46
4.3.3	Multiversal Semantic Maps	49
4.4	Discussion	53
5	Summary of included papers	55
5.1	Paper A: Learning CRFs with data from Semantic Knowledge	55
5.2	Paper B: Joint recognition of objects and rooms	56
5.3	Paper C: Exploiting Semantic Knowledge for a coherent and efficient recognition	56
5.4	Paper D: UPGMpp library for managing PGMs	57
5.5	Paper E: OLT toolkit for managing sequential RGB-D datasets	57
5.6	Paper F: Semantic Map representation handling uncertainty	58
6	Conclusions and future work	59
	Bibliography	65
II	Included papers	79
7	List of papers	81
A	Exploiting Semantic Knowledge for Robot Object Recognition	81
B	Joint Categorization of Objects and Rooms for Mobile Robots	82

C Scene Object Recognition for Mobile Robots Through Semantic Knowl-
edge and Probabilistic Graphical Models 82

D UPGMpp: a Software Library for Contextual Object Recognition . . . 83

E OLT: A Toolkit for Object Labeling applied to robotic RGB-D datasets 84

F Building Multiversal Semantic Maps for Mobile Robot Operation . . 84

Abstract

Semantic maps are world representations that permit a robot to understand not only the spatial aspects of its workspace, but also the meaning of the existing elements (objects, rooms, etc.) and how humans interact with them (*e.g.* functionalities, events, and relations). To achieve this, a semantic map enhances purely spatial representations, like geometric or topological maps, with meta-information concerning the types of elements and relations to be found in the working environment. This meta-information, called *semantic* or *common-sense* knowledge, is typically codified into *Knowledge Bases* (KBs).

An example of a piece of semantic knowledge stored in a KB could be: “refrigerators are big, box-shaped objects normally located in kitchens, which contain pill boxes and perishable food”. Encoding and managing this semantic knowledge enables the robot to reason about the information gathered from a given workspace, as well as to infer new one in order to efficiently accomplish high-level tasks like “hey robot! take the pills to grandma, please”.

This thesis contributes the usage of probabilistic techniques to build and maintain semantic maps, providing three main advantages in comparison with traditional approaches:

- i) to handle uncertainty (coming from inaccurate robot sensors and models),
- ii) to provide coherent environment interpretations by exploiting contextual relations among the observed elements (*e.g.* fridges are usually in kitchens) in a holistic fashion, and
- iii) to yield certainty values that reflect the correctness in the robot understanding of its surroundings.

Specifically, the included contributions can be grouped into two major topics. The first set of contributions focuses on the *scene object and/or room recognition*

problems, since semantic mapping systems must reckon on reliable recognition algorithms for building proper representations. For that, we explore the utilization of *Probabilistic Graphical Models* (PGMs) for exploiting contextual relations among objects and/or rooms dealing with uncertainty, and the utilization of KBs to enhance their performance in different ways, e.g. detecting incoherent results, providing prior information, reducing the complexity of the probabilistic inference, generating synthetic training samples, enabling the learning from experience, etc.

The second group of contributions accommodates the probabilistic outcome of the developed recognition algorithms into a novel semantic map representation, coined *Multiversal Semantic Map* (*MvSmap*). This map manages multiple interpretations of the robot workspace, called *universes*, which are annotated with the probability of being the true ones according to the current knowledge of the robot. Thus, this approach gives a grounded belief about the understanding of the environment, which enables a more coherent and efficient robotic operation.

The proposed probabilistic algorithms have been thoroughly tested against other cutting-edge approaches employing state-of-the-art datasets. Additionally, this thesis also contributes: two datasets, *UMA-Offices* and *Robot@Home*, containing diverse ground truth information and sensory data from different types of devices covering office and home environments, and two software tools, the *Undirected Probabilistic Graphical Models in C++* (UPGMpp) library, and the *Object Labeling Toolkit* (OLT), for working with PGMs and processing datasets respectively.

Acknowledgments

Luckily, it is large the list of people who have been around and helped me, in one way or another, to reach the peak of this sharp mountain called *doctorate*. They were there in both good and not that good moments, and all of them deserve a warm mention. Nevertheless, the space for showing my gratitude is limited, so I will do my best!.

Foremost, I would like to express my special appreciation and thanks to my supervisors Prof. Dr. Javier González Jiménez and Dr. Cipriano Galindo Andrades. Our brainstorming meetings, source of a bunch of ideas, and their constant support, indications, and positive thoughts are responsible to a great extent of this work. I have to say that I consider them the parents of my academic career. Javier is an example of tireless effort, patience, and talent. He is an absolute passionate about his work, and successfully leads the MAPIR group, at the university of Málaga, which is in constant growing. Cipriano is the inspiration personified, always with affectionate words and acts, and fresh ideas to climb right to the top of the mountain. Thank you for believing in me.

To be part of the MAPIR group is an experience itself. All of us, as a team, celebrate the victories and regret the frustrations of others. It is difficult to imagine a more humane, and at the same time skilled group of people, within and outside the workspace. Starting with the *B team*, I have to mention Mariano Tarifa, Francisco Meléndez, Javier G. Monroy, Rubén Gómez, Manuel López, Carlos Sánchez, Jesús Briaes, Andrés Góngora and Ángel Martínez, and the former members Eduardo Fernández, Ana Gago, Emil Khatib, Miguel Algaba and Gregorio Navidad. Thank you guys for being that amazing. I would also like to thank the senior researches at MAPIR, whose words helped me to stay motivated and focused. So Juan A. Fernández, Ana Cruz, Vicente Arévalo, and Francisco Moreno, thank you for that. I cannot forget Jose L. Blanco, a former member of the group, totally in love with research and with a shared passion for music, who gave me practical indications.

During my PhD years I have been in several conferences and schools. In there, I have met great people that were a plus within the study-develop-publish cycle. I also completed a stay at the University of Osnabrück, under the supervision of Prof. Dr. Joachim Hertzberg, a brilliant and close person, where I shared office with my colleague and friend Martin Günther, and I met nice people like Sebastian Stock, Jochen Sprickerhof, Sven Albrecht, Thomas Wiemann, Kai Lingemann, and Astrid Heinze. Thank you all for that enriching adventure, unfortunately my level of German is still low, I promise to improve it!. A special thanks to Bárbara Rotstein, my Spanish girl in Osnabrück, my stay there would had been quite different without her.

Friends have also played a pivotal role during the development of this thesis, specially Cristian F. Segura, Ismael Gutiérrez, José D. Pérez, José D. Sarmiento (*compadre*), Jesús Ramírez, Francisco Jiménez, Francisco A. Nieto, and Laura R., as well as their respective and lovely partners. They forgave my absence from many meetings, and illuminated me with their brilliant careers and growth as people. You fellows are awesome.

My relatives have been like parts of my body, I could not conceive this period of time without them. My mother María Sarmiento, was my heart, and my father José Ruiz, was my mind. My strong brother Juan L. Ruiz, his pretty wife Mónica Gallardo, and my lovely nephew Juan A. Ruiz were my skeleton. My uncles M. Josefa Sarmiento, Inmaculada Sarmiento, Toñi Sarmiento, and Antonio Díaz, and my cousin Samuel D. Díaz were my muscles. I just put the soul, and we all together achieved this goal. I do not forget my grandparents, specially José Sarmiento. I am sure that, wherever you are, you are reading these lines. It does not matter that you did not speak English, the language of love is universal. Heartfelt thanks.

Last but not the least, I would like to thank my *neni*, Rocío, and her family, now also mine, for sharing with me the last two years of this project. You enjoyed my victories like yours, and spent weekends with me at home working in front of a computer screen, patiently waiting for having some leisure time. The effort has now its rewards, and I promise to return all your support and love back, but multiplied by two.

José Raúl Ruiz Sarmiento
Málaga
September 2016

This thesis was partially supported by the Spanish grant program FPU-MICINN 2010, and by the research projects *PROMOVE: Advances in mobile robotics for promoting independent life of elders* (DPI2014-55826-R) and *IRO: Improvement of the sensorial and autonomous capability of Robots through Olfaction* (2012-TEP-530), funded by the the *Spanish Government* and the *Andalucía Regional Government*, respectively.

Introducción

El invierno se acerca. Un robot sirviente detecta que la temperatura está disminuyendo y decide llevarle una manta a una adorable abuela. En el mismo edificio, otro robot encargado de patrullar una planta de oficinas se alerta al detectar una luz encendida en una habitación; rápidamente se percata de que es el compañero del área de investigación, Bob, trabajando hasta tarde por tercera noche en esta semana. Mientras tanto, su hija Alice está triste por la ausencia de sus padres, y su colega robótico, apodado cariñosamente *Roboto*, busca su oso de peluche favorito. Sophie, la madre de Alice, también está contando las horas para verla, y ordena a un robot limpiar las mesas una vez que su restaurante ha cerrado al público.

Estos escenarios son ejemplos donde los robots móviles de hoy en día, en mayor o menor medida, pueden proveer una serie de servicios para mejorar el nivel de vida de la sociedad. Cada vez se vislumbra más claramente que los robots están llegando para quedarse, como se ve en su exitosa aplicación a diversas tareas como vigilancia, cuidado de la salud, compañía, entretenimiento, mantenimiento del hogar, etcétera [97], donde colaboran con humanos o los reemplazan en tediosos o peligrosos quehaceres. Algo común a todas las aplicaciones anteriores es la necesidad de construir representaciones del entorno de trabajo, comúnmente llamadas *mapas*, las cuales permiten a un robot móvil alcanzar un cierto grado de consciencia respecto a sus alrededores para poder, por ejemplo, navegar evitando obstáculos, localizarse a sí mismo con respecto a un sistema de referencia dado, almacenar información relevante sobre los elementos a su alrededor, etc.

Las representaciones tradicionales del entorno de trabajo del robot, como es el caso de mapas geométricos [23, 140], topológicos [110, 109], o híbridos [139, 13], aún son intensivamente usadas gracias a las habilidades básicas con las que dotan

al robot (navegación y localización). A pesar de ello, la ejecución de tareas de alto nivel como las mencionadas en los escenarios anteriores requiere representaciones más sofisticadas, cercanas al modo en el que los humanos interpretan su entorno. Los mapas semánticos (*semantic maps* en inglés) aparecieron para cubrir esta necesidad, permitiendo a un robot no sólo *comprender* los aspectos espaciales de su entorno, sino además el significado de sus elementos (objetos y habitaciones) y cómo los humanos interactúan con ellos, por ejemplo funcionalidades, eventos, y relaciones. Para ello se considera meta-información, comúnmente conocida como *Conocimiento Semántico* (*Semantic Knowledge* o *SK* en inglés¹), sobre los tipos de elementos que se pueden encontrar en el área de trabajo del robot, incluyendo sus relaciones. Esbozos de dicha información, típicamente codificada en una *base de conocimiento* (*Knowledge Base* o *KB* en inglés), pueden ser: las mantas se encuentran habitualmente almacenadas en armarios; las luces de la oficina deben estar apagadas tras la jornada laboral; los osos de peluche mejoran el estado de ánimo; la vajilla frágil debe lavarse en el lavavajillas.

Motivación

Típicamente, los mapas semánticos son poblados² con información exacta, por ejemplo un objeto es una manta o no lo es. Esto se debe a la incapacidad de las representaciones semánticas tradicionales para tratar con resultados inciertos, lo que fuerza la utilización de algoritmos de reconocimiento que provean información exacta, habitualmente mediante la aplicación de umbrales a resultados probabilísticos. Por ejemplo, un algoritmo de reconocimiento³ indicando que un objeto puede ser una manta con una probabilidad de 0.52, y una alfombra con 0.48, podría proveer un único resultado considerando el objeto como una manta y desechando la otra hipótesis, aunque esta es también altamente probable. Este enfoque exacto claramente compromete la operación del robot: la incertidumbre, proveniente de fuentes como el propio sistema de percepción del robot o los modelos empleados para tratar el problema, se ignora al almacenar los resultados de reconocimiento en el mapa semántico. De este modo, aunque los resultados del ejemplo claramente muestran que el reconocimiento es ambiguo, nuestra querida abuela podría terminar con una áspera alfombra encima suya. Este es un escenario de entre los muchos posibles que ponen de manifiesto la necesidad de utilizar técnicas capaces de proveer mediciones de incertidumbre sobre sus resultados para poblar y mantener mapas semánticos – para lo cual la literatura recurre comúnmente a técnicas probabilísticas [141, 65] –, así como de adaptar las representaciones semánticas actuales para poder manejar información incierta. Esto resultaría en una operación más coherente y eficiente por parte del robot móvil.

¹Cuando sea posible, a lo largo de este resumen se utilizarán los acrónimos en inglés de las herramientas utilizadas, por ser su uso más común en la comunidad científica.

²*Poblar* un mapa semántico se refiere al proceso de introducción de los elementos espaciales en el entorno del robot en dicho mapa, comúnmente objetos y habitaciones, percibidos mediante su sistema sensorial.

³Para simplificar la explicación se considera que existen sólo dos tipos de objetos, mantas y alfombras.

Tratando de evidenciar aún más la conveniencia de trabajar con información incierta, supongamos un escenario donde a un robot sirviente, recién aterrizado en su nueva casa desde el laboratorio, se le encomienda el traer las zapatillas a la abuela adorable. En ausencia de información espacial, el robot puede inferir (de acuerdo con la información cargada en su *KB*) que la localización más probable de las zapatillas es un dormitorio. Durante el mapeo inicial de la casa por parte del robot, este reconoció un dormitorio correspondiente a la habitación más lejana con respecto a la posición actual de la abuela con una probabilidad de 0.45, y 0.43 de ser una cocina⁴. Otra habitación cercana a la posición del robot ha sido reconocida como cocina con una probabilidad de 0.48, y como dormitorio con 0.47. La utilización de la interpretación más probable, el *modus operandi* usual cuando se trabaja con mapas semánticos tradicionales, daría lugar a la exploración de la habitación más lejana, con un 45% de probabilidades de ser el lugar correcto, mientras que el considerar ambas interpretaciones produciría un plan más lógico: echar primero un vistazo a la habitación más cercana.

Aunque existen numerosos algoritmos para el reconocimiento de objetos y/o habitaciones que proveen mediciones de incertidumbre sobre sus resultados, estos usualmente trabajan mediante el procesamiento individual de cada elemento espacial de acuerdo con sus características geométricas (forma, tamaño, orientación, etc.) o de apariencia (color, textura, brillo, etc.). En otras palabras, si el tipo más probable para un objeto es *manta*, este es considerado una manta sin tener en cuenta que otros objetos hay a su alrededor ni su localización. Este enfoque ignora la rica información contextual presente en los entornos humanos: la distribución de las habitaciones sigue un cierto orden, y los objetos no están colocados aleatoriamente, sino siguiendo una cierta configuración acorde a su funcionalidad (por ejemplo, un mando a distancia suele estar en el entorno de una televisión, un pasillo conecta habitaciones, o una bañera suele encontrarse en el cuarto de baño) [113, 73, 117]. El modelado y aprovechamiento de esta información contextual puede ser útil, por ejemplo, para clarificar resultados inciertos: siguiendo con el ejemplo anterior, si el objeto se encuentra en un armario, este pertenecerá más probablemente al tipo *manta* que al tipo *alfombra*, el cual se encuentra usualmente sobre el suelo. Este tipo de información puede codificarse de manera natural en las bases de conocimiento, no obstante, su explotación para el reconocimiento contextual de objetos/habitaciones manejando incertidumbre no es simple.

Los *Modelos Gráficos Probabilísticos* (*Probabilistic Graphical Models* o *PGMs* en inglés) [65] son una herramienta ampliamente utilizada para el modelado y la explotación de relaciones de contexto tratando con incertidumbre. Estos modelos trabajan con una representación en forma de grafo, donde los nodos representan variables aleatorias y los arcos conectan variables que tienen algún tipo de relación. Por ejemplo, en el caso del reconocimiento de objetos, cada objeto en la escena es representado como una variable aleatoria que toma valores de entre los tipos de objetos posibles

⁴Nótese que la suma de ambas probabilidades es de 0.88. El resto, hasta sumar 1, se corresponde con las probabilidades de pertenecer a otro tipo de habitación, e.g. pasillo, cuarto de baño, salón, etc.

(mesa, sofá, libro, etc.), mientras que los arcos conectan variables cuyos objetos asociados están situados cerca en la escena. Esta representación soporta la ejecución de algoritmos de inferencia probabilística, los cuales son capaces de proveer los resultados de reconocimiento deseados, junto con mediciones de incertidumbre sobre dichos resultados. Los *PGMs* han sido aplicados con éxito a tareas como eliminación de ruido en imágenes, procesamiento de lenguaje natural, reconocimiento de la actividad en una escena, predicción meteorológica, etc. A pesar de ello, estos modelos muestran una serie de limitaciones que deben ser tratadas antes de ser utilizados para poblar mapas semánticos, a saber: son computacionalmente intratables cuando la complejidad del problema a modelar incrementa (en este caso, cuando el número de objetos/habitaciones en el entorno y sus posibles tipos crece), necesitan una considerable cantidad de datos de entrenamiento para ajustar modelos exitosos, y son incapaces de detectar resultados incoherentes así como de aprender de experiencias pasadas.

Contribuciones

Las contribuciones de la presente tesis tratan de solucionar las limitaciones de los mapas semánticos tradicionales anteriormente comentadas mediante el uso de técnicas probabilísticas. Concretamente, los objetivos de la tesis, que tuvieron como fruto el desarrollo de dichas técnicas, fueron definidos como:

- **Desarrollo de un sistema de reconocimiento completo:** Proveer algoritmos probabilísticos para el reconocimiento de objetos y/o habitaciones manejando información tanto de contexto como incierta, en los cuales también se considere conocimiento semántico, con el objetivo de presentar una serie de características deseables como escalabilidad, detección de resultados erróneos, aprendizaje de experiencias pasadas, etc.
- **Mejora de los mapas semánticos para el manejo incertidumbre:** Acomodar los resultados probabilísticos de dichos algoritmos en una novedosa representación de mapas semánticos, de tal modo que un robot pueda explotarlos para conseguir una noción de la certeza del mismo sobre su comprensión del entorno de trabajo, permitiéndole operar de un modo más coherente.

De este modo, las contribuciones de esta tesis pueden agruparse en dos temas principales: comprensión contextual de la escena, y mapeo semántico de la misma.

Contribuciones a la comprensión contextual de la escena

El primer grupo de contribuciones, presentadas en los artículos [114, 116, 119, 117, 115, 122, 121], se centra en el problema del reconocimiento de objetos y/o habitaciones empleando información contextual. Los *PGMs* en general, y los Campos Aleatorios Condicionales (*Conditional Random Fields* o *CRFs* en inglés) en particular, son usados para modelar este problema desde un punto de vista holístico, considerando las

relaciones de contexto entre objetos y/o habitaciones, y tratando de manera formal la incertidumbre inherente al proceso de reconocimiento. Su aplicabilidad al problema tratado ha sido verificada tras una exhaustiva evaluación de los algoritmos más populares tanto de entrenamiento como de inferencia probabilística sobre dichos modelos.

Estos *CRFs* trabajan en conjunción con *KBs*, lo que permite mantener sus ventajas cuando trabajan por separado a la vez que se mitigan sus limitaciones:

- Las *KBs* dotan a los *CRFs* con capacidades para: reducir su complejidad, explotar información a priori sobre el dominio del problema, verbalizar sus resultados, generar un número aleatorio de ejemplos de entrenamiento sintéticos para su ajuste, detectar resultados incoherentes, y aprender de la experiencia del robot.
- Los *CRFs* permiten a las *KBs* manejar información incierta y explotar relaciones de contexto de acuerdo con una base teórica fundamentada.

Los resultados devueltos durante la evaluación de los métodos desarrollados han sido comparados con los de otras soluciones punteras empleando conjuntos de datos del estado del arte. Además, se ha reunido y hecho público un nuevo repositorio de datos, llamado *UMA-Offices*, consistente en observaciones tridimensionales de 25 habitaciones de nuestro entorno de oficinas. También se ha implementado la librería software de código abierto *Undirected Probabilistic Graphical Models in C++*⁵ (UPGMpp) con el fin de manejar eficientemente los PGMs.

Contribuciones al mapeo semántico

El objetivo del segundo grupo de contribuciones, presentadas en los artículos [120, 118, 123], es el de acomodar los resultados probabilísticos provenientes de las técnicas anteriores en una representación semántica del entorno. Para ello se ha desarrollado la representación *Multiversal Semantic Map (MvSmap)*, la cual permite considerar diferentes interpretaciones del entorno de trabajo del robot en forma de *universos*, también almacenando información sobre la probabilidad de que sean las interpretaciones correctas. Esto permite al robot tener en cuenta no sólo el universo más probable, sino otros que también muestran una alta probabilidad de ser válidos. Este novedoso mapa se acompaña de técnicas para mantener tratable el número de universos considerados, de tal manera que sea aplicable a entornos complejos con numerosos objetos y habitaciones.

La idoneidad de los *MvSmaps*, así como su capacidad para manejar datos inciertos de una manera eficiente, se ha comprobado empleando el novedoso conjunto de datos *Robot@Home*, acumulado por un robot móvil al explorar una serie de entornos domésticos. Además, el conjunto de herramientas *Object Labeling Toolkit*⁶ (OLT), disponible públicamente para la comunidad investigadora, ha sido desarrollado para

⁵<http://mapir.isa.uma.es/work/upgmpp-library>

⁶<http://mapir.isa.uma.es/work/object-labeling-toolkit>

procesar de manera fácil y rápida conjuntos de datos formados por secuencias de información sensorial, como es el caso de *Robot@Home*.

Publicaciones

La presente tesis ha dado lugar a las siguientes publicaciones:

Revistas

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Build-ing Multiversal Semantic Maps for Mobile Robot Operation.* Enviado a Knowledge-Based Systems (2016).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. A Survey on Learning Approaches for Undirected Graphical Models. Application to Scene Object Recognition.* En International Journal of Approximate Reasoning, (aceptado, por aparecer) (2016).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Robot@Home, a Robotic Dataset for Semantic Mapping of Home Environments.* Enviado a International Journal of Robotics Research (2016).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Scene Object Recognition for Mobile Robots Through Semantic Knowledge and Probabilistic Graphical Models.* En Expert Systems with Applications, vol. 42, no. 22, pp. 8805–8816, (2015).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Exploiting Semantic Knowledge for Robot Object Recognition.* En Knowledge-Based Systems, vol. 86, pp. 131–142, (2015).

Conferencias

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Probability and Common-Sense: Tandem Towards Robust Robotic Object Recognition in Ambient Assisted Living.* En 10th International Conference on Ubiquitous Computing & Ambient Intelligence, Las Palmas de Gran Canaria, Spain, (2016).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Joint Categorization of Objects and Rooms for Mobile Robots.* En IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, (2015).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets.* En European Conference on Mobile Robots (ECMR), Lincoln, UK, (2015).

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. UPGMpp: a Software Library for Contextual Object Recognition.* En 3rd. Workshop on Recognition and Action for Scene Understanding (REACTS), Valletta, Malta, (2015).
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. Mobile Robot Object Recognition through the Synergy of Probabilistic Graphical Models and Semantic Knowledge.* En European Conference on Artificial Intelligence, Workshop on Cognitive Robotics (CogRob), Prague, Czech Republic, (2014).

Marco de la tesis

Esta tesis es el resultado de 5 años de trabajo del autor como miembro del grupo *Machine Perception and Intelligent Robotics*⁷ (MAPIR), el cual se encuentra dentro del departamento de *Ingeniería de Sistemas y Automática* de la *Universidad de Málaga*. La investigación realizada ha sido principalmente financiada por el programa de ayudas *Formación de Profesorado Universitario* (FPU), promovido por el *Ministerio de Educación*.

Durante este periodo, el autor completó con éxito el programa doctoral en *Ingeniería Mecatrónica*, coordinado por el mismo departamento del que es miembro, donde obtuvo un conocimiento sólido sobre los pilares fundamentales de la robótica: sistemas de control, sistemas electrónicos, sistemas mecánicos, y ordenadores. Esta educación académica fue completada con distintos cursos, como es el caso de *Writing in the sciences*, impartido por la *Universidad de Stanford*, y la participación en la *Primera Örebro Winter School in Artificial Intelligence and Robotics*, la cual pretende acercar dos campos estrechamente relacionados como son los de la Inteligencia Artificial y la Robótica. Esta escuela también hizo posible el conocer otros investigadores en el mismo campo de estudio, relaciones que se mantienen a día de hoy.

El autor también completó una estancia de tres meses en el *Knowledge-Based Systems Research Group*⁸, en la *Universidad de Osnabrück* en Alemania, durante el año 2014, bajo la supervisión de *Prof. Dr. Joachim Hertzberg*. Durante este tiempo la investigación realizada se centró en el análisis y la implementación de diferentes algoritmos para el manejo eficiente de *PGMs*, así como de su aplicación para el reconocimiento *online* de objetos en robots móviles. En esta gran experiencia también se establecieron colaboraciones con distintos miembros del grupo receptor.

Además, también cabe destacar que el autor ha estado activo en el proceso de revisión de artículos de conferencias y revistas prestigiosas, como es el caso de las conferencias *International Conference on Robotics and Automation* (ICRA, 2014, 2015, 2016), e *International Conference on Intelligent Robots and Systems* (IROS, 2015), o las revistas *Association for the Advancement of Artificial Intelligence* e *Intelligent Service Robotics*.

⁷<http://mapir.isa.uma.es/>

⁸www.inf.uos.de/kbs/

La beca FPU también ofreció al autor la oportunidad de colaborar como profesor asistente con el departamento del que es miembro. Concretamente, impartió docencia en la asignatura de *Robótica* en la *Escuela Técnica Superior de Ingeniería Informática*, en la *Universidad de Málaga*. También supervisó el trabajo fin de grado de un estudiante, David Zúñiga, titulado *Visual SLAM with RGB-D Cameras Based on Pose Graph Optimization*.

Además de la investigación presentada en esta tesis, el autor también ha participado en otros proyectos dentro del grupo MAPIR, algunos de ellos de temática relacionada:

- **TCS: Tunnel Continuous Setout** (Nov'08 – Jul'11): Este proyecto se centró en el desarrollo de un sistema para el replanteo automático de secciones de túneles a ser perforadas. El prototipo del sistema, que toma el mismo nombre que el proyecto, combina una unidad de escaneo que realiza mediciones sobre el frente de excavación y un láser proyector que continuamente muestra la sección del túnel a perforar. La parte más desafiante del proyecto fue la implementación de las técnicas de calibración para localizar con exactitud todos los componentes del sistema dentro de un marco de referencia global.
- **ExCITE: Enabling SoCial Interaction Through Embodiment** (Jul'10 – Jun'13): El rol del autor en este proyecto estuvo relacionado con el desarrollo de mejoras técnicas para la plataforma robótica de telepresencia *Giraff*: un manejo más simple y seguro, detección de obstáculos, y visualización de la posición del robot en un mapa esquemático del lugar visitado. Una arquitectura de control, llamada *Navigation Assistant* (NAS), fue desarrollada para cumplir con estas necesidades especiales.
- **Taroth: New developments toward a Robot at Home** (Ene'12 – Dic'15): Este proyecto persiguió tres objetivos principales: i) aumentar la independencia del robot en cuanto a su movimiento, ii) integrar y explotar información semántica para mejorar la autonomía del robot y permitirle interactuar con humanos, y iii) desarrollar una arquitectura de control robótica para el manejo de servicios de la llamada *Ambient Assisted Living*, como son el entretenimiento, la domótica, las relaciones sociales, la seguridad, etc.
- **IRO: Improvement of the sensorial and autonomous capability of Robots through Olfaction** (Ene'14 – Feb'19): La investigación en este proyecto se orienta al estudio de mecanismos para usar información olfativa en problemas como el reconocimiento de objetos y la interpretación de la actividad en una escena. Dicho estudio presta especial atención al rol de la información semántica en los procesos de percepción por parte del robot y toma de decisiones, persiguiéndose una mejora en términos de eficiencia, autonomía y utilidad.

Del trabajo del autor en estos proyectos se desprendieron una serie de publicaciones adicionales:

Revistas

- *Javier Gonzalez-Jimenez, Vicente Arévalo, Cipriano Galindo, y Jose-Raul Ruiz-Sarmiento. **An Automated Surveying and Marking System for Continuous Setting-out of Tunnels.** En Computer-Aided Civil and Infrastructure Engineering, vol. 31, no. 3, pp. 219–228, (2016).*

Conferencias

- *David Zuñiga-Noël, Jose-Raul Ruiz-Sarmiento, y Javier Gonzalez-Jimenez. **Detección de Lugares con Cámaras RGB-D. Aplicación a Cierre de Bucles en SLAM.** En XXXVII Jornadas de Automática, Madrid, Spain, (2016).*
- *Javier Gonzalez-Jimenez, Jose-Raul Ruiz-Sarmiento, y Cipriano Galindo. **Improving 2D Reactive Navigators with Kinect.** En 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Reykjavic, (Iceland, 2013).*
- *Javier Gonzalez-Jimenez, Cipriano Galindo, Francisco Melendez-Fernandez, y Jose-Raul Ruiz-Sarmiento. **Building and Exploiting Maps in a Telepresence Robotic Application.** En 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Reykjavic, Iceland, (2013).*
- *Javier Gonzalez-Jimenez, Cipriano Galindo, y Jose-Raul Ruiz-Sarmiento. **Technical Improvements of the Giraff Telepresence Robot Based on Users' Evaluation.** En The 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Paris, France, (2012).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. **Cámaras basadas en tiempo de vuelo. Uso en la mejora de métodos de detección de caras.** En XXXII Jornadas de Automática, Sevilla, Spain, (2011).*

Informes técnicos

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, y Javier Gonzalez-Jimenez. **Experimental Study of the Performance of the Kinect Range Camera for Mobile Robotics.** Universidad de Malaga, Andalucia Tech, Departamento de Ingeniería de Sistemas y Automática, (2013).*

Estructura de la tesis

Mas allá del capítulo introductorio (**Chapter 1: Introduction**) el resto de capítulos en la primera parte de esta tesis (**Part I: Thesis description**) están organizados como sigue:

Chapter 2: Theoretical background provee nociones básicas sobre la teoría detrás de dos herramientas intensivamente empleadas en esta tesis: *PGMs* y *KBs*, de tal modo que el lector no experto en estas materias pueda obtener un conocimiento básico para una mejor comprensión de los siguientes capítulos. El autor ha tratado que sea una lectura lo más amena posible.

Chapter 3: Contextual scene understanding describe los enfoques tradicionalmente seguidos para el reconocimiento de objetos y habitaciones por parte de un robot móvil, y de que modo están relacionados con las contribuciones presentadas. También se dan detalles sobre la sinergia entre *PGMs* y *SK* codificado en *KBs* persiguiendo el entendimiento de escenas. Este capítulo también discute los repositorios de datos empleados para evaluar las técnicas desarrolladas, incluyendo *UMA-Offices*, así como el software implementado para manejar *PGMs*.

Chapter 4: Semantic Mapping esboza las representaciones de mapas semánticos comúnmente empleadas en robótica móvil, y describe las contribuciones de esta tesis en relación a una representación capaz de manejar información incierta: el *Multiversal Semantic Map*. Las virtudes de dicho mapa han sido comprobadas empleando un novedoso dataset, *Robot@Home*, cuyas características son descritas en este capítulo, junto con las del software usado para su procesamiento: *Object Labeling Toolkit*.

Chapter 5: Summary of included papers lista los artículos que conforman la segunda parte de esta tesis, **Part II: Included papers**, describiendo brevemente su contenido y el papel del autor en los mismos.

Chapter 6: Conclusions and future work discute las conclusiones que se pueden extraer del trabajo realizado, así como las líneas de investigación que quedan abiertas e interesantes extensiones a dicho trabajo.

Publicaciones incluidas en la Tesis

Esta sección realiza un esbozo de los artículos incluidos en la segunda parte de la tesis, así como las contribuciones del autor a cada uno de ellos.

Artículo A: Aprendiendo *Conditional Random Fields* con datos provenientes de *Semantic Knowledge*

Descripción: Este trabajo estudia la aplicabilidad de *CRFs* entrenados con datos sintéticos, generados a partir de *SK*, al problema del reconocimiento de objetos explotando su contexto. El objetivo de este enfoque para el entrenamiento es el de evitar la recopilación de datos reales para ajustar sistemas de reconocimiento. Dicha recopilación es una tarea pesada que requiere de una alta dedicación temporal, además de

no ser realizable en ciertos entornos, ya que los datos recogidos deben ser suficientemente representativos del dominio del problema. Para solucionar esta cuestión se codifica *SK* en una Ontología, la cual define las clases (o tipos) de objetos del dominio de discurso (por ejemplo, en el dominio del hogar, ejemplos de estos tipos serían *horno*, *microondas*, *salón*, o *cocina*), sus propiedades y sus relaciones, y es usado para generar ejemplos de entrenamiento sintéticos. La conveniencia del método de aprendizaje propuesto debe ser comprobada empleando conjuntos de datos reales, por lo que UMA-Offices y NYUv2 [131] formaron el banco de pruebas necesario para responder a preguntas como: *¿Cuánto contribuyen las relaciones de contexto al éxito del método?*, *¿Cómo afecta el tamaño del conjunto de datos de entrenamiento al rendimiento?*, o *¿Capturan los datos sintéticos generados características y relaciones reales?*.

Contribución del autor: Estudió el estado del arte sobre *PGMs* y *KBs* abordando el problema del reconocimiento de los objetos de una escena. Diseñó el modo de codificar información relevante en la Ontología para su posterior aprovechamiento. Implementó el algoritmo para la generación automática de un número arbitrario de ejemplos de entrenamiento. Procesó el conjunto de datos UMA-Offices, y realizó los experimentos necesarios para demostrar la validez de la propuesta.

Artículo B: Categorización conjunta de objetos y habitaciones

Descripción: En este artículo se extienden los métodos desarrollados en el anterior trabajo para también considerar las habitaciones del entorno. Motivado por estudios recientes que destacan la conveniencia de modelar conjuntamente los problemas de reconocimiento de objetos y habitaciones (dada la influencia mutua que tienen los tipos de los objetos reconocidos y los tipos de las habitaciones), la Ontología definida en el Artículo A es aumentada para también incluir tipos de habitaciones, sus atributos, y relaciones entre ellas así como entre objetos y habitaciones. Un ejemplo de esta información sería que los dormitorios están usualmente conectados con pasillos y suelen contener camas. Los *CRFs* también son convenientemente adaptados para trabajar con diferentes tipos de variables aleatorias (representando categorías de objetos o habitaciones) y relaciones de contexto. Para validar el método se emplean escenas ilustrando entornos domésticos dentro del conjunto de datos NYUv2.

Contribución del autor: Estudió las técnicas en el estado del arte para modelar conjuntamente los problemas de reconocimiento de objetos y habitaciones. Diseñó la expansión de la Ontología en el artículo anterior, así como de los *CRFs* y el algoritmo implementado para la generación de ejemplos sintéticos. Realizó los experimentos que soportan las afirmaciones del trabajo.

Artículo C: Empleando *Semantic Knowledge* para un reconocimiento eficiente y coherente

Descripción: La complejidad de los *CRFs* aumenta considerablemente cuando se aplican a escenarios repletos de objetos. Esto implica la utilización de técnicas de inferencia aproximada para obtener los resultados de reconocimiento, lo que en algunos casos compromete el éxito del método en comparación con el uso de soluciones de inferencia exacta. Este artículo propone la utilización de *SK* para reducir la complejidad del proceso de inferencia. Dicho conocimiento, codificado de nuevo en una Ontología, se aprovecha para generar hipótesis sobre los tipos más probables a los que pueden pertenecer los objetos en la escena, empleando para ello sus características. Estas hipótesis son consideradas por el *CRF* como las únicas categorías candidatas posibles, reduciendo de este modo la complejidad del proceso de inferencia, incluso habilitando en ciertos casos la inferencia exacta. Adicionalmente, también se codifica en la Ontología información a priori sobre la frecuencia de aparición de los distintos tipos de objetos. Esta información muestra que, por ejemplo, en un entorno de oficinas es más probable encontrar un ordenador a un sofá, mientras que es bastante improbable encontrar una tabla de planchar. El artículo también propone una modificación a la formulación usual de los *CRFs* para el aprovechamiento de dicha información. La ganancia en cuanto a la eficiencia y coherencia proporcionada por esta solución es medida con los conjuntos de datos UMA-Offices y NYUv2.

Contribución del autor: Diseñó el marco para, empleando las hipótesis generadas mediante inferencia lógica sobre la Ontología, reducir la complejidad del modelo probabilístico. Adaptó la formulación de los *CRFs* para también considerar información previa sobre la frecuencia de aparición de los diferentes tipos de objetos desde la Ontología. Evaluó la reducción de complejidad conseguida y la mejora en cuanto a la coherencia de los resultados devueltos empleando dos repositorios de datos distintos.

Artículo D: Libería UPGMpp para manejar *Conditional Random Fields*

Descripción: Este trabajo presenta la librería *Undirected Probabilistic Graphical Models in C++* (UPGMpp), un paquete software para trabajar con este tipo de modelos probabilísticos. La librería está especialmente diseñada e implementada para ser eficiente a la hora de tratar el problema del reconocimiento de objetos y/o habitaciones. El artículo describe cómo usar el software para modelar este problema, y presenta sus tres partes fundamentales: *base* (implementa la funcionalidad para construir y manipular modelos gráficos), *training* (permite la definición de conjuntos de datos para entrenar los modelos), e inferencia (implementa algoritmos de inferencia probabilística). Para mostrar la flexibilidad y usabilidad de la librería, este trabajo ilustra los procesos necesarios para entrenar y testear – realizar inferencia sobre – *PGMs*, incluyendo ejemplos de código. También se reportan los resultados de reconocimiento devueltos por distintos métodos de inferencia al tratar con escenas del conjunto de

datos NYUv2, así como el tiempo de ejecución requerido por dichos métodos.

Contribución del autor: Estudió la teoría detrás de los *PGMs* no dirigidos, así como otras librerías relacionadas para tratar con los mismos. Diseñó e implementó las partes de la librería, con el objetivo de que fueran eficientes, versátiles, extensibles, y fáciles de usar. Hizo la librería pública, ejemplificó su uso, y realizó las mediciones sobre tiempos de ejecución y éxito del reconocimiento.

Artículo E: Conjunto de herramientas para el tratamiento de repositorios de datos con información RGB-D

Descripción: En este trabajo se presenta el conjunto de herramientas software *Object Labeling Toolkit* (OLT), desarrollado para el procesamiento eficiente de repositorios de datos compuestos de secuencias de observaciones RGB-D (intensidad, RGB, más profundidad, D), capturadas por un número arbitrario de sensores de este tipo. Para ello, OLT construye una reconstrucción 3D de cada secuencia de observaciones y permite al usuario, mediante una interfaz gráfica, anotar los objetos y habitaciones en dicha reconstrucción con el tipo al que pertenecen (cama, mesa, lámpara, cocina, etc.). El artículo describe sus componentes principales, a saber: pre-procesamiento del conjunto de datos, construcción de mapa 2D, localización de las poses de las observaciones, visualización secuencial, etiquetado de la escena, y propagación automática de etiquetas a cada observación individual, de los cuales sólo el etiquetado de la escena requiere la intervención de un operador humano. También se ejemplifica el uso de OLT para el etiquetado fácil y rápido de dos secuencias de observaciones RGB-D, analizando sus virtudes con respecto a una técnica de etiquetado tradicional.

Contribución del autor: Diseñó el conjunto de herramientas. Estudió e implementó/adaptó las técnicas necesarias para los procedimientos de: procesado de imágenes tanto RGB como de profundidad, construcción de mapas geométricos 2D, reconstrucción de escenas 3D, visualización e interacción con las reconstrucciones, y propagación automática de las anotaciones a través de las secuencias de observaciones. Comparó el tiempo ahorrado empleando OLT con respecto al uso de una técnica de etiquetado típica.

Artículo F: Mapa semántico capaz de manejar incertidumbre

Descripción: En este artículo se propone un mapa semántico novedoso que permite la manipulación de incertidumbre, también aprovechando las relaciones contextuales de los elementos espaciales en el entorno del robot (objetos y habitaciones). Esta representación adopta el nombre de *Multiversal Semantic Map* (*MvSmap*). El artículo proporciona un estudio completo sobre otros enfoques para realizar un mapeo semántico del entorno, así como de técnicas para poblar dichos mapas. Los *MvSmaps* son descritos en detalle y definidos formalmente, incluyendo los algoritmos necesarios para su construcción, donde las técnicas de reconocimiento desarrolladas en trabajos

previos tienen un rol principal. Además, este trabajo estudia algoritmos para tratar eficientemente la incertidumbre modelada en estos mapas. Finalmente, el conjunto de datos Robot@Home [123] es el elegido para evaluar el rendimiento de los distintos sistemas envueltos en la construcción de *MvSmaps*.

Contribución del autor: Diseñó la representación *Multiversal Semantic Map* para el almacenamiento y tratamiento de información incierta. Integró las técnicas de reconocimiento de objetos y habitaciones anteriormente desarrolladas en un sistema para poblar dichas representaciones. Diseñó e implementó el proceso para la construcción de *MvSmaps* de acuerdo a la información percibida por un robot móvil. Procesó el conjunto de datos Robot@Home para que fuera útil durante el testeo de los sistemas en este trabajo.

Conclusiones y líneas futuras

Esta tesis ha explorado y hecho contribuciones al fascinante mundo del mapeo semántico del entorno por medio de un robot móvil. Este tipo de mapas dotan al robot de herramientas para comprender cuales son los elementos y espacios que tiene a su alrededor, así como sus propiedades, lo cual sienta las bases para una operación inteligente, autónoma y eficiente. En la investigación llevada a cabo se ha prestado especial atención a la población de mapas semánticos con información sobre los elementos espaciales en el entorno de trabajo del robot, es decir objetos y habitaciones, a través de la combinación de técnicas de los campos del *Aprendizaje Automático* y la *Inteligencia Artificial*. Estos campos se encuentran actualmente en un momento dulce, donde los estudios y aplicaciones en las que son utilizados sigue creciendo, tal y como apuntó en una reciente entrevista uno de los directivos de Amazon, Ralf Herbrich, afirmando que “*Estamos en una edad dorada para el aprendizaje automático y la inteligencia artificial. Nos encontramos aún lejos de hacer cosas del mismo modo en el que los humanos las hacen, pero estamos solventando problemas increíblemente complejos cada día y consiguiendo un progreso increíblemente rápido*”. En opinión del autor, la investigación de sistemas que aprovechen la sinergia de sendos campos, potenciando sus ventajas y mitigando sus limitaciones, puede llevar a avances notables en la comunidad robótica. Este es el caso de las técnicas desarrolladas en la presente tesis.

Para que un robot móvil alcance un cierto grado de consciencia del entorno en el que se desenvuelve, este debe ser capaz de reconocer los elementos espaciales observados a través de su sistema sensorial. El primer grupo de contribuciones de esta tesis trata este tema, centrándose en la combinación de *Conditional Random Fields* (CRFs), una variante discriminativa no dirigida de los *Probabilistic Graphical Models* (PGMs), y *Semantic Knowledge* (SK) del dominio de discurso codificado en una Ontología. Ambos enfoques han alcanzado un éxito notable en distintos problemas de clasificación.

Por un lado, los CRFs permiten el modelado de relaciones de contexto entre elementos espaciales, al mismo tiempo que maneja la incertidumbre proveniente del

sistema sensorial del robot y de los modelos empleados para definir el problema. Estos modelos también permiten la ejecución de métodos de inferencia probabilística. Precisamente, una de las primeras contribuciones de esta tesis fue la librería *Undirected Probabilistic Graphical Models in C++* (UPGMpp), desarrollada como consecuencia de la ausencia de herramientas software para manejar *PGMs* no dirigidos en general, y *CRFs* en particular, proveyendo las características que demanda un sistema de reconocimiento ejecutándose en un robot móvil (e.g. eficiencia, flexibilidad, o facilidad de integración). Esta librería, disponible públicamente, implementa algoritmos populares para la construcción, aprendizaje e inferencia sobre modelos gráficos. Las posibles combinaciones de métodos para entrenar e inferir información sobre *CRFs* motivó el estudio de diferentes estrategias de aprendizaje, el cual reportó valiosas conclusiones no sólo para la correcta utilización de estos modelos en el resto de contribuciones, sino para su empleo por parte de cualquier miembro de la comunidad robótica que desee configurar rápidamente un sistema de reconocimiento tan exitoso como sea posible.

A pesar de su notoria utilización en distintos campos, los *CRFs* muestran una serie de limitaciones a la hora de ser aplicados al problema de reconocimiento. En primer lugar, para ser correctamente entrenados requieren una considerable cantidad de ejemplos (datos) que, además, cubran por completo los elementos dentro del dominio de trabajo. La recogida de dichos conjuntos de datos es una tarea tediosa y que requiere una alta dedicación temporal, además de ser irrealizable en algunos dominios, tal y como experimentó el autor al procesar el repositorio *UMA-Offices*. Dicho conjunto de datos contiene 25 escenas capturadas por un robot móvil en entornos de oficinas de la Universidad de Málaga, y se recogió con el fin de evaluar las técnicas de reconocimiento desarrolladas – de manera conjunta con otros repositorios del estado del arte. Para evitar la dependencia de conjuntos de datos conteniendo información real, se mostró como *SK*, convenientemente codificado en una Ontología, puede usarse para generar sin esfuerzo una cantidad arbitraria de datos de entrenamiento representativos del dominio de discurso. Las Ontologías suponen una manera natural de codificar *SK*, además de ser compactas, leíbles por un humano, y directamente utilizables en tareas de razonamiento de alto nivel. No obstante, son incapaces de manejar incertidumbre, y es complejo dar el salto de información sensorial de bajo nivel a información codificada sin emplear procesos *ad-hoc*. Su combinación con *CRFs* elimina estas limitaciones, sentando las bases de una relación de beneficio mutuo.

En esta tesis se ha mostrado como las Ontologías que codifican *SK* tienen mucho más que ofrecer en su matrimonio con *CRFs*. Por ejemplo, se han empleado para generar hipótesis sobre los posibles tipos de objetos/habitaciones en una escena, reduciendo drásticamente la complejidad de los *CRFs* cuando modelan dicha escena. Esto incrementa la eficiencia de los métodos de inferencia aproximada sobre *CRFs*, así como amplía el abanico de escenarios donde es posible realizar una inferencia exacta. Nótese que la eficiencia del método de reconocimiento es fundamental para el apropiado funcionamiento del robot, ya que este debe compartir los (usualmente limitados) recursos del robot con otros algoritmos en ejecución, como puedan ser los de navegación o localización. Además, las Ontologías pueden codificar distintos tipos de

información sobre los elementos del dominio, lo cual se ha aprovechado para definir la frecuencia de aparición de los distintos tipos de objetos. La usual formulación de los *CRFs* ha sido consecuentemente adaptada para explotar esta fuente de información, permitiendo a estos modelos alcanzar unos resultados de reconocimiento más coherentes. El *SK* también se ha empleado para la detección de incoherencias en los resultados, y para aprender de las mismas en colaboración con un humano. Este enfoque soluciona la incapacidad de los *CRFs* para aprender de experiencias pasadas, y les permite mejorar su rendimiento y robustez a largo plazo en su aplicación a entornos humanos.

Una vez desarrolladas las técnicas para el reconocimiento, estas fueron integradas en un sistema de mapeo semántico. Para ello se diseñó una novedosa representación del entorno llamada *Multiversal Semantic Map (MvSmap)*, la cual es capaz de acomodar y aprovechar los resultados probabilísticos de los métodos de reconocimiento. Dicho mapa considera diferentes interpretaciones de los elementos espaciales, o *universos*, como instancias de Ontologías, creándose un *multiverso*. Estas Ontologías son además automáticamente anotadas con las probabilidades devueltas por el sistema de reconocimiento, así como con su probabilidad de ser las interpretaciones correctas. De este modo, el desempeño del robot no se limita a la utilización del universo más probable, *modus operandi* de los mapas semánticos tradicionales, sino que también puede considerar otras posibles explicaciones con diferentes interpretaciones semánticas. Además se discutió una estrategia para mantener tratable el número de universos considerados, clave para la eficiencia de esta representación semántica.

También se han hecho públicos dos recursos relacionados con las técnicas de mapeo semántico. El primero se corresponde con el conjunto de datos *Robot@Home*, el cual contiene, entre otros: 87,000+ observaciones recogidas en distintas casas por un robot móvil dotado de un aparejo con 4 cámaras RGB-D y un escáner láser 2D, reconstrucciones tanto en 2D como en 3D de las escenas exploradas, información topológica sobre la conectividad de las habitaciones, y anotaciones sobre los tipos de los objetos y habitaciones percibidos. El repositorio de datos es rico en información contextual de los elementos espaciales antes mencionados, una característica que no se encuentra en la mayoría de los repositorios actuales, lo cual puede ser aprovechado por sistemas de mapeo semántico. La segunda contribución a este respecto es el conjunto de herramientas denominado *Object Labeling Toolkit (OLT)*, diseñado para procesar eficientemente repositorios de datos compuestos de secuencias de observaciones RGB-D. Estas herramientas son altamente personalizables y expansibles, facilitando la integración de algoritmos ya desarrollados, y han mostrado su utilidad para reducir drásticamente el tiempo y esfuerzo necesarios para procesar repositorios conteniendo ese tipo de información. Por ejemplo, OLT fue usado para el procesamiento de *Robot@Home*.

Como observación final, cabe destacar que aunque las técnicas descritas en esta tesis han sido evaluadas con conjuntos de datos provenientes de entornos domésticos y de oficinas, su utilización no se limita a esos dominios, sino que pueden ser empleadas en cualquier escenario que exhiba información semántica como pueda ser el caso de hospitales o centros comerciales. También es interesante añadir que su uso no

está restringido al campo de la robótica móvil, sino que podrían ser exportadas a otros campos que se pudieran beneficiar de la explotación de mapas semánticos tales como asistencia a invidentes o personas mayores, realidad aumentada, y otras aplicaciones por venir en la era de los dispositivos portátiles con gran capacidad de cómputo. Hoy en día, de hecho, nuestros teléfonos móviles son casi tan potentes como los ordenadores de sobremesa. Los esfuerzos en la investigación en mapeo semántico, junto con los avances tecnológicos, nos aseguran la aparición de apasionantes y rompedoras aplicaciones. ¡Manténgase atento!.

Trabajos futuros

El trabajo realizado en la presente tesis deja abiertas una serie de líneas de investigación y expansiones. Algunas de las más relevantes se describen a continuación.

Generación de hipótesis. La generación de hipótesis empleando la información codificada en la Ontología podría ser demasiado restrictiva en algunas situaciones, principalmente con objetos que muestran unas características particulares. Supóngase una escena con un libro en el suelo. En esta situación el razonador lógico no devolvería la clase *libro* como hipótesis, dado que su altura desde el suelo difiere en gran medida de la esperada. Una opción podría ser considerar el resultado del proceso de inferencia lógica como una puntuación a ser considerada en la formulación de los *CRFs*, a expensas de comprometer la opción de inferencia exacta.

Aprovechamiento de los MvSmaps. El potencial real de los *Multiversal Semantic Maps* (en opinión del autor) está aún por verse. Se han diseñado y realizado diversas pruebas de concepto en tareas típicamente robóticas, pero debe estudiarse en mayor detalle el beneficio de estos mapas en problemas reales como navegación eficiente y búsqueda de objetos, localización del robot, planificación de tareas con información incierta/incompleta, etc.

Aprendiendo de experiencias. El sistema propuesto para el aprendizaje en base a la experiencia acumulada puede ser ampliado en diferentes aspectos. Primero, debe realizarse una evaluación rigurosa del sistema empleando complejos *CRFs* y Ontologías, incluyendo información de objetos y habitaciones, a lo largo de extensos periodos de tiempo. También podría estudiarse, dado que un humano forma parte del bucle de aprendizaje, cómo afectan al rendimiento del sistema posibles instrucciones incorrectas por parte del usuario. Además el sistema también se podría beneficiar de un estudio acerca de cuándo sería más apropiado preguntar a dicho humano sobre un resultado incoherente, de tal manera que se le moleste lo mínimo posible.

Posibles desarrollos dentro de UPGMpp. Sería interesante explorar algunas características adicionales relacionadas con el rendimiento de UPGMpp. Por ejemplo, aunque las partes que requieren más tiempo de ejecución han sido paralelizadas empleando *OpenMP*, algunas operaciones repetitivas que utilicen datos de forma ma-

siva podrían beneficiarse de su ejecución en núcleos GPU empleando, por ejemplo, *CUDA* u *OpenCL*. También sería útil el contar con herramientas gráficas para visualizar y modificar los grafos de los *PGMs*, así como para comprender cómo evolucionan en tiempo de ejecución. También se contempla la incorporación de técnicas para la generación de muestras de la distribución de probabilidad definida por un *PGM* (como *Markov Chain Monte Carlo*). Por supuesto, es bienvenida cualquier contribución a esta librería por parte de la comunidad robótica o de visión por computador.

Mejoras a OLT. La incorporación de algoritmos para un registro globalmente consistente de las observaciones RGB-D en una secuencia podría dar lugar a reconstrucciones incluso más precisas. La experiencia de usuario también se podría mejorar considerando otras primitivas geométricas para segmentar y etiquetar escenas, además de las cajas empleadas actualmente, como puedan ser esferas o cilindros. Por último, el tiempo necesario para el etiquetado también podría reducirse si se ofreciera al usuario una segmentación inicial de la escena, así como etiquetas tentativas para los objetos/habitaciones apareciendo en la misma.

Punto y aparte

Esta sección concluye el resumen de la presente tesis, *Probabilistic Techniques in Semantic Mapping for Mobile Robotics*. El lector puede continuar con los siguientes capítulos, en el idioma inglés, donde se describen en mayor detalle las contribuciones de la misma.

Part I

Thesis description

Introduction

Winter is coming. A servant robot senses that the temperature is decreasing and takes a blanket to a lovely grandma. In the same building, another robot patrolling an offices' floor is alerted by a light turned on in a room; rapidly it notices that the research fellow, Bob, is working late in the night, the third time that week. Meanwhile, baby Alice, Bob's daughter, is sad because of the absence of her daddy, and her robotic colleague warmly nicknamed as *Roboto* looks for her favorite teddy. Sophie, Alice's mom, is also counting the hours to see her, and commands a robot to clean the tables once the restaurant she runs is closed to the public.

These scenarios are some examples where mobile robots, to a greater or lesser extent, can provide a number of services for raising the standards of living. Nowadays, it becomes clear that robots are coming to stay, as it is shown by their remarkable application to an increasing number of tasks where they collaborate with humans or release them from tedious or hazardous chores, such as surveillance, health care, companion, entertainment, household maintenance, etcetera [97]. Figure 1.1 depicts some examples of modern robots aimed at performing some of these tasks. Common to all these robotic applications is the necessity of building representations of the working environment, commonly referred to as *maps*, which permit a mobile robot to be aware of its surroundings in order to navigate avoiding obstacles, localize itself with respect to a given reference frame, store relevant information about spatial elements for accomplishing its goals, etc.

Traditional spatial representations, like geometric, topological, or hybrid maps, are extensively used due to the core skills they provide, *i.e.* navigation and localization. Nevertheless, the execution of high-level tasks, like the ones involved in the aforementioned scenarios, calls for more sophisticated representations closer to the way in which humans interpret and behave within their environments. *Semantic maps* came out to cope with this need, permitting a robot to *understand* not only the spatial aspects of its workspace, but also the meaning of its elements (objects and rooms) and how humans interact with them, *e.g.* functionalities, events, and relations. This is

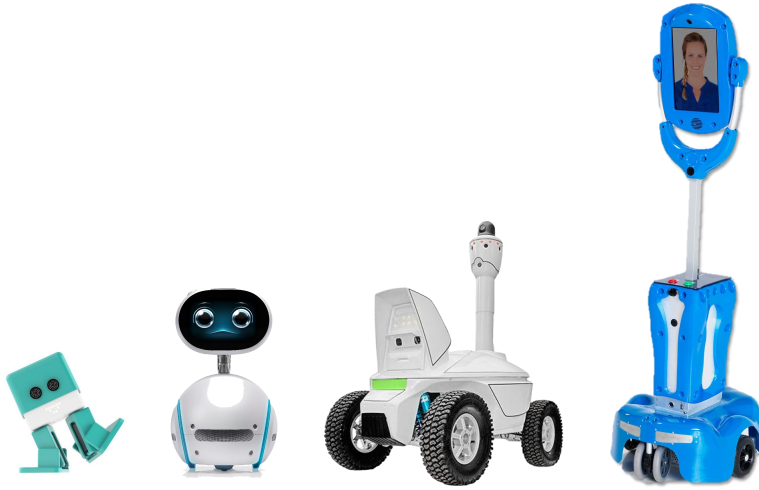


Figure 1.1: Examples of state-of-the-art robots successfully applied to different tasks. From left to right: the educational robot Zowi, the companion and entertainment robot Zenbo, the security patrol robot S5, and the Giraff robot employed in telehealth-care applications.

achieved by considering meta-information, commonly referred to as *common-sense* or *Semantic Knowledge* (SK), concerning the types of elements (and their relations) to be found in the robot workspace. Pieces of this information, typically encoded into a *Knowledge Base* (KB), could be: blankets are often stored in cupboards; lights must be switched off after the working day; teddies make kids happier; fragile crockery should not be cleaned in the dishwasher.

1.1 Motivation

Typically, semantic maps are populated with *crispy information*, *e.g.* an object is a blanket or not. This is due to the weakness of traditional semantic representations to handle uncertainty, which forces the use of recognition algorithms providing a crispy outcome, probably by thresholding a probabilistic result. For example, a recognition algorithm¹ stating that an object can be a blanket with a probability of 0.52, and a carpet with 0.48, might yield a unique outcome by considering the object as a blanket and neglecting the other, high probable, hypothesis. This crispy stance clearly compromises the robot operation: the uncertainty coming from sources like the robot sensory system and the employed models is being disregarded when the recognition results are stored in the semantic map. So, despite the results clamor for a disambiguation, our lovely grandma could end up with a rugged carpet on top of her. Therefore, it becomes clear the necessity of leveraging probabilistic techniques for populating and

¹For the sake of simplicity only two possible object types are considered at this point.

maintaining semantic maps, as well as to adapt semantic representations for managing uncertain information, which would permit a mobile robot to operate in a more coherent and efficient way.

As an illustrative example of the convenience of dealing with uncertain information, let's suppose an scenario where a servant robot right landed from the lab into its new home is commanded to bring the slippers to the grandma. In the absence of spatial information, the robot could infer (according to the loaded KB) that the most probable location for slippers is a bedroom. During the preliminary setup, the robot initially recognized a bedroom corresponding to the farthest room from the current grandma location with a probability of 0.45, and 0.43 of being a kitchen². Another room, close to the robot location, has been recognized as a kitchen with a probability of 0.48, and as a bedroom with 0.47. The utilization of only the most probable interpretation, *modus operandi* of traditional, crispy semantic maps, would lead to the exploration of the farthest room having a 45% of being the correct place, while the consideration of both interpretations would produce the more logical plan (for the robot battery and the grandma patience) of taking a look at the closer room first.

Although there exist numerous algorithms for the recognition of objects and/or rooms that provide uncertainty measurements about their results, they usually work by individually processing each spatial element according to its geometric/appearance features. In other words, if the most probable type of an object is blanket, it is considered a blanket no matter other objects placed nearby nor its location. Nevertheless, human-made environments are rich in contextual information worth to exploit, *i.e.* the room's layout follows a certain order, and objects are not placed randomly but following certain configurations according to their functionality: *e.g.* a remote control is usually found close to a tv, a corridor connects rooms, or bathtubs are (as indicated by its name) placed at bathrooms. Modeling and leveraging context is useful, for example, to disambiguate uncertain results: following the previous example, if the object is found into a wardrobe it would be more probably a blanket than a carpet, which are usually lying on the floor. This kind of information can be naturally encoded in KBs, however, its exploitation for contextual object/room recognition, also managing uncertainties, is not straightforward.

Probabilistic Graphical Models (PGMs) have been a widely resorted tool for modeling and exploiting contextual relations, while dealing with uncertainty. They work with a graph-based representation, where nodes stand for random variables and edges link variables showing some type of relation. For example, in the case of the object recognition problem, each object in the scene is represented by a random variable that takes values from the set of possible object types (table, book, couch, etc.), and nodes whose associated objects are close to each other in the scene are linked by an edge. This representation supports the efficient execution of probabilistic inference methods, which permit us to retrieve the scene object recognition results along with a measure of their uncertainty. PGMs have been successfully applied to tasks

²Notice that the sum of both probabilities is 0.88. The remaining probabilities, up to a total of 1, correspond to other possible room types: corridor, bedroom, living room. etc.

like image denoising, natural language processing, activity recognition, etc. However, they exhibit a number of limitations that could prevent their utilization for populating semantic maps: they become computationally intractable when the complexity of the problem increases, *i.e.* the number of objects/rooms in the environment and their types augments, they need a considerable amount of training data to tune successful models, and they are unable to detect incoherent results as well as to learn from experience.

1.2 Contributions

This thesis contributes to overcome some of the aforementioned limitations of traditional semantic maps by resorting to probabilistic techniques. Concretely, the goals of the thesis, which resulted in the development of those techniques, were stated as:

- **Development of reliable recognition methods:** To provide contextual object/room recognition algorithms able to exploit contextual relations and handle uncertainty, in close synergy with KBs, also offering a number of desirable features like scalability, efficiency, detection of wrong results, learning from experience, etc.
- **Enhancement of traditional representations to manage uncertainty:** To accommodate the probabilistic outcomes of such algorithms into a novel semantic map representation, in such a way that a robot could have a grounded belief about the certainty of its understanding of the surroundings, hence operating in a coherent fashion.

Thereby, the contributions of this thesis can be grouped into two major topics: contextual scene understanding, and semantic mapping.

1.2.1 Contributions to contextual scene understanding

The first set of contributions, presented in the papers [114, 121, 122, 115, 116, 119, 117] focuses on the *scene object and/or room recognition problems*. To overcome these problems is crucial for the proper building of the semantic representations sought. Probabilistic Graphical Models, concretely Conditional Random Fields (CRF), are used to model those issues from a holistic stance, considering the contextual relations among objects and/or rooms, and to natively deal with uncertainty. Their suitability for the problem at hand has been verified through a comprehensive evaluation of PGMs trained and exploited by the most popular learning and probabilistic inference algorithms.

These CRFs work in synergy with KBs, a mutually beneficial relationship which permits to keep their advantages and mitigate their limitations:

- KBs provide CRFs with the capabilities to: reduce their complexity, exploit prior information, verbalize their outcome, generate an arbitrary number of training samples, detect incoherent results, and learn from experience.

- CRFs enables KBs to handle uncertainty and exploit contextual relations in a holistic and principled manner.

The developed algorithms have been compared with other cutting-edge solutions employing state-of-the-art datasets. Additionally, a dataset consisting of 25 rooms from our facilities, called *UMA-Offices*, has been collected and made public. An open-source library, called *Undirected Probabilistic Graphical Models in C++* (UPGMpp), has been also implemented for working with PGMs paying attention to the special requirements of software targeted at robotic applications.

1.2.2 Contributions to semantic mapping

The goal of the second group of contributions, presented in the papers [123, 118, 120], is to accommodate the probabilistic outcome of the previous techniques into a semantic map representation. For that, the so-called *Multiversal Semantic Map* (*MvSmap*) representation has been developed. This map turns such outcome into different interpretations of the robot workspace, coined universes, which are annotated with their probability of being the true ones. This permits the robot to consider not only the most probable universe, but other ones also showing a high probability, hence unlocking a more coherent and efficient operation. Techniques to keep the number of possible universes tractable in complex environments, crowded of objects and rooms, has been also studied.

The suitability of this map as well as its capacity to efficiently handle uncertain information have been tested with a novel dataset, *Robot@Home*, collected by a mobile robot surveying a number of apartments. The *Object Labeling Toolkit* (OLT), publicly available for the researcher community, has been developed to effortlessly process datasets compounded of sequences of sensory information, such as *Robot@Home*.

1.2.3 Publications

The present thesis encompasses the following publications:

Journals

- Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Building Multiversal Semantic Maps for Mobile Robot Operation**. Submitted to Knowledge-Based Systems (2016).
- Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **A Survey on Learning Approaches for Undirected Graphical Models. Application to Scene Object Recognition**. In International Journal of Approximate Reasoning, accepted (2016).
- Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Robot@Home, a Robotic Dataset for Semantic Mapping of Home Environments**. Submitted to International Journal of Robotics Research (2016).

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Scene Object Recognition for Mobile Robots Through Semantic Knowledge and Probabilistic Graphical Models.** In Expert Systems with Applications, vol. 42, no. 22, pp. 8805–8816, (2015).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Exploiting Semantic Knowledge for Robot Object Recognition.** In Knowledge-Based Systems, vol. 86, pp. 131–142, (2015).*

Conference proceedings

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Probability and Common-Sense: Tandem Towards Robust Robotic Object Recognition in Ambient Assisted Living.** In 10th International Conference on Ubiquitous Computing & Ambient Intelligence, Las Palmas de Gran Canaria, Spain, (2016).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Joint Categorization of Objects and Rooms for Mobile Robots.** In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, (2015).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets.** In European Conference on Mobile Robots (ECMR), Lincoln, UK, (2015).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **UPGMpp: a Software Library for Contextual Object Recognition.** In 3rd. Workshop on Recognition and Action for Scene Understanding (REACTS), Valletta, Malta, (2015).*
- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. **Mobile Robot Object Recognition through the Synergy of Probabilistic Graphical Models and Semantic Knowledge.** In European Conference on Artificial Intelligence, Workshop on Cognitive Robotics (CogRob), Prague, Czech Republic, (2014).*

1.3 Thesis framework

This thesis is the result of 5 years of work by the author as a member of the Machine Perception and intelligent Robotics (MAPIR) research group³, part of the Department of System Engineering and Automation of the University of Málaga. This research has been mainly funded by the FPU (*Formación de Profesorado Universitario*) grant program, supported by the Spanish Education Ministry.

³<http://mapir.isa.uma.es/>

During this period, the author successfully completed the doctoral program in Mechatronics Engineering, coordinated by the Department of System Engineering and Automation, where he obtained a strong background knowledge concerning the four fundamental pillars of robotics: control systems, electronic systems, mechanical systems, and computers. This academic education was completed with different courses, like the “Writing in the sciences” course imparted by the Stanford University, and with the participation in the First Örebro Winter School on “Artificial Intelligence and Robotics”, which aimed to bring closer two fields strongly correlated like Artificial Intelligence and Robotics. This school also made possible to meet other researchers in the same and other related fields.

The author also completed a three months research stay at the Knowledge-Based Systems Research Group⁴, of the University of Osnabrück, in 2014, under the supervision of Prof. Dr. Joachim Hertzberg. During this time, research focused on the analysis and implementation of different algorithms for efficiently handling PGMs, as well as in their application to online object recognition in mobile robots. In this great experience, cooperations with researchers of the group were also established.

Besides, it is also worth to mention that the author has been active in the review process of papers/articles from prestigious conferences and journals, like in the case of the International Conference on Robotics and Automation (ICRA, 2014, 2015, 2016), the International Conference on Intelligent Robots and Systems (IROS, 2015), or the Association for the Advancement of Artificial Intelligence and the Intelligent Service Robotics journals.

The FPU grant also offered the opportunity to collaborate as an assistant lecturer with the Department of System Engineering and Automation. Concretely, the author taught on ‘Robotics’ at the faculty of Computer Science, in the University of Málaga. He also co-supervised the bachelor thesis of a student, David Zúñiga Noël, entitled “Visual SLAM with RGB-D Cameras Based on Pose Graph Optimization”.

In addition to the research concerning this thesis, the author has been also involved in other projects within the MAPIR group, some of them with related topics:

- **TCS: Tunnel Continuous Setout** (Nov’08 – Jul’11): this project focuses on the development of a system for the automatic setting-out of tunnel sections to be perforated. The system prototype, which takes the same name as the project, combines a scanning device that surveys the excavation front and a laser projector that continuously displays the actual tunnel section. The most challenging part of the project was the implementation of calibration techniques for retrieving the accurate location of all the system components.
- **ExCITE: Enabling SoCial Interaction Through Embodiment** (Jul’10 – Jun’13): The author’s role in this project was related to the development of technical improvements for the Giraff telepresence platform: a safer and easier driving, including auto-docking to the recharging station, obstacle detection, and displaying the robot position in a sketch map of the visited place. A robotic

⁴www.inf.uos.de/kbs/

architecture called Navigation Assistant (NAS) was also implemented to fulfill these particular needs.

- **Taroth: New developments toward a Robot at Home** (Jan'12 – Dec'15): this project pursues the three following targets: 1) improving dependability of the robot motion, 2) integrating and exploiting semantics to improve robot autonomy and interaction with humans, and 3) developing a robot software architecture that can manage Ambient Assisted Living services related to entertainment, domotics, social networking, safety, etc.
- **IRO: Improvement of the sensorial and autonomous capability of Robots through Olfaction** (Jan'14 – Feb'19): the research in this project is targeted at the investigation of mechanisms to use odor information in problems such as object recognition and scene-activity understanding, paying special attention to the role of semantics within the robot perception and decision-making processes, aiming to improve the robot capabilities in terms of efficiency, autonomy and usefulness.

From the author's work in these projects arose a number of additional publications:

Journals

- *Javier Gonzalez-Jimenez, Vicente Arévalo, Cipriano Galindo, and Jose-Raul Ruiz-Sarmiento. **An Automated Surveying and Marking System for Continuous Setting-out of Tunnels.** In Computer-Aided Civil and Infrastructure Engineering, vol. 31, no. 3, pp. 219–228, (2016).*

Conference proceedings

- *David Zuñiga-Noël, Jose-Raul Ruiz-Sarmiento, and Javier Gonzalez-Jimenez. **Detección de Lugares con Cámaras RGB-D. Aplicación a Cierre de Bucles en SLAM.** In XXXVII Jornadas de Automática, Madrid, Spain, (2016).*
- *Javier Gonzalez-Jimenez, Jose-Raul Ruiz-Sarmiento, and Cipriano Galindo. **Improving 2D Reactive Navigators with Kinect.** In 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Reykjavic, (Iceland, 2013).*
- *Javier Gonzalez-Jimenez, Cipriano Galindo, Francisco Melendez-Fernandez, and Jose-Raul Ruiz-Sarmiento. **Building and Exploiting Maps in a Telepresence Robotic Application.** In 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Reykjavic, Iceland, (2013).*
- *Javier Gonzalez-Jimenez, Cipriano Galindo, and Jose-Raul Ruiz-Sarmiento. **Technical Improvements of the Giraff Telepresence Robot Based on Users' Evaluation.** In The 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Paris, France, (2012).*

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. Cámaras basadas en tiempo de vuelo. Uso en la mejora de métodos de detección de caras.* In XXXII Jornadas de Automática, Sevilla, Spain, (2011).

Technical reports

- *Jose-Raul Ruiz-Sarmiento, Cipriano Galindo, and Javier Gonzalez-Jimenez. Experimental Study of the Performance of the Kinect Range Camera for Mobile Robotics.* Universidad de Malaga, Andalucia Tech, Departamento de Ingenieria de Sistemas y Automatica, (2013).

1.4 Thesis outline

Besides the introductory chapter, the remaining ones in the first part of this thesis, **Part I: Thesis description**, are organized as follows:

Chapter 2: Theoretical background gives brief notions of the theory behind two frameworks constantly resorted in this thesis: Probabilistic Graphical Models and Knowledge Base representations, so the non-expert readers in this field can get the basic background for a proper understanding of the next chapters. The author has tried his best to make the reading of this chapter as pleasant as possible.

Chapter 3: Contextual scene understanding describes the traditional approaches followed for the recognition of objects and rooms by a mobile robot, and how they are related to the presented contributions exploiting contextual information. Details about the synergy of PGMs and Semantic Knowledge for scene understanding are provided. This chapter also discusses the datasets used to test the developed techniques, including the *UMA-Offices* one, as well as the implemented software in this respect: the *Undirected Probabilistic Graphical Models in C++* library.

Chapter 4: Semantic Mapping outlines the semantic map representations traditionally used in mobile robotics, and describes the thesis contribution for a representation handling uncertain information: the *Multiversal Semantic Map*. The virtues of this map have been checked against a novel dataset, *Robot@Home*, whose features are described in this chapter along with those of the software used for its processing: the *Object Labeling Toolkit*.

Chapter 5: Summary of included papers lists the papers that make up the second part of the thesis, **Part II: Included papers**, giving a brief description of their content and contributions.

Chapter 6: Conclusions and future work discusses the conclusions drawn from the work done in this thesis, as well as the research lines still open and possible extensions.

Theoretical background

This chapter briefly covers the theory behind two frameworks that have been essential for the research in this thesis. The first one is Probabilistic Graphical Models, used to holistically model the object and/or room recognition problems from a probabilistic stance. The second framework is Knowledge Bases, employed to encode Semantic Knowledge of the domain at hand for its posterior exploitation with different purposes. The synergy between both frameworks enables the design of sophisticated techniques to manage semantic maps.

2.1 Probabilistic Graphical Models

Probabilistic Graphical Models (PGMs) [65, 12] suppose a widespread framework from the Machine Learning field to efficiently model and exploit contextual relations, aiming to predict multiple, somehow dependent, random variables. These models are usually employed to deal with complex systems that involve uncertainty, which mainly arises from the limitations on the motion and sensory systems of the robot.

PGMs rely on a graph representation $G = (V, E)$, where the set V represents the random variables of the problem as nodes, while the edges $E \subseteq V \times V$ relate variables that are dependent in some way. This graph-based representation permit PGMs to compactly encode complex distributions over high-dimensional spaces, and to support the execution of probabilistic inference techniques for the prediction of the variable values. Thus, PGMs are strongly based on principles from graph theory and probability theory.

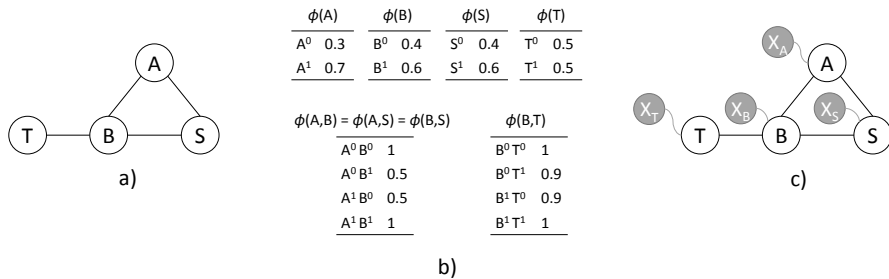


Figure 2.1: a) Graph representation of the MRF *happiness* model. b) Factors defined over such a graph. c) CRF representation including measures about different aspects.

PGMs have been successfully applied to a variety of domains like medicine, computer vision, robotics, etc. Depending on the types of edges, PGMs can be grouped on Directed or Undirected models. On the one hand, Directed Graphical Models, also called Bayesian Networks (BNs) [98], model the dependencies among nodes through directed edges, encoding *causality* relations. These models have been utilized with notable success in problems like medical diagnosis [84], biology [159], weather forecasting [1], or robotic localization and map building [14]. On the other hand, Undirected Graphical Models (UGMs), also called Markov Random Fields (MRFs) [63], employ undirected edges to define *symmetric* relations among random variables. This approach has reached a remarkable success in computer vision [50].

The choice between BNs and MRFs largely depends on the target application, since they are able to encode different types of dependencies (*e.g.* BNs can define induced dependencies, while MRFs are able to represent cyclic dependencies). In the case of the object/room recognition problem, the more suitable framework is such of MRFs, since the nature of the relations among objects and rooms is symmetric, and they can also exhibit loops, which are non trivial to model within the BNs framework. In its turn, the discriminative variant of MRFs, called Conditional Random Fields (CRFs) [70], are more appropriate in classification problems where the random variables are conditioned to observed data [59, 69]. The next section shows an example to illustrate the differences among these models.

2.1.1 The happiness example

Let's suppose the family formed by **Bob**, **Sophie**, and **Alice** presented in the introductory chapter, and a mobile robot with the goal of modeling their *happiness state* through a MRF. As human beings, we empathize with each other, and we are directly affected and affect the well-being and emotional state of our relatives, so it makes sense to take into account these relationships when trying to predict the happiness state of a person. PGMs model this in a principled way. Figure 2.1-a) shows the graph representation exemplifying the relations among the happiness state of each

family member, also including our lovely grandma, called Tess, who have nice conversations with Bob in the elevator. From this representation it can be inferred that the happiness of Alice, Bob, or Sophie directly influences the feelings of the other family members, while Tess has only influence and is influenced by Bob.

At this point, instead of modeling the whole probability distribution $P(\mathbf{y})$ (with $\mathbf{y} = [A, B, S, T]$), MRFs break it down into smaller pieces through the utilization of factors, *i.e.* functions defined over different parts of the graph. The first row of Figure 2.1-b) shows factors defined over the nodes of the graph, which are commonly called *unary factors*, stating the likelihood of these nodes to take certain values. Let's simplify the happiness of a person to two possible states, unhappy (0) and happy (1). Having a closer look at these factors, we can see for example that Alice is more probable to be happy than Tess. In its turn, the second row shows factors defined over pair of nodes, called *pairwise factors*, that set the likelihood about those nodes taking a certain values combination. The defined factors tell us that Bob, Alice and Sophie are prone to share their happiness, and although Bob and Tess are also inclined to have the same state, this influence is weaker. The values defined in a factor have not to sum up 1, since they are not probabilities.

Exhaustively defining $P(\mathbf{y})$ in this toy example requires the codification of $2^4 = 16$ probabilities. In this case, the MRF codification through factors does not save so much work, however, in more realistic scenarios with dozens, hundreds or thousands of random variables their utilization becomes crucial to keep the problem tractable. For example, a scenario with 20 binary random variables entails the definition of $2^{20} \simeq 10^6$ probabilities.

Thus, according to the Hammersley-Clifford theorem [48], the probability $P(\mathbf{y})$ can be factorized over the graph G as a product of factors $\phi(\cdot)$:

$$p(\mathbf{y}) = \frac{1}{Z} \prod_{c \in C} \phi(y_c) \quad (2.1)$$

where C is the set of maximal cliques¹ of the graph G , and $Z(\cdot)$ is the so-called partition function that plays a normalization role so $\sum_{\xi(\mathbf{y})} p(\mathbf{y}) = 1$, being $\xi(\mathbf{y})$ a possible assignment to the variables in \mathbf{y} . Therefore, the computation of the partition function is needed for computing the probability of a given assignment.

This way to define factors is rigid and naive: the happiness of a person can hardly be modeled by writing in stone his tendency to be happy, and it is additionally influenced by a number of (hopefully measurable) daily aspects: the sleeping hours, the success at work, hours spent with family and friends, etc. These aspects could be also included in the MRF graph as additional random variables, although the modeling of their probabilities and relations tend to be needlessly complex. Conditional Random Fields (CRF) [70] avoid the need to model them by conditioning the probability distribution over \mathbf{y} to the values of these aspects, referred to as *features*. Thus, a CRF

¹A maximal clique is a fully-connected subgraph that can not be enlarged by including an adjacent node.

works directly with the distribution $p(\mathbf{y} \mid \mathbf{x})$, where \mathbf{x} is the vector of observed features. Figure 2.1-c) shows the graph representation of a CRF considering this information. Additionally, instead of defining by hand the factors for each possible content of \mathbf{x} , they are parametrized through a vector of weights θ that are learned during the training phase of the CRF. Thus, the probability $p(\mathbf{y} \mid \mathbf{x})$ can be retrieved by:

$$p(\mathbf{y} \mid \mathbf{x}; \theta) = \frac{1}{Z(\mathbf{x}, \theta)} \prod_{c \in \mathcal{C}} \phi(y_c, x_c, \theta_c) \quad (2.2)$$

The parametrized factors can be formulated in different ways depending on the application. For example, in recognition problems, unary factors are often defined as $\phi_u(y_i, x_i, \theta) = \sum_{l \in \mathcal{L}} \delta(y_i = l) \theta_l f(x_i)$, where $f(x_i)$ computes a vector of features that characterizes the object x_i (*e.g.* size, shape, color, etc.), θ_l is the vector of weights for the class l obtained during the training phase, and $\delta(y_i = l)$ is the Kronecker delta function, which takes value 1 when $y_i = l$ and 0 otherwise. Pairwise factors are defined in a similar way, but considering a function that computes a vector of contextual features (*e.g.* difference of color, difference of orientation, etc.).

2.1.2 Learning the models

Training a CRF model for a given domain requires estimating the parameters θ , in such a way that they maximize the likelihood in Eq.2.2 with respect to a certain i.i.d. training dataset $D = [d^1, \dots, d^m]$, that is:

$$\max_{\theta} L_p(\theta : D) = \max_{\theta} \prod_{i=1}^m p(\mathbf{y}^i \mid \mathbf{x}^i; \theta) \quad (2.3)$$

where each training sample $d^i = (\mathbf{y}^i, \mathbf{x}^i)$ consists of a number of observed features from the elements of the problem at hand (\mathbf{x}^i), the people whose happiness is to be estimated in our example, and the corresponding ground truth information about their classification (\mathbf{y}^i), *i.e.* if they are happy (1) or not (0).

The optimization in Eq.2.3 is also known as Maximum Likelihood Estimation (MLE), and requires the computation of the partition function $Z(\cdot)$, which in practice is *NP*-hard, hence an intractable problem. Two major approaches stand out to overcome this concern: (i) the definition of alternative, tractable objective functions, or (ii) the estimation of the likelihood by approximate inference algorithms [68, 66, 96]. The performance of methods from both options highly differs depending on the domain of the problem at hand, *i.e.* the nature and internal structure of the data to work with. Therefore, for a certain application, a thorough study is needed in order to obtain a successful model, which motivates the analysis described in Chapter 3.

2.1.3 Probabilistic inference

Once a CRF is trained, and its graph representation modeling a given problem is built, it can be exploited by probabilistic inference methods to perform different probability

queries. At this point, two types of queries are specially relevant: the *Maximum a Posteriori* (MAP) query, and the *Marginal* query. The goal of the MAP query is to find the most probable assignment $\hat{\mathbf{y}}$ to the variables in \mathbf{y} , *i.e.* :

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} p(\mathbf{y} \mid \mathbf{x}; \theta) \quad (2.4)$$

Once again, the computation of the partition function $Z(\cdot)$ is needed, but since given a certain CRF graph its value remains constant, this expression can be simplified by:

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} \prod_{c \in \mathcal{C}} \exp(\langle \phi(x_c, y_c), \theta \rangle) \quad (2.5)$$

Nevertheless, this task checks every possible assignment to the variables in \mathbf{y} , so it is still unfeasible for real applications. An usual way to address this issue is the utilization of approximate methods, like the *max-product* version of Loopy Belief Propagation (LBP) [150], Iterated Conditional Models (ICM) [11], or Graph Cuts [15].

On the other hand, the Marginal query, which can be performed by, for example, the *sum-product* version of LBP [155], provides us beliefs about the possible assignments to the variables \mathbf{y} . In other words, this query yields the marginal probabilities for each element taking different values, as well as the compatibility of these assignments with respect to the values of contextually related elements. Notice that the most probable MAP assignment to a random variable can differ from the highest marginal probability. Additionally, with this query is also possible to estimate the probability of a certain assignment to the variables in \mathbf{y} .

2.2 Knowledge bases

Knowledge base (KBs) is the term used in Artificial intelligence (AI) to describe one of the two parts of a knowledge-based system, which is in charge of encoding semantic or common-sense knowledge about a particular domain in a computer-readable fashion. The other system part is a reasoning engine able to infer new information or detect inconsistencies in the KB. In the happiness example, a KB could encode the types of relations among persons, the different factors that affect their happiness, etc. (see Section 2.2.2), which are typically modeled through Ontologies. Knowledge-based systems have been a pivotal component for semantic mapping, as they permit a mobile robot to perform efficiently according to the information collected from the environment.

2.2.1 Ontologies

An Ontology is commonly defined as a representation of a conceptualization related to a knowledge domain, which accounts for a number of concepts arranged hierarchically, relations among them, and instances of such concepts, also called individuals [144]. Example of concepts could be Person or Happiness, while Person

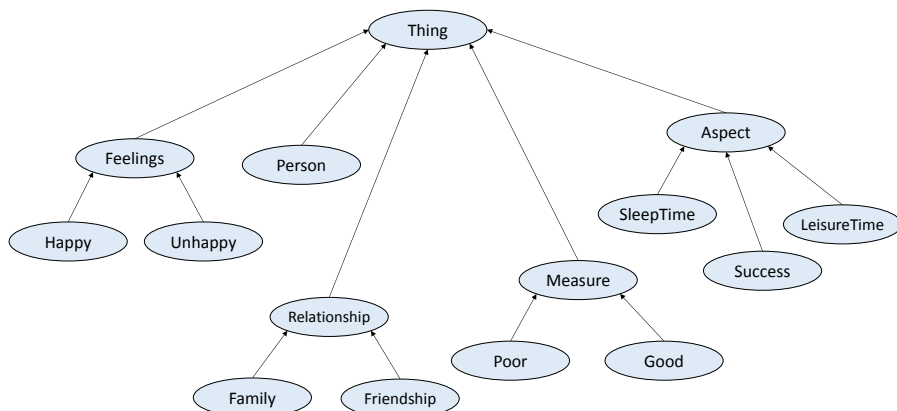


Figure 2.2: Hierarchy of concepts for the *happiness* domain.

`hasState Happiness` could be a relation stating the happiness of a person. Thus, the happiness of Bob, an individual of the concept `Person`, could be codified by `Bob hasState Happy`.

The process of obtaining and codifying Semantic Knowledge can be tackled in different ways. For example, web mining knowledge acquisition systems can be used as mechanisms to obtain information about the domain of discourse [158]. Available common-sense Knowledge Bases, like ConceptNet [134] or Open Mind Indoor Common Sense [46], can be also analyzed to retrieve this information. Another valuable option is the utilization of internet search engines, like Google's image search [29], or image repositories like Flickr [99], for extracting knowledge from user-uploaded information. Semantic Knowledge can be also codified through an human elicitation process, which supposes a truly and effortless encoding of a large number of concepts and relations between them. In contrast to online search or web mining-engine based methodologies, this source of semantic information (a person or a group of people) is trustworthy, so the uncertain about the validity of the information is reduced [119].

2.2.2 Happiness from an Ontological stance

Figure 2.2 shows an example of hierarchy of concepts from an Ontology modeling the *happiness* domain. The root concept is `Thing`, with 5 children codifying information about: the possible states of happiness, the person concept itself, different types of relationships among people, possible aspects that affect happiness, and measurements of those aspects. Using this Ontology, one can define, for example, that a happy person has a `Good SleepTime`, `Success` at work, and `LeisureTime`. Thus, if a `Person` shows these properties, a logical reasoner, like Pellet [133], FaCT++ [143], or Racer [47], can be used to automatically infer that such a person is happy.

Contextual relations among concepts or instances can be also defined. For example, Bob hasFamilyRelation Alice sets that Bob and Alice are relatives. This way of inferring crispy information and defining crispy relations and properties, although useful in some domains, has limitations. The major one is the lack of mechanisms to manage uncertainty or providing beliefs about the inference results, which prevent its application to problems where their consideration is a must.

Contextual scene understanding

This section deals with the developed techniques for contextually recognizing objects and rooms. After an introduction, it discusses the related work that can be found in the literature, describes the datasets used as a testbed to evaluate such techniques, and concludes with the description of the contributions done in this regard.

3.1 Introduction

The ability to be aware of the objects and rooms in the robot surroundings, as well as of their types, is vital for a successful robot operation. Object/room recognition techniques are core components of semantic mapping systems, which are in charge of yielding the type of the spatial elements captured by the robot sensory system. As a consequence of this, a number of recognition approaches have been proposed for populating semantic maps.

Recognition methods often rely on RGB, and more recently on RGB-D information to perceive the robot environment and process the spatial elements therein. For that, the captured images are segmented into such spatial elements, which are individually processed in order to retrieve their type, *e.g.* counter, cabinet, microwave, kitchen, bathroom, etc., through a number of appearance and/or geometric features. The utilization of RGB and depth information entail a number of challenges as changing lighting conditions, cluttered room layouts, occlusions, or changing viewpoints, which can produce ambiguous recognition results. Recognition techniques also face other sources of uncertainty, like those coming from the own sensory system (*e.g.* sensor noise) or from the defined models. Given the effect that ambiguous recognition results stored in a semantic map may have on the robot operation (recall the lovely

grandma and the carpet), recognition techniques integrated into these systems have to tackle them.

In the works that conform this thesis a number of recognition techniques that address this uncertainty issues have been proposed, also striving to decrease the ambiguity of the recognition results by exploiting contextual relations. PGMs are employed for that, in close cooperation with KBs in the form of Ontologies in order to enhance their performance. These techniques are also able to provide a measure about the uncertainty of their results, which is crucial for the semantic mapping framework presented in the next chapter.

3.2 Related work

A vast literature exists around the recognition of objects and/or rooms. This section starts by briefly discussing traditional approaches addressing this issue, and the good reasons for contextually modeling these problems. Then, popular works exploiting context through PGMs are presented, as well as some alternatives exploring the utilization of Semantic Knowledge. Finally, the datasets applicable to the evaluation of the proposed recognition techniques are reviewed, as well as related software applications.

Traditional scene object/room recognition

Scene object recognition is a widely studied topic in computer vision and robotics. Recognition systems have traditionally relied on the features of the objects/room like their geometry or appearance due to their acceptable performance. Regarding object recognition, a popular example is the work by Viola and Jones [146], where an integral image representation is used to encode the appearance of a certain object category, and is exploited by a cascade classifier over a sliding window to detect the occurrences of such object type in intensity images. Another well known approach is the utilization of image descriptors, like Scale-Invariant Feature Transform (SIFT) [74], Speeded-Up Robust Features (SURF) [64], or Local Binary Pattern (LBP) [20], to capture the appearance of objects, and its posterior exploitation by classifiers like Supported Vector Machines (SVMs) [100] or Bag-of-Words [85, 52]. Other works study the automatic learning of low level features, *e.g.* using neuronal networks, as is the case of Bai *et al.* [8]. The work by Zhang *et al.* [157] provides a comprehensive review of methods following this approach.

On the other hand, a considerable number of works also tackle the room categorization problem through the exploitation of their geometry or appearance, like the one by Mozos *et al.* [80] which employs range data to classify spaces according to a set of geometric features. Also popular are works resorting to global descriptors of intensity images, like the *gist* of the scene proposed by Oliva and Torralba [91], those resorting to local descriptors like the aforementioned SIFT and SURF [6, 81], or the works combining both types of cues, global and local, pursuing a more robust performance [149, 101].

Despite the success of local recognition systems for certain applications, their integration into mobile robots arises a number of additional issues to be tackled [119, 93]. One of the most significant ones is the fact that they can lead to ambiguous recognitions, i.e. they are prone to fail in identifying classes with similar features, as analyzed in [92, 19, 38, 115]. This is mainly due to only relying on features of the objects/rooms themselves, disregarding valuable contextual information that is also available. Therefore, a significant, growing body of current research aiming to overcome this issue is considering contextual information of the scene objects in addition to their usually employed individual features. Some works have attempted to exploit this information by providing ad-hoc or preliminary solutions, like in [78], where the co-occurrence of objects appearing in distinct types of rooms are implicitly modeled. However, these works lack a consistent theoretical background, compromising, among others, their comparison, generalization, re-usability, or scalability. Moreover, their output consists of a set of objects' labels, which do not carry any semantic information profitable by high-level AI robotic components. Well grounded alternatives for modeling/exploiting contextual relations are *Probabilistic Graphical Models* and *Semantic Knowledge*, whose combination is exploited in this thesis with the goal of mitigating their drawbacks and boosting their virtues.

Contextual Recognition through PGMs

Probabilistic Graphical Models (PGMs) in general, and Undirected Graphical Models (UGMs) in particular, have become popular frameworks to model and exploit contextual relations in combination with probabilistic inference methods [65]. Contextual relations can be of different nature, involving objects and/or rooms. On the one hand, objects are not placed randomly within the robot workspace, but following configurations that make sense from a human point of view, e.g. carpets are on the floor, remote controls can be found close to televisions, and pillows are normally placed on beds. The earliest works using this information were based on intensity information of the scene, like [152], where the context between pixels in a given RGB image is modeled by a discriminative Conditional Random Field (CRF). Another work, also relying on intensity images, is the presented in [106] that proposes a CRF framework that incorporates hidden variables for part-based object recognition. The work in [79] also builds part-based models of objects, and represents their interrelations with a PGM. More recent is the work presented in [33] which employs stereo intensity images in a CRF formulation. Three-dimensional information from stereo enables the exploitation of meaningful geometric properties of objects and relations. However, stereo systems are unable to perform on surfaces/objects showing an uniform intensity, which can negatively affect the recognition performance.

With the emergence of inexpensive 3D sensors, like Kinect, a new batch of approaches have appeared leveraging the dense and relatively accurate data provided by these devices. For example, the work presented in [4] builds a model isomorphic to a Markov Random Field (MRF) according to the segmented regions from a scene point cloud and their relations. The authors did the tedious work of gathering information

from 24 office and 28 home environments, and manually labeled the different object classes. Interestingly, it is shown in [111] that the accuracy of a MRF in charge of assigning object classes to a set of superpixels increases as the amount of available training data augments. In [145] a meshed representation of the scene is built on the basis of a number of depth estimates, and a CRF is defined to classify mesh faces. CRFs are also used in [60] and [154], where Decision Tree Fields [87] and Regression Tree Fields [56] are studied as a source of potentials for the PGM. The CRF structure for representing the scenes in [154] is similar but less expressive than the one presented here. In that work, a CRF is used to classify the main components of a facility, namely clutters, walls, floors and ceilings.

On the other hand, object–room relations also supposes a useful source of information: objects are located in rooms according to their functionality, so the presence of an object of a certain type is a hint for the categorization of the room and, likewise, the category of a room is a good indicator of the object categories that can be found therein. Thus, recent works have explored the joint categorization of objects and rooms leveraging both, object–object and object–room contextual relations. CRFs have proven to be a suitable choice for modeling this holistic approach, as it has been shown in the works by Rogers and Christensen [113] or Lin *et al.* [73].

Despite their virtues, PGMs shows a number of drawbacks, like the necessity of large and comprehensive datasets for training, their high complexity when modeling real world problems, or their inability to detect incoherent results and learn from experience. The contributions in this section aim to mitigate those issues with the utilization of Semantic Knowledge.

Semantic Knowledge for modeling context

A different trend in the literature resorts to Semantic Knowledge for both recognizing objects and exploiting their contextual information. For example, the work described in Günter *et al.* [45] codifies contextual information in an Ontology, combined with a set of rules defined with the Semantic Web Rule Language [53], to generate objects' candidate classes. These hypotheses are subsequently validated through a matching process with CAD models. Another example is presented in Nüchter and Hertzberg [88], which defines a constraint network in Prolog to classify the main structural surfaces of buildings, i.e. walls, floors, ceiling and doors, using contextual relations like orthogonal, parallel, above, etc. In Galindo *et al.* [35], data codified into an Ontology about scene objects and their relations are used to infer new high-level information. The work introduced by Durand *et al.* [21] recognizes segmented regions that have been previously characterized through a set of features in RGB images. These features are defined in an Ontology, and their usual values for the different object types are learned by symbolic supervised machine learning tools. In this case, a specific procedure matches characterized regions with semantically defined concepts, but although the authors propose the use of contextual relations, they are neither defined nor exploited. An Ontology is also used in Maillot *et al.* [25] for the recognition of isolated objects and their subparts, which manually establishes the as-

sociation between geometric features and numeric values. This Ontology is populated through machine learning techniques like Perceptrons and Support Vector Machines.

A common characteristic of these approaches based on Semantic Knowledge is that they show limitations in quantifying the uncertainty of their results, and in exploiting the encoded contextual relations. The presented contributions face these issues through collaboration with a CRF, which provides the mobile robot with a recognition system endowed with a probabilistic inference mechanism, able to manage uncertainty and adequately exploit contextual relations.

Related software applications

Most contextual-based object recognition works rely on an ad-hoc implementations of both the PGMs framework and inference algorithms [4, 111, 145, 154]. This makes it difficult to conduct a fair comparison between state-of-the-art works, even when they report results resorting to the same dataset. There are some publicly available software libraries implementing this framework [89, 129], but they are not suited for the contextual object recognition problem (e.g. they only handle *chain-structured* models), or their applicability to this issue is limited. Regarding Semantic Knowledge related applications, there exist a number of mature software for codifying and managing this information in Ontologies, as is the case of Protégé [43] or Fluent Editor [17], as well as logical reasoners like Pellet [133], HerMiT [41], FaCT++ [143], or Racer [47].

Applicable RGB-D datasets

The irruption of proposals exploiting RGB-D information has been accompanied with public datasets that offer common benchmarking resources for comparing these works. Among them we can find Berkely-3D [57], Cornell-RGBD [5], NYUv1 [130], NYUv2 [131], TUW [3], SUN3D [153], or ViDRILo [75]. Specially popular are Cornell-RGBD, which is employed in several works aforementioned [4, 60, 54], and NYUv2 used in [151, 119, 116, 117]. The next section reports the datasets employed in this thesis.

3.3 Testbed

Three datasets containing RGB-D information have been used to assess the performance of the contributions in this chapter: UMA-Offices [119], NYUv2 [131] and Cornell-RGBD [5]. This section briefly describe the last two datasets, while details about UMA-Offices are provided in Section 3.4.1.

NYUv2 contains a total of 1,449 labeled pairs of both intensity and depth images, and has been extensively used in the literature (e.g. [151, 119, 116, 117]) due to its challenging, cluttered scenes from commercial and residential buildings. Although the number and type of objects and rooms we have considered differs from one work to other, typically 208 scenes corresponding to home facilities have been employed,

as well as 24 object categories appearing in such environments, e.g: bottle, cabinet, counter, faucet, floor, mirror, sink, toilet, towel, table, sofa, book, etc. It is worth to mention that the provided images only capture a portion of the scene, so the contained contextual relations are somehow limited. An evidence of this is given by the total number of extracted relations, 1,345, when compared with the number of objects, 1,295. This is an average of 6.25 objects and 6.47 relations per scene.

The Cornell-RGBD repository has 24 labeled office scenes and 28 home labeled scenes built from the registration of RGB-D images. As opposed to NYUv2, the provided data inspect a larger portion of the scene, resulting in a richer set of available contextual information. This feature has motivated its utilization in a variety of works (e.g [4, 60, 54]). As before, the home scenes have been selected, which sum up a total of 764 object instances and 2,911 contextual relations among them, averaging 27.29 objects and 103.96 relations per scene. We have used the same 17 categories as in the work that presented this dataset [4].

3.4 Contributions

This section describes the developed techniques for an object/room recognition framework through the synergy of PGMs and Semantic Knowledge. It starts with the description of the UMA-Offices dataset, specially collected for testing such techniques, and continues with an overview of the Undirected Probabilistic Graphical Models in C++ library, implemented for efficiently handling PGMs in robotic applications, as well as an analysis of PGM learning strategies. Then, a brief review of those techniques is provided, along with references to papers and online resources with further information.

3.4.1 UMA-Offices dataset

Office facilities are one of the typical application domains for mobile robots. To test the developed techniques in such environments, the UMA-Offices dataset, compounded of 25 office scenes from the University of Málaga, has been collected. Sensory data included in this dataset was acquired by Rhodon, a mobile robot endowed with an RGB-D device mounted on a Pan-Tilt unit, which permits it to perceive the world from a human-like point of view (see Figure 3.1-left). In this repository, the plane-based mapping algorithm by Fernandez-Moral *et al.* [31] was used to build a 3D representation of the scenes (see Figure 3.1-right), as well as to extract planar patches characterized through a number of features (e.g. size, orientation, position or contextual relations). In total, 170 object instances were labeled from the following categories: floor, wall, table top, table side, chair back rest, chair seat, and computer screen. Table 3.1 lists the features of this and the other two datasets used as testbeds.

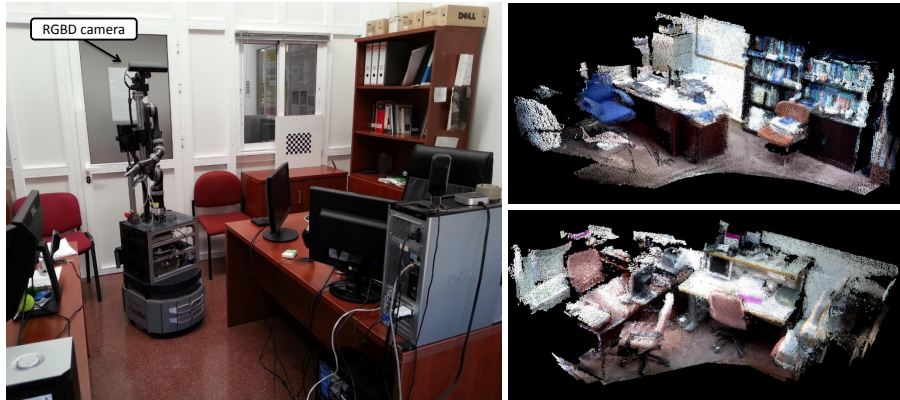


Figure 3.1: Left Rhodon robot from the MAPIR Group capturing RGB-D images from an office. Right, two point clouds from the UMA-Offices dataset.

Table 3.1: Principal characteristics of the three discussed datasets, UMA-Offices, NYUv2 and Cornell-RGBD.

Properties Dataset	UMA-Offices	NYUv2	Cornell-RGBD
#scenes	25	208	24
#obj. categories	7	24	17
#objects	170	1,345	764
#relations	305	1,295	2,911
mean #objects	6.8	6.25	27.29
mean #relations	12.2	6.47	103.96
type of objects	planar surfaces	arbitrary shapes	arbitrary shapes

3.4.2 The UPGMpp library

The study of the software used by state-of-the-art recognition methods employing CRFs arose the lack of public solutions especially focused and optimized for that goal. The utilization of efficient software is a must, since the computational resources in typical robotic platforms are limited given the different modules of the robotic architecture (navigation, localization, etc.) that compete for them.

For that reason, the Undirected Probabilistic Graphical Models in C++ library (UPGMpp, see Figure 3.2) has been developed as open-source¹ for the efficient building, training and managing of undirected PGMs. Its main features are:

- It works with discrete random variables.
- Handles first order (local or unary) and second order (pairwise) relations.

¹<http://mapir.isa.uma.es/work/upgmpp-library>



Figure 3.2: UPGMpp logo.

- Nodes (random variables) of different types can appear and interact in the same PGM (for example, nodes representing objects, rooms, facilities, etc.).
- If the value of a random variable is known, such an evidence can be considered.
- It supports PGMs with an arbitrary structure (including graphs with loops).

UPGMpp is fully implemented in C++, and resorts to the also open-source project libLBFGS [82] for performing numerical optimization, and to the Eigen library [44] for fast matrix operations. Boost library [128] is used to avoid unnecessary re-copy of data across the implemented methods by means of shared smart pointers. This library is also employed for serialization purposes, which adds the possibility of storing/loading graphs from/to files, enabling the long-term life of PGMs beyond execution time. Additionally, the Open Multi-Processing API (OpenMP) [94] was employed to speed-up the execution of a number of algorithms through parallelization techniques. Further implementation information and other details can be found in the work by Ruiz-Sarmiento *et al.* [115], which is included in this thesis.

The methods currently available for managing Undirected PGMs are:

Maximum a Posteriori (MAP) inference: Iterated Conditional Modes (ICM) [11], Greedy ICM, Exact Inference, Loopy Belief Propagation (LBP) [150], Tree Reparametrization Belief Propagation (TRBP) [148], Residual Belief Propagation (RBP) [24], α -expansions and α - β Swaps Graph Cuts [15].

Marginal inference: (sum-product) Loopy Belief Propagation [155], Tree Reparametrization Belief Propagation [148], Residual Belief Propagation [24].

Learning objective functions: Pseudo-likelihood [11], Score-matching [55], Piecewise-likelihood [136, 135], Marginal-based approximation [68], MAP-based approximation [65].

Learning optimization methods: Stochastic Gradient Descent (SGD) [83], quasi-Newton Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [86].

As a proof of the efficiency achieved by the library, and as reported in [115], different inference methods were executed on scenes from the NYUv2 dataset, which averages 6.25 objects and 6.47 contextual relations per scenario (see Table 3.1), reaching the ICM inference method a mean execution time of 0.46ms, the LBP one 2.16ms, and the α -expansions method 7.78ms.

3.4.3 Testing CRF learning approaches

The learning and probabilistic inference methods implemented in UPGMpp have been successfully applied to a variety of problems, however, their performance highly depends on the peculiarities of the application domain [68, 66, 96, 32]. A study of this for the scene object recognition result was missing in the literature, so this gap was covered through an empirical analysis of the most popular strategies. Concretely, two families of objective functions have been explored: pseudo-likelihood, and approximate inference algorithms, including Marginal and Maximum a Posteriori methods: sum-product and max-product LBP, ICM, and Graph-cuts. Two approaches for the optimization of such objectives are also considered: SGD, and L-BFGS.

As a testbed for the conducted analysis the indoor home scenes from the NYUv2 and Cornell-RGBD were employed, with particular features worth to explore: while NYUv2 comprises a high number of labeled images (we have used 208 from home environments) that capture the objects and relations from portions of scenes, Cornell-RGBD provides a lower number of scenes (28 from homes) but fully covering the inspected place, similarly to the contributed UMA-Offices dataset, which results in a considerably larger number of perceived objects and relations.

The conducted study focused on two facets of the learning methods: the recognition performance of the trained CRFs, and the required computational time. To measure the CRFs performance different MAP inference methods were executed over the learned models, and their recognition results compared with the ground-truth information provided by the datasets. The computational time needed by each learning method to converge was also analyzed, studying the advantage of parallelization techniques. Finally, the scalability of the learning methods according to different factors was also studied.

Briefly, the conducted study yielded the following conclusions, which greatly help in deciding the learning strategy to be chosen and the configuration according to the target application (for a complete conclusions' list, please refer to [121]):

- CRF models learned from Cornell-RGBD data were more prone to over-fit their parameters than those working with NYUv2. This is due to the higher complexity of the scenes from the Cornell-RGBD.
- The Marginal inference – SGD strategy yielded the highest recognition performance in both datasets: 79.85% in NYUv2 and 67.27% in Cornell-RGBD.
- The PL – L-BFGS strategy was the most robust, providing acceptable results in all the CRF configurations studied.
- LBP was the winning method for testing, reaching the best results when dealing with CRFs with edges and normalized features.
- In general, the computational time is reduced, ranging from the 24.43s. (on average) with the PL – SGD strategy, up to the 71.03s. with the Marginal inference – SGD one.

- L-BFGS and SGD benefited from parallelization techniques in OpenMP, achieving a speed-up factor of ~ 3.5 for PL – L-BFGS, and ~ 5 for Marginal inference – SGD using 8 CPU-cores.

Concerning the scalability of the studied strategies, it has been analyzed how the utilization of different number of training samples and object categories affect their performance. These experiments reported that the computational time required for learning scales considerably better in both cases when PL – L-BFGS was used, being its growth even sub-linear in some cases. Regarding recognition success, the Marginal inference – SGD option achieved the best outcome.

3.4.4 Exploiting Semantic Knowledge for CRF learning

PGMs in general, and CRFs in particular, need a vast amount of training data in order to reliably encode the gist of the domain at hand. However, the collection of that information is an arduous, time-consuming, and – in some domains – an intractable task that consists of moving the robot from one scene to another, gathering the data, and post-processing it accordingly to the type of information expected by the training algorithms. To face this issue, a framework to codify Semantic Knowledge through human elicitation in an Ontology has been developed, defining the domain object classes, their properties, and their relations. The result is used to generate an arbitrary number of training samples for tuning CRFs. These training samples reify prototypical scenarios where objects are represented by a set of geometric primitives, e.g., planar patches or bounding boxes, that fulfill certain geometric properties and relations, like proximity, difference of orientation, etc. This approach exhibit a number of advantages:

- It eliminates the usually complex and high resource-consuming task of collecting the large number of training samples required to tune an accurate and comprehensive model of the domain.
- Ontologies are compact and human-readable knowledge representations. In that way, extending the problem with additional object classes is just reduced to codify the knowledge about the new classes into the Ontology, generate synthetic samples considering the updated semantic information, and train the CRF. This process can be completed in a few minutes, in contrast to the time needed for gathering and processing real data.
- The recognized objects are anchored to semantically defined concepts, they hence can be straightforwardly incorporated to a semantic map for performing high-level tasks [36, 34, 18].

Thus, the proposed framework follows a top-down methodology (see Figure 3.3). The design starts with the definition of an Ontology for the knowledge domain at hand, e.g. an office environment, through human elicitation, stating the typical objects, their geometrical features, and relations. Then, the encoded Semantic Knowledge is used for generating sets of synthetic samples, which replace the real datasets

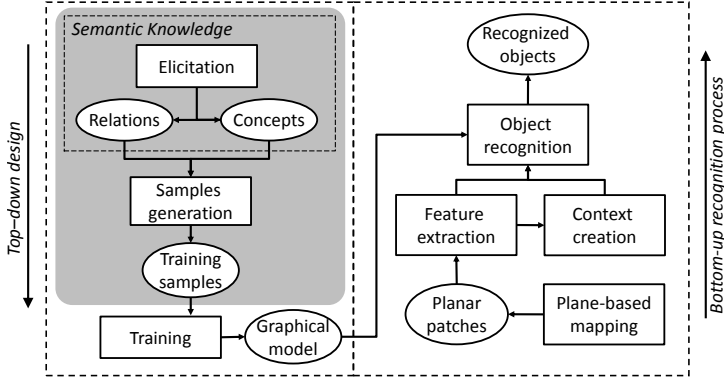


Figure 3.3: Overview of the developed framework for object recognition. The shadowed area delimits the proposed components for the generation of training samples. Boxes represent processes, whereas ovals are generated/consumed data (taken from [116]).

required for training through an algorithm that performs an arbitrary number of times the following steps:

1. **Inclusion of objects in the scene.** The set of objects that appears in the synthetic scene is selected according to their frequency of occurrence codified within the Ontology.
2. **Object characterization.** The geometrical features of the objects included in the previous step, *e.g.* area, centroid height, elongation, orientation, etc. are reified according to their concepts' definitions in the Ontology.
3. **Context creation.** The contextual relations between the included objects are established.
4. **Context characterization.** Different features of those relations are computed, adding valuable contextual information. Examples of these features are: difference between centroid heights, perpendicularity, difference between areas, areas ratio, difference between elongations, etc.

Once the CRF is trained (recall Section 2.1.2), it is integrated into an object recognition framework that works following a bottom-up stance (see Figure 3.3). During the robot operation, a plane-based mapping algorithm [31] extracts planar patches, which are characterized through a number of features, *e.g.*, size, orientation, position or contextual relations. These characterized planar patches feed a probabilistic inference process that yields the recognition results (recall Section 2.1.3).

The results obtained in the conducted evaluations achieved a recognition success of $\sim 90\%$ within the UMA-Offices dataset (see Figure 3.4), and of $\sim 81\%$ and 69.5% using office and home scenes from the NYUv2 dataset respectively, revealing that

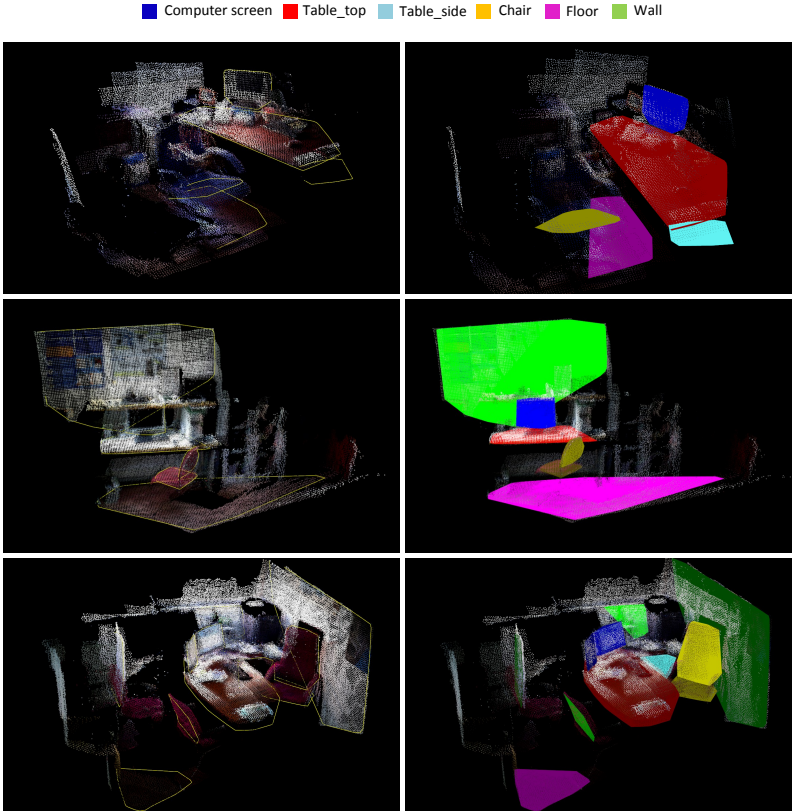


Figure 3.4: Examples of scene object recognitions performed by the proposed framework. Left column, observed scenes from th UMA-Offices dataset with the detected planar patches delimited by yellow lines. Right column, recognition results of such scenes (see [116]).

Semantic Knowledge can be exploited for the suitable training of recognition systems. This approach was also compared with other state-of-the-art approaches based on CRFs, like [154], yielding a substantial improvement.

A number of additional, related issues were also addressed:

- The discriminant capability of different sets of contextual features was studied, showing their positive effect on the system performance.
- The relation between the size of the training datasets and the system performance was analyzed, obtaining the expected conclusions [111]: the larger and the more comprehensive the dataset is, the better the system outcomes are.
- It was also reckoned the computational efficiency, evidencing the suitability of the proposed system for real time robotic applications.

- It was analyzed the time saving gained with the use of human elicitation plus synthetic samples generation processes, resulting 20 times lower than the time spent in collecting real data from the UMA-Offices dataset.

Please refer to [116] for further information about the developed framework, its evaluation, and the reached conclusions.

3.4.5 Including rooms into the equation

The spatial awareness needed by the robot to accomplish high-level tasks must account for the existing close relations among not only objects, but also their typical locations. Thus, the robot should not only tackle the object recognition problem, but also the room recognition one, i.e. to infer the type of space where it is.

Recent publications (e.g. [73, 113]) have shown that the joint modeling of these problems can outperform other methods that address them separately [28, 16, 90, 107, 105]. Holistic approaches exploit the fact that objects are located in rooms according to their functionality, so the presence of an object of a certain type is a hint for the recognition of the room [147, 102, 26]. Likewise, the category of a room is a good indicator of the object types that can be found inside [142]. Besides, objects are not placed randomly, but following configurations that make sense from a human perspective [114, 4, 154]. Thereby, the exploitation of these object-object and object-room contextual clues provides recognition methods with useful information.

For leveraging this information, the framework presented in the previous section has been extended to also consider rooms, recognizing them through the exploitation of their contextual relations. For that, Semantic Knowledge about rooms was codified into the Ontology through human elicitation (see Figure 3.5-top). Figure 3.5-bottom shows the definition of the concept *Microwave* within such Ontology, where we can see, for example, that their orientation is usually horizontal, or that they can be found in kitchens. This Ontology and other resources are available online at: <http://mapir.isa.uma.es/work/objects-rooms-categorization>.

The CRFs employed were also modified in order to consider random variables of different types, e.g. taking values from different object types, or from a set of room types, as well as contextual relations of different nature: object-object and object-room relations.

Thereby, two new steps were added to the four-steps algorithm described in the previous section to also generate room-related data. Concretely the new algorithm is:

1. **Room characterization.** The first step is the computation of the room features which, in the used Ontology, includes its volume (m^3) and color hue variation.
- 2-5. The same four steps as in the original algorithm, but taking into account the type of the room being synthetically generated.
6. **Object-room context characterization.** The relation between the room and its objects is characterized by a fixed value, as it is the training process of the CRF which learns automatically the likelihood of finding an object of a certain

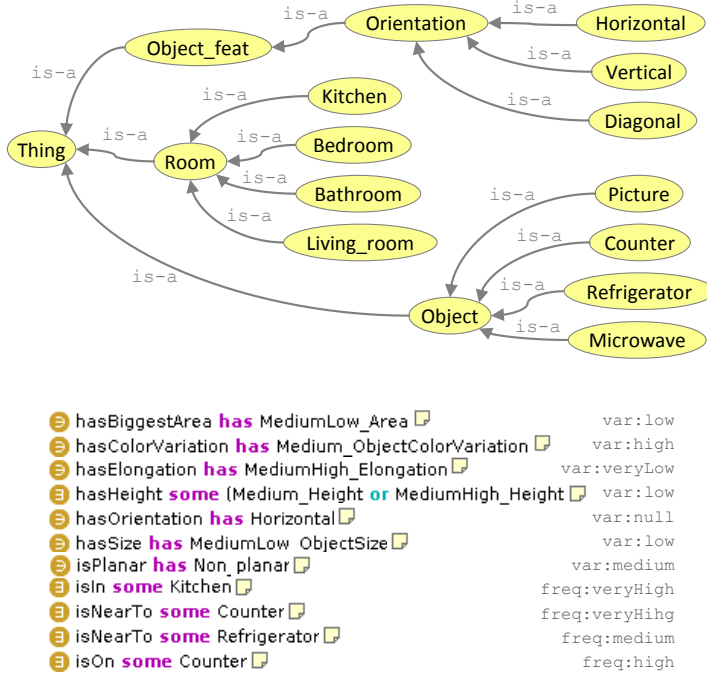


Figure 3.5: Top, excerpt of the Ontology used for the codification of Semantic Knowledge about the home domain. Bottom, definition of the concept Microwave.

type into a kitchen. Notice that the appearance of an object of a certain type in the room depends on previous steps.

In summary, the above six steps yield the objects, room and contextual features needed to feed the unary and pairwise factors during the training of the CRF. The avid reader can find more information about this process in [117].

The approach has been validated against home scenes from the NYUv2 dataset, reaching a categorization success of $\sim 70\%$ for both objects and rooms. The work by Lin *et al.* [73] also employs CRFs and NYUv2 for validation, and although a fair comparison is not possible since the authors consider a different set of object categories and room types, it permits us to qualitatively confirm the promising performance of the proposed approach, since they achieve a success of $\sim 60.5\%$ and $\sim 58.7\%$ recognizing objects and rooms respectively.

It is worth to mention that the applicability of the framework is not limited to robots working at home environments, but it is suitable to perform in other domains which properties and semantics can be defined by human elicitation, e.g. office facilities or hospitals.

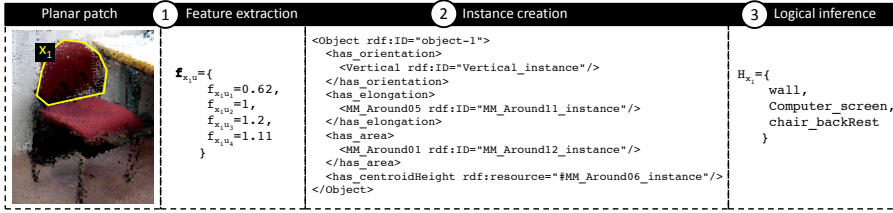


Figure 3.6: Example of hypotheses generation for a given region. New instances are inserted into the Ontology using the OWL language.

3.4.6 Further enhancing CRFs performance: coherence and efficiency

Approximate inference methods executed over CRFs are able to handle complex models and operate in impressive short times, at the expense of a (hopefully) tiny sacrifice in terms of recognition success. Obviously, the utilization of exact inference algorithms is preferable, but the complexity of real models prevents their use. This contribution proposes the exploitation of the Semantic Knowledge encoded in an Ontology to reduce the CRF inference complexity.

Concretely, the Semantic Knowledge is used to generate hypotheses about the most probable belonging classes of the objects according to their features. For example, a horizontal surface with a medium height from the floor could be hypothesized as belonging to the *Chair_seat*, *Table* or *Counter* concepts, but not to *Wall* or *Computer_screen*. These hypotheses are then taken by the CRF as the only possible candidates. This leads to a considerable reduction in the number of combinations, *i.e.* assignments to the random variables, hence decreasing the inference complexity and even enabling, in some cases, exact inference. Moreover, the generation of these hypothesis ensures that the results will be coherent with the information in the Ontology, and consequently, with the Semantic Knowledge that the human encoded about the domain.

The process shown in Figure 3.6 help us to illustrate how hypotheses are generated. First, the object (in this case a chair back) is characterized through a number of features, and a new instance derived from the *Object* concept is inserted into the Ontology, *e.g.* *object-1*, also including a number of properties, or relations, stating such features, *e.g.* *object-1 hasCentroidHeight MM_Around06*. This information is encoded in the Ontology employing the OWL language [10]. Then, a logical reasoner, Pellet [133] in this case, infers a set of concepts that are consistent with the instance definition: *Wall*, *Computer_screen* and *Chair_backRest* in the example. In this way, the CRF only considers that concepts as possible categories for that object, hence decreasing the problem complexity.

Additionally, prior information about the frequency of occurrence of the different object types was also encoded into the Ontology. This type of information permits us

to model that, for example, it is more likely to find a computer than a couch in an office environment, while it is quite unlikely to find an ironing table. This new source of information comes together with a modification to the usual CRF formulation, which can be checked in [114, 119], so it is able to exploit this prior information from the Ontology. This approach enhances even more the expected coherence of the recognition results.

The claimed virtues of these contributions have been thoroughly validated considering the NYUv2 and UMA-Offices datasets. Regarding the recognition success, the evaluation provided the performance of a local object recognition approach as a baseline, which was of $\sim 79\%$ and $\sim 54\%$ for UMA-Offices and NYUv2 respectively, and revealed the progressive increment in the performance and robustness as long as additional information is exploited: contextual information ($\sim 84\%$ and $\sim 59\%$), hypotheses of objects' types ($\sim 93\%$ and $\sim 61\%$), and prior information about object category occurrences ($\sim 94\%$ and $\sim 65\%$).

Moreover, an analysis of the complexity reduction of the probabilistic inference process was carried out by considering the most promising object belonging types, including the feasibility of exact inference for the considered datasets. The yielded results are promising, allowing the system to rely on exact inference in all the scenarios within the UMA-Offices dataset, and in a wider variety of them in NYUv2. Further details in this regard can be found in [114, 119].

3.4.7 Learning from experience

Typically, mobile robots employ CRFs that are pre-tuned with a certain dataset in order to recognize a fixed range of object categories. However, this configuration lacks of the flexibility demanded by robots performing in human-like environments, e.g. it is (of course) unable to recognize new types of objects not appearing in the training dataset, or instances of learned ones showing peculiar features, which can lead to an incoherent performance [93]. This section proposes a recognition framework that relies on (surprise) Semantic Knowledge to detect and learn from incoherent recognition results yielded by inference over a CRF.

For example, it can be defined the concept `Fridge` codifying that they are usually high, box-shaped objects, and the `Pill_box` one, stating that they are small boxes related to fridges by `Pill_box placedInto Fridge`. In the proposed framework, the recognition results yielded by probabilistic inference over the CRF are checked for coherence against the Semantic Knowledge. If any of them is detected as incoherent (for example, a middle-size object is classified as a fridge), then it is annotated for its posterior evaluation by the user through a simple dialog. This human-robot interaction is greatly supported by the Ontology, since its content can be verbalized in a straightforward way. Finally, the feedback from the user is back-propagated in order to tune the CRF and the own Ontology accordingly. It is worth to mention that Ontologies also suppose a basic way to understand the robot workspace, enabling the detection of object configurations that can be hazardous, e.g. the pill box found out of the fridge.

More concretely, the recognition pipeline of the recognition framework starts by capturing an image of the scene to be processed to build its CRF graph representation. This graph, along with the pre-trained CRF parameters, is exploited by a probabilistic inference algorithm to provide a set of tentative object recognition results. These results are then inserted as instances in the Ontology, which checks their consistency with respect to the codified Semantic Knowledge by employing a logical reasoner (Pellet). This permits the robot to detect incoherent results that are subsequently evaluated by the user. The evaluation of a conflicting object starts by showing him/her a cropped image of it. Three different scenarios are then possible:

Case 1: the user determines that the recognition result is right. This means that the CRF performed correctly, but the codified common-sense knowledge was somehow too strict. The Ontology learns from this outcome by relaxing the codified object property that produced the inconsistency.

Case 2: the recognition result is wrong, and:

Case 2.1: the object type is already present in the CRF/Ontology. In this case the CRF misclassified the object. To learn from the mistake, the gathered object information is used to re-tune the CRF parameters.

Case 2.2: the object type is new. The relevant information from the object is used to automatically generate a new concept in the Ontology, and the CRF is also re-trained taking into account this new object type.

To perform a proof-of-concept validation of the framework, a robot was deployed into an apartment and commanded to perform a primary task: to check the configuration of the objects in the kitchen. Concretely, during the robot operation, the RGB-D camera was used to capture both intensity and depth images when reaching certain locations in the kitchen. In that setup, the robot detected an inconsistency, which corresponded to a pill box recognized as a cereal box, since such object type was unknown for the robot. This information was then back-propagated to both: (i) the Ontology, where the system created a new concept `Pill_box`, inheriting from the `Object` one, and described it with the information gathered from the human and from the collected sensory data, and (ii) to the CRF model, which re-tuned its parameters according to the new information. The learning success was evaluated in later observations of pill boxes, where the robot was able to successfully recognize this new type of object.

3.5 Discussion

This chapter has described the thesis' contributions to the contextual object and/or room recognition problem. It started with the *UMA-Offices* dataset, a collection of 3D reconstructions of offices from the University of Málaga, which was necessary for evaluating the developed algorithms. Then, the *Undirected Probabilistic Graphical Models in C++* (UPGMpp) library has been presented, which permits the efficient handling of Undirected PGMs when applied to robotic-related applications.

PGMs in general, and CRFs in particular, have proven to be valuable frameworks for the modeling of recognition problems exploiting contextual information, also dealing with uncertainty. The effort needed for the collection and processing of *UMA-Offices* and other sensory data, along with the hungry for comprehensive and large training datasets exhibited by the learning phase of PGMs, motivated the study of alternative training strategies. This led to the utilization of Semantic Knowledge stored within an Ontology to remove the necessity of a real dataset. This is specially useful in domains where it is difficult, or even infeasible the collection of large amounts of data. Ontologies also provide the recognition system with an structured, human readable representation ready-to-use for high-level robotic tasks.

Semantic knowledge has been further exploited for reducing the complexity of the probabilistic inference processes over the CRFs, as well as to provide prior knowledge about the frequency of occurrence of the object classes of the domain at hand. This information is incorporated into the usual CRF formulation in order to enhance its performance. It has been also leveraged for detecting incoherent recognition results, by considering a logical reasoner that checks the consistency of the CRF outcome with respect to the encoded knowledge. This also allows the recognition system, including an user in the loop (supervised learning), to learn from experience by automatically adapting its internal representations.

These contributions make up a probabilistic recognition system which is able to: (i) exploit contextual relations, (ii) handle uncertainty, (iii) leverage prior knowledge about the domain at hand, (iv) detect incoherent results, (v) learn from experience, and (vi) verbalize its outcome. In addition to these features, the system can also provide a measure about the uncertainty of its results. Finally, the system has been integrated into a semantic mapping framework specially suited for taking advantage of these features, as shown in the next chapter.

Semantic Mapping

This chapter outlines the thesis's contributions to the semantic mapping field. It starts with a brief introduction to the problem and a discussion of relevant works in the literature. Then, it describes: a toolkit for labeling sequential RGB-D datasets, the Robot@Home repository processed by that toolkit, and finally the Multiversal Semantic Map, a novel representation evaluated through Robot@Home.

4.1 Introduction

Despite the possibilities of geometric and/or topological maps when applied to mobile robot applications, the planning and execution of high-level tasks like “bring me the red cup from the kitchen’s counter” or “show the customer off-season clothing, specially pants, please” demands more sophisticated maps. Humans share semantic knowledge about concepts like *red*, *cup*, or *off-season clothing*, which must be transferred to robots in order to successfully face these tasks. *Semantic maps* emerged to cope with this need, providing the robot with the capability to *understand*: (i) the spatial aspects of human environments, (ii) the meaning of their elements (objects, rooms, or facilities), and (iii) how humans interact with them (*e.g.* functionalities, events, or relations).

This feature is distinctive and traversal to semantic maps, being the key difference with respect to maps that simply augment metric/topological models with labels to state the type of recognized objects or rooms [108, 22, 76, 127], *e.g.* saying that a portion of sensory data is a cup, without any other information about the *implications* of that. Contrary, semantic maps handle meta-information that models the properties

and relations of relevant concepts therein the domain at hand, codified into a *Knowledge Base* (KB) and stating that, for example, cups are cylindrical-shaped objects usually found in kitchens and useful for containing liquids. Building and maintaining semantic maps involve the symbol grounding problem [49, 18], *i.e.* linking portions of the sensory data gathered by the robot (percepts), represented by symbols (*e.g.* object-1 or room-1), to concepts in the KB by means of some recognition and tracking method. These representations usually reckon on off-the-shelf recognition methods to individually ground percepts to particular concepts, which disregard the valuable contextual relations between the workspace elements: a rich source of information intrinsic to human-made environments (for example that night-stands are usually in bedrooms and close to beds).

Semantic maps generally support the execution of reasoning engines, providing the robot with inference capabilities for efficient navigation, object search, or proactiveness [36], among others. Typically, such engines are based on logical reasoners that work with crispy information (*e.g.* a percept is identified as a cup or not). The information encoded in the KB, along with that inferred by logical reasoners, is then available for a task planning algorithm dealing with this type of knowledge and orchestrating the aforementioned tasks [35]. Although crispy knowledge-based semantic maps can be suitable in some setups, especially in small and controlled scenarios [156], they are also affected by uncertainty coming from different sources like the robot sensory system or the inaccurate modeling of the elements within the robot workspace.

This chapter presents the contributions done for achieving a semantic map representation able to deal with uncertainty, also managing contextual relations, where the techniques outlined in Chapter 3 play a pivotal role (Section 4.3.3). In addition, given the lack of datasets for evaluating mapping systems with those features, we also describe a repository of information especially collected for that goal, the *Robot@Home* dataset (Section 4.3.2), as well as a toolkit developed for the efficient processing of this type of repositories, the *Object Labeling Toolkit* (Section 4.3.1).

4.2 Related work

This section reviews the most relevant works addressing some issues related to the semantic mapping problem, starting with a discussion about popular semantic representations (Section 4.2), continuing with an analysis of the datasets that are suitable as a testbed for such approaches (Section 4.2), and finishing with a discussion on available tools for managing datasets (Section 4.2).

Semantic mapping approaches

In the last decade, a number of works have appeared in the literature contributing different semantic map representations. One of the earliest works in this regard is the one by Galindo *et al.* [37], where a multi-hierarchical representation models, on the one hand, the concepts of the domain of discourse through an ontology, and on the

other hand, the elements from the current workspace in the form of a spatial hierarchy that ranges from sensory data to abstract symbols. NeoClassic is the chosen system for knowledge representation and reasoning through Description Logics (DL), while the employed recognition system is limited to the classification of simple shape primitives, like boxes or cylinders, as furniture, e.g. a red box represents a couch. The potential of this representation was further explored in posterior works, e.g. for improving the capabilities and efficiency of task planners [35], or for the autonomous generation of robot goals [36]. A similar approach is proposed in Zender *et al.* [156], where the multi-hierarchical representation is replaced by a single hierarchy ranging from sensor-based maps to a conceptual abstraction, which is encoded in a Web Ontology Language (OWL)–DL ontology defining an office domain. To categorize objects, they rely on a SIFT-based approach, while rooms are grounded according to the objects detected therein. In Nüchter and Hertzberg [88] a constraint network implemented in Prolog is used to both codify the properties and relations among the different planar surfaces in a building (wall, floor, ceiling, and door) and classify them, while two different approaches are considered for object recognition: a SVM-based classifier relying on contour-based features, and a Viola and Jones cascade of classifiers reckoning on range and reflectance data.

These works set out a clear road for the utilization of ontologies to codify semantic knowledge, which has been further explored in more recent research. An example of this is the work by Tenorth *et al.* [138], which presents a system for the acquisition, representation, and use of semantic maps called KnowRob-Map, where Bayesian Logic Networks are used to predict the location of objects according to their usual relations. The system is implemented in SWI-Prolog, and the robot's knowledge is represented in an OWL-DL ontology. In this case, the recognition algorithm classifies planar surfaces in kitchen environments as tables, cupboards, drawers, ovens and dishwashers [127]. The same map type and recognition method is employed in Pangercic *et al.* [95], where the authors focus on the codification of object features and functionalities relevant to the robot operation in such environments. The paper by Riazuelo *et al.* [112] describes the RoboEarth cloud semantic mapping which also uses an ontology for codifying concepts and relations, and rely on a Simultaneous Localization and Mapping (SLAM) algorithm for representing the scene geometry and object locations. The recognition method resorts to SURF features, and performs by only considering the object types that are probable to appear in a given scene (the room type is known beforehand). In Günther *et al.* [45], the authors employ an OWL-DL ontology in combination with rules defined in the Semantic Web Rule Language (SWRL) to categorize planar surfaces.

It has been also explored the utilization of humans for assisting during the semantic map building process through a situated dialog. Examples of works addressing this are those by Bastianelli *et al.* [9], Gemignani *et al.* [40], or the aforementioned one by Zender *et al.* [156]. The main motivation of these works is to avoid the utilization of recognition algorithms, given the numerous challenges that they have to face. However, they themselves argue that the more critical improvement of their proposals would arise from a tighter interaction with cutting-edge recognition techniques.

The interested reader can refer to the survey by Kostavelis and Gasteratos [67] for an additional, comprehensive review of semantic mapping approaches for robotic tasks.

The semantic mapping techniques discussed so far rely on crispy categorizations of the perceived spatial elements, *e.g.* an object is a cereal box or not, a room is a kitchen or not, etc., which is typically exploited by (logical) reasoners and planners for performing a variety of robotic tasks. As commented before, these approaches: (i) can lead to an incoherent robot operation due to ambiguous recognition results, and (ii) exhibit limitations to fully exploit the contextual relations among spatial elements. The contributions in the previous chapter propose a solution for probabilistic symbol recognition to cope with both, the uncertainty inherent to the recognition process, and the contextual relations among spatial elements. Perhaps the closest work to this approach addressing semantic mapping is the one by Pronobis and Jensfelt [103], which employs a Chain Graph (a graphical model mixing directed and undirected relations) to model the grounding problem from a probabilistic stance, but that fails at fully exploiting contextual relations. This thesis contributes, among others, a novel representation called Multiversal Semantic Map (*MvSmap*), in order to accommodate and further exploit the outcome of the probabilistic symbol grounding.

Suitable datasets

Datasets containing sensory data are needed for a thorough evaluation of semantic mapping techniques, since they set a common framework for their fair comparison. Mobile robots have traditionally resorted to intensity images to categorize objects and/or rooms, which motivated the collection of datasets providing this kind of information [27, 125, 124]. Nowadays, the tendency is for the datasets to also include depth information [57, 5, 72], given the proved benefits of exploiting morphological and spatial information in assisting recognition methods [114]. These datasets can be roughly classified as: *object-centric*, *view-centric*, and *place-centric*.

Object-centric datasets, like ACCV [51], RGBD Dataset [72, 71], KIT object models [62], or BigBIRD [132], provide RGB-D observations in which a unique object spans over each image. The exploitation of these images for robotic recognition exhibits some drawbacks: (i) they are not representative of the typical images gathered by a robot at a real environment, (ii) they prevent the utilization of valuable contextual information of objects, and (iii) they are not suitable for the room recognition problem. These shortcomings also narrow their utilization by semantic mapping benchmarks.

On the other hand, *view-centric* datasets as Berkeley-3D [57], Cornell-RGBD [5], NYU [130, 131], TUW [3], or UBC VRS [77], consist of isolated RGB-D images, or a sequence of them, which cover a partial view of the working environment. This information permits the exploitation of contextual information but only from a local, reduced perspective, since information of the entire scene is not collected. Therefore, their use for contextual recognition is still limited, as well as their utilization for semantic mapping purposes.

Table 4.1: Summary of related datasets (CR: Collected by a robot, DT: Dataset type, EOC: Enables object context exploitation, ERC: Enables room categorization).

Dataset	CR	DT	EOC	ERC
ACCV [51]		<i>object-centric</i>		
Berkeley-3D [57]		<i>view-centric</i>	✓ (local)	✓ (limited)
BigBIRD [132]		<i>object-centric</i>		
Cornell-RGBD [5]	✓	<i>view-centric</i>	✓ (local)	✓ (limited)
KIT object models [62]		<i>object-centric</i>		
Multi-sensor 3D Object Dataset [39]		<i>object-centric</i>		
NYUv1 [130]		<i>view-centric</i>	✓ (local)	✓ (limited)
NYUv2 [131]		<i>view-centric</i>	✓ (local)	✓ (limited)
RGBD Dataset [72]		<i>object-centric</i>		
RGBD Dataset 2 [71]		<i>object-centric</i>		
TUW [3]	✓	<i>view-centric</i>	✓ (local)	✓ (limited)
SUN3D [153]		<i>place-centric</i>	✓	✓
UBC VRS [77]	✓	<i>view-centric</i>	✓ (local)	
Robot@Home	✓	<i>place-centric</i>	✓	✓

Finally, *place-centric* datasets like SUN3D [153] provide comprehensive information from the inspected room, or even the entire work environment, typically through the registration of RGB-D images. This type of datasets conforms the best option as a testbed for semantic mapping taking advantage of both depth and contextual information, albeit, unfortunately their number is quite limited. A dataset worth to mention at this point is ViDRIO [75], which comprises 5 sequences of RGB-D observations of two office buildings collected by a robot combining *object* and *environment-centric* perspectives. This dataset annotates each observation with its room type and the objects found within it, although this labeling is not per-pixel and the number of object categories is reduced. Table 4.1 shows a summary of datasets applicable to the semantic problem and their characteristics, which also includes the one contributed by this thesis: the *place-centric* Robot@Home dataset.

Available dataset management tools

The tedious object labeling task within RGB-D datasets is carried out in different ways. Some works resort to *Amazon Mechanical Turk* (AMT) to label their intensity images [57, 130, 131], usually through a labeling tool like LabelMe [125], but this merely divides the workload, and the annotated information still needs to be thoroughly checked to fix incoherent labels. Another approach is the manual labeling of *key intensity frames* from a sequence, propagating these labels to the remaining RGB-D observations [77, 153], but this is only suitable for sequences with simple sensor trajectories, and additionally shows the same limitations as the AMT option. There are also works that reconstruct a 3D representation of the inspected scene and annotate the objects appearing on it [5], but there is not a *labeling feedback* to the RGB-D



Figure 4.1: OLT logo.

observations' sequence(s). In the works by Lai *et al.* [72, 71] the ground truth annotations over a reconstructed scene are also propagated to the individual RGB-D observations employing an ad-hoc software which, to the best of the author knowledge, is not publicly available. In the next section it is described an open source solution conveniently divided into configurable components, which provides the robotic community with a number of functionalities towards an efficient labeling of arbitrarily large collections of RGB-D data.

4.3 Contributions

Three contributions are outlined in this chapter, all of them in the scope of the semantic mapping problem. First, the Object Labeling Toolkit (OLT) is described. It consists of a set of software solutions for the labeling of sequential RGB-D datasets, especially relevant to semantic mapping. Then, we describe a novel *place-centric* dataset, named Robot@Home, which contains raw and processed data from domestic settings compiled by a mobile robot. Finally, the *Multiversal Semantic Map* is presented, an environment representation able to handle uncertainty and contextual relations, in which the contributions of the previous chapter are integrated.

4.3.1 The Object Labeling Toolkit

A comprehensive dataset is a valuable benchmark tool for tuning, testing, and comparing robotic algorithms and systems in a convenient and fair way. Although public datasets consisting of intensity images [27, 125, 124] have largely helped researchers to push ahead the state-of-the-art in object recognition or scene interpretation, nowadays new particularly oriented datasets are required given the increasing number of capabilities and applications that are demanded to a mobile robot, e.g. semantic mapping [104], high-level decision making [36], or contextual object recognition [116, 114, 115, 119].

RGB-D cameras have become a key source of information for such *robotic* datasets. Although the sensory data of these datasets may be conveniently gathered by the mobile robot itself, human supervision is still needed to segment objects and to

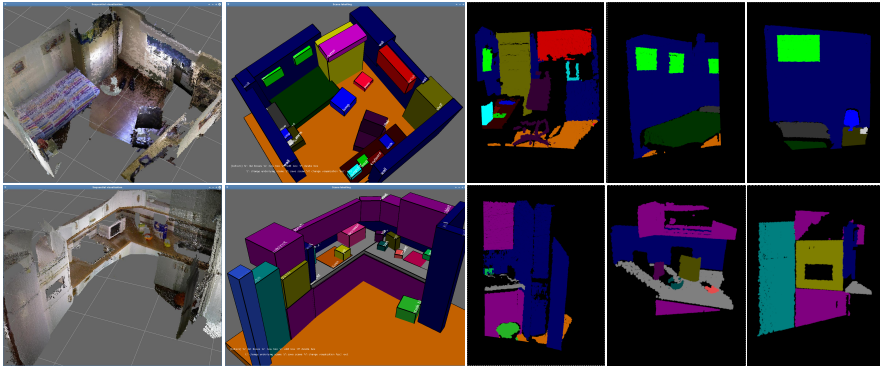


Figure 4.2: First column, reconstructed scenes from two RGB-D sequences. Second column, labeled reconstructed scenes. Third-fifth columns, examples of individual point clouds from RGB-D observations labeled by the propagation of the annotations within the reconstructed scenes.

label them, i.e. to add annotations over portions of the observed data as belonging to a certain object class, e.g. floor, table, lamp, etc. This is the motivation for the development of the *Object Labeling Toolkit* (OLT, see Figure 4.1), i.e. to provide the robotic community with a tool to efficiently label datasets compound of sequences of RGB-D observations, gathered from an arbitrary number of RGB-D sensors. OLT is publicly available under a GNU General Public License at: <http://mapir.isa.uma.es/work/object-labeling-toolkit>.

For achieving such efficient labeling, the toolkit builds a 3D reconstruction of each RGB-D sequence within a given dataset, and allows the user to graphically label objects within that reconstruction (see the two first columns in Figure 4.2). Then, this ground truth annotations are automatically propagated to all the RGB-D observations without requiring human supervision, resulting in a dense labeling of both intensity and depth data (see the three last columns of Figure 4.2). More information about this pipeline can be found in the publication by Ruiz-Sarmiento *et al.* [118].

OLT comprises a number of software components covering the following functionality: i) dataset pre-processing, ii) localization of RGB-D observation poses, iii) 3D scene reconstruction, iv) labeling of the reconstructed scene, and v) automatic propagation of annotated labels. Some of these functionalities can exploit additional information coming from sensors usually present in a robotic platform, e.g. the robot pose estimation computed from 2D laser scans. All the components are highly customizable in order to fit the particularities of robotic datasets, and can be easily expandable to integrate other algorithms of interest. The toolkit resorts to the Mobile Robot Programming Toolkit (MRPT [58]) and the Point Cloud Library (PCL [126]) for point cloud registration and smoothing algorithms, and for data representation and vi-

sualization purposes. The most time-consuming components of OLT have been also parallelized employing OpenMP.

Aiming to illustrate the toolkit suitability, it was utilized for segmenting and labeling a robotic dataset from a home environment (indeed, a part of the Robot@Home dataset, see Figure 4.2) consisting of 77 RGB-D observations. Regarding the time spent in labeling, the human operator needed 2 hours to annotate both the kitchen and the bedroom scenes, spending on average 2 minutes per object (this has been reduced to 1 minute in the last toolkit version). To compare this with the labeling of all the RGB-D observations individually, it was followed the typical intensity image labeling approach and they were annotated 5 non-consecutive observations from each sequence, extrapolating the results to the whole dataset. This yielded a total of ~ 3 hours needed for the labeling of the kitchen sequence, and ~ 7 hours for the bedroom, which clearly illustrated the benefits of the toolkit utilization. When following such a typical approach problems appeared to accurately label the objects' boundaries, and with objects partially occluded and/or with an unclear belonging class, drawbacks that are mitigated with the utilization of the proposed toolkit.

4.3.2 Robot@Home dataset

The Robot-at-Home (Robot@Home) dataset, is a collection of raw and processed data from five domestic settings compiled by the commercial mobile robot Giraff, equipped with 4 RGB-D cameras and a 2D laser scanner. Its main purpose is to serve as a testbed for semantic mapping algorithms through the recognition of objects and/or rooms, so it is publicly available at <http://mapir.isa.uma.es/work/robot-at-home-dataset>. This dataset is unique in three aspects: (i) the sensory system employed for its gathering, (ii) the diversity and amount of provided data, and (iii) the availability of dense ground truth information.

The provided data were captured with a rig of 4 RGB-D sensors with an overall field of view of 180° horizontally and 58° vertically, and with a 2D laser scanner (see Fig. 4.3). In order to yield accurate information within the dataset, the sensors mounted on the robot were calibrated both intrinsically and extrinsically [30, 42, 137]. Detailed information concerning this calibration in particular, and about the dataset in general, can be found in the paper by Ruiz-Sarmiento *et al.* [123].

This robotic platform was employed to explore 5 dwelling apartments, which have been named as *anto*, *alma*, *pare*, *rx2*, and *sarmis*. In this way, a total of 36 rooms were completely inspected (some of them several times), so the dataset is rich in contextual information of objects and rooms. This is a valuable feature, missing in most of the state-of-the-art datasets, which can be exploited by, for instance, semantic mapping systems that leverage relationships like *pillows are usually on beds* or *ovens are not in bathrooms*. This information was processed by OLT, which also supposes a mechanism to conveniently access and manage the data.

The ground-truth information provided by OLT comes in two flavors. On the one hand, it is provided (per-point) annotations of the categories of the main objects and rooms appearing in the scenes reconstructed from the RGB-D sequences (recall the



Figure 4.3: Giraff robot while collecting sensory information. The basic robotic platform was endowed with a rig of 4 RGB-D sensors mounted on the *robot's neck*, and a 2D laser scanner on its base.

second column of Figure 4.2). A total of $\sim 1,900$ objects belonging to 157 different categories were manually labeled from the 36 visited rooms. These rooms are also labeled as belonging to one of 8 possible types: bathroom, bedroom, kitchen, living-room, etc. On the other hand, Robot@Home also includes (per-pixel) annotations of the objects appearing in the 69,000+ gathered RGB-D images. The objects and rooms are also annotated with identifiers, so they can be individually tracked along the video sequences.

Summarizing, the content of the dataset, which comes in different formats accessible by the open source Mobile Robot Programming Toolkit¹ (MRPT), as well as in (human readable) plain text files and PNG images, is as follows:

- **81** sequences of observations containing ~ 75 min. of recorded data. The total number of observations is **87,000+** (18,000+ laser scans and 69,000+ RGB-D images), which are saved in *rawlog* format as well as in plain text (see the three first rows of Figure 4.4).
- **41** 2D geometric maps saved in text files (36 for individual rooms, and 5 maps covering each apartment, see fourth row of Figure 4.4).
- **72** 3D reconstructed scenes in *scene* format and plain text (see fifth row of Figure 4.4).
- **72** Labeled 3D reconstructed scenes in *scene* format and plain text, containing $\sim 1,900$ labeled objects (see sixth row of Figure 4.4).
- **72** Labeled RGB-D sequences in *rawlog* format and plain text (see seventh row of Figure 4.4).

¹<http://www.mrpt.org>



Figure 4.4: Excerpts of information provided by Robot@Home. From top to bottom, examples of 2D laser scans, RGB images, depth images, 2D geometric maps, reconstructed rooms, labeled reconstructed rooms, and labeled depth information. *Taken from [123].*

Moreover, a number of particular characteristics have been intentionally included in each scenario to provide additional data for testing different object recognition algorithms and techniques. Concretely,

- **Inclusion of distinctive objects.** A number of patterns/objects have been placed at different rooms within these houses, concretely: teddies in *alma*, fruits in *anto*, numerical patterns in *pare* and geometric patterns in *rx2*.
- **Varying lighting conditions.** Each of the three sessions in *sarmis* house was conducted at a different time of the day, which means that the objects were visualized under different lighting conditions.
- **Varying sets of objects.** In these three sessions, the set of objects placed in each room from session to session differs, with objects dis/appearing as well as being moved.

Although its main application is the aforementioned semantic mapping, it can be also useful for the recognition of instances of objects/rooms, object segmentation, or data compression/transmission algorithms. Moreover, typical robotic tasks like 3D map building, localization, or SLAM can be tested with Robot@Home, since the robot localization can be accurately estimated from the sequence of 2D scans. Finally, the distinctive patterns and objects placed on purpose can be used, for example, to test object-finding algorithms.

4.3.3 Multiversal Semantic Maps

The third contribution of this chapter is a novel semantic map representation, called *Multiverse Semantic Map* (*MvSmap*). This representation handles uncertainty by considering the different combinations of possible groundings of objects and rooms in the robot workspace, or *universes*, as instances of ontologies with belief annotations on their grounded concepts and relations. These beliefs are provided by the probabilistic recognition techniques described in Chapter 3. According to them, it also encodes the probability of each ontology instance being the right one. Thus, *MvSmaps* can be exploited by logical reasoners performing over such ontologies, as well as by probabilistic reasoners working with the CRF representation. This ability to manage different semantic interpretations of the robot workspace, which can be leveraged by probabilistic conditional planners (*e.g.* those in [61] or [2]), is crucial for a coherent robot operation.

The proposed *MvSmap* (see Figure 4.5) is inspired by the multi-hierarchical semantic map presented in Galindo *et al.* [37]. This map considers two separated but tightly related hierarchical representations containing: (i) the semantic, meta-information about the domain at hand, *e.g.* refrigerators keep food cold and are usually found in kitchens, and (ii) the factual, spatial knowledge acquired by the robot and its implemented algorithms from a certain workspace, *e.g.* obj-1 is perceived and recognized as a refrigerator. These hierarchies are called terminological

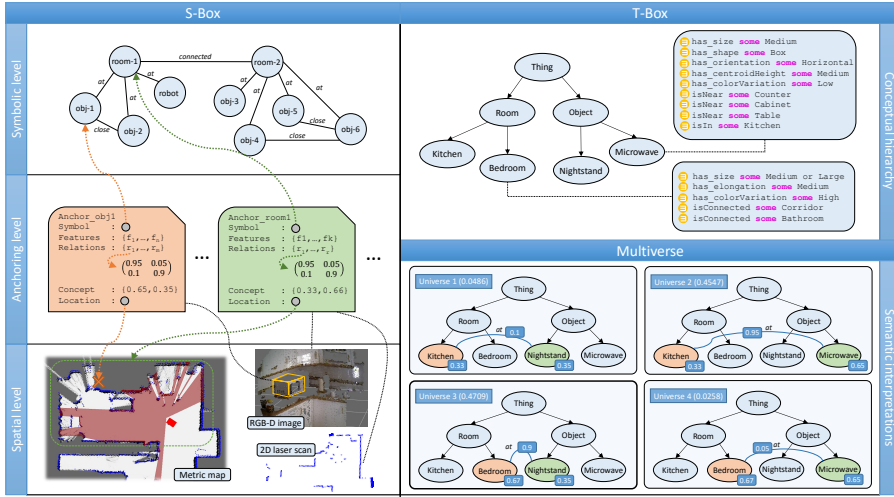


Figure 4.5: Example of Multiversal Semantic Map for a simple scenario.

box (see *T-Box* in Figure 4.5) and spatial box (see *S-Box* in Figure 4.5), respectively, names borrowed from the common structure of hybrid knowledge representation systems [7].

MvSmaps enhance this representation by including uncertainty, in the form of *beliefs*, about the groundings (recognitions) of the spatial elements in the S-Box to concepts in the T-Box. For example, a perceived object, represented by the symbol *obj-1*, could be grounded by the robot as a microwave or a nightstand with beliefs 0.65 and 0.35, respectively, or it might think that a room (*room-1*) is a kitchen or a bedroom with beliefs 0.33 and 0.66. Moreover, in this representation the relations among the spatial elements play a pivotal role, and they have also associated compatibility values in the form of beliefs. To illustrate this, if *obj-1* was found in *room-1*, *MvSmaps* can state that the compatibility of *obj-1* and *room-1* being grounded to microwave and kitchen respectively is 0.95, while to microwave and bedroom is 0.05. These belief values are provided by the proposed probabilistic inference techniques.

Furthermore, *MvSmaps* assign a probability value to each possible set of groundings, creating a *multiverse*, i.e. a set of universes stating different explanations of the robot environment (see *Multiverse* in Figure 4.5). An universe codifies the joint probability of the observed spatial elements being grounded to certain concepts, hence providing a global sense of certainty about the robot understanding of the environment. Thus, following the previous example, an universe can represent that *obj-1* is a microwave and *room-1* is a kitchen, while a parallel universe states that *obj-1* is a nightstand and *room-1* is a bedroom, both explanations annotated with different probabilities. Thereby, the robot performance is not limited to the utilization of

the most probable universe, like traditional semantic maps do, but it can also consider other possible explanations with different semantic interpretations, resulting in a more coherent robot operation.

The symbol grounding problem, *i.e.* linking portions of sensory data, represented by symbols (*e.g.* obj-1 or room-2), to concepts in the KB (*e.g.* Microwave or Kitchen), is faced by an anchoring process [18] that relies on the proposed recognition techniques and a simple tracking algorithm to make the symbols and their groundings consistent over time. In a nutshell, the result of this process is a set of the so-called anchors, which keep geometric/appearance information about the spatial elements (location, features, relations, etc.) and establish links to their symbolic representation. Additionally, in a *MvSmap*, anchors are in charge of storing the beliefs about the grounding of their respective symbols, as well as their compatibility with respect to the grounding of related elements.

Given the ingredients of *MvSmaps* previously provided, a *Multiversal Semantic Map* can be formally defined by the quintuple $MvSmap = \{\mathcal{R}, \mathcal{A}, \mathcal{Y}, \mathcal{O}, \mathcal{M}\}$, where:

- \mathcal{R} is a metric map of the environment, providing a global reference frame for the observed spatial elements (objects and rooms).
- \mathcal{A} is a set of anchors internally representing such spatial elements, and linking them with the set of symbols in \mathcal{Y} .
- \mathcal{Y} is the set of symbols that represent the spatial elements as instances of concepts from the ontology \mathcal{O} .
- \mathcal{O} is an ontology codifying the semantic knowledge of the domain at hand.
- \mathcal{M} encodes the multiverse, containing the set of universes.

Notice that the traditional T-Box and S-Box are defined in a *MvSmap* by \mathcal{O} and $\{\mathcal{R}, \mathcal{A}, \mathcal{Y}\}$ respectively. Since the robot is usually provided with the ontology \mathcal{O} beforehand, building a *MvSmap* consists of creating and maintaining the remaining elements in the map definition.

The suitability of the proposed semantic map representation was assessed with the challenging Robot@Home dataset. On the one hand, the reported success while grounding object and room symbols respectively without considering contextual relations was of $\sim 73.5\%$ and $\sim 57.5\%$, whereas including them these figures increased up to a success of $\sim 81.5\%$ and 91.5% . They have been also evaluated some of the most popular classifiers also resorting to individual object/room features, namely: Supported Vector Machines, Naive Bayes, Decision Trees, Random Forests, and Nearest Neighbors, demonstrating the reported results the higher success of CRF approaches.

On the other hand, they were also shown two sample scenarios of different complexity where it was illustrated the building of *MvSmaps* according to the information gathered by a mobile robot (see Figure 4.6). For a detailed description of this results,

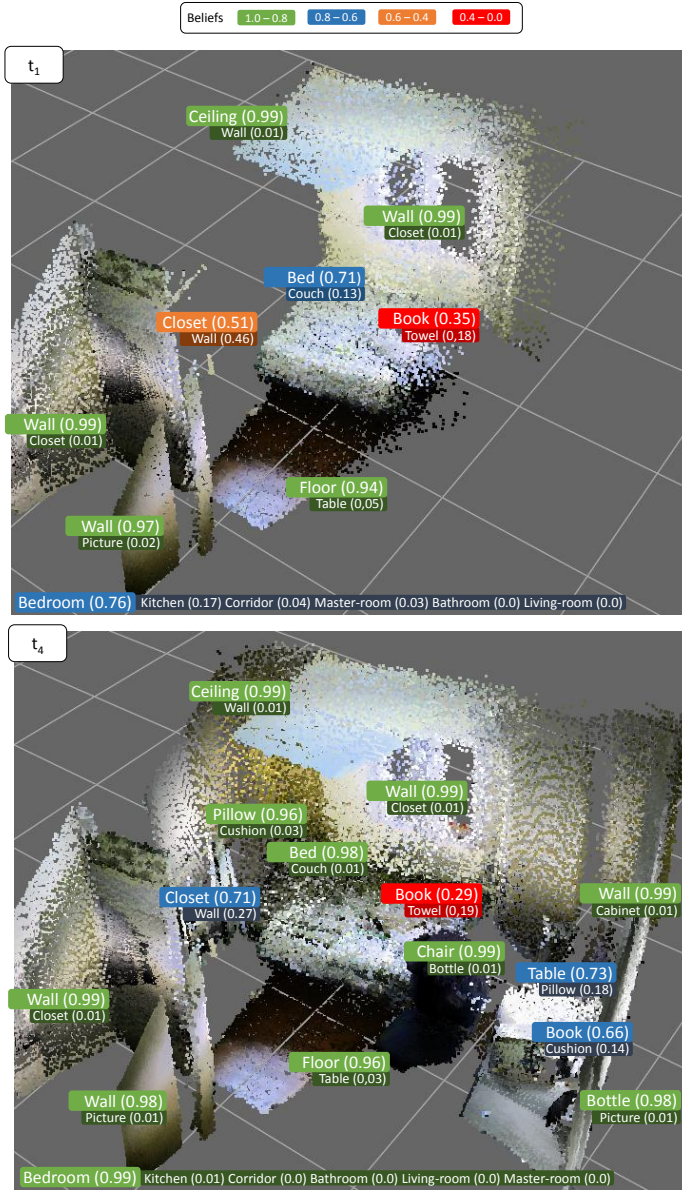


Figure 4.6: Grounding results and their belief values for the spatial elements perceived during the robot exploration of a bedroom from Robot@Home at two time instants: t_1 and t_4 .

as well as of the building of *MvSmaps*, please refer to the work by Ruiz-Sarmiento *et al.* [120].

The main purpose of the proposed *MvSmap* is to provide a mobile robot with a rich representation of its environment, empowering the efficient and coherent execution of high-level tasks. *MvSmaps* can be exploited for traditional semantic map applications by considering only an universe, albeit its potential to measure the (un)certainly of the robot understanding can be exploited for an intelligent, more efficient robotic operation. A clear example of this arises when considering the work by Galindo and Saffiotti [36], which envisages an application of semantic maps where they encode information about how things should be, also called norms, allowing the robot to infer deviations from these norms and act accordingly. The typical norm example is that "towels must be in bathrooms", so if a towel is detected, for example, on the floor of the living room, a plan is generated to bring it to the bathroom. This approach works with crispy information, *e.g.* an object is a towel or not. Instead, the consideration of a *MvSmap* would permit the robot to behave more coherently, for example gathering additional information if the belief of an object symbol being grounded to *Towel* is 0.55 while to *Carpet* is 0.45. In this example, a crispy approach could end up with a carpet in our bathroom, or a towel in our living room. Other applications where *MvSmaps* could be useful are task planning, planning with incomplete information, navigation, object search, human-robot interaction, or robotic localization.

4.4 Discussion

This chapter has outlined the thesis' contributions to the semantic mapping field. A novel semantic representation, called *Multiversal Semantic Map (MvSmap)*, has been described, which was designed to take advantage of the outcome for probabilistic recognition techniques. This permits the robot to propagate the uncertainty coming from different sources like its sensory system, or its internal models of the spatial elements, to the recognition results. *MvSmaps* also allow the tracking and exploitation of contextual relations among the elements in the robot workspace. The utilization of the uncertainty concerning the types of recognized spatial elements enables the robot to consider different semantic interpretations of its environment, resulting in a more coherent operation.

Additionally, it has been also described the *Robot@Home* dataset, a large repository of data collected by a mobile robot in domestic settings. The provided raw data come from two different types of sensors: a 2D laser scan mounted on the robot base, and a rig of 4 RGB-D cameras on the robot's neck. The processed information includes 2D and 3D reconstructions of the fully inspected houses, as well ground truth annotations about the type of the objects and rooms therein. Thus, this dataset is rich in contextual relations among spatial objects given the wide coverage of the provided data, so it is specially suitable for the evaluation of semantic mapping systems. To evaluate the proposed *MvSmaps*, the recognition techniques in the previous chapter

were integrated into a semantic mapping system building those representations, and *Robot@Home* was used as a testbed.

Robot@Home contains a huge number of observations whose processing by traditional techniques is prohibitive. Thereby, it was developed the *Object Labeling Toolkit* (OLT), a set of software components that greatly minimizes the operator intervention for processing sequential RGB-D observations. The developed/integrated algorithms for image processing, point cloud registration, scene reconstruction, scene labeling, and automatic propagation of labels to individual observations, really helped to keep the effort low for processing *Robot@Home*. Both dataset and toolkit are publicly available.

Summary of included papers

*This chapter outlines the content of the included papers, available at the second part of the thesis **Part II: Included papers**, as well as the author's contributions to each of them.*

5.1 Paper A: Learning CRFs with data from Semantic Knowledge

Outline: This paper studies the applicability of CRFs trained with synthetic data, generated from Semantic Knowledge, for contextually modeling the scene object recognition problem. The proposed learning approach aims at avoiding the collection of real data for training object recognition systems, which is a highly time-consuming, cumbersome, and even unfeasible task, since the gathered information must be representative enough of the domain at hand. To face this issue, Semantic Knowledge is represented by means of an Ontology, which defines the domain object classes, their properties, and their relations, and is used to generate synthetic training samples for tuning CRFs. The suitability of the learning approach has to be assessed through real datasets, so UMA-Offices and NYUv2 conformed the benchmark for answering questions like: *How much do the context relations contribute to the recognition performance?*, *How much does the size of the training dataset affect the recognition performance?*, or *Do the generated synthetic data capture actual object properties and relations?*.

Contribution by the author: Studied the state-of-the-art approaches for addressing the scene object recognition problem through Probabilistic Graphical Models or

Semantic Knowledge. Designed the way in which the relevant information can be encoded in an Ontology for its posterior exploitation. Implemented the algorithm for the automatic generation of an arbitrary number of synthetic training samples. Processed the UMA-Offices dataset, and performed the experiments to demonstrate the suitability of the approach.

5.2 Paper B: Joint recognition of objects and rooms

Outline: This work extends the previous one by including rooms in the equation. Motivated by recent studies that highlight the convenience of jointly modeling the object and room recognition problems (in view of the mutual influence between the types of the recognized rooms and the types of the objects therein), the ontology defined in Paper A is augmented to also consider room classes, their attributes, and relations among them as well as among objects and rooms: *e.g.* that bedrooms are usually connected to corridors and beds can be found therein. The CRF models are also conveniently adapted for dealing with different types of random variables (taking values from object or room types) and contextual relations. To validate the approach the paper resorts to home scenes from the NYUv2 dataset.

Contribution by the author: Studied state-of-the-art techniques for jointly modeling the object and room recognition problems. Designed the expansion of the Ontology in the previous paper, as well as of the CRF formulation and the algorithm implemented for generating synthetic training samples. Performed the experiments to support the paper claims.

5.3 Paper C: Exploiting Semantic Knowledge for a coherent and efficient recognition

Outline: The complexity of CRF models increases considerably when applied to cluttered scenarios. This implies the utilization of approximate inference methods for retrieving the recognition results, which in some cases supposes a decrease in the recognition success when compared with exact inference solutions. This paper proposes the utilization of Semantic Knowledge to decrease the CRF inference complexity. This knowledge, encoded in an Ontology, is exploited for the generation of hypotheses about the most probable belonging classes of the objects according to their features. For example, a planar, vertical surface could be a wall or a screen, but not a table. Then, these hypotheses are considered by the CRF as the only possible candidate types. The consequence of this is a considerable reduction in the number of possible assignments, decreasing the inference complexity, even enabling exact inference in some cases. Additionally, prior information about the frequency of occurrence of the different object classes is also encoded into the Ontology. This information reveals that, for example, it is more likely to encounter a computer than a couch in an

office environment, while it is quite unlikely to find an ironing table. A modification to the usual CRF formulation is proposed to exploit such source of prior information. The gain in efficiency and coherence by this approach is measured against the UMA-Offices and NYUv2 datasets.

Contribution by the author: Designed the framework for, employing the hypotheses generated by logical inference over the ontology, reduce the complexity of the CRF model. Adapted the CRF formulation to also consider prior information about the frequency of occurrence of the different object types from the Ontology. Evaluated the achieved complexity reduction and enhanced recognition coherence with two different repositories.

5.4 Paper D: UPGMpp library for managing PGMs

Outline: This paper presents the Undirected Probabilistic Graphical Models in C++ (UPGMpp) library, a software package for working with Undirected PGMs, as is the case of CRFs. The library was specially designed and implemented for efficiently tackling the object/room recognition problem. The paper describes how to apply UPGMpp to this issue, and overviews its three main software packages: *base* (implements the functionality for building and managing PGM graphs), *training* (permits the definition of training datasets to tune a PGM), and *inference* (implements algorithms to perform inference queries over PGMs). To show the flexibility and usability of the library, the paper describes the processes needed for training and testing (performing inference) CRFs, including code snippets, and reports the recognition results yielded by the implemented inference methods dealing with information from the NYUv2 repository. Execution time performance is also discussed.

Contribution by the author: Studied the theory behind Undirected PGMs, as well as related libraries and software solutions for dealing with them. Designed and implemented the library packages, with the goal of being efficient, versatile, extensible, and easy to use. Made the library publicly available. Exemplified how to use the library, and measured its success and execution time performance.

5.5 Paper E: OLT toolkit for managing sequential RGB-D datasets

Outline: In this work it is presented the Object Labeling Toolkit (OLT), a set of software components for the efficient labeling of datasets compound of sequences of RGB-D observations, gathered from an arbitrary number of sensors of that type. For that, the toolkit builds a 3D reconstruction of the scene explored in each RGB-D sequence, and allows the user to graphically label objects within that reconstruction.

Once the scene is labeled, such annotations are automatically propagated to each observation in the sequence. The paper describes its main components, namely: *dataset pre-processing*, *2D map building*, *localization of observation poses*, *sequential visualization*, *scene labeling*, and *labels propagation*, of which only *scene labeling* requires a human operator. It is also depicted the toolkit usage for effortlessly labeling two sequences of observations, also analyzing its virtues with respect to a typical labeling approach.

Contribution by the author: Designed the toolkit and its components. Studied and implemented/adapted techniques for processing RGB and depth images, building 2D geometric maps, building 3D reconstructions, visualizing and interacting with reconstructions, and automatically propagating information through a sequence of sensory data. Compared the time saved when employing the toolkit with respect to a typical labeling approach.

5.6 Paper F: Semantic Map representation handling uncertainty

Outline: This paper proposes a semantic map representation that handles uncertainty, also taking advantage of contextual relations among spatial elements (objects and rooms), coined Multiversal Semantic Map (*MvSmap*). The paper reports a comprehensive survey on semantic mapping approaches, as well as on grounding techniques for populating those maps. *MvSmaps* are described in detail and formally defined, along with the algorithms involved in their building, where the recognition techniques presented in previous works play a pivotal role. Moreover, this paper includes algorithms for efficiently tackling the uncertainty modeled by these maps. The novel Robot@Home dataset is used for both, testing the symbol grounding success, as well as illustrating the building of *MvSmaps* from scenarios with different complexity.

Contribution by the author: Designed the Multiversal Semantic Map representation for storing and managing uncertain information. Integrated the previously developed object and room recognition techniques within a symbol grounding process. Designed and implemented the pipeline for building *MvSmaps* according to the information perceived by a mobile robot. Processed the Robot@Home dataset for being useful for testing symbol grounding algorithms, as well as for illustrating the building of *MvSmaps*.

Conclusions and future work

Reaching the end of the thesis, it is time to draw conclusions and think about the future.

This thesis has explored and made contributions to the fascinating world of semantic mapping applied to mobile robots. This type of maps aims to provide a robot with a *sense of understanding* of what is going on in its surroundings, which sets the basis for an intelligent, autonomous, and efficient operation. Particular emphasis has been placed on the population of semantic maps with information about the spatial elements in the robot workspace, namely objects and rooms, through the combination of techniques from *Machine Learning* and *Artificial Intelligence*. These fields are at a great point, evidenced by a growing number of studies and successful applications, as recently commented by Ralf Herbrich – Amazon’s director of machine learning – “*We’re in a golden age of machine learning and AI, ... , as a scientific community, we are still a long way from being able to do things the way humans do things, but we’re solving unbelievably complex problems every day and making incredibly rapid progress.*”. In the author’s opinion, the research of systems exploiting the synergy of these two fields, boosting their advantages and mitigating their limitations, can lead to remarkable advances profitable by the robotic community. That is the case of the techniques developed in this thesis.

In order to be aware of its surroundings, a mobile robot must be able to recognize the elements that are observed through its sensory system. The second chapter of this thesis described the contributions done in this regard, which focused on the combination of *Conditional Random Fields* (CRFs), a discriminative, undirected variant of

Probabilistic Graphical Models (PGMs), and *Semantic Knowledge* of the domain at hand codified in an *Ontology*. These two frameworks have reached a notable success in different classification applications.

CRFs master the modeling of contextual relations among spatial elements, also handling the uncertainty coming from the robot sensory system and the employed models, and supporting the execution of probabilistic inference methods. Precisely, one of the earliest contributions of this thesis was the *Undirected Probabilistic Graphical Models in C++* (UPGMpp) library, developed as a consequence of the lack of software tools for handling Undirected PGMs in general, and CRFs in particular, providing the features demanded by a recognition system running on board of a mobile robot. This library, which is publicly available, implements popular algorithms for building, learning and performing inference over graphical models. The possible choices of training and inference methods for CRFs motivated the thorough study of different learning strategies, in order to find the most successful configuration for the scene object/room recognition problem. This study provided valuable conclusions, not only for the appropriate utilization of these models in the remaining contributions, but also for those in the robotic community aiming to quickly set-up a working-system as successful as possible for such problem.

Despite their successful utilization in different fields, CRFs exhibit a number of shortcomings when applied to recognition. First of all, to be properly tuned, they require a considerable amount of training data comprehensively covering the elements within the domain at hand. The collection of a dataset is a tedious, heavily time-consuming, and (in some domains) unfeasible task, as the author experienced when processing the *UMA-Offices* dataset. Such dataset, consisting of 25 scenes captured by a mobile robot from office facilities within the *University of Málaga*, was collected to evaluate the developed recognition techniques in conjunction with other state-of-the-art repositories containing information from the trending topic sensors, *#RGB-D_cameras*. To avoid the dependency of datasets containing real data, it was shown how Semantic Knowledge, conveniently codified in an Ontology, can be used to effortlessly generate an arbitrary number of training samples representative of the domain at hand. Ontologies provide a natural way to encode Semantic Knowledge, and suppose a compact, human-readable, and ready-to-use representations in high-level reasoning tasks. However, they are unable to handle uncertainty, and it is difficult to fill the gap between the low level sensory data and the codified information without introducing additional ad-hoc processes. Their synergy with CRFs removes these limitations, setting a mutual benefit relationship.

This thesis has exhibited that Ontologies have much to offer to its marriage with CRFs. For example, they have been employed to generate hypotheses about the possible types of the objects/rooms within a scene, drastically reducing in that way the complexity of the CRFs modeling such scene. This increases the efficiency of approximate inference methods over CRFs, also broaden the scenarios where exact inference is feasible. Notice that the efficiency of the recognition method is key for the proper robot operation, since it must share the (usually limited) robot resources with other algorithms in execution like those performing navigation or localization. On-

tologies may encode different types of information about the elements of the domain of discourse, and this has been leveraged to codify the frequency of occurrence of the different object classes. The usual CRF formulation has been accordingly adapted to exploit this source of prior information, allowing these models to achieve more coherent recognition results. Encoded Semantic Knowledge has been also used to detect incoherences in such results, and learn from them in collaboration with a human. This approach overcomes the CRF inability to learn from experience, and permits it to improve its performance and robustness in the long-term operation within home environments.

Once the mobile robot was able to recognize the elements in its surroundings with guarantees, such recognition framework was integrated into a semantic mapping system. For that, it was designed the *Multiversal Semantic Map (MvSmap)*, a representation of the robot workspace able to accommodate and take advantage of the probabilistic outcome of the developed recognition techniques. This map considers different interpretations of the spatial elements, called *universes*, as instantiations of Ontologies, creating a *multiverse*. These Ontologies are further annotated with the probabilities yielded by the recognition framework, as well as with their probability of being the true one. Thereby, the robot performance is not limited to the utilization of the most probable universe, like traditional semantic maps do, but it can also consider other possible explanations with different semantic interpretations, resulting in a more coherent robot operation. A way to keep the complexity of the multiverse tractable has been also presented, enabling its utilization in complex environments.

Two additional resources related to semantic mapping have been also made public. The first one is a dataset, coined *Robot@Home*, containing among others: 87,000+ time-stamped observations gathered by a mobile robot endowed with a rig of 4 RGB-D cameras and a 2D laser scanner, 3D reconstructions and 2D geometric maps of fully explored houses, topological information about the connectivity of rooms, and ground truth annotations about the type of the surveyed rooms and objects. The dataset is rich in contextual information of the contained spatial elements, a valuable feature missing in most of the state-of-the-art datasets, which can be exploited by semantic mapping systems. The second contribution in this regard is the *Object Labeling Toolkit (OLT)*, a set of software components to efficiently process sequences of sensory information, including RGB-D observations. Such components are highly customizable and expandable, facilitating the integration of already-developed algorithms, and have proven to drastically reduce the time and effort needed for processing that type of datasets.

As a final remark, it is worth to say that although all the techniques described in this thesis have been assessed with data repositories from domestic and office environments, their utilization is not restricted to these domains, but they can be exploited in any scenario exhibiting rich semantic information as hospitals, shopping centers, or other human-like environments. Moreover, their use is not restricted to mobile robot applications, but they could be exported to other fields that would benefit from the exploitation of semantic maps as assistance to visual impaired or elderly people, augmented reality, and more applications to appear in the era of portable devices able

to execute this kind of techniques. Nowadays, in fact, our smartphones are almost as powerful as our desktop computers. The research efforts in semantic mapping, along with the new technological advances, ensure the emergence of breakthrough and exciting applications. Stay tuned!

Future work

The work done in this thesis leaves a number of research lines open. Some of the most interesting ones are outlined below.

Hypotheses generation. The generation of hypotheses employing the information encoded in the Ontology could be so restrictive in some situations, mainly with objects showing unusual properties. Let's suppose a scene with a book placed on the floor. In that situation the logical reasoner does not yield the type Book as a hypothesis, given that its height largely differs from the expected one. An option could be to consider the result of the logical inference as a score to be introduced in the CRF formulation, at the cost of compromising the exact inference option.

Exploitation of MvSmaps. The real potential of *Multiversal Semantic Maps* (in the author's opinion) is still to come. Several proof-of-concept applications have been designed and tested, but it should be studied the benefits of this representation in real world problems like efficient navigation and object search, robot localization, task planning with uncertain/incomplete information, etc.

Learning from experience. There is significant room to explore possible improvements to the proposed system for learning from experience. Firstly, it should be conducted a thorough evaluation of the system with complex CRFs and ontologies, including information from objects and rooms, during long periods of time. Since the human is in the *learning loop*, it could be also studied how possible incorrect indications by the user affect the performance. The system could also benefit from a study of when would be more appropriate to ask the user about inconsistent results in order to not bother him/her.

Further development of UPGMpp. Some additional features regarding the performance of the UPGMpp library could be explored. For example, although the most time-consuming parts of the library are parallelized through OpenMP, some repetitive operations intensively employing data could also benefit for a parallelization at a lower level, aiming to also take advantage of GPU cores through, for example, CUDA or OpenCL. Visualization tools for inspecting the underlying graphs would be also useful for understanding what is on in the code and during execution. The implementation of sampling techniques for drawing samples from the probability defined by a PGM (like Markov Chain Monte Carlo), are also in the spotlight. Of course, any contribution to UPGMpp from the computer vision or robotic communities is welcome.

Improvements in OLT. The incorporation of algorithms for a globally consistent alignment of the RGB-D observations used to reconstruct a scene would lead to even more accurate models. The user experience could be also improved with the addition of geometric primitives like spheres or cylinders to the currently used one (boxes) to segment and label scenes. Moreover, the time needed for labeling would be reduced if an initial segmentation of the scene as well as tentative labels for the objects/rooms therein are provided beforehand.

Bibliography

- [1] B. Abramson, J. Brown, W. Edwards, A. Murphy, and R. L. Winkler. Hailfinder: A bayesian system for forecasting severe weather. *International Journal of Forecasting*, 12(1):57–71, 1996.
- [2] Al-Moadhen, A. Abdulhadi, M. Packianather, R. Setchi, and R. Qiu. Robot task planning in deterministic and probabilistic conditions using semantic knowledge base. *International Journal of Knowledge and Systems Science (IJKSS)*, 7(1):56–77, Jan. 2016.
- [3] A. Aldoma, T. Faulhammer, and M. Vincze. Automation of “ground truth” annotation for multi-view rgb-d object instance recognition datasets. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 5016–5023, Sept 2014.
- [4] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena. Contextually guided semantic labeling and search for three-dimensional point clouds. *In the International Journal of Robotics Research*, 32(1):19–34, Jan. 2013.
- [5] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena. Contextually guided semantic labeling and search for three-dimensional point clouds. *In The International Journal of Robotics Research*, 32(1):19–34, Jan. 2013.
- [6] H. Andreasson, A. Treptow, and T. Duckett. Localization for mobile robots using panoramic vision, local features and particle filter. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 3348–3353, April 2005.

- [7] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2007.
- [8] J. Bai, Y. Wu, J. Zhang, and F. Chen. Subset based deep learning for rgb-d object recognition. *Neurocomputing*, 165(0):280 – 292, 2015.
- [9] E. Bastianelli, D. D. Bloisi, R. Capobianco, F. Cossu, G. Gemignani, L. Iocchi, and D. Nardi. On-line semantic mapping. In *Advanced Robotics (ICAR), 2013 16th International Conference on*, pages 1–6, Nov 2013.
- [10] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL Web Ontology Language reference. W3C Recommendation, 2004.
- [11] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.
- [12] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [13] J. Blanco, J. González, and J.-A. Fernández-Madrigal. Subjective local maps for hybrid metric-topological {SLAM}. *Robotics and Autonomous Systems*, 57(1):64 – 74, 2009.
- [14] J.-L. Blanco, J.-A. Fernández-Madrigal, and J. González-Jiménez. Towards a unified bayesian approach to hybrid metric-topological slam. *IEEE Transactions on Robotics*, 24(2):259–270, 2008.
- [15] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, Nov 2001.
- [16] L. Chang, M. M. Duarte, L. Sucar, and E. F. Morales. A bayesian approach for object classification based on clusters of SIFT local features. *Expert Systems with Applications*, 39(2):1679 – 1686, 2012.
- [17] Cognitum. Fluent editor home page. <http://www.cognitum.eu/semantics/FluentEditor/>, 2016. [Online; accessed 16-September-2016].
- [18] S. Coradeschi and A. Saffiotti. An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2-3):85–96, 2003.
- [19] S. Divvala, D. Hoiem, J. Hays, A. Efros, and M. Hebert. An empirical study of context in object detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1271–1278, June 2009.

- [20] F. Dornaika, A. Bosaghzadeh, H. Salmane, and Y. Ruichek. Graph-based semi-supervised learning with local binary patterns for holistic object categorization. *Expert Systems with Applications*, 41(17):7744 – 7753, 2014.
- [21] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, P. Gancarski, O. Boussaid, and A. Puissant. Ontology-based object recognition for remote sensing image interpretation. In *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*, volume 1, pages 472–479, Oct 2007.
- [22] S. Ekvall, D. Kragic, and P. Jensfelt. Object detection and mapping for service robot tasks. *Robotica*, 25(2):175–187, Mar. 2007.
- [23] A. Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal on Robotics and Automation*, 3(3):249–265, June 1987.
- [24] G. Elidan, I. McGraw, and D. Koller. Residual belief propagation: Informed scheduling for asynchronous message passing. In *Proceedings of the Twenty-second Conference on Uncertainty in AI (UAI)*, Boston, Massachusetts, July 2006.
- [25] N. Eric Maillot and M. Thonnat. Ontology based complex object recognition. *Image Vision Comput.*, 26(1):102–113, Jan. 2008.
- [26] P. Espinace, T. Kollar, A. Soto, and N. Roy. Indoor scene recognition through object detection. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 1406–1413. IEEE, 2010.
- [27] M. Everingham, L. van Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.
- [28] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2010.
- [29] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning object categories from google’s image search. In *IEEE International Conference on Computer Vision (ICCV 2005)*, volume 2, pages 1816–1823 Vol. 2, 2005.
- [30] E. Fernandez-Moral, J. González-Jiménez, P. Rives, and V. Arévalo. Extrinsic calibration of a set of range cameras in 5 seconds without pattern. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, Chicago, USA, September 2014.
- [31] E. Fernandez-Moral, W. Mayol-Cuevas, V. Arevalo, and J. Gonzalez-Jimenez. Fast place recognition with plane-based maps. In *IEEE International Conference on Robotics and Automation (ICRA 2013)*, pages 2719–2724, 2013.

- [32] T. Finley and T. Joachims. Training structural svms when exact inference is intractable. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 304–311, New York, NY, USA, 2008. ACM.
- [33] G. Floros and B. Leibe. Joint 2d-3d temporally consistent semantic segmentation of street scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pages 2823–2830, 2012.
- [34] C. Galindo, J. Fernandez-Madrigal, J. Gonzalez, and A. Saffiotti. Using semantic information for improving efficiency of robot task planning. In *IEEE International Conference on Robotics and Automation (ICRA), Workshop on Semantic Information in Robotics*, Rome, Italy, 2007.
- [35] C. Galindo, J. Fernandez-Madrigal, J. Gonzalez, and A. Saffiotti. Robot task planning using semantic maps. *Robotics and Autonomous Systems*, 56(11):955–966, 2008.
- [36] C. Galindo and A. Saffiotti. Inferring robot goals from violations of semantic knowledge. *Robotics and Autonomous Systems*, 61(10):1131–1143, 2013.
- [37] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2278–2283, Aug 2005.
- [38] C. Galleguillos and S. Belongie. Context based object categorization: A critical survey. *Computer Vision and Image Understanding*, 114(6):712–722, June 2010.
- [39] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, J. Garcia-Rodriguez, J. Azorin-Lopez, M. Saval-Calvo, and M. Cazorla. Multi-sensor 3d object dataset for object recognition with full pose estimation. *Neural Computing and Applications*, pages 1–12, 2016.
- [40] G. Gemignani, D. Nardi, D. D. Bloisi, R. Capobianco, and L. Iocchi. Interactive semantic mapping: Experimental evaluation. In A. M. Hsieh, O. Khatib, and V. Kumar, editors, *Experimental Robotics: The 14th International Symposium on Experimental Robotics*, volume 109 of *Springer Tracts in Advanced Robotics*, pages 339–355. Springer International Publishing, 2016.
- [41] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang. Hermit: an owl 2 reasoner. *Journal of Automated Reasoning*, 53(3):245–269, 2014.
- [42] R. Gómez-Ojeda, J. Briales, E. Fernández-Moral, and J. González-Jiménez. Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners. In *IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, USA, 2015.

- [43] R. Gonçalves, M. Horridge, M. Musen, C. Nyulas, S. Tu, and T. Tudorache. Protégé home page. <http://protege.stanford.edu/>, 2015. [Online; accessed 26-June-2015].
- [44] G. Guennebaud, B. Jacob, et al. Eigen v3. <http://eigen.tuxfamily.org>, 2010.
- [45] M. Günther, T. Wiemann, S. Albrecht, and J. Hertzberg. Building semantic object maps from sparse and noisy 3d data. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013)*, pages 2228–2233, 2013.
- [46] R. Gupta and M. J. Kochenderfer. Common sense data acquisition for indoor mobile robots. In *Proceedings of the 19th National Conference on Artificial Intelligence*, AAAI’04, pages 605–610. AAAI Press, 2004.
- [47] V. Haarslev, K. Hidde, R. Möller, and M. Wessel. The racerpro knowledge representation and reasoning system. *Semantic Web Journal*, 3(3):267–277, 2012.
- [48] J. Hammersley and P. Clifford. Markov fields on finite graphs and lattices, 1971. Unpublished manuscript.
- [49] S. Harnad. The symbol grounding problem. *Phys. D*, 42(1-3):335–346, June 1990.
- [50] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Muller-Gartner. Markov random field segmentation of brain mr images. *IEEE transactions on medical imaging*, 16(6):878–886, 1997.
- [51] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab. Model based training, detection and pose estimation of textureless 3d objects in heavily cluttered scenes. In *Proceedings of the 11th Asian Conference on Computer Vision - Volume Part I, ACCV’12*, pages 548–562, Berlin, Heidelberg, 2013. Springer-Verlag.
- [52] W. L. Hoo, C. H. Lim, and C. S. Chan. Keybook: Unbias object recognition using keywords. *Expert Systems with Applications*, 42(8):3991 – 3999, 2015.
- [53] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, and M. Dean. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member Submission, World Wide Web Consortium, 2004.
- [54] F. Husain, L. Dellen, and C. Torras. Recognizing point clouds using conditional random fields. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 4257–4262, Aug 2014.
- [55] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *The Journal of Machine Learning Research*, 6:695–709, Dec. 2005.

- [56] J. Jancsary, S. Nowozin, T. Sharp, and C. Rother. Regression tree fields - an efficient, non-parametric approach to image labeling problems. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pages 2376–2383, 2012.
- [57] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell. A category-level 3-d object dataset: Putting the kinect to work. In *1st Workshop on Consumer Depth Cameras for Computer Vision (ICCV workshop)*, November 2011.
- [58] J.L. Blanco Claraco. Mobile Robot Programming Toolkit (MRPT). <http://www.mrpt.org>, 2015. [Online; accessed 28-April-2015].
- [59] A. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14:841, 2002.
- [60] O. Kahler and I. Reid. Efficient 3d scene labeling using fields of trees. In *2013 IEEE International Conference on Computer Vision*, pages 3064–3071, Dec 2013.
- [61] L. Karlsson. Conditional progressive planning under uncertainty. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'01*, pages 431–436, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [62] A. Kasper, Z. Xue, and R. Dillmann. The kit object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research*, 31(8):927–934, 2012.
- [63] R. Kindermann, J. L. Snell, et al. *Markov random fields and their applications*, volume 1. American Mathematical Society Providence, RI, 1980.
- [64] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool. Hough transform and 3d surf for robust three dimensional classification. In *Proceedings of the 11th European Conference on Computer Vision: Part VI, ECCV'10*, pages 589–602, Berlin, Heidelberg, 2010. Springer-Verlag.
- [65] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [66] F. Korč and W. Förstner. Approximate parameter learning in conditional random fields: An empirical investigation. In *Proceedings of the 30th DAGM Symposium on Pattern Recognition*, pages 11–20, Berlin, Heidelberg, 2008. Springer-Verlag.

- [67] I. Kostavelis and A. Gasteratos. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66:86–103, 2015.
- [68] S. Kumar, J. August, and M. Hebert. Exploiting inference for approximate parameter learning in discriminative fields: An empirical study. In *Proceedings of the 5th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EMMCVPR'05, pages 153–168, Berlin, Heidelberg, 2005. Springer-Verlag.
- [69] S. Kumar and M. Hebert. Discriminative random fields. *Int. J. Comput. Vision*, 68(2):179–201, June 2006.
- [70] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [71] K. Lai, L. Bo, and D. Fox. Unsupervised feature learning for 3d scene labeling. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 3050–3057, May 2014.
- [72] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1817–1824, May 2011.
- [73] D. Lin, S. Fidler, and R. Urtasun. Holistic scene understanding for 3d object detection with rgb-d cameras. *IEEE International Conference on Computer Vision*, 0:1417–1424, 2013.
- [74] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.
- [75] J. Martinez-Gomez, M. Cazorla, I. Garcia-Varea, and V. Morell. VidriLO: The visual and depth robot indoor localization with objects information dataset. *International Journal of Robotics Research*, 2015.
- [76] D. Meger, P.-E. Forssén, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe. Curious george: An attentive semantic robot. *Robots and Autonomous Systems*, 56(6):503–511, June 2008.
- [77] D. Meger and J. J. Little. The UBC visual robot survey: A benchmark for robot category recognition. In *Experimental Robotics - The 13th International Symposium on Experimental Robotics, ISER 2012, June 18-21, 2012, Québec City, Canada*, pages 979–991, 2012.
- [78] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan. Toward an assisted indoor scene perception for blind people with image multilabeling strategies. *Expert Systems with Applications*, 42(6):2907 – 2918, 2015.

- [79] R. Mottaghi, A. Ranganathan, and A. L. Yuille. A compositional approach to learning part-based models of objects. In *IEEE International Conference on Computer Vision Workshops (ICCV 2011 Workshops)*, pages 561–568, 2011.
- [80] O. Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using adaboost. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1730–1735, April 2005.
- [81] A. C. Murillo, J. J. Guerrero, and C. Sagues. Surf features for efficient robot localization with omnidirectional images. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3901–3907, April 2007.
- [82] J. N. N. Okazaki. libLBFGS: a library of Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS). <http://www.chokkan.org/software/liblbfgs/>, 2015. [Online; accessed 14-September-2015].
- [83] Y. Nesterov. *Introductory lectures on convex optimization : a basic course*. Applied optimization. Springer US, 2004.
- [84] D. Nikovski. Constructing bayesian networks for medical diagnosis from incomplete and partially correct statistics. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):509–516, 2000.
- [85] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2161–2168, 2006.
- [86] J. Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.
- [87] S. Nowozin, C. Rother, S. Bagon, T. Sharp, B. Yao, and P. Kohli. Decision tree fields. In *IEEE International Conference on Computer Vision (ICCV 2011)*, pages 1668–1675, 2011.
- [88] A. Nüchter and J. Hertzberg. Towards semantic maps for mobile robots. *Robots and Autonomous Systems*, 56(11):915–926, 2008.
- [89] N. Okazaki. Crfsuite: a fast implementation of conditional random fields (crfs). <http://www.chokkan.org/software/crfsuite/>. [Online; accessed 28-April-2015].
- [90] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.
- [91] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155:23–36, 2006.

- [92] A. Oliva and A. Torralba. The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12):520–527, Dec. 2007.
- [93] M. Oliveira, L. S. Lopes, G. H. Lim, S. H. Kasaei, A. D. Sappa, and A. M. Tomé. Concurrent learning of visual codebooks and object categories in open-ended domains. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 2488–2495, Sept 2015.
- [94] OpenMP Architecture Review Board: OpenMP API Specification for Parallel Programming. <http://openmp.org/wp/>. [Online; accessed 14-April-2016].
- [95] D. Pangercic, B. Pitzer, M. Tenorth, and M. Beetz. Semantic object maps for robotic housework - representation, acquisition and use. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4644–4651, Oct 2012.
- [96] S. Parise and M. Welling. Learning in markov random fields: An empirical study. In *Proceedings of the Joint Statistical Meeting, JSM2005*, 2005.
- [97] S. Payr, F. Werner, and K. Werner. Potential of robotics for ambient assisted living. Technical report, Austrian Research Institute for Artificial Intelligence, 2015.
- [98] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
- [99] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
- [100] M. Pontil and A. Verri. Support vector machines for 3d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):637–646, Jun 1998.
- [101] A. Pronobis and B. Caputo. Confidence-based cue integration for visual place recognition. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2394–2401, Oct 2007.
- [102] A. Pronobis and P. Jensfelt. Hierarchical multi-modal place categorization. In *European Conference on Mobile Robots (ECMR)*, pages 159–164, 2011.
- [103] A. Pronobis and P. Jensfelt. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3515–3522, May 2012.
- [104] A. Pronobis, P. Jensfelt, K. Sjöö, H. Zender, G.-J. M. Kruijff, O. M. Mozas, and W. Burgard. Semantic modelling of space. In H. I. Christensen, G.-J. M. Kruijff, and J. L. Wyatt, editors, *Cognitive Systems*, volume 8 of *Cognitive Systems Monographs*, pages 165–221. Springer Berlin Heidelberg, 2010.

- [105] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt. Multi-modal semantic place classification. *The International Journal of Robotics Research*, 2009.
- [106] A. Quattoni, M. Collins, and T. Darrell. Conditional random fields for object recognition. In *Advances in Neural Information Processing Systems*, pages 1097–1104. MIT Press, 2004.
- [107] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 413–420. IEEE, 2009.
- [108] A. Ranganathan and F. Dellaert. Semantic modeling of places using objects. In *Robotics: Science and Systems Conference III (RSS)*. MIT Press, 2007.
- [109] A. Ranganathan, E. Menegatti, and F. Dellaert. Bayesian inference in the space of topological maps. *IEEE Transactions on Robotics*, 22(1):92–107, Feb 2006.
- [110] E. Remolina and B. Kuipers. Towards a general theory of topological maps. *Artificial Intelligence*, 152(1):47–104, Jan. 2004.
- [111] X. Ren, L. Bo, and D. Fox. Rgb-(d) scene labeling: Features and algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pages 2759–2766, 2012.
- [112] L. Riazuelo, M. Tenorth, D. D. Marco, M. Salas, D. Gálvez-López, L. Mösenlechner, L. Kunze, M. Beetz, J. D. Tardós, L. Montano, and J. M. M. Montiel. Roboearth semantic mapping: A cloud enabled knowledge-based approach. *IEEE Transactions on Automation Science and Engineering*, 12(2):432–443, April 2015.
- [113] J. G. Rogers and H. I. Christensen. A conditional random field model for place and object classification. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1766–1772, May 2012.
- [114] J. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Mobile robot object recognition through the synergy of probabilistic graphical models and semantic knowledge. In *European Conf. on Artificial Intelligence. Workshop on Cognitive Robotics*, 2014.
- [115] J. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. UPGMpp: a Software Library for Contextual Object Recognition. In *3rd. Workshop on Recognition and Action for Scene Understanding*, 2015.
- [116] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Exploiting semantic knowledge for robot object recognition. *Knowledge-Based Systems*, 86:131–142, 2015.

- [117] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Joint categorization of objects and rooms for mobile robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.
- [118] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets. In *European Conference on Mobile Robots*, 2015.
- [119] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Scene object recognition for mobile robots through semantic knowledge and probabilistic graphical models. *Expert Systems with Applications*, 42(22):8805–8816, 2015.
- [120] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Building Multi-versal Semantic Maps for Mobile Robot Operation. *Submitted*, 2016.
- [121] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. A survey on learning approaches for undirected graphical models. Application to scene object recognition. *International Journal of Approximate Reasoning (Accepted)*, 2016.
- [122] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Probability and common-sense: Tandem towards robust robotic object recognition in ambient assisted living. *10th International Conference on Ubiquitous Computing and Ambient Intelligence*, 2016.
- [123] J. R. Ruiz-Sarmiento, C. Galindo, and J. González-Jiménez. Robot@home, a robotic dataset for semantic mapping of home environments. *Submitted*, 2016.
- [124] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014.
- [125] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, May 2008.
- [126] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.
- [127] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz. Towards 3d point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11):927 – 941, 2008. Semantic Knowledge in Robotics.
- [128] B. Schling. *The Boost C++ Libraries*. XML Press, 2011.
- [129] M. Schmidt. UGM: Matlab Code for Undirected Graphical Models. <http://www.cs.ubc.ca/~schmidtm/Software/UGM.html>, 2015. [Online; accessed 28-April-2015].

- [130] N. Silberman and R. Fergus. Indoor scene segmentation using a structured light sensor. In *Proceedings of the International Conf. on Computer Vision - Workshop on 3D Representation and Recognition*, 2011.
- [131] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor Segmentation and Support Inference from RGBD Images. In *Proc. of the 12th European Conference on Computer Vision (ECCV 2012)*, pages 746–760, 2012.
- [132] A. Singh, J. Sha, K. Narayan, T. Achim, and P. Abbeel. Bigbird: A large-scale 3d database of object instances. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 509–516, May 2014.
- [133] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51–53, June 2007.
- [134] R. Speer and C. Havasi. Conceptnet 5: a large semantic network for relational knowledge. In *The People’s Web Meets NLP. Theory and Applications of Natural Language*, pages 161—176. Springer, 2013.
- [135] C. Sutton and A. McCallum. Piecewise pseudolikelihood for efficient training of conditional random fields. In *Proceedings of the 24th international conference on Machine learning*, pages 863–870. ACM, 2007.
- [136] C. A. Sutton and A. Mccallum. Piecewise Training for Undirected Models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 568–575, 2005.
- [137] A. Teichman, S. Miller, and S. Thrun. Unsupervised intrinsic calibration of depth sensors via slam. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.
- [138] M. Tenorth, L. Kunze, D. Jain, and M. Beetz. Knowrob-map - knowledge-linked semantic object maps. In *2010 10th IEEE-RAS International Conference on Humanoid Robots*, pages 430–435, Dec 2010.
- [139] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21 – 71, 1998.
- [140] S. Thrun. Learning occupancy grid maps with forward sensor models. *Autonomous Robots*, 15(2):111–127, Sept. 2003.
- [141] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. Intelligent robotics and autonomous agents. MIT Press, 2005.
- [142] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 273–280. IEEE, 2003.

- [143] D. Tsarkov and I. Horrocks. *FaCT++ Description Logic Reasoner: System Description*, pages 292–297. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [144] M. Uschold and M. Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 11:93–136, 1996.
- [145] J. Valentin, S. Sengupta, J. Warrell, A. Shahrokni, and P. Torr. Mesh based semantic modelling for indoor and outdoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pages 2067–2074, 2013.
- [146] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, volume 1, pages 511–518, 2001.
- [147] P. Viswanathan, T. Southey, J. Little, and A. Mackworth. Place classification using visual object categorization and global information. In *Computer and Robot Vision (CRV), 2011 Canadian Conference on*, pages 1–7. IEEE, 2011.
- [148] M. Wainwright, T. Jaakkola, and A. Willsky. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Transactions on Information Theory*, 49(5):1120–1146, May 2003.
- [149] C. Weiss, H. Tamimi, A. Masselli, and A. Zell. A hybrid approach for vision-based outdoor robot localization using global and local image features. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1047–1052, Oct 2007.
- [150] Y. Weiss and W. T. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Trans. Inf. Theor.*, 47(2):736–744, Sept. 2006.
- [151] D. Wolf, J. Prankl, and M. Vincze. Fast semantic segmentation of 3d point clouds using a dense crf with learned parameters. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, 2015.
- [152] Y. Xiang, X. Zhou, Z. Liu, T.-S. Chua, and C.-W. Ngo. Semantic context modeling with maximal margin conditional random fields for automatic image annotation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3368–3375, 2010.
- [153] J. Xiao, A. Owens, and A. Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1625–1632, Dec 2013.

- [154] X. Xiong and D. Huber. Using context to create semantic 3d models of indoor environments. In *In Proceedings of the British Machine Vision Conference (BMVC 2010)*, pages 45.1–11, 2010.
- [155] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized Belief Propagation. In *Advances Neural Information Processing Systems*, volume 13, pages 689–695, 2001.
- [156] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493 – 502, 2008. From Sensors to Human Spatial Concepts.
- [157] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 13–13, June 2006.
- [158] K. Zhou, M. Zillich, H. Zender, and M. Vincze. Web mining driven object locality knowledge acquisition for efficient robot behavior. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 3962–3969, 2012.
- [159] M. Zou and S. D. Conzen. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71–79, 2005.

Part II

Included papers

List of papers

A Exploiting Semantic Knowledge for Robot Object Recognition

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Published in: *International Journal of Knowledge-Based Systems*.

Volume: 86.

Pages: 131-142.

Year: 2015.

DOI: <http://dx.doi.org/10.1016/j.knosys.2015.05.032>

Abstract

This paper presents a novel approach that exploits semantic knowledge to enhance the object recognition capability of autonomous robots. Semantic knowledge is a rich source of information, naturally gathered from humans (elicitation), which can encode both objects' geometrical/appearance properties and contextual relations. This kind of information can be exploited in a variety of robotics skills, especially for robots performing in human environments. In this paper we propose the use of semantic knowledge to eliminate the need of collecting large datasets for the training stages required in typical recognition approaches. Concretely, semantic knowledge encoded in an ontology is used to synthetically and effortlessly generate an arbitrary number of training samples for tuning Probabilistic Graphical Models (PGMs). We then employ these PGMs to classify patches extracted from 3D point clouds gathered from office environments within the UMA-offices dataset, achieving a $\sim 90\%$ of recognition success, and from office and home scenes within the NYU2 dataset, yielding a success of $\sim 81\%$ and $\sim 69.5\%$ respectively. Additionally, a comparison with state-of-the-art recognition methods also based on graphical models has been

carried out, revealing that our semantic-based training approach can compete with, and even outperform, those trained with a considerable number of real samples.

B Joint Categorization of Objects and Rooms for Mobile Robots

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Published in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*

Pages: 2523-2528

Year: 2015

DOI: 10.1109/IROS.2015.7353720

Abstract

In general, the problems of objects' and rooms' categorizations for robotic applications have been addressed separately. The current trend is, however, towards a joint modelling of both issues in order to leverage their mutual contextual relations: *object* \rightarrow *room* (e.g. the detection of a microwave indicates that the room is likely to be a kitchen), and *room* \rightarrow *object* (e.g. if the robot is in a bathroom, it is probable to find a toilet). *Probabilistic Graphical Models* (PGMs) are typically employed to conveniently cope with such relations, relying on inference processes to hypothesize about objects' and rooms' categories. In this work we present a *Conditional Random Field* (CRF) model, a particular type of PGM, to jointly categorize objects and rooms from RGBD images exploiting *object-object* and *object-room* relations. The learning phase of the proposed CRF uses *Human Knowledge* (HK) to eliminate the necessity of gathering real training data. Concretely, HK is acquired through elicitation and codified into an ontology, which is exploited to effortlessly generate an arbitrary number of representative synthetic samples for training. The performance of the proposed CRF model has been assessed using the NYU2 dataset, achieving a success of $\sim 70\%$ categorizing both, objects and rooms.

C Scene Object Recognition for Mobile Robots Through Semantic Knowledge and Probabilistic Graphical Models

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Published in: *International Journal of Expert Systems with Applications*

Volume: 42

Issue: 22

Pages: 8805–8816

Year: 2015

DOI: <http://dx.doi.org/10.1016/j.eswa.2015.07.033>

Abstract

Scene object recognition is an essential requirement for intelligent mobile robots. In addition to geometric or appearance features, modern recognition systems strive to incorporate *contextual information*, normally modelled through Probabilistic Graphical Models (PGMs) or Semantic Knowledge (SK). However, these approaches, separately, show some weaknesses that limit their application, e.g., the exponential complexity of the probabilistic inference over PGMs or the inability of SK to handle uncertainty. This paper presents a *hybrid PGM-SK system* for object recognition that integrates both techniques reducing their individual limitations and gaining in probabilistic inference efficiency, performance robustness, uncertainty handling, and providing coherent results according to domain knowledge codified by a human expert. We support this claim with an extensive experimental evaluation according to both recognition success and time requirements in real scenarios from two datasets (NYU2 and UMA-offices). The yielded figures support the suitability of the hybrid PGM-SK recognition system, and its applicability to mobile robotic agents.

D UPGMpp: a Software Library for Contextual Object Recognition

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Published in: 3rd. Workshop on Recognition and Action for Scene Understanding

Year: 2015

DOI: 10.13140/RG.2.2.25749.12006

Abstract

Object recognition is a cornerstone task towards the *scene understanding* problem. Recent works in the field boost their performance by incorporating contextual information to the traditional use of the objects' geometry and/or appearance. These contextual cues are usually modeled through *Conditional Random Fields* (CRFs), a particular type of undirected *Probabilistic Graphical Model* (PGM), and are exploited by means of probabilistic inference methods. In this work we present the *Undirected Probabilistic Graphical Models in C++* library (UPGMpp), an open source solution for representing, training, and performing inference over undirected PGMs in general, and CRFs in particular. The UPGMpp library supposes a reliable and comprehensive workbench for recognition systems exploiting contextual information, including a variety of inference methods based on *local search*, *graph cuts*, and *message passing* approaches. This paper illustrates the virtues of the library, i.e. it is efficient, comprehensive, versatile, and easy to use, by presenting a use-case applied to the object recognition problem in home scenes from the challenging NYU2 dataset.

E OLT: A Toolkit for Object Labeling applied to robotic RGB-D datasets

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Published in: *European Conference on Mobile Robots*

Year: 2015

DOI: 10.1109/ECMR.2015.7324214

Abstract

In this work we present the *Object Labeling Toolkit* (OLT), a set of software components publicly available for helping in the management and labeling of sequential RGB-D observations collected by a mobile robot. Such a robot can be equipped with an arbitrary number of RGB-D devices, possibly integrating other sensors (e.g. odometry, 2D laser scanners, etc.). OLT first merges the robot observations to generate a 3D reconstruction of the scene from which object segmentation and labeling is conveniently accomplished. The annotated labels are automatically propagated by the toolkit to each RGB-D observation in the collected sequence, providing a dense labeling of both intensity and depth images. The resulting objects' labels can be exploited for many robotic oriented applications, including high-level decision making, semantic mapping, or contextual object recognition. Software components within OLT are highly customizable and expandable, facilitating the integration of already-developed algorithms. To illustrate the toolkit suitability, we describe its application to robotic RGB-D sequences taken in a home environment.

F Building Multiversal Semantic Maps for Mobile Robot Operation

Authors: Jose-Raul Ruiz Sarmiento, Cipriano Galindo and Javier Gonzalez-Jimenez.

Submitted to: *International Journal of Knowledge-Based Systems (second revision).*

Year: 2016

Abstract

Semantic maps augment metric-topological maps with meta-information, *i.e. semantic knowledge* aimed at the planning and execution of high-level robotic tasks. Semantic knowledge typically encodes human-like concepts, like types of objects and rooms, which are connected to sensory data when symbolic representations of percepts from the robot workspace are grounded to those concepts. This *symbol grounding* is usually carried out by algorithms that individually categorize each symbol and provide a crispy outcome – a symbol is either a member of a category or not. Such approach is valid for a variety of tasks, but it fails at: (i) dealing with the uncertainty inherent to the grounding process, and (ii) jointly exploiting the contextual relations

among concepts (*e.g.* microwaves are usually in kitchens). This work provides a solution for *probabilistic symbol grounding* that overcomes these limitations. Concretely, we rely on Conditional Random Fields (CRFs) to model and exploit contextual relations, and to provide measurements about the uncertainty coming from the possible groundings in the form of beliefs (*e.g.* an object can be categorized (grounded) as a microwave or as a nightstand with beliefs 0.6 and 0.4, respectively). Our solution is integrated into a novel semantic map representation called *Multiversal Semantic Map* (MvSmap), which keeps the different groundings, or universes, as instances of ontologies annotated with the obtained beliefs for their posterior exploitation. The suitability of our proposal has been proven with the Robot@Home dataset, a repository that contains challenging multi-modal sensory information gathered by a mobile robot in home environments.