

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MATHIAS FASSINI MANTELLI

**Exploiting semantic information in indoor
environments**

Ph.D. Thesis Proposal

Advisor: Profa. Dra. Mariana Luderitz Kolberg
Coadvisor: Prof. Dr. Renan de Queiroz Maffei

Porto Alegre
October 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Prof^a. Patricia Pranke

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. Claudio Rosito Jung

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“If I have seen farther than others,
it is because I stood on the shoulders of giants.”*

— SIR ISAAC NEWTON

AGRADECIMENTOS

Agradeço ao \LaTeX por não ter vírus de macro...

ABSTRACT

Este documento é um exemplo de como formatar documentos para o Instituto de Informática da UFRGS usando as classes \LaTeX disponibilizadas pelo UTUG. Ao mesmo tempo, pode servir de consulta para comandos mais genéricos. *O texto do resumo não deve conter mais do que 500 palavras.*

Keywords: Formatação eletrônica de documentos. Padronização de documentos. Instituto de Informática da UFRGS. \LaTeX . ABNT. UFRGS.

Using L^AT_EX to Prepare Documents at II/UFRGS

RESUMO

This document is an example on how to prepare documents at II/UFRGS using the L^AT_EX classes provided by the UTUG. At the same time, it may serve as a guide for general-purpose commands. *The text in the abstract should not contain more than 500 words.*

Palavras-chave: Electronic formatting of documents. Instituto de Informática da UFRGS. L^AT_EX. ABNT. UFRGS.

LIST OF FIGURES

Figure 1.1 plane.	12
------------------------	----

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

OS Object Search

SLAM Simultaneous Localization and Mapping

CONTENTS

1 INTRODUCTION.....	11
1.1 Outline.....	14
1.2 Contributions.....	14
2 THEORETICAL BACKGROUND.....	16
REFERENCES.....	17
APPENDIX A — RESUMO EXPANDIDO	18

1 INTRODUCTION

The first decades of research in Mobile Robotics, from the beginning until 2004, handled the challenges of connecting efficiency and data association. They introduced probabilistic formulations to path planning, exploration, simultaneous localization and mapping (SLAM), and many other areas. Some of the approaches from these areas are still popular nowadays, such as RaoBlackwellised Particle Filters and Extended Kalman Filters. The majority of them were based on ultrasonic or lidar sensors, as they were the most popular and robust sensors at the time. Consequently, the outcome maps were mostly 2D grid ones, in which the cells represented the free, occupied (obstacles), and unknown regions (CESAR et al., 2016).

After building a solid foundation for many problems with probabilistic approaches, the research community took a forward step. They concentrated on improving the properties of the already proposed and new approaches like observability, convergence, and consistency (CESAR et al., 2016). Simultaneously in this period (2004-2015), visual sensors have been in the spotlight as an alternative to gain information about the environment. Their considerable improvement in data quality and variety (e.g., depth images, point clouds, stereo images) aided their increasing employment. In fact, building 2D and 3D maps from the environment with a visual sensor resulted in a new term, Visual SLAM (SALAS-MORENO, 2014).

Mobile robotics have enjoyed formidable advantages in performing tasks that only expect robots to navigate free spaces and avoid obstacles. Moving items from point A to point B or vacuuming free spaces are examples of robotics tasks with satisfactory solutions. However, the same level of success does not apply so far to many other high-level tasks that robots are supposed to dealing with nowadays. Since mobile robotics shifted its focus from factory floors and assembly lines to everyday living spaces, robots are demanded to perform human-like tasks in different scenarios that are not necessarily as strict and organized as the industrial world (AYDEMIR, 2012). Relying only on purely geometric maps and having limited perceptions that do not allow going beyond basic geometric representations does not allow the robot to obtain a high-level understanding of the environment. This might be the reason for robots not prosper as much in high-level tasks. The robots are deprived of processing the environmental data to infer or estimate extra valuable knowledge useful in various tasks. We then claim that when building autonomous robots, the high-level information inferred from the sensor readings (semantic

information) has to be heavily exploited to complement the robot's perception.

The association of semantic information (or concepts) to geometric entities in the map is called semantic mapping, one of the newest topics the researchers have explored. It enhances the robot's autonomy and robustness in many ways, besides facilitating some high-level tasks (CESAR et al., 2016). Fig. 1.1 is an image from Zoox's autonomous car, and it illustrates the advantage of using semantic information in robotics tasks. If the car maps this scene with its geometric perception, it will have the information that four obstacles are in its front, and two are closer than the other two. With a different perception based on semantic information, the car associates the concept of "person" to the three people and "vehicle" to the truck within the scene. Most importantly, it estimates that one person is distracted using his phone, and another is holding a stop sign. Combining the detection of a walking person and a phone allows the car to estimate the semantic information that this person is likely distracted. Hence, the car should drive itself even more carefully. This whole process is natural for human drivers, but the same can not be said about robots.

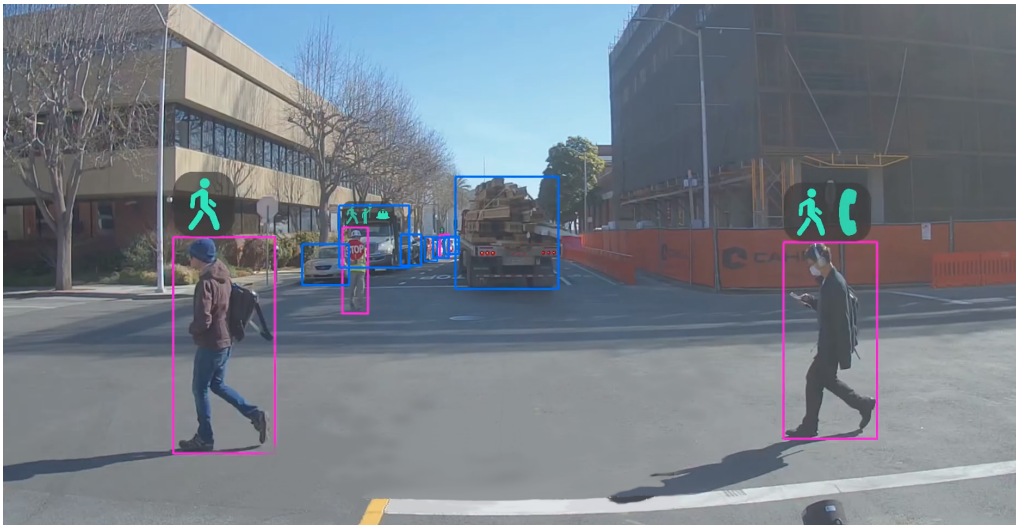


Figure 1.1 – plane.

As semantic information is more like a specific knowledge for each task and inferred from the robot's surroundings than a particular type of data from a sensor reading, several questions must be answered before using it in a robotic task. We see the following as noteworthy challenges:

- Deciding on what type of semantic information is possible to infer and associate to the robot's surroundings that is relevant to the task;
- How to perform the inferring or estimate the semantic information;

- How to use semantic information to improve the robot's performance in a particular task.

Since semantic information is relatively new in the literature, the first point is frequently discussed in the context of geometric information. Briefly, for the context of a robotic task, what information is not explicitly in the environment but could be inferred or estimated to improve the robot's performance? This demands a deep understanding of the task and the general environment characteristics where the robot operates. An inspiration for answering this point is to consider how humans reason under the same circumstance and solve such a task and how we process the environment's information to accomplish the task efficiently.

Second, depending on the needed semantic information, it may be necessary to use methods based on machine learning to estimate it. For example, training a deep learning model for estimating terrain traversability for an outdoor ground robot may provide a suitable result. However, besides the training requirement, the solution's quality depends on the training data, and this approach does not scale well. Probabilistic-based estimations appear as a second option, as it does not require a large set of data for training, and accepts a wide range of different models.

The third and last point, the proper use of the inferred semantic information in the robot's system, is crucial for successful task completion. As the robot gains more information from the environment, it is important to keep updating the estimations, and it is even better if the estimations become more robust over time.

The exploitation of semantic information in robotics is an idea that has recently gained attention from researchers, and thus, most of the challenging problems are still unsolved. A simple way of pushing the limits further and exploring these problems is to study the advantages of semantic information in different areas. In this thesis, we have chosen a task with a high difficulty level that can benefit from semantic information: object search (OS) in indoor and unknown environments, a yet unsolved problem in robotics.

In OS tasks, the robot's goal is to find a target object in the environment with a visual sensor. Usually, the environment is unknown to the robot, and the data it uses for searching are gathered with its own sensors. Since we are complementing the robot's perception with semantic information models for different OS tasks, the extra knowledge from the environment inferred by the robot has to aid the OS searching by reducing the search space. The robot plans a search strategy that estimates the most promising regions

to contain the target object. This thesis exploits the improvements in OS tasks by the use of semantic information inferred from two different data sources disregarded by the research community: text and dynamic obstacles.

1.1 Outline

The outline of this thesis is as follows.

1.2 Contributions

Parts of this thesis have been previously published or submitted as journal articles. The following publications are the results of research carried during this PhD:

1. MANTELLI, M. et al. Temporal object search system based on heat maps. *Journal of intelligent & robotic systems* (in review), Springer, v. 101, n. 2, p. 1–23, 2021.

Summary and Individual contribution: This paper is on how to search a target object in unknown and dynamic environments efficiently. As opposed to other OS works that consider the objects' position static and ignore the human-object interaction, the idea presented in this work is to incorporate a person's routine and habits in the search strategy. This work aimed to model the semantic information of how objects are moved over time within an environment and use the inferred information to reduce the searching space. This idea came from observing how the objects are placed over time and that every person has their own singularities in terms of object placement. The contribution of the author of this thesis is in modeling the semantic information as part of the search strategy and in building a heat map with the inferred data.

2. MANTELLI, M. et al. Semantic active visual search system based on text information for large and unknown environments. *Journal of intelligent & robotic systems*, Springer, v. 101, n. 2, p. 1–23, 2021.

Summary and Individual contribution: This paper is on how to find a target door label based on text analysis. Although humans heavily rely on texts for accomplishing several tasks, text as a data source is not very popular in robotics. In this work, we have argued that texts have a great potential for providing search clues and are often found in man-made environments. This idea came from the human behavior

when searching for a someone's office in an unknown building, and how the door labels are used for estimating whether the current corridor is promising for finding the target office. The search strategy relies on the patterns of door labels in indoor scenarios and it reasons over them to estimate which corridor is more promising for achieving the goal.

2 THEORETICAL BACKGROUND

REFERENCES

AYDEMIR, A. **Exploiting structure in man-made environments**. Thesis (PhD) — KTH Royal Institute of Technology, 2012.

CESAR, C. et al. Simultaneous localization and mapping: Present future and the robust-perception age. **arXiv preprint arXiv: 1606.05830**, 2016.

SALAS-MORENO, R. F. **Dense Semantic SLAM**. Thesis (PhD) — Imperial College London, 2014.

APPENDIX A — RESUMO EXPANDIDO

Resolução 02/2021 – Redação de Teses e Dissertações em Inglês Dissertações de Mestrado e Teses de Doutorado do PPGC, bem como outros trabalhos escritos tais como Proposta de Tese e PEP, poderão ser redigidas em inglês desde que contenham um título e resumo expandido redigidos em português. O resumo expandido deve conter no mínimo duas páginas inteiras, deve aparecer como apêndice e deve conter as principais contribuições e resultados do trabalho.