



HAL
open science

Computer Vision and Robot Navigation in 3D Environments

Panagiotis Papadakis

► **To cite this version:**

Panagiotis Papadakis. Computer Vision and Robot Navigation in 3D Environments. Computer Vision and Pattern Recognition [cs.CV]. Université de Bretagne Occidentale, 2023. tel-04083618v1

HAL Id: tel-04083618

<https://theses.hal.science/tel-04083618v1>

Submitted on 27 Apr 2023 (v1), last revised 26 May 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ACCREDITATION TO DIRECT RESEARCH

UNIVERSITY OF WESTERN BRITTANY (FRANCE)

DOCTORAL SCHOOL 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Field : *Informatics*

By

Panagiotis PAPADAKIS

Computer Vision and Robot Navigation in 3D Environments

Presented and defended in Brest, on the 12/04/2023
Hosting research team : Lab-STICC, UMR CNRS 6285

Reviewers before defense

Antonios GASTERATOS	Professor, Democritus University of Thrace, Greece
Simon LACROIX	Director of Research, LAAS/CNRS, France
Tomas SVOBODA	Professor, Czech Technical University in Prague, Czechia

Jury

Reviewers :	Antonios GASTERATOS	Professor, Democritus University of Thrace, Greece
	Simon LACROIX	Director of Research, LAAS/CNRS, France
	Tomas SVOBODA	Professor, Czech Technical University in Prague, Czechia
Examiners :	Cédric BUCHE	Professor, ENIB, IRL CROSSING
	Cédric DEMONCEAUX	Professor, University of Burgundy, France
	Vincent RODIN	Professor, University of Western Brittany, France

ACKNOWLEDGEMENTS

I address my sincere gratitude to all of those with whom I had the chance to meet and cooperate with during my course in various academic and research institutions or consortia. Every environment from which I traversed during this trajectory, had been an opportunity to reinvent myself and gain in maturity through the renewed challenges and concepts to which I was confronted with. Through the following mention I wish to tribute those that were the most influential to my course to this date.

I was particularly honored to be offered in 2016 a faculty position as *Maître de Conférences* in the department of Informatics of Télécom Bretagne, in Brest. I am thankful to the members of the jury that entrusted me with this role, to the members of the hosting team IHSEV for their reception and investment and more broadly of various associates of CNRS Lab-sticc. Among many other, I wish to especially thank André Thépaut, Ioannis Kanellos, Christophe Lohr, Jérôme Kerdreux, Serge Garlatti, Gilles Coppin, Cédric Buche and more recently Andries Mihai for their involvement and support in diverse activities. Together with the state-of-the-art facilities of IMT Atlantique and team RAMBO today, they form an exceptional environment for pursuing cutting-edge research in response to societal and industrial needs.

I am further thankful to my supervisor David Filliat at the time when I was in ENSTA ParisTech, from whom I could greatly benefit via his supervision and expertise in mobile robotics. In the same occasion, I extend my thanks to Freek Stulp who involved me in his teaching activities at ENSTA, as this allowed me to gain my first teaching experience in the French superior education landscape.

Earlier back to my affiliation with INRIA Sophia Antipolis and Grenoble, I wish to acknowledge Patrick Rives, Anne Spalanzani and Christian Laugier for instilling in me a part of their charisma in pursuing research for autonomous vehicles. Federating a project of bilateral technology transfer between the teams Emotion and Lagadic was an invaluable professional experience, in particular, at the forefront of socially compliant robotics research.

My tribute goes equally to my initiating experience abroad upon the completion of my doctoral studies, in the University of Rome "La Sapienza", in Italy, Department of

Computer, Control and Management Engineering, group ALCOR. I am thankful to Fiorra Pirri, leader of the group at that time, for entrusting me with the task of representing the team in the EU project partnership as well as of the teaching of master courses in the fields of computer vision and pattern recognition. During the same period, I had the chance to meet and work with numerous charismatic people, such as Valsamis Ntouskos, Matia Pizzoli, Mario Gianni, Arnab Sinha and last but not least, EU project leader Geert-Jan Kruijff from DFKI, Saarbrücken, in Germany, among many others.

I should also bestow my acknowledgement to my PhD supervisors for discovering and polishing my inclination towards scientific research and innovation.

Finally, I could not have reached as far without the confidence, support and love of my wife Nikolett who knew to surmount with me the difficulties along the way.

TABLE OF CONTENTS

1	Introduction	7
1.1	Biography sketch	7
1.2	Organization of dissertation	9
2	Vision in 3D environments	11
2.1	3D pose normalization	12
2.1.1	Decision-surface representation of 3D shapes	13
2.1.2	The CAnonically Posed 3D Objects (CAPOD) Dataset	16
2.2	Static environment mapping	17
2.2.1	Object discrimination and multi-hypothesis fusion	18
2.2.2	Semantic segmentation of 3D point clouds	22
2.2.3	Incremental learning of objects	23
2.2.4	Passage detection	24
2.3	Dynamic environment mapping	28
2.3.1	Human interactions representation and mapping	29
2.3.2	Activity recognition	32
3	Navigation in 3D environments	35
3.1	Essay on terrain traversability analysis for UGVs	36
3.2	Navigation in static environment	37
3.2.1	Negative obstacle negotiation	38
3.2.2	Off-line traversability learning	41
3.2.3	Reinforcement-learning based staircase negotiation	43
3.3	Socially-compliant navigation	47
4	Transversal works	53
4.1	Emulated USAR exercises	53
4.2	Human-robot USAR mission in Italy	55
4.3	Assistive robot and smart space integration	57

5	Research Perspectives	59
6	Conclusion	65
	Teaching	67
	Direction	71
	Curriculum Vitae	77
	Appendix - Multimedia	91
	Bibliography	93

INTRODUCTION

This dissertation is positioned in the crossroad between computer vision and robot control and documents my principal research works and accomplishments pertaining to my *Accreditation to Direct Research* - translated from *Habilitation à Diriger des Recherches* (HDR) - covering the period from the beginning of my post-doctorate activities in 2010 up to 2022 and my current enrollment as Associate professor in IMT Atlantique and co-leader of team RAMBO, CNRS Lab-STICC. The scope of the dissertation is encompassed by section 27/Informatics of the French National University Council - *Conseil National des Universités* (CNU) - and more particularly, the domain of artificial intelligence.

1.1 Biography sketch

Figure 1.1 depicts the stages of my career path during the concerned period, namely, from 2010 up to time of writing in 2023. The chronology of these works starts in the field of vision, by the culmination of research in 3D shape recognition, a line of research emanating from my PhD [1]. In the beginning of my first post-doctorate in the University of Rome "La Sapienza" in Italy (group ALCOR), I contributed a 4 degree-of-freedom (for translation and scale) pose normalization approach, inspired from the domain of incremental learning and novelty detection [2], [3]. In essence, it amounted to the detection of minor shape parts and articulations and their treatment as novelties as a pre-requisite for more robust shape matching. This later instigated the creation of the first, publicly available ground-truth dataset for canonical semantic pose normalization and evaluation [4]. In conjunction to this fulfillment, my post-doctorate in Sapienza in the 2010 – 2012 period sets a milestone in the extension of my research horizons to the domain of robotics, via my involvement to the EU project NIFTI, for human robot-teaming in Urban Search and Rescue Scenarios (USAR). In the context of this activity, my research was focused in the domain of traversability analysis for unmanned ground vehicles [5] and negative obstacle negotiation [6]. The former work constituted an extensive literature essay that

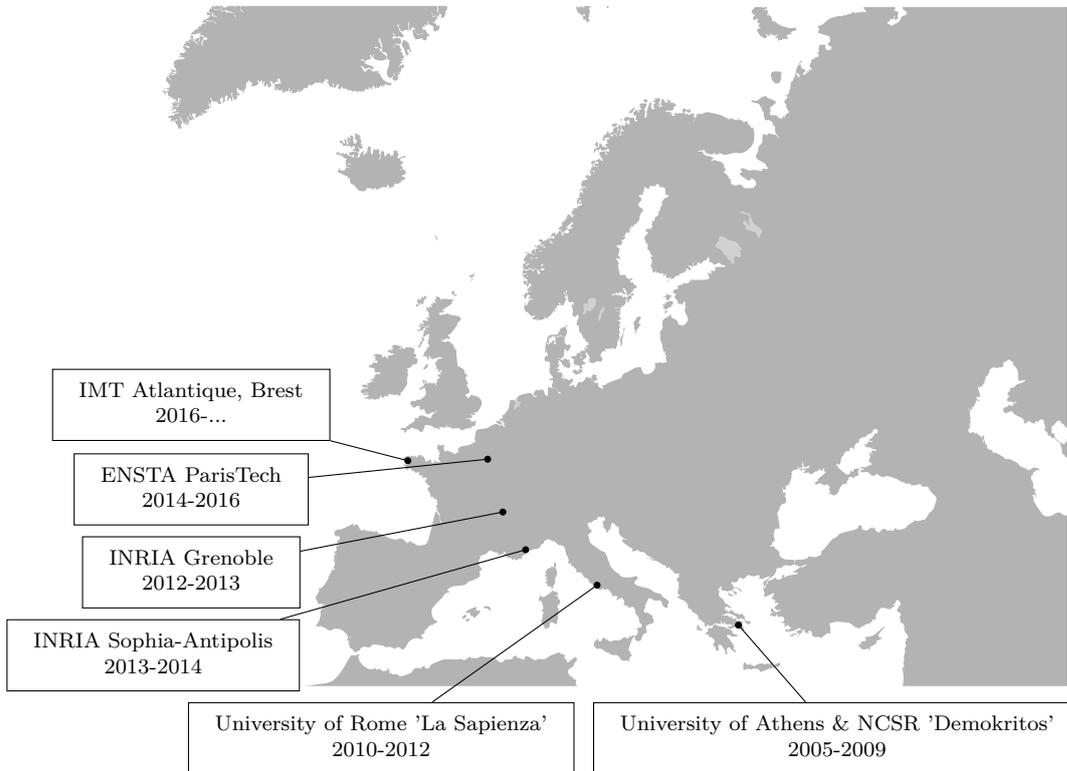


Figure 1.1: Overview of professional sojourns in different research and academic institutions, from doctoral studies up to date.

was undertaken to highlight challenges and promising directions for improving the 3D mobility of ground robots, motivating subsequent project works and further beyond. The pioneering USAR mission of a human-robot team in Europe of which I was the UGV pilot, to the earthquake-hit area of Emilia Romana in Italy [7] in 2012 echoes the eminence of those works in the robotics community and my contribution to their fulfillment.

Following on to 2012 to 2014, I pursued my post-doctorate works within the French research landscape in the context of the multilateral, INRIA research project PAL (Personally Assisted Living), where I federated research activities between the centers of INRIA Grenoble (team EMOTION) and INRIA Sophia-Antipolis (team LAGADIC). The core of my research consisted in the development of socially compliant, indoor robot behaviors, namely, accounting for the presence of humans alongside robot operation. In this context, I proposed an analogy between implicit function learning and human social interaction mapping [8], that could be employed to regulate human-robot encounters [9]. The latter work was seminal in proposing to employ non-stationary, skew-normal probability density

functions for a generic representation of human proxemic distances. On the basis of that framework, I then established a graphical-model based approach [10] for the detection of 3D passages (doors or staircases) in indoor environments, based on spatial human activity cues on top of occupancy maps.

My last post-doctoral sojourn is realized in the period from 2014 to 2016 when I joined ENSTA ParisTech at Palaiseau (team U2IS and team INRIA FLOWERS), conducting research in the context of project ROMEO2, led by Aldebaran Robotics. My involvement contributed to the development of light-weight, unsupervised semantic object mapping [11] that capitalized RGB-D sensory data and to the deduction of scalable, object label fusion under different sensing hypotheses [12].

Since my enrollement in Télécom Bretagne in 2016 and currently IMT Atlantique, my research activities further encompass the use of smart space technologies, deep learning and robot learning more broadly, for vision and navigation in 3D environments. Among them, the most notable recent achievement consists in the development of reinforcement-learning based, staircase negotiation learning and zero-shot transfer to articulated tracked robots [13]. These results are pioneering in terms of task complexity, cross-platform testing and impact in the domain of indoor, 3D navigation learning. Building on top of this work, a software testbed has been developed and made publicly available to the community, allowing to train and benchmark navigation controllers in simulated environments, via incremental domain randomization [14]. Several other parallel works in domains that range from recognition of activities of daily living [15], 3D semantic mapping [16], human posture analysis [17], imitation learning [18] and class-incremental learning [19] attest the central role of deep-learning in my most recent and ongoing research activities.

1.2 Organization of dissertation

The contributions of the dissertation are reflected through its title that reveals from one hand our research scope (robot vision and navigation) and the operational domain of interest (3D environments) on the other hand. The contributions of the manuscript are thus arranged as shown in Figure 1.2.

This arrangement is chosen to facilitate reading based on one’s background, although the limits between vision and navigation tend to become less pronounced in view of the latest scientific advances. This justifies the reason why some contributions presented in Chapter 3 may also concern vision although the task of interest in that chapter is navi-

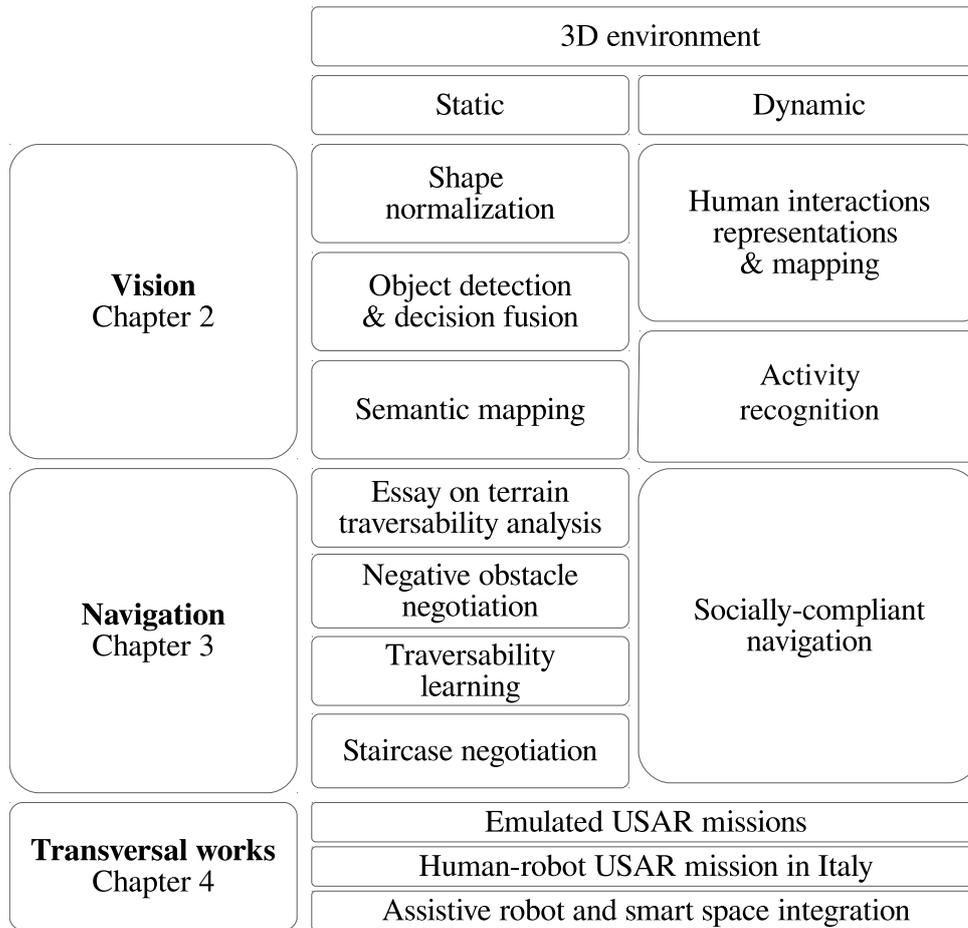


Figure 1.2: Arrangement of the dissertation contributions.

gation and control. Finally, an effort has been made to vulgarize the contents of grouped contributions by employing a presentation structure composed of a list of **Highlights**, a **Problem Summary**, a **Contribution Summary** and a **Results** section, whenever applicable.

Finally, the dissertation presents a description of my research roadmap and perspectives in Chapter 5 followed by a conclusion in Chapter 6, a summary of teaching activities, a summary of coordination-direction activities, a detailed curriculum vitae, a list of multimedia resources and bibliography.

VISION IN 3D ENVIRONMENTS

The works presented in this chapter fall into to the so-called *high-level* vision area of computer vision, as prescribed by the taxonomy set by Forsyth and Ponce in [20], where higher denotes closer to the end of the human (or artificial) visual processing pipeline. Vision in 3D environments covers a broad variety of research topics, from computer graphics, sensing, modelling, rendering, visualization and applications such as autonomous systems, robotics, cultural heritage, virtual and augmented reality. The growing plethora of benchmarks dedicated to 3D data (e.g. [21], [22], [23], etc.), software tools, consumer devices with 3D capturing capabilities or 3D printing, highlight a continuing scientific and industry-wise interest for technologies that rely upon the exploitation of 3D data via vision.

The works presented in this chapter are more closely linked to cognitive tasks such as object detection and recognition and scene understanding more globally. These tasks may not strictly require an accurate 3D reconstruction of the environment but can benefit if it is available. In this context, it is assumed that the reconstruction of 3-D models of objects from a collection of 2D images (cf. Ma et al. [24]) is already available and exploited otherwise vision is performed independently in each scene sample.

The mode of presentation of the contributions has been chosen such that the various contributions are described with an increasing order of challenges encountered in terms of: (i) quality, quantity and dimensionality of the processed data and (ii) degree of uncertainty/entropy. I begin with **Section 2.1** by presenting a novel representation of 3D objects that is more adapted to translation and scale normalization, by being robust to articulations or outlying parts that are known to bias the canonical pose of a 3D object. Object normalization often constitutes a crucial step in tasks such as registration, recognition, view generation/sampling, etc. This contribution follows directly after the end of my PhD and completes my earlier contributions on rotation normalization. The effectiveness of the approach is demonstrated quantitatively, by improving object search and retrieval performance from digital libraries.

Following on to **Section 2.2**, I present earlier works pertaining to robot vision and static environment mapping, by making use of photo-metric and/or geometric sensory data of real scenes, in the context of robot exploration. This part treats aspects related to representation of multi-modal sensory data, noise filtering, recognition, continual learning of objects and accounting for robot or environment motion in the decision process, and serves as a precursor to the last **Section 2.3**, that deals explicitly with mapping of a dynamic environment, as a consequence of human presence and activity alongside robot operation. The term mapping is employed to denote the fact that the vision task is either performed directly by a mobile robot or that the outcome of the mapping task is destined to be used for the purpose of robot navigation in human-robot encounters. We thereby coin the term *social* mapping to differentiate it from metric, topological or semantic mapping that typically concerns the static environment.

2.1 3D pose normalization

3D pose normalization accounts for the normalization of the *translation*, *scale* and *rotation* of 3D objects, namely, of the 7th degree-of-freedom (DOF) object pose represented as a similarity transformation (cf. Table 3.2 in [25]). The need of normalization stems from the fact that the true pose of 3D objects is generally considered unknown or set with respect to an arbitrary reference coordinate frame.

It is therefore necessary, for example, in content-based retrieval when shape matching is based on the establishment of correspondences between the surfaces of objects and achieving invariance to isometric transformations, or further up to non-isometric transformations if scale is irrelevant. Alternatively, it can be used to reduce the search space for obtaining a registration between two shapes, by providing an initial rough solution, favoring the convergence of Iterative-Closest-Point (ICP) methods [26]. Finally, it can also serve in Deep Learning (DL) and neural network architectures operating on voxel-grids that lack the property of invariance to 3D pose.

My contributions in this field concern a method (see section 2.1.1), a prototype ground-truth (see section 2.1.2) and reside on publications [3], [2] and [4].

2.1.1 Decision-surface representation of 3D shapes

Highlights

1. The canonical object translation and scale are computed by discarding object outliers or articulations by one-class SVMs.
2. The improved object surface correspondence favors content-based retrieval in terms of precision and recall.

Problem summary Translation normalization is ordinarily addressed by computing the centroid of a 3D object and by translating it to coincide with the coordinates origin. For the majority of 3D objects, this method is effective and objects are well normalized with respect to translation. However, there are cases where this method is not appropriate such as articulated 3D objects or objects with isolated-outlying or extruding parts. Such cases are demonstrated in Fig. 2.1 (a) and (b) respectively. We would intuitively agree that the centers of the two objects in either case should coincide (in a semantic context). However, by setting the geometric centroid of the surface of an object as its center, we obtain a result that is perceptually inconsistent. As it is evident from the examples, a small change in the

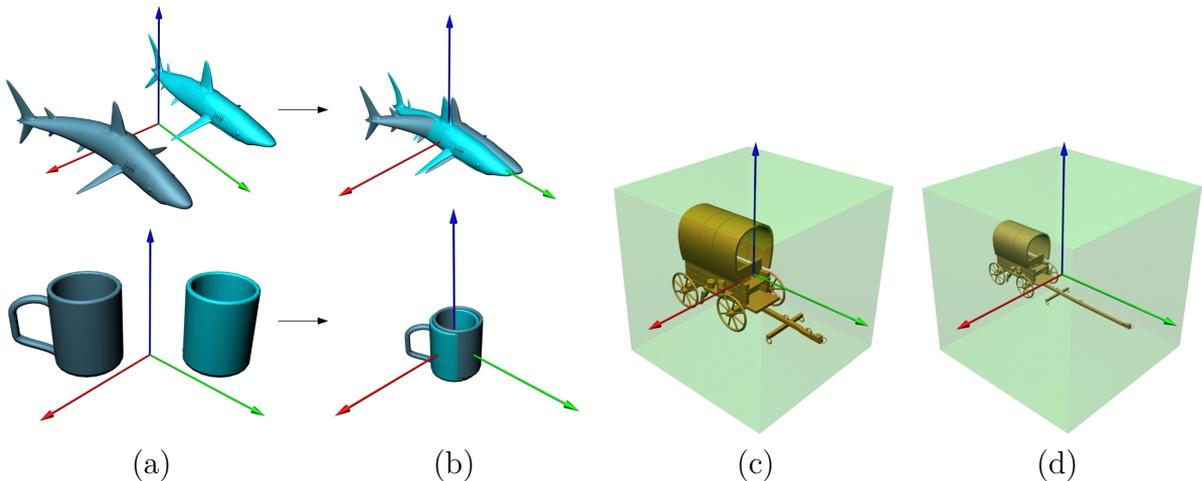


Figure 2.1: (a) 3D object pairs before normalization and (b) after conventional translation and scale normalization, (c) object whose scale is normalized to fit in the unit cube and (d) scale normalized version of its deformed version with elongated carriage axis.

shape of 3D objects can have significant impact on the resulting translation normalization which may in turn negatively affect scale and rotation normalization. Clearly, we expect that the perturbation of the centroid of the surface of an object becomes more pronounced as the amount of articulation or extrusion increases.

Scale normalization is ordinarily addressed by setting the scale of a 3D object to fit within a bounding volume of fixed size (unit cube or unit sphere), or to a size proportional to the average distance of the surface of an object from its centroid. The performance of the first method is problematic in the presence of outlying parts that will cause an undesired shrinking of the entire object in order to make it fit within the bounding volume. A characteristic example is shown in Fig. 2.1 where the objects are scaled so as to fit within the unit cube. The object in Fig. 2.1 (d) is the same in every aspect as the object in Fig. 2.1 (c) not considering the elongated carriage axis. However, the scale normalized versions of the two objects are significantly different.

Contribution summary We posit the following hypothesis:

The vast majority of objects is characterized by a central, dominant component that can be viewed as a single class to which we seek to determine which 3D object points belong, the rest being considered as outliers or novelties.

Therefore, instead of accounting for the entire distribution of 3D points belonging to an object, its center is taken as the centroid of a 3D distribution of points estimated by employing a one-class detection method.

The desired distribution is delimited by the boundaries of a decision function $f_s(\mathbf{x}) = \text{sgn}(f(\mathbf{x}))$ in the observation space \mathbb{R}^3 , where sgn is the sign operator, giving +1 if the input is positive, -1 if it is negative and 0 otherwise.

The centroid of the distribution is obtained by:

$$\mathbf{c} = \frac{1}{V} \int \int \int_{f_s(\mathbf{p}) \geq 0} \mathbf{p} dx dy dz \quad (2.1)$$

where $\mathbf{p} = (x, y, z) \in \mathbb{R}^3$ and $V = \int \int \int_{f_s(\mathbf{p}) \geq 0} dx dy dz$.

We tested this idea using One-Class Support Vector Machines (OC-SVM) [27], though other methods could be analogously employed. In the case of OC-SVM, the unknown function $f(\mathbf{x})$ is obtained by solving for the corresponding optimization problem (cf. eq. (1)-(2) in [3]). The two parameters γ and ν of OC-SVM can be used to control: (i) the

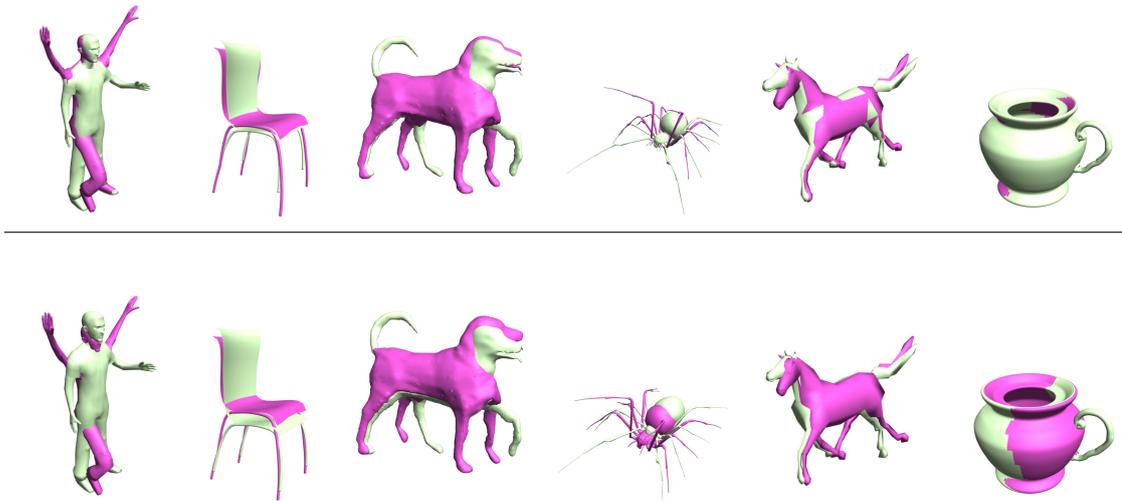


Figure 2.2: Comparison of translation and scale normalization using OCSVM (top row) against conventional use of the centroid and the mean surface distance (bottom row).

over-fitting of the decision surface to the original distribution and (ii) the degree by which we include or discard object outliers, extrusions or articulations (cf. Figure 4 in [3]).

After translation normalization, we can use the computed distribution that corresponds to the 3D object to perform scale normalization. To normalize the scale we estimate the average distance d of the distribution of the object from its center and scale the object so that this distance is unit. The average distance d is computed as:

$$d = \frac{1}{V} \int \int \int_{f_s(\mathbf{p}) \geq 0} \|\mathbf{p} - \mathbf{c}\| dx dy dz \quad (2.2)$$

where $\|\cdot\|$ denotes the L_2 norm.

Results The improved object surface correspondence leads to improved object matching in content-based retrieval in three different benchmarks, with an average performance as shown in Table 2.1 (cf. Table 1 of [3]). Figure 2.2 shows indicative examples comparing conventional translation and scale normalization against the proposed approach.

Method	Nearest-Neighbor	first-tier	second-tier
PANORAMA-OCSVM	92.7%	58.8%	69.9%
PANORAMA	91.4%	56.8%	67.3%

Table 2.1: Average performance of content-based retrieval employing OC-SVM based normalization (PANORAMA-OCSVM) against conventional retrieval (PANORAMA [28]).

2.1.2 The CAnonically Posed 3D Objects (CAPOD) Dataset

Problem summary The gain of performance in content-based retrieval provided an indirect way for promoting my proposal for robust translation and scale normalization. This was due to the fact that at the time of publication of those results, there was a lack of a test-bed or ground-truth for objects of the same class that appear in their canonical 3D pose, that prevented the comparison of alternative methods for pose (translation, scale and rotation) normalization.

Contribution summary To construct such a test-bed, we had to alleviate the subjectivity in semantics and error proneness issues of a conventional approach where 3D objects are first collected, then classified and finally pose normalized. Instead, the construction protocol is summarized as follows:

1. Selection of categories of 3D objects that span a desired semantics range which depends on the desired categorization resolution.
2. Acquisition of a reference 3D object for each category which captures the semantics of the particular class.
3. Generation of 3D object instances for each class by applying permissible modifications of the corresponding reference object, such as articulations, extrusions, suppressions, additions and deletions of shape parts (see Figure 2.3).

The CAPOD dataset [4] contains 180 object instances uniformly distributed in 15 classes and is downloadable from <https://sites.google.com/site/pgpapadakis/CAPOD>.

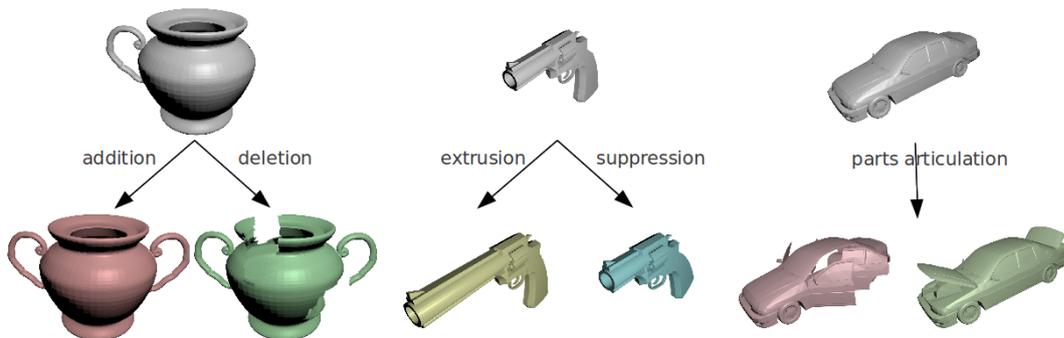


Figure 2.3: Repertoire of shape deformations used to produce object class members.

2.2 Static environment mapping

Mapping an environment allows robots to be deployed in previously unknown workspaces, marking this skill as a primary element in the integration of robots into real-world, human-populated environments. Conventionally, the problem of environment mapping is treated through a decomposition into layers where each represents a desired abstraction level, ranging from the low-level 2D/3D metric Simultaneous Localization and Mapping (SLAM) [29] and topological [30] to the semantic [31] and instance-based level. Also, the process of mapping is incremental following the acquisition of robot data and needs to account for noise, outliers as well as novelties that can be encountered in the unknown environment.

As such, mapping is generally a computationally demanding process and the role of the robot in the respective environment (e.g. obstacle avoidance or negotiation, search or exploration, surveillance or inspection, etc) prescribes a specific, constrained type of mapping to be performed, along with the sensors to be used.

Besides [32] which addresses dense metric SLAM, the various contributions presented in the following sections are all concerned with the highest level of abstraction in environment mapping, namely semantic or instance-based and reside on publications [19], [10], [6], [11], [12], [16], [33] and [34], following my inaugurating work in the domain on topological mapping [35]. They were further promoted via the following national or international presentations:

- P. Papadakis, *Semantic Mapping for Indoor Environments*, Journées Nationales de la Recherche en Robotique (JNRR), 2015, Cap Hornu, France
- P. Papadakis, *From shape descriptors to real world object-based mapping in robotics*, 2015, LAAS-CNRS, Toulouse, France
- P. Papadakis, *Object Classification in Indoor RGB-D Scenes using Projective Light Diffusion Image Expansion*, GDR-ISIS, 2016, Télécom ParisTech, Paris, France
- P. Papadakis, *Object Classification in Indoor RGB-D Scenes using Projective Light Diffusion Image Expansion*, Lab-STICC, 2016, Télécom Bretagne, Brest, France
- P. Papadakis, *Generic Object Discrimination for Mobile Assistive Robots using Projective Light Diffusion & Binding Human Spatial Interactions with Mapping for Enhanced Mobility in Dynamic Environments*, University of Adelaide, Adelaide, Australia, FASIC Workshop 2018.

2.2.1 Object discrimination and multi-hypothesis fusion

Highlights

1. Early fusion; we propose a novel, discriminative object representation termed as projective light diffusion image (PLDI) which can encode surface orientation, depth as well as photometric characteristics.
2. Late fusion; we compare multi-view hypothesis fusion schemes for 3D object classification under observation (in)dependence and in consideration of the robot observational capacities.

Problem summary **3D object discrimination** is indispensable for indoor robots. The fact that skills such as manipulation, semantic mapping and navigation are directly dependent on this capability explains its prevalent role, further promoted by the advent of sensors that provide synchronized photometric and geometric (RGB-D) information. The preponderance of RGB-D data for object perception can be observed by the creation of numerous benchmarks [36], [37] and international contests [38]. Despite the abundance and richness of information captured by RGB-D data, however, there are several challenges to account in RGB-D based object perception by robots, such as differences in domain statistics between colour and 3D data or real and synthetic data, quality and sparsity differences, scene occlusions, not to mention limited on-board computational power. Figure 2.4 shows such problems in different examples.

Differences in domain statistics imply that objects that are similar in 3D geometry might diverge visually and vice versa or that value ranges are different between pixel intensities and depth. Then, the quality of color data is consistently superior compared to 3D and of higher resolution, while 3D data are easily corrupted by environmental noise or sensing conditions, yielding either noisy depth estimations or undetermined (null) values [39]. Finally, a captured color image is intrinsically associated with the 2D coordinate frame of the picture and a regularly spaced grid, whereas a 3D point cloud can be reviewed from any angle and does not impose a space discretization.

These non-trivial differences instigated the majority of researchers to favor distinct pipelines for RGB and 3D data processing, opting for *late fusion* of the two pipelines, i.e.

in the classification stage. Alternatively, to alleviate these challenges a robot could ideally rely on light-weight and at the same time discriminative RGB-D object representations, that capitalize complementarities between colour and geometry and are robust to noise.

Finally, due to limits in the sensor field-of-view, occlusions or obstacles, partial and more broadly multi-view based object discrimination is crucial in robot applications. In turn, inspecting an object from different view-points requires the development of **multi-hypothesis** or **decision-level fusion** mechanisms (alternatively referred as late fusion), as not all viewpoints are equally informative and each introduces its own bias to a certain class. The performance of a late-fusion approach depends on the underlying data distribution and object class semantics but also on the geometric dependency of viewpoints, extracted either via knowledge about sensor or object motion.

Contribution summary For the first problem pertaining to 3D object discrimination, we adopted an *early fusion* approach for color and depth in order to : (i) exploit local correlation existing between image intensities and 3D geometry for noise filtering and edge preservation and (ii) reduce processing time by subsequently processing a single representation for feature extraction instead of separate pipelines.

We use the 4-channel image of an RGB-D frame as a filtering guide (see Figure 2.5 (a)) and employ three kernels that operate within the color, depth and the 2D image domain space respectively. The operation resembles that of the conventional bilateral filtering but differentiates with respect to the input and desired output image, known as joint bilateral filtering (JBF) [40]. We employ JBF by treating the RGB-D image as a guide in order to obtain a reconstructed depth and ignore pixels with undetermined depth (see Figure 2.5

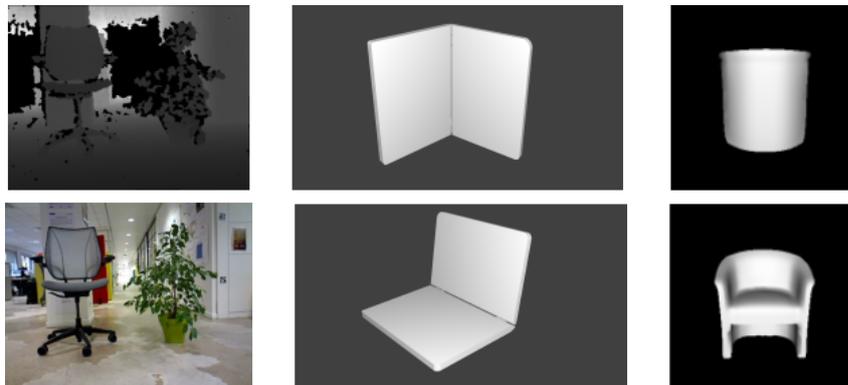


Figure 2.4: Challenges encountered in RGB-D object perception induced by domain differences (left) or viewpoint-dependent class interpretation (middle and right)

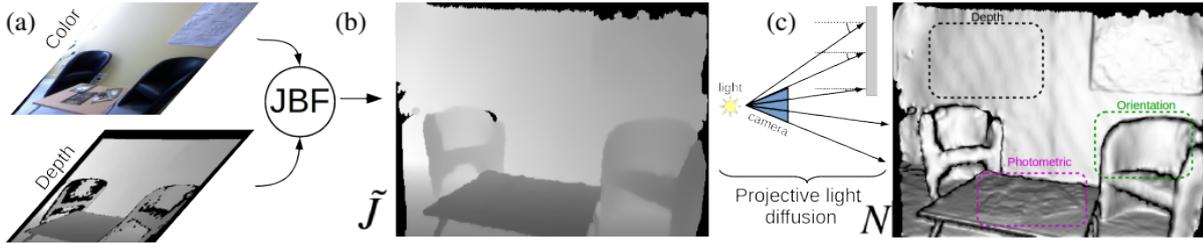


Figure 2.5: Early fusion/filtering pipeline; (a) Raw depth image and RGB image, (b) reconstructed depth image and (c) projective light diffusion image. Emergent depth, photometric and surface orientation scene characteristics captured by PLDI are partially highlighted for emphasis (best viewed in color).

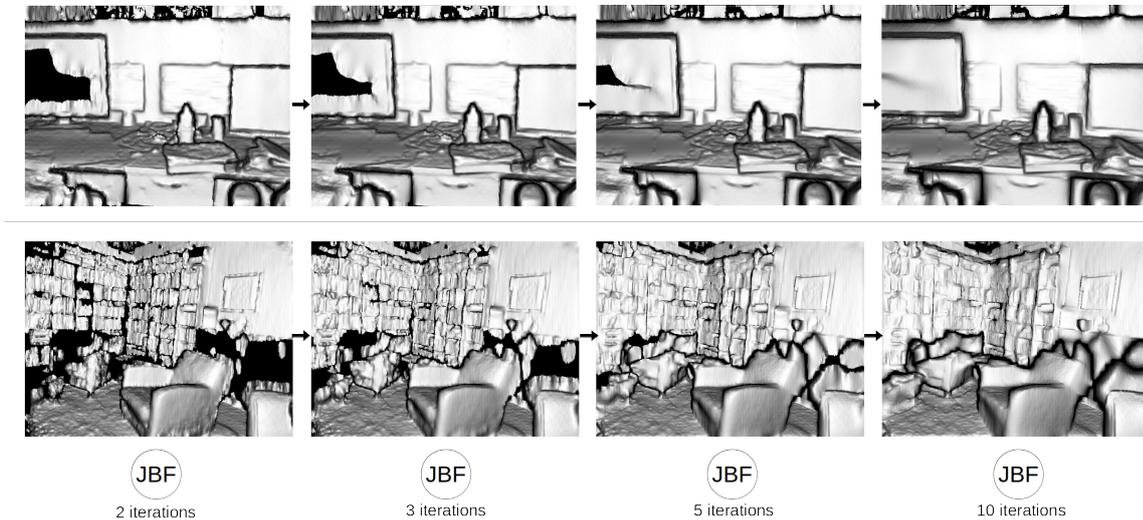


Figure 2.6: Iterative JBF-based filtering for progressive 3D reconstruction

(b)). Multiple iterations of the JBF filtering can be performed in order to extend depth reconstruction until there are no missing depth values (see Figure 2.6). Finally, the 3D surface orientation of the reconstructed depth image is computed by backward projection of depth pixels, computation of the surface orientation tensor and light source emulation at the center of the camera (see Figure 2.5 (c)), making use of the known camera intrinsics.

We term this process as *Projective Light Diffusion* and the resulting image as PLDI. Thanks to the use of RGB and depth kernels in the filtering and depth reconstruction and the derivation of the 3D surface orientation with respect to the camera, we manage to jointly encode depth, photometric and surface orientation features in a single-channel 2D image that we can subsequently use for object discrimination.

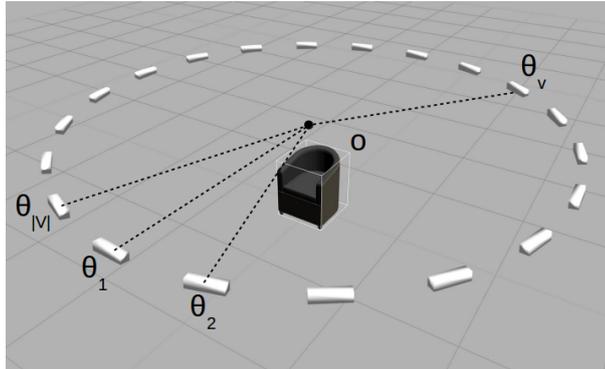


Figure 2.7: Example object and multi-view configuration

For the second problem pertaining to decision-level fusion, we pursued a formulation of the problem in the form of a Hidden Markov Model (HMM). Accordingly, we defined : (i) an *observation model* that jointly assessed the class and the particular viewpoint from which the object is observed, i.e. the state and (ii) a *state transition model* that is normally distributed and centered around the rotational displacement of the sensor around the object or vice versa. We then compared the performance of 4 in total, class inference mechanisms of the HMM for an observed object, using a PLDI-based descriptor.

Results A PLDI-based shape descriptor was proposed and compared against competitive descriptors for 3D shapes, namely, Viewpoint Feature Histogram (**VFH**) [41], Ensemble of Shape Functions (**ESF**) [42] and Signature of Histograms of Orientations (**SHOT**) [43] and **RoPS** (Rotational Projection Statistics) which corresponded to the most informative local 3D shape descriptor based on [44]. Testing the PLDI descriptor on a robot with an RGB-D sensor as well as in **Princeton ModelNet10** [45] revealed a superior discriminative power, in view of its low dimensionality (cf. Tables VI-VII of [11]). Video # 4 of the Appendix - Multimedia provides qualitative examples of performance of the corresponding work.

Using the PLDI-based descriptor for object discrimination in single RGB-D scans, we devised 4 decision fusion schemes that could account for knowledge of robot/object motion and employ a voting or winner-takes-all rule (cf. **HMM-dep**, **HMM-indep**, **MAXS** and **MAXC** schemes in [12]). Experiments were performed on the ModelNet10 dataset and for a range of 1 – 6 viewpoint observations. Knowing that this benchmark contains the 10 most frequently encountered object classes in indoor environments according to [46], the main finding of the evaluation is of particular importance for 3D robot vision, namely :

Optimal multi-view object classification resides on observation dependency for the first couple of views and on observation independence for the rest.

2.2.2 Semantic segmentation of 3D point clouds

3D point cloud data are nowadays ubiquitous in numerous areas such as remote sensing, autonomous vehicles, cultural heritage, video games and more recently, consumer cell phones, largely thanks to the development of lightweight sensors and accelerated embedded processing. A complete semantic analysis of a scene (panoptic segmentation) consists in discovering the underlying structure of classes (semantic segmentation) and object instances (instance segmentation) that yielded a certain spatial configuration [47].

Similar to images, 3D point cloud segmentation consists in appointing a class label to each 3D point of the scene [48]. Besides common challenges with image segmentation, 3D point clouds are characterized by higher levels of noise, non-uniformly distributed data and higher differences in the scale/size of the contained objects. Last, but not least, the task of annotating 3D point cloud data is more arduous which reduces their availability compared to images.

Despite noticeable progress in indoor environments or urban scenes where 3D point cloud data are more abundant, emerging areas of interest such as industrial site inspection or remote sensing need to tackle data scarcity and poorer quality. In [16], we identified such major bottlenecks for the semantic segmentation of 3D point clouds of industrial sites, namely, increased relative object sizes combined with limited or underrepresented scene layouts. The very poor generalization ability of current neural-network architectures suggested that before attempting to learn to segment 3D point clouds, we should revise the use of the underlying ground-truth data.

Towards this goal, in [49] we promoted the use of synthetic 3D models for data augmentation via a protocol that generated more realistic scene layouts and 3D point sampling, combined with an artificial coloring learned from unlabeled, real colored data. In this way, we sought to reduce the domain shift between training data and the targeted application data, tackle data imbalance and instill implicit knowledge (color) in the parameters of a neural-network that could help in shape discrimination in cases where 3D geometry only fails. Quantitative experiments in the **S3DIS** [50] dataset highlighted the overall positive effect of the previous steps in semantic segmentation for indoor scenes while an ablation study revealed the impact of different set-ups. A qualitative example of semantic segmentation using the proposed approach is provided in Figure 2.8.

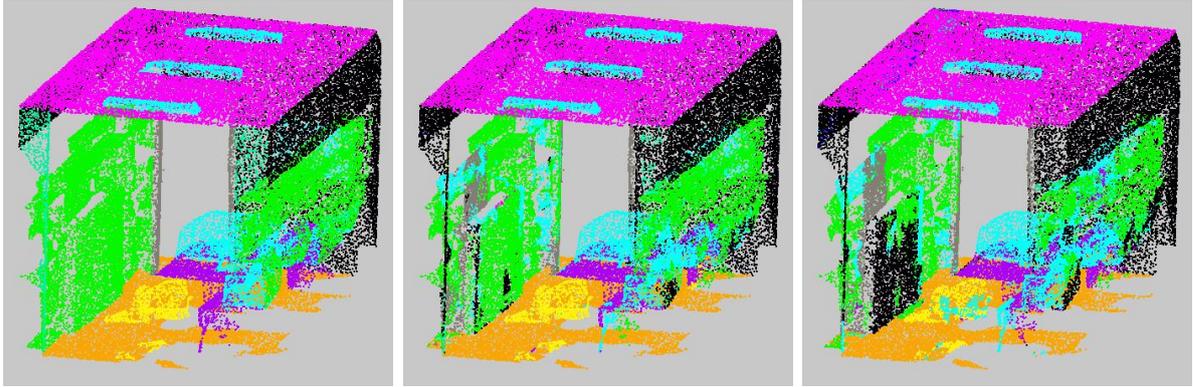


Figure 2.8: Semantic segmentation of an indoor scene. From left to right, ground-truth, segmentation using the artificially colored point cloud as input and segmentation using the colorless point cloud.

2.2.3 Incremental learning of objects

Beyond the discrimination of objects to a predefined set of categories, several real-world applications are characterized by partial knowledge of the set of categories that are encountered and/or by computational constraints that limit the capacity of memorization to a fixed set of exemplars or classes. Typical scenarios concern drones and robots that need to operate with limited on-board resources and/or asynchronous connectivity to external ones. In this context, objects and classes need to be learned incrementally keeping a balance between previously and newly acquired knowledge, a problem in neural networks known as catastrophic forgetting [51]. Various techniques have been since introduced to tackle this problem [52], namely, experience/memory rehearsal, correction of bias due to task recency and knowledge distillation.

We proposed a contrastive learning-based, cross-entropy and knowledge distillation loss used along-side their conventional versions, to learn better features as well as preserve richer knowledge of the previous tasks [19]. This leads to four losses that are linearly combined during each incremental step by accounting for the ratio of the number of old classes to the total number of old and new classes. Experiments performed in CIFAR-100 [53] and ImageNet-100 [54] benchmarks against competitive methods **iCarL** [55], **MDFCIL** [56] and **CO2L** [57] show that the proposed method attains a superior average performance in terms of top-1 accuracy (see Figure 2.9).

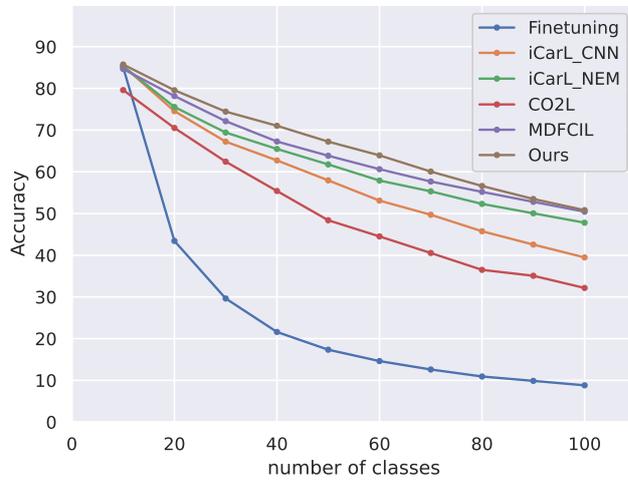


Figure 2.9: Mean accuracy over 10 training trials with random class orderings, as a function of the number of incrementally learnt classes

2.2.4 Passage detection

Highlights

1. We perform global and partial passage detection to deal with variation in object scale or robot proximity.
2. We develop graphical model-based passage detection to jointly account for static map and human activity cues.
3. We develop a gap detection and traversability assessment algorithm.

Problem summary The perception of passages demands particular attention in static scene mapping as such structures are often poorly or partially observed due to their increased size or obstruction by humans. With only partial and noisy data, it can become difficult to assess the true size and geometric pose of a passage such as a door or a staircase, information that is essential in robot navigation. The presence of a passage may also not be bind to salient, visible characteristics but may merely be derived by tracing a human trajectory or be the consequence of an opening towards a spacious area (e.g.

an opened door). In certain robotic applications a *negative* obstacle such as a gap that is sufficiently small to be traversed can also be interpreted as a passage between two disjoint areas. In this sense, passage detection can be a highly contextual task depending on static as well as dynamic surroundings, i.e. human presence. In the latter case, the question to be answered concerns how to make constructive use of human activity cues which unavoidably interfere in the process of static environment mapping.

Contribution summary We investigated different use-cases where a robot perceives conventional passages such as staircases and doors or negative obstacles such as gaps that could be conditionally traversable by a robot.

In the case of conventional passages, we proposed a RANSAC-based approach for detecting dominant staircase structures in 3D [33] and a complementary rule-based algorithm for extracting local, stair-like structures whose clusters can agglomeratively form staircases of more general shape and of arbitrary size [34]. Due to the increased scale of staircases, we note that both paradigms of *global* as well as *local* staircase perception were deemed equally useful, as a robot may encounter a staircase from a far distance as well as come across it at very high proximity before traversal. Another motivation for pursuing both paradigms is that changes in scale as a function of distance further impact the quality and the quantity of 3D data captured by an RGB-D sensor.

Going beyond using cues of the static environment only for perceiving passages, we subsequently proposed graphical-model based passage (staircase and door) detection [10] that jointly accounted for metric map (cf. $M(\mathbf{u})$ function in Figure 2.10) and human activity cues (cf. $IS_{\mathbf{p},t}(\mathbf{u})$ function in Figure 2.10) residing on probabilistic model of proxemics as presented in [9]. Finally, subsidiary probabilistic graphical models were developed, to either update space occupancy from human trajectory tracking or filter false positives of detected humans.

For the case of negative obstacle detection and their potential assessment as passages, we contributed the first method at that time in the domain of Urban Search & Rescue (USAR) [6]. In such adverse, hostile and unconstrained environments, robot operation requires reliable positive as well as negative obstacle perception to avoid accidents. Occasionally however, certain positive or negative obstacles may be negotiated (i.e. be traversed) with safety. To allow real-time performance and timely prevention of unrecoverable robotic states, our approach relied on morphological image processing for noise reduction and border following for the detection and grouping of gaps. Upon the detection

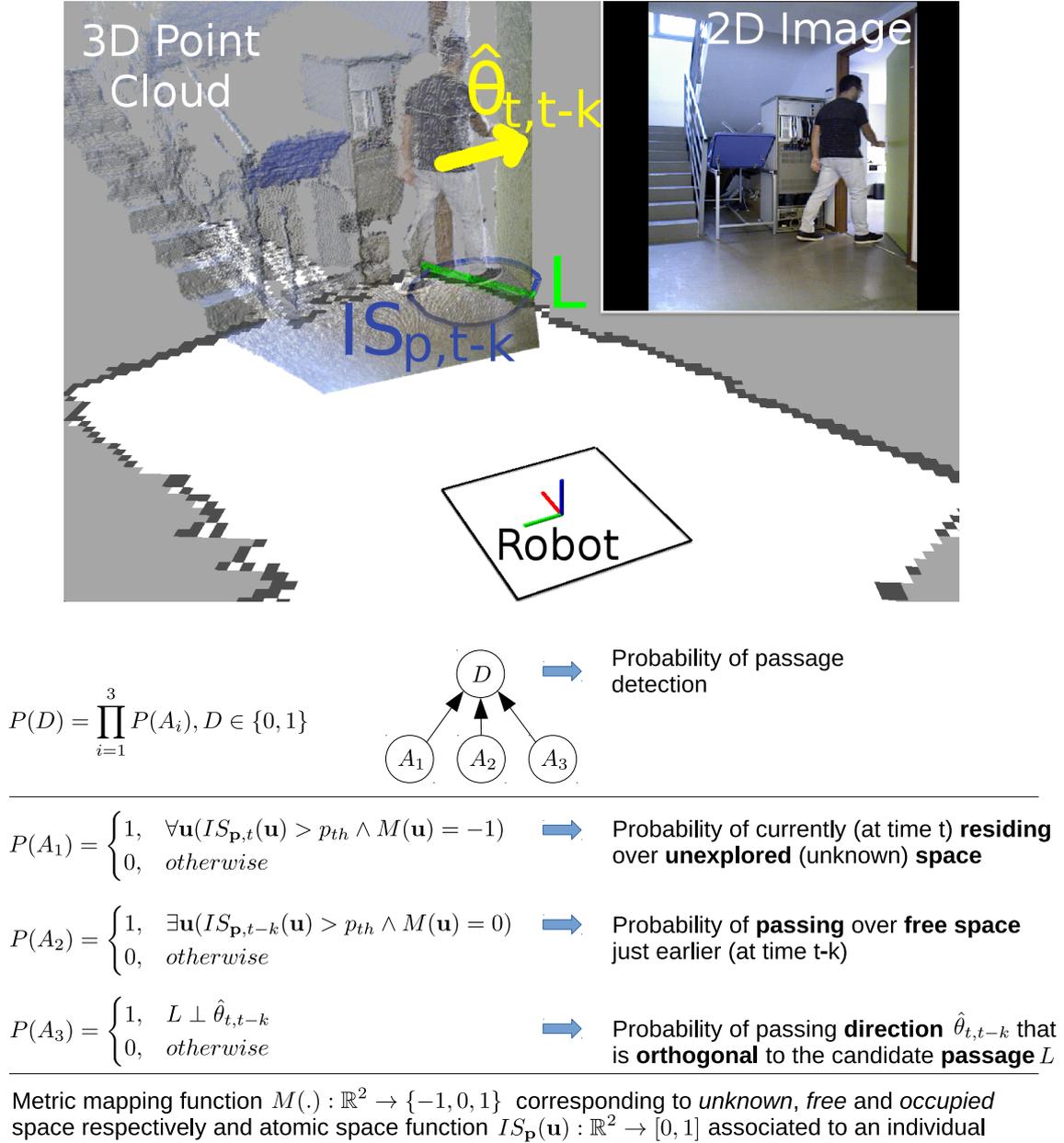


Figure 2.10: Detection of passages via probabilistic graphical models.

of a gap its traversability is assessed as described in Section 3.2.1.

Results Conventional passage detection was performed in a multi-floor office building of INRIA research center in Sophia Antipolis, using a Neobotix MP-500 robot running the

Robot Operating System [58] and equipped with Asus XtionPRO Live cameras capturing synchronized RGB-D images. The accompanied video # 5 of the Appendix - Multimedia shows a part of the experiments on the base floor where the robot detected passages through the probabilistic graphical model-based approach, jointly accounting for the static map and for human activity. Indicative snapshots of staircase perception using all three different approaches are provided in Figure 2.11. Note that in the case where a staircase is detected globally, detection is accompanied with the corresponding inclination and/or size parameters.

Experiments on negative obstacle detection and assessment as potential passages that a tracked mobile robot could traverse, were mostly conducted in the MonteLibretti firefighter training facility in Italy, recreating the conditions of a car accident within a tunnel along with debris (cf. Figure 6 of [6]). Variation of different aspects affecting detection such as processing resolution, distance from sensor, robot pose error or occlusions allowed to assess the robustness of our approach in different conditions.

These experiments yielded a 100% true positive rate, namely, that true gaps were always detected, although occasional false positives could appear depending on noise or other perturbations. This was an acceptable trade-off for the application of interest where missing a true gap could be fatal for the robot while on the other hand false positives could only make the robot more conservative. Figure 2.12 provides a snapshot of video # 8 (see Appendix - Multimedia), visualizing the 3D model of the actual tracked robot within a 3D scan of the tunnel area. It is worth noticing that in contrast to positive obstacles that protrude inside a scene, negative obstacles are invisible and far more subtle.

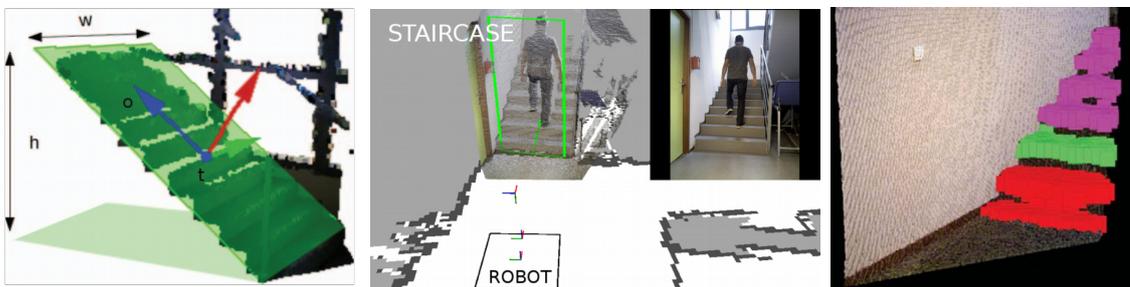


Figure 2.11: Detection of staircases following different approaches: (Left) based on dominant 3D plane extraction, (middle) joint analysis of human trajectory on static map and (right) local-features computation for partial observation.

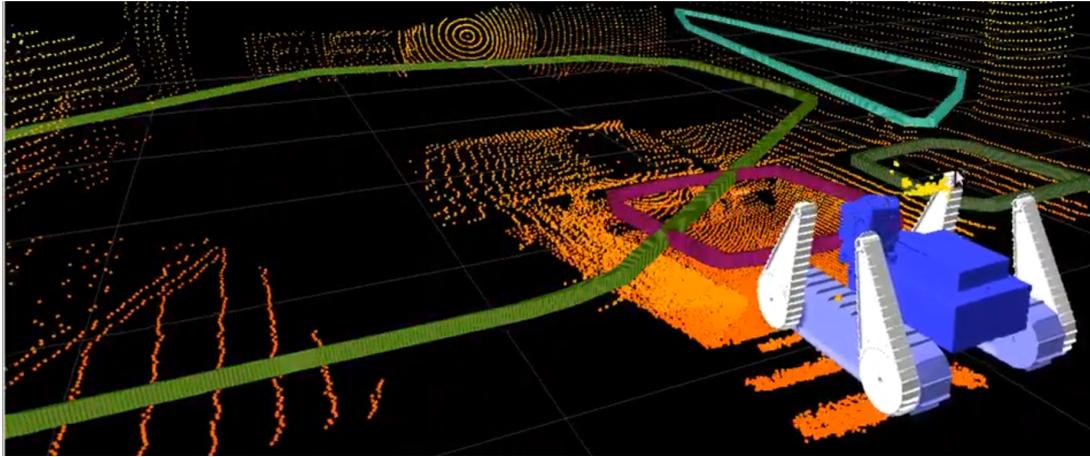


Figure 2.12: 3D visualization of tracked mobile robot and its 3D scanning of a tunneled site, along with detected negative obstacles delimited by different contours.

2.3 Dynamic environment mapping

Following static environment mapping, the next challenge in the deployment of a robot in human populated environments concerns the capacity to perceive and adapt to the dynamics induced by human presence. This is essential for human safety and acceptability of the robot and analogously for ensuring safety and adequate performance of the robot. Beyond dynamics originating to humans, robots and autonomous vehicles more broadly need further account for environment dynamics in case they operate outdoors and for prolonged time periods, for example due to weather or daylight variations. The common denominator is that mapping is a continuous, incremental and interminable process for representing environment geometry and appearance along with human state and intent.

My scope will be on human induced dynamics only, thus mapping the dynamic environment of a robot involves from one hand the problem of detection, tracking and analysis of human individuals or groups and on the other hand transposing this information into robot-centered representations associated to its tasks. This makes it a theme of trans-disciplinary research at the cross-roads of computer vision, sociology, machine learning and robotics, with the potential to foster significant advances in human-robot collaboration in fields such as the Industry of the Future (IoF), Hospital of the Future, Ambient Assisted Living (AAL) and the tertiary sector via service robotics applications.

The contributions presented in the sequel relate to selected matter from publications [8], [9], [15], [59], [60] and [61] and were partially invoked in the next communications :

- P. Papadakis, *Construction de verité terrain d'activités de la vie quotidienne au sein du Living Lab de l'IMT-Atlantique*, Le Couvent des Jacobins - Centre des Congres de Rennes Métropole, Journee "Connecter les Usages" - Connectathon, 2019
- Tébéo - television channel, *Chemins de Traverse* series, episode 20 – 03 – 2019, AMUSAAL (Analyse automatique de Mouvements hUains par décomposition en mouvements Simples pour l'Ambient Assisted Living) project interview

2.3.1 Human interactions representation and mapping

Highlights

1. We employ non-stationary, skew-normal probability density functions for representing dominant proxemic models of atomic social sensitivity.
2. We calculate a social map of human spatial interactions of varying size, shape and cardinality, by an implicit function of proxemics relying on one-class density estimation.

Problem summary Human presence and activity alongside robot operation leads to a shared workspace between the robot and the human. If treated adequately, this can create conditions for *synergistic* operation between humans and robots, otherwise *antagonistic* behaviors will emerge leading unavoidably to failures, discomfort or worstly, accidents.

Favoring synergy between humans and robots is a two-fold problem as the robot needs to be sufficiently competent to perceive and analyze human behavior and by extension, robot behavior should be sufficiently legible and unambiguous to be apprehended by humans. This involves the development of robot perception and action skills that model the factors which govern humans positioning in the shared workspace, in order to enable a robot to either initiate a human-robot interaction or avoid disturbing. Finally, these skills should be applicable to individuals but also generalize to groups of arbitrary size.

Contribution summary Based on studies on social sensitivity (cf. proxemics theory [62]) and its expression in distance preservation among individuals [63], we proposed the use of non-stationary, skew-normal probability density functions [64, 65] for a probabilistic

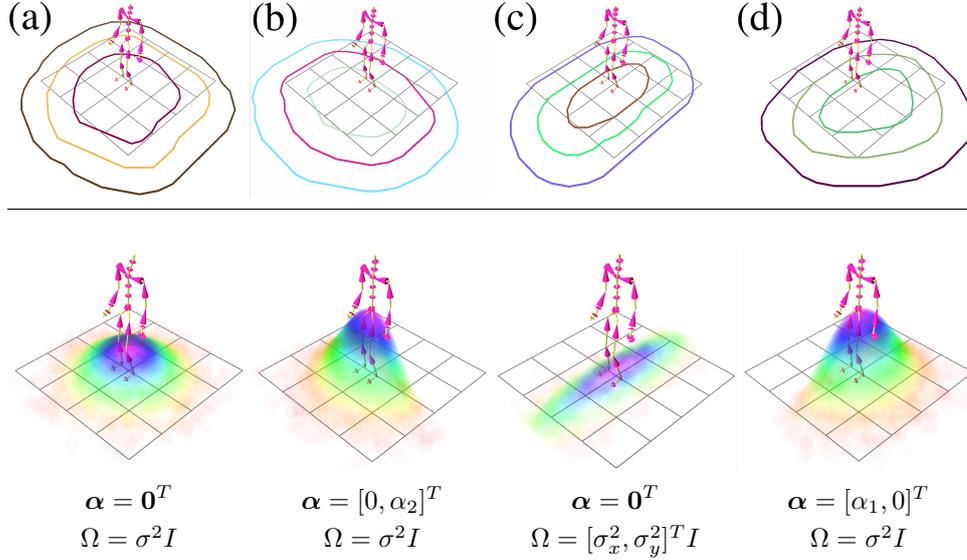


Figure 2.13: Top row; main spacing models for individuals, namely, (a) concentric circles, (b) egg-shape, (c) ellipse and (d) smaller in dominant lateral side. Bottom row; corresponding statistic representation by controlling the parameters of a bi-variate skew-normal probability density function.

treatment of the social spacing models adopted in the majority of human-robot interaction studies, namely: (i) concentric circles [62], (ii) egg-shape [66], (iii) ellipse [67] and (iv) dominant-side [68]. These functions constitute a super-set of the normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Omega)$, while being further parameterized by the *shape* parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$ that controls the amount of skewness (see Figure 2.13) and are denoted as :

$$\tilde{\mathcal{N}}(\mathbf{u}) = 2\phi(\mathbf{u})\Phi(\boldsymbol{\alpha}^T \mathbf{u}) \quad (2.3)$$

where $\mathbf{u} \in \mathbb{R}^N$ (in our case $N = 2$), $\phi(\cdot)$ denotes the normal probability density function $\mathcal{N}(\mathbf{0}, \Omega)$ with covariance matrix Ω and $\Phi(\cdot)$ is the cumulative distribution function of ϕ .

Our approach allows to express the dominant factors that bias an individual's social spacing by a single smooth probability distribution function, in contrast to earlier approaches that were only piece-wise continuous and thus did not necessarily conserve the probabilistic properties of the constituents.

As a means for allowing smooth transitions between pairs of spacing models, we then proposed to linearly combine a skewed, ellipsoidal model (either (b),(c) or (d) of Figure 2.13) with the most generic, non-skewed, circular model, using a smooth transition function (cf. eq.(2) in [9]) that quantified the certainty of perception.

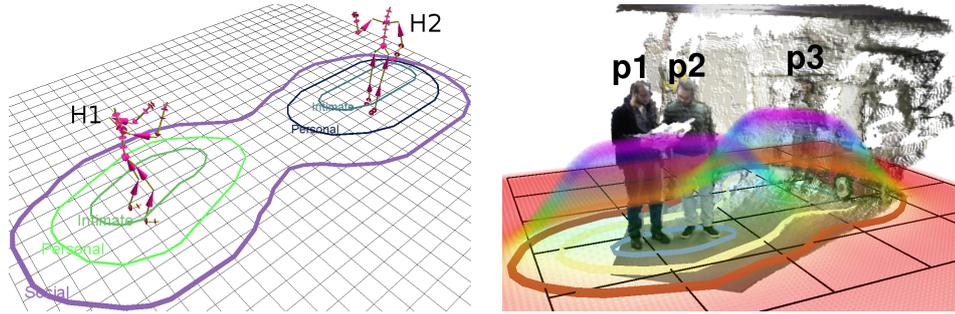


Figure 2.14: Left: Proxemic contours obtained as level-sets of an implicit function of global social sensitivity, obtained via one-class density estimation. Right: Superposition of social map on RGB-D point-cloud containing co-located humans.

Based on the previous formalism of individual social sensitivity, we then devised a framework for representing spatial interactions of multiple people, in the form of what we called **social map** [8]. This was done by drawing an analogy between implicit function learning and human social interaction mapping. In essence, co-located people were treated as sources of social sensitivity whose sum was modeled as an implicit probability density function of arbitrary form. The implicit function was obtained via one-class density estimation and the application of Kernel PCA (Principal Component Analysis) for novelty detection [69]. Our motivation resided on finding socially *charged* areas from areas where the robot could navigate without disturbing. Depending on the intention/task of the robot to either engage in an human-robot interaction or avoid disturbing, the obtained implicit function could then be employed as a potential-field based or navigation-function based approach [70]. Finally, groups of people were identified by delimiting this density through level-sets corresponding to proxemic distances of individuals (see Figure 2.14).

Results The usefulness of the previous contributions can be qualitatively evaluated in the supplementary videos # 6 and #7 of the Appendix - Multimedia, also part of the corresponding articles.

In terms of perception, the merits of representing atomic social sensitivity via non-stationary, skew-normal probability density functions was demonstrated through the association of articulated human motion cues (such as speed, posture and proximity to the nearest obstacle) to corresponding social spacing models. Finally, the social map values were shown useful as features for discriminating human interactions, using sequences of human activities from the CMU dataset of Motion Capture (MOCAP) [71]. The ultimate usage though of the social map was related to the development of socially-compliant robot

behaviors, a subject treated in Chapter 3, section 3.3.

2.3.2 Activity recognition

Problem summary Human action or activity recognition is one of the most challenging applications in the field of computer vision and amounts to inferring a label for each element of a data sequence that captures a human performing a task.

Among the multiple constituents of the problem (refer to [72] or [73] for a global presentation), we restrict our focus to the task of labelling sequences of captured articulated human motion, also known as motion capture (MOCAP), without consideration of context in the form of objects or of human interactions. After nearly two decades focused on video-based analysis, analysing the articulated human motion became massively accessible to researchers thanks to the advent of affordable RGB-D sensing technology and the seminal work of Shotton et al. [39] on human body recognition, followed by deep-learning based approaches [74]. Still, even when the human skeletal pose can be tracked, challenges such as variability in execution speed and style, segmentation among actions, self-occlusions or robot ego-motion interference, temporal context, etc. altogether make action recognition difficult to model and latent in its representation.

Contribution summary Due to the high number of articulations of the human body which leads to a skeletal representation of increased dimensionality, we studied the degree of correlation between the various joints within three major benchmarks of MOCAP sequences [60], namely, HDMO5 [75], CMU [71] and Mocapdata Eyes Japan (MEJ) <http://mocapdata.com/>. This was done via a PCA decomposition to principal components with their corresponding eigenvalues, allowing to conclude that around 95% of the total energy of each activity can be expressed by a total of 10 principal components, among a total of 70 dimensions (see Figure 2.15). Through the same analysis, we also inspected the variable degree of correlation among different classes in order to quantify their complexity.

Following the previous observation, we proposed a Gaussian-process based, latent variable model for representing MOCAP sequences wherein actions of the same category were non-linearly mapped to a lower-dimensional manifold that (i) preserved real-world distances of the original data and (ii) reduced intra-class distances and increased inter-class distances. Our method [59, 61] was thus focused on supervised representation learning that could be afterwards used with any conventional classification scheme. It was then evaluated in benchmarks HDMO5 and CMU and was found superior in terms of accuracy

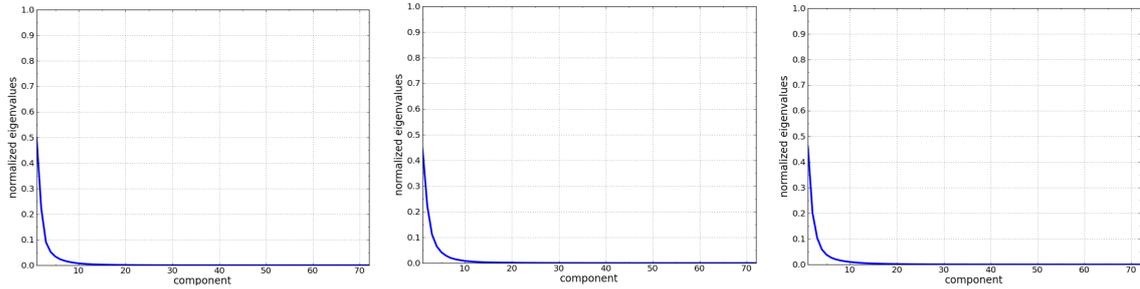


Figure 2.15: From left to right, principal component eigenvalues in decreasing order for a total of 70 MOCAP dimensions from CMU, HDM05 and MEJ datasets respectively.

against discrete time warping (DTW) and V-GPDS [76]. These works were performed in the context of the M.Sc. thesis of Ntouskos [77].

Secondly, we addressed the classification of activities of daily living (ADL) [78] that are more contextual and hierarchical in nature. Towards this goal, we devised a hierarchical deep-neural network architecture [15] composed of two levels of Long-Short Term Memory (LSTM) networks [79], the first fed with motion features of short, temporal sliding windows and the second operating on their hypothesized labels. The proposed approach was evaluated in the *Watch-n-Patch* benchmark [80] and was found superior against state-of-the-art approaches in terms of per frame accuracy of temporal segment, as in contrast to our previous contribution, this method does not need to know beforehand the segments of the sequence to be classified and was thus better suited for real-world application. The per-class performance of the contributed methods is shown in Figure 2.16.

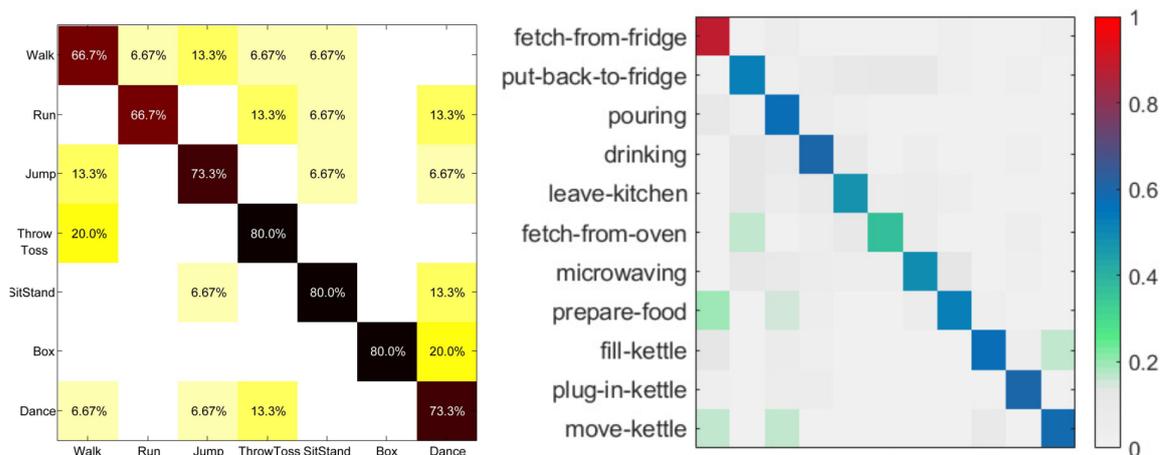


Figure 2.16: Left: confusion matrix for method [59] in CMU dataset. Right: confusion matrix for method [15] in Watch-n-Patch dataset.

NAVIGATION IN 3D ENVIRONMENTS

Autonomous Vehicle (AV) technology is nowadays becoming a reality as manufacturers are empowering modern cars with capabilities of autonomous driving in urban areas. Prior to consumer cars, systems exhibiting 3D navigation capabilities had been sought in applications such as military, space or natural environment exploration, mining and urban search and rescue (USAR). Even so, 3D robot navigation is a relatively recent field considering the birth year of mobile robots to the seventies with robot *Shakey* and robots with legged or more sophisticated locomotion being gradually introduced in the following two decades [81].

Irrespective of the application, enabling a specific robotic vehicle to navigate autonomously in a given 3D environment in an optimized and safe manner, requires the treatment of problems lying at the forefront of various disciplines. The difficult to model dynamics of a vehicle (and perhaps of the environment) in motion, the specificities of terrain interaction in 3D and the various inherent uncertainties of perception, altogether contribute to a risk of accident that can undermine testing and data acquisition.

The contributions presented in this chapter concern applications where mobile robots are destined to navigate in semi-structured indoor environments either for USAR or social robotics applications. Among diverse types platform types and robot locomotions, we focus on articulated tracked robots thanks to their elevated traction and low center of mass which prioritizes robot safety at the expense of reduced maneuverability.

We begin the chapter by presenting in **Section 3.1** our extensive literature survey on terrain traversability analysis for unmanned ground vehicles (UGV) [5], which provided an in-depth understanding of the challenges associated to the problem of 3D navigation in diverse environments as well as the main categories of methodologies. As perception and action are tightly coupled, the various methods that were reported are always conditioned on a robot with specific control and navigation characteristics.

On this basis, in **Section 3.2** we unfold our contributions to the problem of 3D navigation in static environments, namely, for the problem of negotiating negative obstacles

or gaps and positive obstacles in the form of staircases, for ascent and descent, under various degrees of freedom of the platform and constraints. Depending on the case, we adopt conventional planning and control or reinforcement learning (RL)-based control.

Finally, in **Section 3.3** we address the problem of socially-compliant navigation in the presence of humans, on the basis of the previously presented framework of social-mapping (cf. Section 2.3).

3.1 Essay on terrain traversability analysis for UGVs

A systematic and spherical presentation of the problem of safe and effective navigation in 3D environments was lacking, in contrast to the existence of numerous earlier works in various for various applications and robotic platforms. The lack of a central framework and approach to the problem resulted in redundancy in methodology, terminology, lack of insight and potentially conflicting findings. The dispersion of the various contributions in seemingly different domains made difficult to assess the current level of progress and open challenges. Indicatively, 8 terms in total were identified in the literature to refer to the

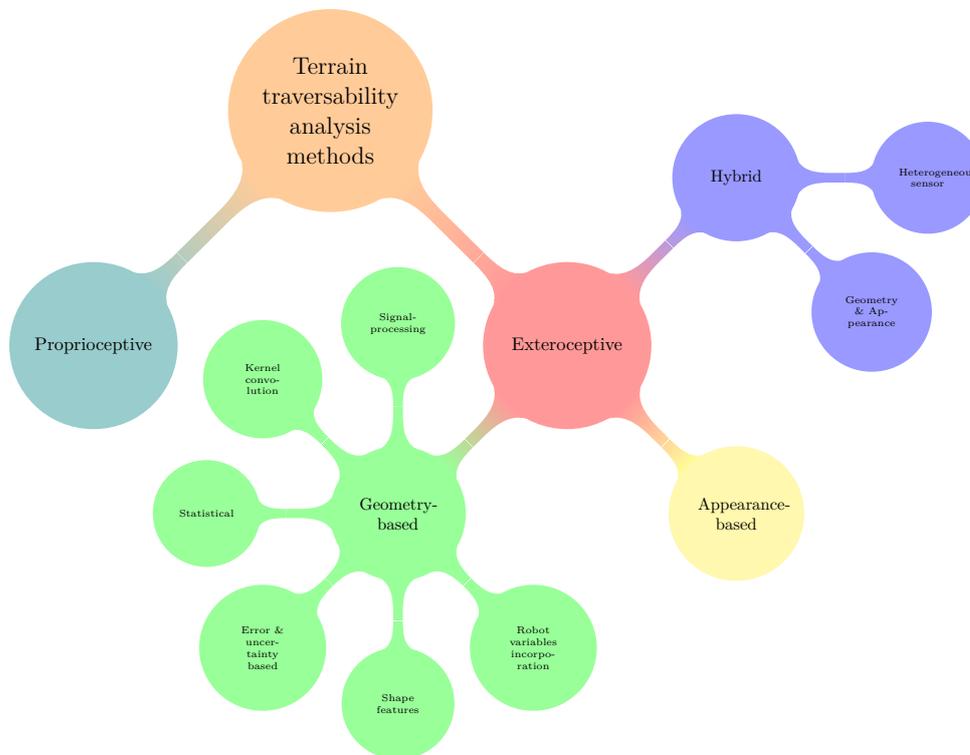


Figure 3.1: Taxonomy of terrain traversability methods for 3D navigation according to [5]

same theme, namely: (i) drivability, (ii) trafficability, (iii) navigability, (iv) coverability, (v) terrainability, (vi) maneuverability, (vii) mobility and (viii) traversability.

In view of that dispersion, we performed a state-of-the-art survey [5] identifying the most suited term, proposed a general definition and a taxonomy of different levels for the main streams of research (see Figure 3.1) and appointed seminal works into their corresponding class. The essay covers a total of 96 seminal articles that were selected from a period of more than 2 decades while our definition of *traversability* was the following:

The capability of a ground vehicle to reside over a terrain region under an admissible state wherein it is capable of entering given its current state, this capability being quantified by taking into account a terrain model, the robotic vehicle model, the kinematic constraints of the vehicle and a set of criteria based on which the optimality of an admissible state can be assessed.

For each individual work, we further identified the targeted application and the degree to which it corresponded to our definition of traversability, alongside a comprehensive description of the methodology. Finally, based on a critical inspection of the state-of-the-art, we were able to highlight open problems, limitations and advantages of sensors or families of methods and highlight promising directions of research.

Indicatively, Figure 3.1 allows the extraction of a key finding, i.e. that the bulk of earlier methods were concerned with geometry-based approaches, namely, active perception of the terrain using LIDAR sensors. On the other hand, appearance-based (or spectroscopic) approaches using passive sensors were largely understudied or overlooked. This changed drastically in the following years driven by the advent of deep-learning and the abundance of annotated image data that could be used for training. Other key findings (cf. Fig. 2, [5]) showed that most geometry-based methods quantified traversability only based on a terrain model with little or no account for robot characteristics (kinematics, model or safety criteria), or that applications mostly concerned planetary exploration and natural, off-road environments compared to urban settings (cf. Table 1, [5]).

3.2 Navigation in static environment

Approaches dealing with the problem of navigation can be traced back to the resolution of the renowned *Piano-Mover's problem* [82]. In the various approaches that have been proposed, a component is tasked with determining whether a given configuration of the robot is admissible, i.e. whether it can be attained while reaching for the goal. This

can be trivially solved for robot configuration spaces of low dimensionality and simple environments but it may require computationally expensive calculations for robots with multiple DOFs that physically interact with an environment of arbitrary complex form, under the influence of gravity and the dynamics of robot motion.

This is in particular the case for articulated tracked robots that are commonly used in USAR, the inspection of buildings or for surveillance, etc. For such mobile robots to fully exploit the available workspace in a 3D environment, they need to negotiate negative obstacles (gaps) as well as positive obstacles (staircases or steps), when possible. In contrast to ordinary, 2D free space where robot pose can be arbitrary without particular incidence to safety or effectiveness, successful negotiation of such obstacles is a necessary condition for the use of any planner.

The contributions presented in the sequel originate from publications [83], [6], [84], [85], [13] and [14] and were further partially covered in the next webinar :

- P. Papadakis, *Robotic Skill Learning and Development in Personal Assistance*, CRNS Webinar of the University of Plymouth, England, 2021, youtu.be/yNEqckC1MJI

3.2.1 Negative obstacle negotiation

Highlights

1. We propose a method for effective negative obstacle detection in USAR.
2. We compute the safest traversal path for negotiable obstacles.

Problem summary By definition, identifying a negative obstacle is a contextual task that requires to determine the space where data should have normally been captured but they are lacking. The absence of sensory data can however be due to various situations depending on the sensor modality and the environment, such as occlusions, field-of-view (FOV) limits or the properties of the sensed material. If certain hypotheses can facilitate the treatment of the problem for ordinary gaps in urban areas, the detection of an arbitrary gap and the question of its traversability by a given robot in USAR applications is not straightforward (see Figure 3.2 for an example of conditions in the case of a tunnel



Figure 3.2: Tunnel car accident scenario conditions for USAR operations

car accident scenario). Earlier research was thus very scarce and solely concerned with binary classification of negative obstacles, completely disregarding the question of optimal negotiation of negative obstacles. Finally, no previous work had dealt with conditions representative of USAR missions.

Contribution summary We developed an effective and efficient signal processing framework that performed negative obstacle localization and contour extraction, along with an assessment of its traversability via an optimally extracted path [6], as summarized in Figure 3.3.

The pipeline is composed of two main blocks of operation, the first assigned with gap perception that is performed by the following steps : (i) obtain a slice of the entire 3D point cloud parallel to the robot footprint, (ii) transform sensor coordinates to 2D image

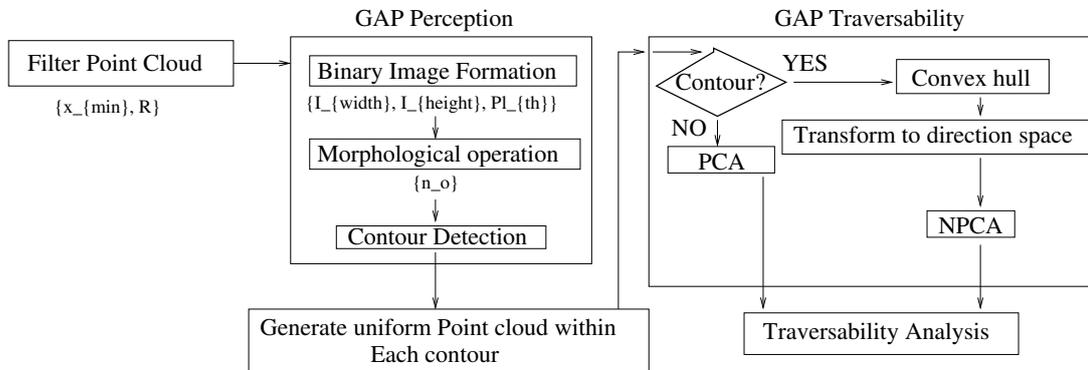


Figure 3.3: Pipeline for gap perception and traversability analysis (cf. Figure 2. [6])

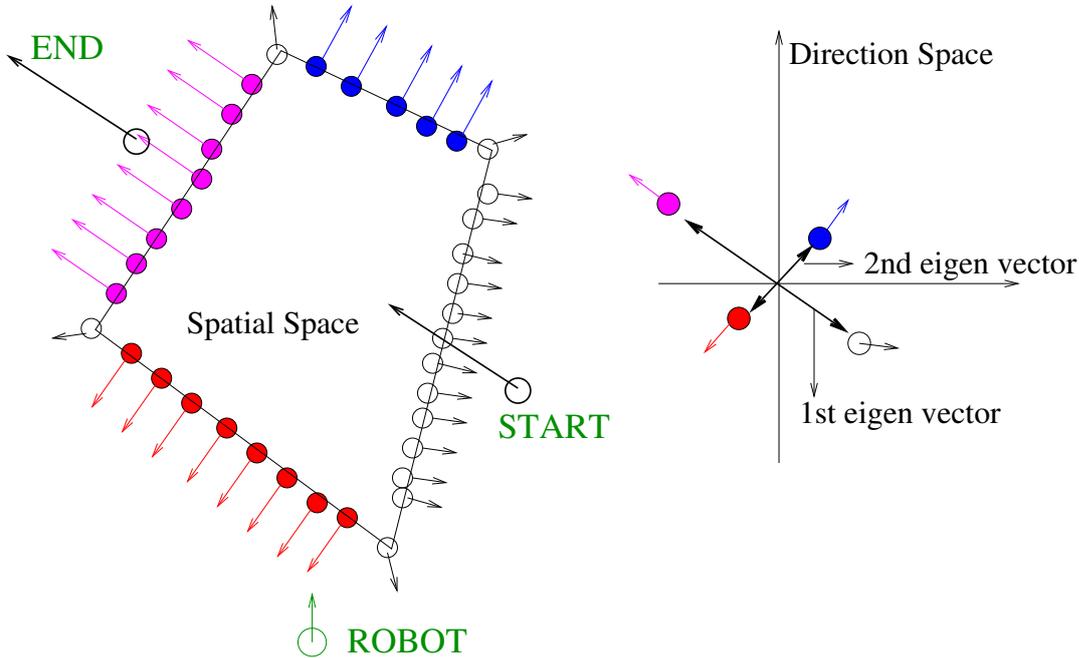


Figure 3.4: The result of employing NPCA to a negative obstacle, in order to determine the starting and ending point of the path to be traversed (cf. Figure 5 [6])

space to account for non-uniform scanning, (iii) filter noise by morphological operations and (iv) compute contours of clusters via border-following. The hypothesis is thus that we deal with relatively planar gaps in the immediate vicinity of the robot since even if it can navigate in 3D over longer distances, the space can be assumed locally planar in the proximity of the robot and tangent to the robot footprint.

The second main block of operation is concerned with calculating a straight path for traversing the negative obstacle that minimizes the loss of contact of the robot with the terrain and therefore its exposure to an accident. Our hypothesis here was that the majority of gaps have a mostly convex, ellipsoidal shape to which we can associate a primary (major) and secondary (minor) principal axis, using Principal Component Analysis (PCA). We then advocated the use of Normals PCA (NPCA) [86] for computing the position and the orientation of the start and end point of the path for negotiating the gap. This idea is illustrated in Figure 3.4.

Results The framework was evaluated on small, middle and big-sized gaps of a simulated tunnel car accident scene and of a building’s inner yard, perceived from different viewpoints and distances of the robot. All true gaps were consistently detected yielding

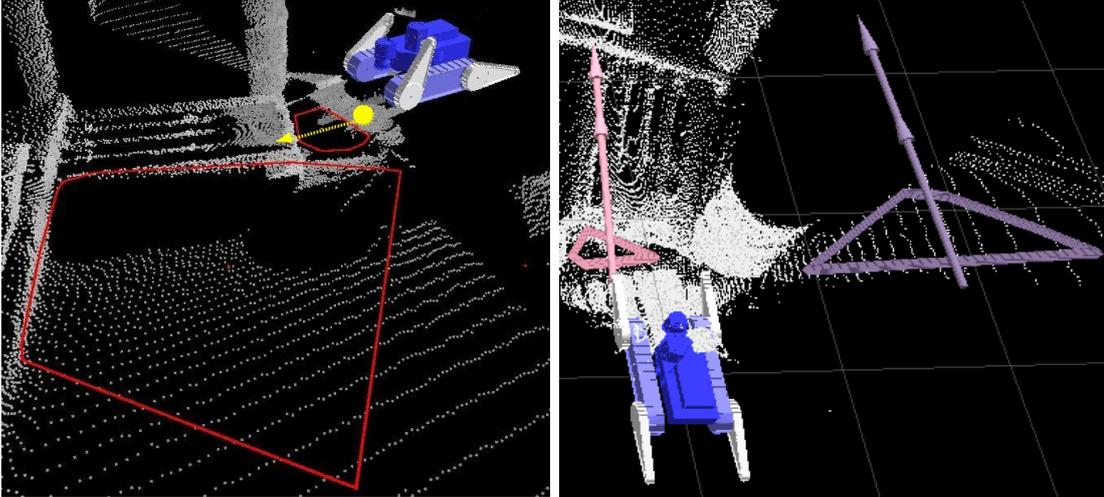


Figure 3.5: Negative obstacle detection and traversability analysis examples. Detected gaps and their traversability is shown by arrows indicating the path to be traversed.

a 100% true positive rate though occasionally false positives could be detected. Those were attributed to errors in the estimated robot state which resulted in erroneous plane estimation and search for gaps. Otherwise, the framework could easily filter out fictitious gaps due to occlusions. Figure 3.5 shows the output of detected gaps accompanied with the corresponding traversability analysis.

3.2.2 Off-line traversability learning

Problem summary The high number of degrees of freedom of articulated tracked robots combined with the dynamics and stochasticity of robot-terrain interaction, can severely hamper real-time operation and safe deployment of such systems. Even if a complete and accurate knowledge of the robot and environment state could be assumed, searching for plans and controlling such robots is subject to particularly costly operations for collision detection and assessment of optimality criteria due to the increased dimensionality of the state, action and search space. Even if the calculation of optimal paths in high-dimensional configuration spaces can be accelerated by sampling-based planners, still, obtaining the pose of an articulated tracked robot on a given 3D terrain patch is computationally demanding.

More concretely, when searching for an optimal path between a start and a goal state for a robot, each state is typically associated with a cost that quantifies the difficulty of visiting that state. This cost is then used when aggregating to obtain the total cost of

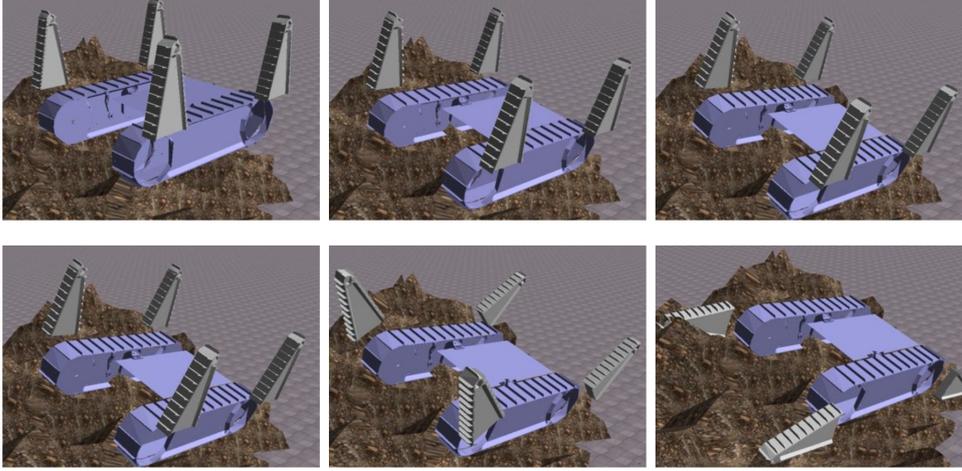


Figure 3.6: Top row; snapshots of robot while settling on top of a 3D terrain patch. Bottom row; adjustment of tracked flippers until ground contact

a candidate solution path, as is conventionally done in graph-search algorithms. Besides binary costs that simply assess the admissibility of a state, for articulated tracked robots that can actively adapt to the terrain it is more pertinent and necessary to regress a single, optimal robot configuration among many alternatives, for a given support terrain.

Contribution summary To mitigate the prohibitive, on-line cost of calculating an optimal robot state and the corresponding cost, we pre-calculated off-line via physics-based simulation the optimal robot pose over a large diversity of terrain patches and learnt to regress the associated cost as a function of terrain complexity [83].

To obtain the ground-truth pose over a given 3D terrain patch, we proposed a two-stage, physics-based robot pose estimation designed to increase the stability and surface contact of the robot footprint with the terrain. In particular, the simulated robot model was left to gently settle on the terrain patch under the influence of gravity until reaching

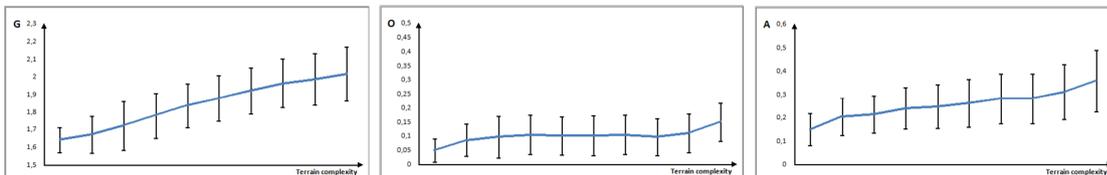


Figure 3.7: Traversability costs as a function of terrain complexity. From left to right, we show the evolution of the Ground clearance G, Orientation O and Angle-stability cost A. The blue lines correspond to the regressed traversability.

stability due to friction (see video # 9 of the Appendix - Multimedia). Once the robot became static, the second stage consisted in pushing its articulated flippers until reaching contact with the ground, yielding the final full articulated pose of the robot (see Figure 3.6).

In terms of criteria allowing to quantify the quality of the optimal pose, we augmented the standard formulations of *ground clearance*, *orientation* and *force angle stability margin* [87] with geometrically better suited estimates for articulated tracked robots (cf. eq. (1)-(3) [83]). By obtaining the ground-truth robot pose over 5000 random 3D terrain patches, we could then trace the mean value of the above measures (see Figure 3.7 with the intention of using it as a cost or heuristic in conventional path planning algorithms.

3.2.3 Reinforcement-learning based staircase negotiation

Highlights

1. Reinforcement-learning based formalization and treatment of staircase ascent and descent for tracked robot manipulators.
2. Transfer of behaviors learnt in simulation to a real robot and staircases and application to a second simulated robot of similar skills.
3. A publicly available software framework for navigation learning and evaluation in 3D indoor environments.

Problem summary Staircase negotiation concerns the capacity of a robot to traverse a staircase in an ascending or descending direction while respecting certain criteria. Formulated as a planning problem, the task consists in calculating and executing a best motion path from a starting point to a goal point. For the types of robots that are designed to address this task, the configuration space has high dimensionality as these robots have additional actuators for locomotion (e.g. flippers) and are often equipped with an arm for object manipulation. Still, even if the additional DOFs can in principle contribute in making negotiation more effective, developing a controller from scratch to be used either in a reactive or a planning process, is neither obvious nor trivial for multiple reasons:

- During staircase traversal a robot is subjected to stochastic perturbations such as sliding or loss of traction that are difficult to be sensed and compensated.
- The dynamics and the motion model of the robot might be unknown and difficult to establish, due to the high number of DOF of the robot and the variation of the 3D surface of a staircase.
- Collision checking in 3D between a highly articulated robot and a staircase is a computationally costly operation.
- Real data collection is costly as errors can severely damage the robot thus training data might be very scarce or impossible to obtain.
- There exists a high plurality of locomotion types and robot platforms, each one requiring its own custom controller.

Contribution summary The problem was addressed in the context of the PhD thesis [88] supervision using the commercial tracked robot manipulator *Jaguar V4 with arm*¹ and concerned the following main contributions:

1. **A reinforcement-learning based formalization and treatment of the complete staircase negotiation task for an actively articulated robot equipped with an arm.** In particular, we successfully tackle the problem of staircase ascent and descent by satisfying the primary goal of traversing the staircase as well as the secondary goal associated to safety constraints, using congruent control of the degrees of freedom associated to the tracks, flippers and arm joints.
2. **Demonstrated, zero-shot transfer of the behaviors learnt in simulation to a real robot and staircases, along with an application to a second simulated robot of similar mobility skills.** The transfer to reality and the application to two robots corroborate the generality and robustness of contribution # 1.
3. **A publicly available software framework for navigation learning and evaluation in 3D indoor environments.** Its architecture allows a user to interface between different RL libraries and algorithm implementations. At the same time,

¹http://jaguar.drrobot.com/specification_V4Arm.asp

learning can be customized to endow specific properties within a control skill. To show its utility, we focus on the case of staircase ascent and descent using depth sensory data while respecting safety via reward function shaping.

As far as contribution #1 is concerned [84, 85], this was performed via a neural-network based, reactive controller obtained via reinforcement-learning in simulation. The proposed solution consisted in using the entire action space of the robot composed of the various joint angles of the flippers and the arm, as well as the platform velocity. With respect to the state, controllers were learnt by using either local estimates of robot-step distances or raw depth-image features, an approach more reminiscent to end-to-end learning. Finally, along with the primary reward function appropriate penalty functions were devised for attaining secondary goals pertaining to the safety of the robot, according to the direction of traversal. In particular, for the ascending direction the safety criterion that was used was designed to respect both the stability margin (SM) [89] and the normalized energy stability margin (NESM) [90]. Conversely, for the descending direction the safety criterion was designed to reduce drop impacts incurred when the platform transitioned between step edges. Figure 3.8 shows the simulated model that was developed and the environment using for training the controller.

With respect to contribution #2 [13], the policies learnt in simulation were transferred to the robot *Jaguar V4 with arm* and tested in staircases of different dimensions, allowing the robot to safely perform the task while actively using its flippers and arm, under the

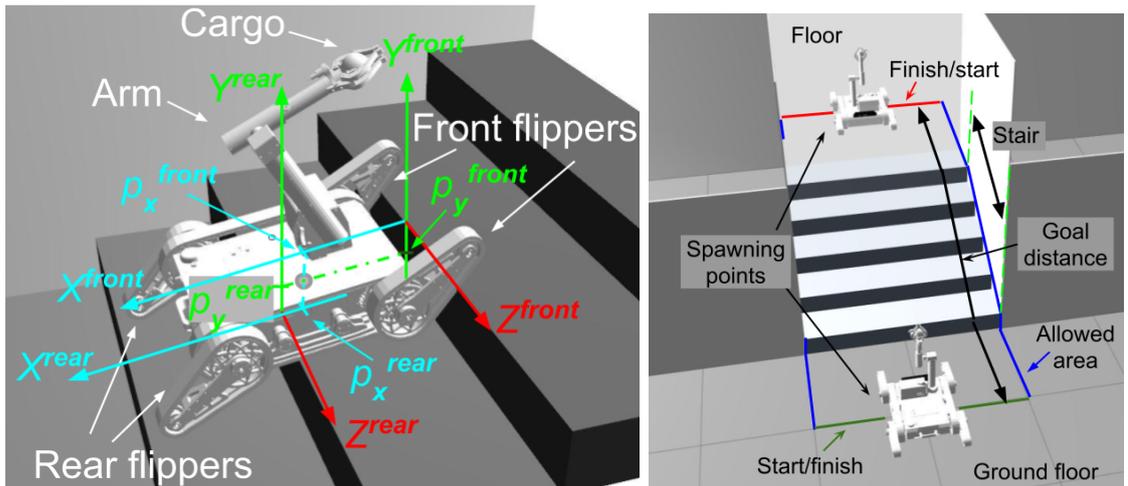


Figure 3.8: Coordinate frames used for the state space of the simulated robot model and training environment used for reinforcement-learning based controller development

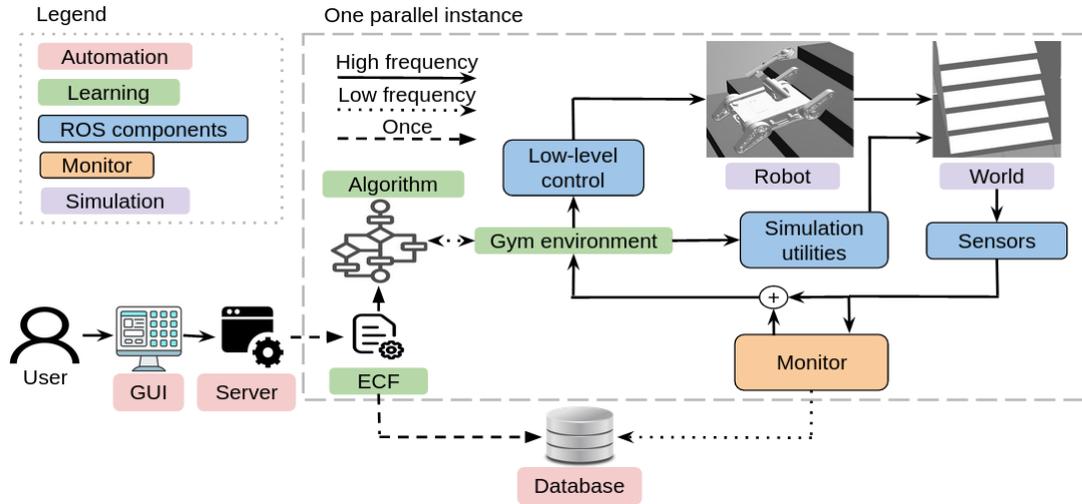


Figure 3.9: Overview of the software framework allowing to train 3D navigation policies via reinforcement-learning in simulation and incremental domain randomization

potential influence of a transported object. The transfer was made possible by training in a diverse range of staircases while adding noise to the state to reduce overfitting. Finally, the developed framework was used in a second, simulated robot of similar locomotion² but of different geometry and mass. This allowed to validate the effectiveness of the approach in a second example and to quantitatively compare the obtained policies between the robots by the Kullback-Leibler divergence of the controller actions for same inputs.

Contribution #3 amounted to the development and dissemination of an open-source, software pipeline for learning navigation of tracked robots in 3D environments. Its novel features concerned: (i) the stochastic generation of 3D environments (corridors or staircases) of different complexity that could be used for incremental learning of policies, an approach known as domain randomization, (ii) the provision of two models of tracked robots and (iii) examples of policies learnt in an end-to-end fashion using depth images. Figure 3.9 shows the structure and components of the developed framework (cf. [14]).

Results Qualitative results of the trained robot behaviors in simulation can be viewed in the videos # 3 and # 2 of the Appendix - Multimedia, attached to publications [84] and [85] respectively. Figure 3.10 shows consecutive snapshots of the real robot Jaguar while performing the ascent and descent task, using the controller learnt previously in simulation. Qualitative results of the robot behavior in reality can be viewed in videos

²https://github.com/mariogianni/trav_nav_indigo_ws

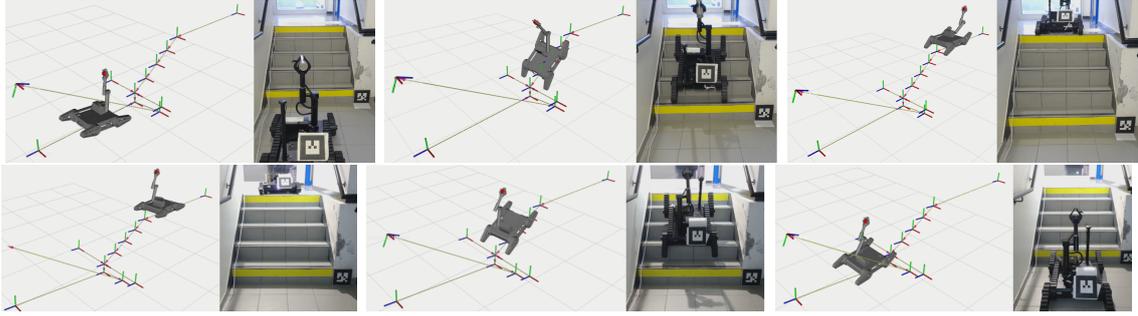


Figure 3.10: Snapshots of staircase negotiation with congruent control of 5 DOF

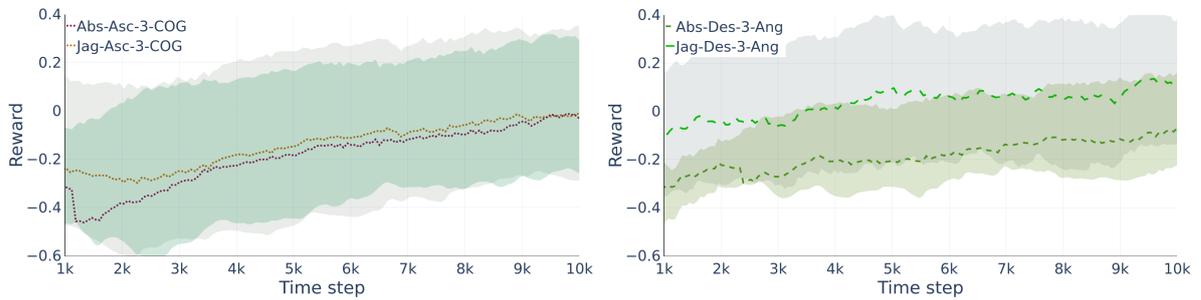


Figure 3.11: Reward evolution during training for ascent (left) and descent (right) and robots Absolem and Jaguar.

1 of the Appendix - Multimedia, accompanying publication [13]. Figure 3.11 shows the obtained rewards while training in simulation for the two different robots, for ascent as well as for descent. Extensive quantitative results provided in the dissertation [88], demonstrated the successful convergence of control policies learnt for variations of the staircase negotiation task, namely, in terms of number of DOFs, direction of negotiation, robot model, levels of noise and reward function used. Figure 3.12 shows samples of environment configurations used for training via incremental domain randomization (DR) [14] and Figure 3.13 shows snapshots of the robot ascending a staircase by employing and policy learned in an end-to-end fashion, using as input depth-image samples.

3.3 Socially-compliant navigation

Problem summary Navigating in a dynamic environment generalizes the problem of navigation in a static environment, a key difference being the time-varying nature of the robot configuration space. The added time dimension implies that motion models need to be employed and updated for the moving obstacles while planning for optimal paths.

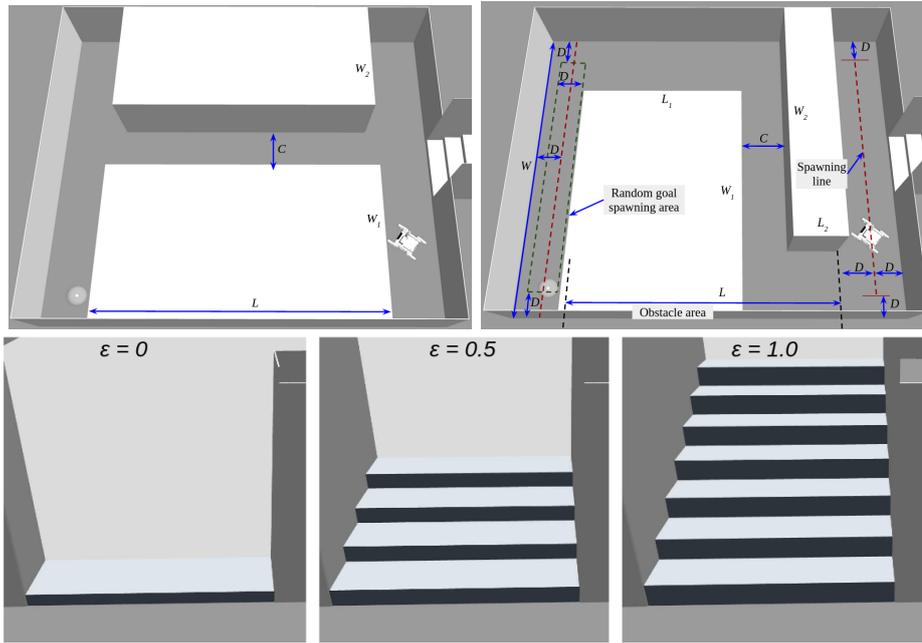


Figure 3.12: Sample environment configurations used in incremental domain randomization for 3D navigation learning

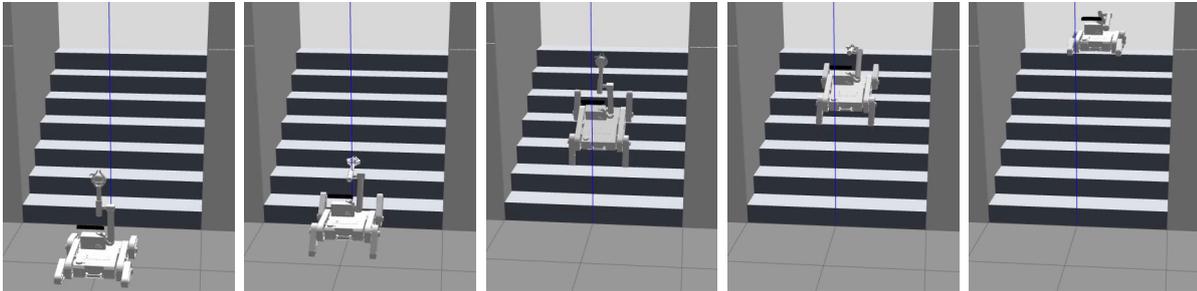


Figure 3.13: Snapshots of robot ascent of a staircase using a policy learned in an end-to-end fashion

Going even further to suppose a similar agency for the moving obstacles, be it humans or robots, i.e. identifying and responding to an agent’s own motion in order to navigate in a conformal way, this introduces the criterion of comprehensibility of navigation.

On this basis, a socially-compliant robot navigation in a dynamic environment need not only be optimal with respect to robot constraints but be further transparent and understood from collocated humans. In essence, this means that robot navigation should respect human conventions of navigation and social spacing.

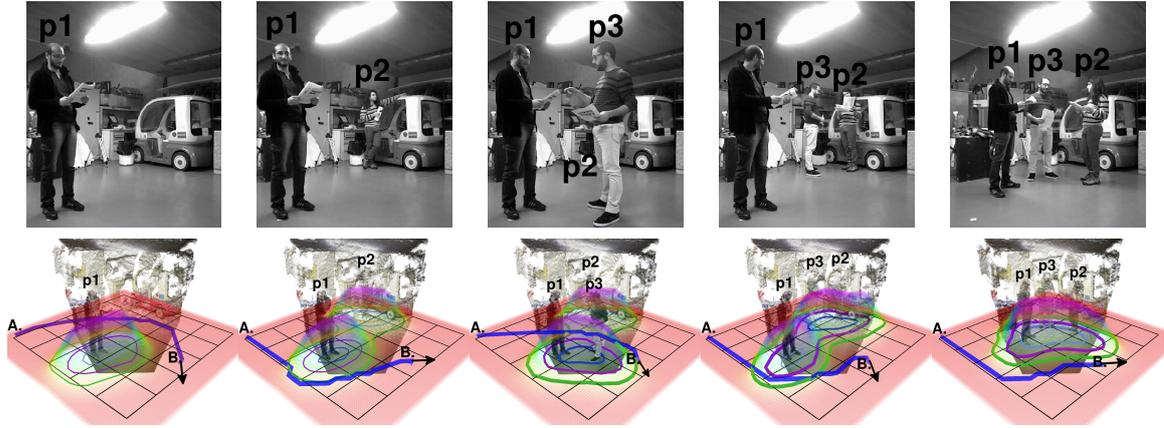


Figure 3.14: Result of mapping the social sensitivity of collocated humans, overlaid with a path plan that would minimally interfere with proxemic levels (cf. Figure 6 of [9])

Contribution summary On the basis of our earlier description in Chapter 2.3 and earlier works [8, 9], we used the dynamically updated *social-map* that was generated by the collective social sensitivity [9], in order to plan socially-compliant 2D paths in scenes populated by arbitrary numbers of interacting humans.

In detail, the problem consisted in calculating a path from a start to a goal using a baseline trajectory planner that could operate in total on 3 distinct levels of increasing degree of interference to human social sensitivity, from *public* until *personal*. As these levels of the social map could be obtained via the level-sets of the implicit function (recall Figure

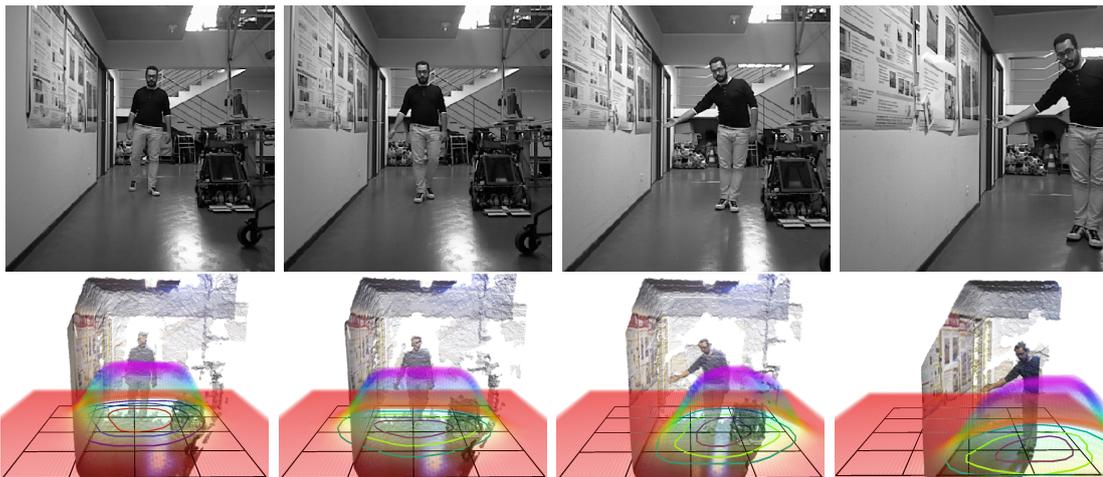


Figure 3.15: Egocentric views of a robot bypassing a human along a corridor, using the concept of adaptive spacing by considering certainty and social cues (cf. Figure 5 of [9])

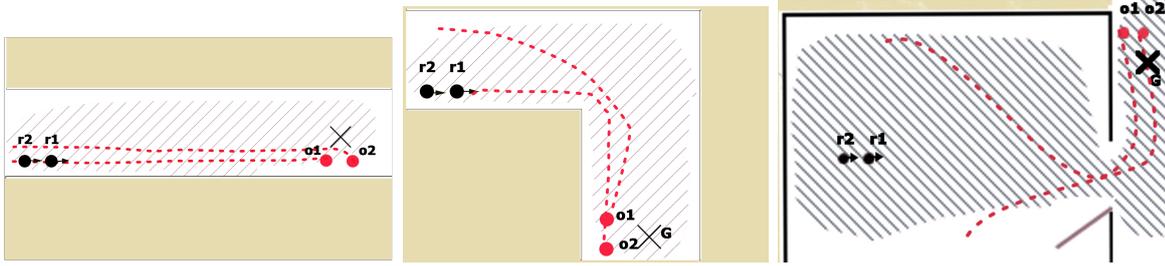


Figure 3.16: Scenarios considered for evaluating a leader-follower platoon robot system denoted by robots $r1$ and $r2$ while encountering dynamic obstacles $o1$ and $o2$. Red-dotted lines indicate obstacle directories, unknown to the platoon.

2.14), path planning was performed to the free space lying on the exterior of a level-set, starting from the first level until the third until a feasible path could be obtained. Figure 3.14 provides characteristic examples of human-interactions at different time moments and the corresponding path plan obtained from the application of the previous algorithm. Figure 3.15 shows an example application of socially-compliant navigation of a robot that traverses a corridor while encountering a human. The entire demonstration sequences can be viewed in video # 6 of the Appendix - Multimedia.

The aforementioned works targeted the navigation of a single robot in dynamic environments possibly populated by multiple colocated humans in different interaction scenarios. In subsequent works, the navigation of multiple robots alongside multiple people was studied. In particular, in [91] we set the foundations of a layered robot architecture to allow coordinated multi-robot deployment alongside humans. Our focus was on the task of item delivery and optimal route finding in conditions that could lead to congestion between robots and the synergies that could be envisioned with ambient intelligence services (see section 4.3).

In a more extensive work [92], the goal consisted in developing a multi-robot system employing a *leader-follower* control scheme moving indoors, while encountering dynamic obstacles in scenarios such as those shown in Figure 3.16. In essence, the scenarios varied the configuration of static obstacles and the extent of maneuvering to avoid collision.

A PID controller was developed for the follower robot with the leader robot using the standard ROS-navigation stack, although this aspect was transparent to the follower which could instead be asked to follow a human walker. The results of this work allowed to assess the portability of baseline algorithms for leader and obstacle detection and avoidance when deployed on the TurtleBot2 robot, using as a measure of performance



Figure 3.17: Leader-follower path along a corridor in the presence of a dynamic obstacle.

the proximity of the follower to the leader along the obtained trajectories. Figure 3.17 shows a series of snapshots of the leader-follower robot team safely bypassing another robot that crosses its path. Overall, however, the experiments revealed that tackling the problem of platooning in a dynamic environment required considerable more attention as a result of unexpected motions, loss of line-of-sight during narrow turns and lack of direct communication between the leader and the follower.

TRANSVERSAL WORKS

This chapter pertains to earlier works concerned with the integration of several individual modules into a single robotic system, capitalizing on vision and navigation skills for 3D environments. During a development life-cycle, the alternation of research with integration and deployment phases, or training and testing, is essential as it allows to evaluate the validity of the hypotheses made during design and to assess the limits of the previously developed methods. Therefore, the essence and importance of such works is that they are interdisciplinary and that they repose on the contributions of several authors and/or partners employing domain-specific terminologies, which renders the harmonization of the total set of constraints into a single, operational system particularly challenging.

4.1 Emulated USAR exercises

In the context of EU project NIFTi (Natural Human Robot Cooperation in Dynamic Environments), highly realistic exercises were conducted consisting of human first-responders (fire-fighters), central-control operators, ground and aerial robots.

A first prototype of a semi-automated, human-robot team was described in [93], presenting the architecture of a differential-drive mobile robot that was capable of car detection, SLAM, topological mapping, high-level human-robot dialogue and task planning, deployed in a parking-lot exploration mission. My role in the respective work consisted in the coordination of the integration efforts between the high-level, logic-based robot control, dialogue and the perception and actuator components of the mobile robot among a total of 4 European partners. Figure 4.1 illustrates the key aspects behind the corresponding framework. The underlying architecture allowed a human operator to provide high-level vocal commands to the robot, such as *go here*, *move left*, *right or forward*, *approach an object* or perform *coverage planning*.

That system served as a baseline for subsequent developments with a custom-made, articulated tracked robotic platform (Absolem-TRADR robot) and a more challenging,

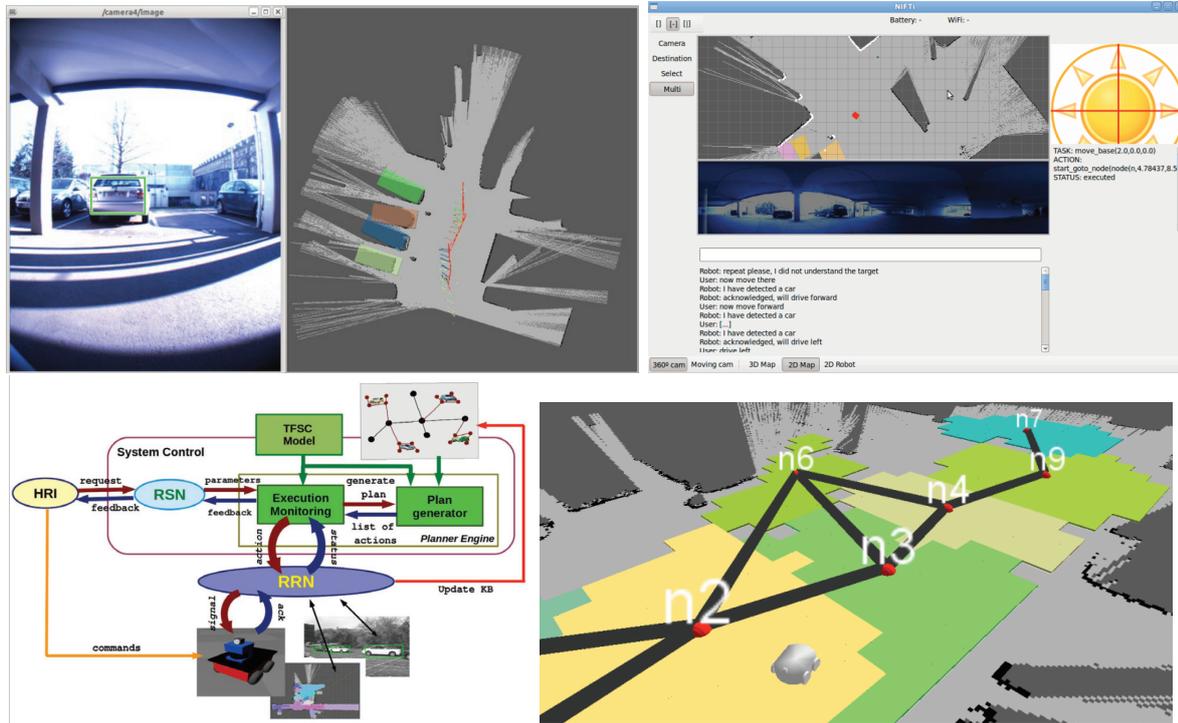


Figure 4.1: Key components of the human-robot architecture of [93]. In row-major order, car-detection in omni-directional images and 3D localization, human-robot interface, planning-engine and topological mapping.

tunnel-car accident scenario in the MonteLibretti firefighter training facility in Italy. Compared to the outdoor, parking lot mission the tunnel-accident scenario was characterized by more harsh conditions of operation for both humans and robots, which underlined the need for more transparent, effective and efficient communication between humans and robots. In this sense, a user-centric design methodology was followed so as to instill as much as possible trust and transparency with regards to the robot operation from the point of view of the humans [94].

This consisted in the directed involvement of end-users from German and Italian firefighter brigades during the entire development cycle, i.e. specifications, development and testing. In this manner, the vision and navigation skills of the robot were developed so as to reflect as best as possible the actual operation protocols of first responders and hence orientate the autonomy and intelligence of the robot accordingly. For example, a command of the type *go to the car* did not merely imply to approach a particular car but to position the robot to a particular vantage point, that would allow its camera to look through its windows and detect potential human victims. This supposes the detection of

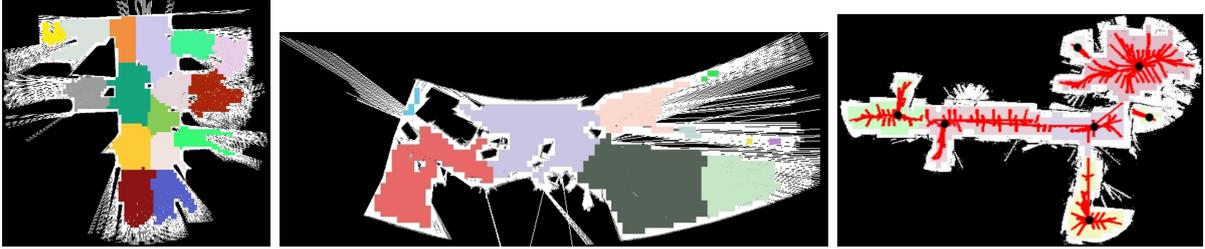


Figure 4.2: Topological mapping and path planning examples from [35]. From left to right: Topological map of a parking lot, topological map of tunnel car accident scenario, topological map and node-level paths of an office environment.

the car along with its pose while further assessing the navigational capacity of the robot to approach that particular area. In doing so, the robot could employ off-line learned, functions of 3D traversability [83], traverse negative obstacles by following an optimal path in terms of safety [6] or plan and execute a path within a topological map of the environment [35].

Aside the first two works presented previously in Chapter 3, our work [35] on topological mapping proposed an approach for obtaining in parallel the topological map of the environment and a set of local, node-level paths allowing to connect any pair of points within the metric map. The idea consisted in employing the mean-shift algorithm for clustering non-occupied cells to a set of clusters and tracing the mean-shift path of each pixel while ensuring that its feasibility by a real robot. Figure 4.2 illustrates the concept on some characteristic examples.

4.2 Human-robot USAR mission in Italy

Two major earthquakes stroke the Emilia-Romagna region in Northern Italy in 2012 followed by further aftershocks and earthquakes in June of the same year, causing multiple casualties and widespread damage to numerous historical buildings. The Italian Fire Brigade deployed disaster response and recovery of people and buildings while in June, they requested the aid of the EU-funded project NIFTi, in order to assess damage of historical buildings and cultural artifacts located therein. The project partners deployed a human-robot team composed of terrestrial and aerial robots in the red-area of Mirandola, Emilia-Romagna, from Tuesday July 24 until Friday July 27, 2012, that worked in collaboration with the rescuers involved in the red area [7] (see Figure 4.3).

In the context of this mission, I was tasked with the coordination and piloting of the

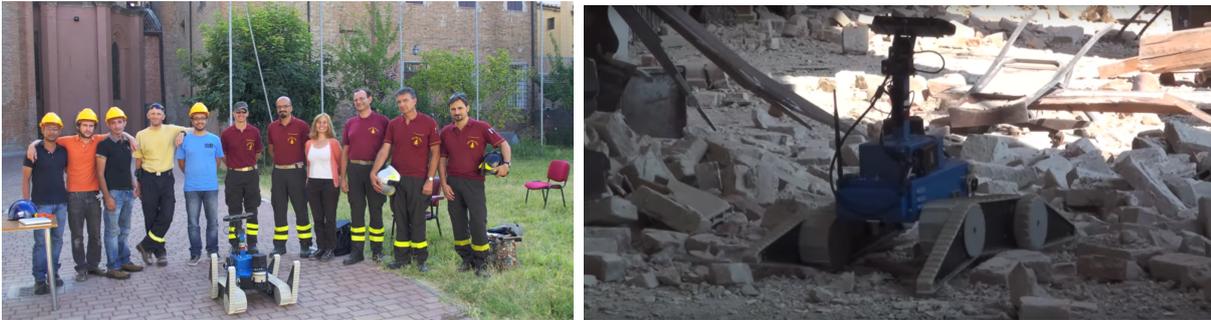


Figure 4.3: Left; project members and firefighters that participated to the USAR mission [7]. Right; snapshot of the tracked robot during operation.

terrestrial vehicle **Absolem**, in order to perform the mapping of two historical buildings. Figure 4.4 gives an idea of the degree of damage and rubble of the two sites, together with the 3D reconstruction that was obtained via the robot operation. That endeavor was characterized by numerous challenges :

- The robot and the entire system infrastructure had never been tested in such extreme conditions in terms of terrain and heat https://ec.europa.eu/commission/presscorner/detail/en/MEMO_12_620
- The area to be explored was blocked to access even to the rescuers due to the elevated accident danger. If the robot got damaged or were unable to safely return to base it would have to be abandoned, scraping an equipment worth of several tens of thousands of euros.
- Timely intervention was critical as the NIFTi team was given a limited time window during which the robots could operate, in order to avoid interference with other ongoing disaster response tasks.

Micro-managing a robot on the rubble of ruined buildings under the above constraints while mapping the environment was pioneering in Europe <http://www.vigilfuoco.it/asp/Notizia.aspx?codnews=16185>. The success of the mission resided on the orchestration of human experts together with the various functionalities of a highly articulated, tracked robotic vehicle <https://spectrum.ieee.org/robot-to-human-trust-me>.

Besides the inspection of damaged cultural heritage art-work via the 3D reconstruction of the corresponding sites, the mission further provided invaluable, real-world training



Figure 4.4: Snapshots of the two destroyed historical sites in which the robot navigated and corresponding 3D maps.

data that could be used for developing controllers for navigating 3D terrain, notably via 3D mobility learning and regression by simulated, physics-based optimization [83].

4.3 Assistive robot and smart space integration

A more recent line of work seeks to leverage the advantages of assistive mobile robots or manipulators and of smart spaces, so as to develop more sophisticated services and technologies for professional or common end-users. Such a synergy seems formidable as smart spaces seek to offer general-purpose services based on a high-level and globally consistent understanding of human activity, while robot mobility and dexterity can be more easily and directly customized according to more specific needs.

Accordingly, we identified in [91] the main functionalities that such an integrated system could provide in connection to previous works and the way by which synergy could be accomplished, from low-level behavioral to higher-level task planning of a multi-layered system architecture (see left of Figure 4.5). The examined scenario concerned a smart hospital equipped with mobile robots operating within it for performing tasks such as meal

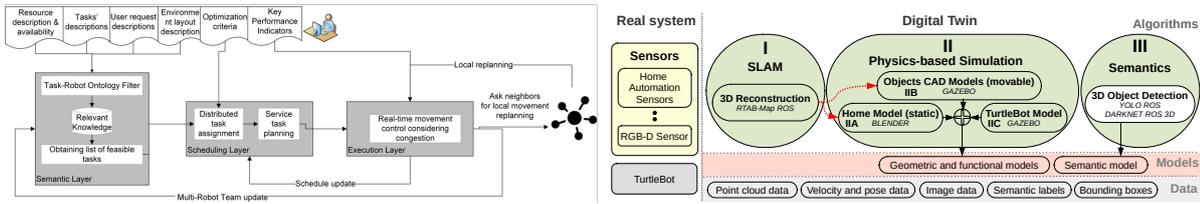


Figure 4.5: From left to right, proof-of-concept schemas for integration of robots with ambient intelligence from [91] and [96] respectively.

or medication delivery, while accounting for human presence during navigation and resolving possible congestions. This pilot study consisted in bridging dispersed methods and systems into a single framework, while accounting for constraints relevant to a potential commercialization, such as inter-operability standards [95] or socio-technical aspects.

More recently in [96], we sought a more generic approach to model the synergy of smart spaces and mobile robots, via the notion of the digital twin (DT). In this regard, we were not concerned with a specific scenario or development of technology for personal assistance but rather with the establishment of a procedure that would allow to create a digital replica of the entire set of physical assets (e.g. smart apartment, mobile robot, human, etc). Such a digital replica could be then used for faster prototyping, training and testing, particularly in what concerns the development of vision and navigation skills of mobile robots, e.g. for object detection and classification, scene segmentation or low-level robot control. The right of Figure 4.5 depicts the structure of the developed DT alongside the physical assets that are modeled. Finally, a correlated work was performed in [97] in the context of the project *TwinCobot*, with the aim of developing AI-enabled digital twins of human operators collaborating with a robot manipulator. The emphasis here was mostly put on the software engineering aspects of the DT and system integration between different hardware and software components.

RESEARCH PERSPECTIVES

My perspectives for further research are orientated towards new paradigms for robot vision, navigation, and by extension to manipulation, that increase the autonomy and robustness of robots in the 3D world whenever they are destined to cooperate with human professionals or provide assistance to individuals. The advent of machine-learning bears particular promise for developing such systems, yet, there is a balance to be sought between data-driven and symbolic approaches as well as between simulated and real-world training and testing. As a matter of fact, systems that attain state-of-the-art performance in predefined benchmarks can easily fall short when faced with the diversity and stochasticity of real-world situations referred to as *open-set* conditions [98], [99].

In this context, via my involvement in a number of ongoing and upcoming projects I seek to address central technological and scientific questions in the domain of vision and navigation in 3D environments, as presented by category in the sequel. For each project, the involved partners and domains of application are further noted so as to highlight the various collaboration initiatives with academic as well as industrial partners.

5.1 Axis I - Meta-learning

Learning to learn or meta-level learning is concerned with learning properties of functions, i.e. learning entire function spaces [100]. Reposing on conventional, base-level learning (such as supervised learning for classification, regression, etc), meta-learning would allow a robot to share knowledge between seemingly different tasks, adapt its knowledge in the light of new evidence, learn continually, i.e. incrementally and hierarchically. Such capacities are crucial for the deployment of robots in the real-world alongside human presence, as base-level learning proves severely limiting in practical applications. The following projects address such challenges for computer vision and robot manipulation tasks:

- Project **ILCOI** - *Incremental Learning for Classification of Objects of Interest* (2020-2023): We focus on the problem of catastrophic forgetting of neural networks [51] in the process of incremental learning of objects or events belonging to new categories and/or new instances of objects or events, not already included in a pre-trained discriminative model. This project aims to develop more robust incremental learning methods for embedded systems in a terrestrial, marine or underwater context, characterized by hard constraints on memory and computational resources. Promising first results have been obtained via joint, incremental contrastive representation and classification learning [19].

Partners: ENSTA Bretagne, Naval Group. **Domain:** Generic & sea vessel images

- Project **ECFvisuL** - *Évaluation automatique continue de la capacité fonctionnelle humaine à domicile à l'aide de la vision par ordinateur* (2022-2024): We seek to develop an automatic solution for continuous Functional Capacity Evaluation (FCE) [101] of individuals in a domestic environment, on the basis of a computer vision system that tracks human body posture. Such a system would acquire more temporal data on the musculoskeletal capacity of a human than a typical FCE session, producing more robust evaluation results. The problem to be addressed can be posed as a meta-learning problem since automating FCE suggests a continual learning problem as the user capacity changes over time, while each individual constitutes a class per se, implying that training data acquired from others to track and analyze human pose might not be directly exploitable or relevant.

Partner: CHU Brest. **Domain:** RGB-D images & human skeleton trajectories

- Project **ABYSSES** - *Annotation Boostée des modèles 3D de monts hydrothermaux Sous-marins par apprentissage profond d'adaptation de domaines* (2022-2023): In the domain of hydrothermal ecosystems mapped by underwater drones, this project aims the development of new digital tools allowing to accelerate the capacity of semantic annotation of very large surfaces of biological, environmental and topographical characteristics of marine depths. The core challenges pertain to the absence of sufficient training data that hinder the application of deep-learning while the particularities of the domain result in annotation protocols that were not designed for use by machine-learning algorithms. This calls for research in meta-learning methodolo-

gies such as self-supervised learning, domain adaptation or transfer learning, so as to exploit abundant, unlabeled data of very high resolution in order to expedite and assist the tedious task of manual annotation by human experts.

Partner: Ifremer. **Domain:** Seabed images

- Project **ROGAN** - *Robot Manipulation Learning Using Generative Adversarial Imitation Learning* (2020-2023): This project aims at developing robot manipulation skills for assistive services, capitalizing on the power of Generative Adversarial Networks (GAN) [102] in order to artificially generate task demonstrations so as to enable more efficient learning while relying less on human experts. The expected contributions reside at the crossroads of reinforcement learning, imitation learning and contrastive learning using as experimental platform a state-of-the-art robot manipulator (Robotnik RB1) [18].

Partner: Australian Institute of Machine Learning (AIML). **Domain:** Robot trajectories

5.2 Axis II - SLAM and Navigation

Metric mapping of an unknown environment resides at the core of numerous applications that range from search & rescue, autonomous cars and digital buildings up to human guiding and assistance. The problem has been at the focus of robotics research by more than 3 decades now, giving rise to a plurality of open-source SLAM libraries. On the other hand, algorithm efficiency, robust performance in challenging environment conditions and employment of deep-learning on resource-constrained hardware can constitute major bottlenecks for the deployment of state-of-the-art SLAM methodologies. A complete SLAM system is typically composed of several components, e.g. feature extraction and matching, data association, loop closure, graph optimization, etc. that need to interact in real-time while the agent is moving. While deep-learning can be auxiliary in different phases of the entire pipeline, it is a computationally-intensive process that can undermine the capacity of the entire system to perform on-line.

Along with the emergence of deep-learning, a new paradigm of navigation known as *end-to-end* navigation [103] aspires to accelerate the perception-action loop via bypassing the stage of mapping and directly map raw sensor readings to control and navigation

commands. This presupposes however massive training and a neural network architecture that is unlikely to be explainable to humans, raising the question of trust and performance guarantees that cannot be obtained. The trade-off between using conventional SLAM with planning-based navigation and the exploitation of DL is thus an active research topic that calls for further research depending on the environment of deployment and the platform into consideration. The following projects investigate such questions either for aerial or terrestrial USAR robots and humans with disabilities.

- Project **HIT** - *Handicap Innovation Territoire* (2020-2027): This multi-partner and multiannual project seeks to capitalize ICT technologies in the development of services and technologies that palliate human impairments/handicaps at various scales and locations, from a person's residence and urban activities upto a psycho-medical institution environment. Among the various project actions, we will be treating the problem of localization and navigation for people with visual impairments, either via wearable sensors or a robot companion.

Partners: Handicap Innovation Territoire consortium. **Domain:** Indoor/outdoor images

- Project **REMED** - *REconstruction Métrique 3D en temps réel* (2020-2023): We are interested in the improvement of metric SLAM solutions for 3D environments, under computational power constraints and by using only passive sensors. Of particular interest is the contribution of deep learning in recovering incomplete 3D models due to occlusions, errors, dynamic objects or luminosity problems while being complementary to conventional geometric SLAM.

Partners: ENIB, Thales Palaiseau, Thales Research & Technology Canada. **Domain:** USAR, aerial images

- Project **LEASARD** - *Low-Energy deep neural networks for Autonomous Search-And-Rescue Drones* (2022-2025): The project will develop more energy-efficient drones with obstacle avoidance and SLAM algorithms that can better exploit the potential of Deep Neural Networks (DNN) in two phases: first, by using energy-efficient imaging sensors whose input is fed to specialized DNNs and second, via energy-efficient DNN processing implemented on energy-friendly hardware as FPGAs and System-on-Chips (SoCs). We will be particularly concerned with develop-

ing energy-aware DNN-improved algorithms for navigation and SLAM for USAR missions, using event-cameras and neuromorphic vision.

Partners: LS2N, IRL CROSSING. **Domain:** USAR, event data & color images

5.3 Axis III - Digital Twin

The definition of the digital twin (DT) can vary depending on the field of application, nevertheless, it refers at its core to a cyber-physical system (CPS) endowed with intelligence and composed of physical and digital entities or processes that are integrated to each other. It is thus a complex digital representation of a physical system which precisely models and is able to predict how the real system will perform. In this respect, it can serve as a tool to bridge the gap between the physical and digital worlds. Crucially, the connection of the DT to the real-world is bidirectional, meaning that real-world changes and effects are propagated to the DT.

The core functionality of the DT that I am interested in concerns its potential to be used as a training environment for robotic tasks pertaining to 3D vision, navigation or manipulation, pertaining to environments such as *iGibson* [104], *AI2THOR* [105] or *Habitat* [106]. This bears enormous advantages in robotic applications as it allows the generation of limitless data, that would otherwise be impossible to obtain or subjected to hardware and various other limitations and costs. Rather than building an ordinary simulated environment that is detached from the real counterpart, I am thus interested in the process of building a DT by observation of the real-system, together with the establishment of metrics for assessing the quality of the DT, be it for 3D vision or navigation tasks. These questions are investigated in the following projects:

- Project **SSS3D** - *Segmentation Sémantique d'une Scène Industrielle 3D* (2020-2023): The scope of the project is the recognition of objects within 3D, colored point clouds obtained from industrial sites, as a step of a DT building process with the goal of inspection for degradations or problems by human experts [16]. The main challenge reposes on the absence of sufficient volume of annotated training data that are mostly proprietary, which hinders the employment of deep-learning. At the same time, the particularities of the scenes of interest challenge the use of transfer learning techniques from other applications where data are more abundant.

Partners: Segula Technologies, ENIB. **Domain:** Industrial areas, 3D point clouds

- Project **undisclosed** (2023-2026): The goal consists in updating in real-time a CAD model of a scene that had been earlier mapped by an agent (human or robot), via observations captured from an augmented reality system. To develop such an application, recent advancements in the field of 3D model representation and rendering could be explored such as Neural Radiance Fields (NERF) [107], [108] and multi-view data fusion.

Partner: undisclosed. **Domain:** Color images, 3D point clouds

It is worth noting that research performed under the theme of DT and its use as training environment for robots is tightly linked to the research theme SLAM and Navigation (recall section 5.2). Indeed, through extensive training and/or testing in representative digital-twin models of robots and the real-world we can develop robotic systems whose SLAM and navigation capabilities are more robust and generalizable to real-world conditions. At the same time, we can better assess the limits of designs to conditions that would otherwise be risky or hard to access realistically, e.g. for USAR missions.

Conversely, research developments under the theme of SLAM and navigation can contribute to the construction and maintenance of DTs of real-environments via the acquisition and processing of various data modalities. Effective SLAM is particularly important for the constitution of a precise 3D representation of the physical asset with as little human-supervision as possible but also for the update of the DT, as for example in applications pertaining to inspection of remote sites (industrial infrastructures, mines, sea-beds, etc) by specialized robots.

CONCLUSION

Following decades of research, development and successful deployment of robotics in industrial assembly lines and more generally in confined workspaces, the use of robots has attracted the interest of several other sectors where human workforce remained dominant, be it to necessity or simply to superior human capacity. In the light of different technological advances in computing and driven by the evolving society needs, however, a number of indicators signal that robots have the potential to match or surpass human performance in cognitive tasks, further beyond manual labor.

Apart from purely tele-operated robots, (semi)autonomous robots inevitably need to rely on proprioceptive and exteroceptive sensing so as to continuously detect and track their state as well as that of the environment in which they operate in. Admittedly, vision is the primary skill required to capture the robot surroundings and, for many, the driving force in artificial intelligence research. **3D robotic vision** in particular encompasses fields that range from simultaneous localization and mapping (SLAM) [109], semantic scene segmentation, object detection [110] or human activity recognition, altogether allowing a robot to interpret the static as well as the dynamic components of the environment. In the earliest and also preponderant paradigm of robot architecture design, 3D vision serves at building internal, higher-level representations of the real world wherein planning and decision can take place before acting. Irrespective of the underlying robot design, sensing and understanding the 3D environment in a reliable and robust manner is the *sine qua non* for a successful robot deployment. 3D vision is thus highly concerned with the resolution of inverse problems of elevated dimensionality that stems from the geometry of a changing and potentially growing environment. The latter point is a direct consequence of robot mobility that allows to expand its workspace and therefore its utility.

Navigating in 3D thus opens new opportunities but also poses challenges related to the integrity of the environment, the surrounding people and the robot itself, not to mention user acceptability concerns. The diversity in robot platforms and designs adds another factor of complexity to the aforementioned problems, as 3D navigation inevitably

adds more degrees of freedom to a robot and increases its exposure to collisions. It further reposes on effective 3D environment sensing coupled with planning and/or control and machine learning, in order to minimize the risk of accidents and energy consumption, increase compliance to human interactions, etc while accounting for stochasticity. Conversely, 3D navigation may deliberately influence the task of 3D vision following the paradigm of active perception [111] which highlights the interest for the understanding and treatment of the two tasks.

Thanks to the advent of affordable graphical processing units (GPU) and realistic simulators combined with the emergence and establishment of deep-learning (DL), the fields of **vision and navigation** have been brought even closer since the last decade via the paradigm of *end-to-end* learning [103]. This paradigm bypasses altogether the intermediate step of constructing observable abstractions of the real-world wherein planning and decision take place, allowing raw sensor readings to be directly mapped to actions. Finally, the use of RGB-D sensors that have become commonplace in robotics has even further boosted research in 3D vision and control be it for navigation or manipulation.

In the previous chapters, I have presented my contributions in the fields of computer vision and robot navigation in 3D environments covering a period of 13 years in total, in domains of application pertaining to urban search and rescue, assistive and field robotics together with digital libraries and information-communication technology (ICT) more broadly. During this course, I had the privilege to discover different styles and practices for pursuing research and innovation, to work with diverse software tools, robot platforms and sensors while developing a spirit of teamwork, extroversion and pursuit of scientific excellence. The reported results were thus to a large extent attained thanks to coordinated and persistent work of various other people that were either directly involved or on whose earlier work I could build upon.

TEACHING

This annex lists my current and earlier teaching activities dating since the beginning of my PhD thesis. During this period, in 2015, I received the *Qualification* from CNU (Conseil National des Universités) for the section 61 - *Génie informatique, Automatique et Traitement du Signal*.

I. University of Athens, Department of Informatics and Telecommunications, Greece In the period 2005 – 2009, I supervised laboratory sessions (TP) of the following undergraduate course:

- **Computer Graphics I¹**

Covered topics: Rendering, visualization, geometric transformations, animation.

Technical skills: OpenGL, C++. Types of training: project-based. Total volume: 45 h.

II. University of Rome "La Sapienza", Rome, Italy Between 2010 – 2012, I realized theoretic sessions (CS) and supervised laboratory sessions (TP) of the following two courses of the M. Sc. in Artificial Intelligence and Robotics:

- **Computer Vision and Perception²**

Covered matter: Multi-view Geometry in Computer Vision. Technical skills: Matlab.

Types of training: project-based. Total volume: 45 h.

- **Pattern Recognition²**

Covered matter: Linear and non-linear classification, probabilistic graphical models.

Types of training: Lecturing. Total volume: 23 h.

III. ENSTA ParisTech, Paris, France Between 2015 and 2016, I supervised laboratory sessions (TP) of the following two courses:

¹Teaching language : Greek

²Teaching language : English

- **Programmation et Algorithmique en Python³** : Types of training: project-based. Total volume: 21 h.
- **Projet Informatique²** : Types of training: project-based. Total volume: 21 h.

IV. IMT Atlantique, Brest, France

Course responsible Since 2017 I have been the coordinator of the following two courses at M.Sc. level :

- **C programming language²** (MSC-INF101B): The course is part of the 1st semester of the MSc IT (Information Technology) of IMT Atlantique, Brest Campus, containing 18 hours in total that are divided into ordinary classes and laboratory sessions. Technical skills : C, gcc, gdb, Make. Types of training : lecturing, project-based. Total volume : 115 h
- **Programming of Robotic Systems²** (PA-DI-BOT-B): The course is provided as an option to the second or third year students of TAFs DCL and OPE, with 40 hours in total divided into ordinary classes and laboratory sessions. The syllabus of the course contains the following themes : (i) Representation & Modeling fundamentals, (ii) Sensing & Observation, (iii) State Estimation, Robot Localization, (iv) Planning, Graph Search, (v) Image representation, features, motion estimation and (vi) Robotic manipulation principles. Each of these is accompanied by a corresponding laboratory project where students learn to program using the ROS (Robot Operating System) middle-ware, in simulation as well as with real mobile robots. Technical skills : ROS, 3D physical simulation, Python/C++. Types of training : lecturing, project-based. Total volume : 226 h

Participation to other courses and diverse implications Apart from the previously mentioned, recurrent responsibilities, I have been implicated at various degrees in the teaching of courses of other departments, in particular :

- **Realization of prescribed TP/TD/CS sessions** - I have contributed in the realization of sessions of the following courses (ordered by decreasing order of volume):

³Teaching language: French

-
- **Génie logiciel et programmation objet**³ (INF-205, resp. Julien MALLET), dept. INFO. Technical skills : Java, UML. Types of training : project-based. Total volume : 62 h.
 - **Bases de données**³ (INF-206, resp. Laurent BRISSON), dept. LUSSE. Technical skills : SQL, Java. Types of training : lecturing, project-based. Total volume : 58 h.
 - **Articulation machine-langage-OS**³ (SIT-151, resp. André THEPAUT), dept. INFO. Technical skills : C. Types of training: project-based. Total volume : 20 h.
 - **Linux & environnement informatique**³ (TC131C, resp. Eric COUSSIN), dept. INFO. Technical skills : System programming, bash. Types of training : Directed work, Total volume : 7 h.
 - **Intelligence Artificielle**² (PA-DI-IAIMP-B, resp. Nicolas FARUGGIA), dept. ELEC. Technical Skills : Python, deep-learning frameworks. Types of training : Lecturing. Volume totale : 2 h.
 - **Creation of new TP/TD/CS sessions**: Since 2019, I have been contributing to the course **Computer Vision**² (PA-DI-COMPVISIO-B) and since 2022 to the course **RAUG - Réalité Augmentée**². The topics covered are monocular and stereoscopic vision and more particularly, single-camera modeling, calibration, two-view geometry and camera-motion estimation. Technical Skills : Python, OpenCV. Types of training : lecturing, project-based. Total volume : 71 h.
 - **Student project supervision**: Proposed and (co)supervised diverse projects of the common education program since 2016, namely: (i) Projet Filière - Systèmes Logiciels et Réseaux - SLR), (ii) Projet (Co)Développement - FISE, (iii) Projet Bibliographique - MSc-IT-Brest, (iv) Projet S4, (v) Projet Recherche. Total volume : the equivalent of 20 h of TP per year.
 - EU project **SmartSoc - Education of Future ICT Experts based on Smart Society Needs**, 2021–2022. Technical supervisor of students developing entrepreneurial cases. Total volume : 15 h.
 - **Seminars**²

- Implicated to the realization of the winter school entitled **WASAA** - Western Alliance for Scientific Actions with Australia, editions 2019, 2020, 2023, organized by IMT Atlantique. My contribution consisted to the realization of theoretic sessions and the proposition and supervision of research projects. Total volume : 30 h.
- Chaire **M@D** (Maintien en Domicile) - 2018 : TP on the middleware ROS (Robot Operating System) to engineers of ENSIBS. Total volume : 3 h.
- **Science popularization**²: Participation to the event **Journée Métiers du numérique** organized by UBL in 2016, by introducing the field of computer science to high-school students via interactive workshops. Total volume : 3 h.

Perspectives and links between teaching and research

The synergy between my teaching and research activities is articulated within the general-engineer curriculum prescribed by IMT Atlantique. The course “Programming of Robotic Systems” serves as initiation to the fundamentals of mobile robotics and will serve as entry to a new course, oriented towards robot-learning and probably more dependent on 3D physical simulation, in order to concentrate more onto the algorithmic and data aspects of robotics. In a more direct connection to my research activities, the monocular and stereo-vision parts of the courses “Computer Vision” and “Augmented Reality” can serve as a basis for the introduction of the theme of SLAM - Simultaneous Localization and Mapping - using passive sensors. Further beyond the perimeter of IMT Atlantique, a second objective of valorization and of promotion of the synergy between teaching and research concerns the publications of tutorials, namely, the publication of a special type of article that is pedagogical and more general public, potentially accompanied by open-source code. Finally, in the upcoming years I will be led to develop the collaboration of IMT Atlantique and the NTUA - National Technical University of Athens - in the domain of education and research on the basis of the underlying mutual engagements.

Coordination

This section lists activities for which I have been involved as (co)supervisor/(co)responsible.

Responsibilities

- **RAMBO** (Robot interaction, Ambient system, Machine learning, Behavior, Optimization) **team co-leader**, Lab-STICC, UMR 6285, France, 2020-...
The team consists of 8 faculty and 11 doctoral and post-doctoral researchers, reuniting IMT Atlantique, ENIB and UBS www.imt-atlantique.fr/en/research-innovation/teams/research-team-rambo. Members of the team have a longstanding implication/participation in the international contest *RoboCup*, having won RoboCup@home education 2020 Standard Platform/Open Category and its co-leader Prof. Dominique Duhaut is the organizer of RoboCup 2023. The team disposes state-of-the-art research facilities and equipment, notably a LivingLab in eHealth www.imt-atlantique.fr/sites/default/files/recherche/plateformes/en/experiment_haal.pdf.
- **Founding member and associate** of International Research Lab (IRL) **CROSSING** (frenCh austRalian labOratory for humanS / autonomouS agents teamING), CNRS - Naval Groupe - Australia, crossing.cnrs.fr/, 2018-...
I was among the IMT Atlantique delegation for the launching of the lab and I am currently co-supervising a joint PhD thesis with University of Adelaide.

Post-doc (co)supervision

- **Maxime Devanne, IMT Atlantique, Brest, 2018**, Project AMUSAAL *Analyse automatique de Mouvements humains par décomposition en mouvements Simples pour l'Ambient Assisted Living* (publication: 19.), cosupervised with Sao Mai Nguyen, [completed](#)

- **Alireza ASVADI**, IMT Atlantique, Brest, 2020-2022, Project HERON *Habitat intelligent Et RObotique d’assistance basés sur le jumeau Numérique* (publications: 14., 48.), cosupervised with Christophe Lohr, [completed](#)
- **Pedro Juan SOTO VEGA**, Ifremer, Brest, 2022-2023, Project ABYSSES *Annotation Boostée des modèles 3D de monts hYdrothermaux Sous-marins par apprentiSSage profond d’adaptation de domainES*, cosupervised with Marjolaine Matabos, [ongoing](#)
- **Ikram KOURBANE**, IMT Atlantique, Brest, 2022-2024, Project ECFvisuL, *Évaluation automatique continue de la capacité fonctionnelle humaine à domicile à l’aide de la vision par ordinateur*, cosupervised with Mihai Andries, [ongoing](#)
- **Hajer FRADI**, IMT Atlantique, Brest, 2023-2024, Project LEASARD *Low-Energy deep neural networks for Autonomous Search-And-Rescue Drones*, cosupervised with Guillaume Moreau, [ongoing](#)

Ph.D. thesis (co)supervision

- **Andrei MITRIAKOV**, IMT Atlantique, Brest, 2018-2022, Thesis: *Reinforcement Learning-based Control for Safe 3D Navigation of Articulated Tracked Robot Manipulators*. Director: Serge GARLATTI (IMT Atlantique), Lab-Sticc (publications: 1., 2., 14., 17., 18., 47.), [completed](#)
- **Mohamed-Khalil JABRI**, IMT Atlantique, Brest, 2020-2023, Thesis: *RObot Manipulation Learning by Demonstration using Generative Adversarial Networks*. Director: Gilles COPPIN (IMT Atlantique) & Javen SHI (University of Adelaide, AU), co-supervisor: Ehsan ABBASNEJAD (University of Adelaide, AU), Lab-Sticc & Australian Institute of Machine Learning (publication: [18]), [ongoing](#)
- **Yassine HABIB**, IMT Atlantique, Brest, 2020-2023, Thesis: *Dense and Real-time 3D Metric Mapping*. Director: Cédric BUCHE (ENIB), co-supervisor: Cédric LE BARZ (Thales Palaiseau), Lab-Sticc (publications: 12., 46.), [ongoing](#)
- **Cédric LE BONO**, IMT Atlantique, Brest, 2020-2023, Thesis: *Proactive Assistance to ADLs by Contextual Robotic Interaction*. Director: Cédric BUCHE (ENIB), co-supervisor: Christophe LOHR (IMT Atlantique) (publication: 45.), [ongoing](#)

-
- **Romain CAZORLA, ENIB, Brest, 2020-2023**, Thesis: *Semantic Segmentation of Industrial 3D Scenes*. Director: Cédric BUCHE (ENIB), co-supervisor: Line POINEL (Segula Technologies) (publications: 13., 16.), ongoing
 - **Quentin FERDINAND, ENSTA Bretagne, Brest, 2020-2023**, Thesis: *Incremental Learning for Classification of Objects of Interest*. Director: Benoît CLEMENT (ENSTA Bretagne), co-supervisors: Gilles LE CHENADEC (ENSTA Bretagne), Quentin OLIVEAU (Naval Group) (publication: 15.), ongoing
 - **Nacer DJERADA, IMT Atlantique, Brest, 2023-2026**, Thesis: *Travel assistance for visually impaired people based on computer vision*. Director: Panagiotis PAPADAKIS (IMT Atlantique), co-supervisor: Christophe LOHR (IMT Atlantique), ongoing

M.Sc. thesis (co)supervision

- **Valsamis NTOUSKOS, University of Rome “La Sapienza”, 2011-2012**, Thesis: *Classification of Human Actions from Motion Capture Sequences* (publications: 24., 30., 31.), [completed](#)
- **Aimenallah SAIAD-SAHRAOUI, IMT Atlantique, 2017-2018**, Thesis: *Dynamic Obstacle Avoidance for Indoor Robot Platooning*, [completed](#)
- **Andreea-Oana PETAC, ENIB, 2017-2018**, Thesis: *Robot detection and localization by machine learning and computer vision algorithms*, [completed](#)

Tutoring : From 2017 up to date, I have been tutor for the final year internships (SFE) of the following M.Sc. students :

Fei Zhao, Michael Matyn, Sekou Traore, Chen Zhen, Thomas Dambricourt, Romain Namyst, Cédric Le Bono, Hanyuan Zhang, Mao Yuan, Mohamed Amine Ben Amira, Quentin Duffeau

(Co)chairing - event organization

- Fête de la science 2015, ENSTA ParisTech
- AAAI Fall Symposium on *Robot Human Teamwork on Dynamic Adverse Environments*, Arlington, USA, 2011, [112]

- International Conference on Computer Vision Theory and Applications, Italy, 2012
- International Conference on Pattern Recognition Applications and Methods, Portugal, 2012
- EU Robotics Week - 2011, University of Rome “La Sapienza”
- Web-site development for 1st Eurographics workshop on 3D Object retrieval 2008 diglib.org/handle/10.2312/313.

Expertise

This section lists my implication in providing scientific expertise.

Mentorship

- Doctorate thesis committee representative (CSI): ENSTA Bretagne & Orange Lannion (CIFRE), 2019-2022, Thesis: *Moteur de dialogue multimodal adaptatif multicapteurs et multi-actionneurs appliqué à une plateforme robotique*
- Doctorate thesis committee member (CSI): IMT Atlantique & CHRU Brest, 2023-2025, Thesis: *Longitudinal follow-up of liver metastases from colorectal cancer using artificial intelligence*
- Technical expert for Entrepreneurial Cases : IMT Atlantique, 2021-2022, EU Project SmartSoc - Education of Future ICT Experts based on Smart Society Needs:
 - * Case 2021: *Safe, effective and sustainable COVID-19 protection operations in congested areas*
 - * Case 2022: *Google Street Map-based Route Analysis for People with Disabilities*
- Robotics & Automation Society, YRP Senior Reviewer

Research proposals reviewer

- Croatian Science Foundation (HRZZ), 2013 Call for Research Projects

-
- CNRS INS2I, International Research Project (IRP) 2022
 - Thesis proposal reviewer: Region of Brittany, France, 2022

Public tender evaluator

- **Evaluator** of public tenders for robots acquired by Télécom-Bretagne, in the context of project **CPER** (Contrat Plein État Région) - **VITAAL** (Vaincre l’Isolement par les TIC par l’ “Ambient Assisted Living”), 2016-2019. Tasked with assessing the quality of received offers for the acquisition of : (i) a tracked robot manipulator, (ii) a humanoid robot, (iii) a tele-presence robot and (iv) a fleet of small robot manipulators https://www.imt-atlantique.fr/fr/recherche-innovation/equipe/rambo?arg=5846_7.

Jury/committee

- Homi Bhabha Institute, India, 2016, **Ph.D** thesis of Mr. V. Satish Kumar Reddy: *Mobile Robot Navigation in an Outdoor Environment*, dissertation **reviewer**
- Electrical Engineering Department of Cybernetics, Czech Technical University in Prague, 2022, **Ph.D** thesis of Mr. Vojtěch Šalanský: *Robot Learning and Perception in Sensory Deprived Environment*, dissertation **reviewer**
- Electrical Engineering Department of Cybernetics, Czech Technical University in Prague, 2023, **Ph.D thesis** of Mr. Teymur Azayev: *Robotic control with deep-learned structured policies*, dissertation **reviewer**
- ENSTA Bretagne, Brest, France, 2023, **Ph.D thesis** of Mr. Rémi RIGAL: *Moteur de dialogue multimodal adaptatif multicapteurs et multi-actionneurs appliqué à une plateforme robotique*, dissertation **examiner**
- Recruitment of Assoc. Professor in the department of Informatics, IMT Atlantique, team RAMBO, 2020, jury member
- Recruitment of Assoc. Professor in the department of Informatics, IMT Atlantique, team RAMBO, 2023, jury member
- INSA Rennes, Master Research Internship - (SIF) - of Mr. Lucas Pélerin: *Simulation de foule par apprentissage*, **M.Sc.** thesis **reviewer**

- Member of Commission de Qualification des Promotions et Changement d'Appellations - MC/CR/CER/IE for IMT Atlantique, 2023-2026
- Department of Production Engineering and Management, Demokritos University of Thrace, Greece - external registry member, 2023

Article reviewer

- **Conferences:** International Conference on Intelligent Robots and Systems (IROS), International Conference on Robotics and Automation (ICRA), Safety Security and Rescue Robotics (SSRR), Field and Service Robotics (FSR), International Conference on Systems, Man and Cybernetics (SMC)
- **Journals:** Journal of Field Robotics (JFR), Pattern Recognition (PR), IEEE Transactions on Intelligent Vehicles (TIV), Mechanism and Machine Theory (MMT), IET Computer Vision, Transactions on Image Processing (TIP), Engineering Applications of Artificial Intelligence (EAAI), Transactions on Cognitive and Developmental Systems (TCDS), Robotics and Autonomous Systems (RAS), Transactions on Multimedia (TM), Image and Vision Computing (IVC), Robotica, International Journal of Digital Earth, International Journal of Machine Vision and Applications, Computer-Aided Design (CAD)

CURRICULUM VITAE

Surname, first name: PAPADAKIS Panagiotis
Telephone: +33 (0) 2 29 00 16 30
E-mail: panagiotis.papadakis@imt-atlantique.fr
Web-page: <https://sites.google.com/site/pgpapadakis/>

Professional affiliation:

Associate professor and team co-leader, Department of Informatics
IMT Atlantique Bretagne/Pays de la Loire
Team RAMBO, CNRS Lab-STICC, UMR 6285
655, Avenue du Technopôle 29280 Plouzané

EDUCATION

Ph.D. in Information Technology (with honours)	2005-2009
Content-based 3D model retrieval considering the user's relevance feedback <i>University of Athens, Informatics & Telecommunications, Greece</i>	
B.Sc. (4 years) in Informatics and Telecommunications	2001-2005
<i>University of Athens, Informatics & Telecommunications, Greece</i>	

CERTIFICATIONS & TRAINING

Qualification: Section 61 - Génie informatique, Automatique et Traitement du Signal
English: Cambridge Certificate of Proficiency
French: Diplôme d' Études en Langue Française 1er Degré
Career seminars: Gestion de projet, gestion de conflit, animation de réunion
Music: Harmony & Proficiency in Organ

R&D PROJECT PARTICIPATION

- LEASARD:** Low-Energy deep neural networks for Autonomous Search- 2022-2025
And-Rescue Drones, ANR CominLabs, France, role : WP leader
- ECFvisuL:** Continuous and automatic evaluation of human functional 2022-2024
capacity at home based on computer vision, France
- TwinCoBot:** Digital twins of human and cobots in smart environments, 2022
Carnot Institute, France
- HERON:** Habitat intelligent Et Robotique d'assistance basés sur le 2021-2022
jumeau Numérique, France, role : PI
- ABYSSES:** Annotation Boostée des modèles 3D de monts hYdrother- 2022-2023
maux Sous-marins par apprentiSsage profond d'adaptation de domainES,
France
- SmartSOC:** Education of Future ICT Experts Based on Smart Society 2019-2022
Needs, EU
- REACT:** Modèles d'Interaction Physique de Robots Compagnons, 2018-2022
France, role : PI
- ROGAN:** Apprentissage par Démonstration de Manipulation Robot via 2020- ...
des Réseaux Adverses Génératifs, France/Australia, role : PI
- SPARC:** aSsistance Proactive aux AVQ par interaction Robotique Con- 2020-2023
textuelle, France
- ILCOI:** Incremental Learning for Classification of Objects of Interest, 2020-2023
CIFRE, France
- SSSI3D:** Segmentation Sémantique d'une Scène Industrielle 3D, Indus- 2020-2023
trial, France
- REMED:** REconstruction Métrique 3D en temps réel, CIFRE, France, 2020-2023
role : PI
- AMUSAAL:** Analysis of Human Movements by Decomposition to 2018
Atomic Actions for Ambient Assisted Living, France, role : PI
- COMRADES:** Coordinated Multi-Robot Assistance Deployment in 2017
Smart Spaces, IMT, role : PI
- ROMEO2:** Humanoid Robot Companion by Aldebaran, France 2014-2016
- P2N:** Perception to Navigation, INRIA, France 2012-2014
- NIFTi:** Natural Human Robot Coop. in Dynamic Environments, EU 2010-2012
- Cognitive search and retrieval of 3D graphical models,** Greece 2006-2009

HELD POSITIONS

ENSTA ParisTech, Palaiseau, France	2014-2016
Post-doctoral fellow, team U2IS & INRIA FLOWERS <i>Perception and mapping of human interactions for assistive robotics</i>	
INRIA Grenoble & Sophia-Antipolis, France	2012-2014
Post-doctoral fellow, team e-Motion & LAGADIC <i>Perception and mapping of human interactions for assistive robotics</i>	
University of Rome "La Sapienza", Department of Computer Control and Management Engineering, Italy	2010-2012
Post-doctoral fellow, team ALCOR, Vision, Perception and Cognitive Robotics Laboratory <i>Robotic vision, motion planning</i>	
Hellenic Army	2009-2010
Research and Informatics Division <i>Military network administration, communications</i>	
University of Athens, Department of Informatics & Telecommunications & NCSR "Demokritos", Institute of Informatics & Telecommunications	2006-2009
Doctoral candidate, Computational Intelligence Laboratory	

VOLUNTEERING

End-user testing	2013-2014
<i>Robot arm control, ENSTA PT, doi: 10.1109/IROS.2015.7353107</i> <i>Robot wheelchair control, INRIA, doi: 10.1109/IROS.2014.6943021</i>	
Olympic Games, Athens	2004
<i>Workstation Technician in Athens Olympic Stadium</i>	

INVITED TALKS & SEMINARS

- *3D terrain perception of mobile robots based on shape and appearance: state-of-the-art, challenges and future trends*, 2012, Foundation for Research and Technology (FORTH), Institute of Computer Science, Crete, Greece
- *Semantic Mapping for Indoor Environments*, 2015, Journées Nationales de la Recherche en Robotique (JNRR), Cap Hornu, France
- *From shape descriptors to real world object-based mapping in robotics*, 2015, LAAS-CNRS, Toulouse, France
- *Object Classification in Indoor RGB-D Scenes using Projective Light Diffusion Image Expansion*, 2016, GDR-ISIS, Télécom ParisTech, Paris, France
- *Object Classification in Indoor RGB-D Scenes using Projective Light Diffusion Image Expansion*, 2016, Télécom Bretagne, Brest, France, Lab-STICC Seminar
- *Generic Object Discrimination for Mobile Assistive Robots using Projective Light Diffusion & Binding Human Spatial Interactions with Mapping for Enhanced Mobility in Dynamic Environments*, University of Adelaide, Adelaide, Australia, FASIC Workshop 2018.
- *Construction de vérité terrain d'activités de la vie quotidienne au sein du Living Lab de l'IMT-Atlantique*, Le Couvent des Jacobins - Centre des Congrès de Rennes Métropole, Journée "Connecter les Usages" - Connectathon, 2019
- *Project interview, AMUSAAL (Analyse automatique de Mouvements hUains par décomposition en mouvements Simples pour l'Ambient Assisted Living)*, Tébéo - television channel, Chemins de Traverse series, episode 20 – 03 – 2019
- *Robotic Skill Learning and Development in Personal Assistance*, [CRNS@Webinar](#) of the University of Plymouth, England, 2021.

Journal articles

1. A. Mitriakov, **P. Papadakis**, S. Garlatti, *An Open-Source Software Framework for Reinforcement Learning-based Control of Tracked Robots in Simulated Indoor Environments*, **Advanced Robotics**, vol. 36, no. 11, pp. 519-532, 2022, [Impact Factor \(JCR\)=2.057](#)
2. A. Mitriakov, **P. Papadakis**, J. Kerdreux, S. Garlatti, *Reinforcement Learning based, Staircase Negotiation Learning in Simulation and Transfer to Reality for Articulated Tracked Robots*, **IEEE Robotics & Automation Magazine**, vol. 28, no. 4, pp. 10-20, 2021, [Impact Factor \(JCR\)=5.229](#), also presented in **IEEE ICRA 2022**
3. **P. Papadakis**, P. Rives, *Binding Human Spatial Interactions with Mapping for Enhanced Mobility in Dynamic Environments*, **Autonomous Robots**, vol. 41, no. 5, pp. 1047–1059, 2017, [Impact Factor \(JCR\)=3.255](#)
4. **P. Papadakis**, *Enhanced Pose Normalization and Matching of Non-Rigid Objects based on Support Vector Machine Modelling*, **Pattern Recognition**, vol. 47, no. 21, pp. 216-227, 2014, [Impact Factor \(JCR\)=8.518](#)
5. A. Sinha, **P. Papadakis**, *Mind the Gap: Detection and Traversability Analysis of Terrain Gaps using LIDAR for Safe Robot Navigation*, **Robotica**, vol. 31, no. 7, pp. 1085-1102, 2013, [Impact Factor \(JCR\)=2.406](#), 33 citations
6. **P. Papadakis**, *Terrain Traversability Analysis Methods for Unmanned Ground Vehicles; a Survey*, **Engineering Applications in Artificial Intelligence**, vol. 26, no. 4, pp. 1373-1385, 2013, [Impact Factor \(JCR\)=7.702](#), 242 citations
7. S. Blisard, T. Carmichael, L. Ding, T. Finin, W. Frost, A. Graesser, M. Hadzikadic, L. Kagal, G.J.M. Kruijff, P. Langley, J. Lester, D.L. McGuinness, J. Mostow, **P. Papadakis**, R. Prasad, S. Stoyanchev, P. Varakantham, *Reports of the AAAI 2011 Fall Symposia*, **AI Magazine**, vol. 33, no. 1, pp. 71-78, 2012 [Impact Factor \(JCR\)=2.524](#)

⁴JCR stands for “Journal Citation Reports” metrics. The number of citations is provided for a selected subset of publications.

8. **P. Papadakis**, I. Pratikakis, T. Theoharis, S. Perantonis, *PANORAMA: A 3D Shape Descriptor based on Panoramic Views for Unsupervised 3D Object Retrieval*, **International Journal of Computer Vision**, Special Issue on: 3D Object Retrieval, vol. 89, no. 2, pp. 177-192, 2010, [Impact Factor \(JCR\)=13.369](#), 239 citations
9. A. Agathos, I. Pratikakis, **P. Papadakis**, N. Sapidis, F. Azariadis, *3D Articulated Object Retrieval using a Graph-based Representation*, **Visual Computer**, Special Issue for 3D Object Retrieval, vol. 26, no. 1, pp. 1301-1319, 2010, [Impact Factor \(JCR\)=2.835](#), 36 citations
10. **P. Papadakis**, I. Pratikakis, T. Trafalis, T. Theoharis, S. Perantonis, *Relevance Feedback in Content-based 3D Object Retrieval: A Comparative Study*, **Computer-Aided Design and Applications Journal**, vol. 5, no. 5, pp. 753-853, 2008, [Impact Factor \(JCR\)=0.557](#)
11. **P. Papadakis**, I. Pratikakis, S. Perantonis, T. Theoharis, *Efficient 3D Shape Matching and Retrieval using a Concrete Radialized Spherical Projection Representation*, **Pattern Recognition**, vol. 40, no. 9, 2007, pp. 2437-2452. [Impact Factor \(JCR\)=8.518](#), 242 citations

Articles in proceedings

12. Y. Habib, **P. Papadakis**, C. Le Barz, A. Fagette, T. Gonçalves, C. Buche, *Densifying SLAM for UAV navigation by fusion of monocular depth prediction*, **IEEE International Conference on Automation, Robotics and Applications (ICARA)**, 2023
13. R. Cazorla, L. Poinel, **P. Papadakis**, C. Buche, *Reducing domain shift in synthetic data augmentation for semantic segmentation of 3D point clouds*, **IEEE International Conference on Systems, Man and Cybernetics (SMC)**, 2022, [doi](#), [B](#)
14. A. Asvadi, A. Mitriakov, C. Lohr, **P. Papadakis**, *Digital Twin Driven Smart Home: A feasibility study*, **IEEE International Conference On Smart Living and Public Health**, 2022, [doi](#)

-
15. Q. Ferdinand, B. Clement, Q. Oliveau, G. Le Chenadec, **P. Papadakis**, *Attenuating Catastrophic Forgetting by Joint Contrastive and Incremental Learning*, **IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)**, Workshop on Continual Learning, 2022, [doi](#), [A*](#)
16. R. Cazorla, L. Poinel, **P. Papadakis**, C. Buche, *Bottleneck Identification to Semantic Segmentation of Industrial 3D Point Cloud Scene via Deep Learning*, **International Joint Conference on Artificial Intelligence (IJCAI)**, 2021, [doi](#), [A*](#)
17. A. Mitriakov, **P. Papadakis**, S. M. Nguyen, S. Garlatti, *Staircase Negotiation Learning for Articulated Tracked Robots with Varying Degrees of Freedom*, **IEEE International Conference on Safety, Security and Rescue Robotics (SSRR)**, 2020, [doi](#)
18. A. Mitriakov, **P. Papadakis**, S. M. Nguyen, S. Garlatti, *Staircase Traversal via Reinforcement Learning for Active Reconfiguration of Assistive Robots*, **IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**, 2020, [doi](#), [A](#)
19. M. Devanne, **P. Papadakis**, S. M. Nguyen, *Recognition of Activities of Daily Living via Hierarchical Long-Short Term Memory Networks*, **IEEE International Conference on Systems, Man and Cybernetics (SMC)**, 2019, [doi](#), [B](#)
20. **P. Papadakis**, D. Filliat, *Generic Object Discrimination for Mobile Assistive Robots using Projective Light Diffusion*, **IEEE Winter Conference on Applications on Computer Vision (WACV)**, Workshop on Computer Vision for Active and Assisted Living (CV-AAL), 2018, [doi](#), [A](#)
21. **P. Papadakis**, C. Lohr, M. Lujak, A.-B. Karami, I. Kanellos, G. Lozenguez, A. Fleury, *System Design for Coordinated Multi-Robot Assistance Deployment in Smart Spaces*, **IEEE International Conference on Robotic Computing**, 2018, [doi](#)
22. M. Lujak, **P. Papadakis**, A. Fernandez, *Endowing Mobile Robot Teams with Ambient Intelligence for Improved Patient Care*, **International Workshop on Artificial Intelligence and Robotics**, 2017, [uri](#)

23. P. Papadakis, *A Use-Case Study on Multi-View Hypothesis Fusion for 3D Object Classification*, **IEEE International Conference on Computer Vision (ICCV)**, Multi-View Relationships in 3D Data workshop, 2017, [doi](#), [A*](#)
24. V. Ntouskos, P. Papadakis, F. Pirri, *Probabilistic Discriminative Dimensionality Reduction for Pose-Based Action Recognition*, **Advances in Intelligent Systems and Computing (selected papers from ICPRAM 2013)**, 2015, [doi](#)
25. R. Drouilly, P. Papadakis, P. Rives, B. Morisset, *Local Map Extrapolation in Dynamic Environments*, **IEEE International Conference on Systems, Man and Cybernetics (SMC)**, 2014, [doi](#), [B](#)
26. A. Sinha, P. Papadakis, M.-R. Elara, *A staircase detection method for 3D point clouds*, **IEEE International Conference on Control Automation Robotics & Vision (ICARCV)**, 2014, [doi](#)
27. P. Papadakis, P. Rives, A. Spalanzani, *Adaptive Spacing in Human-Robot Interactions*, **IEEE International Conference on Intelligent Robots and Systems (IROS)**, 2014, [doi](#), [A](#), 41 citations
28. P. Papadakis, *The Canonically Posed 3D Objects Dataset*, **Eurographics Workshop on 3D Object Retrieval**, 2014, [doi](#)
29. P. Papadakis, A. Spalanzani, C. Laugier, *Social Mapping of Human-Populated Environments by Implicit Function Learning*, **IEEE International Conference on Intelligent Robots and Systems (IROS)**, 2013, [doi](#), [A](#), 40 citations
30. V. Ntouskos, P. Papadakis, F. Pirri, *Discriminative Sequence Back-Constrained GP-LVM for MOCAP Based Action Recognition*, **International Conference on Pattern Recognition Applications and Methods**, 2013, [doi](#)
31. V. Ntouskos, P. Papadakis, F. Pirri, *A Comprehensive Analysis of Human Motion Capture Data for Action Recognition*, **International Conference on Computer Vision Theory and Applications**, 2012, [doi](#), [B](#)

-
32. G.J.M. Kruijff, M. Janicek, S. Keshavdas, B. Larochelle, H. Zender, N.J.J.M. Smets, T. Mioch, M.A. Neerincx, J. van Diggelen, F. Colas, M. Liu, F. Pomerleau, R. Siegwart, V. Hlavac, T. Svoboda, T. Petricek, M. Reinstein, K. Zimmermann, F. Pirri, M. Gianni, **P. Papadakis**, A. Sinha, P. Balmer, N. Tomatis, R. Worst, T. Linder, H. Surmann, V. Tretyakov, S. Corrao, S. P. Wanczura, M. Sulk, *Experience in System Design for Human Robot Teaming in Urban Search & Rescue*, **International Conference on Field and Service Robotics (FSR)**, 2012, [doi](#), [A](#), 127 citations
33. G.-J. Kruijff, F. Pirri, M. Gianni, **P. Papadakis**, M. Pizzoli, A. Sinha, E. Pianese, S. Corrao, F. Priori, S. Febrini, S. Angeletti, V. Tretyakov, T. Linder, *Rescue Robots at Earthquake Hit Mirandola, Italy, a Field Report*, **IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)**, 2012, [doi](#), 157 citations
34. **P. Papadakis**, M. Gianni, M. Pizzoli, F. Pirri, *Constraint-free Topological Mapping and Path Planning by Maxima Detection of the Kernel Spatial Clearance Density*, **International Conference on Pattern Recognition Applications & Methods**, 2012, [doi](#)
35. M. Gianni, **P. Papadakis**, F. Pirri, M. Liu, F. Pomerleau, F. Colas, K. Zimmermann, T. Svoboda, T. Petricek, G. Kruijff, H. Zender, H. Khambhaita, *A Unified Framework for Planning and Execution Monitoring of Mobile Robots*, **AAAI Workshop on Automated Action Planning for Autonomous Robots**, 2011, [uri](#)
36. M. Gianni, **P. Papadakis**, F. Pirri, M. Pizzoli, *Awareness in Mixed Initiative Planning*, **AAAI Fall Symposium on Robot-Human Teamwork in Dynamic Adverse Environments**, 2011, [uri](#)
37. **P. Papadakis**, F. Pirri, *Consistent Pose Normalization of Non Rigid Shapes using One-class Support Vector Machines*, **IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)**, Workshop on Non Rigid Shape Analysis and Deformable Image Alignment, 2011, [doi](#), [A*](#)
38. H. Khambhaita, G.-J. Kruijff, M. Mancas, M. Gianni, **P. Papadakis**, F. Pirri, M. Pizzoli *Help Me to Help You: How to Learn Intentions, Actions and Plans*, **AAAI Spring Symposium**, 2011, [uri](#)

39. A. Agathos, I. Pratikakis, **P. Papadakis**, S. Perantonis, P. Azariadis, N. Sapidis, *Retrieval of 3D Articulated Objects Using a Graph-based Representation*, **Eurographics Workshop on 3D Object Retrieval**, 2009, [doi](#), 56 citations
40. **P. Papadakis**, I. Pratikakis, T. Theoharis, S. Perantonis, *CIL3D: A content-based 3D model search engine*, **ACM International Conference on Digital Interactive Media in Entertainment and Arts**, 2008, [doi](#)
41. **P. Papadakis**, I. Pratikakis, T. Theoharis, G. Passalis, S. Perantonis, *SHREC'08 Entry: Hybrid IIT-NCSR-Demokritos*, **IEEE International Conference on Shape Modeling and Applications**, 2008, [doi](#)
42. **P. Papadakis**, P. Papadakis, I. Pratikakis, T. Theoharis, G. Passalis, S. Perantonis, *3D Object Retrieval using an Efficient and Compact Hybrid Shape Descriptor*, **Eurographics Workshop on 3D Object Retrieval**, 2008, [doi](#), 118 citations
-

International Patents

43. **P. Papadakis**, G. Passalis, I. Pratikakis, T. Theoharis, S. Perantonis, *Hybrid Method and System for Content-based 3D Model Search*, US 11/957,518, 30 citations
44. **P. Papadakis**, I. Pratikakis, T. Theoharis, S. Perantonis, *Method and system for content-based search of 3D models using panoramic views*, EP/09386018.7-2218
-

Posters

45. C. Le Bono, **P. Papadakis**, C. Buche, *Assessment of Conformal Use of Personal Protective Equipment by Object and Human Pose Recognition*, **IEEE International Conference on Safety, Security and Rescue Robotics (SSRR)**, 2020, [HAL](#)

-
46. Y. Habib, **P. Papadakis**, C. Le Barz, C. Buche, A. Fagette, *Metric, dense and real-time 3D Reconstruction for drone autonomous navigation*, **Journées des Jeunes Chercheurs en Robotique (JJCR 2021)**, 2021, [HAL](#)
47. A. Mitriakov, **P. Papadakis**, S.M. Nguyen, S. Garlatti, *Learning-based modelling of physical interaction for assistive robots*, **Plateforme Française de l'Intelligence Artificielle (PFIA)**, 2019, [HAL](#)
48. A. Asvadi, G. Bechu, A. Beugnard, C. G. L. Cao, C. Lohr, **P. Papadakis**, Q. Perez, C. Urtado, S. Vauttier, *Model-driven deployment of Digital Twins for Smart Environments - The HUMAN at home project case study*, **Journées nationales du GDR GPL**, France, 2022, [HAL](#)
-

Accepted or under review

- R. Cazorla, L. Poinel, **P. Papadakis**, C. Buche, *Enhancing Synthetic Data Generation for Semantic Segmentation of Point Clouds*, **Geospatial Informatics XIII, Proceedings of SPIE**, 2023, to be presented
- Y. Habib, **P. Papadakis**, A. Fagette, T. Gonçalves, C. Buche, *Densifying SLAM for UAV navigation*, **Geospatial Informatics XIII, Proceedings of SPIE**, 2023, to be presented
- P. J. Soto Vega, **P. Papadakis**, M. Matabos, L. V. Audenhaege, A. Ramiere, J. Sarrazin, G. A. O. P. da Costa, *Hydrothermal Vents Substrate Classification using Convolutional Neural Networks*, **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, 2023, under review
- Undisclosed author list for anonymity, *Undisclosed title for anonymity*, **IEEE International Conference on Computer Vision (ICCV)**, 2023, under review

Technical Reports

- D. Filliat, **P. Papadakis**, *Context aware obstacle avoidance : spécifications, Livrable 5.5.10, Project ROMEO2*, 2015

- M. Gianni, **P. Papadakis**, F. Pirri, M. Pizzoli, *DR 5.1.1: Domain analysis and specifications: context scenario and skill primitives*, **EU FP7 NIFTi / ICT-247870**, Deliverable Report, 2011, [url](#)

- M. Gianni, **P. Papadakis**, F. Pirri, M. Pizzoli, *DR 5.1.2: Methods and paradigms for skill learning based on affordances and action-reaction observation*, **EU FP7 NIFTi / ICT-247870**, Deliverable Report, 2011, [url](#)

- M. Gianni, **P. Papadakis**, F. Pirri, M. Pizzoli, A. Sinha, *DR 5.2.3: Hierarchical Structure of Learned Skills, Scan-paths, Saliency Map of Activities and Communication Interfaces*, **EU FP7 NIFTi / ICT-247870**, Deliverable Report, 2011, [url](#)

- **P. Papadakis**, I. Pratikakis, S. Perantonis, T. Theoharis, *A concrete radialized spherical projection descriptor for 3D shape retrieval*, **SHREC2006: 3D Shape Retrieval Contest**, 2006, Technical Report UU-CS-2006-030, ISSN: 0924-3275, [url](#)

DISTINCTIONS & PRIZES

- UGV pilot of the team dispatched to the earthquake-hit area of Emilia Romana in Italy in 2012, which was the **pioneering mission** of a human-robot team for USAR in Europe [7]. Press articles: [IEEE Spectrum](#), [European Commission](#), [Italian Fire Brigade](#), [Swiss SRF](#)
- **Top rank** of shape descriptor (PANORAMA) [28] in : (i) computer vision benchmark [113] and (ii) retrieval of macromolecular surfaces [114] by generic descriptors
- **Invited** by IEEE Robotics & Automation journal for oral presentation of [13] in IEEE International Conference on Robotics and Automation (ICRA) 2022
- Honorary **exemption** of conference registration for presentation of article [84]
- Ranked third in International SHape REtrieval Contest [115], 2006
- Third best 2021 entrepreneurial case in EU SmartSoc project, [video](#)
- Earned grants of 600K € in total as principal investigator (PI) or project partner

PUBLIC SOFTWARE, OPEN-SOURCE & DATA

- NPCA rotation normalization, publication 11 :
drive.google.com/file/d/0ByzqyOS0Ur6RRFdnX3JuWmdMShc/view?usp=sharing
- PANORAMA 3D shape descriptor, publication 8 :
drive.google.com/file/d/0ByzqyOS0Ur6RbTBE0TFITVlra3c/view?usp=sharing
- CAPOD dataset, publication 28 :
drive.google.com/file/d/0ByzqyOS0Ur6RT0kwTGhXTVRwSEU/view?usp=sharing
- Reinforcement-learning based robot control, publication 1 :
github.com/gwaxG/robot_ws
- Data augmentation for semantic point cloud segmentation, publication 13 :
github.com/RomainCazorla/Synthetic3DPointCloudDomainShiftReduction

DOCTORATE DISSERTATION SUMMARY

Content-based 3D model retrieval considering the user's relevance feedback

The dissertation dealt with the problem of enabling precise and real-time retrieval of generic 3D models focusing on the problems of rotation normalization, discriminative shape representations and relevance feedback-based retrieval. Considering the first problem, a rotation normalization method was proposed that computes the object's principal axes using the orientation of the object's surface and a hybrid rotation normalization scheme that combines both the surface distribution and the surface orientation distribution. Considering the second problem, a set of 3D shape descriptors was proposed, namely, the Concrete Radialized Spherical Projection (CRSP), the Hybrid and Panoramic Object Representation for Accurate Model Attributing (PANORAMA) shape descriptor. The CRSP descriptor is formed by computing a spherical function-based representation and using the spherical harmonic transform. The Hybrid descriptor is formed by combining the CRSP descriptor with a depth buffer-based representation. The PANORAMA descriptor is formed by projecting the surface of a 3D object and its orientation to the lateral surfaces of a set of cylinders that are centered at the object's centroid and parallel to its principal axes. For each projection the 2D Discrete Fourier Transform and Wavelet Transform is employed. Considering the third problem, a local relevance feedback technique is proposed which is based on shifting the feature vectors of 3D objects closer to their cluster centroid.

Jury members

Theoharis Theoharis, Associate Professor, UOA

Stavros Perantonis, Research Director, NCSR "Demokritos"

Sergios Theodoridis, Professor, UOA

Stefanos Kollias, Professor, NTUA

Andreas-Georgios Stafylopatis, Professor, NTUA

Emmanouil Sagriotis, Associate Professor, UOA

Georgios Papaioannou, Lecturer, AUEB

Summary: www.di.uoa.gr/sites/default/files/documents/grad/phdbook/papadakis.pdf

APPENDIX - MULTIMEDIA

The following videos have been produced to demonstrate the behavior of developed systems and promote the associated research to the community.

Reinforcement Learning based, Staircase Negotiation Learning in Simulation and Transfer to Reality for Articulated Tracked Robots

Andrei Mitriakov, Panagiotis Papadakis,
Jérôme Kerdreux, Serge Garlatti

IMT Atlantique, Lab-STICC, UMR 6285, team RAMBO, F-29238 Brest, France



Video # 1: bit.ly/3NgUdSb



Video # 2: bit.ly/3yhjS99

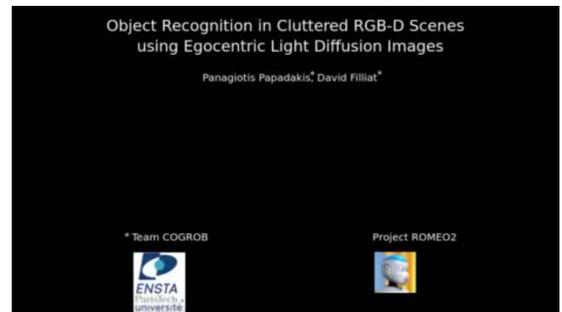
Staircase Traversal via Reinforcement Learning for Active Reconfiguration of Assistive Robots

Andrei Mitriakov, Panagiotis Papadakis,
Sao Mai Nguyen, Serge Garlatti

Lab-STICC, UMR 6285 F-29238, team IHSEV/RAMBO
IMT Atlantique
Brest, France



Video # 3: bit.ly/3tY9VuK



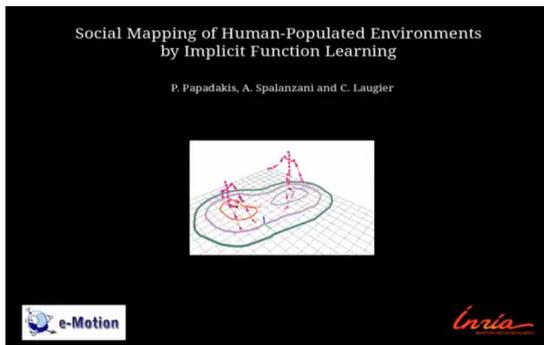
Video # 4: youtu.be/iBDop_BykcA



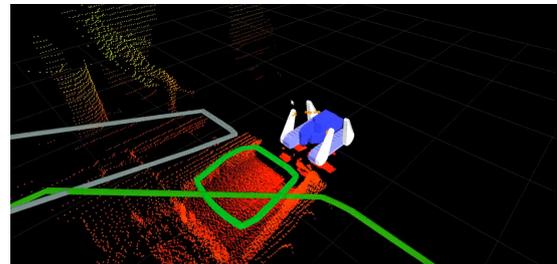
Video # 5: youtu.be/8SSDdBm5sRo



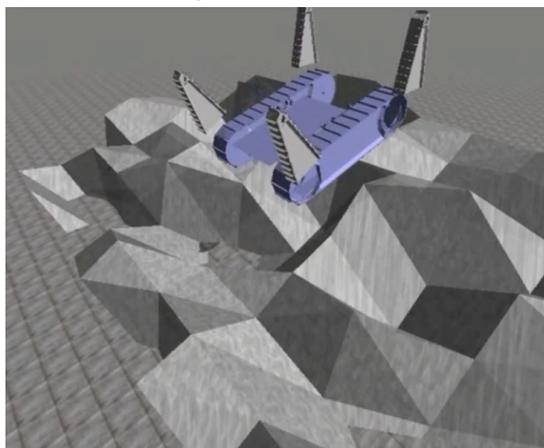
Video # 6: youtu.be/u0_5MEDC57s



Video # 7: youtu.be/Dno7XIoeCR8



Video # 8: youtu.be/9x3NiTtwrq8



Video # 9: youtu.be/y1Yh47SB9Xk



Video # 10: youtu.be/uyh5Ho-xPRc



Video # 11: youtu.be/yNEqckC1MJI

BIBLIOGRAPHY

- [1] P. Papadakis, “Content-based 3d model retrieval considering the user’s relevance feedback,” Ph.D. dissertation, Department of Informatics & Telecommunications, University of Athens, Greece, 2009.
- [2] P. Papadakis and F. Pirri, “Consistent pose normalization of non-rigid shapes using one-class support vector machines,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), NORDIA Workshop*, 2011.
- [3] P. Papadakis, “Enhanced pose normalization and matching of non-rigid objects based on support vector machine modelling,” *Pattern Recognition*, vol. 47, no. 1, pp. 216–227, 2014.
- [4] P. Papadakis, “The Canonically Posed 3D Objects Dataset,” in *Eurographics, Workshop on 3D Object Retrieval*, 2014.
- [5] P. Papadakis, “Terrain traversability analysis methods for unmanned ground vehicles: A survey,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1373–1385, 2013.
- [6] A. Sinha and P. Papadakis, “Mind the gap: detection and traversability analysis of terrain gaps using lidar for safe robot navigation,” *Robotica*, vol. 31, no. 7, p. 1085–1101, 2013.
- [7] G.-J. M. Kruijff, F. Pirri, M. Gianni, P. Papadakis, M. Pizzoli, A. Sinha, V. Tretyakov, T. Linder, E. Pianese, S. Corrao, F. Priori, S. Febrini, and S. Angeletti, “Rescue robots at earthquake-hit mirandola, italy: A field report,” in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2012.
- [8] P. Papadakis, A. Spalanzani, and C. Laugier, “Social mapping of human-populated environments by implicit function learning,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013.
- [9] P. Papadakis, P. Rives, and A. Spalanzani, “Adaptive spacing in human-robot interactions,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [10] P. Papadakis and P. Rives, “Binding human spatial interactions with mapping for enhanced mobility in dynamic environments,” *Autonomous Robots*, vol. 41, no. 5, pp. 1047–1059, 2017.
- [11] P. Papadakis and D. Filliat, “Generic object discrimination for mobile assistive robots using projective light diffusion,” in *IEEE Winter Applications of Computer Vision (WACV), Workshop on Computer Vision for Active and Assisted Living*, 2018.

- [12] P. Papadakis, “A use-case study on multi-view hypothesis fusion for 3d object classification,” in *IEEE International Conference on Computer Vision (ICCV), Workshop on Multi-View Relationships in 3D Data*, 2017.
- [13] A. Mitriakov, P. Papadakis, J. Kerdreux, and S. Garlatti, “Reinforcement learning based, staircase negotiation learning: Simulation and transfer to reality for articulated tracked robots,” *IEEE Robotics & Automation Magazine*, 2021.
- [14] A. Mitriakov, P. Papadakis, and S. Garlatti, “An open-source software framework for reinforcement learning-based control of tracked robots in simulated indoor environments,” *Advanced Robotics*, vol. 36, no. 11, pp. 519–532, 2022.
- [15] M. Devanne, P. Papadakis, and S. M. Nguyen, “Recognition of activities of daily living via hierarchical long-short term memory networks,” in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019.
- [16] R. Cazorla, L. Poinel, P. Papadakis, and C. Buche, “Bottleneck identification to semantic segmentation of industrial 3d point cloud scene via deep learning,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2021.
- [17] C. Le Bono, P. Papadakis, and C. Buche, “Assessment of conformal use of personal protective equipment by object and human pose recognition,” in *IEEE International Conference on Safety, Security and Rescue Robotics (SSRR)*, 2020.
- [18] M. K. Jabri, “Robot manipulation learning using generative adversarial imitation learning,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2021.
- [19] Q. Ferdinand, B. Clement, Q. Oliveau, G. Le Chenadec, and P. Papadakis, “Attenuating catastrophic forgetting by joint contrastive and incremental learning,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Continual Learning*, 2022.
- [20] D. A. Forsyth and J. Ponce, *Computer Vision - A Modern Approach, Second Edition*. Pitman, 2012.
- [21] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, “ShapeNet: An Information-Rich 3D Model Repository,” Tech. Rep. arXiv:1512.03012.
- [22] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgb-d images,” in *European Conference on Computer Vision (ECCV)*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., 2012.
- [23] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nusScenes: A multimodal dataset for autonomous driving,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

-
- [24] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer Verlag, 2003.
- [25] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. USA: Cambridge University Press, 2003.
- [26] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-d point sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, 1987.
- [27] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution,” *Neural Computation*, vol. 13, no. 7, 2001.
- [28] P. Papadakis, I. Pratikakis, T. Theoharis, and S. J. Perantonis, “PANORAMA: A 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval,” *International Journal of Computer Vision*, vol. 89, no. 2-3, pp. 177–192, 2010.
- [29] H. Durrant-Whyte and T. Bailey, “Simultaneous localization and mapping: part i,” *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [30] E. Garcia-Fidalgo and A. Ortiz, “Vision-based topological mapping and localization methods: A survey,” *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, 2015.
- [31] I. Kostavelis and A. Gasteratos, “Semantic mapping for mobile robotics tasks: A survey,” *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, 2015.
- [32] H. Yassine, P. Panagiotis, L. B. Cédric, F. Antoine, G. Tiago, and B. Cédric, “Densifying slam for uav navigation by fusion of monocular depth prediction,” in *IEEE International Conference on Automation, Robotics and Applications (ICARA)*, 2023.
- [33] R. Drouilly, P. Papadakis, P. Rives, and B. Morisset, “Local map extrapolation in dynamic environments,” in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2014.
- [34] A. Sinha, P. Papadakis, and M. R. Elara, “A staircase detection method for 3d point clouds,” in *International Conference on Control Automation Robotics Vision (ICARCV)*, 2014.
- [35] P. Papadakis, M. Gianni, F. Pirri, and M. Pizzoli, “Constraint-free topological mapping and path planning by maxima detection of the kernel spatial clearance density,” in *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 2012.
- [36] M. Firman, “RGBD Datasets: Past, Present and Future,” in *CVPR Workshop on Large Scale 3D Data: Acquisition, Modelling and Analysis*, 2016.
- [37] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, “Scannet: Richly-annotated 3d reconstructions of indoor scenes,” in *IEEE International Computer Vision and Pattern Recognition (CVPR)*, 2017.

- [38] Q.-H. Pham, M.-K. Tran, W. Li, S. Xiang, H. Zhou, W. Nie, A. Liu, Y. Su, M.-T. Tran, N.-M. Bui, T.-L. Do, T. V. Ninh, T.-K. Le, A.-V. Dao, V.-T. Nguyen, M. N. Do, A.-D. Duong, B.-S. Hua, L.-F. Yu, D. T. Nguyen, and S.-K. Yeung, “RGB-D Object-to-CAD Retrieval,” in *Eurographics Workshop on 3D Object Retrieval*, A. Telea, T. Theoharis, and R. Veltkamp, Eds., 2018.
- [39] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images,” in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [40] E. Eisemann and F. Durand, “Flash photography enhancement via intrinsic relighting,” in *ACM SIGGRAPH*, 2004.
- [41] R. Rusu, G. Bradski, R. Thibaux, and J. Hsu, “Fast 3d recognition and pose using the viewpoint feature histogram,” in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2010.
- [42] W. Wohlkinger and M. Vincze, “Ensemble of shape functions for 3d object classification,” in *International Conference on Robotics and Biomimetics*, 2011.
- [43] F. Tombari, S. Salti, and L. Stefano, “Unique signatures of histograms for local surface description,” in *European Conference on Computer Vision (ECCV)*, 2010.
- [44] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, “A comprehensive performance evaluation of 3d local feature descriptors,” *International Journal of Computer Vision*, vol. 116, no. 1, pp. 66–89, 2016.
- [45] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, “3d shapenets: A deep representation for volumetric shapes,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [46] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, “Sun database: Large-scale scene recognition from abbey to zoo,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3485–3492.
- [47] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, “Panoptic segmentation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [48] Y. Xie, J. Tian, and X. X. Zhu, “Linking points with labels in 3d: A review of point cloud semantic segmentation,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 4, 2020.
- [49] R. Cazorla, L. Poinel, P. Papadakis, and C. Buche, “Reducing domain shift in synthetic data augmentation for semantic segmentation of 3d point clouds,” in *IEEE International Conference on Systems, Man and Cybernetics*, 2022.

-
- [50] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, “3d semantic parsing of large-scale indoor spaces,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [51] M. McCloskey and N. J. Cohen, “Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem,” in *Psychology of Learning and Motivation*. Academic Press, 1989.
- [52] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, “Class-incremental learning: survey and performance evaluation,” *arXiv:2010.15277*, 2020.
- [53] A. Krizhevsky, G. Hinton *et al.*, “Learning multiple layers of features from tiny images,” 2009.
- [54] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [55] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [56] B. Zhao, X. Xiao, G. Gan, B. Zhang, and S.-T. Xia, “Maintaining discrimination and fairness in class incremental learning,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [57] H. Cha, J. Lee, and J. Shin, “Co2l: Contrastive continual learning,” in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [58] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng, “Ros: an open-source robot operating system,” in *ICRA Workshop on Open Source Software*, 2009.
- [59] V. Ntouskos, P. Papadakis, and F. Pirri, “Probabilistic discriminative dimensionality reduction for pose-based action recognition,” in *Springer - Advances in Intelligent Systems and Computing*, 2015.
- [60] V. Ntouskos., P. Papadakis., and F. Pirri., “A comprehensive analysis of human motion capture data for action recognition,” in *International Conference on Computer Vision Theory and Applications VISAPP*, 2012.
- [61] V. Ntouskos., P. Papadakis., and F. Pirri., “Discriminative sequence back-constrained gplvm for mocap based action recognition,” in *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 2013.
- [62] E. Hall, *The Hidden Dimension: Man’s Use of Space in Public and Private*. The Bodley Head Ltd, 1966.

BIBLIOGRAPHY

- [63] J. Rios Martinez, “Navigation de robot avec conscience sociale : entre l’évaluation des risques et celle des conventions sociales,” Ph.D. dissertation, Université de Grenoble, 2013. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-00957945>
- [64] A. Azzalini, “A class of distributions which includes the normal ones,” *Scandinavian Journal of Statistics*, vol. 12, no. 2, pp. 171–178, 1985.
- [65] A. Azzalini and A. D. Valle, “The multivariate skew-normal distribution,” *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.
- [66] L. Hayduk, “The shape of personal space: An experimental investigation,” *Journal of Behavioural Science*, vol. 13, pp. 87–93, 1981.
- [67] D. Helbing and P. Molnár, “Social force model for pedestrian dynamics,” *Physical Review E*, vol. 51, pp. 4282–4286, 1995.
- [68] M. Gérin-Lajoie, C. L. Richards, J. Fung, and B. J. McFadyen, “Characteristics of personal space during obstacle circumvention in physical and virtual environments,” *Gait and Posture*, vol. 27, no. 2, pp. 239 – 247, 2008.
- [69] H. Hoffmann, “Kernel pca for novelty detection,” *Pattern Recognition*, vol. 40, no. 3, pp. 863–874, 2007.
- [70] J.-C. Latombe, *Robot Motion Planning*. USA: Kluwer Academic Publishers, 1991.
- [71] (2003) Carnegie-mellon mocap database. [Http://mocap.cs.cmu.edu/](http://mocap.cs.cmu.edu/).
- [72] R. Poppe, “A survey on vision-based human action recognition,” *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.
- [73] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, “Vision-based human activity recognition: a survey,” *Multimedia Tools and Applications*, vol. 79, no. 41-42, pp. 30 509–30 555, 2020.
- [74] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.
- [75] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, “Documentation mocap database hdm05,” Universität Bonn, Tech. Rep. CG-2007-2, June 2007.
- [76] A. C. Damianou, M. K. Titsias, and N. D. Lawrence, “Variational gaussian process dynamical systems,” in *Neural Information Processing Systems*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, Eds., 2011, pp. 2510–2518.
- [77] V. Ntouskos, “Classification of human actions from motion capture sequences,” Master’s thesis, University of Rome “La Sapienza”, 2012.

-
- [78] J. M. Wiener, R. J. Hanley, R. Clark, and J. F. Van Nostrand, “Measuring the Activities of Daily Living: Comparisons Across National Surveys,” *Journal of Gerontology*, vol. 45, no. 6, pp. S229–S237, 1990.
- [79] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [80] C. Wu, J. Zhang, S. Savarese, and A. Saxena, “Watch-n-patch: Unsupervised understanding of actions and relations,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [81] S. G. Tzafestas, *Introduction to mobile robot control*. Elsevier, 2013.
- [82] S. M. LaValle, *Planning Algorithms*. Cambridge, U.K.: Cambridge University Press, 2006.
- [83] P. Papadakis and F. Pirri, “3D Mobility Learning and Regression of Articulated, Tracked Robotic Vehicles by Physics-based Optimization,” in *Eurographics Workshop on Virtual Reality Interaction and Physical Simulation*, 2012.
- [84] A. Mitriakov, P. Papadakis, S. Mai Nguyen, and S. Garlatti, “Staircase traversal via reinforcement learning for active reconfiguration of assistive robots,” in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020.
- [85] A. Mitriakov, P. Papadakis, S. Mai Nguyen, and S. Garlatti, “Staircase negotiation learning for articulated tracked robots with varying degrees of freedom,” in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2020.
- [86] P. Papadakis, I. Pratikakis, S. Perantonis, and T. Theoharis, “Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation,” *Pattern Recognition*, vol. 40, no. 9, pp. 2437–2452, 2007.
- [87] E. Papadopoulos and D. Rey, “A new measure of tipover stability margin for mobile manipulators,” in *IEEE International Conference on Robotics and Automation*, 1996.
- [88] A. Mitriakov, “Reinforcement learning-based control for safe 3D navigation of articulated tracked robot manipulators,” Ph.D. dissertation, Ecole nationale supérieure Mines-Télécom Atlantique, 2022. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-03575229>
- [89] R. B. Mcghee and G. I. Iswandhi, “Adaptive locomotion of a multilegged robot over rough terrain,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 4, pp. 176–182, 1979.
- [90] S. Hirose, H. Tsukagoshi, and K. Yoneda, “Normalized energy stability margin and its contour of walking vehicles on rough terrain,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2001.

BIBLIOGRAPHY

- [91] P. Papadakis, C. Lohr, M. Lujak, A. Karami, I. Kanellos, G. Lozenguez, and A. Fleury, “System design for coordinated multi-robot assistance deployment in smart spaces,” in *IEEE International Conference on Robotic Computing (IRC)*, 2018.
- [92] A. Saiad-Sahraoui, “Dynamic obstacle avoidance for indoor robot platooning,” Master’s thesis, Université de Rennes, 2018.
- [93] M. Gianni, P. Papadakis, F. Pirri, M. Liu, F. Pomerleau, F. Colas, K. Zimmermann, T. Svoboda, T. Petříček, G. M. Kruijff, H. Khambhaita, and H. Zender, “A unified framework for planning and execution-monitoring of mobile robots,” in *Automated Action Planning for Autonomous Mobile Robots*, ser. AAAI Technical Report, vol. WS-11-09. AAAI, 2011.
- [94] G. J. M. Kruijff, M. Janíček, S. Keshavdas, B. Larochelle, H. Zender, N. J. J. M. Smets, T. Mioch, M. A. Neerincx, J. V. Diggelen, F. Colas, M. Liu, F. Pomerleau, R. Siegwart, V. Hlaváč, T. Svoboda, T. Petříček, M. Reinstein, K. Zimmermann, F. Pirri, M. Gianni, P. Papadakis, A. Sinha, P. Balmer, N. Tomatis, R. Worst, T. Linder, H. Surmann, V. Tretyakov, S. Corrao, S. Pratzler-Wanczura, and M. Sulk, *Experience in System Design for Human-Robot Teaming in Urban Search and Rescue*, 2014.
- [95] C. Lohr, P. Tanguy, and J. Kerdreux, “xaal: A distributed infrastructure for heterogeneous ambient devices,” *Journal of Intelligent Systems*, vol. 24, no. 3, pp. 321–331, 2015.
- [96] A. Asvadi, A. Mitriakov, C. Lohr, and P. Papadakis, “Digital twin driven smart home: A feasibility study,” in *IEEE International Conference On Smart Living and Public Health*, 2022.
- [97] A. Asvadi, G. Bechu, A. Beugnard, C. G. L. Cao, C. Lohr, P. Papadakis, Q. Perez, C. Urtado, and S. Vauttier, “Model-driven deployment of Digital Twins for Smart Environments - The HUMAN at home project case study,” 2022. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03701748>
- [98] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, and P. Corke, “The limits and potentials of deep learning for robotics,” *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 405–420, 2018.
- [99] P. Corke, F. Dayoub, D. Hall, J. Skinner, and N. Sünderhauf, “What can robotics research learn from computer vision research?” in *International Symposium on Robotics Research (ISRR)*, ser. Springer Proceedings in Advanced Robotics, vol. 20, 2019.
- [100] S. Thrun and L. Pratt, Eds., *Learning to Learn*. USA: Kluwer Academic Publishers, 1998.
- [101] S. J. Isernhagen, “Functional capacity evaluation: Rationale, procedure, utility of the kinesiophysical approach,” *Journal of Occupational Rehabilitation*, vol. 2, 1992.

-
- [102] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [103] D. C. Guastella and G. Muscato, “Learning-based methods of perception and navigation for ground vehicles in unstructured environments: A review,” *Sensors*, vol. 21, no. 1, 2021.
- [104] B. Shen, F. Xia, C. Li, R. Martín-Martín, L. Fan, G. Wang, S. Buch, C. D’Arpino, S. Srivastava, L. P. Tchapti, K. Vainio, L. Fei-Fei, and S. Savarese, “igibson, a simulation environment for interactive tasks in large realistic scenes,” *arXiv:2012.02924*, 2020.
- [105] “AI2-THOR: An Interactive 3D Environment for Visual AI,” *arXiv:1712.05474*, 2017.
- [106] A. Szot, A. Clegg, E. Undersander, E. Wijmans, Y. Zhao, J. Turner, N. Maestre, M. Mukadam, D. Chaplot, O. Maksymets, A. Gokaslan, V. Vondrus, S. Dharur, F. Meier, W. Galuba, A. Chang, Z. Kira, V. Koltun, J. Malik, M. Savva, and D. Batra, “Habitat 2.0: Training home assistants to rearrange their habitat,” *arXiv:2106.14405*, 2021.
- [107] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *European Conference on Computer Vision (ECCV)*, 2020.
- [108] A. Moreau, N. Piasco, D. Tsishkou, B. Stanciulescu, and A. d. L. Fortelle, “Lens: Localization enhanced by nerf synthesis,” in *Conference on Robot Learning*, 2022.
- [109] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.
- [110] R. Szeliski, *Computer Vision - Algorithms and Applications, Second Edition*, ser. Texts in Computer Science. Springer, 2022.
- [111] R. Bajcsy, Y. Aloimonos, and J. K. Tsotsos, “Revisiting active perception,” *Autonomous Robots*, vol. 42, no. 2, p. 177–196, 2018.
- [112] S. Blisard, T. Carmichael, L. Ding, T. Finin, W. Frost, A. Graesser, M. Hadzikadic, L. Kagal, G.-J. M. Kruijff, P. Langley, J. Lester, D. L. McGuinness, J. Mostow, P. Papadakis, F. Pirri, R. Prasad, S. Stoyanchev, and P. Varakantham, “Reports of the aaai 2011 fall symposia,” *AI Magazine*, vol. 33, no. 1, pp. 71–78, 2012.
- [113] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, M. Burtscher, Q. Chen, N. K. Chowdhury, B. Fang, H. Fu, T. Furuya, H. Li, J. Liu, H. Johan, R. Kosaka, H. Koyanagi, R. Ohbuchi, A. Tatsuma, Y. Wan, C. Zhang, and C. Zou, “A comparison of 3d shape retrieval methods based on a large-scale benchmark supporting multimodal queries,” *Computer Vision and Image Understanding*, vol. 131, pp. 1–27, 2015.

BIBLIOGRAPHY

- [114] M. Machat, F. Langenfeld, D. Craciun, L. Sirugue, T. Labib, N. Lagarde, M. Maria, and M. Montes, “Comparative evaluation of shape retrieval methods on macromolecular surfaces: an application of computer vision methods in structural bioinformatics,” *Bioinformatics*, vol. 37, no. 23, 2021.
- [115] R. C. Veltkamp, R. Ruijsenaars, M. Spagnuolo, R. v. Zwol, and F. t. Haar, “Shrec2006: 3d shape retrieval contest,” Department of Information and Computing Sciences, Utrecht University, Tech. Rep. UU-CS-2006-030, 2006.

Title: Computer Vision and Robot Navigation in 3D Environments

Keywords: Object recognition, obstacle negotiation, unmanned ground vehicles, robot learning

Abstract: This dissertation presents contributions to research problems in the domains of computer vision and navigation in 3D environments for ground robots.

Specifically, Chapter 2 is dedicated to 3D vision and presents machine-learning based methodologies for the semantic understanding of geometric and color data, for applications pertaining to digital libraries of 3D objects and semantic mapping of static or dynamic environments populated by humans and explored by robots. We contribute more robust and effective 3D perception methods inspired from the fields of novelty detection, incremental learning and intelligent, multi-modal data fusion.

In the sequel, Chapter 3 addresses the problem of safe and effective navigation of

ground robots operating in 3D outdoor or indoor environments, focusing on the negotiation of negative and positive obstacles by leveraging the power of robot learning via physics-based simulation. We establish a general framework for modeling robot traversability in 3D terrains that is instantiated to the motion control of articulated, tracked robot manipulators. Diverse motion control behaviors are developed for the purpose of staircase ascent/descent, gap traversal upto socially-compliant navigation, leveraging on advances in the fields of reinforcement learning, end-to-end learning and potential fields.

Finally, Chapter 4 describes the deployment of terrestrial robots from a system engineering perspective in the context of in-field experiments and Chapter 5 unfolds future research directions.