Prediction of Object Position based on Probabilistic Qualitative Spatial Relations

von Malgorzata Goldhoorn

Dissertation

zur Erlangung des Grades eines Doktors der Ingenieurwissenschaften
- Dr.-Ing. -

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Mai 2017

Datum des Promotionskolloquiums: 26. Juli 2017

Gutachter: **Prof. Dr. Frank Kirchner (Universität Bremen)**
Prof. Dr. Joachim Hertzberg (Universität Osnabrück)

Abstract

Due to recent and extensive advancements in the robotic and artificial intelligence fields, intelligent systems can be found, with increasing frequency, in many areas of daily life. From industrial and surgical purposes to space robots, such complex systems are present. However, as demands for robotics systems increase, sophisticated algorithms for use in robotic areas such as perception, navigation, or manipulation are required. Although some algorithms for such purposes exist, there are still open questions and challenges that must be addressed.

Although robots are primarily used in the manufacturing industry, which has since been revolutionized by their precision and speed, there is a growing trend towards using service and personal robotics applications. The latter in particular must interact with humans naturally and effectively manage their environments, such as offices and homes. In contrast to the systems used in an industrial context, systems such as personal robots do not act in a predefined and fixed environment. Rather, these intelligent systems need an intrinsic comprehension of human environments to be able to support people in their daily life and manage common tasks such as preparing a breakfast table or cleaning a room. Crucially, these new robot systems require an entirely new level of capabilities to act in dynamic human environments.

This thesis addresses how qualitative spatial relations can be used to find an object's most probable location and thus guide the search for a sought object. Because current approaches focus mainly on crisp, two-dimensional relations, which are not directly suitable for use in three-dimensional real-world applications, a formalism for a new type of spatial relations is proposed in this work. This theoretical approach is then applied on real-world data to evaluate its applicability for robotics purposes. The resulting validation of the approach demonstrates that the developed method performs well and can be used to enhance search for objects.

Zusammenfassung

Mit dem wachsenden Fortschritt in den Bereichen der Robotik und Künstlicher Intelligenz steigt auch die Anzahl der Anwendungsgebiete für komplexe, autonome Systeme. Beginnend mit der Industrie und Chirurgie bis hin zur Weltraumforschung spielen komplexe System eine immer wichtigere Rolle. Um die Einsetzbarkeit solcher Systeme zu ermöglichen sind geeignete Algorithmen erforderlich, mit deren Hilfe Aufgaben in den Bereichen Wahrnehmung, Navigation oder beispielsweise Manipulation absolviert werden können. Obwohl die Anzahl solcher Algorithmen steigt, gibt es immer noch Herausforderungen in der Entwicklung solcher Systeme.

Seitdem robotische Systeme die industriellen Bereiche mit ihrer Präzision und Geschwindigkeit revolutioniert haben, zeichnet sich ein Trend ab, robotische System auch im Service- und Haushaltbereich einzusetzen. Solche Systeme müssen jedoch über einen erhöhten Fähigkeitsgrad verfügen. Im Gegensatz zu den industriellen Systemen, die in einer eher vordefinierten und geordneten Umgebung eingesetzt werden, müssen die in den personalen Bereichen eingesetzten Systeme ein tiefes Verständnis der menschlichen Umgebung aufweisen, um die geforderten Aufgaben absolvieren zu können. Die bisher in der Literatur existierenden Ansätze berücksichtigen jedoch oft nicht die Mehrdimensionalität und Komplexität der reellen Umgebung.

Diese Dissertation widmet sich der Frage, wie spatial-semantisches Wissen in Form von qualitativen probabilistischen spatialen Relationen benutzt werden kann, um eine Suche nach einem bestimmten Objekt in einer reellen Umgebung zu verbessern. Um dies zu realisieren, werden spatiale Potentialfelder verwendet, auf deren Basis die wahrscheinlichste Position eines gesuchten Objektes bestimmt werden kann. Zudem wird ein Formalismus definiert, der es ermöglicht, spatiales Wissen in einer qualitativen und probabilistischen Weise zu modellieren. Die Leistungsfähigkeit des entwickelten Ansatzes basiert auf evaluierten empirischen Daten, die in einer echten Umgebung aufgenommen wurden. Die aus den Experimenten resultierenden Ergebnisse haben gezeigt, dass die entwickelte Methode vielversprechende Resultate liefert, um in dem Bereich Objektsuche eingesetzt zu werden.

Danksagung

Die Fertigstellung dieser Arbeit wäre ohne Unterstützung vieler Personen, die mich in den letzten fünf Jahren begleitet und gefördert haben, nicht möglich gewesen. Diesen Personen möchte ich an dieser Stelle herzlich dafür danken.

Insbesondere möchte ich meinem Doktorvater Prof. Dr. Frank Kirchner für das in mich gesetzte Vertrauen und die Unterstützung bei der Durchführung der Doktorarbeit danken. Des Weiterem möchte ich Herrn Prof. Dr. Kirchner dafür danken, dass er es mir ermöglicht hat, in einem für mich spannenden Gebiet zu promovieren.

Mein weiterer Dank gilt Prof. Dr. Joachim Hertzberg, Prof. Dr. Ronny Hartanto und Prof. Dr. Diedrich Walter. Ihre fachliche Unterstützung und anregenden Diskussionen spielten eine große Rolle bei der Verfassung dieser Arbeit.

Da diese Arbeit durch das Graduiertenkolleg System Design „SyDe“ aus Mitteln des Zukunftskonzepts der Universität Bremen im Rahmen der Exzellenzinitiative des Bundes und der Länder unterstützt wurde, möchte ich an dieser Stelle Prof. Dr. Rolf Drechsler für die hervorragende Leitung des Kollegs danken.

Mein weiterer Dank gilt meinen Freunden und meiner Familie, die mich in der für mich so wichtigen Lebensphase stets zur Seite standen und unterstützt haben. Ganz besonders möchte ich meinem Mann Matthias Goldhoorn dafür danken, dass er mit mir durch die Höhen und Tiefen gegangen ist und mich stets bei meinen Entscheidungen unterstützt hat. Er war immer in den schwierigen Phasen für mich da. Außerdem möchte ich mich ganz besondes bei Wiebke Wenzel, Florian Keßeler und Christoph Hertzberg für die Anregungen in der Schreibphase bedanken. Ihr hattet immer ein offenes Ohr und wart mir auch eine echte Stütze. Danke dafür!

Des Weiteren möchte ich den Kollegen des DFKI, der Arbeitsgruppe Robotik der Universität Bremen und den Kollegen aus dem Graduiertenkolleg für die tollen Jahre der Zusammenarbeit danken. Ich werde die aufregende Zeit, die ich mit Euch verbracht habe, nie vergessen.

Contents

1 Introduction

This chapter describes the motivation and objectives of the present research. Moreover, the main challenges, previous work, scope, and outline of the thesis are presented.

1.1 Motivation

Nearly all real-world applications require that an intelligent system, such as an autonomous robot, has the capability to find or retrieve objects in the environment. It is also not uncommon that in such scenarios the robot does not have accurate information about object locations. Therefore, the system must, in a way, be able to predict the desired object's location to perform the task properly. Additionally, finding objects in a real-world setting is not a trivial task given the complexity of the human environment.

In many cases, the systems must search for an object that can be of several types and situated in many different locations in the environment. To find the sought after object, the system might search for it using a specified exploration strategy, or in the worst case scenario, perform an exhaustive search. Moreover, this search can also lead to an undesirable result in which the robot does not find the requested object according to the specified requirements. Furthermore, in most scenarios, exhaustive searches require capturing and analyzing a large amount of data, which makes the system unusable and not suitable for practical applications.

The systems require appropriate and precise representation of the environment to perform a given task properly. Such representations could include knowledge about spatial relations between objects, just as humans use such knowledge to structure their environment and place the entities involved in this environment. In this context, spatial relations act as types of abstractions of a configuration in the space of objects. In addition, because the systems should also act within human environments, they can make use of this sort of knowledge. Furthermore, as objects are not typically placed randomly in an environment, the system could learn statistical models over time about the configurations of the objects.

Therefore, the question is why the structure of a given environment, such as the spatial relations between objects are not taken into account to make the search more efficient. Instead of searching for a given object anywhere in the environment, the robot could make use of the knowledge that objects tend to co-occur with other objects and can be found in certain places within the environment with some regularity. Interestingly, there is evidence demonstrating that objects are not placed arbitrarily in the environment, but rather their placement and arrangement consist of certain ordering and patterns. Humans use this sort of *commonsense* knowledge to find and recognize objects in their daily life and are highly successful at doing so. In addition to using such knowledge, the system could also use *intermediate* objects first, which are easy to detect, and then use this evidence to find or recognize the sought after object.

This dissertation presents a novel approach for estimating an object's most probable position with consideration of spatial contexts within the environment and typical spatial relations between objects. These relations can then be used for object search and detection. A new representation form of the spatial semantic knowledge is also developed. This spatial information is formalized and used to formulate hypotheses about possible object positions.

1.2 Previous Work

Over the past decade, object search and recognition has become an increasingly active area of interest in robotics. In the following section, an overview of previous work relevant to the present research is provided.

Spatial Relations as the common representation form of the spatial context

According to Freeman [Fre75], spatial relation specifies how an object is located in space relative to other object. For humans, spatial relations are of great importance for their reasoning and acting capabilities underlying mental representations. In most of the cases, a robot is tasked with finding objects in a previously unknown environment. However, by understanding and exploiting the structure of the environment, the robot can perform tasks more effectively. Additionally, objects used by humans are not placed randomly in an environment. Rather, objects are placed with regards to their role and function, and are structurally organized. Therefore, the robot can use such contextual knowledge from past observations to guide a search for a sought after object. Importantly though, the system needs a representation of the environment and the spatial structures. Apart from using the context information, e.g. which object tends to co-occur with another object, the system could also take advantage of the spatial configuration of objects in the scene, in particular it could exploit the spatial relations between objects.

In the earlier works, spatial relations were used primarily in vision-based systems to provide contextual information. Notably, Divvala et al. [DHH+09] developed an approach in which spatial context configurations were exploited to obtain a better semantic segmentation and annotate the objects in the image. Spatial relations were also used in activity recognition in, for example, [DCH10], where they learned activity phases from videos taken at an airport using spatial relations between objects that were tracked. More recently, the spatial relations have also been used in a three-dimensional (3D) manner, as the 3D representation becomes more common. One of these works is presented by Koppula et al. [KAJS11]. In this research, the authors used geometric information to classify objects in an office and home environment. Likewise, Fisher et al. [FSH11] used contextual information to perform scene similarity and object classification. Later, in [FRS+12], spatial knowledge was used for context-based object search. In Kasper et al. [KJD11], spatial relations between objects are used to develop an empirical base for scene understanding. Then, Ruiz-del Solar et al. [RdSLS13] introduced so-called *masks* based on spatial relations with distance thresholds such as *very far* and *very near* to perform searches for a sought after object. Southey and Litthle [SL07] developed an approach for object recognition based on spatial relations and object types in the scene. They computed distributions of spatial relations between objects using a set of different scenes. Most relevant to the

current research is the approach developed by Kunze et al. [KBA+14], in which spatial relations play a certain role. In this work, the qualitative spatial relations between objects are used for indirect search. The relations are modeled by using the Gaussian Mixture Models (GMM), and these relations are then used to predict the locations of objects. The drawback of the approach is that only two objects are take into account, e.g. a monitor and keyboard, for learning the spatial relations.

Spatial relations in object search processes

Earlier work investigating object search mainly used computer vision, in which the objects should be located in a given image. The problem here is that the search is often limited to finding the objects in the particular part of the image. This approach is subject to an assumption that the target is already in the given field of view. However, in a realistic robotic scenario, this is seldom the case. Rather, the object is located somewhere in the environment and its position must first be estimated for it to be found. A branch of work specifically addresses the problem of object search, and the most relevant findings for the approach developed in this thesis are presented in the following section.

Garvey [Gar76] recognized in 1976 that search for objects in a real environment is a particularly challenging task. However, the vision system he developed was able to find objects using expectations about semantic structures of a given scene. This method introduced by Garvey was termed *indirect search*, and based on the principle of first searching for an *intermediate* object connected to the sought after object via a particular spatial relation to find the sought after object. Garvey illustrated the idea by providing an example where a system looking for a phone first searches for a table on which the object (phone) can be located. In this way, the search is limited to a table top, which reduces the problem. Wixsone et al. [WB94] revived this idea for indirect search and demonstrated that by using this method, greater efficiency can be achieved. He demonstrated this idea theoretically and empirically by compering the indirect search strategy with a common search. Also, other approaches, such as that presented by Reece [Ree92], confirm this idea and show that by using indirect search, useful results can be returned.

Nevertheless, exhaustive search for an arbitrary object, even those that serve as intermediate ones, is still a notably complex and intensive process. Tsotsos [Tso92] evaluated the complexity of searching for an object in 3D space and found it to be NP-complete. To make the search practical in this context, a sort of heuristic is needed to guide the search and make it feasible. One could take humans as an inspiration. Specifically, humans make use of prior knowledge when searching for objects in an environment [BMR82]. Based on the common knowledge that objects almost always co-occur with another objects, and taking into account the objects' relations, the search space could be reduced. Shubina and Tsotsos [ST10] proposed an approach that considers the different spatial relations between objects. The authors used prior knowledge about a potential target's locations encoded as a probability distribution to optimize the search process. These so-called, *hints* contain knowledge about an object's relations, and leads to a higher search probability in specific regions. However, a drawback of this method is that the system knows that the sought after object is placed somewhere on the table and the table's position is known in advanced. Therefore, the object search problem becomes merely a statistical optimization problem.

Some studies have already proven the effectiveness of spatial relations and object location probability distributions. One recent and most frequently quoted work stems from Aydemir et al. [ASJ10], [APS+11], [SAJ12]. These authors have developed a method for object search using explicit qualitative spatial relation to perform visual search more efficiently using a computational model of the spatial relation, *on*, to obtain a probability distribution and guide the robot's camera to the points where probability is high of finding the object [ASJ10]. In further work [APS+11], an approach for object search is presented, in which metric and semantic maps are used to find the place that most probably contain the object. To locate the object, predefined priori knowledge about typical object placement and two spatial relations, *on* and *in*, are used. Later [AGP+11], the authors presented a planning approach for active visual object search. Using this method, the system searches for objects in an office environment by using high-level conceptual and semantic information. Also, in [AGP+11], the authors proposed plan-based object search in which spatial knowledge, such as object co-occurrences and place categorization, are utilized to perform an indirect search. For this approach, spatial relations *on* and *in* are exploited and a switching planner is used. These relations are also applied to perform large-scale visual search for objects, as described in [ASF+11]. For this purpose, a computational model of these relations was proposed and a function for calculating the probability density functions was developed. In more recent works [APJ12] and [AJ12], an active visual search was performed in a large-scale manner and in previously unknown environments. In [APJ12], a method is proposed for active visual search using semantics of the environment to optimize the search strategy. The semantic information is based on unspecific priors and includes knowledge about typical locations for a given object, whereas in [AJ12], a 3D context is used for predicting object locations. This information contains the local 3D shape around an object as an indicator for its placement. In contrast to their previous works where flat horizontal surfaces were considered, the author defined a more general spatial model.

While Aydemir et al. [APS+11] partly defined the objects relations manually, Kollar and Roy [KR09] have utilized object-object and object-scene relations, and extracted these relations from the photos on the website Flickr. In this work, the authors address the challenges of object search in a human environment by utilizing the histogram of the object co-occurrence. The idea of this approach is that objects are located in certain scenes given their type and scene structure and the system can use this context to predict the location of objects. The map and the location from which the search can be started are known as a priori. The application for searching for an object in a large-scale environment presented by Kunze et al. [KBS+12] is based on common-sense knowledge, which is then used to calculate probabilities in a room containing the target object. Bedsides, the a priori knowledge the robot has also includes a floor plan of the building. Although this method seems essential in a large-scale environment, the robot sill needs to perform a rigorous search inside the room. In their further work [KDH14], qualitative spatial relations such as *left-of*, *right-of*, *behind-of* and *in-front-of* represented using Gaussian normal distribution are used to perform indirect search. The find a potential object's position, the multivariate Gaussian distribution is calculated. Additionally, the authors used so-called *static* objects as landmarks to support the search for *dynamic* objects.

In another work, Sjöö et al. [SAJ12] present a method in which topological spatial relations, *on* and *in*, are used to guide visual search and find objects in the scene. The

weakness of this approach is the number of spatial relations required for the search. Elfring et al. [EJvdMS14] proposed an active object search method that uses object co-occurrences and considers the cost to find the intermediate objects. However, they not take into account the spatial relations between the main and intermediate objects. Alternatively, Ekvall et al. [EKJ07] present a method that partitions images in regions to calculate different probabilities for the occurrence of the searched object within them. The actual localization is eventually conducted by searching in the most probable regions.

Spatial relations in object recognition processes

Due to the increasing availability of low-cost RGB-D sensors such as the Microsoft Kinect or Primesense cameras that provide a 3D map of the environment, spatial information about objects relations in 3D space can be extracted more easily. Such information must be represented in a precise way and can be also used to improve the object detection process. Objects play an important role when building a semantic representation and understanding of the function of space [VS08]. A mobile system operating in a human living environment must be able to perform a complex task and, in turn, cope with objects of various types. To perform tasks such as fetching and carrying, the system needs to be capable of finding a recognizing a given object properly.

Common object detection systems rely entirely on object features that describe the certain object class [QT09], [FGMR10]. However, the features are extracted from the noisy and error prone data provided by the sensors. Further issues include the object's classifier being highly dependent on object pose, texture, camera view point, and illumination. Given the existing evidence that objects are not placed randomly in the environment, and that one can make assumptions about scene arrangement and object co-occurrences, object recognition can be improved. By exploiting contextual spatial knowledge, the system could increase detection accuracy. Aydemir et al. [APJ12] have presented a contextually guided semantic labeling and search algorithm, and in this method, a graphical model with geometrical features and contextual relations between objects is used. The model is trained using a maximum-margin learning approach. The authors use merged point clouds from indoor environments obtained with a Kinect RGB-D sensor. In addition, to acquire a better view of occluded objects in the scene, an active object recognition is performed. Similar to this work in [AKJS12], the next best view algorithm is applied to handle occlusion. Other recent work addressing the challenges of object labeling in indoor environments using RGB-D data is presented by Ren et al. [XLF12]. The authors have developed and evaluated a method for scene labeling that combines RGB-D features and contextual models by using Markov Random Fields (MRF) and segmentation. To gain RGB-D features such as gradient, color, and a surface normal, so-called *kernel* descriptors are used. Also, Ali et al. [ASG$^+$14] has proposed the use of a context model to improve the detection results of other related objects in the scene. In this approach, object co-occurrence information is used to perform recognition of the object category. In the work presented by Xiong et al. [XH10], planar patches extracted from a point cloud are labeled with geometric labels such as wall, floor, and ceiling. To model geometrical relationships like orthogonal, parallel, adjacent, and coplanar between objects, Conditional Random Fields (CRF) are used.

1.3 Problem Statement

To manage complex tasks in challenging scenarios, a system must be equipped with strong reasoning capabilities. Typically, objects in human environments are placed in locations with certain likelihoods. Also, there are correlations between object occurrences. For these reasons, the system must be able to use such knowledge to formulate hypotheses about possible object locations and then prioritize the search process while performing a task. However, representing such knowledge in a robot-understandable way and exploiting it during the task are two of the main challenges in object search. The primary problem addressed in the present work is how to use and model such knowledge so the robot can predict possible object locations.

Another existing challenge is unknown object positions. In some cases, the robot does not see the environment in which it should operate prior to working in it. Interestingly, there is evidence demonstrating that humans are highly successful in searching for objects in previously unknown environments or in instances when the locations of the objects are not known in advance. A likely explanation for this ability is that humans can use experiences to reason about object locations, e.g. "plates are often seen on the kitchen table or in the drawer". Furthermore, in a desk scene, for example, people typically expect to see a monitor, keyboard, and mouse arranged in a certain configuration. Therefore, to find and recognize the sought after objects, the system could use spatial context information. In this work, an approach for the representation and learning of this relevant spatial context information in an appropriate way is developed.

In real-world scenes, such knowledge could also be gained regarding the context information about typical object co-occurrences. That is, knowing which object typically occurs with other objects enables the robot to make assumptions about potential object classes. In human psychology, this process is known as *context reasoning*. Thus, using spatial object relations as the context, the search space can be reduced and the search process optimized. This process of using spatial relations is crucial because searching all locations in the environment is highly inefficient.

The third problem addressed in the present work is how to model spatial context information in a precise way. If the robots receive commands such as "bring me the bottle of milk near the fridge", the system must know how to map this semantic statement in the spatial environment. Specifically, the system must know exactly where *near* is referencing. Additionally, the qualitative spatial relation *near* gains an entirely different meaning according to the scale of the environment.

To perform a search using qualitative spatial relations, a robot requires a rich and precise representation of the 3D relations. These relations can be either manually specified based on scene geometry using Qualitative Spatial Reasoning (QSR) calculus [CH01], or learned from observations [SAJ12]. However, existing calculi are not directly suitable to capture variations of object positions with regards to the category and relations. That said, the Region Connection Calculus (RCC) and its variants provide a method for the modeling of qualitative spatial relations between regions. The main problem here is that the relations are strict and represent geometrical, not functional, attributes. Furthermore, because of their geometric nature, which refers to two-dimensional (2D) space, these relations are not applicable for practical real-world scenarios.

The combination of probabilistic methods with qualitative spatial knowledge is still a

highly challenging area of Artificial Intelligence (AI). Moreover, extracting and using such knowledge for robotics purposes is considerably more difficult because of the complexity of real-word human environments. Although some methods for the representation of spatial knowledge are currently available, none of the approaches make use of probabilistic methodology. Even if probability is used in some approaches, such as in [KBH14], its use is negligible compared to the present work. Furthermore, when probability is used, if at all, it is used only for gaining the distribution of so-called landmark objects or to approximate the location of a certain landmark. Critically, previous works do not consider the probabilistic representations of spatial relations intrinsically.

Human-Robot-Interaction

In a typical scenario, domestic service robots would receive command such as "give me the cup from the table" or "bring me the cereal box located near the fridge". Preferably, the human operator provides the command as a spoken statement since this is the natural way people interact with each other. Furthermore, there is evidence suggesting that humans prefer to use qualitative instead of quantitative descriptions to describe entities in their environment [Zim93]. Evidence for these phenomena states that it is more difficult for people to make correct metric statements. For example, instead of saying "the goal is in 50 meters", it is more common to say "the goal is near". Moreover, as revealed from studies in Human Robot Interaction (HRI) and regarding work by Moratz [MTBF03], the acceptability of such systems is strongly correlated with the way the systems are used:

> Non-intuitive styles of interaction between humans and mobile robots still constitute a major barrier to the wider application and acceptance of mobile robot technology. More natural interaction can only be achieved if ways are found of bridging the gap between the forms of spatial knowledge maintained by such robots and the forms of language used by humans to communicate such knowledge.
>
> Moratz, 2003

The second aspect of this statement refers to human robot interaction. Because people naturally use qualitative expressions to describe their actions and everyday items, it would be preferable that a supporting system could interact with humans in a natural way. However, such requirements imply that the system has the capability to map such expressions in the environment in which it acts. In a scenario where a human operator asks the robot to bring an object to them as expressed by linguistic phrases such as "bring me the bottle of milk that is located near the fridge", the system must be able to map such linguistic descriptions onto locations in the environment. Firstly, the system must know which part of the spatial area the qualitative term *near* refers to. This qualitative description is relative because one could say that near depends on the given scenario and the objects involved. Such reasoning means that "near" refers not only to one point in space but also to a whole range of points. The challenge then is to reason which point within this range is relevant for the particular case and scenario. Specifically, the intensification of the relation is relevant. The main challenge here is to map the qualitative description in the quantitative model of the environment. Therefore, one of the questions that must be addressed is how to model such spatial relations precisely. In the present work, this chal-

lenge is approached by extending qualitative spatial relations to probabilistic methodology to model the relations more accurately.

Precise Representation of QSR in a 3D space

In addition to the human interaction aspect, robots can only perform a given task if they are able to decode the semantic meaning of the objects specified in the task and reason where to search for them. Furthermore, the more difficult and complex the tasks are, the more capable the system must be. Many issues must be addressed before such autonomous intelligent system can be an inherent part of our daily lives and one of these issues is that robots must act in an environment that is highly dynamic and designed for humans. In this context, such a system must have the capabilities to exploit the input received by their sensors and amalgamate these data with previously gathered knowledge to reason about possible solutions to the given task. For instance, the system has to extract the objects from the data and ascribe them with appropriate semantic meanings.

Furthermore, to retrieve a given object, the robot must not only be able to detect the given object, but must also be able to reason where to search for it. Without any reasoning capabilities and knowledge about typical object locations, such searches can easily lead to tedious searches, and in the worst case, end with an unfinished task.

Therefore, to limit the search for a given object, an appropriate mechanism that enables the system to make a prediction related to the object's location is needed. Such a mechanism could, for instance, utilize previously gathered knowledge to guide the search. In a human environment, there are regularities regarding the spatial arrangement of objects and strong correlations between object co-occurrences. In this work, these regularities are extracted from the environment and modeled in the form of Probabilistic Qualitative Spatial Relations (PQSR). By using this probabilistic description, the qualitative spatial relations receive a rich representation which, in turn, can be used by the robot to perform reasoning about possible object locations and classes. More precisely, these relations serve as *commonsense* general knowledge for the robots and can be used to find objects in the environment and then ascribe them correct semantic descriptions. Due to the qualitative aspect of this method, more natural HRI can be achieved.

1.4 Scope and Objectives

The following section describes the focus of this thesis and its main objectives.

Scope of this thesis

The present work focuses on the formalization and modeling of PQSR to then use such relations for different robotics purposes. The PQSR specifies the 3D spatial relations between objects in a qualitative and probabilistic manner. In addition, a formalism and an approach for learning qualitative relations from real-world quantitative data is developed. Based on the PQSR, an estimation of the most probable position of the object is obtained and this spatial context information serves as a heuristic assumption to locate a sought after object in an environment. However, only an exemplary search method is presented because the development of an entire search process is beyond the objectives of this thesis.

Instead, the aim of this thesis is development of a clear formalism of the PQSR, which can then be used for different robotics purposes besides object search.

Objectives of this thesis

The objective of this thesis is to contribute to the scientific advancement of intelligent robotics systems by developing a rich 3D spatial representation by which spatial relations between objects are modeled in a qualitative and probabilistic manner. Using these probabilistic and qualitative models, a robotic system should be able to perform an object search more efficiently. Moreover, by providing a formalism, this spatial representation is also applicable in other robotics domains. In this context, the major contributions of this work are:

- ***Modeling qualitative spatial relations using probabilistic methodology***
  One of the main contributions of this dissertation is the modeling and formalization of a new type of qualitative spatial relations, so-called PQSR. These relations are defined in a probabilistic manner to model spatial representations more precisely and represent these relations more appropriately for robotics applications.

  The primary motivator behind this development is the current need for a richer spatial representation of relations between objects that can be used for different robotics purposes. Although common spatial representations such as the RCC and its variants provide a language for expressing qualitative relationships between regions, these approaches are not suitable for real-world applications. As mentioned previously, these methods are of an "all-or-nothing nature" and rely only on pure geometric information, typically in two dimensions, which makes them unusable for realistic robot scenarios. Furthermore, the models assume that the objects are positioned axis-parallel to each other, which is unusual in a real 3D environment. Consequently, these restrictions render the models unsuitable for robotics applications.

  In contrast, this thesis presents spatial relations that are modeled with regard to a real robotics scenario. The spatial relations presented here constitute a powerful model to not only represent the probabilistic spatial relations in general, but also to capture the multiple relations that objects can have with their environment and other objects. Due to this probabilistic aspect, it is possible to define exactly how probable a spatial relation holds between two types of objects in general, and additionally, at which position in a scene and a certain relation holds the most considering the objects within the scene.

- ***Designing and implementing new type of spatial representation using Spatial Potential Fields***
  In this work, probabilistic spatial relations are modeled using so-called Spatial Potential Fields (SPF). These fields are used to map a given probabilistic qualitative spatial relation to the environment. Using these fields, spatial relations can be modeled in a probabilistic manner in the environment. More precisely, SPF enable the mapping of relations with high accuracy by defining the area in which the relation

is valid. Using this probability, the strength of the relation's validity can be calculated. For instance, if the system should search for an object located near the fridge, the robot needs to know where the "near" relation holds. Otherwise, the system would not know where it should focus to locate the sought after object. In cases when the robot has already focused on the space in which the object is assumed to be located, it might not be too difficult to locate the object. However, if a part of the scene must be determined first, a precise model of the spatial representation is required. Ultimately, such spatial knowledge enables the robot to find objects even in an environment it has not viewed previously. Depending on the scene, a given relation such as *near* can range from *very-near* to *very-far*. Such variation in the relation-related space must be defined and importantly, some regularities regarding the spatial relations are present and can be exploited. In this work, these regularities are learned from real-world data based on the PQSR's formalism. With this approach, a statistical model of the common spatial relations between objects can be developed.

- ***Designing algorithm to use PQSR for object position prediction based on Field Intensity and Maximum Field Intensity***
  For humans, general knowledge such as the common relations between objects and object co-occurrences are useful not only for object search but also for object recognition. For robots, making use of such knowledge can lead to predictions about possible object locations or object class. However, in a hypothetical environment, an object can be found at any location with the same probability, which makes the search unsustainable. Such an assumption is also crucial for real-world applications and can result in exhaustive searches and in the worst case, an unaccomplished task. Therefore, a system requires heuristics to guide and prioritize the search. In this work, a method is developed that uses SPF to inform the search strategy. Then, several SPF are combined with an Field Intensity (FI) using this method. The FI provides information about possible object position given its relations to other objects. By using FI, a prediction of the object's location is calculated even if a system acts in an environment it has not seen before. Consequently, the most probable object location can be obtained based on the FI.

- ***Validation of the approaches under consideration using different data sources***
  The fourth objective of this thesis is the validation of the theoretical and algorithmic concept of the PQSR by conducting experiments that consider several aspects of the developed approach. To determine the applicability of the developed method for robotics purposes, many experiments were preformed on different data sets. Since the method was developed with a focus on its applicability for robotics scenarios, the data sets were gathered using real sensors such as those provided by the Microsoft Kinect camera. The experiments conducted investigate different aspects of the approach such as precision of the PQSR-based object position prediction and the method's reliability in real-world scenes. Further experiments refer to the PQSR formalism, such as the experiments for learning of the PQSR from the real-world data.

1.5 Outline

The structure of this thesis is provided in Figure 1.1. In the first chapter 1, the motivation and objectives of the thesis are presented. Additionally, this chapter describes previous work and the current challenges in robotics addressed by this thesis. The second chapter 2 introduces spatial relations theory and presents the formalism developed in the present work. The algorithm that uses the PQSR for object position estimation is also detailed. Chapter 3 describes the experiments by which the developed method is evaluated. The experiments and their results are accompanied by detailed discussion and analysis. Finally, chapter 4 provides the conclusion of the developed approach. This chapter also presents a detailed overview of possible future work and alterations that could be implemented to improve the developed approach.

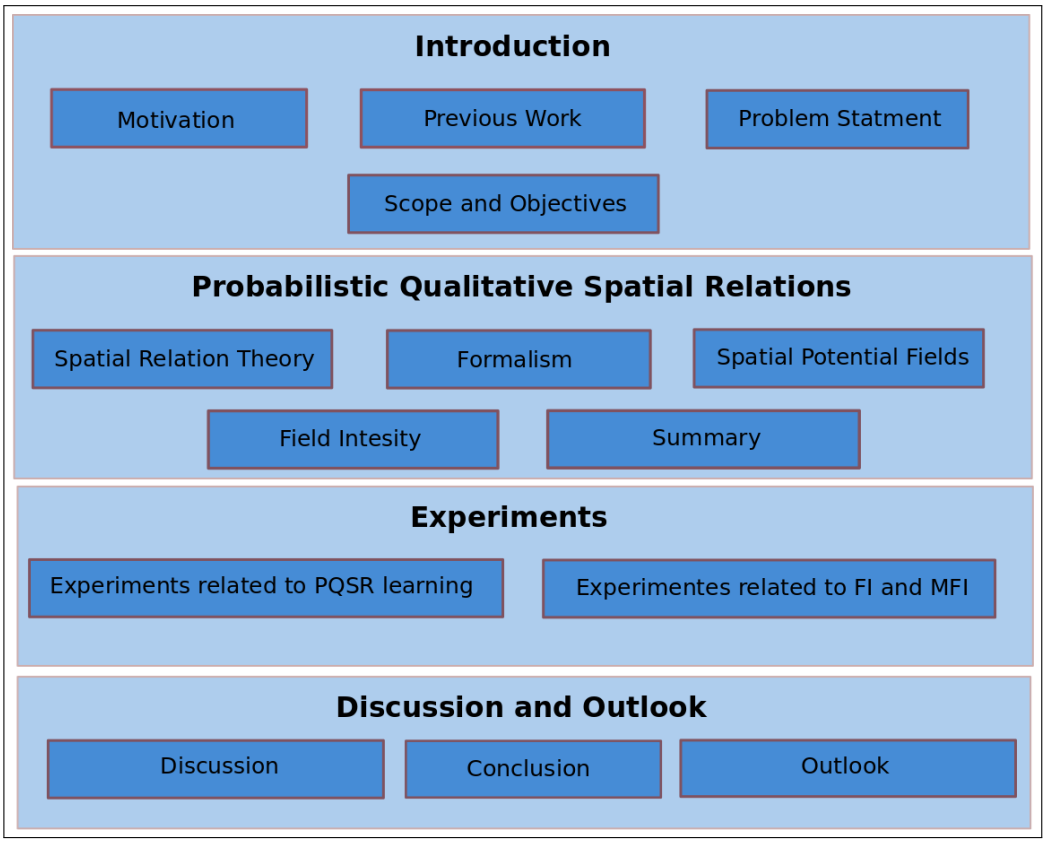


Figure 1.1: The structure of the thesis.

2 Probabilistic Qualitative Spatial Relations

Physical space and its properties are important for human decision making and actions in everyday life [Fre92]. Furthermore, people do not need exact information such as metric descriptions of their environment to manage common tasks. Their reasoning is also of a symbolic and qualitative nature [Fra92]. Because human are successful at these actions, it is desirable to make use of such representations in an artificial system, which also needs suitable representations of space. Additionally, a system that should support people must interact with the person in a human and natural way.

However, the current challenge is to equip the robot with reasoning capabilities, which can enable it to use qualitative statements and map them onto its metric world model [MTBF03]. In other words, gathering the qualitative relations from the metric data. Qualitative spatial information is already used in many areas within artificial intelligence and robotics, from navigation [WH05], planning [LD04], [HTD90] to natural language [Her86] and vision [SS03]. The most common spatial representations, such as the RCC and its variations, are often based on geometric forms and used to perform reasoning at a symbolic level, but foremost, they are typically 2D and crisp. Because a robot should perform tasks in 3D space, i.e. the human living environment, a suitable and adequate representation of such space is required.

In this work, a probabilistic spatial representation form of common relations between objects is developed and proposed. This representation is modeled in a probabilistic and qualitative manner to gain a reach formalism that can be used for robotics purposes, such as object search. These qualitative spatial relations are calculated from the quantitative data obtained from 3D real-world data sets. The major difference between the existing common spatial representations and the one developed in this thesis is that the latter is modeled using probabilistic methodology. This approach enables definition of the spatial relations between objects more precisely and uses these relations in the real-world scenarios. Due to this extension, the spatial relations between objects can be defined more accurately by describing how *likely* a given relation is. Furthermore, by implementing the probabilistic extension of the spatial representation, crisp relations are transferred into a more flexible model. This approach makes the representation suitable for real-world applications, because they are not crisp but rather uncertain and variable. One such application of this model could be a system tasked with finding an object and retrieving it.

In physical space, objects are connected by spatial relations, or more precisely, a spatial relation holds between objects. Generally, objects are complex entities with various properties and several types of relations can be observed in the human environment. Referring to Freeman's observations [Fre75], the most commonly used relations in a semantic content are as follows: *left-of*, *right-of*, *in-front-of*, *behind-of*, *on*, *above* and *near*. Many object arrangements in a scene can be described as a combination of these relations. For example, a statement: "bring me the remote control, which is on the table and right of the TV"

consists of the combination of the two spatial relations *on* and *right-of*. Considering the spatial relations in a 2D context only, some cannot be represented, or rather, their meaning depends on the perspective from which the relations are observed. However, many approaches such as those from the qualitative representation and reasoning field, focus on 2D space. Since relations such as *in-front-of* and *behind-of* require some 3D space to be represented, they can not directly be modeled using 2D-based methods. Considering that the human environment, in which a robot should perform tasks, is 3D, it is desirable to also model spatial relations in a 3D manner.

Furthermore, qualitative spatial relations can be interpreted differently depending on the context and spatial space to which they refer because one can not say exactly to which range a given qualitative relation is related to without limiting or specifying its corresponding area. For instance, when considering two buildings located close to each other or a mouse and keyboard, the spatial relation *near* is related to vastly different distances in both scenarios. That is, the spatial space between two buildings is typically much wider than between smaller objects like office supplies. Hence, a qualitative spatial relation is not related to one point in the environment but refers to an area in space.

In this dissertation, a 3D probabilistic model of the common qualitative spatial relations is formalized. Furthermore, an approach for using such relations in an object search process is proposed. In this work, the spatial relations are modeled from quantitative data and used as a meaningful spatial description of the relationships between objects. The probabilistic nature of the spatial relations enables definition of the corresponding area in which a given relation is valid and this approach is conducted in a qualitative manner. The qualitative relations are represented in a real metric environment that consists of quantitative aspects. This representation, in turn, can be used for real-world robotics applications in which the system must handle numbers whilst interacting with humans for humans in a natural way.

This chapter outlined the formalism for modeling and learning of the probabilistic qualitative spatial relations. Furthermore, a new method for using those relations in robotics applications was presented. In Section 2.1, a comparison between the spatial representation developed in this work and existing approaches is presented. Section 2.2 contains formal definitions of the PQSR and further formalisms referring to these relations. The learning method of the co-occurrence probabilities and typical distances between objects is presented in Section 2.2.3.1. This information is followed by Sections 2.3 and 2.4. In these sections, a new method for applying the PQSR in real application by so-called SPF is presented. The Section 2.4 provides a formal definition of the so-called FI, which can be used to determine the object's most probable position.

2.1 Spatial Relations Theory

Spatial relations describe the co-relations between spatial entities. These relations are often related to certain activities or functional objectives, as humans design and organize space in ways that serve various different purposes. These relations also refer to the placement of objects in the environment, that is, objects are not placed randomly but rather in the context of certain activities and application. However, it is not only object's placement that is not random but also the choice of spatial relation type for describing objects arrangement.

As mentioned previously, there are several spatial relations that serve as a basis for humans to describe a given configuration of space [Fre75]. Such relations are *left-of*, *right-of*, *in-front-of*, *on*, *near* and *behind-of*. Using these main relations and various other relations, which can hold between objects, provide descriptions, such as "*on* the table and to the *right-of* of the monitor". A common aspect of these spatial relations is that they are of a qualitative nature, or more precisely, without quantitative terms.

In Section 1.2, some previous work related to spatial relations in the context of robot search and recognition was presented. In this section, the most relevant works are selected and discussed in more detail. The works were selected based on their relevance to the approach developed in this thesis.

2.1.1 Related Work

In the existing literature, the is a large number of works related to qualitative spatial relations. Some of these address spatial relations in the context of psycholinguistics and cognitive science [Her87], [CG04], [Lev96]. For instance, in work by [O'K99], the author attempts to define how spatial relations can be qualified based on spatial information stored in the hippocampus. O'Keefe proposes several factors that can be used to define spatial relations, such as *near* or *behind-of*. Regier and Carlson [RC01] have investigated how accurate particular spatial relations are for scene description using so-called *Attention Vector Sum*. In the work by [LFHU06], a system is developed that learns to make distinctions between spatial relations such as *in*, *on*, *above*, and *left*.

Qualitative spatial relations are also extensively studied in the field of Qualitative Spatial Reasoning [CBGG97]. In this research area, spatial reasoning is performed to handle common-sense knowledge without using quantitative models. The common-sense knowledge is instead represented as qualitative categories such as qualitative spatial relations. This approach is one of the principles of qualitative approaches that does not use a vague or probabilistic method. The main point of qualitative reasoning is to perform on a symbolic level. That is, qualitative reasoning is intended to represent continuous attributes using discrete symbols.

In 1992, Randell, Cui, and Cohn [RCC92] introduced RCC (known as the RCC-8 calculus) as a method of qualitative spatial calculus for reasoning about relations between regions. The relations are distinguished based on their connectedness and mereological properties. There are eight base relations, as follows: disconnected (DC), externally connected (EC), partially overlap (PO), tangential proper and its inverse (TPP, TPP^{-1}), non-tangential proper part and its inverse (NTPP, $NTPP^{-1}$), and equivalent (EQ). All relations are jointly exhaustive and pairwise disjoint (JEPG). These properties mean that exactly one relation between given spatial entities holds. Figure 2.1 illustrates the RCC-8.

The extension of the RCC calculus, so-called "egg-yolk" calculus [LC94], is used to model uncertainty and vague relationships. For this, non crisp regions are defined additionally to the crisp regions. By using this calculus, regions with indeterminate boundaries can be modeled as a pair of regions. Notably, these spatial regions can have uncertain boundaries. The egg is the maximal extent of the vague region, whilst the yolk is its minimal extent. To be more precise, the certain region is located inside the uncertain region. The crisping space denotes the area of indeterminacy. Figure 2.2 illustrates the egg-yolk structure.

Space and its properties is also an important aspect of developing approaches in artificial

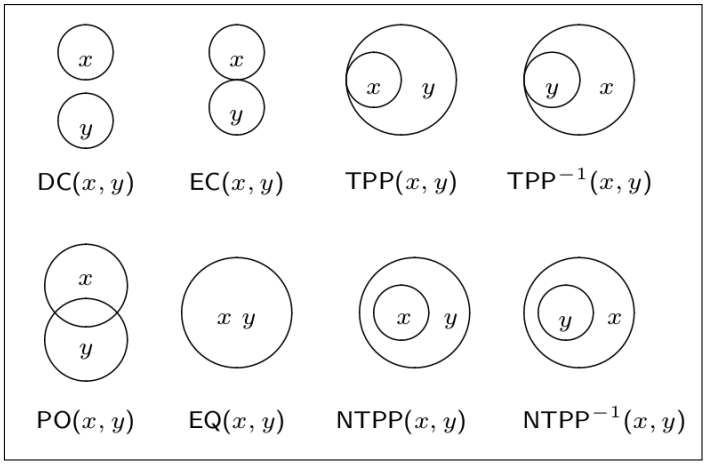


Figure 2.1: RCC-8 calculus with its eight binary base relations between region x and y (source: [Ren02]).

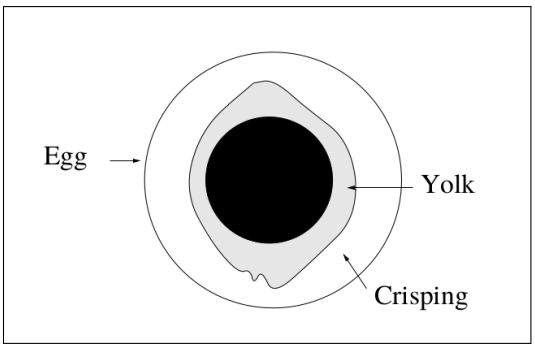


Figure 2.2: An egg-yolk interpretation of regions with indeterminate boundaries (source: [CH01]).

intelligence or for robotics purposes because the systems are acting in a real environment and handling objects located in space relative to other objects. Previous works in which qualitative spatial relations are used in a robotics domain are mainly related to object search and recognition. In the work by [KDH14], qualitative spatial relations are used for an indirect object search. The authors define two types of spatial relations, the so-called directional and distance relations. The directional relations contain relations such as *left-of*, *right-of*, *in-front-of*, and *behind-of* and the distance relations *close-to* and *distant-to*. Furthermore, the objects are divided into two groups, landmark and simple objects. While the landmark objects are static objects, the position of simple objects may change over time. To define the QSR, the *qualitative positional calculus* based on *ternary relations* [MTBF03] is used. In this calculus method, three positions are considered and referred by origin, relatum, and referent. To specify the positional calculus, those positions correspond to the robot, landmark, and sought after object, respectively. A given spatial relation is calculated by determining the partition that contains the object with respect

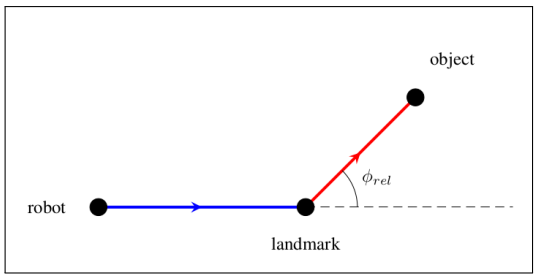


Figure 2.3: The 2D illustration of the reference axis specified by the robot, landmark and object. Regarding to the angle, the relations *left* of and *behind* are pictured (source: [KDH14]).

to the reference axis specified by the robot and landmark axis. Figure 2.3 displays this constellation with the resulting reference axis.

The directional relations are represented by Gaussian normal distribution with means 0, $\frac{1}{2}\phi$, ϕ, and $\frac{3}{4}\phi$. As a result, four directional relations are specified, these are: *behind-of*, *left-of*, *in-front-of*, and *right-of*. The sampling of the positions for qualitative spatial relations are based on the GMM with their mean on the corresponding angle of the qualitative spatial relations. To obtain the distance relations, the distance is calculated as the ratio between the object - landmark and the landmark - robot. The smaller the ratio, the closer the relation. If the ratio exceeds the given threshold, the relation is then classified as distant. The set of used qualitative spatial relations is represented as a 2D GMM. Based on the qualitative spatial relations between objects an indirect object search is performed. During the search, the position for the robot is calculated at which the sought after object can be seen. The search for an object is performed by utilizing so-called *supporting planes*, on which all objects must be located. Further, *view cones* are calculated with the corresponding voxels, and only voxels that are part of the supporting plane are considered. To find the object, the view cones considered the most probable for containing the sought after object at respective poses are selected and the robot can locate the pose most suitable for the optimal view of the object.

Another work in the context of using spatial relations for object search is by Ruiz et. al [RdSLS13], who presents a method for informed object search. In this method, a Bayesian framework that combines observation likelihoods and a so-called *spatial relation mask* for estimating a probability map of the search object. The given spatial relation is defined as a probability distribution of poses of the sought after object and the fixed poses of a second object. In this approach, four basis spatial relations such as: *very near*, *near*, *far*, and *very far* are defined by using spatial relation masks. The authors distinguished between two types of these masks: the *hard* and *soft* masks. The hard masks specify hard and the soft masks specify soft relations. Each of these masks are defined by threshold values, and the hard mask contains two threshold values whereas the soft mask contains four. To define the masks, four circles with different radii are used. The radius values, which are the parameters of the spatial relations masks, are predefined manually as the radius of 0-0.6 m for the relation very near, 0.6-1.0 m for near, 1.0-1.5 m for far, and 1.5 m for very far. To obtain the co-occurrence statistic between objects, the labeled images from the web

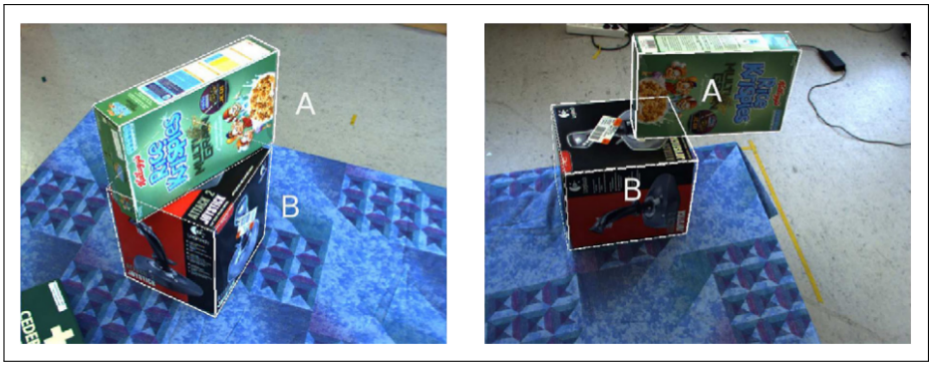


Figure 2.4: Illustration of the spatial relation *on* (source: [SAJ12]).

are used. Two object are in the given spatial relation if the measured distance between them corresponds to the mask with the given threshold value. In this way, it is determined whether the given sample belongs to the category. For example, if the probability to find a monitor near a keyboard is 77%, in the data, these two objects are located within the distance for the very near relation. The use of these spatial relations enable the system to obtain the probability distribution from samples of relative positions of objects. The search for a given object is performed by selecting poses that maximize the probability of finding a certain object at a given pose.

Sjöö and Aydemir [SAJ12] have used topological spatial relations to improve a visual object search, and the authors define two topological spatial relations, *on* (as illustrated in Figure 2.4) and *in*, to guide the search. For this purpose, an *applicability function* is calculated based on the pose and geometry of two objects, that is, *supporting* (reference) and *trajector* (target) object. This function measures the degree to which a given relation is applicable to the configuration of the model. For the *on* relation, the assumption that the target object is on the reference object is made if the reference object physically supports the target object. Therefore, if the reference object is removed, the target object falls. Further, two distances between these two objects are specified. The first value refers to the vertical distance, which indicates whether the target object is located too deep in the supporting object, and the second value specifies the distance in the horizontal direction, that is, how far the target object is located from the supporting object. Too great a distance would lead to the target object falling from the supporting object and consequently, no longer being located on top of it. Notably, in the calculation of the spatial relation *in*, the volume of the objects are considered. The spatial relation *in* refers to the ratio between the contained volume and total volume of the sought after object, that is, how much volume of the object is within the volume of the reference object. To locate the desired object in the environment, a probability distribution is calculated based on the spatial relations. This probability distribution specifies where the most probable location of the given object is, and based on this information, the next best view is selected to find the object.

Work by [KJD11] Kasper et al. introduces an approach for scene understanding in which spatial relations of objects are used. The authors recorded spatial relations from real office scenes to obtain statistical knowledge about average object sizes or appearance

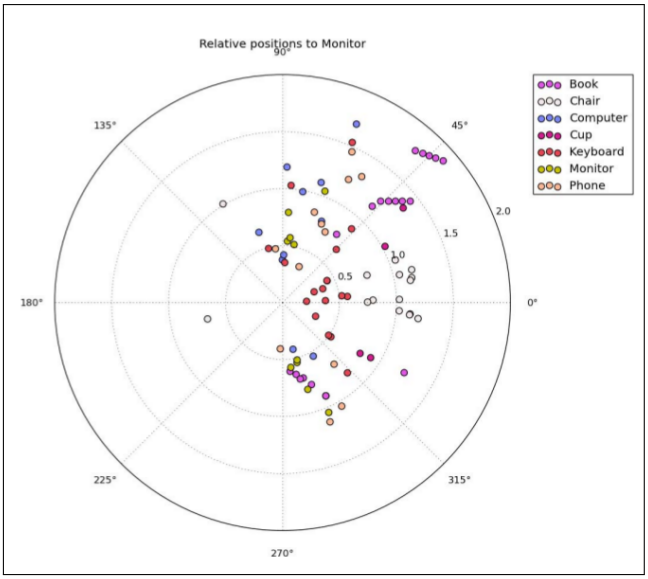


Figure 2.5: Relative position of the given objects to the query object *monitor* (source: [KJD11]).

probabilities of objects in the given scene type. Based on this spatial information, such as the objects' relationships, the structure of the scene is specified. The authors termed this *spatial proximity*. To annotate the data, bounding boxes are placed around the objects and then labeled with appropriate object classes. In this annotation process, the orientation of the bounding box is provided by the expert who labels the scenes. Importantly, this orientation depends on the object's functionality. The labeled data are then stored in a database, and the spatial relations are calculated by determining the Euclidean distance between two objects, i.e., their bounding boxes and relative position of the objects to each other. Figure 2.5 illustrates a relative position of the given objects to the object "monitor". For the specification of the distance relations, the object is projected onto the x-y plane and the Euclidean distance between these objects in the plane is calculated. To obtain a relative position, a so-called *query* object is selected and the positions of the remaining objects are calculated by considering their orientation to the query object. As a result, relative positions to the query object are calculated. This position is defined by the angle between the direction of the query object and the relative position of the objects. As a result, the 0 degree value specifies that an object is located *in-front-of* the query object and in cases of 180 degrees, *behind* it. The average size of an object type is learned by statistical measurement of the objects' bounding boxes. To classify a given scene, the probability distribution across all objects is calculated, and according to the obtained probability distribution, the scene type is classified. This classification is conducted by searching for objects that are typically observed in a given scene type.

In [EJvdMS14], the authors have proposed an approach for indirect object search based on probabilistic object-object relations. In this method, the authors use probabilistic statistical knowledge about object co-occurrences. The probability value specifies how likely an object is to co-occur with another object. To obtain this statistical information,

2D labeled camera images are used. The learned values then enable assumptions to be made about the likelihood of finding a given object under consideration by using objects currently present in the view. The author termed the reference objects "intermediate objects" and used these for active visual search. Figure 2.6 provides an example presented by the authors. To search for the target object, the probability of locating this object is combined with the travel cost, and the search is performed by considering the intermediate objects, which are possibly connected to the target object, until the sought after object is found.

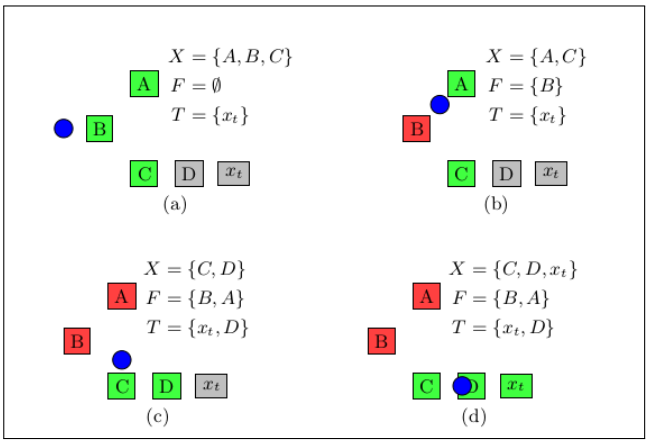


Figure 2.6: An example of the search strategy in which intermediate objects are used to locate the sought after object x_t. The blue dot represents the robot, the green objects can be seen by the robot, and red denotes locations behind the view area of the robot (source: [EJvdMS14]).

Additionally, Thippur et al. [TBK+15], [KBS+12] have used qualitative spatial relations for scene understanding. In this work, metric and qualitative spatial relations are used as spatial context information about the given scene and the objects contained in it for object recognition purposes. The spatial relations are learned from 3D labeled data [TAA+14]. Importantly, the authors distinguish between metric and qualitative spatial relations. The metric spatial relations are features based on relationships between pairs of objects. These so-called *object pair features* denote the spatial distribution of objects pairs and consists of the Euclidean distance between object centroids in an x-y plane. To learn these metric relations, GMM are applied to encode the probability distributions of the features. Based on the spatial relations between objects and typical object co-occurrences, a so-called *voting scheme* is applied to calculate a score value. This value specifies the assignment of a given object to possible objects classes. For the qualitative spatial relations, 12 relations are considered: 4 directional, 3 distance, 3 size, and 2 projective relations. The directional relations *behind*, *in-front-of*, *left-of*, and *right-of* are calculated using *ternary point calculus* in the same way as described in [KBS+12]. The distance relations *very-close-to*, *close-to*, and *distant-to* are obtained by clustering the metric distance relations into boundaries between given clusters, which correspond to a certain qualitative relation. Deepening on which cluster belongs to a given geometric relation, the associated qualitative relation is

assigned to the object involved. Size relations such as *shorter-than*, *narrower-than*, and *thinner-than* refers to the difference in each dimension of two objects. The projective relations are determined by applying *Allen's Interval Caluclus* [All83]. This method is conducted by using the projection on two objects' axis-aligned bounding boxes onto the x or y plane. The projective connectivity, such as the overlaps predicate, is then extracted for each axis of the object. Based on the qualitative spatial relations, a probabilistic reasoning method is used to derive the types of objects in a given scene. To achieve this, the prior portability of object types received from the perception system is combined with the probability that a given spatial relation holds between the given objects.

2.1.2 Summary

In the previous section, related works in which qualitative spatial relations are used for different purposes were presented. Although most of the aforementioned works were related to robotics applications such as object recognition or object search, some of these methods used qualitative spatial relations to perform spatial reasoning in general. Notably, these works are often based on pure geometry and related to 2D space. Moreover, by using many the approaches available, it is still not possible to determine how probable a given relation holds in the environment. Due to this disadvantage, the previously discussed methods are not directly suitable for real-world robotic applications.

In [CBGG97], [RCC92], the spatial relations are used as logical predicates with true/ false values and crisp definitions, such as the RCC. Region Connection Calculus and its variants are used to express qualitative relations between regions. However, the relations are of an "all-or-nothing" nature, for example, stating that a given region is touching another region. Furthermore, some base relations in RCC are possible, but there is no information regarding which of those relations are more likely. In the worst case, all the base relations are equally as probable.

Furthermore, when regarding the properties of qualitative binary spatial relations, given two spatial entities, only one of the basic relations can hold. Although such assumptions are particularly useful for symbolic reasoning, proof creation, or generally for formal systems, these concepts make the methods not directly suitable for robotics applications because real human environments are more vague than crisp, and although people are adept at handling uncertainties and ambiguities, systems also need methods to manage such issues. The further disadvantage of these type of relations is that the objects must lie axis-parallel to each other. In 3D space this is seldom the case, so this restriction reduces applicability for real-world purposes.

Another important aspect of real-world robotic applications is the handling of uncertainty and incomplete knowledge. Although the authors of [LC94] attempted to modeled vagueness by applying the "egg-yolk" method, this method does not enable the system to clearly state how vague the given assumption is. Therefore, this approach is not suitable for use in a real scenario.

Further work, in which the spatial relations are treated as a 2D relations, has been presented [KDH14]. The GMM used and the qualitative spatial relations are two-dimensional. Therefore, it is not possible to define relations such as *on* or *above*. Moreover, the environment is known in advanced as the 2D and 3D maps of the environment and the possible qualitative spatial relations between objects are provided to the robot. The robot

also always knows the full pose of the landmark objects. In contrast to the PQSR-based approach developed in this thesis, the occurrence of the relations can change and the spatial relations are learned from real 3D data. In [KDH14], only certain arrangements of objects are possible and those are strict. Moreover, the knowledge about the qualitative spatial relations is given in advance and the relations are crisp, also only the distribution of the relations is probabilistic and provided by the GMM. In this way, it seems that the distance relations depend on the robot's position. Specifically, the author makes the assumption that the robot is located in front of the table while calculating the distance relations. In [RdSLS13], the search space is also reduced to a two-dimensional space since the relations are defined in a 2D grid. The typical distances between objects for the spatial relations are learned from 2D images obtained from the LabelMe [RTMF08] and Flickr websites. Additionally, the threshold values for a given spatial relation, to be valid, are predefined and fixed. In [SAJ12], spatial knowledge about the relations is predefined and not learned. Moreover, the robot possesses the complete knowledge of which relations hold. This knowledge is not probabilistic. Thus, it cannot be determined how probable the given relation holds using this method. That is, the authors calculate only where the relation is fulfilled the most considering the pose and geometry of the given object. However, this value can be viewed more as an applicability measurement of the relation rather than probability, with 1 if the relation is completely fulfilled and 0 if it does not apply at all. The probability is then only used for searching for the target object by prediction places where the relation would most hold. Furthermore, only two spatial relations are considered in this method and voxels are used to define the probability for a given relation. In contrast, the PQSR proposed in this thesis are calculated in a continuous manner. Although, it the work by [SAJ12] it may appear that the pose of the reference object, e.g. table, is not known in advanced, the search space for the table is highly limited. Therefore, the robot receives the occupancy map of the environment in which the table can be assumed as the middle-law region in the map. In addition, the orientation of the table is fixed and points upward. The spatial relations introduced in [KJD11] are also of an "all or nothing" nature. Furthermore, the objects are selected and annotated by placing bounding boxes around them, and their orientation is provided by the human expert. Also, the approach in the present work does not contain any information about how likely a certain relation holds between given objects, since the relations are strict. The author takes only the Euclidean distance in the x-y plane and the relative position of the objects to each other into consideration. Because the average size of the objects is calculated from the bounding boxes and not the particular objects, this is a stark simplification and thus, not applicable for various different objects in the real world. Although the data are 3D, the relation calculations are primarily made in 2D. Moreover, only two relation types are used, that is, the relations based only on the distance between objects in the x-y plane and their relative position based on the orientation. The author also calculates the Euclidean distance between bounding boxes in 3D space, but there is no indication why this distinction is made. Similar to [RdSLS13], in [EJvdMS14], object co-occurrences are learned from 2D camera images. Although the method uses object co-occurrences, the authors do not directly consider the spatial relations between objects. Moreover, because learning the probability labeled images are used, this knowledge refers to objects in 2D space, and therefore, it is not possible to learn relations such as *behind* or *in-front-of*. The probabilistic statistical knowledge specifies the probability to locate a given object with

another object. Importantly, in the approach developed in this thesis, the probabilistic spatial relations are learned from 3D data and not 2D images. Comparable to [KBS+12] and [RdSLS13], and also in the work by [TBK+15], the spatial relations are learned in a two dimensional manner. Furthermore, the metric spatial relations are learned from real data but the qualitative relations are not, as the qualitative spatial relations are "extracted" by humans according to the definitions of qualitative relation type. Note that this factor constitutes a strong simplification and weakness of this approach. In a real-world scenario, the system does not have human classifiers and should preferably be capable of extracting relations without human support.

Taken together, it can be stated that although the works related to qualitative spatial relations are intended to be used in robotics applications, they do not consider some crucial aspects such as accuracy or missing knowledge. Due to simplifications such as two-dimensionality or predefined and crisp relations, the aforementioned methods are not directly suitable for real-world scenarios. Nevertheless, these works provide a basis for further development of approaches that utilize the qualitative spatial relations for more reliable object recognition and search processes. In this thesis, the qualitative spatial relations are used in a probabilistic manner. In this context, the goal of the present work is to provide a meaningful model of such relations that are suitable for real-world robotics applications.

2.2 Formalism

One of the primary goals of the thesis is to provide an appropriate and accurate formalism for PQSR based on extended methods and approaches that can be developed for different purposes. This section describes such a formalism, which combines the theories of the theoretic and symbolic approaches for qualitative spatial relations from the QSR field with probabilistic methodology from the AI field. However, because the PQSR are the new representation form of spatial relations, the corresponding formalism specifies basic rules for these new relations. Furthermore, the formalism conveys the theoretical approach for probabilistic relations with substantial rules, which must be followed to apply the PQSR properly. By using these rules, a reasoning technique based on the PQSR for different robotics approaches is developed. In the following section, the overall formalisms of the PQSR is presented. This formalism contains the axioms, definitions, and conventions valid for all further methods developed in this thesis, and the mathematical definitions for each spatial relation and its properties are also provided. The definitions and axioms serve as a basis for further calculations such as the SPF or the FI.

First, the basic definitions related to the PQSR are outlined in Section 2.2.1. This description is followed by definitions of the axioms and features of the PQSR. In Section 2.2.2, the formal definitions of spatial relations are presented. Following this, the learned method for the PQSR, including the related formulas, are described in Section 2.2.3.

2.2.1 Basic Definitions

This section describes the basic terminology used in the present work and these terms are valid for all methods provided in subsequent descriptions. In addition, the overall basic concepts and conventions, on which the formalism is based, are provided.

Object class vs. object instance Since the term *object* can have different meanings depending on the application or research field in which it used, the definition of an object used in this formalism is provided by Def. 2.1. In this work, an object is a spatial entity in a 3D space with given properties that underlie certain spatial rules such as the Axioms 1-3.

Furthermore, a distinction between an object's *class* and an *instance* is made. The object class describes a possible object type such as *table*, *monitor*, or *keyboard*, which can be found in a finite domain. In other words, an object class specifies a general concept of a given object from the environment. A specific object such as a particular table in a scene is then an instance of an object class.

Definition 1 *Let $\mathcal{L}$ be a finite set of possible object labels from the given domain, $\mathcal{C} \subseteq \mathbb{R}^3 \times \mathbb{R}^3$ a finite set of possible coordinate systems in which a given object can be located, and $\mathbb{P}(\Psi) \subseteq \mathbb{R}^3 \times [0;1]^3$ a finite power set of three-dimensionally colored points. Then, a finite set of spatial objects $\mathcal{D}$ is defined as follows:*

$$\mathcal{D} \subseteq \mathcal{L} \times \mathcal{C} \times \mathcal{P}(\Psi) \tag{2.1}$$

Types of spatial relations In this formalism, there are two types of the probabilistic spatial relations: the *binary* and *projective binary* relations. This distinction is based on the fact that the spatial meaning of some relations depend on the perspective from which the objects are observed, (as illustrated in Figure 2.7 and described in Example 1), whereas other relations still have the same meaning regardless of the view of the objects the relations contain. These perspective-independent relations are *binary relations* (Def. 2), which include spatial relations such as *near*, *above* and *on*. In turn, the *projective binary relations* (Def. 3) consist of relations such as *left-of*, *right-of*, *in-front-of*, and *behind-of*. The binary and projective binary relations comply with the defined coordinate system according to the Axioms 1-3.

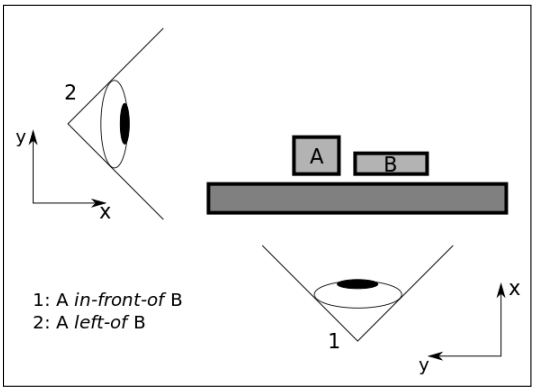


Figure 2.7: Illustration of the exemplary projective relations *in-front-of* and *left-of*.

In this formalism, each spatial relation holds between two objects. Since objects play a certain role in the given relation [JSZ+13], [Fre75], a distinction between a so-called *target* and *reference* object is made. The target object specifies the central object in the given

relation, whereas the reference object can be viewed as supporting object. As illustrated in Figure 2.8, the spatial relation refers to the position of a certain object, for example, the circle relative the reference object (the square). However, changing the role of the objects also changes the relation type. In the following definitions, the order of the object's type is fixed, as the first object is the target and second is the reference object. Both objects are taken from a finite set of all known objects in the domain.

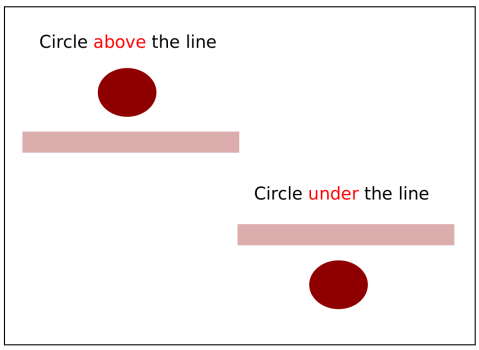


Figure 2.8: Illustration of the target and reference object's roles in a spatial relation. As can be observed in the given relation, the smaller object has been considered as a target object.

Example 1 *Consider a table desk with a monitor, a keyboard, and a phone observed by a robot. The keyboard and phone are in the relation* ***near*** *when the robot is located in front of the desk, as well as if it stands behind the desk. However, from the perspective of standing in front of the desk, the keyboard is located* ***in-front-of*** *the monitor, while when viewed from behind the desk, it is located* ***behind-of*** *the monitor. As* ***near*** *is independent from the robot's position, the relation is a binary relation, while* ***in-front-of*** *and* ***behind-of*** *are projective binary relations.*

Definition 2 *Let $\mathcal{D}$ be a finite set of all possible objects in the domain. Then, the set of* ***probabilistic binary relations*** *$\mathcal{R}_\beta$ is given by:*

$$\mathcal{R}_\beta \subseteq \{\mathcal{D}^2 \to [0;1]\} \tag{2.2}$$

Definition 3 *Again, let $\mathcal{D}$ be a finite set of all possible objects in the domain and $\mathcal{C}$ a set of possible coordinate systems that refer to reference view points. Then, the set of* ***projective binary probabilistic relations*** *$\mathcal{R}_\tau$ is defined as follows:*

$$\mathcal{R}_\tau \subseteq \{\mathcal{D}^2 \times \mathcal{C} \to [0;1]\} \tag{2.3}$$

Properties of the spatial relations Binary and projective binary relations can have several properties (see Def. 4), such as *symmetry*, *asymmetry*, *reflexivity*, *irreflexivity*, and *conversity*. However, these properties refer only to object instances and not classes. That

is, for properties, two particular objects from object classes are considered. Example 2 illustrates the property *symmetry*.

Example 2 *Consider a kitchen environment with two object instances such as* `fridge` *and* `cereals`*. These two particular objects are located* `near` *each other. In this case, the objects* `fridge` *and* `cereals` *denote the instance of the fridge's and cereals' classes between which the* `near` *relation holds. In this context, the* `near` *relation is* `symmetric` *because the cereals can be found near the fridge with the same probability as finding the fridge near the cereals. Although the near relation between these two objects is symmetric, the near relation in general does not have to be symmetric. Specifically, the probability of locating any cereals near any fridge can be symmetric, but this does not need to be equal in both directions. Therefore, if the probability of finding a fridge near the cereals is 80%, it does not necessarily mean that any cereals can also be found near a fridge with 80% probability.*

As illustrated in Example 2, the properties of the binary and projective binary relations are related exclusively to the object instances and not to the object classes. The reason for this is that the probability for a given spatial relation in general might vary even for the same object classes. Similarly, although the probability of finding any *mouse near* any *keyboard* is 90%, it does not mean that the probability of finding any *keyboard near* any *mouse* must also be 90%. Whereas the probability of finding a particular *mouse* (this particular mouse on a particular table) *near* a particular *keyboard* is equal to the probability to finding exactly this particular *keyboard near* this particular *mouse*.

Definition 4 *Let $r \in \mathcal{R}$ be a single relation between two object instances $s, t \in \mathcal{D}$. Then, a probabilistic spatial relation r is deemed:*

- ***symmetric**, if:*

$$\forall s, t \in \mathcal{D} : r(t, s) = r(s, t) \tag{2.4}$$

- ***asymmetric**, if:*

$$\forall s, t \in \mathcal{D} : r(t, s) \neq 0 \Rightarrow r(s, t) = 0 \tag{2.5}$$

- ***reflexive**, if:*

$$\forall s, t \in \mathcal{D} : r(t, s) = 1 \Rightarrow t \equiv s \tag{2.6}$$

- ***irreflexive**, if:*

$$\forall s, t \in \mathcal{D} : r(t, s) = 0 \Rightarrow t \equiv s \tag{2.7}$$

- ***converse** to relation r_1, if:*

$$\forall s, t \in \mathcal{D} : r(t, s) = r_1(s, t) \tag{2.8}$$

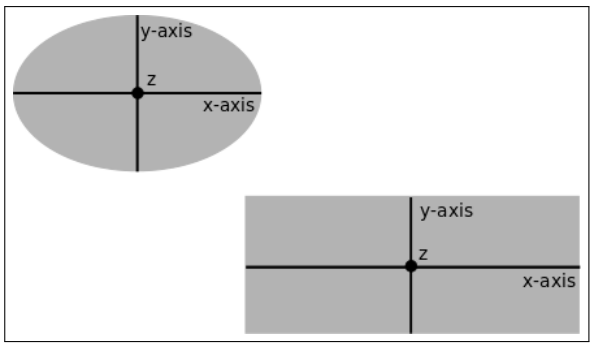


Figure 2.9: Illustration of object's axis.

2.2.1.1 Definition of the axioms

The PQSR must comply with the Axioms 1-2 to be valid and hold between two objects. These axioms define the right-handed coordinate system and object coordinates used in the entire formalism. Figures 2.9 and 2.10 visualize the coordinate system used and the object's coordinates. These axioms serve as the basis for further methods developed in this thesis. Importantly, the coordinate systems used in this approach must satisfy the following axioms:

Axiom 1 *Each coordinate system c from the set of all possible coordinate systems* $\mathcal{C} = \mathcal{R}^3 \times \mathbb{H}$ *must be right-handed. This rule refers to objects, worlds, and the robot's coordinate systems.*

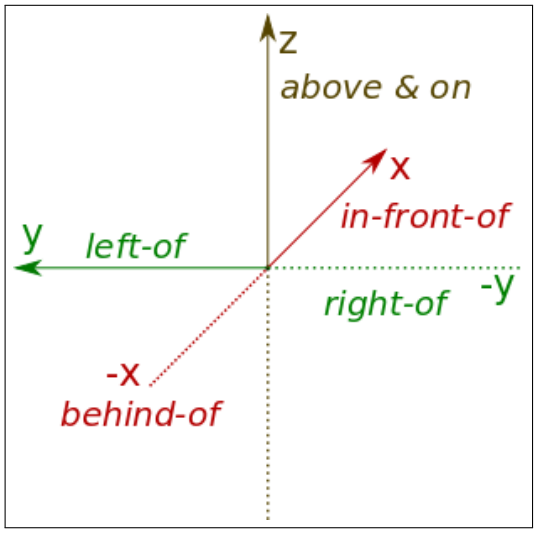


Figure 2.10: Illustration of a right-handed coordinate system.

Axiom 2 *The x-axis of the world coordinate system $c_\phi \in \mathcal{C}$ refers to the North-West-Up (NWU) convention, where the x-axis points in the north direction, the y-axis in the west direction, and the z-axis pointing upwards (opposite to the gravity vector).*

Axiom 3 *The coordinate system of an object is defined as follows: the vector $\vec{x}$ points to the direction of the longest inertia axis, the vector $\vec{y}$ points in the direction of the second longest inertia axis, and the vector $\vec{z}$ points in the direction of the shortest inertia axis. The position $p \in \mathbb{R}^3$ of an object's coordinate system equals to the center of axis-aligned cuboid in the world coordinate system.*

2.2.2 Modeling of the PQSR

Modeling of PQSR is conducted by calculating the probabilities for each spatial relation between objects extracted from real data under the consideration the previously learned average distance values typical for a given relation. This distance value is calculated from the instances of the objects' classes and describes the average distance to find the objects' classes in a given relation. The learned distance value for a relation depends on the object type. For instance, the distance value for a monitor located on a table differ from those for a keyboard and the "on" relation. A detailed description of how these relations and the distances are learned is provided in Section 2.2.3.

Because the learned distance value is related to a probability value for a certain relation, the spatial relation is most probable at this distance. With increasing distance from this position, the probability decreases. In general, all spatial relation calculations refer to object instances and not object classes. In the following section, modeling of the PQSR is described and the corresponding definitions for the spatial relations are provided.

For all following definitions, these terms are used:

- t_w, t_h, s_w, s_h, with $w, h \in \mathbb{R}$
  denote the target t or reference s object's width and depth, respectively. The object's width corresponds to the expansion in the object's x-axis direction, whereas the depth of the object's corresponds to the expansion in the object's y-axis direction.
- t_p, s_p with $p \in \mathbb{R}^3$
  denotes the position of the target t or reference s object. The position of the objects equal to the center of their axis-aligned cuboid.
- t_o, s_o with $o \in \mathbb{R}^{3\times3}$
  denotes the orientation of the target t or reference s object. The orientation of the objects correspond to the orientation of their axis-aligned cuboid.
- t_l, s_l with $l \in \mathcal{L}$
  denote the label of the target and reference object, from the finite set of all possible labels of the domain, respectively.

2.2.2.1 Binary Spatial Relations

Binary spatial relations specify probabilistic qualitative relations between two objects located in a 3D space. These include *near*, *above* and *on* relations. These relations are all translation and rotation invariant and, with the exception of the *on* relation, are also scale invariant. The exception refers to the maximum allowed distance value that limits the spatial relation. Because the *near* and *above* relation values are calculated according to the target and reference objects' width and depth, these relations are scale-invariant. In contrast, the *on* relation value is a constant threshold number and not related to the object features.

The property of the invariance in translation and rotation has the advantage that the objects are still in a given relation independently of how they are oriented and placed in the space. More precisely, the orientation and position of the objects to each other do not influence the probability for the given spatial relation.

Furthermore, in contrast to relations from related work discussed previously in section 2.1, the binary relations introduced in this thesis are calculated in a probabilistic manner and refer to 3D space.

Spatial relation near The spatial relation *near* describes the distance between two objects located in a 3D space. The resulting probability value indicates how probable the *near* relation holds between these two considered objects. For the *near* relation, only the Euclidean distance between the objects has an impact on the resulting probability value. Moreover, if this distance is too high, the relation does not hold. In contrast, the orientation of the objects does not influence the relation. Because the considered objects are located in 3D space, the *near* relation is of the 3D nature, and a visualization of the *near* relation in a real scene is provided in Figure 2.11.

Definition 5 *Given the reference and target objects $s, t \in \mathcal{D}$, the qualitative spatial relation r_{near} is defined by:*

$$r_{near}(t,s) = \begin{cases} 1 - d^{\Delta}_{near}(t,s), & \textit{if } \; d^{\Delta}_{near}(t,s) < 1 \\ 0, & \textit{otherwise} \end{cases} \tag{2.9}$$

The $d^{\Delta}_{\text{near}}(t,s)$ describes the scale invariant distance resulting from the ratio of the actual distance $d_{\text{near}}(t,s)$ between the target t and reference s objects and the maximum allowed distance $\hat{d}_{\text{near}}(t,s)$:

$$d^{\Delta}_{\text{near}}(t,s) = d_{\text{near}}(t,s) / \hat{d}_{\text{near}}(t,s) \tag{2.10}$$

The $d_{\text{near}}(t,s)$ denotes the Euclidean distance between the Center of Gravity (CoG) of the cuboid of each object with respect to the learned distance $\ell_{near}(\tilde{t}, \tilde{s})$ Def. 13 for the near relation:

$$d_{\text{near}}(t,s) = ||t_p - s_p| - \ell_{near}(\tilde{t}, \tilde{s})| \tag{2.11}$$

The $d_{\text{near}}(t, s)$ value has to be smaller than the maximum allowed distance $\hat{d}_{\text{near}}(t, s)$. This maximum allowed distance between the objects t and s is calculated by the sum of the objects' width and depth.

$$\hat{d}_{\text{near}}(t, s) = s_w + s_h + t_w + t_h \tag{2.12}$$

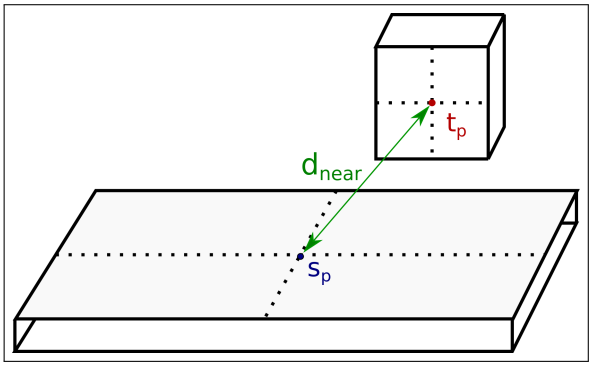


Figure 2.11: Visualization of the *near* relation with corresponding terminology.

According to Def. 5, the resulting probability value for the *near* relation can only be obtained if the relative distance $d^{\Delta}_{\text{near}}(t, s)$ 2.10 is smaller than 1. If this is the case, this value indicates that the target object is located within the allowed area for the *near* relation. Otherwise, the target object is beyond the acceptance area and the probability value for the considered objects in the near relation is equal to 0. Hence, the objects are too far from each other to be in the *near* relation. Figure 2.11 lists the terminology according to the Def. 5 of the *near* relation.

Due to the qualitative nature of the spatial relation and the quantity of the real-world data, a value must be specified that limits the spatial relation. As a *near* term can be matched to a different extension of a physical space, a given area for the relation to be valid must be defined. In regards to the calculation 2.12, the allowed distance is the result of the sum of the object's width and depth, and thus, depends on their sizes. That is, the area (or rather space) in which the given objects are located must be at least equal or higher than the sizes of the objects. For example, if only a fixed threshold value is taken into account as a maximum allowed distance, this could lead to undesirable effects. In cases of a too small threshold value, one of the considered objects could be too large and as a consequence, the objects are no longer in a *near* relation even if they are located in close proximity to each other. Additionally, the *near* relation refers to a space that is relative and depends on the object types involved in the relation. For instance, the maximum allowed value of, for example, 0.10 m can influence the area relation according to the target object's size. Although the area of 0.10 m close to a mouse, given its small size, would be reasonable, the same area for a table would be too small for the *near* relation. Therefore, if only the size of one object, such as the reference object, is considered, an inconsistency of the relation could result.

By having two objects such as a laptop and a mouse, with respect to the mouse size only, the laptop would be outside the acceptance area and thus, not near mouse. By

taking into account the sizes of both objects for the maximum allowed value calculation in this context, the symmetry of the relation can be retained.

Importantly, the probability for the *near* relation decreases with increasing Euclidean distance from the position where the relation is most probable until the maximal allowed distance $\hat{d}_{\text{near}}(t, s)$ 2.12 is reached. Since the probability for two objects is equal if the objects are reversed, the *near* relation is symmetric according to Def. 4-Eq. 2.4. A further property of this relation is reflexivity Def. 4-Eq. 2.6, as the given object's instance is near itself with 100% probability.

Spatial relation above The spatial relation *above* specifies whether a target object is located above the reference object. For the above relation however, it must be ensured that the target object is located higher than the reference object and within the reference object's plane area. Although the relation does not depend on the orientation of both objects, the position of the objects is a critical factor for the relation to be valid. In Figure 2.12 a visualization of the *above* relation in a scene is provided.

Definition 6 *Reference and target objects $s, t \in \mathcal{D}$ as well as the world coordinate system $c_\phi \in \mathcal{C}$ (that obeys axioms 1-3), the spatial relation $r_{above} \in \mathcal{R}$ are defined as:*

$$r_{above}(t, s) = \begin{cases} 1 - d^{\Delta}_{above}(t, s), & \textit{if } \Lambda^{a}_{above}(t, s) \ \wedge \ d^{\Delta}_{above}(t, s) < 1 \\ 0, & \textit{otherwise} \end{cases} \tag{2.13}$$

The $d^{\Delta}_{\text{above}}(t, s)$ is obtained by dividing the actual distance between the objects $d_{\text{above}}(t, s)$ by the maximum allowed distance $\hat{d}_{\text{above}}(t, s)$.

$$d^{\Delta}_{\text{above}}(t, s) = d_{\text{above}}(t, s) / \hat{d}_{\text{above}}(t, s) \tag{2.14}$$

The distance $d_{\text{above}}(t, s)$ denotes the *projective* distance between the objects' center of gravity in the z-axis dimension. To calculate this projective distance, the learned distance $\ell_{above}(\tilde{t}, \tilde{s})$ Def. 13 for the above relation is taken into account.

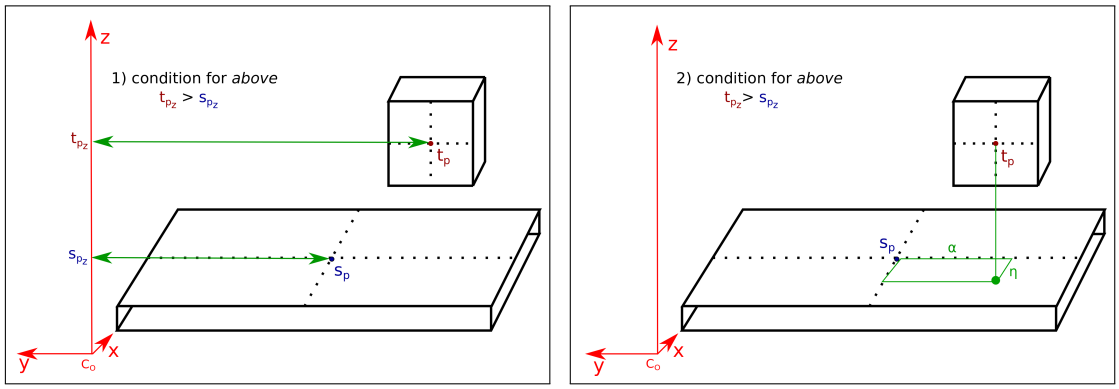
$$d_{\text{above}}(t, s) = |(t_{p_z} - s_{p_z}) - \ell_{above}(\tilde{t}, \tilde{s})| \tag{2.15}$$

The maximum allowed distance $\hat{d}_{\text{above}}(t, s)$ is provided by the sum of the objects' width and depth:

$$\hat{d}_{\text{above}}(t, s) = s_w + s_h + t_w + t_h \tag{2.16}$$

For the relation *above* to be valid, the $\Lambda^{a_1}_{\text{above}}(t, s)$ must be fulfilled. This condition requires that the value of the target object's z-position t is greater than the value of the z-position of the reference object s. As a result, the target object is located higher than the reference object.

$$\Lambda^{a_1}_{\text{above}}(t, s) : t_{p_z} > s_{p_z} \tag{2.17}$$



(a) The first constrain of the *above* relation (b) The second constrain of the *above* relation

Figure 2.12: Visualization of the *above* relation with corresponding terminology.

Regarding the condition $\Lambda^{a_2}_{\text{above}}(t,s)$, the target object must be located within the area above the surface of the reference object. This condition means that if the target object "falls", then it will fall into the reference object. In the formula of the condition $\Lambda^{a_2}_{\text{above}}(t,s)$ 2.18, o_x and o_y denote the inertia axis of the reference objects in the x- and y-directions respectively. The ϕ_{o_z} refers to the world coordinate system axis in the z-direction.

$$\Lambda^{a_2}_{\text{above}}(t,s) : \exists \alpha, \eta, \gamma : (s_p + \alpha\ \vec{s_{o_x}} + \eta\ \vec{s_{o_y}} = t_p + \gamma\ \vec{c_{\phi_{o_z}}}) \\ \wedge\ (|\alpha| \leq \frac{s_w}{2}, |\eta| \leq \frac{s_h}{2})) \tag{2.18}$$

The term $\Lambda^{a}_{\text{above}}(t,s)$ includes the given conditions $\Lambda^{a_1}_{\text{above}}(t,s)$ 2.17 and $\Lambda^{a_2}_{\text{above}}(t,s)$ 2.18. Regarding the $\Lambda^{a}_{\text{above}}(t,s)$, both conditions must be satisfied so the target object is located above the reference object:

$$\Lambda^{a}_{\text{above}}(t,s) : \Lambda^{a_1}_{\text{above}}(t,s) \wedge \Lambda^{a_2}_{\text{above}}(t,s) \tag{2.19}$$

According to Def. 6, a target object is located *above* the reference object if the relative distance value is smaller than 1 and the object is located within the area of the reference object. More precisely, an object is located higher than the reference object. The physical space in which the target object is located must refer to the area provided by the reference object plane. Figure 2.12 illustrates the conditions and provides visual representations of the terms used in the definition. Importantly, the target object must be located within the area above the surface of the reference object according to the condition $\Lambda^{a_2}_{\text{above}}(t,s)$ 2.18. Furthermore, for the relation, the projective distance between the objects $d_{\text{above}}(t,s)$ 2.15 should not be greater than the average results from the sum of the reference and target objects' sizes $\hat{d}_{\text{above}}(t,s)$ 2.16. The value of the target object's z-position t must be higher than the value of the z-position of the reference object s, which follows the condition $\Lambda^{a_1}_{\text{above}}(t,s)$ 2.17. This relation is asymmetric (Def. 4 - Eq. 2.5) and irreflexive (Def. 4 - Eq. 2.7).

Spatial relation on The *on* relation indicates whether a target object is located *on* the reference object. In this context, an object is on another object if the distance between this object and the plane of the reference object is smaller than a fixed value $\lambda \in \mathcal{R}^+$. Furthermore, the target object must be located higher than the reference object. In contrast to the *above* relation, the orientation of a reference object influences the *on* relation. Due to this aspect, the *on* relation may relate to not only the horizontal but also the vertical directions [LFHU06]. As a result, it can be defined that an object such as a picture is located on the wall. Figures 2.13 and 2.14 illustrate the possible orientations for the *on* relation and the corresponding terminology.

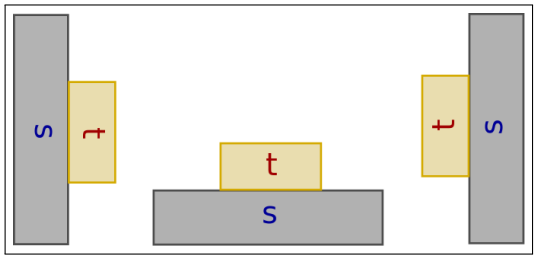


Figure 2.13: Simplified illustration of three possible orientations for a target object t in an *on* relation with a reference object s.

Definition 7 *Given reference and target object $s, t \in \mathcal{D}$, the probabilistic spatial relation $r_{on} \in \mathcal{R}$ is defined, as follows:*

$$r_{on}(t,s) = \begin{cases} 1 - d^{\Delta}_{on}(t,s), & \text{if } \Lambda^{a}_{on}(t,s) \wedge d^{\Delta}_{on}(t,s) < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.20}$$

The $d^{\Delta}_{\text{on}}(t,s)$ value is calculated by dividing the distance between the objects in the z-axis direction with the $\lambda \in \mathcal{R}^+$ value. The λ provides the threshold value for the maximum allowed distance, so the object is still in the *on* relation. For the calculation of $d^{\Delta}_{\text{on}}(t,s)$, the learned distance of $\ell_{on}(\tilde{t},\tilde{s})$ Def. 13 the *on* relation is considered:

$$d^{\Delta}_{\text{on}}(t,s) = (|u_{\text{on}}(t,s)_z - \ell_{on}(\tilde{t},\tilde{s})|)/\lambda \tag{2.21}$$

The $u_{\text{on}}(t,s)$ denotes the vector between the target and reference object in the coordinate system of the reference object:

$$u_{\text{on}}(t,s) = s_o(t_p - s_p) \tag{2.22}$$

The condition $\Lambda^{a_1}_{\text{on}}(t,s)$ describes the distance in the z-axis direction between the target and reference object in the coordinate system of the reference object. In this condition, the value must be greater or equal to 0. Specifically, the target object must be located higher than the reference object:

$$\Lambda^{a_1}_{\text{on}}(t,s) : u_{\text{on}}(t,s)_z \geq 0 \tag{2.23}$$

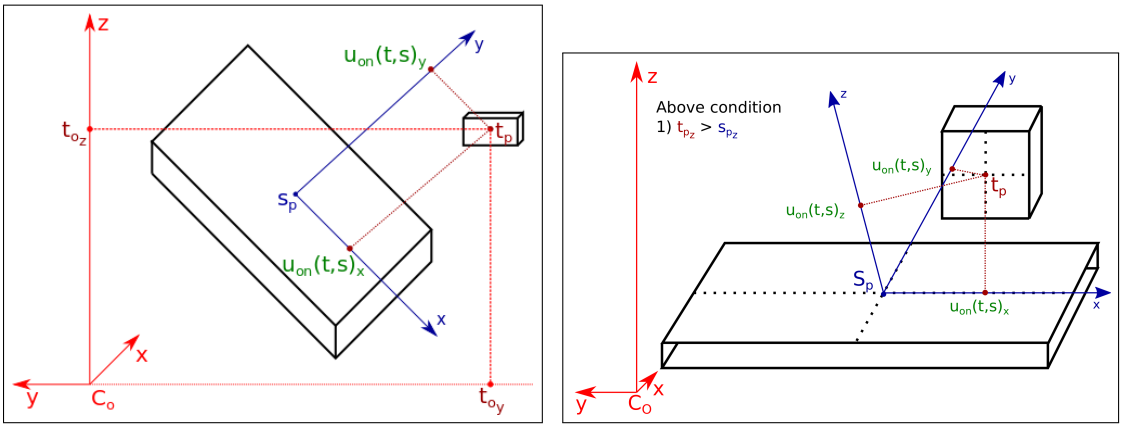
The further $\Lambda^{a_2}_{\text{on}}(t,s)$ and $\Lambda^{a_3}_{\text{on}}(t,s)$ conditions for the *on* relation refer to the area in which the target object must be located. That is, the target object must be located within the area of the reference object:

$$\Lambda^{a_2}_{\text{on}}(t,s) : |u_{\text{on}}(t,s)_x| < \frac{s_w}{2} \tag{2.24}$$

$$\Lambda^{a_3}_{\text{on}}(t,s) : |u_{\text{on}}(t,s)_y| < \frac{s_h}{2} \tag{2.25}$$

The last condition $\Lambda^{a}_{\text{on}}(t,s)$, summarizes the previous conditions $\Lambda^{a_1}_{\text{on}}(t,s)$ 2.23, $\Lambda^{a_2}_{\text{on}}(t,s)$ 2.24, and $\Lambda^{a_3}_{\text{on}}(t,s)$ 2.25 for the overall condition:

$$\Lambda^{a}_{\text{on}}(t,s) : \Lambda^{a_1}_{\text{on}}(t,s) \wedge \Lambda^{a_2}_{\text{on}}(t,s) \wedge \Lambda^{a_3}_{\text{on}}(t,s) \tag{2.26}$$



(a) First step of the coordinates transformation of the target to reference objects for the *on* relation

(b) Second step of the coordinates transformation of the target to reference objects for the *on* relation

Figure 2.14: Visualization of the *on* relation with corresponding terminology.

According to Def. 7, two objects are in the spatial *on* relation if they are located no further away than λ and if the target is located within the area described by the plane of the reference object. Since the position of both objects is provided in the coordinate system of the reference object, the spatial relation *on* can hold in horizontal as well as vertical alignments. In contrast to the *above* relation, the maximum allowed distance λ is a fixed value and does not depend on the objects' sizes. Specifically, for the *on* relation to be valid, the allowed distance value must be correspondingly small, and this can not be ensured by the value that results from the objects' sizes. In this case, the allowed distance for an object being *on* another object is too high and would reject the expectation for the relation where an object is located very near to another object. On the other hand, the assumption that an object must touch another object would not be suitable for real-world applications, as the sensory data are noisy and therefore the distances would vary

depending on the quality of the data. Figure 2.14 illustrates the terminology used in the definition of the *on* relation.

2.2.2.2 Projective Spatial Relations

According to Def. 3, a projective binary relation describes a spatial relationship between two objects with respect to the given view point from which the objects have been observed. This view can be related to the robot's coordinate system or a camera's view. In contrast to the binary spatial relations, the projective relations are only scale invariant. This term means that the scale of the object does not influence the result of the projective relation. The projective relations include 3D relations such as *left-of*, *right-of*, *in-front-of* and *behind-of*. Example 3 illustrates the perspective dependency in the projective relations.

Example 3 *Consider a table desk with a monitor, keyboard, and phone observed by a robot. The keyboard and phone are in the relation* `near` *despite if the robot is located* `in-front-of` *the desk or if it stands* `behind` *the desk. However, from the perspective of standing in front of the desk, the keyboard is located* `in-front-of` *the monitor but when seen from the back of the desk, it is located* `behind-of` *the monitor. As* `near` *is independent from the robot's position, the relation is a binary relation, while* `in-front-of` *and* `behind-of` *are considered projective binary relations.*

Projective spatial relation left-of The *left-of* relation specifies whether a target object is located to the left of a reference object with respect to the view from which the objects are observed. The *left-of* relation is calculated by comparing the projected y-position values of the target and reference objects. An object is located to the left of the reference object if its y-position value is smaller than the value of the y-position of the reference object for a given reference coordinate system.

Definition 8 *Given reference and target objects $s, t \in \mathcal{D}$ as well as the reference coordinate system $c \in \mathcal{C}$, the spatial projective relation $r_{left\text{-}of} \in \mathcal{R}$ is defined, as follows:*

$$r_{left\text{-}of}(t,s,c) = \begin{cases} 1 - d^{\Delta}_{left\text{-}of}(t,s,c), & \text{if } \Lambda^{a}_{left\text{-}of}(t,s,c) \wedge d^{\Delta}_{left\text{-}of}(t,s,c) < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.27}$$

The $d^{\Delta}_{\text{left-of}}(t,s,c)$ value denotes the difference between the distance of the target and reference objects in the y-axis direction and the maximal allowed distance $\hat{d}_{\text{left-of}}(t,s)$ 2.29. The term c refers to the coordinate system, i.e., the projection from which the relation is being observed. In this calculation, the learned distance $\ell_{\text{left-of}}(\tilde{t},\tilde{s},c)$ Def. 14 is considered.

$$d^{\Delta}_{\text{left-of}}(t,s,c) = ||(t_{\text{left-of}}(c)_y - s_{\text{left-of}}(c)_y)| - \ell_{\text{left-of}}(\tilde{t},\tilde{s},c)|/\hat{d}_{\text{left-of}}(t,s) \tag{2.28}$$

The value $\hat{d}_{\text{left-of}}(t,s)$ is calculated by the sum of the object's width and depth.

$$\hat{d}_{\text{left-of}}(t,s) = s_w + t_w + s_h + t_h \tag{2.29}$$

The terms $t_{\text{left-of}}(c)$ and $s_{\text{left-of}}(c)$ 2.31 refer to the rotation of the CoG of the target and reference object in the projective coordinate system c. This projection is required because the *left-of* relation is a projective relation and thus depends on the view from which the relation is considered.

$$t_{\text{left-of}}(c) = c_o t_p \tag{2.30}$$
$$s_{\text{left-of}}(c) = c_o s_p \tag{2.31}$$

According to the condition $\Lambda^{a_1}_{\text{left-of}}$, the target object is located to the left of the reference object if its value of the y-axis is smaller than the value of the y-axis of the reference object.

$$\Lambda^{a_1}_{\text{left-of}}(t,s,c) : t_{\text{left-of}}(c)_y > s_{\text{left-of}}(c)_y \tag{2.32}$$

The condition $\Lambda^{a_2}_{\text{left-of}}$ requires that the maximum allowed distance is greater than the value obtained from the Euclidean distance between the CoG of the target and reference objects.

$$\Lambda^{a_2}_{\text{left-of}}(t,s,c) : |t_p - s_p| < \hat{d}_{\text{left-of}}(t,s) \tag{2.33}$$

The condition $\Lambda^{a}_{\text{left-of}}(t,s,c)$ includes the previous conditions $\Lambda^{a_1}_{\text{left-of}}(t,s,c)$ 2.32, and $\Lambda^{a_2}_{\text{left-of}}(t,s,c)$ 2.33.

$$\Lambda^{a}_{\text{left-of}}(t,s,c) : \Lambda^{a_1}_{\text{left-of}}(t,s,c) \wedge \Lambda^{a_2}_{\text{left-of}}(t,s,c) \tag{2.34}$$

According to the Def. 8, the probability value for the relation is valid if the constraints are satisfied and the relative distance $d^{\Delta}_{\text{left-of}}(t,s,c)$ 2.28 is smaller than 1. Of most importance for this relation is the constraint that the value of the y-position of the target object is greater than the y-position of the reference object. If this is the case, the object is located *left-of* of the other object. The *left-of* relation is asymmetric Def. 4-Eq. 2.5, irreflexive Def. 4-Eq. 2.7, and converse Def. 4-Eq. 2.8 to the relation *right-of*. Figure 2.15 displays the *left-of* and *right-of* relations with the corresponding terminology.

Projective spatial relation right-of Similar to relation *left-of* the *right-of* relation describes the target object's position with respect to the reference object under consideration in the given view. In contrast to the *left-of* relation however, the y-position value of a target object must be smaller than the y-position value of the reference object. Furthermore, the distance between both objects shall not exceed the maximum allowed value.

Definition 9 *Given the reference and target objects $s, t \in \mathcal{D}$ and the coordinate system $c \in \mathcal{C}$, the projective spatial relation $r_{right\text{-}of} \in \mathcal{R}$ is defined, as follows:*

$$r_{right\text{-}of}(t,s,c) = \begin{cases} 1 - d^{\Delta}_{right\text{-}of}(t,s,c), & \text{if } \Lambda^{a}_{right\text{-}of}(t,s,c) \wedge d^{\Delta}_{right\text{-}of}(t,s,c) < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.35}$$

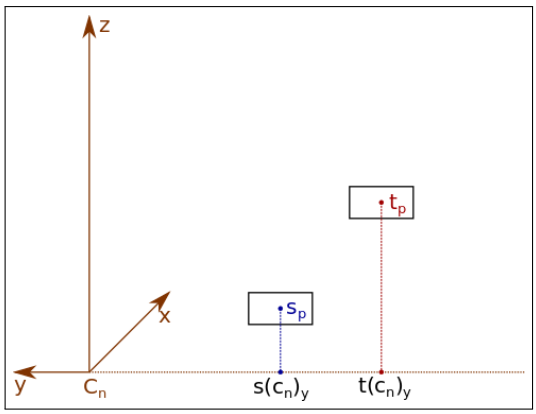


Figure 2.15: Visualization of the *left-of* and *right-of* relations with corresponding terminology.

The $d^{\Delta}_{\text{right-of}}(t, s, c)$ of the *right-of* relation is calculated in a similar way to the Formula 2.28 of the *left-of* relation. However, the learned distance $\ell_{\text{right-of}}(\tilde{t}, \tilde{s}, c)$ Def. 14 refers to the *right-of* relation.

$$d^{\Delta}_{\text{right-of}}(t, s, c) = ||(t_{\text{right-of}}(c)_y - s_{\text{right-of}}(c)_y)| - \ell_{\text{right-of}}(\tilde{t}, \tilde{s}, c)|/\hat{d}_{\text{right-of}}(t, s, c) \quad (2.36)$$

Similar to the *left-of* relation, the CoG of the target and reference objects are first projected onto the coordinate system c of a given view. The terms $t_{\text{right-of}}(c)$ and $s_{\text{right-of}}(c)$ denote the projected objects.

$$t_{\text{right-of}}(c) = c_o t_p \quad (2.37)$$
$$s_{\text{right-of}}(c) = c_o s_p \quad (2.38)$$

The maximum allowed distance $\hat{d}_{\text{right-of}}(t, s)$ results from the sum of the widths and depths of the target and reference objects.

$$\hat{d}_{\text{right-of}}(t, s) = s_w + t_w + s_h + t_h \quad (2.39)$$

For a target object to be in the *right-of* relation with the reference object, its y-value must not exceed the y-value of the reference object according to the condition $\Lambda^{a_1}_{\text{right-of}}(t, s, c)$.

$$\Lambda^{a_1}_{\text{right-of}}(t, s, c) : t_{\text{right-of}}(c)_y < s_{\text{right-of}}(c)_y \quad (2.40)$$

The second condition $\Lambda^{a_2}_{\text{right-of}}(t, s, c)$ requires that the maximum allowed distance must be greater than the value obtained from the Euclidean distance between the center of gravity of the both the target and reference objects.

$$\Lambda^{a_2}_{\text{right-of}}(t,s,c) : |t_p - s_p| < \hat{d}_{\text{right-of}}(t,s) \tag{2.41}$$

The condition $\Lambda^{a}_{\text{right-of}}(t,s,c)$ includes all previous conditions $\Lambda^{a_1}_{\text{right-of}}(t,s,c)$ and $\Lambda^{a_2}_{\text{right-of}}(t,s,c)$.

$$\Lambda^{a}_{\text{right-of}}(t,s,c) : \Lambda^{a_1}_{\text{right-of}}(t,s,c) \wedge \Lambda^{a_2}_{\text{right-of}}(t,s,c) \tag{2.42}$$

Accordingly, the $d^{\Delta}_{\text{right-of}}(t,s,c)$ must be smaller than 1. In addition, the maximum allowed value must be greater than the Euclidean distance between the considered objects. The maximum allowed value $\hat{d}_{\text{right-of}}(t,s)$ is calculated by the sum of the objects' widths and depths and thus depends on their sizes. The *right-of* relation is asymmetric Def. 4-Eq. 2.5, irreflexive Def. 4-Eq. 2.7, and converse Def. 4-Eq. 2.8 to the relation *left-of*.

Projective spatial relation in-front-of The *in-front-of* relation provides the probability value for a target object to be located *in-front-of* of the reference object. In this relation, the distance between two objects in the x-axis direction and the reference view are considered. Figure 2.16 illustrates the *in-front-of* and *behind-of* relations.

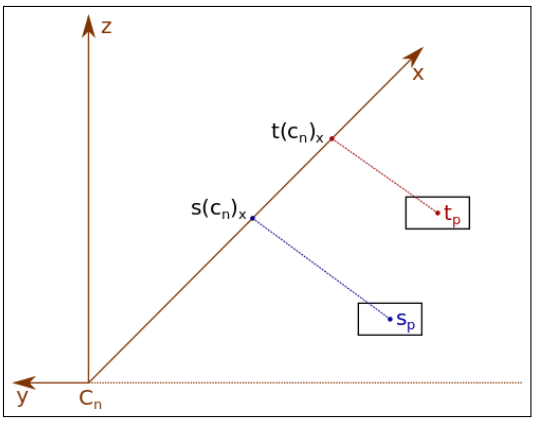


Figure 2.16: Visualization of the *in-front-of* and *behind-of* relations with corresponding terminology.

Definition 10 *Given reference and target objects $s, t \in \mathcal{D}$ as well as the coordinate system $c \in \mathcal{C}$, the spatial projective relation $r_{in\text{-}front\text{-}of} \in \mathcal{R}$ is defined, as follows:*

$$r_{in\text{-}front\text{-}of}(t,s,c) = \begin{cases} 1 - d^{\Delta}_{in\text{-}front\text{-}of}(t,s,c), & \text{if } \Lambda^{a}_{in\text{-}front\text{-}of}(t,s,c) \wedge d^{\Delta}_{in\text{-}front\text{-}of}(t,s,c) < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.43}$$

The $d^{\Delta}_{\text{in-front-of}}(t,s,c)$ is provided by the difference between the distance of the target and reference objects in the x-axis direction and the maximum allowed value $\hat{d}_{\text{in-front-of}}(t,s)$

calculated from the width and depth of both objects. This value is calculated with respect to the learned distance value for the $\ell_{\text{in-front-of}}(\tilde{t}, \tilde{s}, c)$ Def. 14 relation.

$$d^{\Delta}_{\text{in-front-of}}(t, s, c) = \quad ||(t_{\text{in-front-of}}(c)_x - s_{\text{in-front-of}}(c)_x)| - \ell_{\text{in-front-of}}(\tilde{t}, \tilde{s}, c)|/\hat{d}_{\text{in-front-of}}(t, s) \tag{2.44}$$

Similar to the maximum allowed distances of the other projective relations, $\hat{d}_{\text{in-front-of}}(t, s)$ is obtained from the sum of the target and reference objects' width and depth.

$$\hat{d}_{\text{in-front-of}}(t, s) = s_w + t_w + s_h + t_h \tag{2.45}$$

Because the *in-front-of* relation depends on the view point, the center of gravity of the both objects t and s are rotated to the orientation of the reference coordinate system c. This results in $t_{\text{in-front-of}}(c)$ and $s_{\text{in-front-of}}(c)$.

$$t_{\text{in-front-of}}(c) = c_o t_p \tag{2.46}$$
$$s_{\text{in-front-of}}(c) = c_o s_p \tag{2.47}$$

For condition $\Lambda^{a_1}_{\text{in-front-of}}(t, s, c)$, a target object is located in front of a reference object if the value of the target object's x-axis is smaller than the x-axis value of the reference object.

$$\Lambda^{a_1}_{\text{in-front-of}}(t, s, c) : t_{\text{in-front-of}}(c)_x < s_{\text{in-front-of}}(c)_x \tag{2.48}$$

The second $\Lambda^{a_2}_{\text{in-front-of}}(t, s, c)$ requires that the maximum allowed distance $\hat{d}_{\text{in-front-of}}(t, s)$ 2.45 must be greater than the value obtained from the Euclidean distance between the center of gravity of both the target and reference objects.

$$\Lambda^{a_2}_{\text{in-front-of}}(t, s, c) : |t_p - s_p| < \hat{d}_{\text{in-front-of}}(t, s) \tag{2.49}$$

The $\Lambda^{a}_{\text{in-front-of}}(t, s, c)$ condition is satisfied if the $\Lambda^{a_1}_{\text{in-front-of}}(t, s, c)$and $\Lambda^{a_2}_{\text{in-front-of}}(t, s, c)$ conditions hold.

$$\Lambda^{a}_{\text{in-front-of}}(t, s, c) : \Lambda^{a_1}_{\text{in-front-of}}(t, s, c) \wedge \Lambda^{a_2}_{\text{in-front-of}}(t, s, c) \tag{2.50}$$

Analogous to the relations *left-of* and *right-of* the maximum allowed distance $\hat{d}_{\text{in-front-of}}(t, s)$ between two objects results from their sizes. The *in-front-of* relation is asymmetric Def. 4-Eq. 2.5, irreflexive Def. 4-Eq. 2.7, and converse Def. 4-Eq. 2.8 to the relation *behind-of*.

Projective spatial relation behind-of The *behind-of* relation specifies whether a target object is located *behind-of* the reference object. Analogous to all remaining relations, the resulting probability value provides the probability value of how how likely it is that the relation holds.

Definition 11 *Given reference and target objects $s, t \in \mathcal{D}$ as well as the reference coordinate system $c \in \mathcal{C}$, the probabilistic projective relation $r_{behind\text{-}of} \in \mathcal{R}$ is defined, as follows:*

$$r_{behind\text{-}of}(t,s,c) = \begin{cases} 1 - d^{\Delta}_{behind\text{-}of}(t,s,c), & \text{if } \Lambda^{a}_{behind\text{-}of}(t,s,c) \wedge d^{\Delta}_{behind\text{-}of}(t,s,c) < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.51}$$

$d^{\Delta}_{\text{behind-of}}(t,s,c)$ denotes the difference between the object's distance in the x-axis direction and the maximum allowed distance $\hat{d}_{\text{behind-of}}(t,s)$. In this calculation, the learned distance $\ell_{\text{behind-of}}(\tilde{t},\tilde{s},c)$ Def. 14 for the behind relation is considered.

$$\begin{aligned} d^{\Delta}_{\text{behind-of}}(t,s,c) \quad &= ||(t_{\text{behind-of}}(c)_x - s_{\text{behind-of}}(c)_x)| - \\ &\quad \ell_{\text{behind-of}}(t,s,c)|/\hat{d}_{\text{behind-of}}(t,s) \end{aligned} \tag{2.52}$$

The terms $t_{\text{behind-of}}(c)$ and $s_{\text{behind-of}}(c)$ refer to the rotation of the CoG of both objects in the orientation of the reference coordinate system c.

$$t_{\text{behind-of}}(c) = c_o t_p \tag{2.53}$$

$$s_{\text{behind-of}}(c) = c_o s_p \tag{2.54}$$

The relation $\hat{d}_{\text{behind-of}}(t,s)$ is provided by the sum of the target and reference objects' widths and depths, and describes the maximum allowed distance for the relation *behind-of*.

$$\hat{d}_{\text{behind-of}}(t,s) = s_w + t_w + s_h + t_h \tag{2.55}$$

The constraint $\Lambda^{a_1}_{\text{behind-of}}(t,s,c)$ is satisfied if the x-axis value of the target objects is greater than the x-axis value of the reference object.

$$\Lambda^{a_1}_{\text{behind-of}}(t,s,c) : t_{\text{behind-of}}(c)_x > s_{\text{behind-of}}(c)_x \tag{2.56}$$

The second condition $\Lambda^{a_2}_{\text{behind-of}}(t,s,c)$ requires that the maximum allowed distance is greater than the value obtained from the Euclidean distance between the CoG of the target and reference objects.

$$\Lambda^{a_2}_{\text{behind-of}}(t,s,c) : |t_p - s_p| < \hat{d}_{\text{behind-of}}(t,s) \tag{2.57}$$

The relation $\Lambda^{a}_{\text{behind-of}}(t,s,c)$ includes all conditions defined previously, which must be satisfied for the behind relation to be valid.

$$\Lambda^{a}_{\text{behind-of}}(t,s,c) : \Lambda^{a_1}_{\text{behind-of}}(t,s,c) \wedge \Lambda^{a_2}_{\text{behind-of}}(t,s,c) \tag{2.58}$$

For the *behind-of* relation, the objects' widths and depths are considered and specify the maximum allowed distance $\hat{d}_{\text{behind-of}}(t,s)$ 2.55 between these objects. According to Def. 11, this maximum allowed distance must be higher than the distance between them. Similar to previous projective relations, the main term is the view from which the objects are observed. For a target object to be *behind-of* the reference object, its value in the x-axis direction must be greater than the value in the x-axis of the reference objects. This condition is based on the right-handed coordinate system and refers to the Axiom 1. The *behind-of* relation is asymmetric Def. 4-Eq. 2.5, irreflexive Def. 4-Eq. 2.7, and converse Def. 4-Eq. 2.8 to the relation *in-front-of*.

2.2.3 Learning of the PQSR

The PQSR specifies how probable two objects are in a certain relation. According to Def.5-Def.11 and to calculate the probability for a given binary or projective binary relation, the typical distance between two object classes must be considered. In an environment, there are strong correlations between objects and regularities referring to the object's positions. In turn, such spatial context enable assumptions to be made about typical object co-occurrences. In this work, spatial knowledge is captured by calculating an average distance between two objects' classes for a particular relation. Therefore, the distance refers to statistical knowledge about typical spatial information between an object's class pair.

One of the most important reasons for using such learned distance is that objects can be located in varying distances from each other depending on the environment. By applying this knowledge, the probabilities of the relations can be calculated more robustly by using only the actual distance value. This calculation is performed on the assumption that the distance between object pairs differs depending on the object classes, even if the given relation holds. For instance, a computer mouse can be located 0.2 m *right-of* of a keyboard, but a (free-standing) house can be located about 6.0 m *right-of* of other house. In both cases, the relation *right-of* holds, but refers to a different range. Therefore, the assumption for this approach states that the distances are bound to a relation type and an object pair. The learned distance can be considered mathematically as an average or expected value for the distance between particular object classes. This distance information is then used to learn the co-occurrences probability, which can be gathered by applying the formulas from Def. 5-Def. 11. In turn, probabilities of the PQSR specify how probable the given relation between object pair is under consideration of the learned average value and can be viewed as a deviation from the expected value. The closer the distance between the objects' instances is to the learned distance, the more probable it is that the given relation holds.

Although the learning of specific knowledge might provide the impression that the algorithm can be over-trained and works only in already known environments, this is not the case. Due to the probabilistic interpretation, the algorithm is more robust than the non-probabilistic approaches discussed in Section 3. Using this method, the objects can still be in a certain relation (but with less probability) even if the resulting value deviates

from the expected value. In this way, the approach can detect atypical arrangements of objects with reduced probability.

Furthermore, the learned distance between objects and the statistical value denoting how often these objects have been seen in a particular relation influence the resulting learned probability. Moreover, the learned probability value may change depending on the given scene. Therefore, it is desirable to consider such variations. For example, object arrangement in an environment of a right-handed person may differ from the that of left-handed person. In general, it can be said that object co-occurrences and arrangements depend on the user's preferences. By learning such information, the PQSR can be adapted with respect to a given environment. A robotics system acting in a certain environment could gather such information over time and learn the spatial context information typical for this environment. Although this knowledge could also by provided manually, as demonstrated in work from [TBK+15], it would be difficult to predefine the entire scene and possible object arrangements because human environments are highly dynamic and may changed over time. Even if the expert could cover the most probable cases, this would be still a tedious and error-prone process. So, to be more precise and less biased (such as is the case when knowledge information is provided by an expert), a learning approach is developed and used in this thesis.

In this work, statistical knowledge about different geometric configurations of objects and the probability of certain relations between them is extracted from real-world data. The registered data contain scenes of office environments including various object classes such as tables, monitors, keyboards, etc. In the scenes, several arrangement variations of office objects are represented. Each scene might contain more or no instances of the known object classes.

In the following section, the learning process of the statistical knowledge regarding the typical distances between objects and excepted probability values is presented. First, the definition of a statistical co-occurrence probability between two objects is explained. This information is followed by definitions of the learned distances including the distance formulas for certain spatial relations.

2.2.3.1 Learning co-occurrence probability

In this thesis, seven spatial relations such as *on*, *above*, *right-of*, *left-of*, *near*, *in-front-of*, *behind-of* are learned. The learning process commences with the search for the instances of known object classes in the previously labeled data. These data serves as the main input for the knowledge learning process and build the knowledge model for all further calculations. In the following step, the distance between each object's instances is calculated from the data with regard to the distance function of a given spatial relation. These distances are then used to calculate the statistical co-occurrence probability.

In the world model of a given environment, the statistical knowledge about the PQSR is represented by the functions f_β 2.59 and f_τ 2.60. These functions specify the probability value for an object being in a certain relation with another object considering the typical distance between these objects' classes. Although this knowledge is obtained based on the object's instances the data contains, the probability value denotes the likelihood of finding two object classes in a given spatial relation. Therefore, this learned knowledge provides general information because it refers to object classes and not instances.

In any singular scene, particular object instances are presented. However, because all scenes and thus the sum of the object's instances are considered in the learned process, the resulting value refers to an object's class. In this way, the evidence about which objects co-occur can be derived from all instances.

Definition 12 *Let $\tilde{\mathcal{D}} = \mathcal{L}$ be a finite set of possible object classes, $\tilde{t}, \tilde{s} \in \tilde{\mathcal{D}}$ the target and reference objects' class, $t, s \in \mathcal{D}$ the instances of the given object's classes, $c \in \mathcal{C}$ the reference coordinate system, and $r \in \mathcal{R}$ a certain spatial relation. Then, the statistical co-occurrence probability for the binary f_β and projective binary relation f_τ is provided as:*

$$f_\beta : \mathcal{R} \times \tilde{\mathcal{D}}^2 \rightarrow [0;1] \tag{2.59}$$

$$f_\tau : \mathcal{R} \times \tilde{\mathcal{D}}^2 \times \mathcal{C} \rightarrow [0;1] \tag{2.60}$$

and calculated by:

$$f_\beta(r, \tilde{t}, \tilde{s}) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} r(t,s)}{\sum_{t \in \mathcal{D}[\tilde{t}]} 1} \tag{2.61}$$

$$f_\tau(r, \tilde{t}, \tilde{s}, c) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} r(t,s,c)}{\sum_{t \in \mathcal{D}[\tilde{t}]} 1} \tag{2.62}$$

To calculate the statistical co-occurrence probability, the sum of the occurrences where the given relation holds between the reference and target objects' instances is divided by the number of occurrences of the target object's instances. The division by the occurrences of the target object's builds the statistical value for the object's relation. If it is assumed that each relation holds with 100% probability, the sum of the probabilities must by divided by the number of occurrences of the target object to obtain correct results. Example 4 illustrates the calculation of the co-occurrence probability. Furthermore, depending of the number of target object occurrences, the learned result can, but does not necessarily have to, be symmetric. That is, the learned value refers to object classes and not instances. According to Def. 4-Eq. 2.4, the given spatial relation between two objects is symmetric if the symmetry constraint is satisfied. This rule is correct because the symmetry property refers to object instances and not classes.

Example 4 *Consider three scenes of a kitchen-like environment. In the scenes, there are three object instances of the object's classes:* `kitchen table (tb)`, `refrigerator (rf)`, *and* `cereal box (cb)`. *In the first scene, the cereal box is located* **`near`** *the fridge with a probability of 0.6 (60%), in the second scene* **`on`** *the kitchen table with 0.8 (80%), and in the last scene,* **`near`** *the fridge with a 0.9 (90%) probability. Given the defined formalism and co-occurrence probability (Def. 12), the following information about the probabilistic relations for the target object* `cereal box` *is obtained:*

$$f_\beta(near, \tilde{cb}, \tilde{rf}) = \frac{0.6 + 0.9}{3} = 0.5 \ (50\%) \tag{2.63}$$

$$f_\beta(on, \tilde{cb}, \tilde{tb}) = \frac{0.8}{3} = 0.26\overline{6} \ (26.\overline{6}\%) \tag{2.64}$$

As described in Example 4, the probability of finding an object class *cereal box near* an object class *fridge* is 50% and the probability of finding an object class *cereal box on* an object class *kitchen table* is 26%. Calculation of the co-occurrence probability according to Def. 12 provide these values. In the first scene, the spatial relation *near* holds with 60% probability and in the third scene with 90% for the object instances cereal box and fridge. Therefore, the resulting average probability of finding any cereal box *near* any fridge denotes 50% probability. Because three instances of the cereal box exist in all scenes, the resulting probability value is divided by three, even if only the cereal box was in the *near* relation with the fridge in two cases. Although this probability refers to those two object classes, it does not mean that the value is symmetric and the invert probability of finding a fridge near a cereal box must also be 50%. Importantly, and as discussed previously, the co-occurrence probability refers to the object classes ($\tilde{cb}$ and $\tilde{rf}$) and not instances (cb and rf).

2.2.3.2 Learning average distances

As discussed in the previous section, to obtain the statistical probability value for a probabilistic spatial relation, the average distance is considered. The distance denotes the expected distance between two given object classes and relations. In Section 2.2, definitions of the PQSR are provided, and in these definitions, the learned distance has been considered.

In this chapter, the method for learning the distance value is presented. Importantly, the calculation of the distance differs with respect to a given spatial relation. The final distance value of the binary $\ell_r(\tilde{t}, \tilde{s})$ Def. 13, and projective binary $\ell_r(\tilde{t}, \tilde{s}, c)$ Def. 14 relation results from all single distance calculations d_r of all known object instances given a particular relation.

This value denotes an average distance between two object classes in a certain relation. In this calculation, only the valid distance values are taken into account. The distance value is valid if the constraints for the particular relation, such as the maximum allowed distance, are satisfied.

Definition 13 *Let $r \in \mathcal{R}$ be a binary spatial relation type, $\tilde{t}, \tilde{s} \in \tilde{\mathcal{D}}$, a target, and a reference object's classes. Then, the average distance value $\ell_r(\tilde{t}, \tilde{s})$ for two object classes and the binary spatial relation is calculated, as follows:*

$$\ell_r(\tilde{t}, \tilde{s}) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} \begin{cases} \ell_{d_r}(t, s), & \text{if } \ell_{d_r}(t, s) \neq -1 \\ 0, & \text{otherwise} \end{cases}}{\nu_r(\tilde{t}, \tilde{s})} \tag{2.65}$$

ν_r denotes the number of valid distance values between both object's classes provided by the sum of the valid values.

$$\nu_r(\tilde{t}, \tilde{s}) = \sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} \begin{cases} 1, & \text{if } \ell_{d_r}(t, s) \neq -1 \\ 0, & \text{otherwise} \end{cases} \tag{2.66}$$

Definition 14 *Again, let $r \in \mathcal{R}$ be a binary spatial relation type, $\tilde{t}, \tilde{s} \in \tilde{\mathcal{D}}$, a target, and a reference object's class. Then, the average distance value $\ell_r(\tilde{t}, \tilde{s}, c)$ for a projective spatial relation is calculated in consideration of the reference coordinate system $c \in \mathcal{C}$.*

$$\ell_r(\tilde{t}, \tilde{s}, c) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} \begin{cases} \ell_{d_r}(t, s, c), & \text{if } \ell_{d_r}(t, s, c) \neq -1 \\ 0, & \text{otherwise} \end{cases}}{\nu_r(\tilde{t}, \tilde{s}, c)} \tag{2.67}$$

ν_r denotes the number of valid distance values between both object classes provided by the sum of the valid values.

$$\nu_r(\tilde{t}, \tilde{s}, c) = \sum_{t \in \mathcal{D}[\tilde{t}]} \sum_{s \in \mathcal{D}[\tilde{s}]} \begin{cases} 1, & \text{if } \ell_{d_r}(t, s, c) \neq -1 \\ 0, & \text{otherwise} \end{cases} \tag{2.68}$$

The $\ell_r(\tilde{t}, \tilde{s})$ 2.65 and $\ell_r(\tilde{t}, \tilde{s}, c)$ 2.67 result from the sum of all valid distances $d_r(t, s)$ i.e. $d_r(t, s, c)$ divided by the number of valid distance values $\nu_r(\tilde{t}, \tilde{s})$ i.e. $\nu_r(\tilde{t}, \tilde{s}, c)$. A single distance $\ell_{d_r}(t, s)$, $\ell_{d_r}(t, s, c)$ for a given spatial relation is valid, if its value differs from -1. This is because, the value -1 specifies that the distance between two object's classes is greater than the allowed distance or the relation between two objects does not hold at all.

Learning object size For the statistical probabilities $f_\beta(r, \tilde{t}, \tilde{s})$, $f_\tau(r, \tilde{t}, and \tilde{s}, c)$ and the average distance between object classes $\ell_r(\tilde{t}, \tilde{s})$, $\ell_r(\tilde{t}, \tilde{s}, c)$, the width and depth of both objects must be provided [GH14a]. Because only the sizes of the reference object instances are obtained during the learning process, a *virtual* target object is defined. However, the width and depth of the virtual object must still be defined. This value could be, for example, provided manually by an expert user, but (as opposed to a learned knowledge) this would be a particularly tedious process. Furthermore, one cannot consider all possible sizes of the object's classes in the domain. To overcome this issue in the present work, the typical target object widths and depths are learned from the collected data.

Definition 15 *Let $\tilde{t} \in \tilde{\mathcal{D}}$ be a given target object class. Then, the learned width ℓ_w and depth ℓ_h of a* `virtual` *target object are calculated, as follows:*

$$\ell_w(\tilde{t}) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} t_w}{\sum_{t \in \mathcal{D}[\tilde{t}]} 1} \tag{2.69}$$

$$\ell_h(\tilde{t}) = \frac{\sum_{t \in \mathcal{D}[\tilde{t}]} t_h}{\sum_{t \in \mathcal{D}[\tilde{t}]} 1} \tag{2.70}$$

As discussed previously, the learned average distance has been considered in the calculation of each PQSR because a spatial relation refers to space within it that is valid. Depending on the object's class, the distance for a given relation can change. Furthermore, the learned average distance denotes the typical area for a relation in which an object is located. Thereby, the distance specifies the position where the object can typically be found with respect to the given relation. For example, a mouse is typically located *near* a

keyboard but not necessarily directly next to keyboard. Thus, the value indicates where the target object could be located relative to the reference object.

In the following paragraph, the definitions for learning distances $d_r(t,s)$ and $d_r(t,s,c)$ between two object instances and a given spatial relations are provided. The sum of these valid distances is then used to calculate the average distance $\ell_r(\tilde{t},\tilde{s})$, $\ell_r(\tilde{t},\tilde{s},c)$ for a given PQSR.

Distance for the spatial relation near

Definition 16 *Let $t,d \in \mathcal{D}$ be the instances of the target and reference object's class. Then, the distance between these two object instances for the spatial relation* `near` *is provided by:*

$$\ell_{d_{near}}(t,s) = \begin{cases} d_{near}(t,s), & \text{if } d_{near}(t,s) < \hat{d}_{near}(t,s) \\ -1, & \text{otherwise} \end{cases} \tag{2.71}$$

In the formula $\ell_{d_{near}}(t,s)$, the term $d_{\text{near}}(t,s)$ refers to the Euclidean distance between the target and reference objects, which is analogous to the calculation of $d_{\text{near}}(t,s)$ 2.11 but without the subtraction of the $\ell_{d_{near}}(t,s)$ and is not considered as an absolute value. $\hat{d}_{\text{near}}(t,s)$ denotes the maximum allowed distance and can be obtained from the formula 2.12. According to $\ell_{d_{\text{near}}}(t,s)$ 2.71, the resulting distance must be smaller than $\hat{d}(t,s)$.

Distance for the spatial relation above

Definition 17 *Let $t,s \in \mathcal{D}$ be the instances of the target and reference objects' classes. Then, the distance for the relation* `above` *between two object instances is calculated, as follows:*

$$\ell_{d_{above}}(t,s) = \begin{cases} d_{above}(t,s), & \text{if } \Lambda^a_{above}(t,s) \wedge d_{above}(t,s) < \hat{d}_{above}(t,s) \\ -1, & \text{otherwise} \end{cases} \tag{2.72}$$

The distance $d_{\text{above}}(t,s)$ between a target and reference object is calculated similar to the distance of the above relation $d_{\text{above}}(t,s)$ 2.15 but does not include the learned distance $\ell_{d_{\text{above}}}(t,s)$. The constraints that must be satisfied for the distance calculation are the same as the constraints for the relation *above* $\Lambda^a_{\text{above}}(t,s)$ 2.19. To obtain a valid distance value, the actual distance $d_{\text{above}}(t,s)$ must be smaller than the maximal allowed distance $\hat{d}_{\text{above}}(t,s)$.

Distance for the spatial relation on

Definition 18 *Let $t,s \in \mathcal{D}$ be instances of the target and reference objects' classes. Then, the distance for the relation* `on` *between these two object instances is calculated, as follows:*

$$\ell_{d_{on}}(t,s) = \begin{cases} d_{on}(t,s), & \text{if } \Lambda^a_{on}(t,s) \wedge d_{on}(t,s) < \lambda \\ -1, & \text{otherwise} \end{cases} \tag{2.73}$$

To obtain the learned distance for the *on* relation, the $d_{\text{on}}(t,s)$ must be known. This value is calculated in a similar way to $d^{\Delta}_{\text{on}}(t,s)$ 2.21 of the *on* relation. The difference in this calculation is that the value is not divided by the λ threshold and the $\ell_{d_{\text{on}}}(t,s)$ is not subtracted. The λ is equal to the λ value of the *on* Def. 7 relation. The constraint $\Lambda^{a}_{\text{on}}(t,s)$ denotes the same constraints as for the *on* relation $\Lambda^{a}_{\text{on}}(t,s)$ 2.26.

Distance for the projective spatial relation left-of

Definition 19 *Let $t,s \in \mathcal{D}$ be instances of a target and reference objects' classes and $c \in \mathcal{C}$ a reference coordinate system. Then, the distance for the projective relation* `left-of` *between two object instances is calculated, as follows:*

$$\ell_{d_{\textit{left-of}}}(t,s,c) = \begin{cases} d_{\textit{left-of}}(t,s,c), & \textit{if } \Lambda^{a}_{\textit{left-of}}(t,s,c) \wedge d_{\textit{left-of}}(t,s,c) < \hat{d}_{\textit{left-of}}(t,s) \\ -1, & \textit{otherwise.} \end{cases} \tag{2.74}$$

In this formula, $d_{\text{left-of}}(t,s,c)$ is calculated in the same way as in formula $d^{\Delta}_{\text{left-of}}(t,s,c)$ 2.28 for the relation *left-of* but does not include subtraction of $\ell_{d_{\text{left-of}}}$ and is not an absolute value. The $\hat{d}_{\text{left-of}}(t,s)$ is calculated similarly to $\hat{d}_{\text{left-of}}(t,s)$ 2.29. For the $\ell_{d_{\text{left-of}}}(t,s,c)$ 2.74 calculation, the same constraints $\Lambda^{a}_{\text{left-of}}(t,s,c)$ as for the *left-of* relation $\Lambda^{a}_{\text{left-of}}(t,s,c)$ 2.34 must be satisfied.

Distance for the projective spatial relation right-of

Definition 20 *Let $t,s \in \mathcal{D}$ be instances of target and reference objects' classes and $c \in \mathcal{C}$ a reference coordinate system. Then, the distance for the projective relation* `right-of` *between two object instances is calculated, as follows:*

$$\ell_{d_{\textit{right-of}}}(t,s,c) = \begin{cases} d_{\textit{right-of}}(t,s,c), & \textit{if } \Lambda^{a}_{\textit{right-of}}(t,s,c) \wedge d_{\textit{right-of}}(t,s,c) < \hat{d}_{\textit{right-of}}(t,s) \\ -1, & \textit{otherwise.} \end{cases} \tag{2.75}$$

The distance $d_{\text{right-of}}(t,s,c)$ of the *right-of* relation is calculated in a similar way to $d^{\Delta}_{\text{right-of}}(t,s,c)$ 2.36. In this formula, the distance between the target and reference object in the direction of the y-axis is considered, the subtraction of $\ell_{d_{\text{right-of}}}(t,s,c)$ does not occur, and the result is not an absolute value. The $\hat{d}_{\text{right-of}}(t,s)$ 2.75 is calculated like the $\hat{d}_{\text{right-of}}(t,s)$ 2.39.The $\Lambda^{a}_{\text{right-of}}(t,s,c)$ refers to the constraints $\Lambda^{a}_{\text{right-of}}(t,s,c)$ 2.42 defined for the *right-of* relation.

Distance for the projective spatial relation in-front-of

Definition 21 *Let $t,s \in \mathcal{D}$ be instances of target and reference object classes and $c \in \mathcal{C}$ a reference coordinate system. Then, the distance for the projective relation* `in-front-of` *between two object instances is calculated, as follows:*

$$\ell_{d_{\textit{in-front-of}}}(t,s,c) = \begin{cases} d_{\textit{in-front-of}}(t,s,c), & \textit{if } \Lambda^{a}_{\textit{in-front-of}}(t,s,c) \wedge \\ & d_{\textit{in-front-of}}(t,s,c) < \hat{d}_{\textit{in-front-of}}(t,s) \\ -1, & \textit{otherwise.} \end{cases} \tag{2.76}$$

The distance $d_{\text{in-front-of}}(t,s,c)$ for the *in-front-of* relation is calculated in a way similar way to $d^{\Delta}_{\text{in-front-of}}(t,s,c)$ 2.44. The difference in the calculation of $d_{\text{in-front-of}}(t,s,c)$ 2.76 is the subtraction of $\ell_{d_{\text{in-front-of}}}(t,s,c)$ and omitted from the absolute value. $\hat{d}_{\text{in-front-of}}(t,s)$ is calculated like the $\hat{d}_{\text{in-front-of}}(t,s)$ 2.45, and $\Lambda^{a}_{\text{in-front-of}}(t,s,c)$ refers to the constraints $\Lambda^{a}_{\text{in-front-of}}(t,s,c)$ 2.50 defined for the *right-of* relation.

Distance for the projective spatial relation behind-of

Definition 22 *Let $t,s \in \mathcal{D}$ be instances of target and reference object classes and $c \in \mathcal{C}$ a reference coordinate system. Then, the distance for the projective relation* `behind-of` *between these two object instances is provided by:*

$$\ell_{d_{behind\text{-}of}}(t,s,c) = \begin{cases} d_{behind\text{-}of}(t,s,c), & \text{if } \Lambda^{a}_{behind\text{-}of}(t,s,c) \wedge \\ & d_{behind\text{-}of}(t,s,c) < \hat{d}_{behind\text{-}of}(t,s) \\ -1, & \text{otherwise.} \end{cases} \tag{2.77}$$

The distance $d_{\text{behind-of}}(t,s,c)$ of the *behind-of* relation is calculated similarly to $d^{\Delta}_{\text{behind-of}}(t,s,c)$ 2.52. The difference in the calculation of $d_{\text{behind-of}}(t,s,c)$ is the subtraction of $\ell_{d_{\text{behind-of}}}(t,s,c)$ without the absolute value. The $\hat{d}_{\text{behind-of}}(t,s)$ is calculated like the $\hat{d}_{\text{behind-of}}(t,s)$ 2.55 and the $\Lambda^{a}_{\text{behind-of}}$ refers to the constraints $\Lambda^{a}_{\text{behind-of}}$ 2.50 defined for the *behind-of* relation.

2.3 Spatial Potential Fields

In the previous Section 2.2, the formalism of the PQSR was presented. Based on this formalism, the probability value for given objects and relations can be obtained. Furthermore, the statistical knowledge about typical object co-occurrences can be learned from the environments as described in Section 2.2.3.

In this section, a new representation form, SPF, is introduced. This form combines the learned knowledge and the probability values of the PQSR to provide information about how likely it is that a given relation holds between two object instances at a certain position in the scene with respect to the learned knowledge. Thereby, the SPF are calculated for each spatial relation and with consideration to all the reference object's instances present in the scene. As a result, a 3D model of a given PQSR in the scene is calculated. The SPF are then used in subsequent steps of the algorithm to determine the most probable positions of the target object.

According to the definitions Def. 2 and Def. 3 of probabilistic spatial relations, an object is in a given spatial relationship to other spatial object with a certain probability. This probability value specifies how probable this relation holds between those objects. Specifically, the probability value denotes how strong a relation holds. Depending on the distance between the objects and their corresponding constraints, this value may vary from 0 to 100%. For calculations of the probability for a given relation, two objects and their positions in space are considered. As previously discussed, the definitions Def. 5-Def. 11 are used for calculation of the PQSR and the Def. 13-Def. 14 for learning of these

relations. The main difference between these definitions is that the learning relations refer to the statistical probability (the co-occurrence) of finding two objects' classes in a particular relation. In contrast, the probability value resulting from the definition of the given relation describes how strongly the relation holds between two object instances.

To estimate the overall probability for two objects and a certain relation, the probability for a given relation between two object instances and the statistical probability from the learned knowledge for the object's classes are combined for an *overall* probability. This combination is performed because the assumption is that the actual arrangement between the observed object's instances is relevant for the probability of it being in a particular relation, and that the general knowledge about how likely the objects are to co-occur is important and must be considered.

The SPF are developed to achieve this combination. A spatial potential field is a representation of a given PQSR in a qualitative and 3D manner, and specifies the probability of an object to be in a given relation with another object by considering the statistical co-occurrence probability gathered from the extracted knowledge. Each PQSR has a corresponding SPF, and these SPF are used to represent the PQSR in the environment, for example, the given domain in which the robot is operating.

The concept of SPF [GH14b] was inspired by the field method, which has been used in the areas such as robots navigation [BK91] for obstacle avoidance. In this field, the method Vector Filed Histogram (VFH) is used for the detection of obstacles and avoiding collisions by a mobile robot while it performs a task. The field histogram is calculated from the detected objects of a given map of the environment, and the magnitude of a field represents the distance of the objects to the robot. The calculated potential fields are used for avoiding obstacles in the robot's pathway. In contrast, SPF are used to guide the robot to the position at which the object can most likely be found.

2.3.1 Definition of the SPF

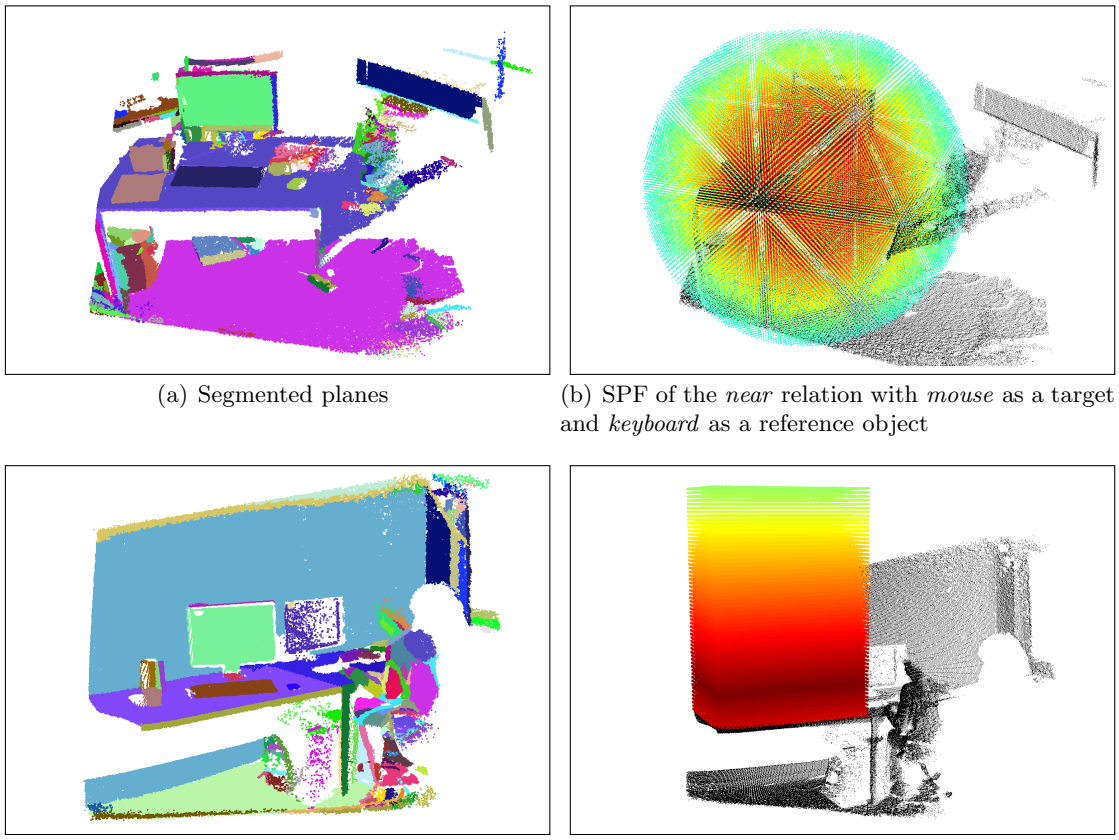
SPF are generated by applying Def. 5-Def. 11 for PQSR calculation. While calculating the SPF, knowledge extracted from real-world data, as described in Section 2.2.3, is used. To obtain the probability value within the SPF, the intensity of the relation and probability values gathered from the extracted knowledge are combined. The calculation of the SPF is performed for certain objects and a relation.

Definition 23 *Let $\tilde{t}, \tilde{s} \in \tilde{\mathcal{D}}$ be the target and reference object classes, $t, s \in \mathcal{D}$ the instances of those object classes, $r \in \mathcal{R}$ a given relation, and $c \in \mathcal{C}$ a view point of the system. Then, the SPF for a binary SPF_β and projective binary SPF_τ relations are defined, as follows:*

$$SPF_\beta(r, t[\tilde{t}_l], s[\tilde{s}_l]) = f_\beta(r, \tilde{t}, \tilde{s}) \cdot r(t, s) \tag{2.78}$$

$$SPF_\tau(r, t[\tilde{t}_l], s[\tilde{s}], c = f_\tau(r, \tilde{t}, \tilde{s}, c) \cdot r(t, s, c) \tag{2.79}$$

In the formulas 2.78 and 2.79, the $f_\beta(r, t, s)$, $f_\tau(r, t, s, c)$ values denote the statistical co-occurrence probability provided by Def. 2.59 and Def. 2.60. The values $r(t, s)$ and $r(t, s, c)$ refer to the probability for a relation at a given position in the scene, and are calculated by applying the formulas of the PQSR calculations as described in Section 2.2.



(a) Segmented planes

(b) SPF of the *near* relation with *mouse* as a target and *keyboard* as a reference object

(c) Segmented planes

(d) SPF of the *above* relation with *keyboard* as a target and *table* as a reference object

Figure 2.17: SPF of the relations *near* and *above*.

SPF have different forms depending on the relation type they represent. Figures 2.17-2.19 provide a visualization of the corresponding SPF. In these figures, the colors of the fields denote the scale of their intensity, with red being a higher and blue being a lower probable value.

2.3.2 Calculating the SPF in a real scene

In Section 2.3.1, the definition of the SPF was presented and the general concept behind this representation form was explained. This section focuses on the calculation of the SPF in a scene gathered from a real environment. As previously discussed, one SPF specifies the probability for a given spatial relation in consideration of the learned co-occurrence probability for this particular relation within a given scene. This, in turn, enables definition of how probable two objects are in a spatial relation by considering their general co-occurrence.

Because each SPF is calculated for two objects' instances, i.e., the target and reference object, to calculate the SPF in the environment, these objects must be known in advanced.

Although the data already contain labeled reference objects, the information about the target object is missing but still required. In fact, by searching for a target object, this information is not available initially. Nevertheless, for the probability calculation, both the position and size of the target object are required. Because the target object is the sought after object in the scene, this information must be assumed.

To provide a potential target object, two assumptions about the object are considered. First, a target object can, theoretically, be located anywhere in the environment. Second, its size depends on the class to which the object belongs. Regarding the size of the object, the learned knowledge about the object's typical width and depth can be used. To handle the missing position information, a *grid* of a scene is calculated and used. This grid divides a certain scene in fixed pieces, which are termed *cells*. Each cell contains an *index*, based on which the position of each cell can be calculated. By determining this position, it can be estimated that the target object is located on each known position, that is, a cell in the grid of the given scene. In this way, the position of the target object can be assumed as the current position of the cell. In this context, the SPF value is calculated for each cell of the grid. The number of the cells depend on the resolution of the grid, and this resolution can be altered according to its purpose. Figure 2.20 provides a grid of an exemplary scene.

Definition 24 *Let $e \in \mathbb{R}$ be a single cell of a grid and $k \in \mathbb{N}_+$ an index of the given cell. Then, the grid Θ is given by:*

$$\Theta = [e]^k \tag{2.80}$$

To calculate the SPF at the given cell e with the index k, the position of the cell must be known. The functions 2.81, 2.82, and 2.83 are used to calculate the position of a grid's cell.

Definition 25 *Let $\rho \in \mathbb{R}^3$ be the size of the grid Θ, $\omega \in \mathbb{R}^3$ the resolution of the grid, and k denoted the given index of a cell e. Then, the position of the cell e at the index k is calculated by the following function:*

$$f_x(k) = (k \bmod (\rho_x/\omega_x)) \cdot \omega_x \tag{2.81}$$

$$f_y(k) = \lfloor (k/(\rho_x/\omega_x)) \rfloor \bmod ((\rho_y/\omega_y)) \cdot \omega_y \tag{2.82}$$

$$f_z(k) = \lfloor k/\ ((\rho_x/\omega_x) \cdot (\rho_y/\omega_y)) \rfloor \cdot \omega_z \tag{2.83}$$

To obtain the x-position $f_x(k)$ of a cell, the x-index of the k-cell must first be known. The x-index results from modulo between the k-th cell's index and the number of cells in the x-dimension. Then, this x-index is multiplied with the grid's resolution in the x-dimension. In this way, the x-position of the k-th cell is obtained. The following example describes the calculation:

Example 5 *Consider a grid with a width in the x-dimension of $p_x = 10m$ and resolution in the x-dimension $\omega_x = 0.1m$. Then, the x-dimension contains $p_x/\omega_x = 10m/0.1m = 100$ cells. For a cell with, for example, the index $k = 78430$ according to the formula 2.81, the x-position is calculated as follows:*

$$78430 \bmod 100 = 30 \tag{2.84}$$

$$30m \cdot 0.1m = 3m \tag{2.85}$$

Calculation of the y-position $f_y(k)$ of the cell is conducted in a similar way to the x-position. The x-position results from overflow between the k-th cell's index and the number of cells in the x-dimension. More precisely, the x-position is obtained by calculating how often the index k fits into the number of x-cells. For instance, in the Example 5, the x-dimension consists of 100 cells. Following this example, the index $k = 78430$ of the cell fits 784 ($\lfloor 78430/100 \rfloor$) times. Assuming the y-dimension consists, again, of 100 cells, the result of the modulo between the 784 and 100 ($784 \bmod (\rho_y/\omega_y)) = 784 \bmod 100 = 84$) would be 84 and denotes the y-index. Finally, to obtain the position in the y-dimension, this y-index 84 is multiplied by the resolution $84 \cdot \omega_y = 84 \cdot 0.1 = 8.4$. Therefore, the resulting value 8.4 is the y-position of the k-th cell.

To determine the z-position of the k-th cell, similarly to previous calculations of the x- and y-position, the z-index must be specified. This index can be obtained by the overflow of the x- and y-dimensions $|cells_x| \cdot |cells_y| = (\rho_x/\omega_x) \cdot (\rho_y/\omega_y)$ divided by the current index k. For example, the k-th cell's index is 78430 and the number of the cells in the x- and y-dimension is 100, then, the z-index of the k-th cell is calculated as follows: $\lfloor 78430/(100 \cdot 100) \rfloor = \lfloor 7.843 \rfloor = 7$. Following the multiplication of the z-index 7 with the z-resolution of the grid $\omega_z = 0.1$, $7 \cdot 0.1 = 0.7m$, the z-position 0.7 is determined.

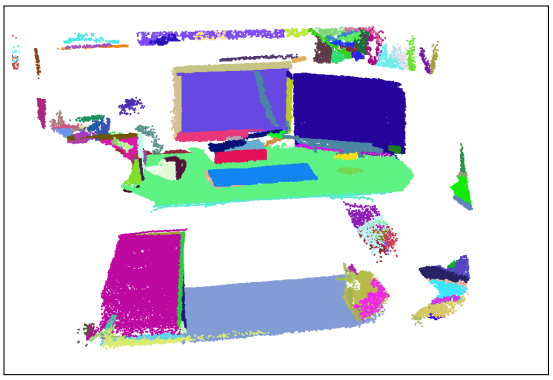
Finally, the coordinates of the given cell $k = 78430$ result in $[3, 8.4, 0.7]$. By determining the position of the cell, a virtual target object at the given position can be defined.

Definition 26 *Let k be the index of a given cell e, $\tilde{t} \in \tilde{\mathcal{D}}$ the target object class, and $x, y, z, w, h \in \mathbb{R}$ denoting the object's x-, y-, z-position, width, and depth, respectively. Then, the virtual target object t_v is defined, as follows:*

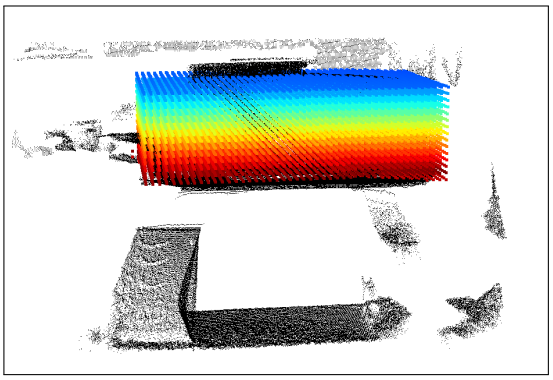
$$t_v = [x, y, z, w, h] \tag{2.86}$$

$$t_v(k, \tilde{t}) = [f_x(k), f_y(k), f_z(k), \ell_w(\tilde{t}), \ell_h(\tilde{t})] \tag{2.87}$$

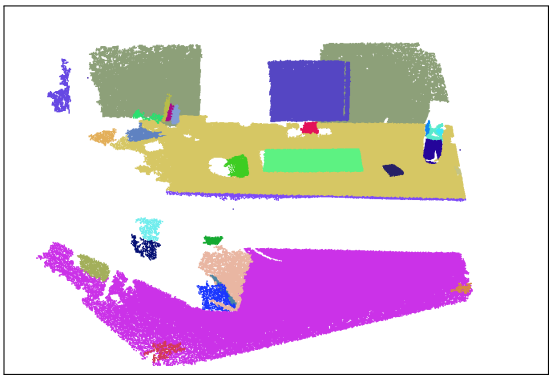
By applying the formula 2.87, the virtual object t_v can be represented. The position of this object corresponds to the position of the current cell $f_x(k)$. The width and depth of the virtual object are provided by the learned width $\ell_w(\tilde{t})$ and depth $\ell_h(\tilde{t})$ of the class to which the object belongs. With the virtual target object and reference objects from the considered scene, the SPF can be calculated.



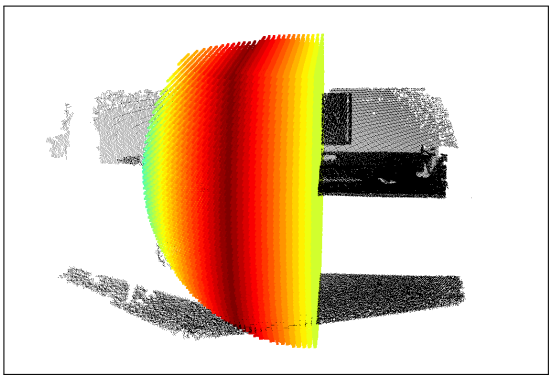
(a) Segmented planes



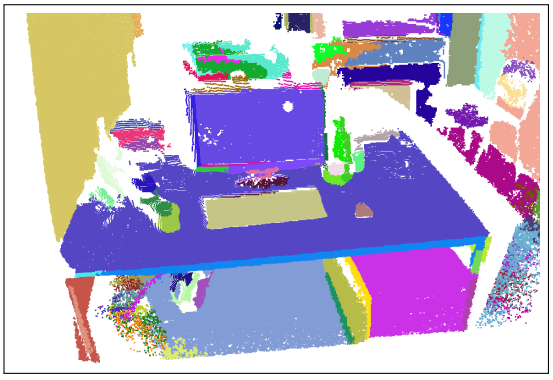
(b) SPF of the *on* relation with *keyboard* as a target and *table* as a reference object



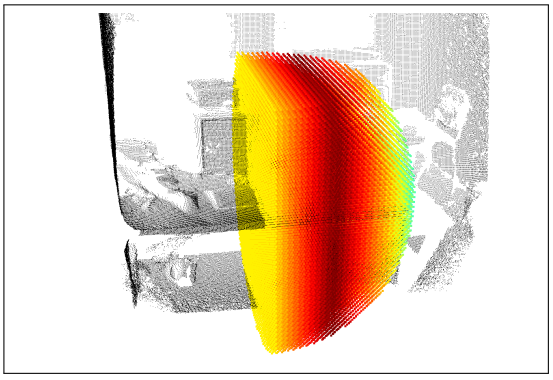
(c) Segmented planes



(d) SPF of the *left-of* relation with *mug* as a target and *keyboard* as a reference object

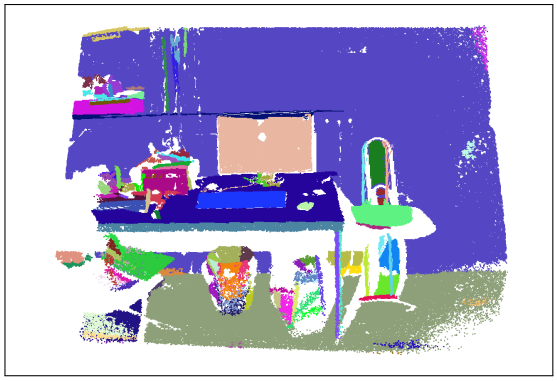


(e) Segmented planes

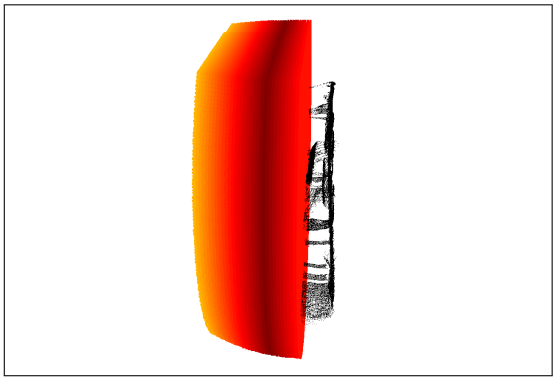


(f) SPF of the *right-of* relation with *mouse* as a target and *keyboard* as a reference object

Figure 2.18: SPF of the relation *on* and the projective *left-of*, *right-of* relations.

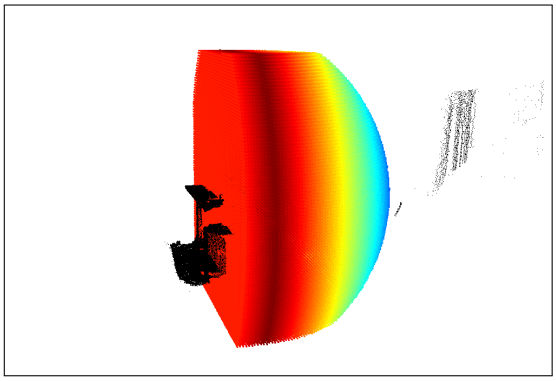


(a) Segmented planes

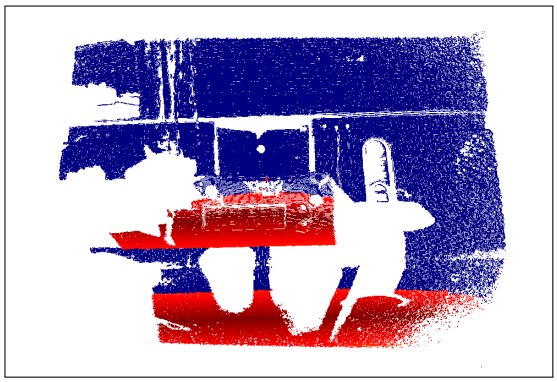


(b) SPF of the *in-front-of* relation with *keyboard* as a target and *monitor* as a reference object

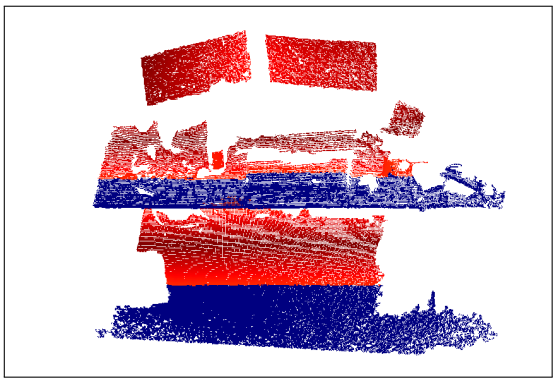
(c) Segmented planes



(d) SPF of the *behind-of* relation with *monitor* as a target and *keyboard* as a reference object



(e) 2D visualization of the SPF of the *in-front-of* relation



(f) 2D visualization of the SPF of the *behind-of* relation

Figure 2.19: SPF of the projective relations *in-front-of* and *behind-of*.

Definition 27 *Let $r \in \mathcal{R}$ be a given spatial relation, $\tilde{t}, \tilde{s} \in \tilde{\mathcal{D}}$ the target and reference objects' class, $s \in \mathcal{D}$ the reference object, e_k the current cell of the grid Θ with the index k, and $t_v(k, \tilde{t})$ the virtual object at this cell. Then, the SPF for a given relation and cell are calculated by:*

$$\forall e_k \in \Theta$$

$$SPF_\beta(r_\beta, t_v(k,\tilde{t}), \tilde{t}, s, \tilde{s}) = f_\beta(r_\beta, \tilde{t}, \tilde{s}) \cdot r_\beta(t_v(k,\tilde{t}), s) \tag{2.88}$$

$$SPF_\tau(r_\tau, t_v(k,\tilde{t}), \tilde{t}, s, \tilde{s}, c) = f_\tau(r_\tau, \tilde{t}, \tilde{s}, c) \cdot r_\tau(t_v(k,\tilde{t}), s, c) \tag{2.89}$$

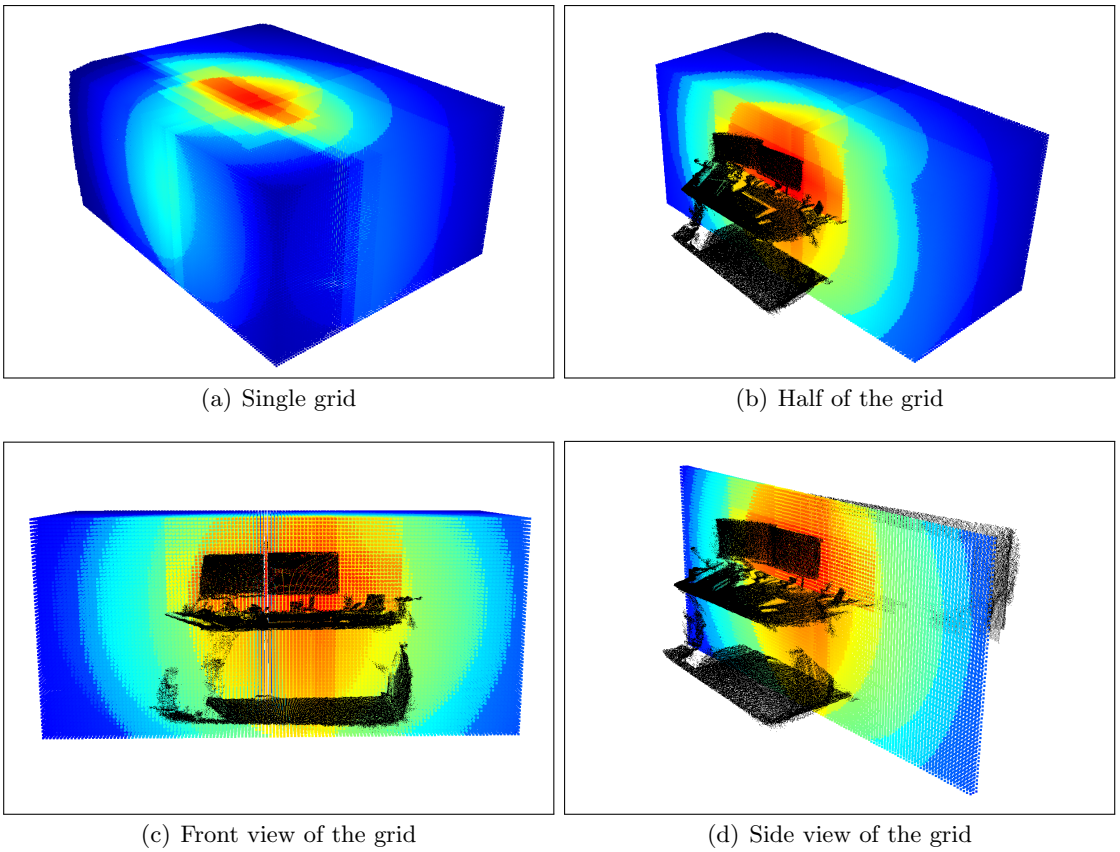
Due to the SPF calculation in a given grid Θ, each cell $e \in E$ contains a probability value at a given position within the current index. Each cell of the grid, or rather, its position, is used as a assumed position of the *virtual* target object. Therefore, the probability value for a given object at each position in the scene can be calculated. For the calculation, the labeled reference objects (preselected manually, as described in 2.2.3) are used as reference objects. As in the relation calculation, the size of a target object must be known, this information is obtained from the learned knowledge. According to Def. 15, the typical size of the possible object's classes is learned. This information serves as input for the calculation of the virtual object and relation probability.

The resulting probability value provided by the corresponding SPF specifies how likely the sought after object (referring to a virtual object) can be found at the given (cell-) position in the scene with respect to the considered reference object. However, this value refers only to one relation between the target and a certain reference object. As a given target object can be in different relations with several reference objects, a further step of the algorithm is performed in which the resulting SPF are combined to determine a final probability value.

2.4 Field Intensity

Based on the formalism provided in Section 2.2 and the definition of the statistical co-occurrence probability Def. 12, which specifies how to obtain this probability from real-world environments, a method that can be used as a heuristic to guide an object search based on PQSR is proposed. In this method, the generated knowledge is used to improve the object search process by providing estimations for the most probable object positions in a given scene.

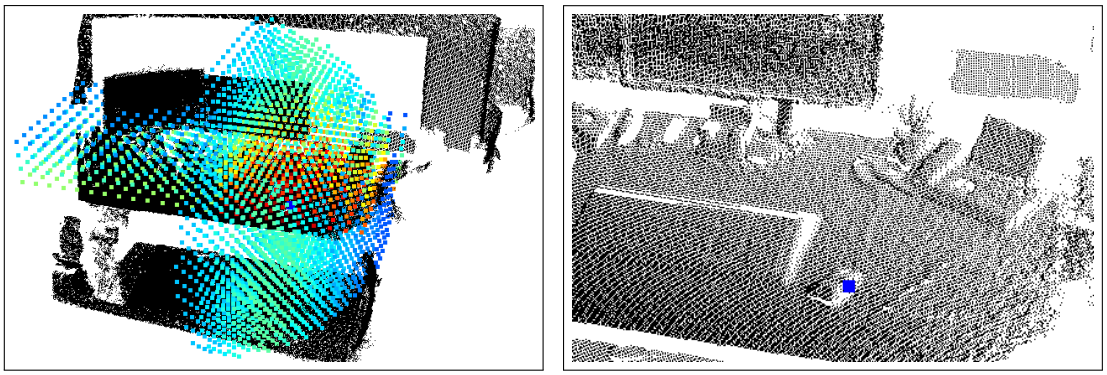
As discussed in Section 2.3, the SPF provide information about possible object positions considering *one* spatial relation between two given objects. A single SPF already contains the learned information about the object co-occurrences and probability of finding two given objects in a certain relation. In turn, the *Field Intensity* (FI) is a combination of all *Spatial Potential Fields* and is used to ultimately determine the most probable object location across all possible relations and reference objects in the scene. In contrast to the SPF, the FI is the result of the combination of all relations between all reference objects and the sought after object at the hypothetical position p, the position of the actual cell with the index k. Therefore, the FI value of any point denotes the probability that a searched object is located at this position given the previously learned knowledge. Figure 2.21 provides an exemplary FI.



(a) Single grid (b) Half of the grid

(c) Front view of the grid (d) Side view of the grid

Figure 2.20: Visualization of a grid used for SPF calculation with a grid resolution of 0.03 meters and the searched object (monitor). The grid cells containing the highest probability are red and cells with the smallest probability values are blue.



(a) SPF mouse *on* table and mouse *right-of* keyboard (b) The FI value for these two spatial relations

Figure 2.21: FI for two spatial relations *on* and *right-of*. The target object is a *mouse* and the reference objects are a *keyboard* and *table*.

Definition 28 *Let $\tilde{t} \in \tilde{\mathcal{D}}$ be the target object class, $\tilde{s} \in \tilde{\mathcal{D}}$ the reference object class, $s \in \mathcal{D}$ the instance of the reference object class, e_k the k-th cell of the grid, t_v the virtual target object, and $c \in \mathcal{C}$ a viewpoint of the system. Then, the FI of given SPF is calculated, as follows:*

$$FI(\tilde{t}, e_k, c) = \frac{FI_\beta(\tilde{t}, e_k) + FI_\tau(\tilde{t}, e_k, c)}{\Xi_\beta + \Xi_\tau} \tag{2.90}$$

$$FI_\tau(\tilde{t}, e_k, c) = \sum_{r_\tau \in \mathcal{R}} \sum_{s \in \mathcal{D}} SPF_\tau(r_\tau, t_v(k, \tilde{t}), \tilde{t}, s, \tilde{s}, c) \tag{2.91}$$

$$FI_\beta(\tilde{t}, e_k) = \sum_{r_\beta \in \mathcal{R}} \sum_{s \in D} SPF_\beta(r_\beta, t_v(k, \tilde{t}), \tilde{t}, s, \tilde{s}) \tag{2.92}$$

Definition 29 *Again, let $\tilde{t} \in \tilde{\mathcal{D}}$ be the target object class, $\tilde{s} \in \tilde{\mathcal{D}}$ the reference object class, $s \in \mathcal{D}$ the instance of the reference object class, f_β and f_τ the learned knowledge, and $c \in \mathcal{C}$ a viewpoint of the system. Then, the number of the valid relations Ξ is given by:*

$$\Xi_\beta(\tilde{t}) = \sum_{r_\beta \in \mathcal{R}} \sum_{s \in \mathcal{D}} \begin{cases} 1, & \textit{if } \ f_\beta(r_\beta, \tilde{t}, \tilde{s}) \neq 0 \\ 0, & \textit{otherwise} \end{cases} \tag{2.93}$$

$$\Xi_\tau(\tilde{t}, c) = \sum_{r_\tau \in \mathcal{R}} \sum_{s \in \mathcal{D}} \begin{cases} 1, & \textit{if } \ f_\tau(r_\tau, \tilde{t}, \tilde{s}, c) \neq 0 \\ 0, & \textit{otherwise} \end{cases} \tag{2.94}$$

The formula $FI(\tilde{t}, e_k, c)$ 2.90 denotes calculation of the overall Field Intensity. The resulting Field is the average intensity value obtained from the sum of all single Fields for binary FI_β and projective binary FI_τ relations. The FI_β and FI_τ sum all SPF values for binary $r_\beta \in \mathcal{R}$ and projective binary relations $r_\tau \in \mathcal{R}$ and reference objects $s \in \mathcal{D}$ from the scene. In this formula, the sum of the FI_β and FI_τ is divided by the number of the relations that are valid Ξ. The term *valid* means that a given relation exists between the reference $\tilde{s}$ and target object $\tilde{t}$ classes with respect to the learned knowledge. The target object is given by $t_v(k, \tilde{t})$ according to the Def. 26. The calculation of the number of the valid relations is determined by the formulas 2.93 and 2.94.

Example 6 *Consider the following scenario. In an office scene* `A`*, there are two reference object instances, as follows: an* `office desk` *and a* `monitor`*. The search object belongs to the object class* `mouse` *and can potentially be located anywhere in the scene. Therefore, the virtual target object (assumed to be a* `mouse`*) has the position of the cell* `e` *with the index* `k`*. Considering this information with respect to the learned knowledge f_β and f_τ, the following SPF are calculated:*

$$f_\beta(r_{near}, mouse, office\ desk) = 0.9 \tag{2.95}$$

$$f_\beta(r_{near}, mouse, monitor) = 0.8 \tag{2.96}$$

$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, office\ desk) = 0.9 \cdot 0.6 = 0.54 \tag{2.97}$$

$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, monitor) = 0.8 \cdot 0.7 = 0.56 \tag{2.98}$$

Then, the FI results in:

$$FI(mouse, e_k) = \frac{0.54 + 0.56}{2} = 0.55 \tag{2.99}$$

In this scenario, the probability of finding a mouse (given the spatial relation *near*, the learned knowledge, and two reference objects) is 55%.

Example 7 *Consider a second scenario. In an office scene* `B`*, there are three object instances, as follows: an* `office desk`*, a* `monitor`*, and a* `keyboard`*. Similar to Example 6, the search object is also a computer* `mouse`*. The virtual target object* `mouse` *has the position of the cell e with the index k. Considering this information and the learned knowledge f_β and f_τ, the following SPF are calculated:*

$$f_\beta(r_{near}, mouse, office\ desk) = 0.9 \tag{2.100}$$
$$f_\beta(r_{near}, mouse, monitor) = 0.8 \tag{2.101}$$
$$f_\tau(r_{right\text{-}of}, mouse, keyboard) = 0.8 \tag{2.102}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, office\ desk) = 0.9 \cdot 0.6 = 0.54 \tag{2.103}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, monitor) = 0.8 \cdot 0.7 = 0.56 \tag{2.104}$$
$$SPF_\tau(r_{right\text{-}of}, mouse(e_k, mouse), mouse, keyboard, c) = 0.8 \cdot 0.8 = 0.64 \tag{2.105}$$

Then, the FI results in:

$$FI(mouse, e_k) = \frac{0.54 + 0.56 + 0.64}{3} = 0.58 \tag{2.106}$$

In this example, the presence of a keyboard in the scene resulted in a 58% probability of finding a mouse at the position of the cell e, which is higher than the probability value in Example 6.

Example 8 *Consider a third scenario. In an office scene* `C`*, there are three object instances, as follows: an* `office desk`*, a* `monitor`*, and a* `keyboard`*. The virtual target object* `mouse` *has the position of the cell e with the index k. Considering this information and the learned knowledge f_β and f_τ, the following SPF are calculated:*

$$f_\beta(r_{near}, mouse, office\ desk) = 0.9 \tag{2.107}$$
$$f_\beta(r_{near}, mouse, monitor) = 0.8 \tag{2.108}$$
$$f_\tau(r_{right\text{-}of}, mouse, keyboard) = 0.8 \tag{2.109}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, office\ desk) = 0.9 \cdot 0.6 = 0.54 \tag{2.110}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, monitor) = 0.8 \cdot 0.7 = 0.56 \tag{2.111}$$
$$SPF_\tau(r_{right\text{-}of}, mouse(e_k, mouse), mouse, keyboard, c) = 0.8 \cdot 0 = 0 \tag{2.112}$$

Then, the FI results in:

$$FI(mouse, e_k) = \frac{0.54 + 0.56 + 0}{3} = 0.36\overline{6} \tag{2.113}$$

In this example, the probability of finding a mouse, given these two spatial relations and three reference objects, decreased compared to the probability in Examples 6 and 7.

Example 9 *Consider the following scenario. In an office scene* `D`*, there are four object instances, as follows: an* `office desk`*, a* `monitor`*, a* `keyboard`*, and a* `phone`*. The virtual target object* `mouse` *has the position of the cell e with the index k. Considering this information and the learned knowledge f_β and f_τ, the following SPF are calculated:*

$$f_\beta(r_{near}, mouse, office\ desk) = 0.9 \tag{2.114}$$
$$f_\beta(r_{near}, mouse, monitor) = 0.8 \tag{2.115}$$
$$f_\tau(r_{right\text{-}of}, mouse, keyboard) = 0.8 \tag{2.116}$$
$$f_\tau(r_{in\text{-}front\text{-}of}, mouse, phone) = 0.0 \tag{2.117}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, office\ desk) = 0.9 \cdot 0.6 = 0.54 \tag{2.118}$$
$$SPF_\beta(r_{near}, mouse(e_k, mouse), mouse, monitor) = 0.8 \cdot 0.7 = 0.56 \tag{2.119}$$
$$SPF_\tau(r_{right\text{-}of}, mouse(e_k, mouse), mouse, keyboard, c) = 0.8 \cdot 0.4 = 0.32 \tag{2.120}$$
$$SPF_\tau(r_{in\text{-}front\text{-}of}, mouse(e_k, mouse), mouse, phone, c) = 0.0 \cdot 0.9 = 0 \tag{2.121}$$

Then, the FI results in:

$$FI(mouse, e_k) = \frac{0.54 + 0.56 + 0.32 + 0}{3} = 0.47\overline{3} \tag{2.122}$$

In this example, the probability of finding a mouse, given these three spatial relations and four reference objects, is $0.47\overline{3}\%$.

These Examples 6, 7 and 8 demonstrate the influence of valid relations and number of considered reference objects on the resulting overall probability value. As presented in Example 8, the presence of a keyboard in the scene with a probability of 0% of being in the spatial *right-of* relation with the target object, reduced the likelihood of finding a *mouse* at the given position. Therefore, it can be stated that all reference objects present in a given scene must be in a valid relation with the virtual target object to obtain a high probability value. In contrast, non-presence of reference objects had no negative impact on the probability value. As illustrated in Example 9, the presence of a reference object did not influence the FI calculation if the learned knowledge was equal to 0, and thus, the relation between the two objects did not exist.

After calculating the FI, the grid is normalized to obtain comparable results between different scenes.

Definition 30 *The grid Θ is normalized by the following function:*

$$\forall e \in \Theta: \quad e = \frac{e - min(\Theta)}{max(\Theta) - min(\Theta)} \tag{2.123}$$

2.4.1 Using FI for predicting the most probable position of an object

In this thesis, an approach for using FI to locate a given object is proposed. Because the primary objective of this thesis is the development of the PQSR, this method is used

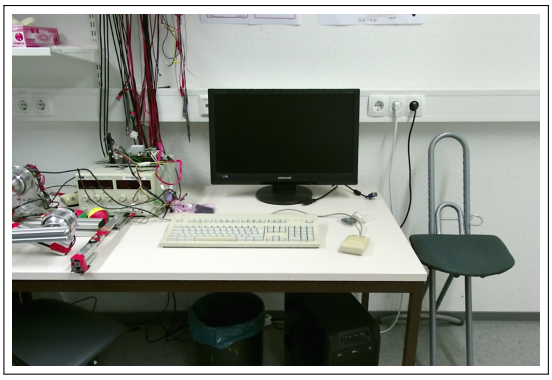
to illustrate how PQSR can be applied to enhance object search. Nevertheless, further possible applications for PQSR with respect to object search are discussed in Section 4.

In the previous Section 2.4, the formal definitions of Field Intensity and some related examples were provided. According to this definition, FI describes probability of a target object being in a certain spatial relation with a reference object at given position in the scene with respect to the learned probabilistic spatial knowledge about object co-occurrences and the current position of the target object. More precisely, the field intensity value indicates how likely the given object can be found at the position in the environment by considering all reference objects present in the scene and all possible spatial relations. In the robotics domain, to accomplish a given task and locate a sought after object, the system must be able to reason about probable object positions and where a given object can most likely be located. By referring to the current literature, it has been argued that an exhaustive search is not suitable for a human environment due to large search spaces. In this work, the FI is used as a heuristic for guiding search and, thus, reducing the search space. This spatial context of the environment is utilized to support the object search process by providing estimations for typical object positions in a given scene. By having this valuable information, the systems can commence the search by first inspecting positions with the highest FI values. FI in the scene can have the same values, and therefore, the Maximum Field Intensity (MFI) is calculated to obtain the most probable object position.

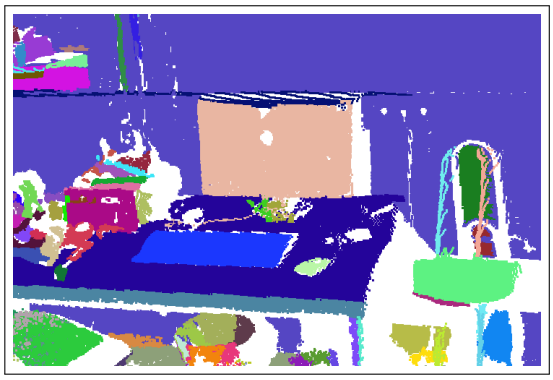
2.4.1.1 Maximum Field Intensity

After applying the FI formula 2.90, the field intensity in each cell of the grid is calculated. The FI corresponds to the probability value, which indicates how probable it is that the sought after object is located at a given position in the scene considering all known relations from the learned knowledge between all reference objects. By having these probabilities, a search strategy can be developed by which given objects can be found. Because each cell of the grid contains a FI value, the search can be commenced at the position where the FI is highest, and in cases where the object is not located at this position, proceed at the less probable position. However, according to the FI calculation, it is possible that several FI are equal, or rather, have the same probability value. To address this issue, a Maximum Field Intensity is defined. The MFI refers to an average position obtained from the positions of the highest FI. Figure 2.22 provides an exemplary scene with several highest FI and the corresponding MFI.

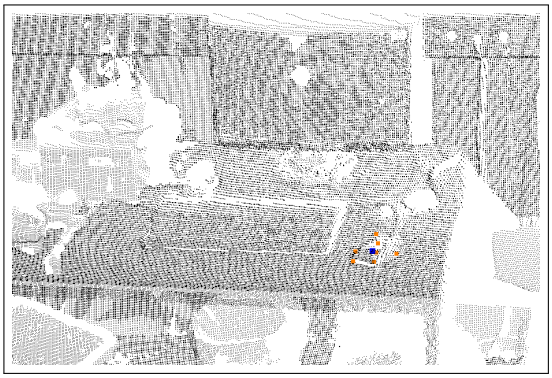
The algorithm 1 describes the process of locating the most probable object positions as denoted by the MFI value. Once the MFI is identified, the search for the object can be performed. In instances when the object is not present at the MFI position, the search is then conducted at the position with the second highest FI. This procedure can be carried out until the object has been found or all positions with the highest FI have been examined.



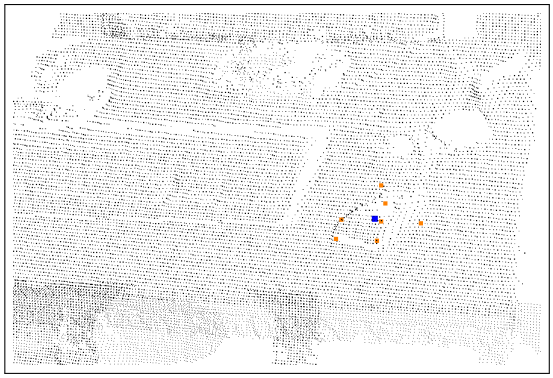
(a) Exemplary office scene of the DFKI-Lab



(b) Segmented planes of the office scene



(c) FI (orange) with the highest probability values



(d) The MFI (blue) of the FI from the left figure

Figure 2.22: An exemplary scene of a desk with the corresponding FI and MFI as the most probable *mouse* position.

$MaxValue \leftarrow 0$;
$Counter \leftarrow 0$;
$\Sigma \leftarrow 0.01$;
$SumPosition \leftarrow [0, 0, 0]$;
forall the $e \in \Theta$ **do**
  if $e > MaxValue$ **then**
    $MaxValue \leftarrow e$;
  end
end
forall the $e \in \Theta$ **do**
  if $e - MaxValue < \Sigma$ **then**
    $SumPosition \leftarrow SumPosition + f(e)$;
    $Counter \leftarrow Counter + 1$;
  end
end
return $SumPosition/Counter$;

Algorithm 1: Search for most probable object position

2.5 Summary

In this chapter, the formalism of the PQSR and the corresponding terms and definitions were provided. Furthermore, a new representation form SPF and resulting FI were presented. As described in the corresponding sections, the PQSR are first learned from real-world data and used to estimate the most probable position of the sought after object in a given environment. Thereby, this position is calculated in consideration of all possible spatial relations that can hold at the given position in the scene and of all reference objects in the environment. Compared with other methods (such as those described in Section 2.1), the PQSR provide a richer model of spatial relations and are suitable for different robotics applications such as object search or recognition. In this context, an exemplary approach for finding the most probable position of the object was proposed. However, because the primary objective of this thesis is to provide a formalism for the probabilistic model of the qualitative spatial relations, this method serves only as an example how the PQSR could be applied to guide search.

The PQSR have several advantages. First, by applying the PQSR, the spatial relations can be learned in a qualitative and probabilistic manner from quantitative data. These aspects are beneficial with regards to object search because they enable definition of the spatial relations in more precisely and in greater detail. Second, by determining the FI, one can estimate at which position a given object class is most likely to be found, even if the information about the relations is provided in a qualitative manner. This characteristic makes the PQSR more suitable for use in a man-machine interaction context.

The next chapter presents the results of the experiments related to the formalism described in this section. The experiments were performed to evaluate the applicability of the PQSR-based object position prediction in the context of object search.

3 Experiments

In this chapter, the theoretical concept of probabilistic qualitative spatial relations is verified by several experiments that address different aspects of the proposed method. The results of the respective experiments provide a basis for the evaluation and measurement of the applicability of the method with regards to the search domain. The chapter is grouped as follows: section 3.1 presents and discusses the experiments related to learning and modeling of the probabilistic qualitative spatial relations; followed by experiments in section 3.2 that were conducted to investigate how precisely the sought after object position can be predicted based on the field intensity method; and during the evaluation, the merits of spatial potential fields and field intensity is analyzed with regards to their usability for object search purposes.

Experiment overview

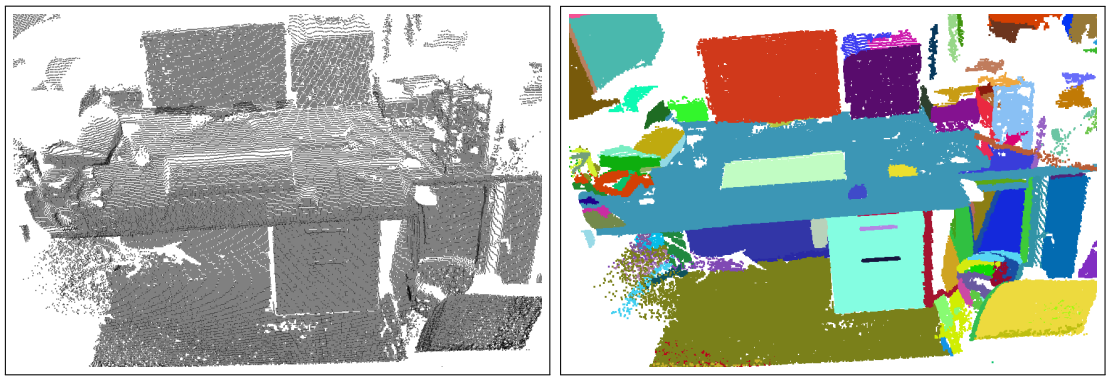
To evaluate the theoretical method developed in this thesis, several experiments were performed. Some experiments were conducted based on a real-world data from several indoor scenes, and for others, artificial data were used. The experiments were performed with a focus on an office environment, and part of the real-world data was acquired from different offices in the German Research Center for Artificial Intelligence - Robotics Innovation Center (DFKI-RIC) institute. Since the experiments relate to the evaluation of the theoretical concept of the PQSR rather than their application in an end-to-end system, the real-world data of the DFKI-RIC were acquired by using a Microsoft Kinect RGB-D sensor (second version) mounted on a tripod, as illustrated in Figure 3.1. Using this small and portable setup, the data could be acquired more easily, compared to data acquisition with a mobile robot system, as the equipment could be transported to various locations and the height of the tripod adjusted. This setup enabled data to be obtained from different views, which is not always possible due to limitations of the robot's size. For data acquisition, the rock robotics [roc] framework was used. The applications and algorithms developed in this thesis were performed on a Dell Inspiron 7000 Series notebook with an Intel Core i7-4500U processor, NVIDIA GeForce GT 750M, and 8GB RAM.

The collected scans focused on an office desk and its surrounding area. The DFKI-RIC data contains 26 scans with 26 office scenes, and was annotated manually with 13 different object classes. For the annotation, and to provide the *ground truth* information, the scans were segmented into planes using the region growing method described in [ED10], [EDK10]. An exemplary segmentation result of a point cloud is presented in Figure 3.3. Because each researcher arranged the objects on their office desk based on their preferences, the objects were arranged differently in each office. Consequently, not all offices contained the discussed object classes. Apart from the DFKI-RIC data, the external data set KTH-3D-TOTAL [TAA+14] from the web repository of the KTH Royal Institute of Technology (KTH) university group, was used. These data contain 495 scenes of 20 office desks

gathered in a context of long-term observations (each scene was scanned three times a day over 19 days). The data were manually annotated with 27 different object classes.

For additional experiments, an artificial scene of a large office room was created using of the Dia Diagram Editor (DIA) [dia] application (as illustrated in Figure 3.4). The resulting SVG file was then transformed to a point cloud using an application developed within this thesis. Furthermore, a real-world merged office scene of the DFKI-RIC institute composed of single scans was used. The merged process was performed using the SLAM6D [Nüc09] approach. A detailed description of the experiments and corresponding data is presented in their respective sections.

Figure 3.1: The acquisition setup used consisted of the Microsoft Kinect Camera v.2, tripod, and Dell Inspirion Notebook.



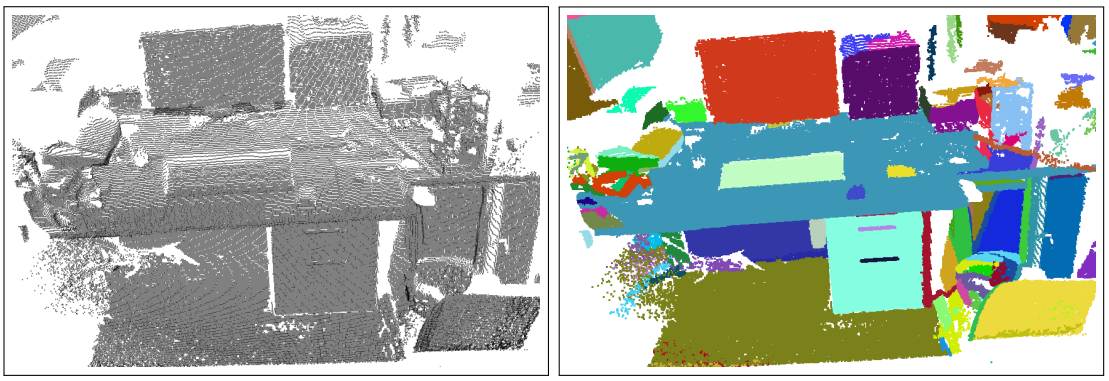
(a) Point cloud of an office scene (b) Corresponding planes

Figure 3.2: An exemplary scan of an office scene taken in the DFKI-RIC institute, and the segmentation result.

3.1 Experiments related to the PQSR learning

The probabilistic spatial relations are modeled as described in Section 2.2.2 and are based on the formalisms described in Section 2.2. In this section, the results from the learning of

the PQSR from real-world data are presented. The most relevant results are evaluated and discussed with respect to their formal definitions and spatial meanings. The remaining results are provided in Appendix A.1.



(a) Point cloud of an office scene (b) Corresponding planes

Figure 3.3: An exemplary scan of an office scene taken in the DFKI-RIC institute, and the segmentation result.

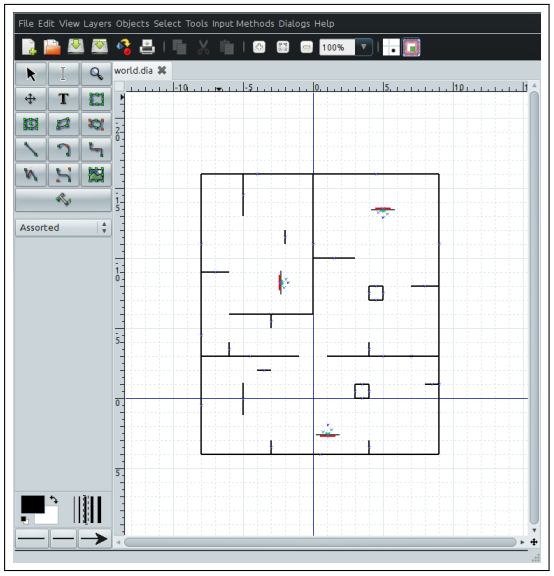


Figure 3.4: An artificial office scene created with the "Dia" application.

To evaluate the results of the PQSR learning process concerning different data, the experiments were performed using both the DFKI-RIC and the KTH data sets. By using external data from the KTH university, the learning of the spatial models from different environments was analyzed. The DFKI-RIC data set contains 26 office scenes as 3D point clouds and has been annotated with 13 object classes: *monitor, keyboard, table, mouse, mug, bottle, cupboard, phone, book, wall, ceiling, notebook, and floor.* The KTH data set contains 459 table-tops scenes taken from 20 offices and has been annotated with 27

object classes, for example, *ball, book, bottle, cellphone, CPU, external, flask, folder, glass, headphones, highlighter, jug, keyboard, keys, lamp, notebook, notepad, marker, monitor, mouse, mug, papers, pen, pencil, pen stand, phone, and rubber*. The annotated data were then used as input for the extraction of the spatial relations and further experiments.

Because the data were also used in the experiments related to the object search, the corresponding scans, which served as an input for the evaluation, were removed from the data used in the learning process according to the *leave-one-out-folding* method. As a result, 26 knowledge files were generated, whereby each piece of knowledge was learned in the absence of a given scan. Depending on the scan acting as input for the field intensity related experiments described in Section 3.2, the corresponding knowledge file (learned without this particular input scan) was used. In this way, it was assured that the results were not positively influenced by the data.

In the following Sections 3.1.1, 3.1.2 and 3.1.3, the results for the learning of the spatial distribution of considered object classes and their relations are presented. The results were evaluated according to the formal definitions of the PQSR 2.2.2. Furthermore, it was investigated whether the results meet expectations regarding the typical spatial arrangements of office objects. The overall summary of the learning results is provided in Section 3.1.4.

3.1.1 Learning of object sizes

To model the PQSR and calculate the SPF, the widths and depths of the considered objects must be known. Importantly, according to the formal definitions of spatial relations (apart from the *on* relation), the maximum allowed distance between the objects depends on their widths and depths. For learning of the distances and probabilities, the annotated data can directly be used because it already contains the object classes (including the target objects) and their sizes. However, for further experiments, such as those referring to the FI calculation, the size of the target object cannot be assumed to be provided and must be estimated. In real environments, the size of a particular sought after object is not always known in advanced. Hence, to perform the test under realistic conditions, the typical size of a given object class was also learned from the data.

According to Definition 15, the width and depth of an object class are calculated as an average value of all known sizes for a certain object class. Given this value, the width and depth of any target object of the known object classes can be estimated and thus, the spatial relation can be calculated. Table 3.1 provides the results of learning the widths and depths for all object classes annotated in the DFKI-RIC data.

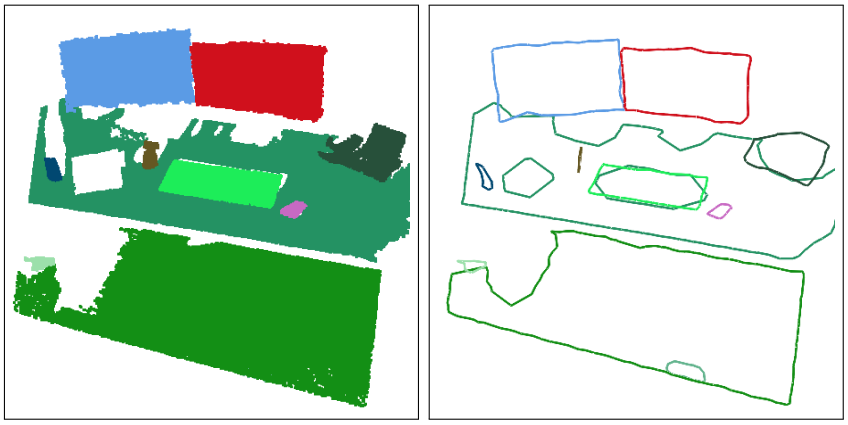
From Table 3.1, it can be observed that most values met expectations regarding the objects' sizes. For instance, a *keyboard* is two times as broad as its depth. Considering an average keyboard in an office, it can be argued that this result supports the observation. However, the notable results are those that refer to the usually large objects such as ceilings, floors, or walls. For these objects, the learned values appear to be too small. Moreover, for example, a floor has nearly the same depth as table, which does not appear to be correct. However, these results comply with the data and are correct. The reason for this atypical outcome lies, on the one hand, in the type of data used, and on the other hand, the method by which the object sizes are calculated.

Because the data correspond to the desk scenes, building structures such as floors or

Table 3.1: Learned average widths and depths of object classes annotated in the data (provided in meters).

Name	Width	Depth
Cupboard	0.535611	0.401712
Ceiling	3.71156	1.6601
Floor	2.2127	1.0901
Wall	2.99418	1.24084
Table	1.82113	1.03222
Keyboard	0.488412	0.242522
Monitor	0.592941	0.369981
Book	0.31991	0.186112
Mouse	0.107362	0.0671557
Notebook	0.44326	0.359984
Phone	0.629662	0.264798
Mug	0.152116	0.0960976
Bottle	0.154919	0.0983261

walls are only partially present in the data. Figure 3.5 provides an example of such a scene. From this figure, it can be seen that only part of the floor has been captured during data acquisition. As a result, the depth of the object (floor) is similar to the depth of a table. The second aspect that influenced the results was that according to Def. 2.2.2, referring to the calculation of the object's width and depth (which obeys the Axiom 3), the width and depth are extended along with the longest and second longest inertia axis, respectively. As a consequence, the value of the object's width is always greater than its depth. The results of learning the objects' sizes, presented in the Table 3.1, underline this assumption.



(a) Segmented planes of the office objects (b) Corresponding hulls

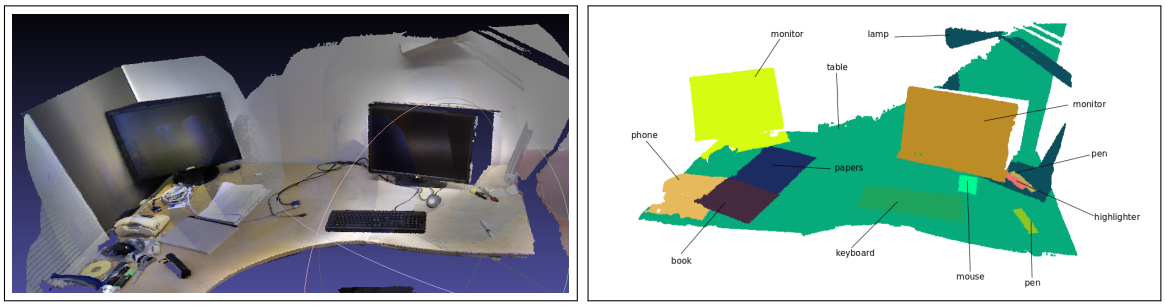
Figure 3.5: An exemplary scan with segmented planes of the office objects, i.e., floor, table, and phone with their hulls.

3.1.2 PQSR learning based on the DFKI-RIC and the KTH data sets

The following Tables 3.2-3.36 list the results of learning the seven probabilistic qualitative spatial relations: *above*, *on*, *near*, *in-front-of*, *behind-of*, *left-of*, and *right-of* from the DFKI-RIC and KTH data sets. As previously discussed, the DFKI-RIC data set includes 13 object classes. In the following experiments, the abbreviations for these classes are as follows: *MT-monitor, KB-keyboard, TA-table, MO-mouse, MU-mug, BT-bottle, CB-cupboard, NB-notebook, PH-phone, BO-book, WL-wall, CI-ceiling, and FL-floor*. The KTH data set contains 27 object classes with the following abbreviations: *KB-keyboard, MT-monitor, BO-book, MO-mouse, NB-notebook, PH-phone, MU-mug, BT-bottle, LA-laptop, CP-cellphone, HP-headphones, PC-pencil, PA-papers, PN-pen, HI-highlighter, MA-marker, FA-Flask, PS-pen stand, LP-lamp, FL-folder, GL-glass, JU-jug, RU-rubber, EX-external, NP-notepad, BA-ball, and CU-CPU*. The main difference between the data sets is that in the KTH data set, many small objects have been annotated compared to the DFKI-RIC data. Also, the object class *table* exists in the data but it has not been labeled (as highlighted in Figure 3.6). Importantly, because not all objects in the data are in a certain relation, the learning results do not contain all object pairs.

For each spatial relation, three tables were calculated. The first table includes the learned average probability values for the spatial relation between the object classes under consideration and the learned average distance values. The second table provides occurrences of the target object in the data compared with the number of cases where the relation was valid between the target and the reference object. Hereby, it was assumed that a target object occurred, at most, one time in a given scene, and this instance of a possible reference object class, located closest to the considered target object, was then considered. The third table presents the corresponding average distance values learned for two objects. For clarity and conciseness, not all results of learning the PQSR are presented in this section, but those of particular important for the evaluation are provided. The supplementary tables can be found in Appendices A.1.1 and A.1.2.

In the *horizontal* rows of the following tables, the *target* objects are listed, whereas the *columns* present the *reference* objects in the given spatial relation with the target object. The probability values provided in the tables specify how probable it was that a given object class could be found, on *average*, in a given relation with another object under consideration in the learned average distance. In this context, it should be noted that if, for instance, the probability of finding a keyboard and mouse *above* a floor is 91%, then it is not a given that the objects are at the same height above the floor. Importantly, the values refer to the learned position at which an object is expected to be located. However, the higher the probability value, the more likely it is that a given object at the learned position in the spatial relation would be found. For instance, if the probability of finding a mug above a floor is 70% and a monitor 80%, then it is generally more likely that the monitor, rather than the mug, would be found above the floor.



(a) Colored point cloud of an office scene (b) The annotated objects

Figure 3.6: An exemplary scan of a table top from the KTH data set (source:[kth]) with the corresponding annotated planes. The table has not been annotated in the data.

Results of learning the spatial relation above

DFKI-RIC data set Table 3.2 lists the learned probabilities for objects in the spatial relation *above* by taking into account the learned average distance values provided in Table 3.4. In Table 3.3, the ratio between the target object's occurrence (represented by a hash character) and the frequency of how often the above relation holds between the target and reference object, is presented. First, from Table 3.2, it can be observed that the values differ depending on the object class, as some of the target objects were more likely to be found above other objects. Interestingly, the probability of finding a *notebook* and *cupboard above* the *floor* amounted to 100%. This outcome can also be observed for a *book* and *notebook* above a *table*. It seems unlikely that these results are correct, but

Table 3.2: Learned average probabilities for objects in the spatial relation *above* (provided in percentages).

	Floor	Table	Keyboard
Cupboard	100.00	0.00	0.00
Wall	16.46	8.33	0.00
Table	95.80	0.00	0.00
Keyboard	91.86	99.82	0.00
Monitor	83.87	91.45	3.85
Book	0.00	100.00	0.00
Mouse	91.73	99.75	0.00
Notebook	100.00	100.00	0.00
Phone	49.58	89.59	0.00
Mug	71.71	93.78	0.00
Bottle	99.30	99.58	0.00

they conform with the formal definition of learning the *above* relation. According to the formula for learning the co-occurrence probability 12, the average probability results from the sum of all valid probability values for a given object and relation divided by the number

of occurrences of the target object. From Table 3.3, it can be observed that the objects *cupboard*, *notebook*, and *book* occurred only once in the data. Moreover, there is only one observation where the relation *above* holds between the target and the corresponding reference objects. As a result, the average probability for a *notebook* and a *cupboard* being above the *floor* was equal to 100%. This was also true for a *book* and *table*, as illustrated in Table 3.2.

Table 3.3: Target object occurrences (represented by a hash character) vs. the number of valid *above* relations between an object pair.

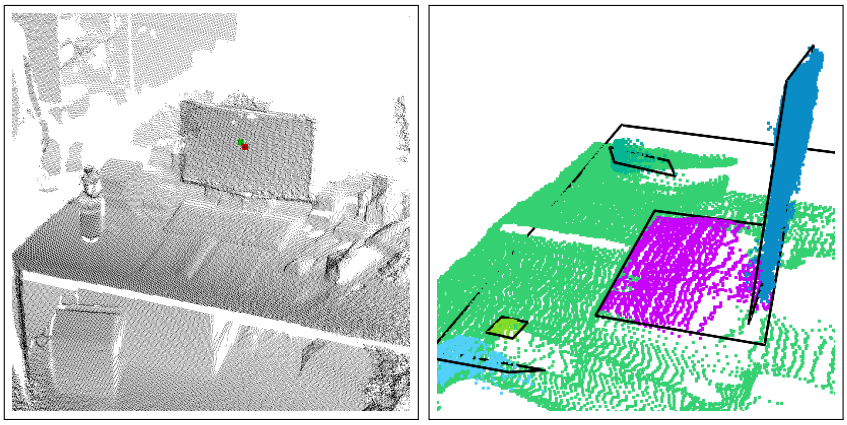
	#	Floor	Table	Keyboard
Cupboard	1	1	0	0
Wall	12	2	1	0
Table	26	25	0	0
Keyboard	26	24	26	0
Monitor	26	22	24	1
Book	1	0	1	0
Mouse	26	24	26	0
Notebook	1	1	1	0
Phone	10	5	9	0
Mug	18	13	17	0
Bottle	8	8	8	0

Table 3.4: Learned average distances for objects in the spatial relation *above* (provided in meters).

	Floor	Table	Keyboard
Cupboard	0.33	-	-
Wall	1.16	0.53	-
Table	0.74	-	-
Keyboard	0.78	0.03	-
Monitor	1.03	0.28	0.24
Book	-	0.04	-
Mouse	0.77	0.03	-
Notebook	0.87	0.14	-
Phone	0.82	0.07	-
Mug	0.80	0.07	-
Bottle	0.81	0.07	-

In addition to the target object occurrences, how often a given relation was valid between two objects also had an influence on the resulting probability values. For instance, from Table 3.2, it can be observed that the *phone* is located *above* the *floor* with 49.58% probability and *above* the *table* with 89.59% probability. The difference in the probability values lies in the number of observations in which the object *phone* was in an *above* relation

with the *floor* and *table*, respectively. Because a phone is more likely to be located above a table than above a floor, this resulted in the lower number of observations in which the relation *above* held for phone and floor (five and nine, respectively) as illustrated in Table 3.3.

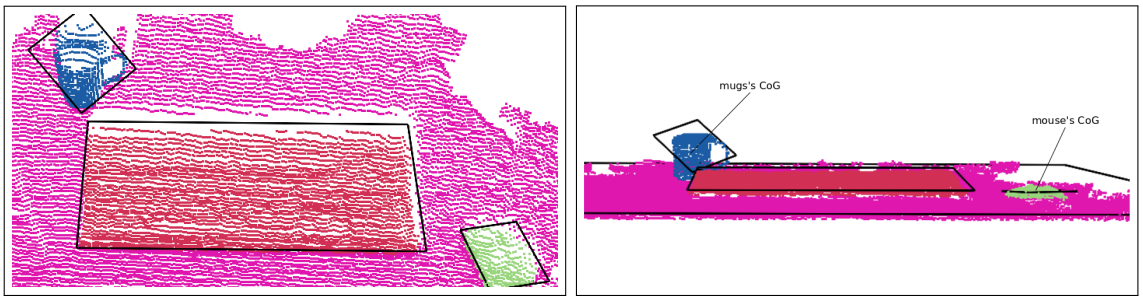


(a) CoG (red) of the object monitor (b) Hulls of the segmented objects

Figure 3.7: An exemplary scan with segmented planes of office objects, i.e., floor, table, and phone, and their hulls. It can seen that the keyboard is located exactly under the monitor.

Interestingly, the table reveals that a monitor can be also found above a *keyboard* with 3.85% probability. The reason for this unexpected result is that in one scene, a keyboard was observed as located below the monitor and very near it. Since the distance constraints (as calculated by the Formula 2.16) for the above relation were satisfied, the monitor was formally located above the keyboard. Figure 3.7 provides an exemplary scene that illustrates this particular case. As can be viewed in this figure, the hull of the object monitor (and its CoG) is located within the plane of the keyboard.

Table 3.4 presents the learned average distances for the above relation. By considering the values, the expectations regarding the spatial arrangement of the object classes, with respect to the above relation and the distances between the objects, are met. As expected, a cupboard was on average considered as closest to the floor, whereas a wall was the farthest. Furthermore, objects such as a notebook, mouse, mug, and phone were located at almost the same heights. The minor difference in the distance resulted, on the one hand, from the noise in the data, and on the other hand, from the fact that the CoG of the mug was higher than that of the mouse, as illustrated in Figure 3.8.



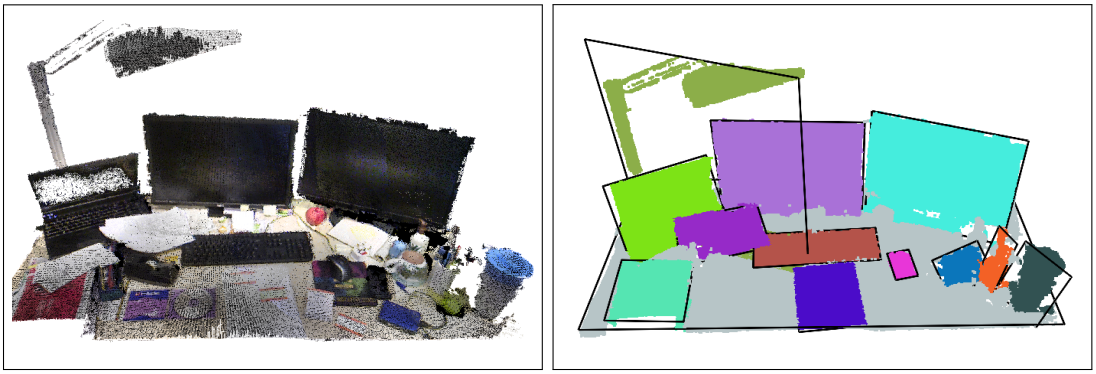
(a) Segmented planes of a table top (top view) (b) CoG of the objects (front view)

Figure 3.8: The center of gravity of the objects mouse, keyboard, and mug.

KTH data sets Tables 3.5 and 3.6 present the average probabilities for the *above* relation learned from the KTH data. From these tables, it can be observed that the resulting probability values were smaller than those learned from the DFKI-RIC data. These values were returned because in the KTH data, neither tables nor floors were selected and annotated. Since most objects in the table top scenes were in the *above* relation with the table, the probabilities for the remaining object classes were correspondingly low (or more precisely, they ranged from 0.2% to 6.2%). Moreover, by considering the information in Tables 3.7 and 3.8 regarding the learned objects occurrences, it is striking that the number of object classes present in the data was higher than the number of valid above relation between them. The highest probability for the *above* relation were observed for target object *lamp(LP)* and reference object *notebook(NB)*. This result was caused by the inaccurate segmentation of the object (lamp). Figure 3.9 highlights that the object lamp was segmented as a squared plane and its CoG was located higher than the CoG of the notebook. Moreover, the lamp was located within the surface of the notebook. Therefore, the above relation was valid for these two objects. The same applied for the reference object *headphones(HP)*, because most target objects were in the *above* relation with this reference object (as illustrated in Tables 3.5 and 3.6) However, the probabilities of findings these objects above the headphones was small, as the corresponding values did not exceed 5.9%.

Table 3.5: The first part of the learned average probabilities for objects in the spatial relation *above* from the KTH data set (provided in percentages).

	KB	MT	BO	MO	NB	MU	CP	HP	PA	PN
KB	0	0.2	0	0.7	0.2	0.2	0	0.2	0	0
MT	2.1	0	0	0.2	0	0.4	0.2	0.4	0	0
BO	0.6	0	0	0.6	0	0	0	0.6	0	0
MO	0.4	0	0	0	0.2	0.4	0	0.2	0.2	0
NB	1	1	0	0.5	0	0	0.5	0.5	0.5	0
MU	1.2	0	0.4	0.8	0	0	0	1.2	0	0
BT	0	0	0	0	0	0	0	0	0	0
CP	2.5	0	2.5	2.5	0	0	0	0	2.5	0
HP	0.8	0	0.8	0	0	0	0	0	0.8	0
PC	0	0	0	0	0	0	0	0	3.1	0
PA	0.6	0	0	0	0	0.3	0	0.3	0	0.3
PN	0	0	0	0	0	0	0	0	1.6	0
HI	0	0	0	0	0	0	0	1.3	0	0
MA	1.4	0	1.4	1.4	0	1.4	0	2.9	0	0
PS	0	0	0	0	0	0.7	0	0.7	0	0
LP	3.6	0	1.1	0	6.2	0	0.7	0.3	0.3	0
FA	2.8	5.7	0	0	0	2.8	0	5.7	2.8	0
JU	3.9	0	0	0	0	4	0	5.9	2	0
RU	0	0	0	0	0	0	0	3.5	3.5	0
NP	0	0	0	0	0	0	0	1.5	0	0



(a) Colored scan of a table top (b) Segmentation result with corresponding hulls

Figure 3.9: Point cloud of the KTH office scene (source: [kth]) and the resulting objects. As can be observed, the hull of the lamp was almost calculated as a square.

Table 3.6: The second part of the learned average probabilities for objects in the spatial relation *above* from the KTH data set (provided in percentages).

	MA	FL	PS	LP	FA	JU	NP
KB	0	0	0	0	0	0	0
MT	0.2	0	0	0.2	0	0.2	0
BO	0	0.6	0	0.6	0	0	0.6
MO	0	0	0	0	0	0	0
NB	0	0	0	0.5	0	0	0
MU	0.4	0	0.4	0	0	0	0
BT	0	0	0	0	1	0	0
CP	0	0	0	0	0	0	0
HP	0	3.4	0	0	0	0	3.4
PC	0	0	0	0	0	0	0
PA	0.3	0.3	0	0	0	0	0
PN	0	0	0	0	0	0	3.3
HI	0	0	0	0	0	0	1.3
MA	0	1.4	0	0	0	1.4	0
PS	0	0	0	0	0	0	0
LP	0	0.3	0	0	0	0	0.7
FA	2.8	0	0	0	0	5.6	0
JU	2	0	0	0	0	0	0
RU	0	0	0	0	0	0	0
NP	0	0	0	0	0	0	0

Table 3.7: The first part of the target object occurrences and the number of valid *above* relations between the object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	MU	CP	HP	PA	PN	MA	FL
KB	410	0	1	0	3	1	1	0	1	0	0	0	0
MT	452	10	0	0	1	0	2	1	2	0	0	1	0
BO	163	1	0	0	1	0	0	0	1	0	0	0	1
MO	409	2	0	0	0	1	2	0	1	1	0	0	0
NB	196	2	2	0	1	0	0	1	1	1	0	0	0
MU	233	3	0	1	2	0	0	0	3	0	0	1	0
BT	95	0	0	0	0	0	0	0	0	0	0	0	0
CP	40	1	0	1	1	0	0	0	0	1	0	0	0
HP	117	1	0	1	0	0	0	0	0	1	0	0	4
PC	32	0	0	0	0	0	0	0	0	1	0	0	0
PA	320	2	0	0	0	0	1	0	1	0	1	1	1
PN	119	0	0	0	0	0	0	0	0	2	0	0	0
HI	74	0	0	0	0	0	0	0	1	0	0	0	0
MA	68	1	0	1	1	0	1	0	2	0	0	0	1
PS	133	0	0	0	0	0	1	0	1	0	0	0	0
LP	260	10	0	3	0	17	0	2	1	1	0	0	1
FA	35	1	2	0	0	0	1	0	2	1	0	1	0
JU	50	2	0	0	0	0	2	0	3	1	0	1	0
RU	28	0	0	0	0	0	0	0	1	1	0	0	0
NP	64	0	0	0	0	0	0	0	1	0	0	0	0

Table 3.8: The second part of the target object occurrences and the number of valid *above* relations between the object classes learned from the KTH data set.

	PS	LP	FA	JU	NP
KB	0	0	0	0	0
MT	0	1	0	1	0
BO	0	1	0	0	1
MO	0	0	0	0	0
NB	0	1	0	0	0
MU	1	0	0	0	0
BT	0	0	1	0	0
CP	0	0	0	0	0
HP	0	0	0	0	4
PC	0	0	0	0	0
PA	0	0	0	0	0
PN	0	0	0	0	4
HI	0	0	0	0	1
MA	0	0	0	1	0
PS	0	0	0	0	0
LP	0	0	0	0	2
FA	0	0	0	2	0
JU	0	0	0	0	0
RU	0	0	0	0	0
NP	0	0	0	0	0

Results of learning the spatial relation on

DFKI-RIC data set In Table 3.9, the resulting average probability values for four reference and ten target objects in the spatial relation *on* are presented. The small number of reference objects resulted from the spatial relation only being valid for those objects, despite the learning process of all reference objects in the scenes being considered. Similar to the *above* relation, the probability values depended on the object types between which the spatial relation *on* held. For example, as listed in Table 3.9, the average probability of finding a *mouse* and *mug* *on* the *table* was 98.66% and 91.86%, respectively. These high values indicate that those objects were located closely to the learned average distance and thus, the probability of finding them in the on relation and at the learned position was high. Furthermore, according to the formal Def. 7 of the *on* relation, the distance between two objects must not exceed a given threshold. In this case, the distance was smaller than the threshold and thus, the distance constraints for this relation were satisfied.

Additionally, values listed in Tables 3.9 and 3.10 reveal that the smaller the ratio between the occurrence of the target object and the ratio where the on relation was valid, the higher the average probability value was. For instance, a *keyboard* and a *monitor* occurred 26 times in the data. Although the occurrence of these objects in the data set was the same, the number of observations where the *on* relation was valid between these objects and a *table* was greater for a *keyboard* than for a *monitor*. Therefore, the probability for a keyboard being on a table with respect to the learned average distances was higher than that for a monitor.

An interesting and unexpected result was found regarding the target object *monitor* and reference object *table*. One would expect that the probability for a monitor to be on a table should be similarly high for a keyboard or mouse. However, Table 3.9 reveals that this was not that case. Although in the data, the *monitor* occurred 26 times, the *on* relation was valid in only 24 cases between these objects and the *table*. The explanation for this outcome can be obtained by looking at Figure 3.10. From this figure, it can be observed that the monitor is not located on the table, as the surface of the table ends before the monitor surface begins. Consequently, and with respect to the spatial on relation, the monitor is not located on the table. According to the formal definition of the *on* relation, it is required that a target object is located within the surface of a reference object, which is not the case in this scene. The same is also true for the *mug* and *phone*. Nevertheless, the results presented in Table 3.9 correspond in general to the expectations regarding the given objects and the on relation.

However, as illustrated in this table, the target objects *floor*, *monitor*, *mouse*, and *mug* are located *on* the *wall* with 3.85% and 5.56% probability, respectively. Although initially the correlation between these objects and the wall appears to be surprising, the results are correct and comply with the definition of the *on* relation. For the *on* relation, the distance and the z-axis of the reference object is considered, regardless of the orientation of the target and reference objects, therefore, this is a correct result. Accordingly, an object such as a *monitor* is located on a *wall* if the distance between the objects is small enough and thus, the maximum allowed distance constraint, given by a threshold value, is satisfied.

Similar to the results of learning the *above* relation, a *monitor* was found *on* a *keyboard* with a probability of 3.85% in the *on* relation. By comparing information in Tables 3.4 and 3.11, it is noticeable that the distance between these objects for both relations was

the same. This finding indicates that the same scene caused this outcome.

The result suggesting that the table was not located on the floor is somewhat surprising. However, the finding is correct and was caused by the too small threshold value chosen for the on relation. Since the value for an on relation has to be relatively small to distinguish between the on and the above relations, this value was much smaller than the distance between the objects floor and table. As a result, the on relation was not valid between the floor and the table.

Table 3.9: Learned average probabilities for objects in the spatial relation *on* (provided in percentages).

	Floor	Wall	Table	Keyboard
Cupboard	100.00	0.00	0.00	0.00
Floor	0.00	3.85	0.00	0.00
Keyboard	0.00	0.00	98.61	0.00
Monitor	0.00	3.85	85.97	3.85
Book	0.00	0.00	100.00	0.00
Mouse	0.00	3.85	98.66	0.00
Notebook	0.00	0.00	100.00	0.00
Phone	0.00	0.00	86.89	0.00
Mug	0.00	5.56	91.86	0.00
Bottle	0.00	0.00	97.07	0.00

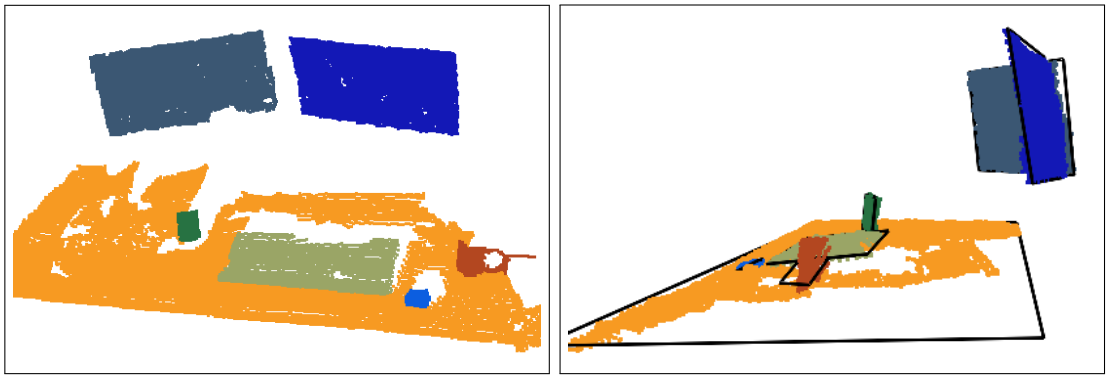
Table 3.10: Target object occurrences (represented as hash character) vs. the number of valid *on* relations between the object classes.

	#	Floor	Wall	Table	Keyboard
Cupboard	1	1	0	0	0
Floor	26	0	1	0	0
Keyboard	26	0	0	26	0
Monitor	26	0	1	24	1
Book	1	0	0	1	0
Mouse	26	0	1	26	0
Notebook	1	0	0	1	0
Phone	10	0	0	9	0
Mug	18	0	1	17	0
Bottle	8	0	0	8	0

KTH data set Tables 3.12 and 3.13 reveal that the *on* relation learned from the KTH data held on average lower probabilities than the *on* relation extracted from the DFKI-RIC data set. More precisely, the probability values for the objects ranged between 2% and 22.1%. However, in comparison with the DFKI-RIC results (provided in the Table 3.9), more object classes were observed in the KTH data. Nevertheless, the important

Table 3.11: Learned average distances for objects in the spatial relation *on* (provided in meters).

	Floor	Wall	Table	Keyboard
Cupboard	0.33	-	-	-
Floor	-	0.28	-	-
Keyboard	-	-	0.03	-
Monitor	-	0.23	0.29	0.24
Book	-	-	0.03	-
Mouse	-	0.22	0.03	-
Notebook	-	-	0.14	-
Phone	-	-	0.07	-
Mug	-	0.20	0.06	-
Bottle	-	-	0.07	-



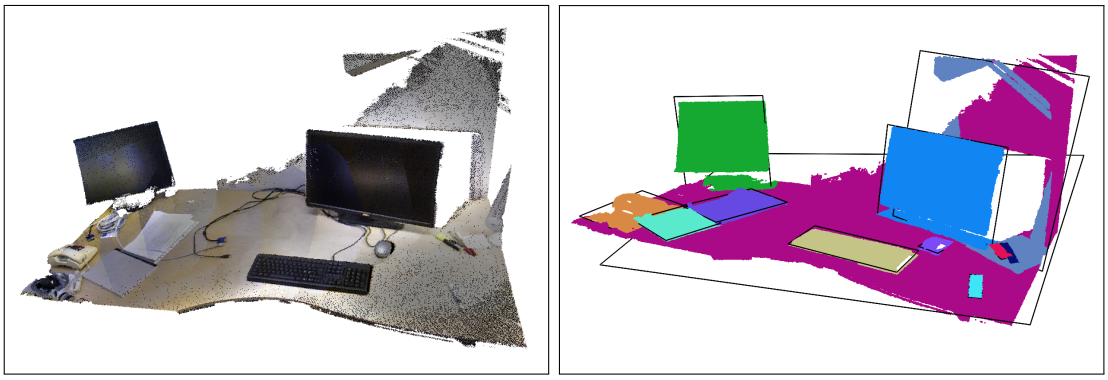
(a) Planes of the office objects (b) Their corresponding hulls

Figure 3.10: An exemplary scan in which the object monitor is not located on a table, since the table ends before the plane of the monitor begins.

objects for the *on* relation (table and floor) are not present or annotated in the KTH data. As a result, the probability values for the remaining objects were correspondingly small. Furthermore, when considering Tables 3.12 and 3.13, it is striking that the highest probabilities that could be observed were between objects and the reference object *lamp*. Similar to the *above* relation also learned from the KTH data set, this finding is partly caused by inaccuracies in the object's segmentation (in this particular case by the square-like hull of the lamp) but also because many small objects were located very close to the object (lamp). For the on relation, only the maximum allowed distance (given by a threshold value) and not the orientation of the objects (in the world coordinate system) is a crucial factor for the relation to be valid, which lead to the results of smaller objects such as a mouse or pen being located on the lamp. Figure 3.11 displays a table desk with the segmented lamp on it. From this figure, it can be observed that the lamp has not been segmented properly. Consequently, the hull of the lamp is almost a square. Moreover, the small objects are located near the lamp.

Table 3.12: The first part of the learned average probabilities for objects in the spatial relation *on* from the KTH data set (provided in percentages).

	KB	MT	BO	MO	NB	MU	BT	CP	HP	PA	HI	MA
KB	0	5.5	0	1.2	0.4	0.4	0	0	0	0	0	0
MT	3.4	0	0	0.3	0	0.7	0	0	0.4	0.2	0	0.2
BO	0.6	4.1	0	0.6	0	0.6	0	0	0	0	0	0
MO	0.2	4.8	0	0	0.2	0.4	0	0	0	0.2	0	0
NB	1	3.9	0	0.5	0	0.5	0.5	0.5	0.5	0.5	0	0
PH	0	0	0	0	0	0	1.1	0	0	0	0	0
MU	1.5	7.3	0.4	1.1	0.4	0	1.2	0	0.4	0	0	0.8
BT	1	3.9	0	0	0	0	0	0	0	0	0	0
CP	2.5	6	0	2.5	0	2.5	0	0	0	2.5	0	0
HP	0.8	3.5	0.8	0.8	0	1.6	0	0	0	1.6	0	0.8
PC	0	4	0	0	0	0	0	0	0	0	0	0
PA	0.6	1	0	0.6	0	0.6	0.3	0	0	0	0	0.6
PN	0	2.5	0	0	1.6	0.8	0	0	0	1.6	0	0
HI	0	3.7	0	0	0	2.5	1.3	0	0	0	0	0
MA	2.9	5.3	0	1.4	0	0	0	0	0	0	0	0
FL	0	0	1.5	0	0	0	0	0	0	0	0	0
PS	1.5	10.3	0	0	0	7	0	0	0.7	0	0	0
LP	3	11.1	1.2	0.3	3.1	0.3	0	0.6	0.3	0	0.3	0
FA	2.8	5.4	0	2.8	0	5.6	4.2	0	5.5	4.7	0	2.8
GL	0	0	0	0	0	0	3.3	0	0	0	0	0
JU	3.9	2	0	3.7	0	4	0	0	3.9	2	0	3.9
RU	0	9.5	0	0	0	0	3.5	0	3.5	3.5	0	0
NP	0	1.5	0	0	0	0	0	0	1.5	0	0	0



(a) Exemplary point cloud (b) Their corresponding hulls

Figure 3.11: A segmentation result of an object lamp from the KTH data set. The hull of the lamp has an almost square-like form and some small objects are located nearby.

Table 3.13: The second part of the learned average probabilities for objects in the spatial relation *on* from the KTH data set (given in percentages).

	FL	PS	LP	FA	GL	JU	CU	NP
KB	0	0	5.1	0	0	0	0	0
MT	0	0	2.1	0	0	0.4	0	0
BO	0.6	0.6	9.9	0	0.6	0	0	0.6
MO	0	0	3.6	0	0	0	0	0
NB	0	0	13.5	0	0	0	0	0
PH	0	3.3	0	0	0	0	0	0
MU	0	0.4	5.8	0.4	0	0.8	0	0
BT	0	2.9	7.6	1	0	2	3.6	0
CP	0	0	16.2	0	0	0	0	0
HP	3.3	0	2.4	0	0	2.4	0	3.4
PC	0	0	5.7	0	0	0	0	0
PA	0.3	0	1.6	0	0	0	0	0
PN	0	0	5.5	0	0	0	0	3.3
HI	0	0	15.8	0	0	0	0	1.3
MA	1.4	0	3.5	0	0	0	0	0
FL	0	0	1.5	0	0	0	0	0
PS	0	0	4.8	0.7	0.7	2.8	0	0
LP	0.3	0	0	1.8	0.3	0	0	0.5
FA	0	2.8	22.1	0	0	5.5	0	0
GL	0	3.3	10.5	0	0	3.3	0	0
JU	0	7.3	2	0	0	0	0	0
RU	0	0	3.5	0	0	0	0	0
NP	4.5	0	2.1	0	0	1.5	0	0

Results of learning the spatial relation near

DFKI-RIC data set When considering the results listed in Table 3.14, which presents the learned average probabilities for the spatial relation *near*, it is noteworthy that the number of target and reference objects for which the *near* relation held was higher than for the *on* and *above* relations. This finding is likely due to rules defined in Def. 5, that is, the *near* relation is less constrained as the remaining spatial relations.

From this Table, it is apparent that the *near* relation held most frequently between the target object *floor* and reference object *table* because for the *near* relation, the considered maximum allowed distance is calculated from the object's width and depth, as given by the formula 2.12. Since a *floor* and *table* are rather large entities compared to other objects, the maximum allowed distance between them was correspondingly high. However, as revealed in Table 3.14, the probability of, for instance, a *bottle* being in the *near* relation with a *floor* (96.2%) was slightly higher than with a *table* (93.8%). The small difference in the probabilities was caused by the fact that due to the resulting maximum allowed distance, the spatial relation (the corresponding SPF reaches a more spatial space) and thus, the range of the relation, was wider. By considering the width and depth values

from Table 3.1, it can be observed that a *floor* is larger than a *table*.

As previously discussed, the learned probability value specifies how likely it is that a given target object class is located at the learned average distance. For instance, as listed in Tables 3.14 and 3.16, a *mug* can be found with 96.7% probability at 0.9 meters *near* the *floor* and with 94.1% at 0.5 meters to the *table*. Although the probability for a floor was higher, the distance to the table was smaller than to the floor. This exemplary result indicates that the probabilities did not have any direct correlation to the distance values.

It is remarkable that the values of the table in the diagonal line were equal to zero. This finding met the expectations that an object cannot be in the near relation with itself, which is correct and conforms to the definition of the *near* relation.

According to Def. 5, the spatial relation *near* is symmetric because the symmetry refers to the object's instances and not classes. Since the learned probability values correspond to the objects classes, the learned probability can, but does not necessarily, have to be symmetric. The symmetry depends on the number of the target object occurrences and the number of valid *near* relations between it and the given reference object. The results provided in Table 3.14 include both cases. For instance, the target object *keyboard* can be found at 0.2 m *near* the *table* with 96,9% probability, and the probability of finding the *table* at 0.2 m *near keyboard* was also 96,9%. This exemplary result demonstrates that the values are equal and thus the probability of finding these objects near each other is symmetric. By considering the corresponding vales from Table 3.15, it can be noted that both objects occurred 26 times in the data and the near relation held between them 26 times. This result indicates that in this particular case, the same objects' instances were taken into account for the relation calculation. In contrast, the probability of finding a *mug near* a *monitor* was 73.3% and finding a *monitor near* a *mug* corresponded to 50.7%, which was lower. In this case, the learned relation *near* was not symmetric. Table 3.15 reveals that the values were 18 and 15 and 26 and 15, respectively. Therefore, the difference in the probability values was caused by the unequal number of target object occurrences.

Table 3.14: Learned average probabilities for objects in the spatial relation *near* (given in percentages).

	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	0	0	100	0	100	100	100	0	100	0	0	0	0
CI	0	0	100	100	100	100	100	0	100	0	0	100	0
FL	3.8	3.8	0	41.1	97.3	96.7	97.7	3.8	93.9	3.8	35.7	66.9	29.6
WL	0	8.3	89.2	0	88.1	81.1	82.5	8.3	77.6	0	21.8	52.8	21.1
TA	3.8	3.8	97.3	40.7	0	96.9	98.2	3.8	95.3	3.8	36.5	65.1	28.8
KB	3.8	3.8	96.7	37.4	96.9	0	96.4	3.8	92.4	3.8	33.7	61.9	27.9
MT	3.8	3.8	97.7	38	98.2	96.4	0	3.8	86.7	3.8	34.7	50.7	27.5
BO	0	0	100	100	100	100	100	0	100	0	100	0	0
MO	3.8	3.8	93.9	35.8	95.3	92.4	86.7	3.8	0	3.8	11.4	20.4	13.5
NB	0	0	100	0	100	100	100	0	100	0	0	0	0
PH	0	0	92.8	26.2	94.9	87.6	90.3	10	29.6	0	0	33.6	13.2
MU	0	5.5	96.7	35.2	94.1	89.5	73.3	0	29.5	0	18.6	0	9.2
BT	0	0	96.2	31.7	93.8	90.8	89.5	0	43.9	0	16.5	20.8	0

Table 3.15: Occurrences of target objects in the data and number of valid *near* relations between target and reference objects.

	#	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	1	0	0	1	0	1	1	1	0	1	0	0	0	0
CI	1	0	0	1	1	1	1	1	0	1	0	0	1	0
FL	26	1	1	0	12	26	26	26	1	26	1	10	18	8
WL	12	0	1	12	0	12	12	12	1	12	0	3	8	3
TA	26	1	1	26	12	0	26	26	1	26	1	10	18	8
KB	26	1	1	26	12	26	0	26	1	25	1	10	18	8
MT	26	1	1	26	12	26	26	0	1	24	1	10	15	8
BO	1	0	0	1	1	1	1	1	0	1	0	1	0	0
MO	26	1	1	26	12	26	25	24	1	0	1	5	6	4
NB	1	0	0	1	0	1	1	1	0	1	0	0	0	0
PH	10	0	0	10	3	10	10	10	1	5	0	0	5	3
MU	18	0	1	18	8	18	18	15	0	6	0	5	0	2
BT	8	0	0	8	3	8	8	8	0	4	0	3	2	0

Table 3.16: Learned average distances for objects in the spatial relation *near* (given in meters).

	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	-	-	0.6	-	0.4	0.5	0.9	-	0.5	-	-	-	-
CI	-	-	3.3	1.6	2.5	2.8	2.3	-	2.9	-	-	2.9	-
FL	0.6	3.3	-	2.3	0.9	0.9	1.1	1.6	1	1	1.2	0.9	1
WL	-	1.6	2.3	-	1.8	1.9	1.5	2.3	2	-	1.7	1.8	2.1
TA	0.4	2.5	0.9	1.8	-	0.2	0.4	0.7	0.4	0.6	0.7	0.5	0.6
KB	0.5	2.8	0.9	1.9	0.2	-	0.4	0.6	0.3	0.4	0.8	0.4	0.5
MT	0.9	2.3	1.1	1.5	0.4	0.4	-	0.8	0.6	0.5	0.7	0.5	0.6
BO	-	-	1.6	2.3	0.7	0.6	0.8	-	0.2	-	0.3	-	-
MO	0.5	2.9	1	2	0.4	0.3	0.6	0.2	-	0.7	0.8	0.2	0.2
NB	-	-	1	-	0.6	0.4	0.5	-	0.7	-	-	-	-
PH	-	-	1.2	1.7	0.7	0.8	0.7	0.3	0.8	-	-	0.7	1
MU	-	2.9	0.9	1.8	0.5	0.4	0.5	-	0.2	-	0.7	-	0.3
BT	-	-	1	2.1	0.6	0.5	0.6	-	0.2	-	1	0.3	-

KTH data set The learned average probability values from the KTH data set for the *near* relation were considerably higher compared to those for the *on* or *above* relations (as illustrated in Tables 3.17 and 3.18). Similar to the results from the DFKI-RIC data set, the *near* relation held between all objects occurring in the data. This finding occurred because, as previously mentioned, the *near* relation is less constrained than the *above* and *on* relations, since for the near relation only the maximal allowed distance is considered. In contrast, for instance in the on relation for the target object must additionally be located

within the reference object plane (compare Def. 7 and Def.5 for more details). From Table 3.17, it can be observed that the highest probabilities for the *near* relation refer to the reference object *monitor* and the corresponding target objects. This finding is related to the fact that the reference object monitor was larger than the remaining object classes such as the pen or highlighter. Therefore, the corresponding maximum allowed distance, which is calculated from the objects' widths and depths, was also the highest. Thus, the objects were still in a near relation, even if they were not located in close proximity to the monitor. In contrast, the objects such as the rubber and highlighter must be located close to each other to be in the spatial relation near. Nevertheless, the results of learning the near relation met the expectation of object arrangements and are comparable to those obtained from the DFKI-RIC data set.

Table 3.17: The first part of the learned average probabilities for objects in the spatial relation *near* from the KTH data set (given in percentages).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	0	96	27.8	81.1	38.4	11.8	41.3	16.2	6.1	0	23.6	6.2	46.2	20.8
MT	87	0	29.5	75.6	40.9	14.5	41.8	16.4	7.7	0.2	24.1	6.3	59.3	21.2
BO	70.1	81.8	0	18.1	29.9	1.7	18.6	4.8	3.7	0	27.3	3.4	42.7	12.7
MO	81.3	83.5	7.2	0	25.4	5.9	29	10.6	2.5	0	4.3	2.5	20.8	10.2
NB	80.4	94.4	24.8	53.1	0	13	26.8	8.2	11.2	0.5	18.8	6.4	46.3	23.5
PH	56.5	76.2	3.3	28.2	29.7	0	4.4	11.2	7.3	1.1	2.3	4.6	23.4	7.7
MU	72.7	81.1	13	51	22.5	1.6	0	9	1.7	0	15.9	2.1	28.3	12.6
BT	70.1	78	8.3	45.6	17	10.2	22.2	0	1.9	0	3.7	4.9	29.8	11.9
CP	63.3	87.1	15.3	25.6	54.8	15.8	10.2	4.6	0	0	13.5	4.3	28.9	18.5
KS	0	100	0	0	100	100	0	0	0	0	0	0	0	0
HP	83	93.3	38	15.1	31.5	1.7	31.7	3	4.6	0	0	9.8	54.9	18.7
PC	80.1	89.3	17.6	32.1	39.6	12.4	15.2	14.6	5.4	0	36.1	0	30.6	25.6
PA	59.2	83.7	21.7	26.6	28.4	6.2	20.6	8.8	3.6	0	20	3	0	11.3
PN	71.9	80.7	17.4	35.3	38.8	5.6	24.8	9.5	6.2	0	18.4	6.9	30.4	0
HI	68.5	74.4	20.6	31.1	20.9	9.3	21	7.3	1.3	0	22.6	5	35.4	12.2
MA	70.6	80.1	24.4	32.8	14.2	5.1	13.8	4.7	1.4	0	27.4	3.8	32.4	9.9
FL	49.5	78.6	20.2	23.4	36.5	26	10	3.3	2.1	1.5	22.2	1.6	37	7.5
PS	66.3	81.2	12.4	47.9	33.5	30.2	17	17.8	6.1	0.7	10.5	3.5	26	8.2
LP	88	91.2	33.2	71.5	39.7	11.3	46.5	21.7	9.3	0.3	30.5	4.8	63.4	27.2
FL	75.7	89.8	12.3	29.5	12.1	7.9	29.1	23.9	7.2	0	52.1	4.6	50.6	15.8
GL	64.7	89.1	22.6	44.8	32.7	0	17.2	19.4	0	0	24.3	0	37.7	7.2
JU	77.1	89.1	17.8	52.6	11.2	0	45.5	13.9	0	0	30.5	6.4	33	14.9
BA	55.8	97.8	0	0	80.8	0	0	0	0	0	50.7	0	0	0
CU	99.2	99.2	0	82.9	16.6	0	0	64.2	0	0	0	16.6	0	40
RU	73.3	83.5	17.1	47.6	31.5	9.5	12.2	10.6	3.5	0	23.5	26.4	25.1	23.8
EX	0	0	0	0	0	99.4	0	0	0	0	0	0	0	0
NP	76.2	91.4	32.8	36.5	34.8	5.3	37	10.4	4.8	0	28	4	51.1	19.7

Table 3.18: The second part of the learned average probabilities for objects in the spatial relation *near* from the KTH data set (given in percentages).

	HI	MA	FL	PS	LP	FL	GL	JU	BA	CU	RU	EX	NP
KB	12.3	11.7	7.9	21.5	55.8	6.4	4.7	9.4	0.4	1.4	5	0	11.8
MT	12.1	12	11.4	23.8	52.4	6.9	5.9	9.8	0.6	1.3	5.1	0	12.9
BO	9.3	10.1	8.1	10.1	53	2.6	4.1	5.4	0	0	2.9	0	12.9
MO	5.6	5.4	3.7	15.5	45.4	2.5	3.2	6.4	0	1.2	3.2	0	5.7
NB	7.9	4.9	12.3	22.7	52.7	2.1	5	2.8	1.2	0.5	4.5	0	11.3
PH	8	4	20	46.7	34.2	3.2	0	0	0	0	3.1	2.3	3.9
MU	6.6	4	2.8	9.7	51.9	4.3	2.2	9.7	0	0	1.4	0	10.1
BT	5.6	3.4	2.3	25	59.4	8.8	6.1	7.3	0	4	3.1	0	7
CP	2.5	2.5	3.4	20.4	60.6	6.3	0	0	0	0	2.5	0	7.7
KS	0	0	100	100	100	0	0	0	0	0	0	0	0
HP	14.3	15.9	12.5	12	67.9	15.6	6.2	13	1.3	0	5.6	0	15.3
PC	11.6	8.2	3.3	14.6	39	5.1	0	10	0	3.1	23.1	0	8
PA	8.2	6.8	7.6	10.8	51.5	5.5	3.5	5.1	0	0	2.2	0	10.2
PN	7.5	5.7	4.2	9.2	59.4	4.6	1.8	6.2	0	2	5.6	0	10.6
HI	0	4.3	7.5	0	62.2	7.6	0	12	0	0	1.3	0	10.7
MA	4.7	0	7	5.8	57	8.7	2.7	11.9	0	0	0	0	8.8
FL	8.5	7.2	0	40.6	50.4	5.7	3.7	1.5	0	0	1.5	0	16.3
PS	0	3	20.1	0	58.5	3.7	5.4	7.1	0	0	2.7	1.4	9
LP	17.7	14.9	12.8	29.9	0	9.1	8.2	12	1	1.4	5.2	0	17.9
FL	16.1	17	10.8	14.1	67.8	0	9.9	27.5	0	0	2.8	0	13.5
GL	0	6.2	8.2	24.1	71.5	11.5	0	9.5	0	0	0	0	12.7
JU	17.7	16.2	2	19	62.8	19.3	5.7	0	0	0	0	0	13.7
BA	0	0	0	0	90.3	0	0	0	0	0	0	0	0
CU	0	0	0	0	64.3	0	0	0	0	0	23.1	0	16.6
RU	3.5	0	3.5	13	48.6	3.5	0	0	0	4.9	0	0	19
EX	0	0	0	97.6	0	0	0	0	0	0	0	0	0
NP	12.3	9.4	16.8	18.8	72.8	7.4	6	10.7	0	1.5	8.3	0	0

Table 3.19: The first part of the occurrences of target objects and the number of valid *near* relations between the objects from the KTH data set.

	#	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC
KB	410	0	409	138	350	171	57	203	77	30	0	108	29
MT	452	409	0	153	365	195	76	210	82	40	1	117	31
BO	163	138	153	0	37	61	4	38	10	7	0	54	7
MO	409	350	365	37	0	133	30	135	52	13	0	25	13
NB	196	171	195	61	133	0	30	68	20	25	1	42	15
PH	86	57	76	4	30	30	0	4	10	7	1	2	4
MU	233	203	210	38	135	68	4	0	25	5	0	48	6
BT	95	77	82	10	52	20	10	25	0	2	0	5	5
CP	40	30	40	7	13	25	7	5	2	0	0	6	2
KS	1	0	1	0	0	1	1	0	0	0	0	0	0
HP	117	108	117	54	25	42	2	48	5	6	0	0	13
PC	32	29	31	7	13	15	4	6	5	2	0	13	0
PA	320	227	299	81	104	110	24	80	34	14	0	73	12
PN	119	100	106	24	51	60	8	39	13	9	0	27	10
HI	74	63	62	20	26	19	8	22	6	1	0	20	5
MA	68	59	62	20	27	13	4	11	4	1	0	22	4
FL	66	39	64	15	20	28	18	8	3	2	1	18	2
PS	133	111	131	20	72	54	44	26	30	10	1	19	5
LP	260	251	258	109	224	120	40	145	69	29	1	94	15
FA	35	32	33	5	13	5	3	12	11	3	0	20	2
GL	30	25	30	9	16	12	0	7	7	0	0	9	0
JU	50	47	50	11	29	8	0	26	8	0	0	17	4
BA	3	2	3	0	0	3	0	0	0	0	0	2	0
CU	6	6	6	0	5	1	0	0	4	0	0	0	1
RU	28	24	26	6	15	11	3	5	4	1	0	8	11
EX	2	0	0	0	0	0	2	0	0	0	0	0	0
NP	64	59	64	25	29	27	4	31	8	4	0	23	3

Table 3.20: The second part of the occurrences of target objects and the number of valid *near* relations between the objects from the KTH data set.

	PA	PN	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU
KB	227	100	63	59	39	111	251	32	25	47	2	6	24
MT	299	106	62	62	64	131	258	33	30	50	3	6	26
BO	81	24	20	20	15	20	109	5	9	11	0	0	6
MO	104	51	26	27	20	72	224	13	16	29	0	5	15
NB	110	60	19	13	28	54	120	5	12	8	3	1	11
PH	24	8	8	4	18	44	40	3	0	0	0	0	3
MU	80	39	22	11	8	26	145	12	7	26	0	0	5
BT	34	13	6	4	3	30	69	11	7	8	0	4	4
CP	14	9	1	1	2	10	29	3	0	0	0	0	1
KS	0	0	0	0	1	1	1	0	0	0	0	0	0
HP	73	27	20	22	18	19	94	20	9	17	2	0	8
PC	12	10	5	4	2	5	15	2	0	4	0	1	11
PA	0	43	30	26	31	40	197	20	14	22	0	0	8
PN	43	0	13	8	6	13	83	7	3	9	0	3	8
HI	30	13	0	4	7	0	57	6	0	11	0	0	1
MA	26	8	4	0	6	5	48	7	2	10	0	0	0
FL	31	6	7	6	0	31	39	4	3	1	0	0	1
PS	40	13	0	5	31	0	90	6	10	11	0	0	4
LP	197	83	57	48	39	90	0	29	27	38	3	4	17
FA	20	7	6	7	4	6	29	0	4	11	0	0	1
GL	14	3	0	2	3	10	27	4	0	3	0	0	0
JU	22	9	11	10	1	11	38	11	3	0	0	0	0
BA	0	0	0	0	0	0	3	0	0	0	0	0	0
CU	0	3	0	0	0	0	4	0	0	0	0	0	2
RU	8	8	1	0	1	4	17	1	0	0	0	2	0
EX	0	0	0	0	0	2	0	0	0	0	0	0	0
NP	39	17	10	8	13	14	54	5	4	8	0	1	6

Results of learning the spatial relations in-front-of and behind-of

DFKI-RIC data set Table 3.21 presents the results of learning the *in-front-of* relation from the DFKI-RIC data set. Note that the object class *wall* is not listed as a target object in the table, which meets the expectations that most objects in a room are located in front of the wall. Furthermore, the results demonstrate that object such as a *keyboard* or *mouse* are located *in-front-of* of the monitor with 95.6% and 86.5% probabilities. Also, these results corresponded with general expectations about typical placement of such objects.

A more interesting result refers to the target object *monitor*. From Table 3.22, it can be observed that there are three cases in which the monitor was located *in-front-of* of a table. This results from the definition of the *in-front-of* relation. According to Def. 10, the distance between the target and reference object is calculated by considering the CoG of both objects. Because of this, a monitor is located *in-front-of* of table if it is located in the front area beginning from the table's CoG. Intuitively, one can expect that the relation *in-front-of* of a table refers to the area of the front edge of the table, which does not comply with the definition of the *in-front-of* relation. In fact, in this case, the front of relation corresponds to the front area of the table starting from its CoG. Figure 3.12 provides a visualization of a table and the four projective relations with respect to the table's CoG.

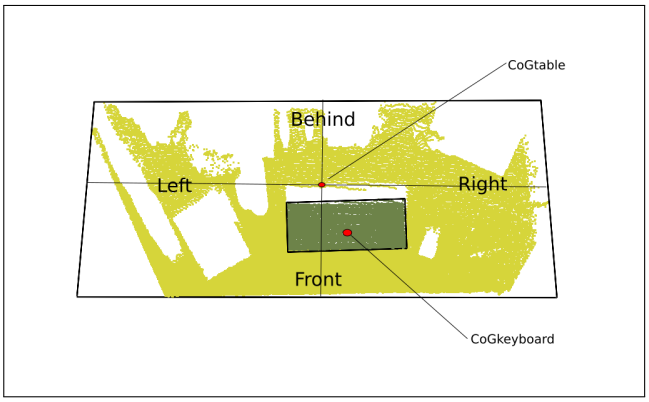


Figure 3.12: Visualization of a table with its projective spatial relations according to its CoG.

In Table 3.23, the results of learning the spatial relation *behind-of* are presented. When considering the learned average probability, it can be argued that most of the values meet the expectation regarding the behind relation. For instance, the probabilities of finding a *keyboard* behind a *monitor*, *bottle*, or *notebook* were equal to zero. In turn, a monitor was located behind a keyboard, mouse, and mug, which conformed to the results of the opposite case. Furthermore, Table 3.23 reveals that the *behind-of* relation held most probable, that is, 95.6%, between the target object (monitor) and reference object (keyboard). This finding suggests that these objects were located very near to the average distance learned from the data for the *behind-of* relation. The overall results indicate that a monitor and a wall are most likely to be found behind all remaining objects. Therefore, it follows that a monitor and a wall are located farthest to the back of the scene.

Table 3.21: Learned average probabilities for objects in the spatial relation *in-front-of* (given in percentages).

	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	0	0	100	0	100	100	100	0	0	0	0	0	0
CI	0	0	0	100	0	0	0	0	0	0	0	0	0
FL	0	3.8	0	40.6	48.4	15.1	81.9	3.8	7.6	0	24.7	18.9	7.6
TA	0	3.8	48.7	40.7	0	3.8	86.5	0	3.8	3.8	16.9	11.4	3.8
KB	0	3.8	81.5	37.7	93.9	0	95.6	3.8	3.8	3.8	32.8	36.1	14.5
MT	0	3.8	11	38.4	11.3	0	0	0	0	0	8.5	3.8	0
BO	0	0	0	100	100	0	100	0	0	0	100	0	0
MO	3.8	3.8	88.2	36.6	92.6	86.6	86.5	3.8	0	3.8	11	10	13.9
NB	0	0	100	0	0	0	100	0	0	0	0	0	0
PH	0	0	28.2	26.8	48.4	0	56.6	0	0	0	0	10	10
MU	0	5.5	68.4	34.3	79.5	41.2	69.7	0	15.2	0	12.4	0	5.5
BT	0	0	72.8	34.5	83.7	46.7	93.8	0	0	0	13.1	12.5	0

Table 3.22: Occurrences of the target objects vs. the number of valid spatial relations *in-front-of* between the given objects.

	#	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	1	0	0	1	0	1	1	1	0	0	0	0	0	0
CI	1	0	0	0	1	0	0	0	0	0	0	0	0	0
FL	26	0	1	0	12	13	4	22	1	2	0	7	5	2
TA	26	0	1	13	12	0	1	23	0	1	1	5	3	1
KB	26	0	1	22	12	25	0	26	1	1	1	10	10	4
MT	26	0	1	3	12	3	0	0	0	0	0	3	1	0
BO	1	0	0	0	1	1	0	1	0	0	0	1	0	0
MO	26	1	1	24	12	25	24	24	1	0	1	5	3	4
NB	1	0	0	1	0	0	0	1	0	0	0	0	0	0
PH	10	0	0	3	3	5	0	6	0	0	0	0	1	1
MU	18	0	1	13	8	15	8	14	0	3	0	4	0	1
BT	8	0	0	6	3	7	4	8	0	0	0	2	1	0

By comparing Tables 3.21 and 3.23, a correlation between *behind-of* and *in-front-of* relations can be observed. For instance, according to results from Table 3.21, the probability of finding a *monitor in-front-of* a *keyboard*, *mouse*, or *mug* was equal to zero. In contrast, with respect to the probability values from Table 3.23, a *monitor* was located *behind-of* these objects with probabilities 95.6%, 86.5%, and 48.2%, respectively. These results correspond to Def. 11 and Def. 10 and the properties of the *behind-of* and *in-front-of* relations. This is because, according to the Def. 10 and 11 the front and behind relations are converse to each other[1].

[1]It should be noted that these relations are converse because in their definitions, object instances and not classes are taken into account.

Table 3.23: Learned average probabilities for objects in the spatial relation *behind-of* (given in percentages).

	CB	CI	FL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	0	0	0	0	0	0	0	100	0	0	0	0
CI	0	0	100	100	100	100	0	100	0	0	100	0
FL	3.8	0	0	48.7	81.5	11	0	88.2	3.8	10.8	47.4	22.4
WL	0	8.3	88.1	88.3	81.8	83.3	8.3	79.3	0	22.4	51.5	23
TA	3.8	0	48.4	0	93.9	11.3	3.8	92.6	0	18.6	55	25.7
KB	3.8	0	15.1	3.8	0	0	0	86.6	0	0	28.5	14.3
MT	3.8	0	81.9	86.5	95.6	0	3.8	86.5	3.8	21.7	48.2	28.8
BO	0	0	100	0	100	0	0	100	0	0	0	0
MO	0	0	7.6	3.8	3.8	0	0	0	0	0	10.5	0
NB	0	0	0	100	100	0	0	100	0	0	0	0
PH	0	0	64.3	44.1	85.2	22.1	10	28.7	0	0	22.3	10.5
MU	0	0	27.4	16.4	52.1	5.5	0	14.4	0	5.5	0	5.5
BT	0	0	24.8	12.5	47.3	0	0	45.4	0	12.5	12.5	0

Table 3.24: Occurrences of the target objects vs. the number of valid spatial relations *behind-of* between the given objects.

	#	CB	CI	FL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	1	0	0	0	0	0	0	0	1	0	0	0	0
CI	1	0	0	1	1	1	1	0	1	0	0	1	0
FL	26	1	0	0	13	22	3	0	24	1	3	13	6
WL	12	0	1	12	12	12	12	1	12	0	3	8	3
TA	26	1	0	13	0	25	3	1	25	0	5	15	7
KB	26	1	0	4	1	0	0	0	24	0	0	8	4
MT	26	1	0	22	23	26	0	1	24	1	6	14	8
BO	1	0	0	1	0	1	0	0	1	0	0	0	0
MO	26	0	0	2	1	1	0	0	0	0	0	3	0
NB	1	0	0	0	1	1	0	0	1	0	0	0	0
PH	10	0	0	7	5	10	3	1	5	0	0	4	2
MU	18	0	0	5	3	10	1	0	3	0	1	0	1
BT	8	0	0	2	1	4	0	0	4	0	1	1	0

However, the learned average probability for these relations can, but does not necessarily have to be, converse. This property is only given if the considered objects are the same instances of the given object classes in both relations. For example, by comparing Tables 3.21 and 3.23, it can be observed that a keyboard was found *in-front-of* a monitor with 92.6%, and the probability of finding a monitor *behind-of* a keyboard was also 92.6%. Conversely, a mug could be found *in-front-of* a keyboard with 41.2% probability, but the probability of finding the keyboard *behind-of* a mug was only 28.5%. An analysis of the results reveals that there were only 18 mugs, but 26 keyboards, in the DFKI-RIC data set

(as listed in Tables 3.22 and 3.24). Although in both cases the relations were valid eight times, the number of target object occurrences was not the same. Therefore, the resulting probability was not equal. However, apart from the same number of valid relation, the average distances for these relations were also equal (as provided in Tables A.1 and A.2 in Appendix A.1.1). This finding indicates that the valid relations include the same object instances. Based on this result, it has been demonstrated that the *behind-of* and *in-front-of* relations can, but do not necessarily have to be, symmetric. Similar to the results of the *near* relation, the symmetry depends on the number of target object occurrences in the data.

KTH data sets Tables 3.25 and 3.26 present the outcomes of learning the *in-front-of* relation from the KTH data. When considering the probability values from these tables, it is remarkable that the values were, on average, highest for target objects and the reference object *monitor*. These results are comparable with those from the DFKI-RIC data set because, for instance, the probability of finding the target objects *in-front-of* of a monitor was also correspondingly high. Furthermore, many of the target objects were located *in-front-of* of the reference object *lamp*. This was because, as illustrated in Figure 3.6, the lamp is located in the back half of the table.

The learned probabilities from the KTH data set for the *behind-of* relation are provided in Tables 3.27 and 3.28.From these tables, it can be observed that, unlike the results from the DFKI-RIC data set, there are several target objects that were located *behind-of* a *monitor*. This is related to the amount and type of objects that were annotated. Because the KTH data set contains considerably more smaller objects than the DFKI-RIC data set, it is reasonable that the number of valid *behind-of* relations between the *monitor* and those objects was also correspondingly high. Furthermore, due to the given objects' functions, which also affect their arrangement, objects such as a pen or rubber were in general more likely to be located *behind-of* a monitor than, for instance, objects such as the keyboard or mug.

Table 3.25: The first part of learned average probabilities for objects in the spatial relation *in-front-of* from the KTH data set (given in percentages).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	HP	PC	PA	PN	HI
KB	0	91.6	11.6	25.6	20.4	5.3	23	6.9	1.8	8.7	1.5	21.7	7.7	6.7
MT	0.4	0	4.3	0.8	3.6	3.9	0.2	0.6	0.2	0.4	0.6	3.4	1.5	1.7
BO	45.7	65.3	0	13.2	20.9	1.2	11.4	3.1	1.7	15.1	2.1	14.6	4.6	6.6
MO	49.8	82.2	2.8	0	24.9	1.4	20.7	6.5	1.4	3.5	0.9	10.2	4.8	4.3
NB	33.9	75.2	9.2	10.3	0	4.8	6	2.1	1.8	5.5	1.8	11.6	7.2	2.4
PH	37.8	62.6	2.1	23.8	21.5	0	4.1	11.1	2.2	2.3	1.1	23.3	6.1	5.8
MU	37.2	76.2	5.1	14.9	19.1	0	0	7.8	1.2	6.1	1.4	13.8	5.7	2.3
BT	38.8	66.2	2.8	13.2	14.4	0	5	0	1	4.1	1	9.9	5.3	4.1
CP	45.4	86.2	8.8	10.2	46.6	9.1	4.6	2.5	0	9.1	2.5	14.4	15.4	0
KS	0	100	0	0	100	100	0	0	0	0	0	0	0	0
HP	51.4	92.6	16.3	6.7	23.7	0	21.9	0.8	1.3	0	2.4	22.4	6.1	5.1
PC	63.8	79.1	7.9	21.7	30.2	9.3	5.5	11.8	3.1	28.2	0	10.6	21.1	5.8
PA	23.2	59.8	9	9.8	14.9	0	7.4	4.5	1.1	6	0.9	0	3.9	3
PN	46.4	73.8	9.9	17.4	33.3	1.4	15.8	4.4	1.1	14.2	1.6	13.1	0	6.2
HI	35	65	7.4	5.9	12.7	2.4	13.8	0	1.3	13.7	4	11.5	3.8	0
MA	42.1	73.1	6.9	4.4	13.8	2.7	5.5	4.2	1.4	12.4	1.4	12.1	2.5	4
FL	15.4	60.6	11.3	7.4	17.1	17.9	0	1.5	0	8.9	1.5	13.4	1.5	2.9
PS	35.2	89.2	5.2	7.7	18.3	14.1	7.1	7.8	1.3	3.4	0.7	8.3	1.4	0
LP	8.7	53.8	13.4	4.5	14.4	1.4	10.7	4.9	2.2	7.1	1.4	4.2	2.9	4
FL	33.6	73.4	11.5	2.8	5.4	0	5	15.4	2.8	10.5	2.8	5.4	5	3.8
GL	47.9	80.5	12.5	21.5	27.4	0	15.7	7.8	0	14.8	0	15.7	3.3	0
JU	49.7	76.3	5.4	27.4	8.5	0	25.4	11.1	0	9.2	0	11.3	3.1	6.3
BA	0	33.3	0	0	91.7	0	0	0	0	0	0	0	0	0
CU	0	99.3	0	0	0	0	0	0	0	0	16.6	0	33.2	0
RU	55	84.7	6.5	26	26.2	3.5	8.6	0	3.5	11.6	24.3	6.4	14.2	0
EX	0	0	0	0	0	99.6	0	0	0	0	0	0	0	0
NP	52.8	83.3	22	22.1	29.2	1.5	26.6	9.8	2.9	24.5	2.6	35.1	6	11.2

Table 3.26: The second part of learned average probabilities for objects in the spatial relation *in-front-of* from the KTH data set (given in percentages).

	MA	FL	PS	LP	FL	GL	JU	BA	CU	RU	NP
KB	5	5.8	13.3	51.7	4.1	1.3	4.3	0.4	1.4	1.6	4.3
MT	0.8	3.4	0.2	21.6	0.8	0.4	1.3	0.4	0	0	1.4
BO	4.2	3.8	5.7	37.8	0	1.6	3.2	0	0	2.2	4.2
MO	4.5	3.2	12.1	47.8	1.2	1	3.3	0	1.2	1.2	2.5
NB	0.9	6.8	12.4	38.2	1.2	0.9	1.5	0	0.5	1.4	2.3
PH	1.7	6.2	26.4	37.4	3.3	0	0	0	0	2.3	3.4
MU	2.2	3.1	5.9	45	3.6	0.8	4.8	0	0	0.8	4.2
BT	1	2	16.3	47.5	4	3.6	2.1	0	3.9	3.9	1
CP	0	3.7	17.6	51.1	4.6	0	0	0	0	0	4.4
KS	0	100	100	100	0	0	0	0	0	0	0
HP	4.4	9.1	10.7	60	12.8	3.3	9.1	1.4	0	3.3	4.6
PC	6.8	3.1	10.7	32.1	3.1	0	10.6	0	0	8.5	3.1
PA	1.9	4.6	5.3	36.9	3.2	1.6	3	0	0	1.2	1.9
PN	4.3	3.6	7.1	56.7	3.3	0.8	4.9	0	0.8	1.5	9.2
HI	1.3	6.3	0	52	5.2	0	6.6	0	0	0	2.4
MA	0	6.2	5.8	50.4	8.3	1.4	8.8	0	0	0	2
FL	0	0	6.7	39	2.9	0	0	0	0	1.5	0
PS	0.7	17.4	0	46.9	0.7	4.2	1.4	0	0	0	2
LP	2.5	4.3	8.2	0	1.1	3	2.5	1.1	0	1.8	4
FL	2.8	4.4	12.7	70.6	0	10.2	2.8	0	0	0	5
GL	3.3	8.2	11.7	58.7	0	0	3.3	0	0	0	8
JU	2	2	17.1	56.5	18.3	3.9	0	0	0	0	12.6
BA	0	0	0	0	0	0	0	0	0	0	0
CU	0	0	0	64	0	0	0	0	0	33.1	0
RU	0	0	12.7	40.3	3.5	0	0	0	0	0	7.1
EX	0	0	99	0	0	0	0	0	0	0	0
NP	8.4	16.8	15.8	62	3.9	1.5	1.5	0	1.5	4.8	0

Table 3.27: The first part of learned average probabilities for objects in the spatial relation *behind-of* from the KTH data set (given in percentages).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	0	0.4	18.1	49.7	16.2	7.9	21.1	9	4.4	0	14.6	4.9	18.1	13.4
MT	83.1	0	23.5	74.4	32.6	11.9	39.3	13.9	7.6	0.2	23.9	5.6	42.3	19.4
BO	29.3	11.9	0	7.1	11.1	1.1	7.3	1.6	2.1	0	11.7	1.5	17.7	7.2
MO	25.7	0.9	5.2	0	4.9	5	8.5	3	1	0	1.9	1.7	7.7	5
NB	42.7	8.4	17.3	51.9	0	9.4	22.7	6.9	9.5	0.5	14.1	4.9	24.4	20.2
PH	25.6	20.6	2.3	6.7	11	0	0	0	4.2	1.1	0	3.4	0	2
MU	40.5	0.4	7.9	36.3	5	1.5	0	2	0.8	0	11	0.7	10.1	8.1
BT	30	3.1	5.4	28.3	4.3	10	19.2	0	1	0	1	3.9	15.3	5.6
CP	18.6	2.5	6.9	15.1	8.9	4.8	7.1	2.5	0	0	3.8	2.5	8.9	3.2
HP	30.7	1.6	21.1	12.5	9.2	1.6	12.2	3.3	3.1	0	0	7.7	16.4	14.4
PC	20.2	9.3	10.7	12.5	11.3	3.1	10.7	3.1	3.1	0	9	0	9	6.1
PA	27.8	4.9	7.4	13	7.1	6.2	10	2.9	1.8	0	8.2	1	0	4.8
PN	26.7	5.7	6.3	16.7	11.8	4.4	11.2	4.3	5.2	0	6	5.7	10.5	0
HI	37.1	10.4	14.7	24.1	6.5	6.7	7.5	5.3	0	0	8.2	2.5	13	10
MA	30.6	5.7	10.1	27.1	2.8	2.1	7.7	1.4	0	0	7.6	3.2	9.3	7.5
FL	36	23.7	9.5	19.9	20.2	8.1	11.1	2.9	2.2	1.5	16.2	1.5	22.4	6.5
PS	41.1	0.7	7	37.3	18.3	17	10.4	11.6	5.3	0.7	9.4	2.5	12.9	6.4
LP	81.5	37.6	23.7	75.2	28.8	12.3	40.4	17.3	7.8	0.3	27	3.9	45.5	25.9
FL	48.8	11.2	0	14.7	6.8	8.2	24.4	10.9	5.3	0	42.9	2.8	29.3	11.2
GL	18.8	6.4	9	14.4	6.1	0	6.5	11.4	0	0	13.1	0	17.8	3.3
JU	35.7	11.8	10.5	27.1	5.8	0	22.7	3.9	0	0	21.4	6.8	19.4	11.8
BA	61.6	66.5	0	0	0	0	0	0	0	0	57.6	0	0	0
CU	99.1	0	0	82.7	16.6	0	0	63.1	0	0	0	0	0	16.6
RU	23.8	0	13.1	18.2	9.8	7.1	6.7	13.4	0	0	13.8	9.7	13.8	6.7
NP	28	10.5	10.7	16.1	7.1	4.6	15.5	1.5	2.7	0	8.4	1.5	9.6	17.1

Table 3.28: The second part of learned average probabilities for objects in the spatial relation *behind-of* from the KTH data set (given in percentages).

	HI	MA	FL	PS	LP	FL	GL	JU	BA	CU	RU	EX	NP
KB	6.3	6.9	2.4	11.4	5.5	2.8	3.5	6	0	0	3.7	0	8.2
MT	10.6	11	8.8	26.2	30.9	5.6	5.3	8.4	0.2	1.3	5.2	0	11.8
BO	3.4	2.8	4.6	4.2	21.4	2.4	2.3	1.6	0	0	1.1	0	8.6
MO	1	0.7	1.2	2.5	2.9	0.2	1.5	3.3	0	0	1.7	0	3.4
NB	4.8	4.7	5.7	12.4	19.1	0.9	4.2	2.1	1.4	0	3.7	0	9.5
PH	2.1	2.1	13.7	21.8	4.2	0	0	0	0	0	1.1	2.3	1.1
MU	4.3	1.6	0	4	11.9	0.7	2	5.4	0	0	1	0	7.3
BT	0	3	1	10.9	13.4	5.6	2.4	5.8	0	0	0	0	6.6
CP	2.5	2.5	0	4.6	14.5	2.5	0	0	0	0	2.5	0	4.6
HP	8.6	7.2	5	3.9	15.9	3.1	3.8	3.9	0	0	2.7	0	13.4
PC	9.3	3.1	3.1	3.1	12.1	3.1	0	0	0	3.1	21.3	0	5.3
PA	2.6	2.5	2.7	3.4	3.4	0.5	1.4	1.7	0	0	0.5	0	7
PN	2.4	1.4	0.8	1.6	6.4	1.4	0.8	1.3	0	1.6	3.3	0	3.2
HI	0	3.6	2.6	0	14.3	1.8	0	4.3	0	0	0	0	9.7
MA	1.4	0	0	1.4	9.8	1.4	1.4	1.4	0	0	0	0	7.9
FL	7.1	6.4	0	35.2	17.1	2.3	3.7	1.5	0	0	0	0	16.2
PS	0	2.9	3.3	0	16.1	3.3	2.6	6.4	0	0	2.6	1.4	7.6
LP	14.8	13.1	9.9	24	0	9.5	6.7	10.8	0	1.4	4.3	0	15.2
FL	11	16.2	5.6	2.8	8.1	0	0	26.2	0	0	2.8	0	7.1
GL	0	3.3	0	18.8	26.1	11.9	0	6.5	0	0	0	0	3.3
JU	9.8	12	0	3.9	13.4	2	2	0	0	0	0	0	2
BA	0	0	0	0	97.8	0	0	0	0	0	0	0	0
CU	0	0	0	0	0	0	0	0	0	0	0	0	16.6
RU	0	0	3.5	0	17.3	0	0	0	0	7.1	0	0	11.1
NP	2.8	2.1	0	4.1	16.3	2.7	3.7	9.8	0	0	3.1	0	0

Results of learning the spatial relations left-of and right-of

DFKI-RIC data sets Tables 3.29 and 3.31 list the results of learning the projective spatial relations *left-of* and *right-of* from the DFKI-RIC data set. From these tables, it can be observed that, for instance, an average probability of finding a *mouse left-of* a *keyboard* was 3.8%, whereas the probability of finding a *mug left-of* a *keyboard* was 55.2%. These results indicate that an object mug was located more to the *left-of* a keyboard than the mouse due to the higher probability value between the mug and keyboard. However, this conclusion would be not correct because both probability values are not directly related, since for each relation only two objects and the distance between them are considered. Therefore, the results demonstrate how probable it is that an object can be found *left-of* another object at a learned distance. Although it was more likely that a *mug* would be found *left-of* a *keyboard* than a *mouse* would, this is no indication that the *mug* was placed more *left-of* the *keyboard* than the *mouse*.

Table 3.29: Learned average probabilities for objects in the spatial relation *left-of* (given in percentages).

	CB	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CI	0	100	100	100	100	100	0	100	0	0	100	0
FL	3.8	0	29.3	51.3	53.7	50.4	3.8	80.1	3.8	25.2	35.5	14.4
WL	0	31	0	48.2	37.4	30.9	8.3	52.2	0	16.4	8.3	12.8
TA	3.8	45.1	22.2	0	51.9	33.7	3.8	92	0	22.8	28.8	14.4
KB	3.8	36.7	22.3	41.3	0	29.7	3.8	88	0	21.6	24.7	14.1
MT	3.8	22.1	18.7	26.1	26.5	0	3.8	67.8	0	21.3	14.5	13.6
BO	0	0	0	0	0	0	0	0	0	100	0	0
MO	3.8	7.5	18.2	0	3.8	3.8	3.8	0	0	17.6	16.7	13.2
NB	0	0	0	100	100	100	0	100	0	0	0	0
PH	0	27.5	10	38.3	38.2	34.2	0	0	0	0	18.4	10
MU	0	41.1	36.4	53	55.2	34	0	5.5	0	14.2	0	5.5
BT	0	46.8	12.5	48.1	44.7	32.9	0	0	0	17.3	12.5	0

Table 3.29 reveals that there was a correlation between values if the object classes were conversely considered. For instance, the probability of finding a *keyboard left-of* of a *mouse* was 88%. In this case, a keyboard was the target and the mouse the reference object. In turn, by changing the roles of the objects in the same relation, the probability of finding a *mouse left-of* of the *keyboard* was only 3.8%. Furthermore, and similar to other spatial relations, the occurrence of the target object impacted the resulting relation value. Table 3.29 reveals that a bottle was located to the *left-of* a keyboard with 44.7% probability. By considering the results oppositely, one can observe that the keyboard was located with 14.01% probability to the *left-of* the bottle. However, by taking into account the occurrence values from Table 3.30, the bottle occurred only one time in the data, whereas the keyboard was present 26 times. This difference in the occurrences caused the discrepancy in the probabilities.

The learned average probability for the *right-of* relation is provided in Table 3.31. When considering the results, it is striking that a mouse was more likely than a mug

Table 3.30: Target object occurrences vs. the number of valid *left-of* relations between the given objects.

	#	CB	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CI	1	0	1	1	1	1	1	0	1	0	0	1	0
FL	26	1	0	8	14	15	14	1	23	1	7	10	4
WL	12	0	4	0	6	5	4	1	7	0	2	1	2
TA	26	1	12	6	0	14	9	1	25	0	6	8	4
KB	26	1	10	6	11	0	8	1	24	0	6	7	4
MT	26	1	6	5	7	7	0	1	19	0	6	4	4
BO	1	0	0	0	0	0	0	0	0	0	1	0	0
MO	26	1	2	5	0	1	1	1	0	0	5	5	4
NB	1	0	0	0	1	1	1	0	1	0	0	0	0
PH	10	0	3	1	4	4	4	0	0	0	0	2	1
MU	18	0	8	7	10	11	7	0	1	0	3	0	1
BT	8	0	4	1	4	4	3	0	0	0	2	1	0

(88%) to be found *right* from keyboard. In contrast, and with regards to the Table 3.29, the probability of finding the mug *left-of* the keyboard was higher than *right-of* it, that is 55.2% and 35.7%, respectively. Such result could be used not only for finding the most probable object locations but also to enhance an object recognition process [GH17], [GK15]. By having such knowledge, the probability of a given object belonging to a certain object class could be adjusted. However, such clear assignment to one of these relations can not be observed for each object classes. For instance, for the object class bottle, it is apparent that a bottle was located to the *left* and to the *right* of the keyboard with almost the same probability, 44.7% and 46.7% respectively.

Table 3.31: Learned average probabilities for objects in the spatial relation *right-of* (given in percentages).

	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	0	100	0	100	100	100	0	100	0	0	0	0
FL	3.8	0	14.3	45.1	36.7	22.1	0	7.5	0	10.5	28.4	14.4
WL	8.3	63.6	0	48.2	48.3	40.7	0	39.4	0	8.3	54.6	8.3
TA	3.8	51.3	22.2	0	41.3	26.1	0	0	3.8	14.7	36.7	14.8
KB	3.8	53.7	17.2	51.9	0	26.5	0	3.8	3.8	14.7	38.2	13.7
MT	3.8	50.4	14.2	33.7	29.7	0	0	3.8	3.8	13.1	23.5	10.1
BO	0	100	100	100	100	100	0	100	0	0	0	0
MO	3.8	80.1	24.1	92	88	67.8	0	0	3.8	0	3.8	0
NB	0	100	0	0	0	0	0	0	0	0	0	0
PH	0	65.5	19.7	59.4	56.2	55.5	10	45.7	0	0	25.6	13.9
MU	5.5	51.3	5.5	41.6	35.7	21	0	24.2	0	10.2	0	5.5
BT	0	46.9	19.2	46.7	46.1	44.2	0	42.9	0	12.5	12.5	0

Table 3.32: Target object occurrences vs. the number of valid *right-of* relations between the given objects.

	#	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	1	0	1	0	1	1	1	0	1	0	0	0	0
FL	26	1	0	4	12	10	6	0	2	0	3	8	4
WL	12	1	8	0	6	6	5	0	5	0	1	7	1
TA	26	1	14	6	0	11	7	0	0	1	4	10	4
KB	26	1	15	5	14	0	7	0	1	1	4	11	4
MT	26	1	14	4	9	8	0	0	1	1	4	7	3
BO	1	0	1	1	1	1	1	0	1	0	0	0	0
MO	26	1	23	7	25	24	19	0	0	1	0	1	0
NB	1	0	1	0	0	0	0	0	0	0	0	0	0
PH	10	0	7	2	6	6	6	1	5	0	0	3	2
MU	18	1	10	1	8	7	4	0	5	0	2	0	1
BT	8	0	4	2	4	4	4	0	4	0	1	1	0

KTH data sets The results of learning the projective spatial relations *left-of* and *right-of* from the KTH data set are provided in Tables 3.33, 3.34, 3.35, and 3.36. These results demonstrate that although on an average the values were similar to those obtained from the DFKI-RIC data set, some differences in the spatial arrangement of the objects can be observed. For instance, the probability of finding a *mug left-of* a *keyboard* was lower than to the *right*. By comparing this result with the result from the DFKI-RIC data set, the opposite is the case because the mug was much more likely to be located *left-of* the keyboard than *right-of* it. In contrast, by considering Tables 3.31, 3.29 3.35, 3.36, 3.33, and 3.34, the probability of finding a mouse *right-of* a keyboard learned from both data sets was much higher than *left-of* it.

Furthermore, by comparing the results from both data sets, it is also striking that the spatial distribution of the object phone on a table differed depending on the data source. Thus, from the KTH data sets, the phone was most likely to be found *left-of* a keyboard than *right-of* it. In contrast, results from the DFKI-RIC data set indicated that the phone was more likely to be found *right-of* a keyboard than *left-of* it. In this context, it can be argued that the spatial arrangement of objects differed depending on the data source. Nevertheless, some basic office objects such as a keyboard or monitor, as expected, tended to be located nearly at the same position regardless of the data source.

Table 3.33: The first part of learned average probabilities for objects in the spatial relation *left-of* from the KTH data set (given in percentages).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	0	26.4	5.4	72.7	7.6	5.2	24.6	11.9	0.9	0	1.7	2.7	12.4	9.4
MT	18.3	0	5	51.9	7.4	7.8	18.5	10.1	2.1	0	2.3	2	12.9	6
BO	52.1	60.1	0	12.4	18.1	0.6	13.1	1.8	3	0	18.1	1.6	17.3	7.9
MO	0.2	0.7	1.9	0	6.3	5.8	21.7	7.6	0.7	0	1.1	0.6	9.1	6.9
NB	59.2	65	7.7	41.9	0	7.1	16.3	4.1	8.4	0	8.4	3.4	18.7	15.6
PH	33.3	23.9	1.9	1.1	12.6	0	0	5.7	3.8	1.1	0	0	19.2	1.1
MU	22.6	21.7	3.3	9.7	7.6	1.5	0	2.5	1.4	0	3.9	0.7	8.8	4.5
BT	9.1	12.7	6	1.6	6.5	4	12.8	0	1	0	1	3.6	13.9	6
CP	44	38.6	2.5	14	9.9	6.3	2.5	2.5	0	0	4.9	0	9.4	10.4
KS	0	100	0	0	100	0	0	0	0	0	0	0	0	0
HP	73.4	76.4	10.2	12.6	16.6	1.7	22.7	2.1	3.3	0	0	8.4	16.6	14.2
PC	40.4	35.6	9.2	24.4	15	12.4	10.9	3.1	5.1	0	5.7	0	14	13.5
PA	26.3	29.1	5.4	7.8	10.9	1.6	10.6	2.6	1.5	0	6.9	1	0	3.9
PN	30.8	31.1	5.4	9.3	10.1	4.8	14.7	3.6	2.9	0	4.7	3.2	9.2	0
HI	26.9	30.4	6.3	14.7	13.5	7.5	15	2.6	0	0	5.2	5	9.1	7.8
MA	30.5	33.7	5.2	7.8	7.4	2.9	9.1	2.2	0	0	10.7	2.9	12.1	4.9
FL	36.9	52.2	9.5	9.1	25.1	19	5.6	1.5	2.4	1.5	10.5	1.5	16	4.4
PS	24	22.5	3.5	15	10.7	17	2.6	12.1	2.6	0	0	0	9	3.6
LP	53.2	59.3	11.4	54.8	8.6	0	36.9	12.8	5.3	0	15.5	2.6	19.8	17.8
FL	50	57.4	4.1	6.2	0	4.6	20.1	11.2	2.8	0	38.7	2.8	17.6	12.1
GL	12.5	23.4	12	6.5	8.7	0	6.2	5.2	0	0	8.4	0	5.8	3.3
JU	48.8	37.3	9.8	23.6	7.5	0	18.3	3.9	0	0	11.8	6.9	7.4	14.4
BA	0	33.3	0	0	0	0	0	0	0	0	0	0	0	0
CU	0	0	0	0	0	0	0	16.6	0	0	0	0	0	0
RU	25.8	20	4.7	11.7	14.6	9.6	7.8	10.4	3.5	0	12.1	10.3	3.5	6.2
EX	0	0	0	0	0	98.4	0	0	0	0	0	0	0	0
NP	43.4	34.8	11.8	18	11.8	1.5	25	10.3	1.5	0	13.9	3.7	19.5	12.4

Table 3.34: The second part of learned average probabilities for objects in the spatial relation *left-of* from the KTH data set (given in percentages).

	HI	MA	FL	PS	LP	FL	GL	JU	BA	CU	RU	EX	NP
KB	4.5	4.9	1.5	13.7	17.8	1.8	2.9	3.6	0.2	1.4	3.3	0	3.8
MT	2	2.8	2.5	17.6	14.2	1.1	2.8	4	0	1.3	3.2	0	3.7
BO	6.7	5.7	3.7	6.6	30.4	1.4	2	3.5	0	0	2.1	0	6.8
MO	2	3	2.1	10.3	5.9	0.7	1.8	3.6	0	1.2	2.1	0	2.5
NB	2.2	2.1	2.8	14.7	38.9	1.8	3	1.3	1.1	0.5	1.8	0	7
PH	1.7	1.9	5.2	20.8	35.1	1.1	0	0	0	0	0	0	3.3
MU	2.9	1.2	1.2	8.2	9.7	0.6	1.3	6	0	0	0.8	0	2.8
BT	3.8	2	1.8	7.6	14.3	4.2	4.8	5.2	0	3.1	1	0	0
CP	2.5	2.5	0	10.5	27.7	4.7	0	0	0	0	0	0	3.4
KS	0	0	0	100	100	0	0	0	0	0	0	0	0
HP	9.2	9.4	6.4	11.6	34.6	3.6	3.5	8	1.3	0	2.5	0	7.6
PC	0	6.1	3.1	14.4	16.8	3.1	0	0	0	3.1	13.6	0	0
PA	2.6	2.3	2.5	4.4	16.6	1.2	1.9	1.8	0	0	1.6	0	4.7
PN	4.2	2.9	1.6	5.1	15.6	1.6	0.8	0.8	0	1.9	4.1	0	3
HI	0	3.3	2.6	0	12	1.3	0	2.6	0	0	0	0	1.3
MA	1.4	0	1.4	4.3	24.3	2.9	1.4	5.7	0	0	0	0	3.7
FL	5.5	3.3	0	32.9	37.9	1.5	2.2	0	0	0	0	0	8
PS	0	0.7	3.4	0	42.8	2.3	3.2	3.6	0	0	0	1.4	4.2
LP	12.2	8.4	3	7.9	0	4.2	3.6	6.7	0	1.4	4.8	0	9.5
FL	13.9	10.1	6.5	5.5	31.8	0	2.8	23.5	0	0	0	0	7.8
GL	0	0	0	15.5	36.9	9.5	0	9.4	0	0	0	0	3.3
JU	17.1	10.3	2	9.1	26.4	2	0	0	0	0	0	0	3.2
BA	0	0	0	0	88.8	0	0	0	0	0	0	0	0
CU	0	0	0	0	0	0	0	0	0	0	0	0	0
RU	0	0	3.5	12.8	2.2	3.5	0	0	0	4.7	0	0	3.5
EX	0	0	0	0	0	0	0	0	0	0	0	0	0
NP	10.5	4.8	8.9	10.1	28.9	2.9	3.8	8.2	0	1.5	6.6	0	0

Table 3.35: The first part of learned average probabilities for objects in the spatial relation *right-of* from the KTH data set (provided in percentages).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	0	20.2	20.7	0.2	28.3	7	12.8	2.1	4.3	0	20.9	3.1	20.5	8.9
MT	24	0	21.6	0.6	28.2	4.5	11.2	2.6	3.4	0.2	19.7	2.5	20.6	8.2
BO	13.8	14.1	0	4.8	9.3	1	4.8	3.5	0.6	0	7.3	1.8	10.6	3.9
MO	72.8	57.4	4.9	0	20.1	0.2	5.5	0.3	1.3	0	3.6	1.9	6.1	2.7
NB	16	17.1	15.1	13.1	0	5.5	9	3.1	2	0.5	9.9	2.4	17.8	6.1
PH	24.7	41.2	1.2	27.5	16.3	0	4.2	4.4	2.9	0	2.3	4.6	6.1	6.7
MU	43.3	36	9.2	38.1	13.7	0	0	5.2	0.4	0	11.4	1.5	14.6	7.5
BT	51.5	48.2	3.1	32.9	8.6	5.2	6.3	0	1	0	2.6	1	9	4.5
CP	10.1	23.7	12.3	7.9	41.2	8.3	8.2	2.5	0	0	9.8	4.1	12	8.6
KS	0	0	0	0	0	100	0	0	0	0	0	0	0	0
HP	5.9	9.2	25.2	4	14.2	0	7.8	0.8	1.6	0	0	1.5	19	4.8
PC	35.6	29.2	8.5	7.9	21	0	5.1	10.8	0	0	30.9	0	10	12
PA	15.9	18.2	8.8	11.7	11.4	5.1	6.4	4.1	1.1	0	6	1.4	0	3.4
PN	32.5	23.1	10.8	23.7	25.7	0.8	8.8	4.8	3.5	0	13.9	3.6	10.6	0
HI	25	12.6	14.7	11.1	5.9	2	9.2	4.9	1.3	0	14.5	0	11.4	6.8
MA	29.6	18.6	13.7	18.1	6.1	2.5	4.2	2.8	1.4	0	16.2	2.9	11.2	5.1
FL	9.6	17.3	9.1	13.1	8.4	6.8	4.2	2.6	0	0	11.4	1.5	12.4	2.9
PS	42.4	60	8.1	31.8	21.7	13.5	14.4	5.4	3.1	0.7	10.2	3.4	10.6	4.6
LP	28.1	24.8	19.1	9.4	29.3	11.6	8.6	5.2	4.2	0.3	15.5	2	20.4	7.1
FL	21.1	14.8	6.7	8.4	10.2	2.8	4.5	11.4	5.4	0	12.3	2.8	11.2	5.5
GL	40.9	42.9	11.1	24.9	19.9	0	10.3	15.1	0	0	13.6	0	20.5	3.3
JU	29.9	36.1	11.6	29.7	5.2	0	27.9	10	0	0	18.9	0	12	2
BA	33.3	0	0	0	77.7	0	0	0	0	0	53.5	0	0	0
CU	99	99.1	0	82.8	16.6	0	0	49.8	0	0	0	16.6	0	39.4
RU	48.8	51.9	12.2	31.9	13.2	0	6.7	3.5	0	0	10.7	15.6	18.2	17.7
EX	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NP	24.5	26.4	17.4	16.4	21.5	4.5	10.2	0	2.1	0	14	0	23.7	5.7

Table 3.36: The second part of learned average probabilities for objects in the spatial relation *right-of* from the KTH data set (given in percentages).

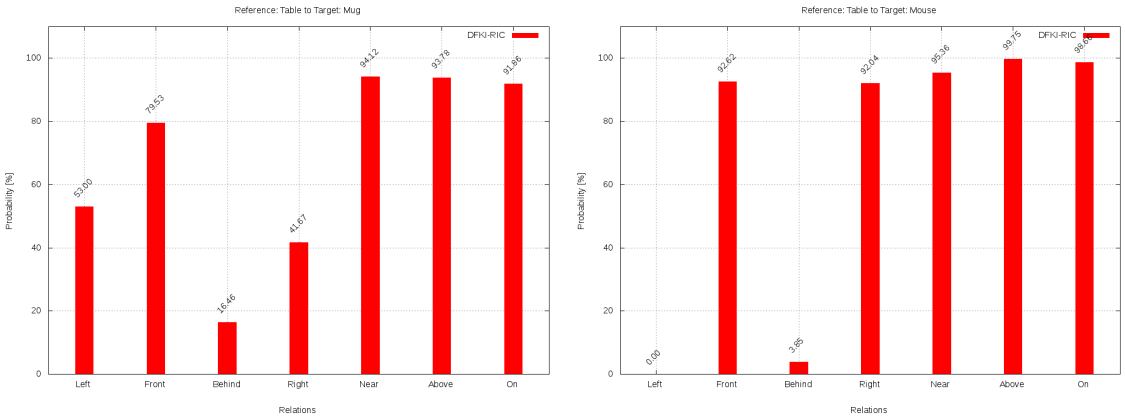
	HI	MA	FL	PS	LP	FL	GL	JU	BA	CU	RU	EX	NP
KB	4.8	5	5.9	7.8	33.7	4.2	0.9	5.9	0	0	1.7	0	6.7
MT	4.9	5	7.6	6.6	34.1	4.4	1.5	4.1	0.2	0	1.2	0	4.9
BO	2.8	2.1	3.8	2.9	18.2	0.8	2.2	3	0	0	0.8	0	4.6
MO	2.6	1.3	1.4	4.9	34.8	0.5	0.4	2.8	0	0	0.8	0	2.8
NB	5.1	2.5	8.4	7.2	11.4	0	1.3	1.9	0	0	2	0	3.8
PH	6.4	2.3	14.6	26.3	0	1.9	0	0	0	0	3.1	2.2	1.1
MU	4.7	2.6	1.6	1.5	41.2	3	0.8	3.9	0	0	0.9	0	6.8
BT	2	1.6	1	17	35	4.1	1.6	2	0	1	3	0	6.9
CP	0	0	3.9	8.8	34.4	2.5	0	0	0	0	2.5	0	2.5
KS	0	0	100	0	0	0	0	0	0	0	0	0	0
HP	3.3	6.2	5.9	0	34.5	11.6	2.1	5	0	0	2.9	0	7.6
PC	11.6	6.2	3.1	0	21.1	3.1	0	10.9	0	0	9	0	7.4
PA	2.1	2.5	3.3	3.7	16.1	1.9	0.5	1.1	0	0	0.3	0	3.9
PN	4.8	2.8	2.4	4.1	38.9	3.5	0.8	6	0	0	1.4	0	6.7
HI	0	1.3	4.9	0	42.9	6.6	0	11.6	0	0	0	0	9
MA	3.6	0	3.2	1.4	32.1	5.2	0	7.5	0	0	0	0	4.5
FL	3	1.5	0	7	11.9	3.4	0	1.5	0	0	1.5	0	8.6
PS	0	2.2	16.3	0	15.5	1.4	3.5	3.4	0	0	2.7	0	4.8
LP	3.4	6.3	9.6	21.8	0	4.2	4.2	5	1	0	0.2	0	7.1
FL	2.8	5.6	2.8	8.7	31.5	0	8.1	2.8	0	0	2.8	0	5.4
GL	0	3.3	4.8	14.1	31.3	3.3	0	0	0	0	0	0	8.3
JU	3.9	7.8	0	9.7	35.1	16.5	5.6	0	0	0	0	0	10.6
BA	0	0	0	0	0	0	0	0	0	0	0	0	0
CU	0	0	0	0	64.3	0	0	0	0	0	21.9	0	16.6
RU	0	0	0	0	45.2	0	0	0	0	0	0	0	15.1
EX	0	0	0	97.7	0	0	0	0	0	0	0	0	0
NP	1.5	3.9	8.3	8.9	38.5	4.2	1.5	2.5	0	0	1.5	0	0

3.1.3 Comparing the learned PQSR per object pairs (based on the DFKI-RIC and KTH data sets)

The aim of this evaluation is to compare the results from the learning of the PQSR from both data sets considering only the two object classes and all spatial relations that held between them. This enables analysis of whether and in what manner the spatial arrangement of objects with respect to the PQSR changed depending on the data source. Furthermore, based on this evaluation, the most probable spatial relation for a given target and reference object can be identified. As in the previous Section 3.1.2, all spatial relations and the given objects are presented. In this section, only two object classes are compared, and in this way, a clearer presentation of the learning results can be achieved.

In the following graphs, the average probabilities of finding two object classes in all known spatial relations are presented. The values were calculated based on the same formulas used in the experiments discussed in the previous Section. The horizontal axis of the graph represents the given spatial relation, and the vertical its average probability value. The numbers above the bars denote the exact learned average probability of the given relation.

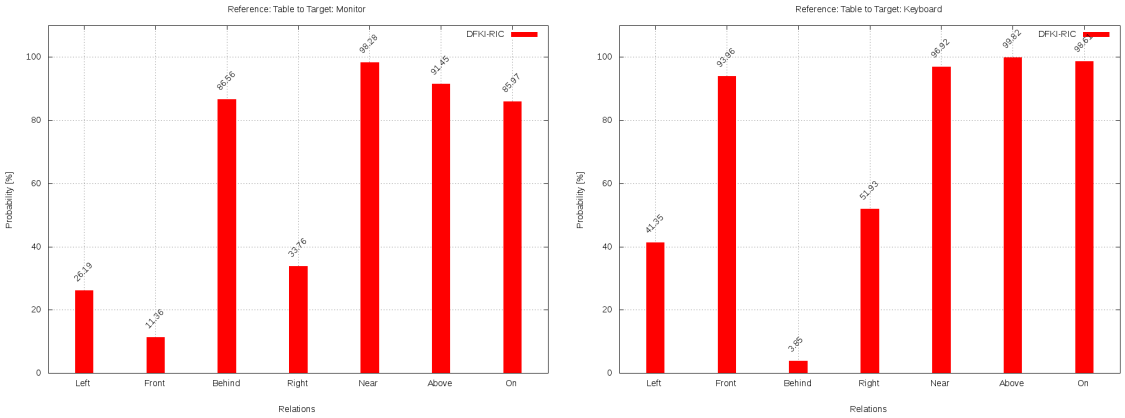
In this evaluation, six object classes, such as table, mug, mouse, monitor, keyboard, and phone are compared. The red boxes present the results obtained from the DFKI-RIC data set and green from the KTH data set. Because the object class table in the KTH data set has not been annotated, only the results from the DFKI-RIC data set for this object are presented.



(a) (b)

Figure 3.13: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *table* and the target objects are (a) a *mug* and (b) a *mouse*.

By considering the results referring to a table as a reference object, some regularities according to the objects' arrangements can be observed, which were true for all considered object classes. For instance, Figures 3.13-3.25 reveal that a mug, mouse, monitor, keyboard, phone, and bottle were more likely to be found *on*, *near* and *above* a table



(a) (b)

Figure 3.14: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *table* and the target objects are (a) a *monitor* and (b) a *keyboard*.

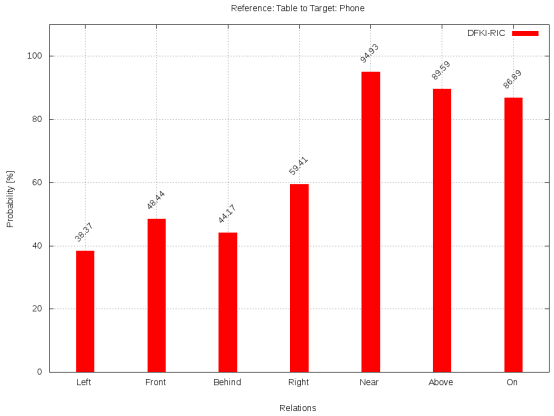


Figure 3.15: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *table* and the target object is a *phone*.

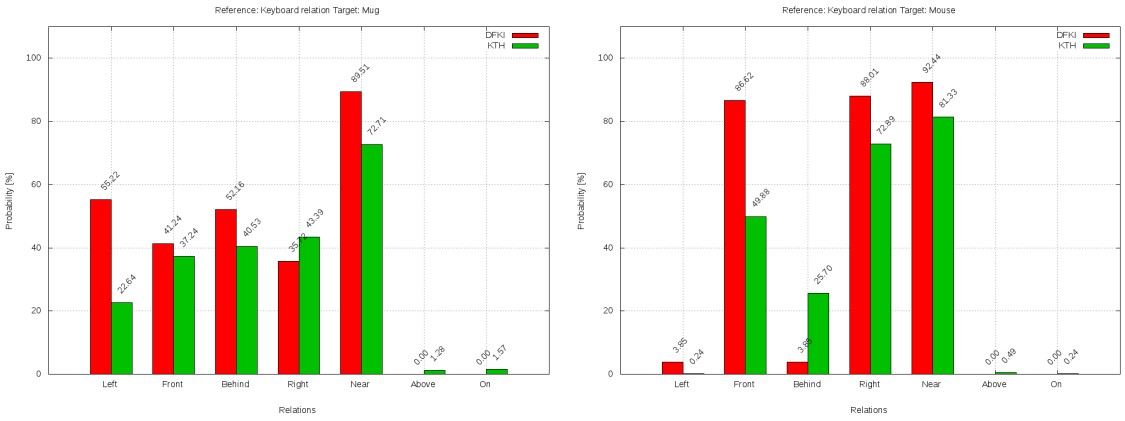
with probabilities of more than 85%. As expected, a mouse was significantly more likely, on average, to be located in the *front* area of the table [2] than *behind* it. Interestingly, the mouse could only be found in the *right* part of the table. This result demonstrates that, according to the data, there were clear user preferences with respect to the mouse's placement on the table. In turn, a mug could also be found in *in-front-of* a table, but the average probability of finding it in the *left* area of the table was higher than in the *right*

[2] It should be noted that, according to Def. 10 and Def. 11, the projective binary relations refer to the CoG of the objects. Therefore, the front of the area of a table was located in the front part from the CoG of the table to its edge.

area. This knowledge could be used to reduce the search space for a given object and, therefore, enhance the search process.

According to both data sets, a monitor and a keyboard had the tendency to be located more in the *right* area of the table, though, a keyboard had a higher probability than the monitor. However, the main difference between the location of these two objects relative to a table refers to the *behind* and *in-front-of* relations. Thus, a monitor was located *behind* the table with 86% probability, and conversely, a keyboard was likely to be located *in-front-of* the table with 93% probability. The spatial distribution of the phone relative to the table was relatively balanced, as no strong tendency for the *right-of*, *left-of*, *in-front-of*, and *behind-of* relations were observed.

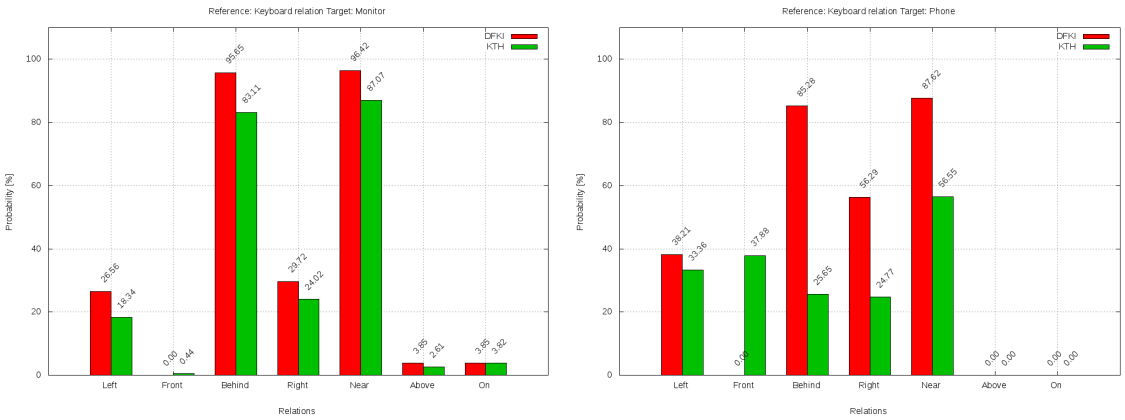
By considering the keyboard as a reference object, the most probable relation between it and the remaining target objects was the *near* relation. This finding was true for both results based on the DFKI-RIC and KTH data sets. However, from the left Figure 3.16, it can be observed that with respect to the DFKI-RIC data set, the mug was more likely to be located to the *left-of* a keyboard, whereas referring to the KTH data set results, the mug was more likely to be located to the *right-of* a keyboard. This result is quite surprising, as both data sets refer to the same environment type (an office environment). Therefore, this result indicates that there are differences in the object's arrangements depending on the data source. Nevertheless, even if there are some differences, in general, the placement of objects did not vary much in both data sets. For instance, the probability of a mug being *behind* a keyboard was higher than it being *in-front-of* the keyboard. Thereby, the difference in the probability values between the *in-front-of* and *behind-of* relations was only 10.02% and 3.92% for the DFKI-RIC and KTH results, respectively.



(a) (b)

Figure 3.16: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *keyboard* and the target objects are (a) a *mug* and (b) a *mouse*.

Interestingly, results obtained from both data sets reveal that an object *mouse* is located significantly more probable to the *right* and *in-front-of* the keyboard than *left* and *behind* it, as shown in Figure 3.16. In turn, from Figure 3.17, it can be observed that the *monitor*



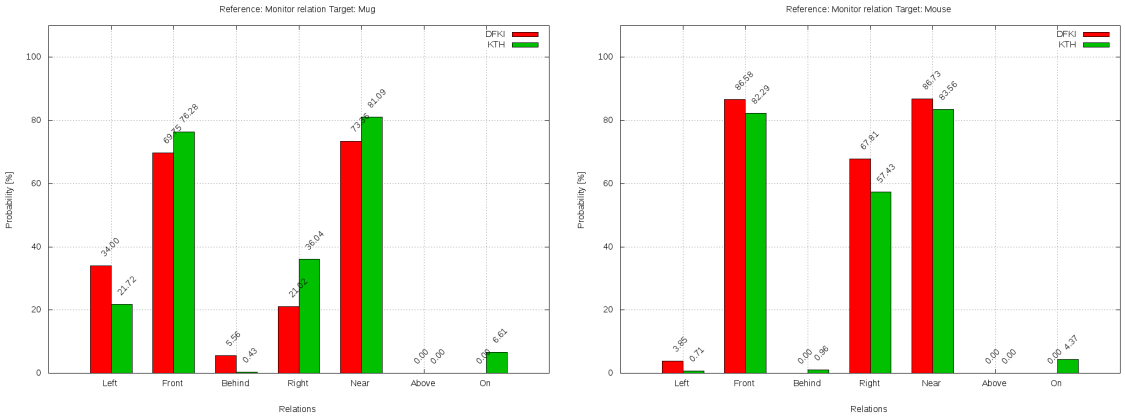
(a) (b)

Figure 3.17: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *keyboard* and the target objects are (a) a *monitor* and (b) a *phone*.

was more likely to be located *behind* and to the *right-of* the keyboard. Moreover, there was a small probability of finding the monitor *above* and *on* the keyboard. As discussed in the previous section, these unlikely results were caused by the scene in which the keyboard was located nearly under the monitor. Therefore, the relations on and above held between these two objects. From the Figure 3.17(a), it is also apparent that the probability values learned from both data sets are similar. Conversely, the probability of finding a phone more *left* or *right* to the keyboard, differed depending on the corresponding data set. Thus, based on the results from the DFKI-RIC data, a phone was more likely to be located *behind* and to the *right-of* the keyboard. In contrast, from the KTH data set, it was observed that the phone was more likely to be located in *front* and to the *left-of* the keyboard.

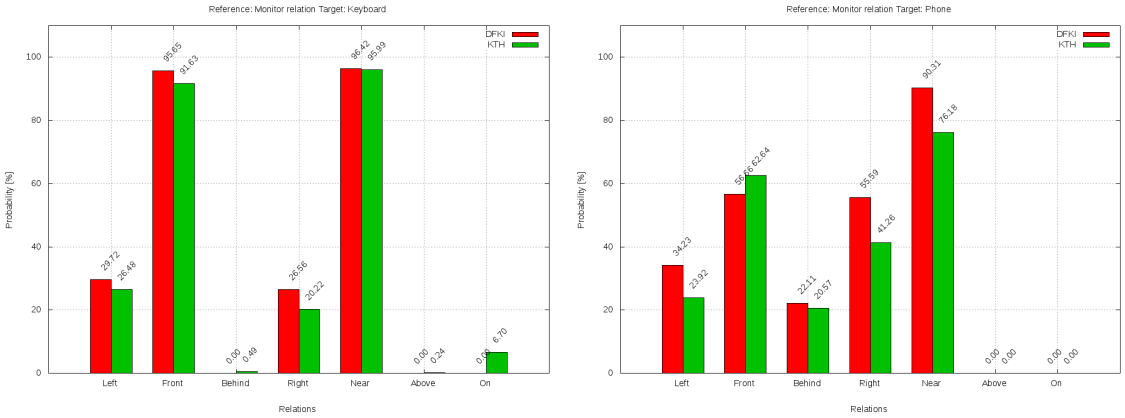
When considering the results with the *monitor* as a reference object based on both data sets, it was apparent that most target objects were located *near* and *in-front-of* it. The results indicate, as illustrated in Figures 3.18-3.19, that the monitor was most likely to be located farther *behind* compared to mug, mouse, keyboard and phone. With respect to the DFKI-RIC data set, the object *mug* was located more likely *left* than *right-of* the monitor. In contrast, in the results from KTH data set, the opposite was the case, as the mug was more likely to be located to the *right-of* a monitor. In Figure 3.18(b), there is a clear trend of finding the *mouse right-of* a monitor in both data sets. Interestingly, Figure 3.19 illustrates that the *keyboard* was located to the *left-of* a monitor whereas the *phone* was located to the *right-of* the monitor, and this was the case in both data sets.

In Figures 3.20 and 3.21, results from the DFKI-RIC and KTH data sets, in which a *mouse* was selected as reference object, are presented. Figure 3.20 illustrates the spatial distribution of the *mug* and *monitor* relative to the *mouse*. This figure reveals that the average probability of finding a mug *near* and *right-of* the mouse was the highest, although the overall probability dis not exceed 52%. The fact that a mug was located to the *right-of* a mouse may appear to be incorrect because, for instance, in results from the DFKI-RIC



(a) (b)

Figure 3.18: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *monitor* and the target objects are (a) a *mug* and (b) a *mouse*.



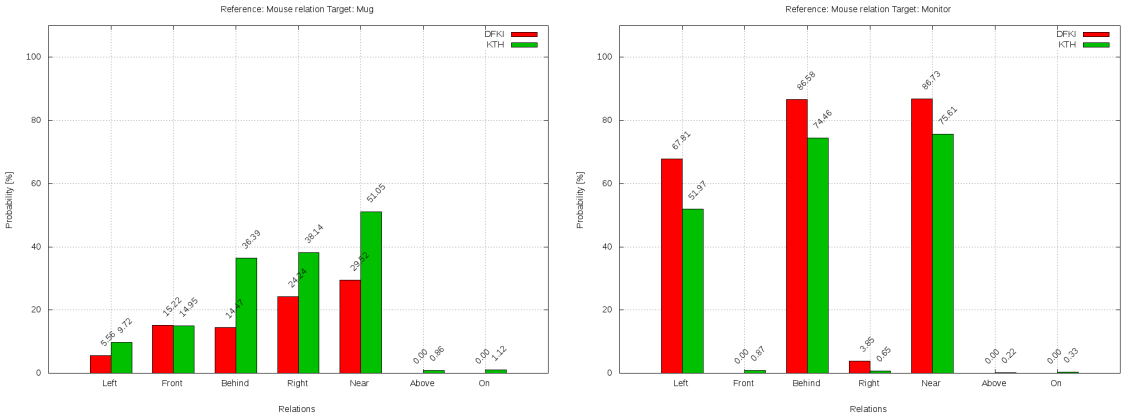
(a) (b)

Figure 3.19: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *monitor* and the target objects are (a) a *keyboard* and (b) a *phone*.

data set, the probability of finding a mug *left-of* keyboard was higher than *right-of* it but at the same time, a mouse was more likely to be located to the *right-of* keyboard. However, because the results in the Figure 3.20(a), only the objects mug and mouse are considered, therefore, this result is valid. As the mug was most likely to be located to the *left-of* keyboard and a mouse to the *right*, the mugs were too far away from the mouse and thus, were not being considered for the relation. As a result, the mugs located within the maximum allowed distance to the mouse were more likely to be to the *right-of* it. This

exemplary result indicates that a spatial reasoning about object's placement can not be directly performed, since only two objects were considered in a certain spatial relation.

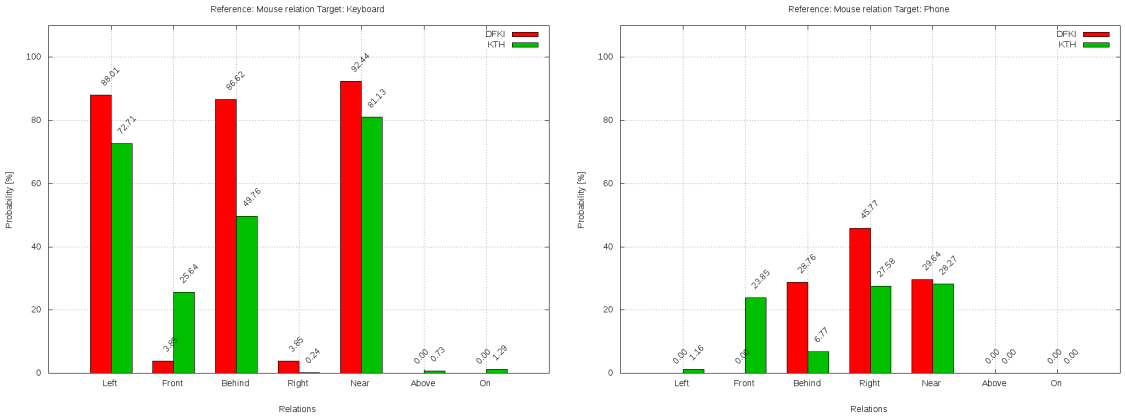
By considering the Figure 3.20(b), it can be observed that from both data, the *left-of* and *behind-of* relations for a *monitor* relative to *mouse* held most probable. By comparing these results with those from Figure 3.18, the spatial correlation between the projective relations can be observed. As described previously, the mouse was most likely located to the *right* and *in-front-of* the monitor, whereas the monitor was located *behind* and to the *left-of* the mouse. This result demonstrates the spatial correlation of the projective relations. The same correlation can be observed in the results for the *keyboard* and *mouse*. For the keyboard relative to mouse, the probabilities of the *left* and *behind* relations were the second highest (after the *near* relation) and this was the case for both data sets. In contrast, although in both data sets the object *phone* was more likely to be found to the *right-of* a keyboard, in the DFKI-RIC data, it was more likely located, on average, *behind-of* and *in-front-of* of the mouse in the KTH data set.



(a) (b)

Figure 3.20: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *mouse* and the target objects are (a) a *mug* and (b) a *monitor*.

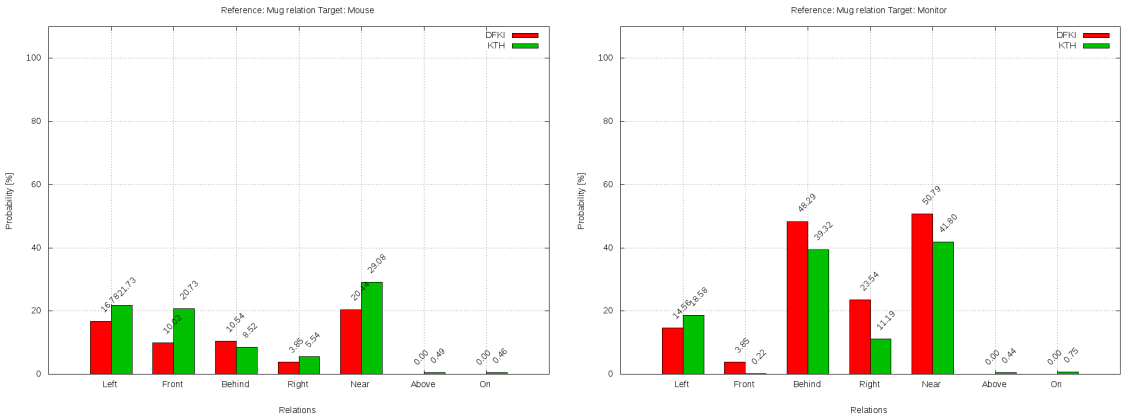
Figures 3.22 and 3.23 present the spatial distribution of target objects relative to the reference object *mug* with respect to the DFKI-RIC and KTH data sets. Figure 3.22, based on the results from the DFKI-RIC data set, reveals that a *mouse* could most likely be found to the *left* and *behind* a mug. In comparison, from the KTH data set, it was demonstrated that a *mouse* was also most likely to be located *left-of* the mug and more to its *front* as *behind* it. In the Figure 3.22(b), the spatial distribution of the *monitor* relative to the *mug* is presented. The results illustrate that the monitor was most likely to be located *behind* the mug, with the difference in the DFKI-RIC data set being that the monitor was more likely to be found *right-of* the mug and, according to KTH data, to the *left-of*. Additionally, the *keyboard* was more likely to be placed *in-front-of* mug. However, from the KTH data set, results indicate that the *keyboard* was more likely to be *left*, and from the DFKI-RIC data set, to the *right-of* the mug. Regarding the objects



(a) (b)

Figure 3.21: Learned average probability values for all spatial relations with a given target and reference object. Thereby, the reference object is a *mouse* and the target objects are (a) a *keyboard* and (b) a *phone*.

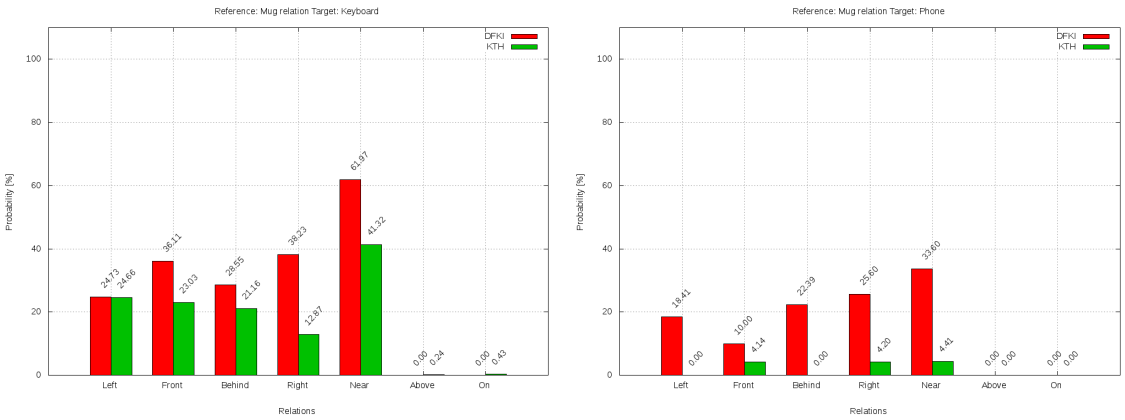
phone and *mug*, it is interesting to note that in the KTH data set, the phone was not observed located either to the *left* or *behind* the mug. In turn, the learned probabilities from the DFKI-RIC data set were the highest for the *right* and *behind* relations.



(a) (b)

Figure 3.22: Learned average probability values for all spatial relations with a given target and reference object. Thereby the reference object is a *mug* and the target objects are (a) a *mouse* and (b) a *monitor*.

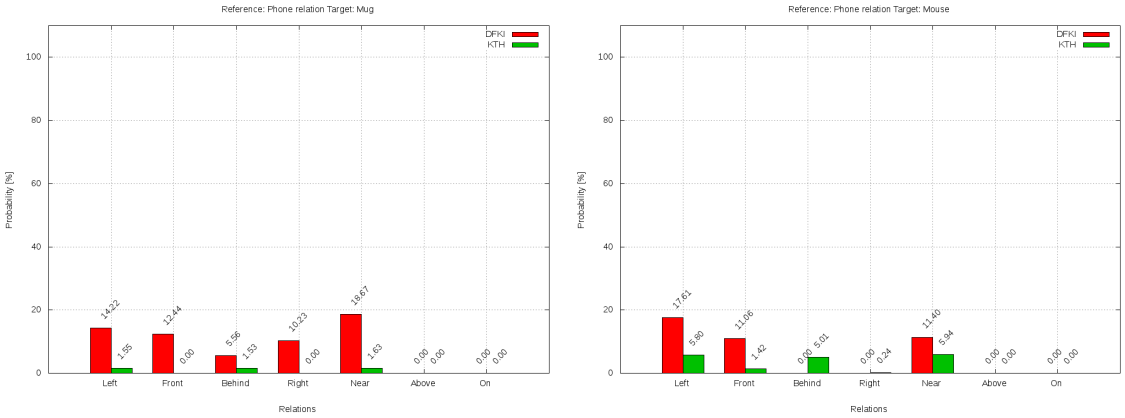
By considering the reference object *phone*, it can be observed that the resulting probability values were much smaller than in the cases of other object pairs. Moreover, the resulting values of the KTH data set were generally small, which indicates that it was



(a) (b)

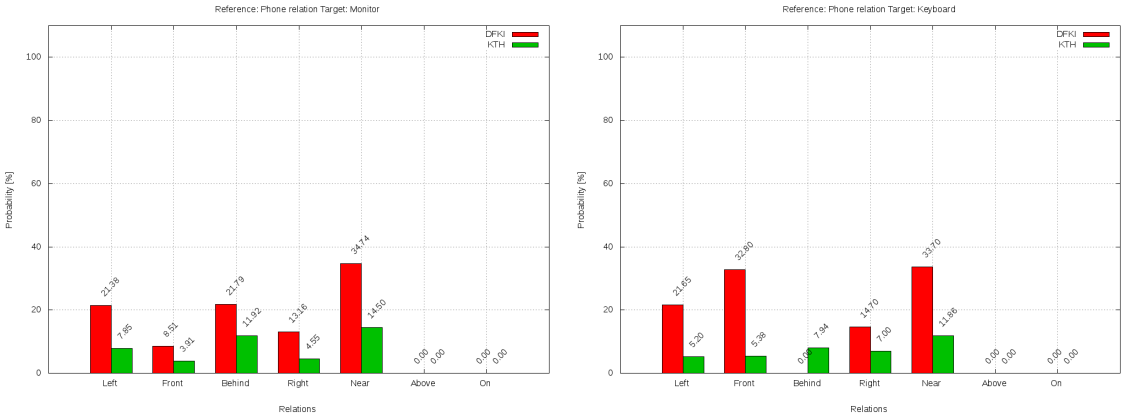
Figure 3.23: Learned average probability values for all spatial relations with a given target and reference object. Thereby the reference object is a *mug* and the target objects are (a) a *keyboard* and (b) a *phone*.

not likely that a phone would be found in the given spatial relations with other objects. Although the probability values were small, some relations were still more probable than others. For instance, based on the results obtained from the DFKI-RIC data set, a *mug* was more likely to be located to the *left-of* the phone and to its *front*. In contrast, based on the KTH data set, the *mug* was located more to the *in-front-of* a phone and *behind* it. The same relation distribution was true for the *mouse* and the reference object *phone*. In instances of the *monitor*, there was a tendency to find the monitor to the *left* and *behind* the phone, as illustrated in Figure 3.25. For the *keyboard* and reference object *phone*, the *near*, *in-front-of*, and *left-of* relations held most probable, whereas in the KTH data set, the keyboard was placed more *behind* and to the *right-of* the phone.



(a) (b)

Figure 3.24: Learned average probability values for all spatial relations with a given target and reference object. Thereby the reference object is a *phone* and the target objects are (a) a *mug* and (b) a *mouse*.



(a) (b)

Figure 3.25: Learned average probability values for all spatial relations with a given target and reference object. Thereby the reference object is a *phone* and the target objects are (a) a *monitor* and (b) a *keyboard*.

3.1.4 Summary of the experiments for PQSR learning

In Sections 3.1.2 and 3.1.3, the results from the learning of the PQSR were presented. To evaluate the change in the PQSR according to different data sets, the DFKI-RIC and the external KTH data were used. This enabled examination of how the different data sources influence the resulting relations and their probabilities. In summary, the results revealed that in both data sets, there were regularities in the spatial arrangement of objects. Fur-

thermore, by analyzing the results, the correctness of the formal definitions of the spatial relation and its properties were also verified. The overall results demonstrate that the PQSR were learned properly according the their formal definitions. Taken together, the results meet the expectations that, even if some objects are placed differently depending on the user's preferences, the are some basic office object classes that underlie strict placement order. These regularities were observed in both the DFKI-RIC and external KTH data sets. However, some crucial aspects of learning the PQSR must be emphasized:

Number of object instances of an object class in the data As demonstrated in the experiments, if only one target object instance of a given object class exits in the data, a 100% probability for all valid spatial relations referring to this object is calculated. By considering only one object for a given relation and the previously learned distance value for this relation, the object matched the learned distance perfectly. Therefore, in such cases, the relation holds with 100% probability. A possible improvement of this outcome is discussed in Section 4.1.

Number of valid relations in the data vs. the probability values of the relations The number of valid spatial relations in the data influenced the resulting average probability more than the probability value of the particular relations itself. For instance, although the probabilities for a certain spatial relation were generally high, this relation did not hold in every scene. Thus, the resulting average probability could become lower than in cases where the relation held in every scene but with lower probability. In Section 4.1, some suggestions for improvement of this issue are provided.

Segmentation failure and valid relations According to the results obtained in this work, the resulting average probabilities did not meet expectations, in some of the cases, regarding typical object placement in an office scene. For instance, according to the learning results the object monitor was located on a keyboard, which is an unlikely position for a monitor. However, as previously described, this result is correct according the formal definitions of the on relation. A further result with this issue was that an object lamp was located above some small objects such as a pen or highlighter. This result was caused by inaccuracies in the segmentation of the object lamp.

Redundancy of on and above relations Considering the results of the on relation, it could be argued that some of the results are unexpected and somewhat redundant to the results for the above relation. However, the relations are redundant only if the reference object is horizontally aligned. The on relation is intended to specify that an object can be located on another object even if it is not oriented horizontally, for instance, like a picture on the wall. Unfortunately, due to the type of data used, this case could not be learned, as this arrangement did not occur in the scenes. Moreover, further experiments in 3.2.3 demonstrate the positive influence of the on relation on the overall target object position estimation.

Additional data from different domains Although the results of the experiments reveal that the PQSR were learned properly, additional data would make the results less over-

trained. Furthermore, the learned knowledge would become more general and not so domain related. Nevertheless, the results demonstrate that even in a small set of data there are regularities and differences in the PQSR learned from the KTH and DFKI-RIC data sets. This finding refers, for instance, to objects such as the mug, which was more likely to be found left or right to the keyboard depending on the data source. This observation indicates that additional domain-specific data are needed to obtain precise results for specific object arrangements.

3.2 Experiments related to FI and MFI

This section describes the experiments related to the Field Intensity and Maximum Field Intensity evaluation. As discussed in Section 2.4, the FI value contains knowledge about the PQSR and can be used to improve or support the search for a sought after object. To evaluate the usability of the theoretical FI approach, several test were preformed. These tests were performed under considerations of the different aspects of the data, such as single view vs. merged scans or real vs. artificial data. The real-world data included the DFKI-RIC and KTH data sets. In addition, a merged scan of a whole office room was used to evaluate the FI-based search for a target object in such a large scene.

Section 3.2.1 describes experiments performed on single scans. These scans focused on the area of a table desk and its immediate surroundings. Within this Section 3.2.1, the results of distance-based evaluation are presented in Section 3.2.1.1, followed by experiments in which the change in probability values was analyzed in a context of different sought after object positions in Section 3.2.1.2. Section 3.2.2 presents the results after different grid resolutions were applied. In Section 3.2.3, the influence of removing different spatial relations from the FI calculation is presented. The results related to the single scans are summarized and concluded in Section 3.2.1.3.

The second part of the experiments in Section 3.2.4 addresses evaluation of the FI-based method in the context of merged scans that contain several scenes. In Section 3.2.4.1, the results of a distance-based FI-evaluation performed on a real-data are provided and Section 3.2.4.2 presents the corresponding evaluation results performed on artificially created data. Finally, Section 3.2.5 provides a summary of the overall experiments.

3.2.1 Single view scenes

In the following experiments, real-world data consisting of single office scenes is used. Each scene includes a view of a table desk and its surrounding. On the table, typical office objects such as a monitor, keyboard, and mouse are placed. The arrangement of objects depended on the preferences of the researcher preferences sitting in the given office room. In some scenes, building structures such as the floor and walls are partly present. The acquired scans consist of non-merged point clouds taken from one perspective using the setup displayed in Figure 3.1.

3.2.1.1 Distance-based FI and MFI evaluation

To analyze the precision of the position prediction based on the FI approach, the distance between the real position of the target object and its predicted position resulting from

the MFI was evaluated. For better measurement of the distance values, the results were compared with the expected average failure resulting from the average distance from the sought after object's position to all cells of the grid. Even if there were general expectations that the method performed better than the average failure, the average failure in the distance enabled the performance of the FI approach to be estimated.

Thereby, the average failure approach referred to an average distance between the actual position where the sought after object was located and all cells from the grid. Therefore, the resulting value is an expected failure and can be obtained by applying the formula:

$$\frac{\sum_{i=0}^{(\rho_x/\omega_x)(\rho_y/\omega_y)(\rho_z/\omega_z)} |p - \begin{pmatrix} f_x(i) \\ f_y(i) \\ f_z(i) \end{pmatrix}|}{(\rho_x/\omega_x)(\rho_y/\omega_y)(\rho_z/\omega_z)} \tag{3.1}$$

Results based on the DFKI-RIC data set In Table 3.37, the results of the average distance calculation are presented. As previously discussed, the results obtained are not intended to serve as a benchmark for the evaluation of the FI-based method, but rather a measurement of the expected failure. The scene number with the corresponding object's instances are listed in the rows of the table. The columns list the distance between the average position to all cells and the real position of the given object. In these experiments, four object's classes were taken into account: keyboard, monitor, mouse, and table. The values specified how far away an object was from an average position obtained from all cells' positions. Because in some scenes more than one object instance of a given object class was present, the additional object was also considered by the evaluation. These scenes are marked by a sub-number related to the given scene, for instance 3.1. The minus indicate that there was no further object of this particular object class in the given scene.

As illustrated by Table 3.37, the deviation between the real and average position for the *keyboard* ranged from 0.9 to 3.03 meters. In turn, the average deviation for the object class keyboard was 1.89 meters. Values for the object *monitor* reveal that the deviation was between 1.0 and 2.96 meters. The average deviation value for the monitor was 1.89 meters, which was similar to the keyboard.

The minimum distance between the average and real *mouse* position was 0.94 meters, whereas the maximum distance to the mouse was 3.20 meters. On average, the position deviated from the real value by 1.95 meters. The values for the largest object class *table* show that the distances ranged from 0.88 to 2.89 meters. The average distance for the object table was 1.82 meters. Overall, the distance between the considered objects and all grid's cells was on average about 1.9 meters, which is notably high.

Table 3.37: Distances between real sought after object's position and an average distance from its CoG to all cells of the grid obtained from the DFKI-RIC data (given in meters).

Scene	Dist. Keyboard	Dist. Monitor	Dist. Mouse	Dist. Table
1	2.44234	2.42419	2.54584	2.5092
2	1.85295	1.79688	1.91473	1.81823
3	2.20177	2.13088	2.36636	2.01904
3.1	-	2.00388	-	-
4	3.03948	2.96427	3.20168	2.88936
4.1	-	2.88805	-	-
5	1.61162	1.61017	1.70816	1.5706
6	0.903571	1.02345	0.942635	0.889846
6.1	-	1.01357	-	-
7	1.91063	1.82631	2.02769	1.8489
8	1.00758	1.04611	1.07185	0.978824
9	1.61326	1.57777	1.67842	1.57194
9.1	-	1.63703	-	-
10	2.19934	2.14965	2.2911	2.13335
11	2.55562	2.36623	2.57323	2.50974
11.1	-	2.53318	-	-
12	1.21439	1.34294	1.16665	1.17894
13	2.79294	2.67135	2.89208	2.73965
13.1	2.83215	2.67588	2.78977	-
14	1.63032	1.65688	1.6168	1.64047
15	2.23653	2.18297	2.36773	2.17333
16	2.74832	2.66389	2.78715	2.64831
16.1	-	2.54951	-	-
17	1.69652	1.65623	1.62952	1.72545
17.1	-	1.59543	-	-
18	1.22096	1.29495	1.24024	1.21305
18.1	-	1.37395	-	-
19	1.25669	1.44901	1.344	1.24599
19.1	-	1.35985	-	-
19.2	-	1.32794	-	-
20	2.95252	2.77928	3.01635	2.89829
20.1	-	2.79334	-	-
21	1.11286	1.1275	1.15419	1.1178
22	2.03367	2.06974	2.12326	2.08427
23	1.1681	1.23368	1.21469	1.15053
24	2.27605	2.05862	2.27629	2.18765
24.1	-	2.1507	-	-
25	1.25286	1.28086	1.34325	1.2062
26	1.4899	1.55622	1.49122	1.47105
Average	1.89826	1.89339	1.95463	1.82385

Table 3.38: Distances between real target object's position and its predicted position based on the MFI-method based on the DFKI-RIC data (given in meters).

Scene	Dist. Keyboard	Dist. Monitor	Dist. Mouse	Dist. Table
1	0.0607578	0.12695	0.0842961	0.445023
2	0.0528113	0.0748308	0.0870983	0.147316
3	0.0273771	0.0782597	0.135258	0.42351
3.1	-	0.599785	-	-
4	0.117317	0.145782	0.0424162	0.353155
4.1	-	0.755629	-	-
5	0.0706204	0.176787	0.147424	0.136928
6	0.126789	0.409818	0.117169	0.0814048
6.1	-	0.255323	-	-
7	0.0639314	0.19734	0.157997	0.140313
8	0.0926369	0.137827	0.1322	0.140996
9	0.0234388	0.0550485	0.0607488	0.0876882
9.1	-	0.593149	-	-
10	0.056185	0.0973032	0.0288848	0.0903582
11	0.0859322	0.133876	0.0897404	0.170426
11.1	-	0.612002	-	-
12	0.0476658	0.248899	0.0662121	0.338555
13	0.0278045	0.0779956	0.244378	0.329199
13.1	0.775217	0.70654	0.543443	-
14	0.02727	0.0710752	0.0181803	0.0791559
15	0.071108	0.277966	0.155763	0.460509
16	0.417783	0.779592	0.73847	0.354047
16.1	-	0.264719	-	-
17	0.090066	0.0880812	0.107866	0.100493
17.1	-	0.760922	-	-
18	0.115873	0.0387125	0.059151	0.169385
18.1	-	0.63167	-	-
19	0.0441124	0.41634	0.0391635	0.0632642
19.1	-	0.0515904	-	-
19.2	-	0.617139	-	-
20	0.0816147	0.229584	0.104618	0.264905
20.1	-	0.393544	-	-
21	0.0868299	0.15314	0.0841239	0.17679
22	0.0409902	0.175313	0.107748	0.393273
23	0.101748	0.225087	0.138591	0.396133
24	0.158006	0.245244	0.22724	0.148385
24.1	-	0.408547	-	-
25	0.279492	0.122724	0.0524564	0.422433
26	0.0474388	0.347512	0.0762752	0.389348
Average	0.118178	0.302094	0.142478	0.242423

Table 3.38 provides the results for the deviation between the real position of a given object and the predicted position based on the MFI method. From the table, it can be observed that the average distance value ranged from 0.11 meters in the case of the *keyboard* and up to 0.30 meters in cases of the *monitor*. Moreover, and as expected, the values were significantly smaller than those listed in Table 3.37. By considering the first object, the keyboard, it can be observed that the values ranged from 0.02 to 0.77 meters. In this context, the highest distance value was smaller than the smallest value for the keyboard as illustrated in Table 3.37. The same finding was also true for all the remaining objects. For instance, the average deviation of the object monitor was 0.30 meters. This value was also smaller than the minimum distance given by the average cell distance, 0.30 and 1.0 meters, respectively. The results from Table 3.38 demonstrate that the predicted position for a *mouse* was, on average, close to its real position. Moreover, the highest value predicted for a mouse based on the MFI method was smaller than the smallest value obtained from the average cell calculation. The average distance for a mouse, based on the MFI, was 0.14 meters. By comparing the results from both tables and the object class *table*, it is revealed that the values resulting from the MFI calculation were, as expected, considerably smaller that those listed in Table 3.37, as the average distance for the table was 0.24 meters.

The results listed in Table 3.38 reveal that the variance between the object's real and predicted position did not exceed 0.30 meters. In contrast, by considering the average distance over all cells the resulting deviation value was about 1.9 meters. Together, these results indicate that the MFI-based position prediction provides promising results towards the reduction of the search space for a given object, since the deviation in the distance for all object's classes was within 0.3 meters.

KTH data set Experiments referring to the distance between an object's real and predicted position were also performed on the KTH data set. These were conducted to evaluate the performance of the MFI-based prediction with respect to external data. Table 3.39 presents the average distance values for the objects keyboard, monitor, and mouse which were calculated from 495 scenes. Since in the KTH data set the object *table* had not been annotated, no distance value for this object was calculated. For the sake of clarity, only the average distance values for the aforementioned objects are presented.

According to the results listed in Table 3.39, it can be observed that the average values were slightly higher than those calculated from the DFKI-RIC data set (shown in Table 3.38). Notably, the average values referring to the expected average value were significantly smaller than those based on the DFKI-RIC data set. This finding is related to the scene type and object classes the KTH data set contains. Since only the table top of the desk and not its surroundings were present in the data, the distances between the office objects were correspondingly small compared to those of the DFKI-RIC in which more spatial space was considered during the experiments.

However, the values obtained from the MFI-based method did not differ much from those resulting from the DFKI-RIC data set and were still smaller than the average values resulting from the average expected value calculation. This finding indicates that the MFI-based method of position estimation of the target object performed well even when testing external data.

Table 3.39: Average distance values for an average expected object's position resulting from the average position to each cell, together with the distance values between the object's real and predicted position based on the MFI-method for the three objects: keyboard, monitor and mouse. The results refer to the KTH data set (provided in meters)

Average	Dist. Keyboard	Dist. Monitor	Dist. Mouse	Dist. Table
KTH average	0.906588	0.898633	0.959471	-
KTH mfi	0.21764	0.318831	0.217467	-

3.2.1.2 Signal-noise ratio evaluation based on probability values of different object positions

In this section, the MFI-based position estimation is analyzed with regards to its applicability for purposes such as object search. To evaluate this method, the probability for an object being found at a given position in the scene, based on the MFI method, is compared with an average probability of finding the object anywhere in the scene and the probability value at the object's real position. For the experiments, scans consisting of 26 office scenes from the DFKI-RIC data set were used. Similar to the experiments described in the previous section, the four objects keyboard, monitor, mouse, and table, were considered in this evaluation. It should be noted that to ensure the knowledge resulting from the same scan did not positively influence the results, each evaluated scan was removed from the knowledge used for the experiments. Furthermore, in the following experiments, a grid resolution of 0.05 meters was used. Figures 3.29-3.33 display the results of the signal-noise ration evaluation. During the evaluation, three probability values for each object and scene were compared:

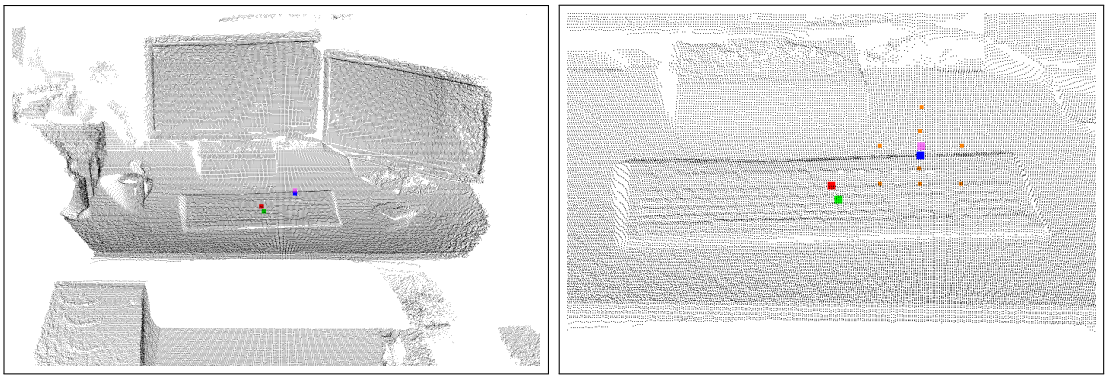
1. The *average* probability of finding a sought after object anywhere in the scene,
2. the probability of finding an object at the *most likely* position and
3. the probability value at the object's *real* position.

The average probability 1 was calculated by summing the FI values of all grid cells and dividing these by their number. In this way, an average probability of finding an object anywhere in the scene was obtained. In the following graphs, this value is termed *Average* and marked with red color. The probability value at the MFI-based predicted position 2 was calculated as described in Section 2.4.1.1 and corresponds to the green bars in the graphs. The probability value at the object's real position 3 corresponds to the FI value at the actual sought after object's position and in the graphs, this number is represented as blue, violet, and light blue bars with the title GT.

In the evaluation process, the MFI position corresponds to the position of a cell located closest to the MFI position, and the GT position to the cell's position located closest to the object's CoG. It should be noted that in some scans more than one object from the object class had been annotated. Therefore, in some scenes, two or three bars for ground truth (GT) values are presented. In contrast to the distance-based experiments described

in 3.2.1.1, the results did not provide any information about the distances between these positions. However, the values provided the ratio of the probability between the position where the given object was located, the average position in the scene, and the assumed position in which the object was meant to be according to the MFI calculation.

Figure 3.29 demonstrates that the values at the predicted position were always significantly higher than the average. This finding was also true for the results from Figures 3.30, 3.32, and 3.33. This result indicates that the method provides reliable results due to a good signal-noise ratio. Furthermore, the probabilities of finding *keyboard* at the predicted and real position did not differ much on average. Interestingly, in some scans, for instance, scans 6, 16, or 25, the probabilities resulting from the MFI calculation were higher than those at the keyboard's real position. These results are related to the learned knowledge about typical object relations and that the spatial object arrangements in these scenes do not comply with the learned knowledge. If the object's real position in the particular scene differs greatly from the position based on the MFI (which contains the learned knowledge), the FI value at the object's real position is lower than at the MFI. Figure 3.26 provides a case where the keyboard's real position is beyond the cell's with the highest FI values. However, although these two probability values vary widely with respect to Table 3.38, the distance between the keyboard's predicted and real position in the 6th scan was only 0.12 meters. This result demonstrates that the resulting probability values refer only to the probability of finding the object at the given position but do not indicate anything about the distance between these positions.

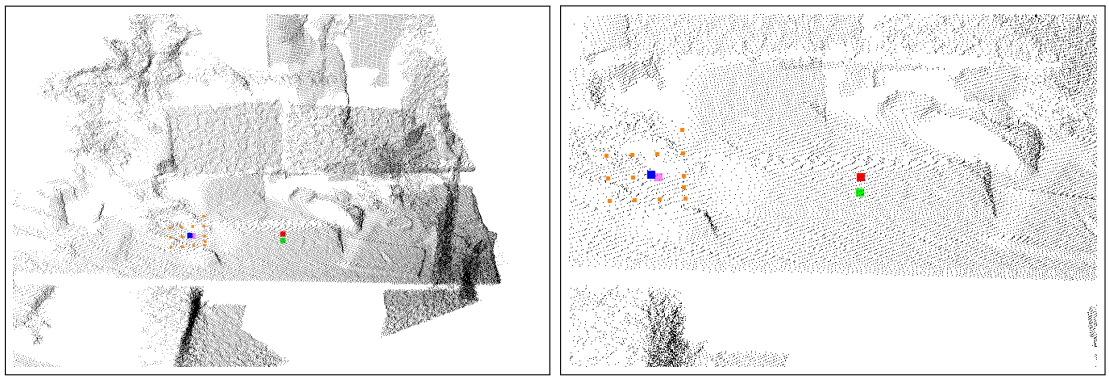


(a) Scan 6 with keyboard's CoG (red) and MFI (blue) and their closest cells (marked green and pink, respectively)

(b) Zoom of the area where the keyboard is located and the highest FI (marked as orange points)

Figure 3.26: The 6th scan of a table desk with searched object *keyboard*. As can be seen the closest cell of the keyboard's CoG is located beyond the area with the highest FI values.

Figure 3.27 provides a further aspect related to the result in which the probability at the MFI position was higher than at the GT position. As can be observed from this figure, the object *mouse* was located to the *left-of* the keyboard in the 16th scan. The relation between these two object classes is somewhat unusual when considering the results provided in Figure 3.16. According to the learned knowledge listed in Tables 3.29 and



(a) Scan 16 with the keyboard's CoG (red) and MFI (blue) and their closest cell (marked as green and pink, respectively)

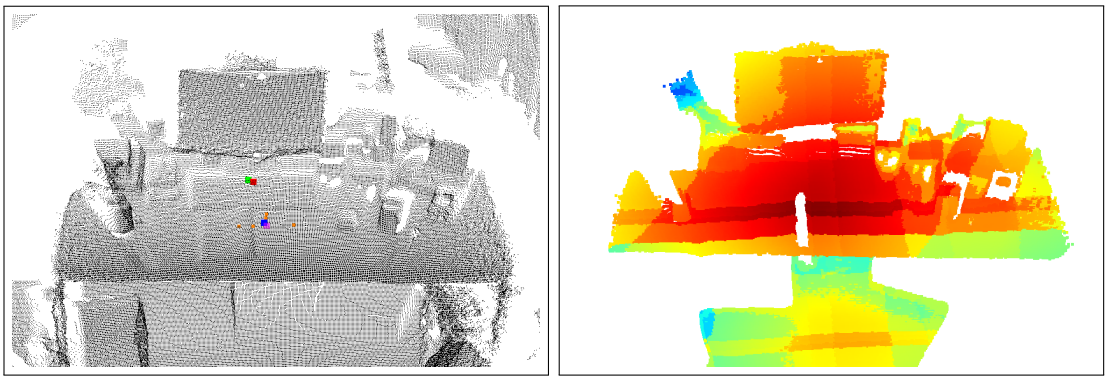
(b) Zoom of the area, with the highest FI (marked as orange points)

Figure 3.27: The 16th scan of a table desk with searched object *keyboard*. The object's position does not corresponds with the most probable keyboard's position from the learned knowledge, since the keyboard is located to the right of a mouse.

3.31 that refer to the objects keyboard and mouse, it is revealed that the *keyboard* was more likely to be found to the *left-of* the *mouse*. Since the mouse is located to the left of keyboard in the 16th scan, the probability value at the resulting MFI position was higher than at the keyboard's actual position. Also, the position of the keyboard did not correspond with the learned most probable keyboard position in the 25th scan. Figure 3.28 illustrates that due to an open book in front of the keyboard, the keyboard had been placed in the area directly near monitor. As a result, the probability at the keyboard's actual position was lower than at the predicted location. These exemplary results highlight the influence of learned knowledge on MFI-based position estimation.

The results of the probability on finding a *monitor* at different positions is compared in Figure 3.30. Similar to the results from 3.29, there was a significant difference between the probability values at the MFI, GT, and average positions. Thus, the probability of a monitor being found at the MFI and GT positions was much higher than at the position referring to the average probability in all scans. As illustrated in Figure 3.30, more than one monitors were present in the 12 scans. However, even in cases where several monitors existed, the probabilities at the GT and MFI positions were significantly higher than those corresponding to the average position.

Interestingly, even though more than one monitor was present in the scan, the probability values at their ground truth (GT) positions were still high. This finding is a promising result, as the probability at the MFI position refers to only one most probable monitor position in the given scan. Therefore, the remaining monitors were not taken into account for the MFI calculation. However, the graphs show that the probability values for the remaining monitors were over 80%. Only in the 17th scan the probability value for a monitor at the GT2 position was 66%. By analyzing the corresponding scan, it is striking that neither the keyboard nor the mouse were located *near* the second to right monitor



(a) Scan 25 with the keyboard's CoG (red) and MFI (blue) and their closest cell (marked as green and pink, respectively). The points with the orange color denote the highest FI values.

(b) A 2D visualization of the most probable keyboard's position under consideration the actual scene and the learned knowledge

Figure 3.28: Scan 25 of a table desk with searched object *keyboard*. In this scene, the keyboard is located under the monitor, as the most likely keyboard's position is occupied by a book.

Figure 3.29: Probabilities to find the object *keyboard* at different positions.

present in the scene, as illustrated in Figure 3.31. Moreover, the monitor was not located *above* or *on* a table. Therefore, the relevant relations such as on or above table did not hold at this position. However, the still high probability at the right monitor's position was due to the near relation, which was valid at this position. Figure 3.31 shows that the SPF of the *near* relation reached the second monitor.

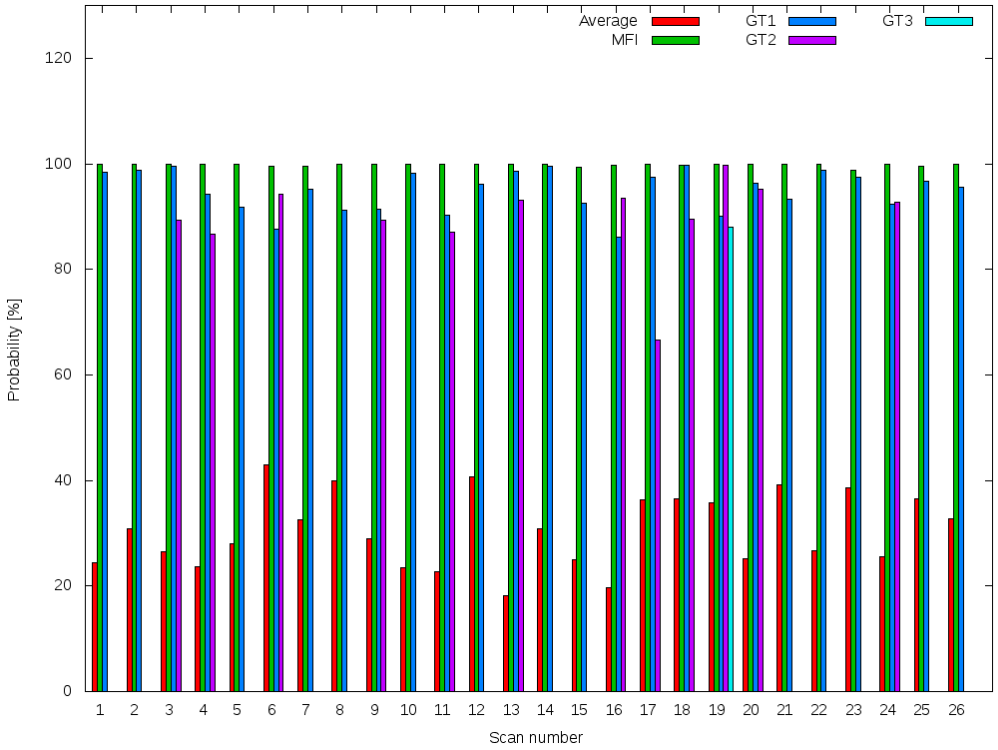
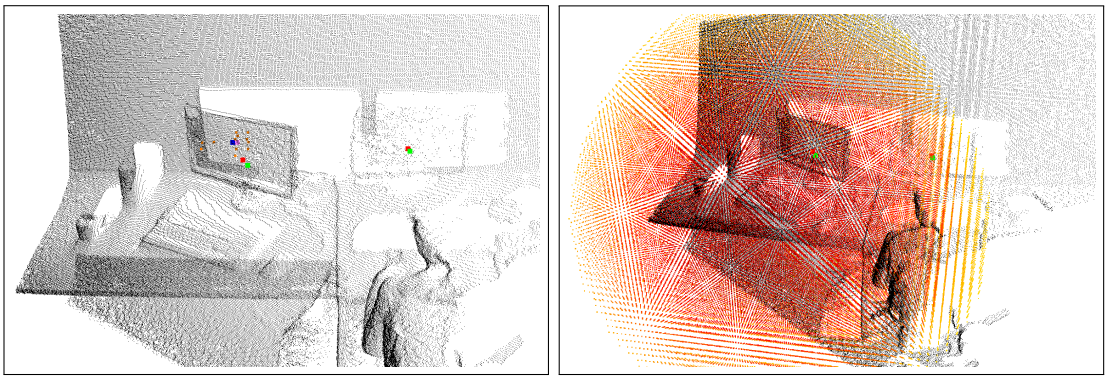


Figure 3.30: Probabilities to find the object *monitor* at different positions.

Figure 3.32 reveals that the probability of finding the *mouse* at different positions do not differ much from the results obtained regarding the object *keyboard* because, again, the probability values at the average position were significantly smaller than at the MFI and GT positions. In turn, the probability at the MFI position were higher than at the GT positions in most scans. For instance, in the 16th scan the probability for the *mouse* at the GT position was lower than at the MFI, as the mouse was located to the *left-of* the keyboard (as shown in Figure 3.27) and this did not correspond with the learned knowledge about typical spatial relations for the mouse and other objects. Furthermore, by looking in more detail at the average probability values, it is striking that the average probability of finding a mouse anywhere in the scene was lower than in cases of the table, by comparing Figures 3.32 and 3.33. These results indicate that in cases of larger objects, such as a table, there were more higher FI values than those with smaller objects. This finding is due to the maximum allowed distance considered by the calculation of the spatial relations. Since larger objects occupy more space in the scene than smaller objects, the resulting FI values were dispersed over a wider area.



(a) Scan 17 with two monitors and their two CoGs (red), one MFI(blue) and the closest cell to the CoGs (green) and the MFI (pink)

(b) The SPF *near* with at least 66%

Figure 3.31: Table desk scene of the 17th scan with two monitors and the SPF of the near relation. In the right figure it can be seen that the *near* range up to the right monitor.

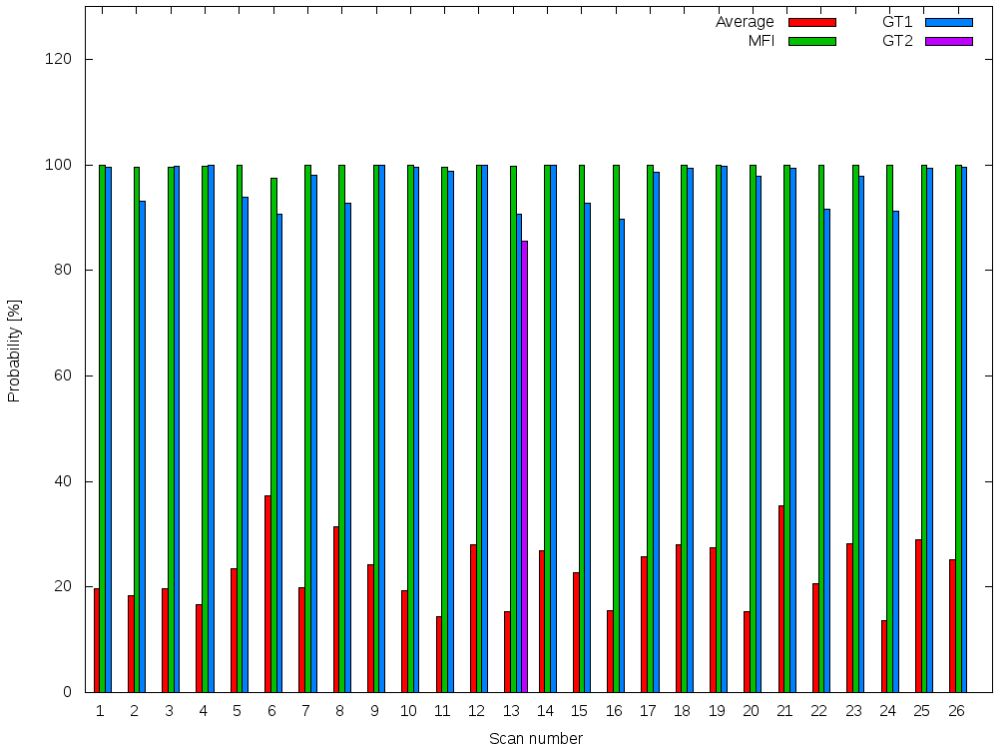


Figure 3.32: Probabilities to find the object *mouse* at different positions.

Results presented in Figure 3.33 reveal that, in the 22th scan, the probability obtained

from the MFI was significantly higher than at the ground truth (GT) position. Figure 3.34 shows the corresponding point cloud of the 22th scan. As illustrated by this figure, the hull of the table is somewhat large because two table instances had been segmented as one. Moreover, due to the size of the table's plane, the CoG of the table (red) was located *in-front-of* the keyboard. According to the results of the PQSR learning (Tables 3.23 and 3.21), the table was most likely to be located *in-front-of* a monitor and *behind* a keyboard. Since the opposite was the case in this scan, the probability resulting of the MFI calculation was higher than the probability at the table's real position. This is because, in this scan, the position of the table did not comply with the position resulting from the learned spatial relations. Moreover, Table 3.38 reveals that the distance value between the table's predicted and real position in scan 22 was 0.39 meters.

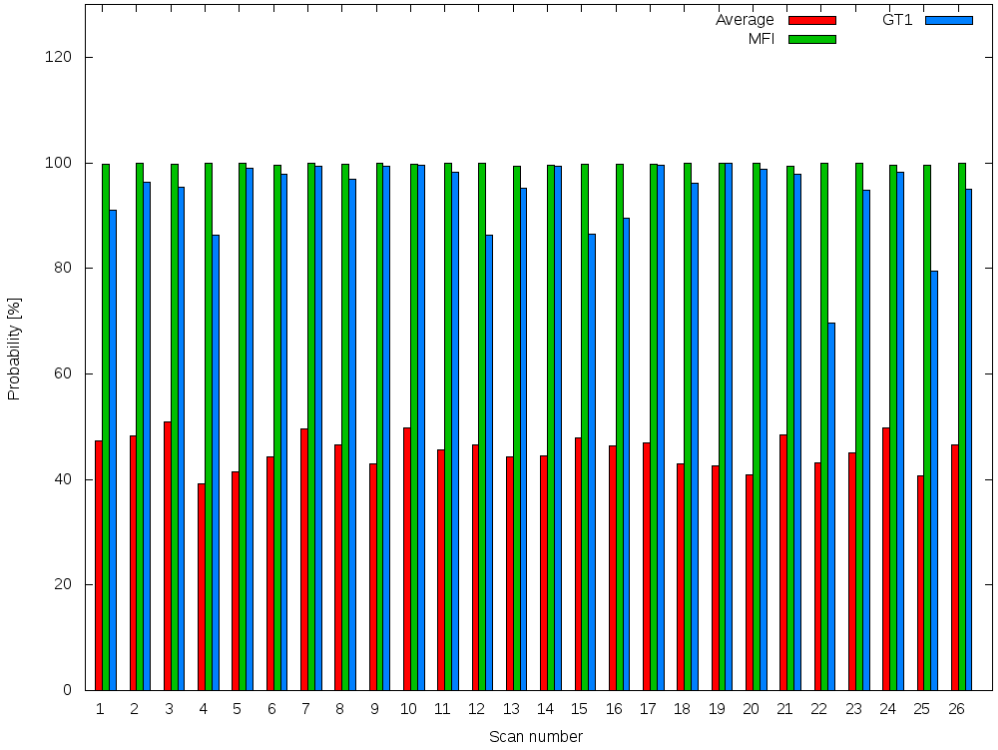
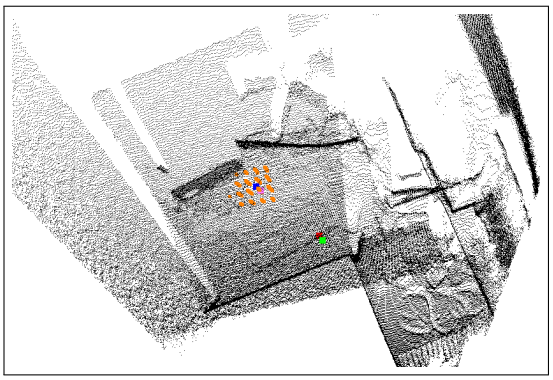
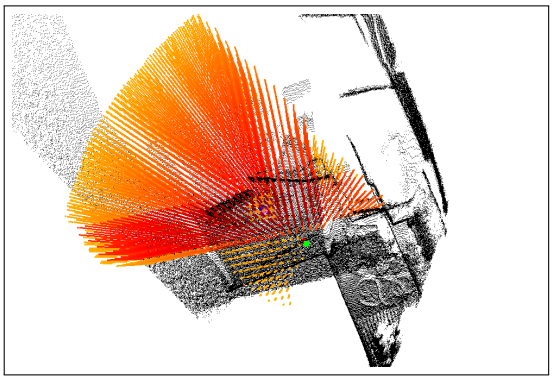


Figure 3.33: Probabilities to find the object *table* at different positions.

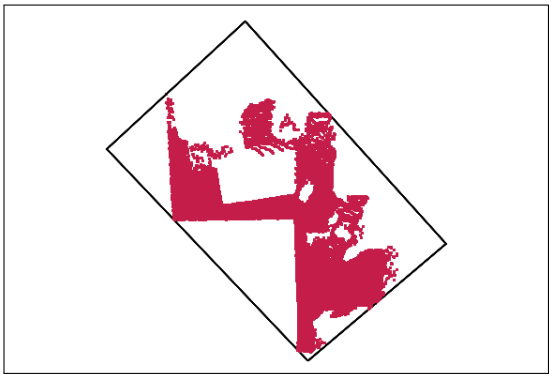
KTH data set Figure 3.35 and Table 3.39 show the average probability and distance values for all KTH scenes at different object positions. Because the KTH data consist of 495 scenes, for clarity, only the average results of all scans are presented. By comparing these results with those from the DFKI-RIC data set, it is evident that the average probability values from the KTH data were significantly higher than those from the DFKI-RIC data. This finding is related to the different scene type in both data sets. In the KTH data, the scans contain only the table top view of the scene, and consequently, the resulting SPF held with a higher probability value. Therefore, the probabilities at each cell position were



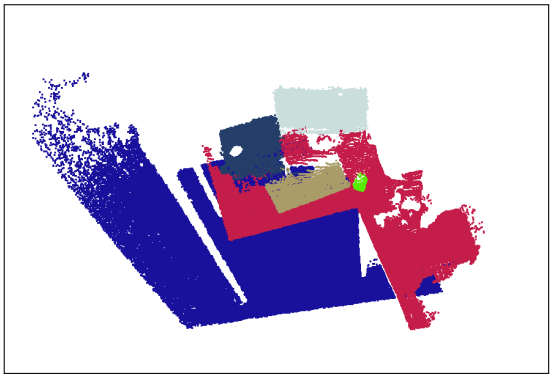
(a) Scan 22 with a table, its CoG(red) and MFI(blue) and their closest cells (marked as green and pink, respectively)



(b) The SPF for the behind and front relations



(c) The hull of the table's plane



(d) The segmentation result of the scan

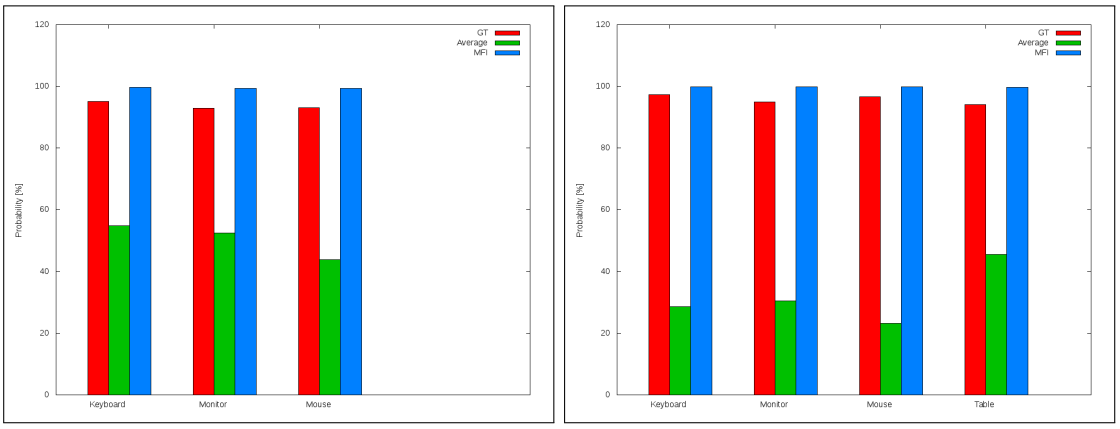
Figure 3.34: The 22th scan with an segmented table and the most probable table's position. As it can be seen the two tables have been segmented as a single table. As a result, the estimated table position does not comply with this based on the learned knowledge.

correspondingly high. In this context, these increased average values in the results from the KTH data are reasonable. Moreover, due to the restricted search area, the probability of finding the sought after object anywhere in the scene was higher.

In contrast, the resulting average probabilities at the MFI and GT positions did not change significantly in the KTH data compared with the DFKI-RIC data. Because the results for both KTH and DFKI-RIC data sets are similar, this finding indicates that the algorithm performed well even in cases of different data sources.

3.2.1.3 Summary of the results referring to the single view scenes

In this section, the usability of the FI-based approach in the context of object search was analyzed. The evaluation presented was performed on both the DFKI-RIC and KTH data sets. In the experiments, two aspects of the FI method were considered. First,



(a) Average probabilities at different object's positions from KTH data set

(b) Average probabilities at different object's positions from DFKI-RIC data set

Figure 3.35: Comparison of the average results of probabilities at different object's positions from the KTH and the DFKI-RIC data sets.

how well the method performs for predicting the most possible object position given the learned spatial knowledge and scan was evaluated by comparing the distances between the object's predicted and actual positions. The corresponding results reveal that, in general, the method provided satisfactory results, as the distances were in the acceptance range of a few centimeters. Furthermore, the probability values at the positions based on the MFI were correspondingly high compared with the probabilities of finding the object anywhere in the scene. Because the evaluation was also performed on the KTH data, the corresponding results indicate that the method also worked well on different data sets.

With respect to the MFI calculation, the results indicate that the MFI-based approach provides promising results for data containing a single view. For instance, an object mouse was found with high accuracy. However, even if the method provides promising results, by having a single MFI, only one most probable object position can be estimated. However, more than one sought after object can exist in the scene. Therefore, it is desirable to predict several probable object positions and, thus, to calculate several MFI.

The results also highlight that incorrect segmentation of the reference objects could lead to inaccurate estimation of the object's position. This problem occurs if, for example, several instances of an object class are segmented as one, such as four tables being segmented as one large table. As a consequence, the maximum allowed distance used for the SPF calculation did not correspond to the real object size and this, in turn, resulted in imprecise distribution of the spatial relation.

Nevertheless, one of the main objectives of the this thesis is to evaluate the developed method - FI-based position prediction. As described in Section 2.4.1.1, the MFI serves primarily as a supporting method to identify the most probable position in instances where more than one highest FI exists in the scene. Because the MFI position results as an average position between several highest FI, this position can deviate significantly from the object's real position. Moreover, in some cases, the probability at the GT position can be slightly higher than at those provided by the MFI, as illustrated in Figure 3.29.

However, since the probability at the GT position is also based on the FI, it has been demonstrated that the method is promising for improving tedious search for a sought after object.

3.2.2 Influence of resolution on position prediction

As previously discussed, the experiments were performed only on the grid with a resolution of 0.05 meters. In this section, the influence of different grid resolutions on the results of MFI-based position estimation is evaluated. For the distance based evaluation, the real position of the sought after object's CoG and the estimated MFI position are compared. In contrast, in the experiments referring to the probabilities at different object positions in the scene, the closest cell to the object's CoG and the predicted MFI position were considered. In the experiments described in this section, similar to the signal-noise evaluation, the positions of the closest cells to the MFI and CoG rather than the real MFI and CoG positions are considered during evaluation.

Lower resolution of the grid provides less precise estimation of the object's position. Therefore, it is important to identify the resolution from which the results become significantly inaccurate. On the other hand, a too high resolution can lead to unfeasible processing time. To perform the evaluation, the FI were calculated using the following four resolutions: 0.05, 0.01, 0.1 and 0.5 meters.

Table 3.40 provides the results referring to the distance between a target object's real and predicted positions by taking into consideration different grid resolutions. The distance values are the average distances calculated from considering 26 DFKI-RIC scans. In the table, the distance values and the corresponding runtimes for each resolution are provided. The experiments were performed using an Intel(R) Core(TM) i7-5820K CPU 3.30GHz, 64 bit processor with 66 GB RAM and 512GB SSD, and the 64bit Debian Jessie operating system.

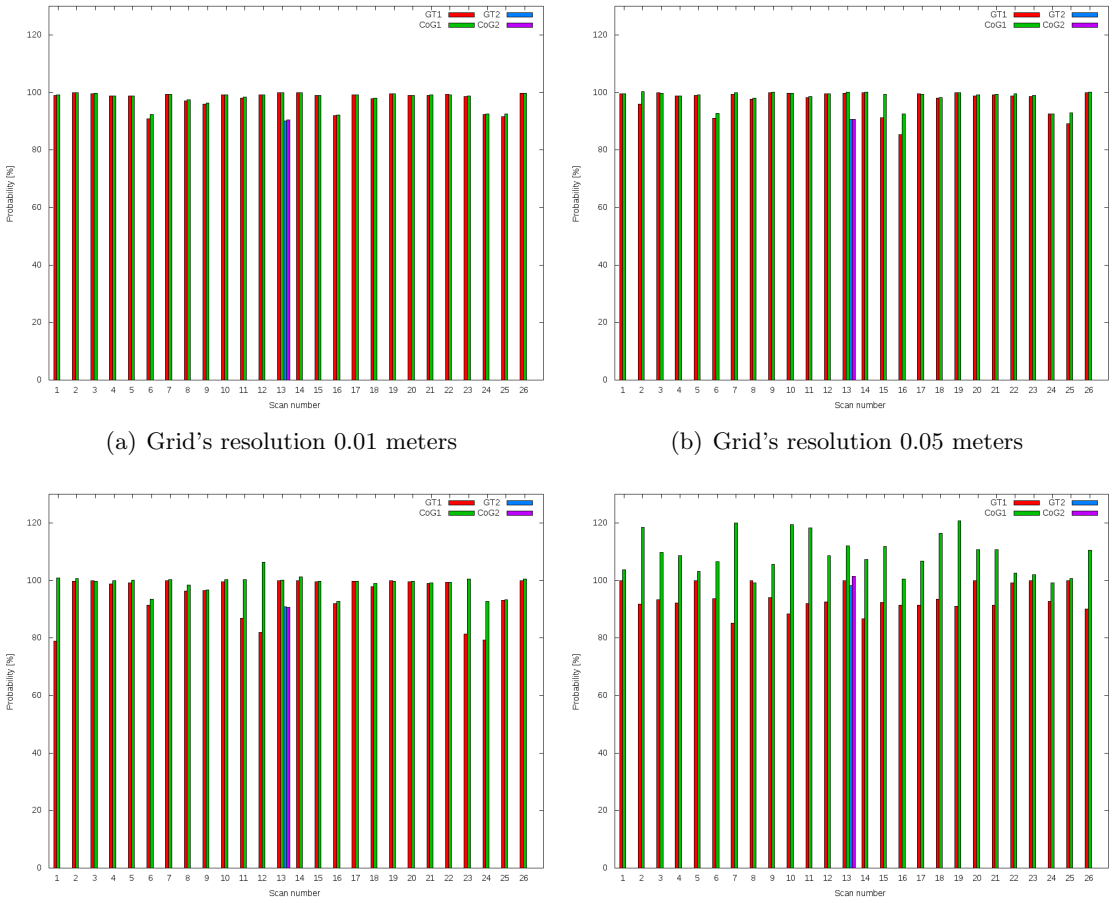
Table 3.40: The average distances between the real and predicted positions of four object classes under consideration different grid's resolutions (given in meters) and the corresponding runtime (with d-days, h-hours and m-minutes).

Resolution [m]	CPU time	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
0.010000	44.656061d	0.117631	0.312843	0.142538	0.23936
0.050000	8.666356h	0.118178	0.302094	0.142478	0.242423
0.100000	1.119511h	0.130057	0.317459	0.148574	0.244138
0.500000	2.477833m	0.361784	0.44795	0.39023	0.331648

As evident from the results in Table 3.40, in cases of the lowest resolution (0.5 meters) the average failure of all four target objects was significantly higher than for the remaining resolutions. By comparing the average distances between the 0.05 and 0.5 meters resolutions, the value varied from 0.09 to 0.25 meters for the *table* and the *mouse* or *keyboard*, respectively. In contrast, for the resolutions 0.01 and 0.05 meters, the difference between the real and predicted positions for all objects was the smallest. However, the table also highlights that the runtime of the processing time increased with higher resolutions. For

instance, it took about 41 days to calculate the FI with a grid resolution of 0.01 meters, but for a 0.5 meters resolution, the calculation took only 1.65 minutes.

It is important to emphasize that the results referring to the 0.01 and 0.05 meters grid resolutions did not differ markedly. This finding indicates the algorithm achieved the optimal results with respect to runtime and precision with the 0.05 meters resolution. Moreover, for resolutions smaller than 0.05 meters, no notable increase in the distance was observed. Therefore, the results indicate that from a 0.05 meter grid resolution, the discretization does not negatively influence the results. In contrast, with decreasing resolution, the average distance between the object's predicted and real positions increased, as illustrated in Table 3.40.



(a) Grid's resolution 0.01 meters

(b) Grid's resolution 0.05 meters

(c) Grid's resolution 0.1 meters

(d) Grid's resolution 0.5 meters

Figure 3.36: Probabilities at keyboard's real position (marked as CoG) and at the position of the closest cell to the real position (marked as GT). Thereby, the results refer to four different grid's resolutions: 0.01, 0.05, 0.1 and 0.5 meters.

Different grid resolutions influenced not only the resulting distances but also caused a change in the probabilities at the position of the closest cell to the object's CoG. In

Figure 3.36, the probability values at the keyboard's CoG position and the closest cell to the CoG for resolutions 0.01, 0.05, 0.1, and 0.5 are displayed. Further results for the objects: *mouse*, *monitor* and *table* are provided in Appendix A.2. By comparing of these results, it can be observed that in cases of the highest resolution, the values were similar or equal at both positions. However, by increasing the resolution, the difference in the values became more significant. For instance, when comparing the grid's 0.01 and 0.05 meter resolutions, the probability at the closest cell to the keyboard's CoG decreased for 0.05 meters resolution in the 15th and 16th scans. Figure 3.37 illustrates this particular case.

In instances of lower resolutions at 0.1 and 0.5 meters, the difference between the probabilities was even greater, as revealed in the Figure 3.38. From this figure, it can be seen that the distance between the object's CoG and the closest cell to the CoG was considerably higher in cases of 0.1 meter resolutions. This results confirmed the expectations that the lower the resolution, the greater the difference between the probabilities and consequently, the imprecise position estimation. However, a more relevant outcome is that the resolution of 0.05 meters lead to the most optimum result with respect to runtime and prediction precision.

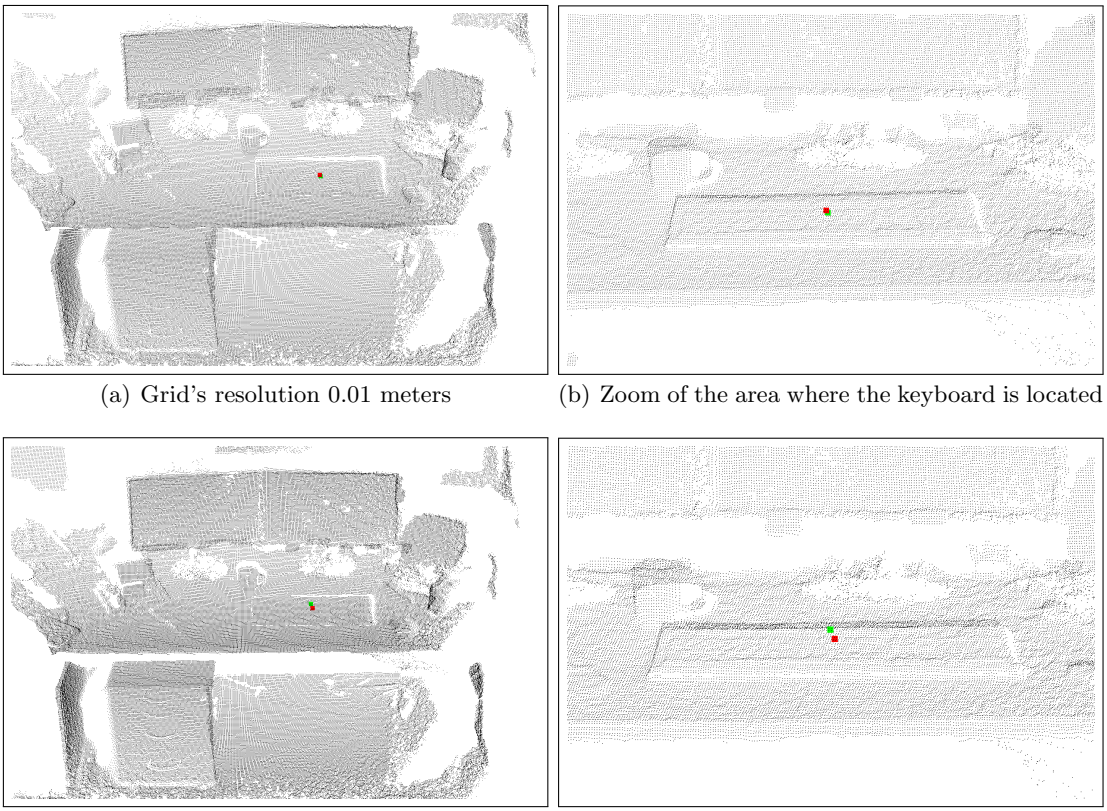
However, the difference between the probabilities decreased with lower resolutions in some scans because by changing the resolution, the cells could become more closest to the CoG. The most striking results are presented in the bottom right of Figure 3.36, as there are several probabilities at the CoG positions higher than 100%. This finding is correct and caused by the fact that the FI value of the real CoG position is not part of the grid and, therefore, was not considered in the grid's normalization process. Therefore, by having a lower resolution (which leads to a greater distance between the object's CoG on their closest cell) the highest FI values from the grid can be smaller than the FI at the object's CoG position. Since in the normalization process, the highest FI was matched to a 100%, which can lead to the FI value at the CoG position exceeding the 100% threshold.

3.2.2.1 Summary of the different resolution results

Taken together, the experiments confirmed the expectations that with decreasing resolution the distance values would increase. Moreover, the experiments revealed that the resolution of 0.05 meters was an adequate trade-off between runtime and precision by estimating the object position. The probability graphs demonstrate that from the 0.05 meter resolution, no significant improvement of the precision could be achieved.

3.2.3 Influence of the relations on position prediction

This section focuses on the evaluation of how absence of certain spatial relation or relation combinations influence the MFI-based position estimation. To evaluate this influence, the deviation between the object's real and predicted position was examined under consideration of the different combinations of spatial relations. Similar to the experiments discussed in the previous section, the following four objects were take into account: keyboard, monitor, mouse, and table. In Table 3.41, the resulting average distance values are presented. These values refer to the results based on the 26 DFKI-RIC office scans calculated with 0.05 meter grid resolution.



(a) Grid's resolution 0.01 meters

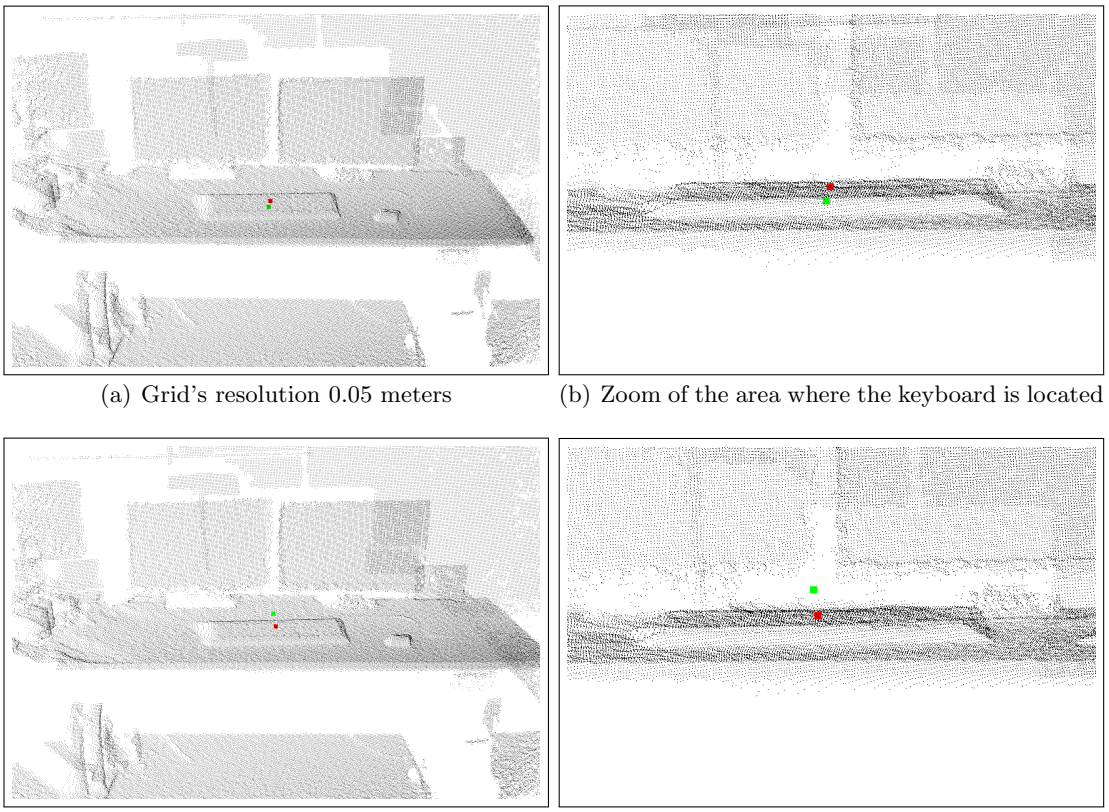
(b) Zoom of the area where the keyboard is located

(c) Grid's resolution 0.05 meters

(d) Zoom of the area related to the keyboard's position

Figure 3.37: An exemplary table desk scene with the keyboard's real position (marked as red) and the closest cell to this position (marked as green). From the (a) and (b) it can be seen that by the grid's resolution of 0.01 meters, the real keyboard's position matched the position of the closest cell.

In the first row of the Table 3.41, the average distances for the four target objects under consideration of all seven spatial relations are provided. These results serve as a basis for further comparison, as the quality of the results was measured by comparing each value with those corresponding to all spatial relations. In the table columns, the spatial relations are listed, which were removed during MFI calculation. In this way, the spatial relation combination that most influenced the results of the MFI-based position prediction for a given object class could be analyzed. The first seven rows of the table (beginning from the row with the above relation) include results of removing one spatial relation for a given object during the MFI calculation. In instances involving the keyboard, absence of the projective spatial relation *left-of* caused on average the highest deviation between the MFI and keyboard's real position. This finding indicates that the spatial relation *left-of* had a significant impact on the predicted keyboard's position. This result is reasonable, as according to Table 3.29, the spatial relation *left-of* held with high probability between the keyboard and reference objects. Strikingly, results for the object *monitor* demonstrate that



(a) Grid's resolution 0.05 meters

(b) Zoom of the area where the keyboard is located

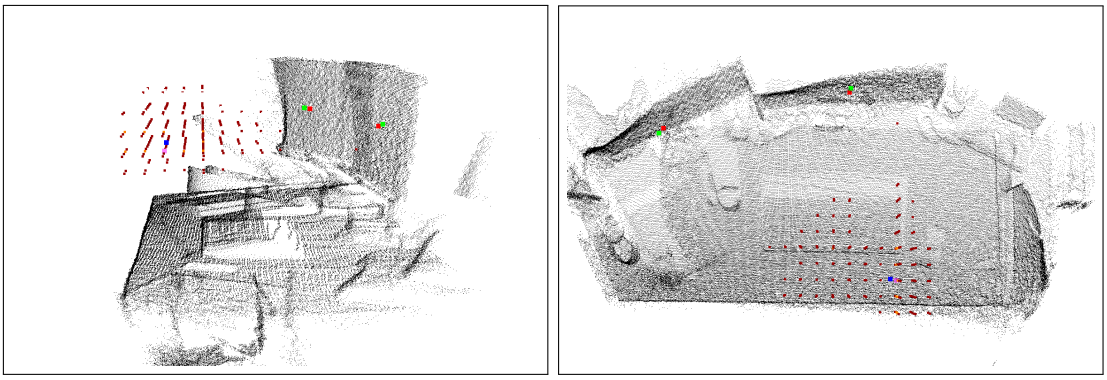
(c) Grid's resolution 0.1 meters

(d) Zoom of the area related to the keyboard's position

Figure 3.38: An exemplary table desk scene with a keyboard's real position (marked red) and the closest cell to this position (marked green) for 0.05 and 0.1 meters grid's resolution. As it can be seen the distance between these positions is much higher in cases of lower grid resolution.

the highest distance resulted by removing the spatial relation *behind-of*. From Table 3.23, it can be observed that the monitor was highly likely to be located behind a keyboard, mouse, and table. By removing this relation, important evidence for the monitor is no longer available. Therefore, using only the behind relation, the object's position can be estimated precisely. In Figure 3.39, the resulting FI and MFI values are provided for an exemplary scan and the object monitor. This figure illustrates that the highest FI were concentrated at the front of the table and not at the back where the monitors actually were.

The greater influence on the distance between the real and predicted *mouse* position after removing *one* of the spatial relations referred to the *right-of* relation. By removing this relation, the average distance between the object's real and predicted positions was consequently the highest. In turn, for the object *table*, the resulting average distance value was the highest when the spatial relation *left-of* was removed. According to the learned knowledge for the projective relations 3.29 and 3.31, the mouse was likely to be located



(a) Side view (b) Top view

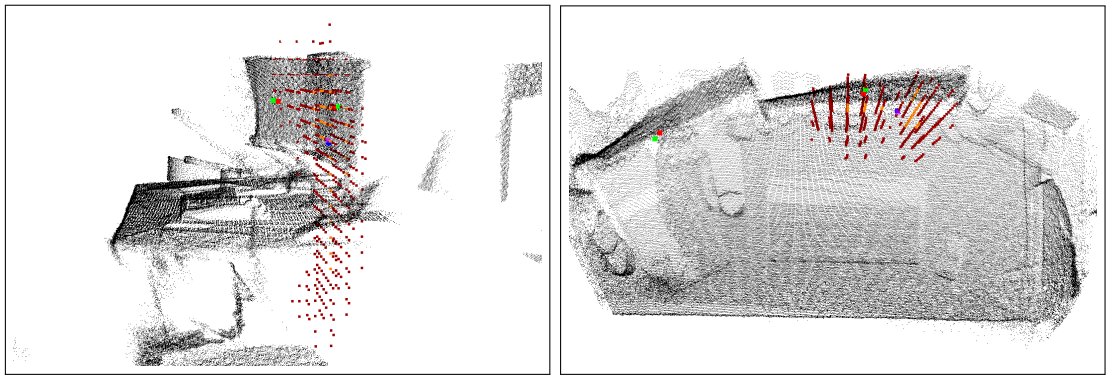
Figure 3.39: Results of predicted monitor's position after removing the spatial relation *behind-of* from the FI calculation. The predicted monitor's position is marked blue, and the real monitors' position red. The dark red points shows the highest FI.

to the *right-of* a keyboard and the table to the *left-of* a monitor.

In the second part of Table 3.41, the results of removing *two* spatial relations from the FI calculation are presented. In cases involving a *keyboard*, the absence of the *above* and *on* relations causes the highest deviation between the keyboard's real and predicted positions. When considering the object *monitor*, the distance was the highest when removing the *above* and *on* relations and not the *in-front-of*, *behind-of* relation pair. This results is surprising, since individually the *behind-of* relation had the greatest impact on distance. However, the distance value for *above* and *on* was still smaller than the value resulting from removing only the *behind-of* relation. As demonstrated by Figure 3.40 and in comparison with Figure 3.39, by removing the *on* and *above* relations, the highest FI values were concentrated in the area around the monitor, but they were distributed more along the z-axis (high). In turn, by analyzing the results of removing the *in-front-of* and *behind-of* relations and as revealed in Figure 3.41, it is striking that the MFI was closer to the monitor's real position in the case of removing only the *behind* relation. Furthermore, the highest FI values were distributed over the entire table top, which, should have significantly increased the inaccuracy in the prediction of possible object positions. However, to obtain the MFI value, an average of the highest FI values was calculated, which lead to a shift of the MFI closer to the monitor's real position, even if the algorithm still had a higher uncertainty.

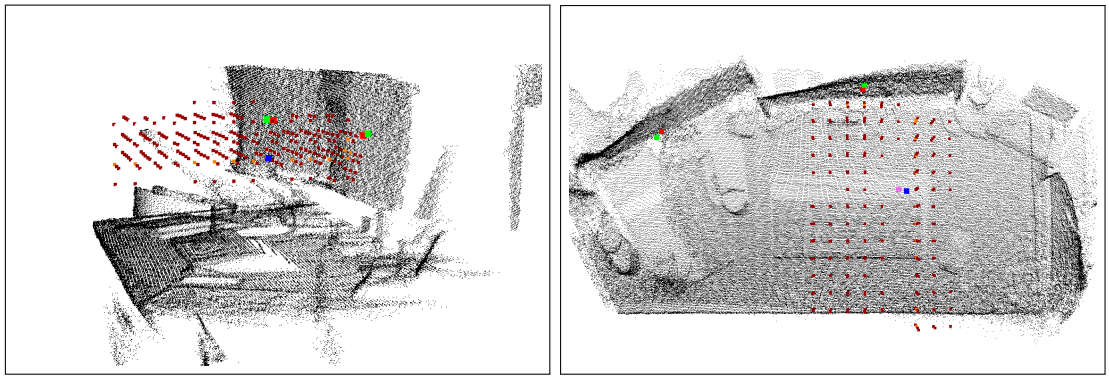
Table 3.41: Average distances between object's real and predicted position resulting after removing different spatial relations (given in meters) with grid's resolution of 0.05 meters and based on the DFKI-RIC data set.

Removed Relation	Average Dist. Keyboard	Average Dist. Monitor	Average Dist. Mouse	Average Dist. Table
-	0.118178	0.302094	0.142478	0.242423
Above	0.1178	0.309872	0.137621	0.251911
On	0.156627	0.3039	0.170113	0.242423
Near	0.146813	0.321387	0.138926	0.242161
Behind	0.153652	0.51905	0.14156	0.369632
Front	0.191254	0.306875	0.240216	0.316448
Right	0.196809	0.324382	0.49543	0.359664
Left	0.263226	0.3553	0.143217	0.441128
On, Near	0.159993	0.323645	0.14942	0.242161
Above, Near	0.149274	0.324857	0.136595	0.32261
Above, On	0.191605	0.349172	0.18846	0.251911
Front, Behind	0.144707	0.31933	0.161325	0.260737
Left, Right	0.106205	0.29615	0.416056	0.237645
Above, On, Near	0.183439	0.496161	0.19156	0.32261
Front, Behind, Above	0.142108	0.3046	0.161073	0.292131
Front, Behind, Near	0.202412	0.473005	0.266549	0.267223
Front, Behind, On	0.157171	0.328333	0.170025	0.260737
Left, Right, Above	0.10173	0.296278	0.412687	0.253858
Left, Right, Near	0.178947	0.3129	0.460426	0.351569
Left, Right, On	0.139644	0.298008	0.433481	0.237645
Left, Right, Front, Behind	0.138036	0.342234	0.469366	0.25377
Left, Right, On, Above	0.17396	0.335764	0.448698	0.253858
Front, Behind, On, Above	0.18497	0.383677	0.222399	0.292064
Front, Behind, Left, Right, Above	0.142216	0.319813	0.4413	0.286245
Front, Behind, Left, Right, On	0.150368	0.351859	0.460099	0.253768
Front, Behind, Left, Right, Near	0.283996	0.456141	0.490374	0.409512
Front, Behind, Left, Right, Above, Near	0.414939	0.468127	0.544418	0.420718
Front, Behind, Left, Right, Above, On	0.200022	0.441061	0.469442	0.286245
Front, Behind, Left, Right, On, Near	0.230709	0.457942	0.456964	0.409512



(a) Side view (b) Top view

Figure 3.40: Results for an object monitor after removing the spatial relations *on* and *above* from the FI calculation. The dark red points denote the area with the high FI values.

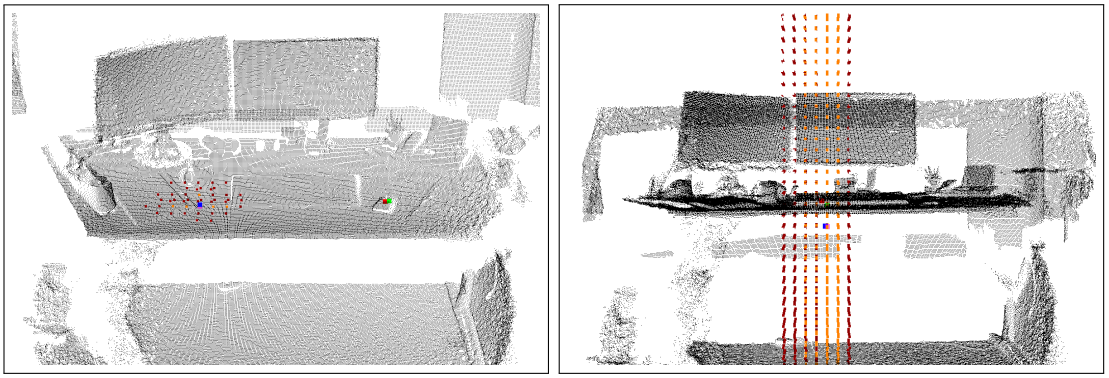


(a) Side view (b) Top view

Figure 3.41: Results for an object monitor after removing the spatial relation *in-front-of* and *behind-of* from the FI calculation. The dark red points denote the high FI values.

For the object *mouse*, removing the *left-of* and *right-of* relations caused the highest distance deviation between its real and predicted positions, whereas absence of the *above* and *near* relations most influenced the resulting distance values for the *table*. These results are reasonable because the mouse was most likely located to the *right-of* of a keyboard. Also, removing information about the *left-of* and *right-of* relations had a negative influence on the prediction result (as illustrated in the right Figure 3.42). In instances involving the table, the *above* and *near* relations were those that held with highest probability between the table and other objects. Without this information, the algorithm could not determine at which height the table was most likely located, as illustrated in the left Figure 3.42. From this figure, it can be seen that the highest FI values were distributed from the floor

to the area above the monitor. However, the relatively small deviation in the distance value was the result of the MFI calculation, which resulted from the average of all highest FI.



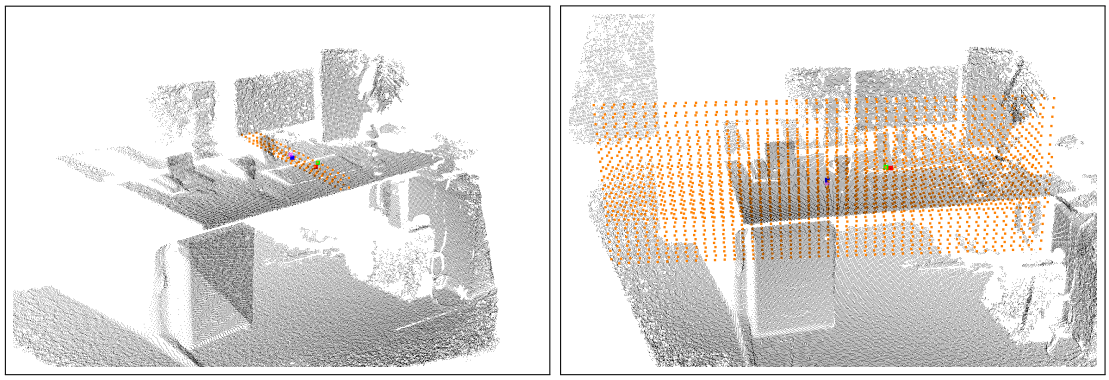
(a) Removed relations: left, right by searching for an object mouse

(b) Removed relations: above, near by searching for an object table

Figure 3.42: Predicted mouse and table positions after removing of different spatial relations during the FI calculation. The dark red points show the probability values above 90%, whereas the orange points show the highest probability values.

The third part of the table lists changes in the distance values when the *three* spatial relations were removed. When considering the first target object *keyboard*, the distance value was the highest if the relations *in-front-of*, *behind-of* and *near* were not taken into account by calculating the FI. This is because, if the information about the front and behind area in which a keyboard is likely to be found is missing, the most probable object position can only be predicted based on the remaining knowledge. The Figure 3.43(a), reveals that the highest FI were distributed through the entire depth of the table top. Again, the relatively small deviation in the keyboard position resulted from the MFI calculation method. In the cases of the *monitor* and *mouse*, the highest distances were caused by removing the relations *above*, *on*, *near* and *left-of*, *right-of*, and *near*. These results correspond with the previous results for monitor and mouse, as the same relations were included in the current combination and lead to the highest distance. For the *table*, the combination of the *left-of*, *right-of*, and *near* relation caused the highest average distance between the table's real and predicted positions.

If *four* spatial relations were removed during the FI calculation, one relation combination had a major influence on the resulting distance values. This combination consisted of the *in-front-of*, *behind-of*, *on*, and *above* relations and caused the highest deviation in the distance values for the *keyboard*, *monitor*, and *table*. In the Figure 3.44(a), an exemplary scene with highest FI values is displayed, which reveals that the most probable *monitor* position was above and to the right of the ground truth monitor in the scene. This result was caused by the missing information about the *on*, *above*, *in-front-of*, and *behind-of* relations, which were important for this object class. Consequently, only the *left-of*, *right-of*, and *near* relations were considered to estimate the monitor's position.



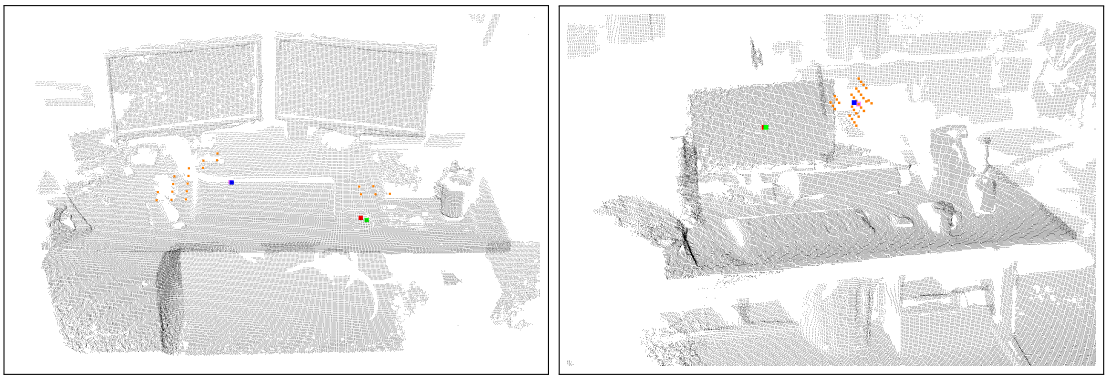
(a) Removed relations: front, behind, near for a keyboard

(b) Removed relations: left, right, near for a table

Figure 3.43: Two exemplary scans with results of removing different spatial relations during the FI calculation. The orange points denote the highest FI values.

Interestingly, in the case of the *mouse*, another combination lead to a highest distance between its real and predicted positions. That is, by removing the projective spatial relations of *left-of*, *right-of*, *in-front-of*, and *behind-of*, an increasing distance was observed. Figure 3.44 provides the highest FI for the mouse in an exemplary scene after removing these projective relations. As can be observed, the highest FI values are concentrated in the right and left part of the table and near to the keyboard. This result is related to the learned knowledge for the sought after object mouse and remaining reference objects, and the fact that only the *on*, *above*, and *near* relations were considered. Because the *on* and *above* relations held on the table surface, the *near* relation had an sizable impact on the final position estimation. According to Table 3.14, the mouse was besides objects such as the keyboard or table, and also near the mug and cupboard. Therefore, in the absence of additional information, both objects influenced the mouse position estimation.

The most striking result from the Table 3.41 is that by removing *five* and *six* spatial relations, the same relations had an influence on the increasing distance values for all four objects. In the absence of five spatial relations, the average distance values between the real and predicted object's positions were the highest by removing the spatial relations *in-front-of*, *behind-of*, *left-of*, *right-of*, and *near*. In turn, if only one relation was considered in the calculation of the FI, the combination of the removed six relations *in-front-of*, *behind-of*, *left-of*, *right-of*, *above*, and *near* had the greatest impact on the deviation in the position. Moreover, the difference in the resulting average distances were significantly greater than when removing one, two, or three relations. The reason for this lies in the missing essential information lost with a decreasing number of relations considered by the FI calculation, since the most probable position must be obtained based only on two or rather one relation. In Figure 3.45, the results of removing five and six relations for the *keyboard* and *monitor*, respectively, are presented. In the Figure 3.45(a) and (b), five relations were removed and only the *above* and *on* relations were taken into account when searching for the keyboard. These figures show that the highest FI were distributed along the table top. In turn, in the Figure 3.45(c) and (d), only the *on* relation was



(a) Removed relations: left, right, front, behind (b) Removed relations: behind, front, on, above

Figure 3.44: Results for finding a mouse after removing different spatial relations. Thereby, the orange points denote the highest FI values.

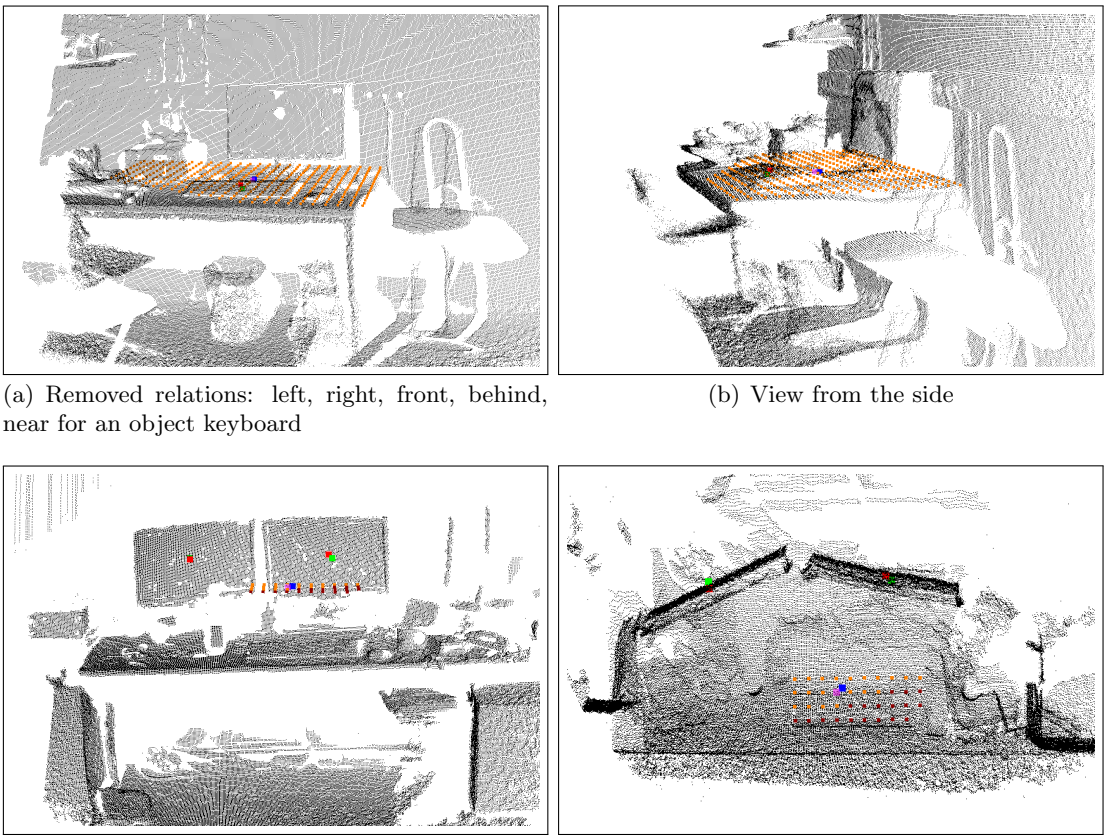
considered by the estimation of the monitor's position. Interestingly, this result lead to an accumulation of the highest FI in the area above the keyboard because according to the learned knowledge for the *on* relation (Table 3.9), the monitor was located with 85% probability *on* the table, but also with 3.8% *on* the wall and keyboard. Since the wall is not present in the scene, the most probable position for the monitor was *on* the table and keyboard. This result caused the FI to be considered as located above the keyboard and not in the area on the entire table top, as in cases of the above relation.

3.2.3.1 Summary of results related to the influence of different spatial relations on position prediction

Taken together, the results indicate that all relations are important for the prediction of an object's most probable position. That is, there is no one particular relation that is equally unimportant for all objects, as the non-presence of different relations had a negative influence on the resulting distances. Moreover, the assumption that some of the relations are redundant was not confirmed by the results. Even if the relation *near* appeared to cover spatially the *left-of*, *right-of*, *in-front-of*, and *behind-of* relations, the experiments revealed that the absence of the *near* relation had a negative influence on the estimated distances. In addition, the near relation is less constrained that the projective spatial relations and thus, the near relation is relevant if the spatial distance between the objects in general is considered.

Regarding the *on* and *above* relations, the results indicate that the *on* relation had a more negative impact on the distance than the *above* relation, or that the above relation is less important than the on relation. By removing of the *on* relation, the difference in the resulting distance was greater than case of removing the *above* relation. This results is due to the fact that for the considered objects classes in the experiments, the *on* relation was more influential as three of the four objects were located on the table.

Furthermore, the results align with the learned knowledge about PQSR, as it is probable that a given object class in a particular relation with other object would be found. Hence,



(a) Removed relations: left, right, front, behind, near for an object keyboard

(b) View from the side

(c) Removed relations: left, right, front, behind, near, above for an object monitor

(d) View from the top

Figure 3.45: Two different scans in which five and six spatial relations were removed by calculating the FI value. The orange points denote the highest FI values.

the absence of this relation caused a higher deviation in the distance between the object's real and predicted positions. More precisely, the important information was missing and the position of the corresponding object could not be accurately estimated.

The results demonstrate that, in general, all relations are important and the more relations that are taken into account by the FI calculation, the better the overall results. According the these findings, by removing five and six of the seven possible relations, the average distance values were significantly higher for all four objects than for removing a relations' combination consisting of only one, two, three, or four relations.

3.2.4 Experiments performed on merge scans

To evaluate the performance of the FI-based approach within larger scenes gathered from merged scans, experiments on the merged scene were conducted. The objective of the experiments was to analyze the accuracy of predicting the most probable sought after object position by using the MFI method in a large scale context. For this evaluation,

one merged office scan consisting of several single scans and an artificial office scene were used.

3.2.4.1 A merged real office scene

The large office scene was obtained by merging single scans taken of an office at the DFKI-RIC. The data were acquired using the Kinect v2 camera mounted on a tripod. The single scans were then merged using the SLAM6d [Nüc09] and MeshLab[CNR] applications. The resulting scan was annotated with five object classes, as follows: table, monitor, keyboard, mouse, and mug. Figure 3.46 displays the merged scan used for the experiments.

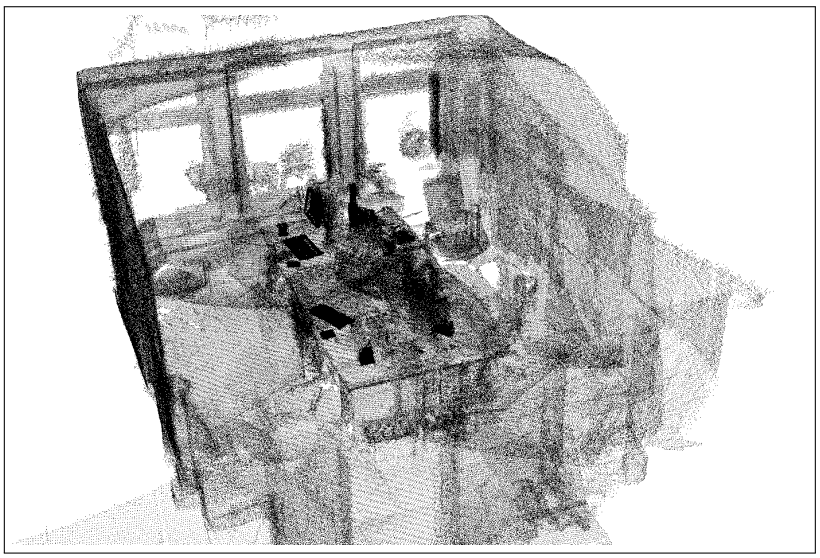


Figure 3.46: The merged point cloud scene used in the experiments.

In the previous experiments, the reference coordinate system for the projective spatial relations was equal to the camera coordinate system used during the acquisition of the point cloud. This assumption was considered to be valid because in a single scan, only one view was provided. In a merged scan, more than one single view of the scene is considered because of its size, and therefore, different coordinate systems must be provided. More precisely, for each possible position from which the robot can observe the scene, a reference coordinate system must be available. To provide this information, the given reference coordinate systems were added to the scan, and represent the different views of the robot. Figure 3.47 illustrates the used reference coordinate systems including their poses.

For calculation of the FI value, each spatial relation shown to hold between the sought after and the given reference object and the corresponding reference coordinate systems must be considered. For the FI calculation, the reference coordinate system was used, which was located *cosest* to the considered reference object (not those located nearest to the given grid cell for which the actual FI was calculated). The reason for this is that the projective spatial relations refer to the view of the reference and not target object. For instance, in an verbal statement “mouse is located right of the keyboard”, the projective

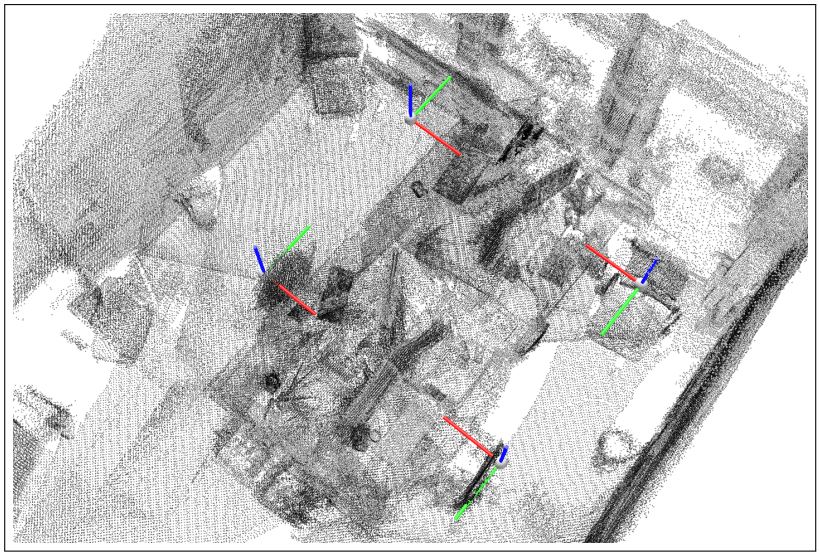


Figure 3.47: Different views of the merged scene used for the projective spatial relations (pictured as poses).

relation *right-of* refers to the direction right from the keyboard. Thereby, in this spatial relation, the keyboard is a reference object.

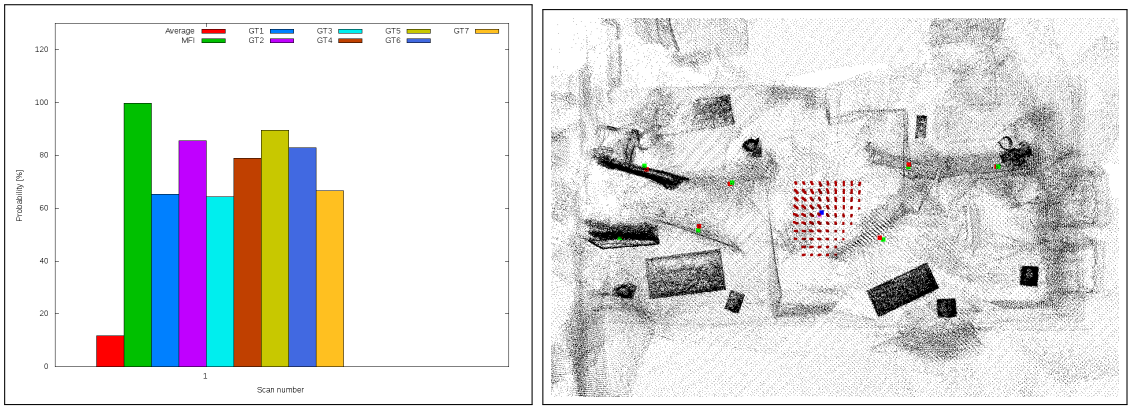
Table 3.42: Distance values between the object's real and predicted position in the given merged scene (given in meters).

Instances of the object class	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
1	1.25114	1.13057	0.307133	1.02742
1.1	0.961069	0.603597	1.43043	-
1.2	0.941687	1.12533	1.34285	-
1.3	0.992965	0.628562	1.10423	-
1.4	-	0.403028	-	-
1.5	-	0.764991	-	-
1.6	-	1.26633	-	-

Table 3.42 provides the results for the distance deviations between the predicted object's position based on the MFI value and the object's real position. Because more than one object instance of the given object class were present in the scene, the distances for all these instances are also listed in the table. However, the resulting distance values refer to the distance between the position of a single MFI and all object instances of a given object class in the scene. This is because, in the MFI-based position prediction, only the most probable object position (single MFI) was calculated. This table highlights that the distance for the object keyboard differ by approximately one meter from the keyboards' real and predicted positions. In turn, the distance values for the monitor, mouse, and

table deviated from 0.3 to 1.43 meters.

When considering the corresponding Figure 3.48, which displays the resulting probabilities at different *monitor* positions, it can be observed that only one value for the predicted monitor position was calculated, whereas seven monitors were present in the scene. The reason for this is that, in the approach so far, only a single MFI value had been calculated regardless of the number of object instances of a given object class in the data. As in the single scans, no more than two instances of a given object's class were present, and the view of the scene was restricted to the table top, which had no negative influence on the results. However, in the larger merged scans in which several table tops exist and the distances between the objects was greater, inaccurate results were returned. Moreover, the resulting probability values ranged from 60 to 99% because only one monitor was located near the single predicted position.



(a) Probability values at objects real positions with a single MFI

(b) The predicted monitor position based on the MFI. The dark red points denote the highest FI values

Figure 3.48: Real-world merged office scene with single MFI in the middle of the scene (marked blue) and the corresponding graph with the probabilities at the monitor's different positions.

As described in Section 2.4.1.1, the MFI is calculated to obtain the most probable object position in instances of more than one highest FI of the same probability value. For a single scene with a limited view of the table top, having one MFI value seemed comprehensible as a limited amount of instances of a given object class were expected to be present. However, in larger scenes merged from several scans including many table tops, there were more possible object instances of a given class and thus, more possible positions of those objects. For this reason, several MFI were calculated in the following experiments to analyze the MFI-based method in a large scale context.

To obtain a number of MFI in a such a scene, the highest FI located not further than one meter from the corresponding reference coordinate system were considered. In this way, for each reference coordinate system, a MFI could be calculated and thus, a more precise evaluation of the MFI based prediction for a target object could be achieved for this type of scene. As a result, as many MFI were calculated as there were reference coordinate systems in the scene. Figure 3.49 displays the merged office scene with six MFI for the

monitor's positions. This figure demonstrates that if only a single MFI was calculated, it was not possible to find all instances of the sought after object's class. In contrast, by having the MFI in each area related to a certain reference coordinate system, more than one object instance could be found.

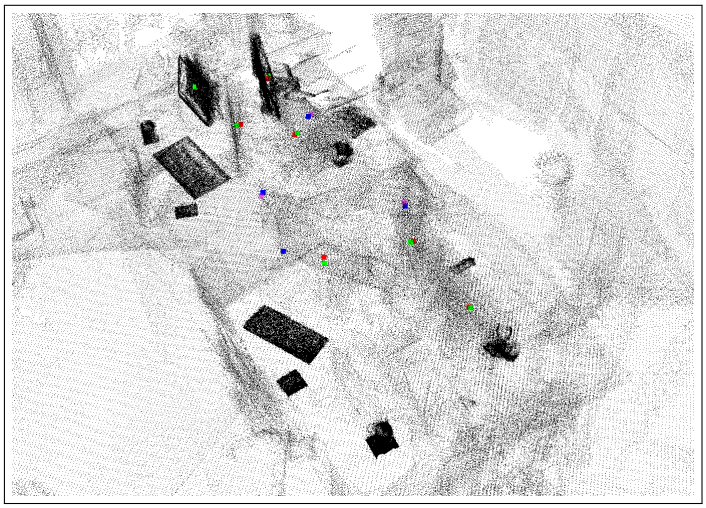


Figure 3.49: The merged scene used for the experiments with four MFI at different possible monitor's positions.

Table 3.43 presents the distance results under consideration in seven MFI. By comparing these results with those from Table 3.42, one can see that for some object's instances the distance values decreased. However, the decrease of the distance values could only be observed in object instances located more than one meter from the single MFI in the previous experiments. That is, the remaining distance values did not changed significantly as most of the values were already within the one meter area. Regarding the probabilities at the GT positions, no changes in the values were observed (3.50) compared with the results present Figure 3.48. This is because the FI values at the given cell positions did not change by the estimation of several MFI-based positions. However, the additional MFI values were lower than the single MFI value in Figure 3.48. This result was caused by the fact that by searching for the MFI, only the highest FI were considered, which were located within a radius of 1.0 meters from the reference coordinate systems. Nevertheless, because of the limitation of the considered area by searching for the MFI, it could not be guaranteed that these MFI values included the global MFI, which was calculated without limitations for the entire grid. Moreover, the distance results from Table 3.43 were between 0.3 and 0.97 meters. Since the search area was already limited to one meter, these results indicate that there was still a high deviation between the object's real and predicted positions.

In addition, by comparing the distances of the single scans, it can be observed that the distances resulting from the large scale experiments were significantly higher than those from the previous experiments described in Section3.2. By analyzing the FI values from the right Figure 3.50, it is striking that the highest FI accumulated in the area not expected to contain the given object. Further analysis revealed that the most probable positions for the sought after object were those in which the most valid relations to all reference

objects existed, regardless of the position where the particular relation most held. As a result, the number of valid relations at a given position was crucial in searching for the target object. Therefore, it can be argued that not only the number of MFI, but also the number of reference objects taken into account in the FI calculation, influence the results.

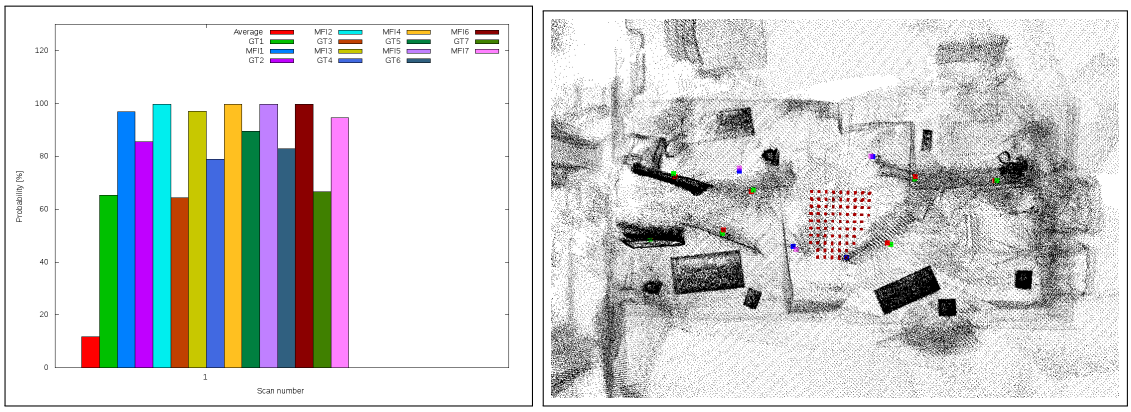
Table 3.43: Distance values at different object's positions under consideration of several MFI.

Instances of the object class	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
1	0.851605	0.971103	0.307133	0.954816
1.1	0.961069	0.603597	0.604131	-
1.2	0.941687	0.943055	0.561891	-
1.3	0.989841	0.628562	0.979971	-
1.4	-	0.403028	-	-
1.5	-	0.764991	-	-
1.6	-	0.955044	-	-

According to formula 2.90 for the FI calculation, each reference object present in the scene is considered. As a result, for each reference object and spatial relation, a SPF is calculated, even if the given relation does not hold at the position with this particular reference object. Moreover, since the SPF are summed and divided by their number, the resulting FI value can decrease due to SPF that is, in fact, not valid at the given position. Because of this, in a large scene (such as in Figure 3.46), taking into consideration all reference objects can negatively influence the resulting FI and, in turn, the MFI. A reason for a relation to be not valid at the given position can be, for instance, greater distance between the reference and possible target object, that is, the reference object might be too far away to be included in a relation with the target object. However, if this is the case, the reference object appears not to be relevant for a given target object position and should not be considered during the FI calculation.

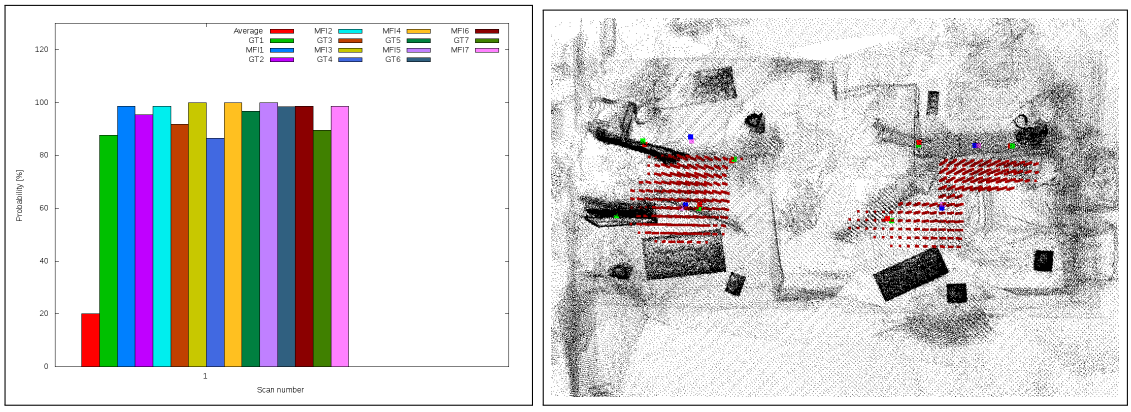
Table 3.44: Distance values between objects' real and predicted positions under consideration several MFI and one reference object per given object class (given in meters).

Instances of the object class	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
1	0.895189	0.418901	0.109687	0.495749
1.1	0.975394	0.356708	0.401755	-
1.2	0.151936	0.250018	0.271652	-
1.3	0.190547	0.370364	0.170683	-
1.4	-	0.591847	-	-
1.5	-	0.0909906	-	-
1.6	-	0.432625	-	-



(a) Resulting probabilities for seven MFI (b) Resulting highest FI with probability from 95%

Figure 3.50: Large scale office scene with visualization of FI of 95% probability.



(a) Probability values at different object's positions (b) Highest FI with probability of more than 93%

Figure 3.51: Probability values for several MFI at different object's positions and under consideration of one reference object per given object class (given in percentages).

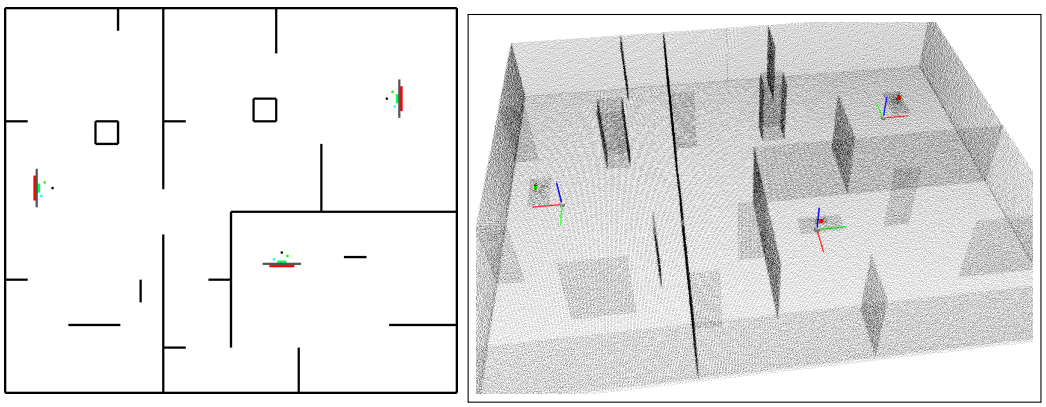
In the following experiments for each FI calculation, only the reference object of a given object class located closest to the actual cell of a grid was considered. In this way, only the closest reference object (of a given type) to the potential target object's position was examined. Table 3.44 presents the distances returned after reducing the number of considered reference objects. By comparing these results with those from Tables 3.42 and 3.43, it can be observed that the distance values between the objects' real and predicted positions decreased. For instance, the distance value for the third *keyboard* in the scene decreased by 0.79 meters, specifically, from 0.94 to 0.15 meters, respectively. The same can been observed in the case of the *monitor* and *mouse*. Figure 3.51 demonstrates that the highest FI were located in the areas closer to the objects' real positions than those in

Figure 3.50. Regarding the probability values at the MFI and monitor's real positions, a significant increase compared withe the previous results can be observed. These results confirm the idea that the consideration of only the closest reference objects improves overall position estimation.

3.2.4.2 An artificial office scene

In the previous Section 3.2.4.1, the performance of the FI-based most likely object position was analyzed in the context of a merged scans consisting of several scenes. The objective of these experiments was to evaluate how the method performs in larger scenes than those used in the previous experiments. To conduct the evaluation, a large scale artificial scene was created. The scene was designed using the DIA[dia] and a self-developed application for this purpose. In the scene, three tables, each with a monitor, keyboard, mouse and mug were included. Additionally, a floor and several walls were added to the scene to simulate the building structure of a larger scene. Figure 3.52 displays the resulting DIA scene.

For the projective spatial relations, the view from which the objects are being observed must be known, and three reference coordinate systems were defined in the scene, as illustrated in the right Figure 3.52. Details about position, width, depth, and orientation (rotation) of a particular object, as well as the reference coordinate systems, are provided in Table 3.45. In these experiments, a grid resolution of 0.1 meters was used, and the planes in the artificial scene had a point density of 0.05 meters. Since real data are often noisy, a normal distributed Gaussian noise of 0.02 meters was added to each point of the plane to create a more realistic representation of the planes.



(a) Artificial scene: DIA (b) Reference coordinate systems

Figure 3.52: Visualization of the artificial scene.

Similar to the experiments described in the previous Section 3.2.4.1, the FI-based object position prediction was evaluated by considering several different aspects. However, for both experiments, the MFI located within three meters of the given reference coordinate system were selected. This enable different possible object positions to be obtained, which was desirable because of the scene type (as discussed in Section 3.2.4.1).

Table 3.45: Elements of the artificial scene with their features.

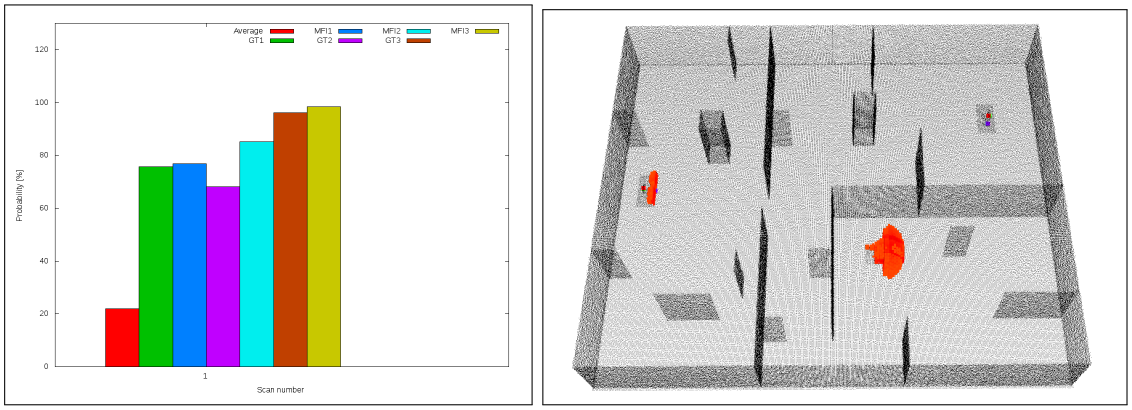
Type	x	y	z	width	height	yaw
Perspective	5	12.95	-	-	-	-
Perspective	1.05	-1.85	-	-	-	180
Perspective	-1.8	8.3	-	-	-	90
Monitor	13.5485	5.015	1.06504	0.450067	0.300192	0.348877
Mug	13.1236	5.23312	0.90415	0.120201	0.0702383	-67.909
Table	13.45	5.02494	0.799788	1.65004	0.79991	4.77927e-06
Keyboard	13.37	5.0152	0.900785	0.450898	0.201879	0.00691089
Mouse	13.2386	4.59312	0.90415	0.120201	0.0702383	-67.909
Monitor	-2.70147	1.065	1.06504	0.450067	0.300192	0.348877
Mouse	-2.26139	1.24312	0.90415	0.120201	0.0702383	67.909
Table	-2.59988	1.075	0.799788	1.64991	0.799979	-4.77929e-06
Keyboard	-2.47998	1.0652	0.900785	0.450898	0.201879	-0.00691089
Mug	-2.42639	0.633117	0.90415	0.120201	0.0702383	-67.909
Monitor	8.265	-2.39854	1.06505	0.450067	0.300192	89.9996
Mug	8.47361	-2.01688	0.90415	0.120201	0.0702383	67.909
Table	8.275	-2.3	0.799788	1.64991	0.799979	-90
Keyboard	8.26502	-2.22	0.900786	0.450898	0.201879	-89.9701
Mouse	7.88861	-2.15688	0.90415	0.120201	0.0702383	67.909
Floor	5.96901	0.481973	1.80978e-05	20.1	17.1	-90

In the first trial, the MFI were calculated under consideration of all reference objects present in the scene. The resulting distance values are listed in Table 3.46, and demonstrate that the distance values resulting from the MFI calculation deviated from the object's real position being between 0.15 and 0.99 meters. Especially in cases of objects located on the second table, it is striking that the prediction of these positions deviated the most from the real positions. These results are surprising, since the objects' arrangement on each table did not differ to a large extent. The reason for this strong deviation lies in the number of relations that held at the given object positions with respect to the second table. As can be observed in the right Figure 3.53, the probabilities for the monitor to be located on the table in the right upper corner did not exceed 80%. Conversely, the probability of finding the monitor located on the third table, which was near the middle of the scene, was the highest. By analyzing these results and considering the information in the right Figure 3.53, one can see that the third table was located closer to most walls in the scene, whereas for the second table, only five walls were close. However, not all of the five walls were in a valid relation with the table and the objects located on it. As the number of valid relations influence the probability values resulting from the MFI calculation, the most probable object position is where the most relations hold. Therefore, the probabilities at the GT positions for the third table were the highest, as illustrated in the left Figure 3.53.

Table 3.47 provides the results for distances in the same scene, but under consideration of only one closest reference object of a given object class. Results in this table reveal

Table 3.46: Distance values between the object's real and predicted position calculated under consideration of all reference objects from the scene.

Instances of the object class	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
1	0.150982	0.197225	0.122952	0.182536
1.1	0.439237	0.498003	0.683259	0.991503
1.2	0.155043	0.315879	0.48346	0.542777



(a) Probabilities at different positions

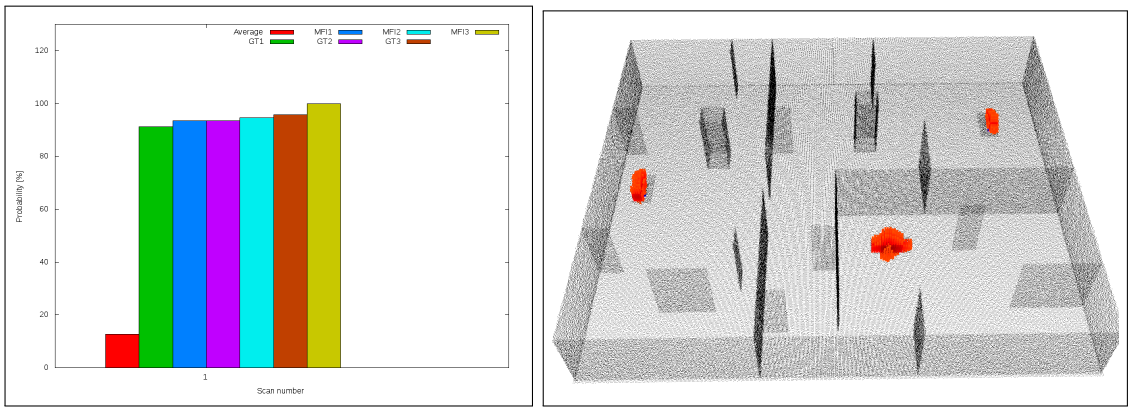
(b) FI visualization

Figure 3.53: Probabilities at the most probable monitor positions and the corresponding FI visualization with probability of at least 0.8 meters under consideration all objects in the scene.

that the distance values for all objects decreased. In particular, distances to the objects located on the second table became smaller than those listed in Table 3.46. Information in Figure 3.54 compared to Figure 3.53 reveal that the probabilities of finding a *monitor on* each table increased and exceeded 80% for each table. These results indicate that by removing the number of reference objects considered by the FI calculation, a positive influence on the resulting distance values can be achieved. The same was observed in the experiments performed on the large scale real data 3.2.4.1

Table 3.47: Distance values between object's real and predicted position under consideration only one reference object of a given object class

Instances of the object class	Distance Keyboard	Distance Monitor	Distance Mouse	Distance Table
1	0.140375	0.166755	0.122952	0.0911042
1.1	0.161673	0.142073	0.207063	0.197757
1.2	0.136553	0.192004	0.174273	0.0990414



(a) Probabilities at different positions

(b) FI visualization

Figure 3.54: Probabilities at the most probable monitor positions and the corresponding FI visualization with probability of 0.8 meters under consideration only one reference object of each type in the scene.

3.2.4.3 Summary of the experiments related to merged and large scenes

In summary, the results for both, the real and artificial large-scale data, have shown that there is correlation between the number of reference objects being taken into account by the FI calculation and the distance deviation between the object's real and predicted position. The smaller the number of the reference objects of a given object classes considered by the FI calculation, the smaller the distance deviation between the object's predicted and real position is in the end.

Furthermore, compared to the distance results for a single table top scene, it has been shown that for a large scale scenes it is required to estimate more than one probable object's position. The reason for this is, that in a large scale scene the objects are present several times. Although in the single scenes an instance of a target object can be present also more than one time, the experiments have shown that already by searching for one possible object's position (single MFI) sufficient results can be achieved. In contrast, in a large scale context it is desirable to have an local MFI and not global, as for a robotic system acting in a large scale environment it is more efficient to search locally for a sought after object than on more distant areas.

3.2.5 Summary of the Field Intensity-related experiments

In Section 3.2, experiments and results related to the FI-based object position estimation were presented. In these experiments, the distances between the sought after object's predicted and real positions, as well as the probabilities at different positions in the scene were analyzed. For the evaluation, both the DFKI-RIC and KTH data sets were used to investigate how well the algorithm performed on data from different sources. With regards to different possible scene types, a single view and large scale scans were considered in the evaluation.

In addition, the influence of different qualitative spatial relations on the resulting dis-

tance values was investigated by comparing the results obtained with consideration of all spatial relations with those that resulted after removing one or several relations.

The results presented in this chapter indicate that the FI-based position prediction is a promising approach for predicting an object's most probable position and thus, supporting the object search. The experimental results have demonstrated that by using of the FI method, the most probable position of the sought after object can be estimated reliably. Furthermore, an improvement in the distance deviation between the object's real and estimated position was observed. Furthermore, the distances resulting from the FI method were significantly smaller than the average distance failures of the real position. Therefore, the FI-based method is suitable for reducing the search space compared to an exhaustive search. Since the FI-method was evaluated using real-world data, which also contained noise, the usability of this approach for real-world applications was also evaluated.

In addition to the distance-based evaluation, probability at different positions were also analyzed. The corresponding results illustrate how well the FI-based method performs in a more general manner and whether it is suitable for object search in a three-dimensional, real indoor environment. In this context, it has been shown that by applying the learned knowledge about typical objects relations, it can be estimated how probable it is that a sought after object is found at certain positions in the scene and thus, locate the object's most probable position. Moreover, by comparing the probability values of finding the desired object at the FI-based predicted position with those of finding the objects anywhere in the scene, the probability at the FI-position was always significantly higher than the average. Although in some cases the probability at the object's ground truth (GT) position was higher than at the MFI, the overall approach provided reliable results. This is because the values at the object's real position were obtained using the FI method. Moreover, the MFI was intended to find only the highest FI value in instances when several highest FI existed. In this context, the position estimation based on the PQSR and the SPF was promising for finding a sought after object.

Results from the relations evaluation demonstrate that all the PQSR were important for the position estimation and did not negatively influence the resulting distance values. In turn, the experiments revealed that the more relations considered by the FI calculation the better the results. However, depending on the object class, different relations and relation combinations had a different impact on the results.

In these experiments, the influence of different grid resolutions on the prediction was also investigated. The evaluation demonstrated that the resolution of 0.05 meters was suitable for achieving precise prediction results. Moreover, from this point onwards, no significant improvement of the prediction could be observed.

During the experiments related to the merged and large scale scenes, it was observed that the number of relations that did not hold between the target and reference objects negatively influenced the prediction results. As a result, the number of considered objects for the FI calculation was adapted to only one reference object per object class. Corresponding experiments revealed that because of this adaptation, the influence of the invalid relations returned was reduced. Consequently, the probability distribution of a large object scene was calculated more precisely.

Although according to the corresponding results, the MFI method produced suitable results for a single view scenes, it became evident that for scenes in a large scale context, an adjustment of the MFI method was required. Crucially, in the MFI-based method,

a single MFI was calculated. As a consequence, if more than one object instance of a given object class were present in the scene not all objects could be found by applying this approach. To make the method more reliable for large scenes, a search for a local MFI with respect to the given reference coordinate system was preformed. After applying this change, the method provided comparable results to that of single view scenes.

4 Discussion and Outlook

In this chapter, the results presented in this thesis are summarized and discussed in context of the research objectives described in the first chapter 1.4. Thereby, the results are subject to critical analysis so possible improvements to the approach can be identified. In addition to this discussion, an outlook of further developments in robotics, such as object search, is presented.

4.1 Discussion

The primary objective of this thesis is to develop a formal concept of probabilistic qualitative spatial relations and apply these relations to support a robotic system as it searches for an object in an environment. For that purpose, qualitative spatial relations were extracted from quantitative real-world data such as point clouds or images from a Microsoft Kinect camera. To achieve this objective, a formal definition of PQSR was elaborated by which relations are learned from the data and then used for estimating the most probable position of a sought after object. To this end, the spatial representation form (Spatial Potential Fields) and their most meaningful structure (Field Intensity) were defined and used.

A particular advantage of this approach is that the PQSR are defined in a probabilistic manner and therefore enable estimation of an object's position more reliably by using approaches based only on crisp relations. Moreover, due to the three-dimensionality of the relations, a more precise estimation can be achieved. By using the PQSR, the typical object co-occurrences can also be identified in an exact three-dimensional spatial manner. Furthermore, the developed approach comprises two different aspects, the estimation of the most probable object position and a probability calculation for each object occurrence at any position in the environment. Notably, based on the FI it can be estimated where a sought after object is most likely to be found in the scene and how probable it is that an arbitrary object is located at any position in the three-dimensional space.

During the experiments using two different data sets, the theoretical concept of the relations was evaluated with respect to their applicability for object search purposes. Thereby, the qualitative spatial relations were learned from the quantitative data based on their formal definitions described in 2.2.2. The corresponding results indicate that the PQSR are a rich form suitable for representing spatial relations between objects in a qualitative and probabilistic manner. The overall results presented in the previous chapter 3 reveal that the developed approach is a promising new method for supporting object search by using this method as a heuristic to estimate a target object's most probable position. The experiments also examined how precisely the prediction could be performed by comparing an object's actual and estimated position, as well as how reliable the FI calculation is in context of probabilities at different object positions. In these experiments, the effect of

removing particular spatial relations or different relation combinations from the FI calculation on the overall prediction results was also evaluated. In further experiments, the influence of the grid resolution on the resulting precision was analyzed. The evaluation process was conducted under consideration of several different aspects of the data such as the size of the scene, type of data (real vs. artificial), and views contained within the data (single view vs. merged scans). Ultimately, the experiments confirmed the expectation that a suitable predestination of an object's position can be estimated by applying the FI method, and as a result the effort required for the object search can be reduced. Based on the FI method, the most probable object position can be precisely estimated, and thus, the search space can be reduced to an area where the object is expected to be found. Although the developed approach provides promising results for object search purposes, there are some aspects of the method that must be discussed and improved in further work.

The Probabilistic Qualitative Spatial Relations and Field Intensity method

As the main part of this thesis refers to the formalization of the PQSR and their usability for position estimation, this section outlines the relevant findings regarding the definitions and learning of the PQSR and then provides suggestions for improvement.

During the learning process of the PQSR from the provided data, the given target and reference objects of all known object classes are considered. Thereby, one target object per scan is considered regardless of the number of target object instances annotated in the scan. Furthermore, the reference object of a given object class located closest to the considered target object is included during learning of each PQSR. Since a scan can have more target object instances, part of the information becomes lost. Therefore, improvement of the assessment of which reference objects are important for the given relation and consequent adaption of the learning process with respect to this gained knowledge is required. For instance, only the reference objects that are spatially near the current target objects could be considered, irrespective of whether they belong to the same object class. To this end, a clustering method could be used. By applying such a method, the scene becomes divided into areas related to the existing instances of the target object. In this way, relations are learned between the current target and reference objects located in close proximity. Importantly, the reference objects located beyond this area would not be taken into account. As a result, PQSR for each target object in the scene and their nearest reference objects would be learned.

Since the data used in the learning process are related to office scenes gathered from the same institute, the resulting knowledge might become over-trained by the use of similar environments. Additional and different data consisting of varied office settings could reduce such a over-training effect. However, the aim of this thesis to develop an approach for object position estimation that uses spatial knowledge from the data and uses this information in a probabilistic manner to calculate how likely it is that an object would be found at any position in the scene. As previously described, the experimental results demonstrate that by using knowledge already learned from the data available during this work, a reliable object position estimation could be achieved. Furthermore, these results indicate that the developed approach is suitable for the learning of object co-occurrences and spatial relations between objects.However, this thesis is not primarily intended to

obtain statistical representations of object occurrences, but rather to evaluate the basic theory behind the developed method.

The method developed in this thesis requires that for learning the PQSR data annotated with reference objects must be provided. Because point cloud data are used in this approach, the objects must first be classified before they can be annotated. However, this classification can have a negative effect on the results because the object's features are used in the relation calculations. As demonstrated in the experiments 3.1, the reference objects had not always been classified accurately in the DFKI-RIC and KTH data sets. Unfortunately, inaccuracy in the classification can negatively influence the resulting relations because in their formal definitions, the width, depth, and CoG of the reference objects are considered. To improve the PQSR learning in future developments, the segmentation method could be replaced by a more precise algorithm. Importantly, the better the classification of objects the more accurately the relations can be learned.

The experimental results presented in this thesis demonstrate that by applying formal definitions of the PQSR, knowledge about spatial relations between objects can be extracted from metric real-world data in a reliable way. However, some formal concepts of the relations could be improved to make the relations more valid generally. Currently, the CoG of a target and reference object are considered during the PQSR calculation. For some spatial relations, the maximum allowed distance is compared with the distance between an object's CoG, and this might be a crucial factor for a given relation to be valid. For instance, according to Def. 7 of the *on* relation, a target object is located on the reference object if the distance between the CoG of the target and the surface of the reference object does not exceed a given threshold. With respect to this definition, a larger object such as monitor would not be located on a reference object. Hence, the size of a given object and, more precisely, its CoG influence the validity of the relation. Improvement in the formal definition of the *on* relation should prevent this effect. One possible option would be to consider the point of the target object's surface located closest to the surface of the reference object rather than target objects CoG. If this rule was applied, larger objects could also be classified as being located on a reference object if further constraints of the on relations are satisfied.

In the developed method, learned knowledge about typical object relations is used in calculation of the SPF, which also contains the probability of finding a target object in a given relation with a reference object at a particular position in the scene. These SPF are then used to obtain the overall probability of finding a given object at a particular position in the scene. For this calculation, the SPF for all reference objects are summarized to an FI that specifies the overall probability of a target object being located at the position under consideration, the learned knowledge, and all valid spatial relations in the scene. Because only reference objects existing in the scene are considered, the reference objects not present in the scene do not influence the resulting FI values. This outcome is also the case if a certain relation holds at a given position in the scene but does not exist according to the learned knowledge. In contrast, if it has been learned that the given relation holds between the target and reference object, but the relations are not valid in the scene (for instance, because the reference object is located too far from the target object) then this assessment impacts the final probability value as the resulting FI decreases at this position.

Because the number of valid relations has an impact on the resulting FI value, the most probable object position in the scene is where the FI is the highest irrespective of the

number of valid SPF at this position. For instance, if a certain position in the scene has two SPF with 80% probability, each hold and in turn, only one SPF is valid at another position, but with 90%. As a result, and with respect to the FI, the position where only one relation holds becomes more probable of containing the target object as this point is where two of the possible SPF are valid. This result arises from the SPF being summarized and divided by their number during calculation of the FI. Ultimately, the position that contains the FI with the highest probability, regardless of the number of valid SPF at this position, is considered as the most probable. Although this result is correct according to the formal definition of the FI, further investigation is needed to analyze how the FI calculation affects the position estimation.

Furthermore, the experiments related to the large-scale office scene demonstrate that by applying the FI method, it was not possible to obtain feasible results in a larger scene because too many reference objects irrelevant to finding a target object's position were considered. The evaluation of single view scans reveal that the method was sufficient for estimating the most probable object position, but the experiments in the large-scale context highlighted that an adjustment to the method is needed to achieve reliable results. To this end, only reference objects considered to be located closest to the actual position of the virtual target object were considered during the FI calculation. As a result, the FI values accumulated on different positions in the scene according to the existing target object instances in the best case scenarios. Consequently, this modification lead to more reliable results in a large scene context.

A further improvement of the developed approach refers to the MFI calculation. According to the MFI method, in instances when several FI have the same probability value, the MFI is calculated as an average position of these FI when more than one FI has the same probability value. As in the experiments performed on the single view scans, the distance between the objects are smaller than in scenes containing large office rooms, the resulting distance values of the single view scans are correspondingly small. In instances when the highest FI are spread across the entire area of the single view scene being considered, the MFI position resulting from an average of these FI can comply with the center of the scene. As a result, the position of the MFI value might be located close to the object's real position, even if the decision is based on the available knowledge, a clear estimation of the position cannot be achieved. Therefore, the position of the objects placed closer to the center of the scene can still be estimated properly despite the missing knowledge. To make the resulting MFI less randomly selected, an appropriate clustering algorithm could be added to the MFI calculation.

4.2 Conclusion

Based on the results presented in the previous chapter 3.2.4 and considering the objectives stated in the thesis introduction 1.3 and 1.4, the following contributions are summarized:

Representation of spatial relations in a probabilistic and qualitative manner
In this thesis, the spatial relations are modeled from 3D data in a qualitative and probabilistic manner according to the formalism described in sec 2.2. By applying the formalism, the qualitative relations can be calculated from metric data, which makes

them suitable to be used for robotics applications regardless of their qualitative characteristic. Furthermore, the PQSR specify how probable a given spatial relation holds between two objects. The probabilistic aspect of the relations enables specification of how probable an object is in a given relation with another object, and not just whether a relation exists between the objects.

Precise definition of qualitative spatial relations
By applying the formalism defined in this thesis, the spatial relations can be modeled in a highly precise way. The PQSR and resulting FI enable calculation of the exact position in the environment at which a given spatial relation holds between two objects. Furthermore, the PQSR specify a 3D area in the environment to which a certain qualitative spatial relation refers to. Moreover, the PQSR enable estimation of how probable it is that a given object can be found at a particular position in the environment generally.

Learning of object occurrences from 3D data in a probabilistic manner
The formalism of the PQSR cannot only be applied to calculate the probability of a given spatial relation, but can also be used for learning this relation from the data in a probabilistic manner. The resulting co-occurrence probability specifies how probable a certain object class can be found in a certain relation with another object class. This knowledge is then used to calculate the SPF values in a real scene.

Prediction of object positions in an unknown environment
By applying the method developed in this thesis, prediction of the most probable object positions in an environment can be performed. This estimation can be achieved using the Field Intensity, which can be considered a combination of the learned knowledge about object co-occurrences and the probability of finding two given objects in a certain relation. Based on the FI value, the position in the environment most likely to contain the target object can be estimated. Since the FI already contains the general learned knowledge about object co-occurrences, the estimation can be calculated even in a newly encountered environment.

PQSR in a man-machine interaction context
By determining the highest FI value, the position in the environment at which a given object class is most likely to be found can be estimated. Thereby, the information about spatial relations is provided in a qualitative manner. Since humans prefer using qualitative instead of metric statements, the PQSR are more suitable for use in a man-machine interaction context.

4.3 Outlook

The work conducted within the scope of this thesis refers to the use of probabilistic qualitative spatial relations to predict the most probable position of a sought after object to consequently support object search. As described in the previous section, the developed method is evaluated with regards to its applicability for object search. The results presented in this thesis demonstrate that the method performs well in this context. However, there are several ways in which this method could be applied and used in contexts outside of object search purposes.

As the developed method provides promising results on a real-world data, the next evaluation step is to integrate the approach into a real robotic system. Since the algorithm was developed as an independent library, the method can be used as a component in an overall system such as a service or domestic robot. Importantly, the system must be equipped with sensors that enable three-dimensional data to be acquired, for example, a point cloud of the environment.

In consideration of the method's usability for a real system, the FI method could be extended to support navigation to the object's most probable position. For this feature, a cost function that includes information such as the cost to reach the object's estimated position can be calculated that includes information such as the the cost to reach the position where the object is estimated to be located. This cost could refer, for instance, to the shortest path to the closest MFI, and the MFI could then be reorganized depending on the path and not necessarily their value. The cost function could also be calculated with respect to the potential obstacles along the path to the position from which the area related to the MFI can be observed. Additionally, the best view of the area related to the estimated position could be considered by calculating the cost function. Further factors for the cost of reaching the object could also be related to the effort the robot has to expend to investigate the FI, for example, the number of way points the robot needs to drive through to reach the position. For instance, if the highest FI is located five meters from the robot's current position and to reach the second highest FI the robot only needs to turn in a given direction, it would be reasonable to first search for the object at the second highest FI position. This information would be considered by calculating the cost function.

Seven PQSR are defined and used in this thesis, but further work could investigate more spatial relations and then analyze how these additions influence the resulting position estimation. For instance, further conceivable relations could be in, below, and covered. The experiments performed in the present work demonstrate that by increasing the number of spatial relations considered in the FI calculation, a positive impact on the position estimation can be observed. However, a sufficient training data set is required to evaluate the exact influence of these additional relations on the overall method.

Because the PQSR investigated in this thesis include knowledge about typical object relations, this spatial information could also be considered in other robotics applications. For example, the approach developed in this thesis could also be used to enhance an object detection system. Complementary to the present work, there is already a body of research in which semantic-spatial knowledge is used to improve object recognition processes. According to the learned knowledge, some relations are more likely for particular object classes than others. This additional knowledge could be used together with an object detection system for more robust recognition of objects. For instance, the probability of a given object belonging to a object class increases if a required relation with a reference object at the object's position holds with high probability.

Another possible application for the method developed in this work is related to a placement task. The most probable object position can be estimated based on the PQSR and this estimated position can be used to find an object of a given object class as well as determining the location the object belongs with respect to the known spatial relations. By having this spatial semantic knowledge, the object could be placed at the estimated position. These kinds of tasks are related to the domain of domestic service robot where

the robot must handle the acts of locating and rearranging objects.

In the development of the current approach, annotated real data are used to estimate the object's most probable position. Additionally, the data contain the reference objects considered by the position calculation. In the current work, this information is presumed to be provided. However, in a real-world scenario, this important information is not always available and must first be extracted from the given data. Because the learned PQSR specify the probability of finding a target object in a given spatial relation with a certain reference object, this knowledge could also be used to identify and label the reference objects. To this end, the object cluster extracted from the data would first be randomly labeled with the possible object classes. In the subsequent step, all possible relation configurations are checked for the given labels. As a result, the probabilities for these labeled objects considering the learned PQSR are then calculated. Finally, an average probability is determined. During the calculation, each object of the current configuration is considered as the target and the remaining objects as reference objects. By permuting the labels of the objects in this way, it would be possible to find the object's most probable constellation, and thus, the labels of the reference object.

Bibliography

[AGP+11] A. Aydemir, M. Göbelbecker, A. Pronobis, K. Sjöö, and P. Jensfelt. Plan-based object search and exploration using semantic spatial knowledge in the real world. *Proc. of the European Conference on Mobile Robotics (ECMR 2011), Orebro, Sweden*, 2011.

[AJ12] A. Aydemir and P. Jensfelt. Exploiting and modeling local 3d structure for predicting object locations. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2012.

[AKJS12] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena. Contextually guided semantic labeling and search for three-dimensional point clouds. *The International Journal of Robotics Research*, 2012.

[All83] J. F Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[APJ12] A. Aydemir, A. Pronobis, and P. Jensfelt. Active visual search in unknown environments using uncertain semantics. *Transactions in Robotics*, 1:2329–2335, 2012.

[APS+11] A. Aydemir, A. Pronobis, K. Sjöö, M. Göbelbecker, and P. Jensfelt. Object search guided by semantic spatial knowledge. In *RSS 2011: Workshop on Grounding Human-Robot Dialog for Spatial Tasks*. RSS, 2011.

[ASF+11] A. Aydemir, K. Sjöö, J. Folkesson, A. Pronobis, and P. Jensfelt. Search in the real world: Active visual object search based on spatial relations. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2818–2824. IEEE, 2011.

[ASG+14] H. Ali, F. Shafait, E. Giannakidou, A. Vakali, N. Figueroa, T. Varvadoukas, and N. Mavridis. Contextual object category recognition for rgb-d scene labeling. *Robotics and Autonomous Systems*, 62(2):241 – 256, 2014.

[ASJ10] A. Aydemir, K. Sjöö, and P. Jensfelt. Object search on a mobile robot using relational spatial information. *Proc. Int. Conf. on Intelligent Autonomous Systems*, pages 111–120, 2010.

[BK91] J. Borenstein and Y. Koren. The vector field histogram-fast obstacle avoidance for mobile robots. *Robotics and Automation, IEEE Transactions on*, 7(3):278–288, Jun 1991.

[BMR82] I. Biederman, R. J. Mezzanotte, and J. C. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, 14(2):143–177, 1982.

[CBGG97] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Representing and reasoning with qualitative spatial relations about regions. In *Spatial and temporal reasoning*, pages 97–134. Springer, 1997.

[CG04] K. R. Coventry and S. C. Garrod. *Saying, seeing and acting: The psychological semantics of spatial prepositions.* Psychology Press, 2004.

[CH01] A. G. Cohn and S. M. Hazarika. Qualitative spatial representation and reasoning: An overview. Fundamenta Informaticae, 2001.

[CNR] Visual Computing Lab ISTI CNR. Meshlab. http://meshlab.sourceforge.net/.

[DCH10] K. SR. Dubba, A. G. Cohn, and D. C. Hogg. Event model learning from complex videos using ilp. In *ECAI*, volume 215, pages 93–98, 2010.

[DHH+09] S. K. Divvala, D. Hoiem, J. H. Hays, A. Efros, M. Hebert, et al. An empirical study of context in object detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1271–1278. IEEE, 2009.

[dia] Dia. https://wiki.gnome.org/Apps/Dia.

[ED10] M. Eich and M. Dabrowska. Semantic labeling: Classification of 3d entities based on spatial feature descriptors. In *Best Practice Algorithms in 3D Perception and Modeling for Mobile Manipualtion. IEEE International Conference on Robotics and Automation (ICRA-10), May 3, Anchorage, United States*. o.A., 5 2010.

[EDK10] M. Eich, M. Dabrowska, and F. Kirchner. 3d scene recovery and spatial scene analysis for unorganized point clouds. *Proceedings of the 13th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR-10)*, pages 21–28, 2010.

[EJvdMS14] J. Elfring, S. Jansen, R. van de Molengraft, and M. Steinbuch. Active object search exploiting probabilistic object–object relations. In *RoboCup 2013: Robot World Cup XVII*, pages 13–24. Springer, 2014.

[EKJ07] S. Ekvall, D. Kragic, and P. Jensfelt. Object detection and mapping for service robot tasks. *Robotica*, 25(02):175–187, 2007.

[FGMR10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[Fra92] A. U. Frank. Qualitative spatial reasoning about distances and directions in geographic space. *Journal of Visual Languages & Computing*, 3(4):343 – 371, 1992.

[Fre75] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156 – 171, 1975.

[Fre92] C. Freksa. *Using orientation information for qualitative spatial reasoning.* Springer, 1992.

[FRS+12] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan. Example-based synthesis of 3d object arrangements. *ACM Transactions on Graphics (TOG)*, 31(6):135, 2012.

[FSH11] M. Fisher, M. Savva, and P. Hanrahan. Characterizing structural relationships in scenes using graph kernels. In *ACM Transactions on Graphics (TOG)*, volume 30, page 34. ACM, 2011.

[Gar76] T. D. Garvey. *Perceptual Strategies for Purposive Vision.* PhD thesis, Stanford, CA, USA, 1976. AAI7613006.

[GH14a] M. Goldhoorn and R. Hartanto. Semantic labelling of 3d point clouds using spatial object constraints. In *In Special Session on Active Robot Vision (WARV 2014) of the 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP-2014)*, Lisbon, Portugal, 05–09 January 2014. IEEE Computer Society.

[GH14b] M. Goldhoorn and R. Hartanto. Semantic perception using spatial potential fields. In *In The 9th International Workshop on Cognitive Robotics (CogRob-2014) of the 21st European Conference on Artificial Intelligence (ECAI-2014), Prague, Czech Republic, 18–22 August, 2014*, 2014.

[GH17] M. Goldhoorn and R. Hartanto. Enhancing object detection by using probabilistic spatial-semantic knowledge. *Journal of Computers*, 12(1):68–75, January 2017.

[GK15] M. Goldhoorn and F. Kirchner. Semantic object recognition based on qualitative probabilistic spatial relations. In *Formal Modeling and Verification of Cyber-Physical Systems*, pages 278–280. Springer, 2015.

[Her86] A. Herskovits. *Language and Spatial Cognition. An Interdisciplinary Study of Prepositions in English.* Cambridge University Press, 1986.

[Her87] A. Herskovits. *Language and spatial cognition.* Cambridge University Press, 1987.

[HTD90] J. A. Hendler, A. Tate, and M. Drummond. AI planning: Systems and techniques. *AI magazine*, 11(2):61, 1990.

[JSZ+13] K. Johannsen, A. Swadzba, L. Ziegler, S. Wachsmuth, and J. P. De Ruiter. A computational model for reference object selection in spatial relations. In *Spatial Information Theory*. Springer, 2013.

[KAJS11] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena. Semantic labeling of 3d point clouds for indoor scenes. In *Advances in Neural Information Processing Systems*, pages 244–252, 2011.

[KBA+14] L. Kunze, C. Burbridge, M. Alberti, A. Tippur, J. Folkesson, P. Jensfelt, and N. Hawes. Combining top-down spatial reasoning and bottom-up object class recognition for scene understanding. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 2910–2915. IEEE, 2014.

[KBH14] L. Kunze, C. Burbridge, and N. Hawes. Bootstrapping probabilistic models of qualitative spatial relations for active visual object search. In *AAAI Spring Symposium*, pages 24–26, 2014.

[KBS+12] L. Kunze, M. Beetz, M. Saito, H. Azuma, K. Okada, and M. Inaba. Searching objects in large-scale indoor environments: A decision-theoretic approach. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4385–4390. IEEE, 2012.

[KDH14] L. Kunze, K. K. Doreswamy, and N. Hawes. Using qualitative spatial relations for indirect object search. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 163–168. IEEE, 2014.

[KJD11] A. Kasper, R. Jäkel, and R. Dillmann. Using spatial relations of objects in real world scenes for scene structuring and scene understanding. In *Advanced Robotics (ICAR), 2011 15th International Conference on*, pages 421–426. IEEE, 2011.

[KR09] T. Kollar and N. Roy. Utilizing object-object and object-scene context when planning to find things. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, ICRA'09, pages 4116–4121, Piscataway, NJ, USA, 2009. IEEE Press.

[kth] Kth-3d-total. http://www.cas.kth.se/data/kth-3d-total/.

[LC94] F. Lehmann and A. G. Cohn. The egg/yolk reliability hierarchy: Semantic data integration using sorts with prototypes. In *Proceedings of the third international conference on Information and knowledge management*, pages 272–279. ACM, 1994.

[LD04] J. Liu and L. K. Daneshmend. *Spatial Reasoning and Planning: Geometry, Mechanism, and Motion; with 6 Tables*. Springer Science & Business Media, 2004.

[Lev96] S. C. Levinson. Language and space. *Annual review of Anthropology*, pages 353–382, 1996.

[LFHU06] K. Lockwood, K. Forbus, D Halstead, and J. Usher. Automatic categorization of spatial prepositions. In *Proceedings of the 28th annual conference of the cognitive science society*, pages 1705–1710, 2006.

[MTBF03] R. Moratz, T. Tenbrink, J. Bateman, and K. Fischer. Spatial knowledge representation for human-robot interaction. In *Spatial cognition III*, pages 263–286. Springer, 2003.

[Nüc09] A. Nüchter. *3D robotic mapping: the simultaneous localization and mapping problem with six degrees of freedom*, volume 52. Springer Verlag, 2009.

[O'K99] J O'Keefe. The spatial prepositions. The MIT Press, 1999.

[QT09] A. Quattoni and A. Torralba. Recognizing indoor scenes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[RC01] T. Regier and L. A. Carlson. Grounding spatial language in perception: an empirical and computational investigation. *Journal of experimental psychology: General*, 130(2):273, 2001.

[RCC92] D. A. Randell, Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. *KR*, 92:165–176, 1992.

[RdSLS13] J. Ruiz-del Solar, P. Loncomilla, and M. Saavedra. A bayesian framework for informed search using convolutions between observation likelihoods and spatial relation masks. In *Advanced Robotics (ICAR), 2013 16th International Conference on*, pages 1–8. IEEE, 2013.

[Ree92] D. A. Reece. Selective perception for robot driving. Technical report, DTIC Document, 1992.

[Ren02] J. Renz. *Qualitative spatial reasoning with topological information*. Springer-Verlag, 2002.

[roc] ROCK, the Robot Construction Kit. http://www.rock-robotics.org.

[RTMF08] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173, 2008.

[SAJ12] K. Sjöö, A. Aydemir, and P. Jensfelt. Topological spatial relations for active visual search. *Robotics and Autonomous Systems*, 2012.

[SL07] T. Southey and J. J. Little. Learning qualitative spatial relations for object classification. In *IROS 2007 Workshop: From Sensors to Human Spatial Concepts*, 2007.

[SS03] P. Santos and M. Shanahan. A logic-based algorithm for image sequence interpretation and anchoring. In *IJCAI*, page 1408, 2003.

[ST10] K. Shubina and J. K. Tsotsos. Visual search for an object in a 3d environment using a mobile robot. *Computer Vision and Image Understanding*, 114(5):535 – 547, 2010. Special issue on Intelligent Vision Systems.

[TAA+14] A. Thippur, R. Ambrus, G. Agrawal, A. G. Del Burgo, J. H. Ramesh, M. K. Jha, M. B. S. S. Akhil, B. S. Nishan, J. Folkesson, and P. Jensfelt. Kth-3d-total: A 3d dataset for discovering spatial structures for long-term autonomous learning. In *Proc. of the International Conference on Automation, Robotics and Computer Vision (ICARCV 2014)*, 2014.

[TBK+15] A. Thippur, C. Burbridge, L. Kunze, M. Alberti, J. Folkesson, P. Jensfelt, and N. Hawes. A comparison of qualitative and metric spatial relation models for scene understanding. In *Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15), January 25–30, 2015, Austin Texas, USA*, 2015.

[Tso92] J. K. Tsotsos. On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, 7(2):127–141, 1992.

[VS08] S. Vasudevan and R. Siegwart. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *Robotics and Autonomous Systems*, 56(6):522–537, June 2008.

[WB94] L. E. Wixson and D. H. Ballard. Using intermediate objects to improve the efficiency of visual search. *International Journal of Computer Vision*, 12(2-3):209–230, 1994.

[WH05] T. Wagner and K. Hübner. An egocentric qualitative spatial knowledge representation based on ordering information for physical robot navigation. In *RoboCup 2004: Robot Soccer World Cup VIII*, pages 134–149. Springer, 2005.

[XH10] X. Xiong and D. Huber. Using context to create semantic 3d models of indoor environments. In *Proceedings of the British Machine Vision Conference (BMVC)*, September 2010.

[XLF12] R. Xiaofeng, B. Liefeng, and D. Fox. Rgb-(d) scene labeling: Features and algorithms. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2759–2766, June 2012.

[Zim93] K. Zimmermann. Enhancing qualitative spatial reasoning — combining orientation and distance. In AndrewU. Frank and Irene Campari, editors, *Spatial Information Theory A Theoretical Basis for GIS*, volume 716 of *Lecture Notes in Computer Science*, pages 69–76. Springer Berlin Heidelberg, 1993.

Appendices

A Additional Experimental Data

A.1 Learned PQSR from DFKI and KTH data sets

A.1.1 Learned PQSR from the DFKI data set

Table A.1: Learned average distances for objects in the spatial relation *in-front-of* (provided in meters).

	CB	CI	FL	WL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	-	-	0.1	-	0.1	0	0.3	-	-	-	-	-	-
CI	-	-	-	0.7	-	-	-	-	-	-	-	-	-
FL	-	2	-	1.6	0.2	0.1	0.2	0.1	0	-	0.4	0.1	0.1
TA	-	1.7	0.1	1.5	-	0.1	0.1	-	0	0.3	0.5	0.1	0
KB	-	2.1	0.2	1.7	0.1	-	0.3	0	0	0.2	0.3	0.1	0
MT	-	1.5	0.2	1.4	0.1	-	-	-	-	-	0.4	0	-
BO	-	-	-	2.1	0.5	-	0.3	-	-	-	0.3	-	-
MO	0.1	2.2	0.3	1.8	0.2	0.1	0.4	0	-	0.3	0.7	0.1	0
NB	-	-	0.4	-	-	-	0	-	-	-	-	-	-
PH	-	-	0.3	1.6	0.1	-	0.2	-	-	-	-	0.1	0
MU	-	2	0.2	1.5	0.2	0.1	0.3	-	0	-	0.6	-	0.1
BT	-	-	0.1	1.6	0.2	0.1	0.3	-	-	-	1.1	0.1	-

Table A.2: Learned average distances for objects in the spatial relation *behind-of* (provided in meters).

	CB	CI	FL	TA	KB	MT	BO	MO	NB	PH	MU	BT
CB	-	-	-	-	-	-	-	0.1	-	-	-	-
CI	-	-	2	1.7	2.1	1.5	-	2.2	-	-	2	-
FL	0.1	-	-	0.1	0.2	0.2	-	0.3	0.4	0.3	0.2	0.1
WL	-	0.7	1.6	1.5	1.7	1.4	2.1	1.8	-	1.6	1.5	1.6
TA	0.1	-	0.2	-	0.1	0.1	0.5	0.2	-	0.1	0.2	0.2
KB	0	-	0.1	0.1	-	-	-	0.1	-	-	0.1	0.1
MT	0.3	-	0.2	0.1	0.3	-	0.3	0.4	0	0.2	0.3	0.3
BO	-	-	0.1	-	0	-	-	0	-	-	-	-
MO	-	-	0	0	0	-	-	-	-	-	0	-
NB	-	-	-	0.3	0.2	-	-	0.3	-	-	-	-
PH	-	-	0.4	0.5	0.3	0.4	0.3	0.7	-	-	0.6	1.1
MU	-	-	0.1	0.1	0.1	0	-	0.1	-	0.1	-	0.1
BT	-	-	0.1	0	0	-	-	0	-	0	0.1	-

Table A.3: Learned average distances for objects in the spatial relation *on* (provided in meters).

	FL	WL	TA	KB
CB	0.3	-	-	-
FL	-	0.2	-	-
KB	-	-	0	-
MT	-	0.2	0.2	0.2
BO	-	-	0	-
MO	-	0.2	0	-
NB	-	-	0.1	-
PH	-	-	0	-
MU	-	0.2	0	-
BT	-	-	0	-

A.1.2 Learned PQSR from the KTH data set

Table A.4: The first part of the learned distances for objects in the spatial relation *on* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	MU	BT	CP	HP	PA	HI	MA	FL	PS
KB	-	0.1	-	0	0	0.2	-	-	-	-	-	-	-	-
MT	0.2	-	-	0.2	-	0.2	-	-	0	0.1	-	0.2	-	-
BO	0	0.3	-	0	-	0.1	-	-	-	-	-	-	0	0.2
MO	0.1	0.3	-	-	0	0.1	-	-	-	0	-	-	-	-
NB	0.2	0.2	-	0.4	-	0.2	0.3	0.2	0	0	-	-	-	-
PH	-	-	-	-	-	-	0.1	-	-	-	-	-	-	0.2
MU	0	0.2	0	0.1	0	-	0.1	-	0	-	-	0	-	0
BT	0.1	0.4	-	-	-	-	-	-	-	-	-	-	-	0.2
CP	0	0.3	-	0.1	-	0.3	-	-	-	0	-	-	-	-
HP	0.2	0.1	0	0.3	-	0	-	-	-	0	-	0.1	0	-
PC	-	0.2	-	-	-	-	-	-	-	-	-	-	-	-
PA	0.1	0.3	-	0.1	-	0	0.1	-	-	-	-	0.1	0.1	-
PN	-	0.2	-	-	0	0	-	-	-	0	-	-	-	-
HI	-	0.2	-	-	-	0	0.4	-	-	-	-	-	-	-
MA	0	0.2	-	0.1	-	-	-	-	-	-	-	-	0	-
FL	-	-	0.3	-	-	-	-	-	-	-	-	-	-	-
PS	0.1	0.2	-	-	-	0	-	-	0	-	-	-	-	-
LP	0.2	0.2	0.1	0.3	0.1	0.4	-	0.1	0.2	-	0.4	-	0.1	-
FA	0.3	0	-	0.4	-	0.2	0.1	-	0.1	0.1	-	0.3	-	0.1
GL	-	-	-	-	-	-	0.3	-	-	-	-	-	-	0
JU	0.1	0.4	-	0.2	-	0	-	-	0	0	-	0.1	-	0.1
RU	-	0.2	-	-	-	-	0.1	-	0	0	-	-	-	-
NP	-	0.3	-	-	-	-	-	-	0	-	-	-	0.1	-

Table A.5: The second part of the learned distances for objects in the spatial relation *on* from the KTH data set (given in meters).

	LP	FA	GL	JU	CU	NP
KB	0.1	-	-	-	-	-
MT	0.1	-	-	0	-	-
BO	0.2	-	0.2	-	-	0
MO	0.1	-	-	-	-	-
NB	0.1	-	-	-	-	-
PH	-	-	-	-	-	-
MU	0.1	0.2	-	0.1	-	-
BT	0.1	0.3	-	0.4	0.1	-
CP	0.1	-	-	-	-	-
HP	0.3	-	-	0.1	-	0
PC	0.1	-	-	-	-	-
PA	0.2	-	-	-	-	-
PN	0.2	-	-	-	-	0
HI	0	-	-	-	-	0
MA	0.1	-	-	-	-	-
FL	0.3	-	-	-	-	-
PS	0.2	0.2	0	0.1	-	-
LP	-	0.2	0.4	-	-	0.2
FA	0.2	-	-	0.1	-	-
GL	0.2	-	-	0	-	-
JU	0.1	-	-	-	-	-
RU	0	-	-	-	-	-
NP	0.3	-	-	0.1	-	-

Table A.6: The first part of the target object occurrences and the number of valid *on* relation between object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	MU	BT	CP	HP	PA	HI	MA
KB	410	0	28	0	6	2	2	0	0	0	0	0	0
MT	452	16	0	0	2	0	4	0	0	2	1	0	1
BO	163	1	8	0	1	0	1	0	0	0	0	0	0
MO	409	1	24	0	0	1	2	0	0	0	1	0	0
NB	196	2	11	0	1	0	1	1	1	1	1	0	0
PH	86	0	0	0	0	0	0	1	0	0	0	0	0
MU	233	4	20	1	3	1	0	3	0	1	0	0	2
BT	95	1	4	0	0	0	0	0	0	0	0	0	0
CP	40	1	3	0	1	0	1	0	0	0	1	0	0
HP	117	1	5	1	1	0	2	0	0	0	2	0	1
PC	32	0	2	0	0	0	0	0	0	0	0	0	0
PA	320	2	4	0	2	0	2	1	0	0	0	0	2
PN	119	0	4	0	0	2	1	0	0	0	2	0	0
HI	74	0	3	0	0	0	2	1	0	0	0	0	0
MA	68	2	4	0	1	0	0	0	0	0	0	0	0
FL	66	0	0	1	0	0	0	0	0	0	0	0	0
PS	133	2	15	0	0	0	10	0	0	1	0	0	0
LP	260	10	35	4	1	10	1	0	2	1	0	1	0
FA	35	1	2	0	1	0	2	2	0	2	2	0	1
GL	30	0	0	0	0	0	0	1	0	0	0	0	0
JU	50	2	1	0	2	0	2	0	0	2	1	0	2
RU	28	0	3	0	0	0	0	1	0	1	1	0	0
NP	64	0	1	0	0	0	0	0	0	1	0	0	0

Table A.7: The second part of the target object occurrences and the number of valid *on* relation between object classes learned from the KTH data set.

	FL	PS	LP	FA	GL	JU	CU	NP
KB	0	0	25	0	0	0	0	0
MT	0	0	11	0	0	2	0	0
BO	1	1	21	0	1	0	0	1
MO	0	0	18	0	0	0	0	0
NB	0	0	31	0	0	0	0	0
PH	0	3	0	0	0	0	0	0
MU	0	1	17	1	0	2	0	0
BT	0	3	10	1	0	2	4	0
CP	0	0	8	0	0	0	0	0
HP	4	0	4	0	0	4	0	4
PC	0	0	2	0	0	0	0	0
PA	1	0	6	0	0	0	0	0
PN	0	0	8	0	0	0	0	4
HI	0	0	13	0	0	0	0	1
MA	1	0	3	0	0	0	0	0
FL	0	0	1	0	0	0	0	0
PS	0	0	8	1	1	4	0	0
LP	1	0	0	5	1	0	0	2
FA	0	1	9	0	0	2	0	0
GL	0	1	4	0	0	1	0	0
JU	0	4	1	0	0	0	0	0
RU	0	0	1	0	0	0	0	0
NP	3	0	2	0	0	1	0	0

Table A.8: The first part of the learned distances for objects in the spatial relation *behind-of* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	-	0	0.2	0.1	0.1	0.1	0.1	0.1	0.2	-	0.1	0.1	0.1	0.1
MT	0.1	-	0.3	0.2	0.2	0.3	0.2	0.2	0.2	0.5	0.2	0.2	0.2	0.2
BO	0.1	0	-	0.1	0.2	0.1	0.2	0.1	0	-	0.1	0.1	0.1	0.1
MO	0	0	0	-	0.1	0.1	0	0.1	0.1	-	0	0.1	0	0.1
NB	0.1	0	0.2	0.2	-	0.2	0.1	0.2	0.2	0.3	0.2	0.1	0.2	0.2
PH	0.2	0	0	0	0.1	-	-	-	0.3	0.5	-	0	-	0.1
MU	0.1	0.1	0.1	0.1	0.1	0.1	-	0	0.1	-	0.1	0.1	0.1	0.1
BT	0	0.1	0	0.1	0.2	0.1	0.1	-	0.1	-	0.2	0	0.1	0.1
CP	0	0	0.2	0.1	0.1	0	0.1	0	-	-	0.2	0	0.1	0.2
HP	0.1	0	0.1	0.1	0.1	0	0.1	0.1	0.1	-	-	0.1	0.1	0.1
PC	0.2	0	0.1	0.2	0.2	0.1	0	0.2	0	-	0	-	0	0
PA	0.1	0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	-	0.1	0.2	-	0.1
PN	0.1	0	0.1	0.1	0	0.1	0	0.1	0	-	0.1	0.1	0.1	-
HI	0.1	0	0.1	0.1	0	0.1	0.1	0.2	-	-	0.1	0.1	0.2	0.1
MA	0.1	0	0.1	0.1	0	0.2	0.1	0.1	-	-	0	0.1	0.1	0.1
FL	0.1	0	0.2	0.1	0.1	0.2	0.1	0	0.2	0.4	0.2	0.1	0.2	0.2
PS	0	0	0.2	0.1	0.1	0.2	0	0	0.2	0.3	0.1	0.1	0.2	0.1
LP	0.1	0	0.3	0.2	0.2	0.2	0.2	0.2	0.3	0.6	0.2	0.2	0.2	0.2
FA	0.1	0	-	0.2	0.2	0.1	0.1	0.1	0	-	0.1	0.1	0.1	0.2
GL	0.1	0.1	0.1	0.1	0.4	-	0	0.2	-	-	0.2	-	0.2	0
JU	0.1	0	0.3	0	0.1	-	0	0.4	-	-	0.2	0.4	0	0.3
BA	0.2	0	-	-	-	-	-	-	-	-	0.2	-	-	-
CU	0.1	-	-	0.1	0	-	-	0.1	-	-	-	-	-	0.3
RU	0.2	-	0.1	0	0.1	0	0	0.1	-	-	0	0	0.1	0.1
NP	0.2	0.1	0.1	0	0.1	0.1	0.1	0.1	0	-	0.2	0.2	0.2	0.1

Table A.9: The second part of the learned distances for objects in the spatial relation *behind-of* from the KTH data set (given in meters).

	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	EX	NP
KB	0.1	0.1	0.1	0.1	0	0.1	0.2	0.2	-	-	0.1	-	0.1
MT	0.2	0.2	0.1	0.1	0.1	0.2	0.3	0.2	0	0.1	0.2	-	0.2
BO	0.1	0.2	0.2	0.3	0.1	0.2	0	0.2	-	-	0.1	-	0.2
MO	0	0.1	0	0	0	0	0.1	0.1	-	-	0.1	-	0
NB	0.1	0.1	0.2	0.2	0.1	0.4	0.2	0.2	0	-	0.1	-	0.2
PH	0.1	0	0	0.2	0.1	-	-	-	-	-	0.1	0.1	0.2
MU	0.1	0	-	0	0.1	0.1	0	0.1	-	-	0.1	-	0.1
BT	-	0	0.2	0.1	0	0.1	0.2	0.3	-	-	-	-	0.2
CP	0	0	-	0.1	0.1	0	-	-	-	-	0	-	0.3
HP	0.1	0.1	0	0	0	0.1	0.1	0.1	-	-	0.1	-	0.1
PC	0	0	0	0.2	0	0	-	-	-	0.1	0.1	-	0.2
PA	0.2	0.1	0.1	0	0.1	0.2	0.1	0.1	-	-	0.1	-	0.1
PN	0	0	0	0.1	0	0.1	0.1	0.2	-	0.1	0	-	0
HI	-	0	0.1	-	0	0.2	-	0	-	-	-	-	0.1
MA	0	-	-	0	0	0.1	0	0	-	-	-	-	0.1
FL	0.1	0.1	-	0.1	0.1	0.3	0.1	0.1	-	-	-	-	0.2
PS	-	0	0.1	-	0	0.1	0.2	0.1	-	-	0.2	0.3	0.2
LP	0.2	0.2	0.1	0.2	-	0.2	0.3	0.3	-	0.2	0.1	-	0.2
FA	0.2	0.2	0	0	0.1	-	-	0	-	-	0	-	0.2
GL	-	0	-	0	0	0.1	-	0	-	-	-	-	0.2
JU	0.3	0.2	-	0.1	0.1	0.1	0.1	-	-	-	-	-	0.1
BA	-	-	-	-	0	-	-	-	-	-	-	-	-
CU	-	-	-	-	-	-	-	-	-	-	-	-	0.3
RU	-	-	0	-	0.1	-	-	-	-	0	-	-	0.1
NP	0.1	0.2	-	0.1	0.1	0.1	0.1	0.1	-	-	0.2	-	-

Table A.10: The first part of the target object occurrences and the number of valid *behind-of* relation between object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC
KB	410	0	2	79	222	71	34	94	41	20	0	65	22
MT	452	393	0	116	360	154	58	194	69	39	1	115	27
BO	163	52	20	0	13	21	2	14	3	4	0	22	3
MO	409	110	4	23	0	22	22	37	14	5	0	9	8
NB	196	91	17	37	110	0	20	49	15	21	1	30	11
PH	86	23	18	2	6	10	0	0	0	5	1	0	3
MU	233	101	1	21	96	13	4	0	5	2	0	29	2
BT	95	30	3	6	33	5	10	20	0	1	0	1	4
CP	40	8	1	3	7	4	2	3	1	0	0	2	1
HP	117	39	2	27	16	12	2	16	4	4	0	0	10
PC	32	7	3	4	5	4	1	4	1	1	0	3	0
PA	320	96	16	27	49	24	22	36	10	7	0	28	4
PN	119	35	7	9	24	15	6	15	6	7	0	8	8
HI	74	31	8	12	20	5	6	9	6	0	0	7	2
MA	68	23	4	8	22	2	2	6	1	0	0	6	3
FL	66	26	16	7	15	14	6	8	2	2	1	12	1
PS	133	58	1	10	60	26	25	16	17	8	1	14	4
LP	260	226	100	66	212	80	36	114	49	23	1	75	11
FA	35	18	4	0	6	3	3	10	4	2	0	16	1
GL	30	6	2	3	5	2	0	2	4	0	0	4	0
JU	50	19	6	6	14	3	0	13	2	0	0	12	4
BA	3	2	2	0	0	0	0	0	0	0	0	2	0
CU	6	6	0	0	5	1	0	0	4	0	0	0	0
RU	28	7	0	4	6	3	2	2	4	0	0	4	3
NP	64	20	7	8	11	5	3	11	1	2	0	6	1

Table A.11: The second part of the target object occurrences and the number of valid *behind-of* relation between object classes learned from the KTH data set.

	PA	PN	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU
KB	79	60	28	31	11	53	23	13	16	28	0	0	17
MT	205	97	53	54	43	128	144	29	27	43	1	6	26
BO	32	14	7	6	8	9	37	5	4	3	0	0	2
MO	34	23	5	4	5	12	12	1	7	15	0	0	9
NB	52	45	11	10	12	28	39	2	9	5	3	0	8
PH	0	2	2	2	12	19	4	0	0	0	0	0	1
MU	27	23	13	4	0	10	29	2	5	13	0	0	3
BT	16	6	0	3	1	12	13	7	3	6	0	0	0
CP	4	2	1	1	0	2	6	1	0	0	0	0	1
HP	21	19	11	9	6	5	19	4	5	5	0	0	4
PC	3	2	3	1	1	1	4	1	0	0	0	1	8
PA	0	18	10	9	10	12	12	2	5	6	0	0	2
PN	14	0	3	2	1	2	8	2	1	2	0	2	5
HI	11	10	0	3	2	0	11	2	0	4	0	0	0
MA	7	6	1	0	0	1	7	1	1	1	0	0	0
FL	16	5	5	5	0	25	12	2	3	1	0	0	0
PS	19	11	0	4	5	0	22	5	4	9	0	0	4
LP	126	75	42	37	27	68	0	26	19	31	0	4	12
FA	11	5	4	6	2	1	3	0	0	10	0	0	1
GL	6	1	0	1	0	6	8	4	0	2	0	0	0
JU	10	7	7	7	0	2	7	1	1	0	0	0	0
BA	0	0	0	0	0	0	3	0	0	0	0	0	0
CU	0	1	0	0	0	0	0	0	0	0	0	0	0
RU	4	2	0	0	1	0	5	0	0	0	0	2	0
NP	7	13	2	2	0	3	11	2	3	7	0	0	2

Table A.12: The first part of the learned distances for objects in the spatial relation *in-front-of* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	HP	PC	PA	PN	HI
KB	-	0.1	0.1	0	0.1	0.2	0.1	0	0	0.1	0.2	0.1	0.1	0.1
MT	0	-	0	0	0	0	0.1	0.1	0	0	0	0	0	0
BO	0.2	0.3	-	0	0.2	0	0.1	0	0.2	0.1	0.1	0.1	0.1	0.1
MO	0.1	0.2	0.1	-	0.2	0	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
NB	0.1	0.2	0.2	0.1	-	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0	0
PH	0.1	0.3	0.1	0.1	0.2	-	0.1	0.1	0	0	0.1	0.2	0.1	0.1
MU	0.1	0.2	0.2	0	0.1	-	-	0.1	0.1	0.1	0	0.1	0	0.1
BT	0.1	0.2	0.1	0.1	0.2	-	0	-	0	0.1	0.2	0.1	0.1	0.2
CP	0.2	0.2	0	0.1	0.2	0.3	0.1	0.1	-	0.1	0	0.1	0	-
KS	-	0.5	-	-	0.3	0.5	-	-	-	-	-	-	-	-
HP	0.1	0.2	0.1	0	0.2	-	0.1	0.2	0.2	-	0	0.1	0.1	0.1
PC	0.1	0.2	0.1	0.1	0.1	0	0.1	0	0	0.1	-	0.2	0.1	0.1
PA	0.1	0.2	0.1	0	0.2	-	0.1	0.1	0.1	0.1	0	-	0.1	0.2
PN	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0	0.1	-	0.1
HI	0.1	0.2	0.1	0	0.1	0.1	0.1	-	0	0.1	0	0.2	0	-
MA	0.1	0.2	0.2	0.1	0.1	0	0	0	0	0.1	0	0.1	0	0
FL	0.1	0.1	0.2	0	0.2	0	-	0.2	-	0	0	0.1	0	0.1
PS	0.1	0.1	0.3	0	0.2	0.2	0	0.1	0.1	0	0.2	0	0.1	-
LP	0	0.1	0.1	0	0.1	0.1	0.1	0	0.1	0	0	0.1	0	0
FA	0.1	0.2	0.2	0	0.4	-	0.1	0.1	0	0.1	0	0.2	0.1	0.2
GL	0.2	0.3	0	0.1	0.2	-	0	0.2	-	0.1	-	0.1	0.1	-
JU	0.2	0.2	0.2	0.1	0.2	-	0.1	0.3	-	0.1	-	0.1	0.2	0
BA	-	0	-	-	0	-	-	-	-	-	-	-	-	-
CU	-	0.1	-	-	-	-	-	-	-	-	0.1	-	0.1	-
RU	0.1	0.2	0.1	0.1	0.1	0.1	0.1	-	0	0.1	0.1	0.1	0	-
EX	-	-	-	-	-	0.1	-	-	-	-	-	-	-	-
NP	0.1	0.2	0.2	0	0.2	0.2	0.1	0.2	0.3	0.1	0.2	0.1	0	0.1

Table A.13: The second part of the learned distances for objects in the spatial relation *in-front-of* from the KTH data set (given in meters).

	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	NP
KB	0.1	0.1	0	0.1	0.1	0.1	0.1	0.2	0.1	0.2	0.2
MT	0	0	0	0	0	0.1	0	0	-	-	0.1
BO	0.1	0.2	0.2	0.3	-	0.1	0.3	-	-	0.1	0.1
MO	0.1	0.1	0.1	0.2	0.2	0.1	0	-	0.1	0	0
NB	0	0.1	0.1	0.2	0.2	0.4	0.1	-	0	0.1	0.1
PH	0.2	0.2	0.2	0.2	0.1	-	-	-	-	0	0.1
MU	0.1	0.1	0	0.2	0.1	0	0	-	-	0	0.1
BT	0.1	0	0	0.2	0.1	0.2	0.4	-	0.1	0.1	0.1
CP	-	0.2	0.2	0.3	0	-	-	-	-	-	0
KS	-	0.4	0.3	0.6	-	-	-	-	-	-	-
HP	0	0.2	0.1	0.2	0.1	0.2	0.2	0.2	-	0	0.2
PC	0.1	0.1	0.1	0.2	0.1	-	0.4	-	-	0	0.2
PA	0.1	0.2	0.2	0.2	0.1	0.2	0	-	-	0.1	0.2
PN	0.1	0.2	0.1	0.2	0.2	0	0.3	-	0.3	0.1	0.1
HI	0	0.1	-	0.2	0.2	-	0.3	-	-	-	0.1
MA	-	0.1	0	0.2	0.2	0	0.2	-	-	-	0.2
FL	-	-	0.1	0.1	0	-	-	-	-	0	-
PS	0	0.1	-	0.2	0	0	0.1	-	-	-	0.1
LP	0	0.1	0	-	0.1	0	0.1	0	-	0.1	0.1
FA	0.1	0.3	0.1	0.2	-	0.1	0.1	-	-	-	0.1
GL	0	0.1	0.2	0.3	-	-	0.1	-	-	-	0.1
JU	0	0.1	0.1	0.3	0	0	-	-	-	-	0.1
BA	-	-	-	-	-	-	-	-	-	-	-
CU	-	-	-	0.2	-	-	-	-	-	0	-
RU	-	-	0.2	0.1	0	-	-	-	-	-	0.2
EX	-	-	0.3	-	-	-	-	-	-	-	-
NP	0.1	0.2	0.2	0.2	0.2	0.2	0.1	-	0.3	0.1	-

Table A.14: The first part of the target object occurrences and the number of valid *in-front-of* relation between object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	PH	MU	BT	CP	HP	PC	PA
KB	410	0	393	52	110	91	23	101	30	8	39	7	96
MT	452	2	0	20	4	17	18	1	3	1	2	3	16
BO	163	79	116	0	23	37	2	21	6	3	27	4	27
MO	409	222	360	13	0	110	6	96	33	7	16	5	49
NB	196	71	154	21	22	0	10	13	5	4	12	4	24
PH	86	34	58	2	22	20	0	4	10	2	2	1	22
MU	233	94	194	14	37	49	0	0	20	3	16	4	36
BT	95	41	69	3	14	15	0	5	0	1	4	1	10
CP	40	20	39	4	5	21	5	2	1	0	4	1	7
KS	1	0	1	0	0	1	1	0	0	0	0	0	0
HP	117	65	115	22	9	30	0	29	1	2	0	3	28
PC	32	22	27	3	8	11	3	2	4	1	10	0	4
PA	320	79	205	32	34	52	0	27	16	4	21	3	0
PN	119	60	97	14	23	45	2	23	6	2	19	2	18
HI	74	28	53	7	5	11	2	13	0	1	11	3	10
MA	68	31	54	6	4	10	2	4	3	1	9	1	9
FL	66	11	43	8	5	12	12	0	1	0	6	1	10
PS	133	53	128	9	12	28	19	10	12	2	5	1	12
LP	260	23	144	37	12	39	4	29	13	6	19	4	12
FA	35	13	29	5	1	2	0	2	7	1	4	1	2
GL	30	16	27	4	7	9	0	5	3	0	5	0	5
JU	50	28	43	3	15	5	0	13	6	0	5	0	6
BA	3	0	1	0	0	3	0	0	0	0	0	0	0
CU	6	0	6	0	0	0	0	0	0	0	0	1	0
RU	28	17	26	2	9	8	1	3	0	1	4	8	2
EX	2	0	0	0	0	0	2	0	0	0	0	0	0
NP	64	36	57	16	15	21	1	19	7	2	17	2	25

Table A.15: The second part of the target object occurrences and the number of valid *in-front-of* relation between object classes learned from the KTH data set.

	PN	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	NP
KB	35	31	23	26	58	226	18	6	19	2	6	7	20
MT	7	8	4	16	1	100	4	2	6	2	0	0	7
BO	9	12	8	7	10	66	0	3	6	0	0	4	8
MO	24	20	22	15	60	212	6	5	14	0	5	6	11
NB	15	5	2	14	26	80	3	2	3	0	1	3	5
PH	6	6	2	6	25	36	3	0	0	0	0	2	3
MU	15	9	6	8	16	114	10	2	13	0	0	2	11
BT	6	6	1	2	17	49	4	4	2	0	4	4	1
CP	7	0	0	2	8	23	2	0	0	0	0	0	2
KS	0	0	0	1	1	1	0	0	0	0	0	0	0
HP	8	7	6	12	14	75	16	4	12	2	0	4	6
PC	8	2	3	1	4	11	1	0	4	0	0	3	1
PA	14	11	7	16	19	126	11	6	10	0	0	4	7
PN	0	10	6	5	11	75	5	1	7	0	1	2	13
HI	3	0	1	5	0	42	4	0	7	0	0	0	2
MA	2	3	0	5	4	37	6	1	7	0	0	0	2
FL	1	2	0	0	5	27	2	0	0	0	0	1	0
PS	2	0	1	25	0	68	1	6	2	0	0	0	3
LP	8	11	7	12	22	0	3	8	7	3	0	5	11
FA	2	2	1	2	5	26	0	4	1	0	0	0	2
GL	1	0	1	3	4	19	0	0	1	0	0	0	3
JU	2	4	1	1	9	31	10	2	0	0	0	0	7
BA	0	0	0	0	0	0	0	0	0	0	0	0	0
CU	2	0	0	0	0	4	0	0	0	0	0	2	0
RU	5	0	0	0	4	12	1	0	0	0	0	0	2
EX	0	0	0	0	2	0	0	0	0	0	0	0	0
NP	4	8	6	12	11	43	3	1	1	0	1	4	0

Table A.16: The first part of the learned distances for objects in the spatial relation *left-of* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	-	0	0.5	0.2	0.4	0.6	0.4	0.4	0.3	-	0.2	0.2	0.5	0.3
MT	0	-	0.3	0.2	0.4	0.5	0.2	0.4	0.5	-	0.1	0.2	0.4	0.3
BO	0.5	0.5	-	0.5	0.5	0.5	0.3	0.4	0.1	-	0.3	0.1	0.2	0.3
MO	0	0.1	0.3	-	0.2	0.4	0.1	0.2	0.1	-	0.1	0	0.3	0.1
NB	0.4	0.3	0.5	0.5	-	0.8	0.5	0.4	0.3	-	0.2	0.4	0.5	0.4
PH	0.9	0.5	0.4	0.8	0.5	-	-	0	0.4	0	-	-	0.2	0.6
MU	0.3	0.3	0.2	0.2	0.3	0.5	-	0.2	0.1	-	0.2	0.1	0.3	0.2
BT	0.5	0.4	0.3	0.1	0.2	0	0.1	-	0.4	-	0.2	0	0.2	0.1
CP	0.3	0.3	0.2	0.3	0.2	0.5	0.6	0.4	-	-	0.1	-	0.3	0.1
KS	-	0.8	-	-	0.3	-	-	-	-	-	-	-	-	-
HP	0.4	0.4	0.4	0.4	0.3	0.8	0.4	0.3	0.2	-	-	0.3	0.3	0.2
PC	0.2	0.2	0.2	0.1	0.5	0.6	0.3	0	0	-	0.2	-	0.1	0.1
PA	0.5	0.4	0.2	0.3	0.5	0.6	0.2	0.4	0.1	-	0.3	0.2	-	0.1
PN	0.2	0.3	0.3	0.2	0.3	0.4	0.1	0.2	0	-	0.1	0.1	0.2	-
HI	0.3	0.2	0.1	0.1	0.4	0.5	0.1	0	-	-	0.2	0.1	0.1	0
MA	0.3	0.3	0.1	0.1	0.3	0.4	0.1	0.3	-	-	0.1	0	0.3	0
FL	0.5	0.7	0.2	0.5	0.6	0.3	0.5	0.9	0.4	0.3	0.2	0.9	0.5	0.3
PS	0.6	0.6	0.1	0.1	0.3	0.2	0.1	0.2	0.3	-	-	-	0.3	0.2
LP	0.4	0.4	0.4	0.6	0.4	-	0.7	0.7	0.2	-	0.4	0.7	0.6	0.5
FA	0.5	0.3	0.5	0.3	-	0.1	0.2	0.2	0.3	-	0.1	0.1	0.2	0.2
GL	0.2	0.4	0.1	0.1	0.1	-	0.3	0.2	-	-	0.2	-	0.2	0.2
JU	0.4	0.4	0.4	0.4	0.2	-	0.1	0.1	-	-	0	0.2	0.4	0.2
BA	-	0.2	-	-	-	-	-	-	-	-	-	-	-	-
CU	-	-	-	-	-	-	-	0	-	-	-	-	-	-
RU	0.1	0.3	0.1	0.1	0.2	0.5	0.3	0	0.1	-	0.2	0.1	0.2	0.1
EX	-	-	-	-	-	0	-	-	-	-	-	-	-	-
NP	0.3	0.4	0.3	0.4	0.5	0.7	0.4	0.4	0.2	-	0.2	0.3	0.3	0.2

Table A.17: The second part of the learned distances for objects in the spatial relation *left-of* from the KTH data set (given in meters).

	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	EX	NP
KB	0.3	0.3	0.5	0.4	0.3	0.4	0.3	0.6	0	0.5	0.3	-	0.3
MT	0.5	0.3	0.6	0.4	0.3	0.3	0.3	0.2	-	0.5	0.2	-	0.3
BO	0.3	0.2	0.2	0.3	0.5	0.2	0.4	0.1	-	-	0.4	-	0.4
MO	0.1	0.1	0.5	0.3	0.2	0.2	0.2	0.3	-	0.2	0.1	-	0.2
NB	0.5	0.7	0.9	0.5	0.4	0.5	0.3	0.7	0.5	1	0.5	-	0.6
PH	0.2	0.2	0.3	0	0.9	0	-	-	-	-	-	-	0.2
MU	0.2	0.1	0.4	0.1	0.5	0.2	0.2	0.1	-	-	0.1	-	0.5
BT	0.2	0.1	0.2	0.3	0.4	0.2	0	0.2	-	0.1	0.5	-	-
CP	0.2	0.1	-	0.1	0.4	0.1	-	-	-	-	-	-	0.3
KS	-	-	-	0	0.6	-	-	-	-	-	-	-	-
HP	0.2	0.3	0.1	0.3	0.9	0.3	0.3	0.2	0.3	-	0.2	-	0.4
PC	-	0.1	0.1	0.3	0.8	0.2	-	-	-	0	0.1	-	-
PA	0.2	0.2	0.2	0.2	0.5	0.3	0.3	0.4	-	-	0.1	-	0.3
PN	0	0	0.1	0.1	0.4	0.1	0.3	0.2	-	0.2	0	-	0.1
HI	-	0.1	0.1	-	0.5	0	-	0.2	-	-	-	-	0
MA	0	-	0	0.3	0.8	0.1	0.1	0.2	-	-	-	-	0.2
FL	0.3	0.3	-	0.4	0.8	0.3	0.4	-	-	-	-	-	0.4
PS	-	0.2	0.5	-	0.4	0.2	0.3	0	-	-	-	0.3	0.2
LP	0.4	0.4	0.7	0.7	-	0.2	0.6	0.7	-	0.8	0.7	-	0.6
FA	0.2	0.3	0.3	0	0.4	-	0.5	0.3	-	-	-	-	0.3
GL	-	-	-	0	0.4	0.3	-	0	-	-	-	-	0.3
JU	0.2	0.3	0.3	0.2	0.6	0.3	-	-	-	-	-	-	0.1
BA	-	-	-	-	0.5	-	-	-	-	-	-	-	-
CU	-	-	-	-	-	-	-	-	-	-	-	-	-
RU	-	-	0.2	0.2	0.7	0	-	-	-	0.4	-	-	0.2
EX	-	-	-	-	-	-	-	-	-	-	-	-	-
NP	0.3	0.2	0	0.3	0.6	0.2	0.3	0.4	-	0.5	0.3	-	-

Table A.18: The first part of the target object occurrences and the number of valid *left-of* relation between object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC
KB	410	0	111	27	315	33	26	117	57	5	0	8	14
MT	452	87	0	25	252	35	41	94	51	12	0	12	11
BO	163	104	116	0	25	39	2	29	4	6	0	37	3
MO	409	1	3	10	0	27	29	104	41	4	0	5	3
NB	196	131	139	20	105	0	17	47	12	19	0	19	8
PH	86	31	28	2	1	13	0	0	5	4	1	0	0
MU	233	65	58	9	28	20	4	0	8	4	0	11	2
BT	95	10	13	6	2	8	4	15	0	1	0	1	4
CP	40	21	18	1	6	5	3	1	1	0	0	2	0
KS	1	0	1	0	0	1	0	0	0	0	0	0	0
HP	117	97	97	15	20	23	2	36	4	4	0	0	11
PC	32	15	12	4	10	6	4	4	1	2	0	2	0
PA	320	108	109	20	33	41	6	43	12	6	0	26	4
PN	119	44	44	8	14	15	7	23	6	4	0	6	5
HI	74	22	25	5	12	13	6	14	2	0	0	5	5
MA	68	27	27	4	7	7	2	8	2	0	0	8	2
FL	66	29	45	7	7	18	13	5	1	2	1	9	1
PS	133	40	38	6	23	16	25	4	20	5	0	0	0
LP	260	158	168	40	175	27	0	118	38	15	0	48	9
FA	35	20	22	2	3	0	2	9	6	1	0	15	1
GL	30	5	8	4	2	3	0	2	2	0	0	3	0
JU	50	31	20	5	12	4	0	11	2	0	0	6	4
BA	3	0	1	0	0	0	0	0	0	0	0	0	0
CU	6	0	0	0	0	0	0	0	1	0	0	0	0
RU	28	8	7	2	4	5	3	3	3	1	0	4	4
EX	2	0	0	0	0	0	2	0	0	0	0	0	0
NP	64	36	28	10	15	8	1	22	8	1	0	11	3

Table A.19: The second part of the target object occurrences and the number of valid *left-of* relation between object classes learned from the KTH data set.

	PA	PN	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU
KB	59	48	24	24	8	71	81	10	16	16	1	6	16
MT	66	31	12	14	14	93	72	6	15	20	0	6	16
BO	34	15	15	12	7	13	65	3	5	6	0	0	4
MO	47	33	10	15	11	49	26	4	8	17	0	5	11
NB	49	42	6	5	7	37	92	5	9	4	3	1	5
PH	18	1	2	2	5	19	40	1	0	0	0	0	0
MU	26	14	8	3	3	21	27	2	5	15	0	0	2
BT	15	7	4	2	2	8	18	5	5	6	0	3	1
CP	4	5	1	1	0	5	14	2	0	0	0	0	0
KS	0	0	0	0	0	1	1	0	0	0	0	0	0
HP	23	21	12	13	9	19	46	5	5	11	2	0	4
PC	5	5	0	2	1	5	6	1	0	0	0	1	7
PA	0	15	10	9	9	17	64	5	8	9	0	0	6
PN	13	0	6	4	2	7	23	2	1	1	0	3	6
HI	8	6	0	3	2	0	12	1	0	2	0	0	0
MA	10	4	1	0	1	3	21	2	1	4	0	0	0
FL	16	4	5	3	0	25	28	1	2	0	0	0	0
PS	14	6	0	1	6	0	64	4	5	5	0	0	0
LP	66	58	41	26	11	26	0	14	12	20	0	4	15
FA	7	5	5	4	3	2	15	0	1	10	0	0	0
GL	2	1	0	0	0	5	14	3	0	3	0	0	0
JU	5	8	9	6	1	6	18	1	0	0	0	0	0
BA	0	0	0	0	0	0	3	0	0	0	0	0	0
CU	0	0	0	0	0	0	0	0	0	0	0	0	0
RU	1	2	0	0	1	4	2	1	0	0	0	2	0
EX	0	0	0	0	0	0	0	0	0	0	0	0	0
NP	15	10	9	4	6	7	23	2	3	6	0	1	5

Table A.20: The first part of the learned distances for objects in the spatial relation *right-of* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	-	0	0.5	0	0.4	0.9	0.3	0.5	0.3	-	0.4	0.2	0.5	0.2
MT	0	-	0.5	0.1	0.3	0.5	0.3	0.4	0.3	0.8	0.4	0.2	0.4	0.3
BO	0.5	0.3	-	0.3	0.5	0.4	0.2	0.3	0.2	-	0.4	0.2	0.2	0.3
MO	0.2	0.2	0.5	-	0.5	0.8	0.2	0.1	0.3	-	0.4	0.1	0.3	0.2
NB	0.4	0.4	0.5	0.2	-	0.5	0.3	0.2	0.2	0.3	0.3	0.5	0.5	0.3
PH	0.6	0.5	0.5	0.4	0.8	-	0.5	0	0.5	-	0.8	0.6	0.6	0.4
MU	0.4	0.2	0.3	0.1	0.5	-	-	0.1	0.6	-	0.4	0.3	0.2	0.1
BT	0.4	0.4	0.4	0.2	0.4	0	0.2	-	0.4	-	0.3	0	0.4	0.2
CP	0.3	0.5	0.1	0.1	0.3	0.4	0.1	0.4	-	-	0.2	0	0.1	0
KS	-	-	-	-	-	0	-	-	-	-	-	-	-	-
HP	0.2	0.1	0.3	0.1	0.2	-	0.2	0.2	0.1	-	-	0.2	0.3	0.1
PC	0.2	0.2	0.1	0	0.4	-	0.1	0	-	-	0.3	-	0.2	0.1
PA	0.5	0.4	0.2	0.3	0.5	0.2	0.3	0.2	0.3	-	0.3	0.1	-	0.2
PN	0.3	0.3	0.3	0.1	0.4	0.6	0.2	0.1	0.1	-	0.2	0.1	0.1	-
HI	0.3	0.5	0.3	0.1	0.5	0.2	0.2	0.2	0.2	-	0.2	-	0.2	0
MA	0.3	0.3	0.2	0.1	0.7	0.2	0.1	0.1	0.1	-	0.3	0.1	0.2	0
FL	0.5	0.6	0.2	0.5	0.9	0.3	0.4	0.2	-	-	0.1	0.1	0.2	0.1
PS	0.4	0.4	0.3	0.3	0.5	0	0.1	0.3	0.1	0	0.3	0.3	0.2	0.1
LP	0.3	0.3	0.5	0.2	0.4	0.9	0.5	0.4	0.4	0.6	0.9	0.8	0.5	0.4
FA	0.4	0.3	0.2	0.2	0.5	0	0.2	0.2	0.1	-	0.3	0.2	0.3	0.1
GL	0.3	0.3	0.4	0.2	0.3	-	0.2	0	-	-	0.3	-	0.3	0.3
JU	0.6	0.2	0.1	0.3	0.7	-	0.1	0.2	-	-	0.2	-	0.4	0.2
BA	0	-	-	-	0.5	-	-	-	-	-	0.3	-	-	-
CU	0.5	0.5	-	0.2	1	-	-	0.1	-	-	-	0	-	0.2
RU	0.3	0.2	0.4	0.1	0.5	-	0.1	0.5	-	-	0.2	0.1	0.1	0
EX	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NP	0.3	0.3	0.4	0.2	0.6	0.2	0.5	-	0.3	-	0.4	-	0.3	0.1

Table A.21: The second part of the learned distances for objects in the spatial relation *right-of* from the KTH data set (given in meters).

	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	EX	NP
KB	0.3	0.3	0.5	0.6	0.4	0.5	0.2	0.4	-	-	0.1	-	0.3
MT	0.2	0.3	0.7	0.6	0.4	0.3	0.4	0.4	0.2	-	0.3	-	0.4
BO	0.1	0.1	0.2	0.1	0.4	0.5	0.1	0.4	-	-	0.1	-	0.3
MO	0.1	0.1	0.5	0.1	0.6	0.3	0.1	0.4	-	-	0.1	-	0.4
NB	0.4	0.3	0.6	0.3	0.4	-	0.1	0.2	-	-	0.2	-	0.5
PH	0.5	0.4	0.3	0.2	-	0.1	-	-	-	-	0.5	0	0.7
MU	0.1	0.1	0.5	0.1	0.7	0.2	0.3	0.1	-	-	0.3	-	0.4
BT	0	0.3	0.9	0.2	0.7	0.2	0.2	0.1	-	0	0	-	0.4
CP	-	-	0.4	0.3	0.2	0.3	-	-	-	-	0.1	-	0.2
KS	-	-	0.3	-	-	-	-	-	-	-	-	-	-
HP	0.2	0.1	0.2	-	0.4	0.1	0.2	0	-	-	0.2	-	0.2
PC	0.1	0	0.9	-	0.7	0.1	-	0.2	-	-	0.1	-	0.3
PA	0.1	0.3	0.5	0.3	0.6	0.2	0.2	0.4	-	-	0.2	-	0.3
PN	0	0	0.3	0.2	0.5	0.2	0.2	0.2	-	-	0.1	-	0.2
HI	-	0	0.3	-	0.4	0.2	-	0.2	-	-	-	-	0.3
MA	0.1	-	0.3	0.2	0.4	0.3	-	0.3	-	-	-	-	0.2
FL	0.1	0	-	0.5	0.7	0.3	-	0.3	-	-	0.2	-	0
PS	-	0.3	0.4	-	0.7	0	0	0.2	-	-	0.2	-	0.3
LP	0.5	0.8	0.8	0.4	-	0.4	0.4	0.6	0.5	-	0.7	-	0.6
FA	0	0.1	0.3	0.2	0.2	-	0.3	0.3	-	-	0	-	0.2
GL	-	0.1	0.4	0.3	0.6	0.5	-	-	-	-	-	-	0.3
JU	0.2	0.2	-	0	0.7	0.3	0	-	-	-	-	-	0.4
BA	-	-	-	-	-	-	-	-	-	-	-	-	-
CU	-	-	-	-	0.8	-	-	-	-	-	0.4	-	0.5
RU	-	-	-	-	0.7	-	-	-	-	-	-	-	0.3
EX	-	-	-	0.3	-	-	-	-	-	-	-	-	-
NP	0	0.2	0.4	0.2	0.6	0.3	0.3	0.1	-	-	0.2	-	-

Table A.22: The first part of the target object occurrences and the number of valid *right-of* relation between object classes learned from the KTH data set.

	#	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC
KB	410	0	87	104	1	131	31	65	10	21	0	97	15
MT	452	111	0	116	3	139	28	58	13	18	1	97	12
BO	163	27	25	0	10	20	2	9	6	1	0	15	4
MO	409	315	252	25	0	105	1	28	2	6	0	20	10
NB	196	33	35	39	27	0	13	20	8	5	1	23	6
PH	86	26	41	2	29	17	0	4	4	3	0	2	4
MU	233	117	94	29	104	47	0	0	15	1	0	36	4
BT	95	57	51	4	41	12	5	8	0	1	0	4	1
CP	40	5	12	6	4	19	4	4	1	0	0	4	2
KS	1	0	0	0	0	0	1	0	0	0	0	0	0
HP	117	8	12	37	5	19	0	11	1	2	0	0	2
PC	32	14	11	3	3	8	0	2	4	0	0	11	0
PA	320	59	66	34	47	49	18	26	15	4	0	23	5
PN	119	48	31	15	33	42	1	14	7	5	0	21	5
HI	74	24	12	15	10	6	2	8	4	1	0	12	0
MA	68	24	14	12	15	5	2	3	2	1	0	13	2
FL	66	8	14	7	11	7	5	3	2	0	0	9	1
PS	133	71	93	13	49	37	19	21	8	5	1	19	5
LP	260	81	72	65	26	92	40	27	18	14	1	46	6
FA	35	10	6	3	4	5	1	2	5	2	0	5	1
GL	30	16	15	5	8	9	0	5	5	0	0	5	0
JU	50	16	20	6	17	4	0	15	6	0	0	11	0
BA	3	1	0	0	0	3	0	0	0	0	0	2	0
CU	6	6	6	0	5	1	0	0	3	0	0	0	1
RU	28	16	16	4	11	5	0	2	1	0	0	4	7
EX	2	0	0	0	0	0	0	0	0	0	0	0	0
NP	64	19	19	13	12	18	3	9	0	2	0	12	0

Table A.23: The second part of the target object occurrences and the number of valid *right-of* relation between object classes learned from the KTH data set.

	PA	PN	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU
KB	108	44	22	27	29	40	158	20	5	31	0	0	8
MT	109	44	25	27	45	38	168	22	8	20	1	0	7
BO	20	8	5	4	7	6	40	2	4	5	0	0	2
MO	33	14	12	7	7	23	175	3	2	12	0	0	4
NB	41	15	13	7	18	16	27	0	3	4	0	0	5
PH	6	7	6	2	13	25	0	2	0	0	0	0	3
MU	43	23	14	8	5	4	118	9	2	11	0	0	3
BT	12	6	2	2	1	20	38	6	2	2	0	1	3
CP	6	4	0	0	2	5	15	1	0	0	0	0	1
KS	0	0	0	0	1	0	0	0	0	0	0	0	0
HP	26	6	5	8	9	0	48	15	3	6	0	0	4
PC	4	5	5	2	1	0	9	1	0	4	0	0	4
PA	0	13	8	10	16	14	66	7	2	5	0	0	1
PN	15	0	6	4	4	6	58	5	1	8	0	0	2
HI	10	6	0	1	5	0	41	5	0	9	0	0	0
MA	9	4	3	0	3	1	26	4	0	6	0	0	0
FL	9	2	2	1	0	6	11	3	0	1	0	0	1
PS	17	7	0	3	25	0	26	2	5	6	0	0	4
LP	64	23	12	21	28	64	0	15	14	18	3	0	2
FA	5	2	1	2	1	4	14	0	3	1	0	0	1
GL	8	1	0	1	2	5	12	1	0	0	0	0	0
JU	9	1	2	4	0	5	20	10	3	0	0	0	0
BA	0	0	0	0	0	0	0	0	0	0	0	0	0
CU	0	3	0	0	0	0	4	0	0	0	0	0	2
RU	6	6	0	0	0	0	15	0	0	0	0	0	0
EX	0	0	0	0	0	2	0	0	0	0	0	0	0
NP	19	5	1	3	6	7	30	3	1	2	0	0	1

Table A.24: The first part of the learned distances for objects in the spatial relation *above* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	MU	CP	HP	PA	PN	MA	FL	PS	LP
KB	-	0	-	0	0	0	-	0	-	-	-	-	-	-
MT	0.1	-	-	0	-	0	0	0	-	-	0	-	-	0
BO	0	-	-	0	-	-	-	0	-	-	-	0	-	0
MO	0	-	-	-	0	0	-	0	0	-	-	-	-	-
NB	0.1	0	-	0.2	-	-	0.1	0	0	-	-	-	-	0
MU	0	-	0	0	-	-	-	0	-	-	0	-	0	-
BT	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CP	0	-	0	0	-	-	-	-	0	-	-	-	-	-
HP	0	-	0	-	-	-	-	-	0	-	-	0	-	-
PC	-	-	-	-	-	-	-	-	0	-	-	-	-	-
PA	0	-	-	-	-	0	-	0	-	0	0	0	-	-
PN	-	-	-	-	-	-	-	-	0	-	-	-	-	-
HI	-	-	-	-	-	-	-	0	-	-	-	-	-	-
MA	0	-	0	0	-	0	-	0	-	-	-	0	-	-
PS	-	-	-	-	-	0	-	0	-	-	-	-	-	-
LP	0.2	-	0.2	-	0.1	-	0.1	0.3	0.2	-	-	0.1	-	-
FA	0.1	0	-	-	-	0.1	-	0.1	0	-	0	-	-	-
JU	0	-	-	-	-	0	-	0	0	-	0	-	-	-
RU	-	-	-	-	-	-	-	0	0	-	-	-	-	-
NP	-	-	-	-	-	-	-	0	-	-	-	-	-	-

Table A.25: The second part of the learned distances for objects in the spatial relation *above* from the KTH data set (given in meters).

	FA	JU	NP
KB	-	-	-
MT	-	0	-
BO	-	-	0
MO	-	-	-
NB	-	-	-
MU	-	-	-
BT	0	-	-
CP	-	-	-
HP	-	-	0
PC	-	-	-
PA	-	-	-
PN	-	-	0
HI	-	-	0
MA	-	0	-
PS	-	-	-
LP	-	-	0.2
FA	-	0	-
JU	-	-	-
RU	-	-	-
NP	-	-	-

Table A.26: The first part of the learned distances for objects in the spatial relation *near* from the KTH data set (given in meters).

	KB	MT	BO	MO	NB	PH	MU	BT	CP	KS	HP	PC	PA	PN
KB	-	0.3	0.6	0.3	0.4	0.8	0.4	0.4	0.3	-	0.4	0.3	0.5	0.3
MT	0.3	-	0.6	0.4	0.4	0.6	0.3	0.4	0.5	1	0.5	0.4	0.5	0.4
BO	0.6	0.6	-	0.4	0.6	0.5	0.4	0.4	0.2	-	0.4	0.2	0.3	0.3
MO	0.3	0.4	0.4	-	0.5	0.4	0.2	0.2	0.2	-	0.3	0.2	0.3	0.2
NB	0.4	0.4	0.6	0.5	-	0.7	0.5	0.5	0.4	0.5	0.4	0.5	0.6	0.4
PH	0.8	0.6	0.5	0.4	0.7	-	0.5	0.1	0.6	0.5	0.8	0.6	0.4	0.5
MU	0.4	0.3	0.4	0.2	0.5	0.5	-	0.2	0.3	-	0.4	0.3	0.3	0.2
BT	0.4	0.4	0.4	0.2	0.5	0.1	0.2	-	0.4	-	0.4	0.1	0.3	0.2
CP	0.3	0.5	0.2	0.2	0.4	0.6	0.3	0.4	-	-	0.2	0.1	0.3	0.1
KS	-	1	-	-	0.5	0.5	-	-	-	-	-	-	-	-
HP	0.4	0.5	0.4	0.3	0.4	0.8	0.4	0.4	0.2	-	-	0.3	0.3	0.2
PC	0.3	0.4	0.2	0.2	0.5	0.6	0.3	0.1	0.1	-	0.3	-	0.2	0.2
PA	0.5	0.5	0.3	0.3	0.6	0.4	0.3	0.3	0.3	-	0.3	0.2	-	0.2
PN	0.3	0.4	0.3	0.2	0.4	0.5	0.2	0.2	0.1	-	0.2	0.2	0.2	-
HI	0.3	0.4	0.3	0.2	0.5	0.4	0.2	0.3	0.2	-	0.2	0.1	0.2	0.1
MA	0.3	0.4	0.2	0.2	0.5	0.4	0.2	0.2	0.1	-	0.2	0.1	0.2	0.1
FL	0.6	0.7	0.3	0.5	0.7	0.3	0.5	0.5	0.5	0.6	0.3	0.5	0.4	0.4
PS	0.5	0.5	0.4	0.3	0.5	0.3	0.1	0.3	0.3	0.3	0.4	0.4	0.3	0.2
LP	0.5	0.4	0.6	0.7	0.5	1	0.7	0.6	0.5	0.9	0.8	0.8	0.6	0.6
FA	0.5	0.4	0.4	0.3	0.7	0.2	0.3	0.3	0.2	-	0.2	0.2	0.3	0.2
GL	0.4	0.4	0.3	0.2	0.4	-	0.3	0.3	-	-	0.3	-	0.3	0.2
JU	0.5	0.4	0.4	0.4	0.6	-	0.2	0.4	-	-	0.2	0.4	0.4	0.3
BA	0.3	0.2	-	-	0.5	-	-	-	-	-	0.3	-	-	-
CU	0.6	0.5	-	0.3	1	-	-	0.2	-	-	-	0.2	-	0.3
RU	0.3	0.4	0.4	0.1	0.4	0.5	0.3	0.2	0.1	-	0.2	0.1	0.2	0.1
EX	-	-	-	-	-	0.1	-	-	-	-	-	-	-	-
NP	0.4	0.5	0.4	0.3	0.6	0.4	0.5	0.4	0.3	-	0.4	0.4	0.4	0.2

Table A.27: The second part of the learned distances for objects in the spatial relation *near* from the KTH data set (given in meters).

	HI	MA	FL	PS	LP	FA	GL	JU	BA	CU	RU	EX	NP
KB	0.3	0.3	0.6	0.5	0.5	0.5	0.4	0.5	0.3	0.6	0.3	-	0.4
MT	0.4	0.4	0.7	0.5	0.4	0.4	0.4	0.4	0.2	0.5	0.4	-	0.5
BO	0.3	0.2	0.3	0.4	0.6	0.4	0.3	0.4	-	-	0.4	-	0.4
MO	0.2	0.2	0.5	0.3	0.7	0.3	0.2	0.4	-	0.3	0.1	-	0.3
NB	0.5	0.5	0.7	0.5	0.5	0.7	0.4	0.6	0.5	1	0.4	-	0.6
PH	0.4	0.4	0.3	0.3	1	0.2	-	-	-	-	0.5	0.1	0.4
MU	0.2	0.2	0.5	0.1	0.7	0.3	0.3	0.2	-	-	0.3	-	0.5
BT	0.3	0.2	0.5	0.3	0.6	0.3	0.3	0.4	-	0.2	0.2	-	0.4
CP	0.2	0.1	0.5	0.3	0.5	0.2	-	-	-	-	0.1	-	0.3
KS	-	-	0.6	0.3	0.9	-	-	-	-	-	-	-	-
HP	0.2	0.2	0.3	0.4	0.8	0.2	0.3	0.2	0.3	-	0.2	-	0.4
PC	0.1	0.1	0.5	0.4	0.8	0.2	-	0.4	-	0.2	0.1	-	0.4
PA	0.2	0.2	0.4	0.3	0.6	0.3	0.3	0.4	-	-	0.2	-	0.4
PN	0.1	0.1	0.4	0.2	0.6	0.2	0.2	0.3	-	0.3	0.1	-	0.2
HI	-	0.1	0.3	-	0.6	0.3	-	0.4	-	-	0.1	-	0.3
MA	0.1	-	0.3	0.2	0.7	0.3	0.1	0.4	-	-	-	-	0.3
FL	0.3	0.3	-	0.5	0.8	0.4	0.4	0.3	-	-	0.2	-	0.4
PS	-	0.2	0.5	-	0.6	0.2	0.2	0.2	-	-	0.3	0.4	0.3
LP	0.6	0.7	0.8	0.6	-	0.5	0.6	0.8	0.7	0.8	0.7	-	0.7
FA	0.3	0.3	0.4	0.2	0.5	-	0.4	0.4	-	-	0.1	-	0.3
GL	-	0.1	0.4	0.2	0.6	0.4	-	0.1	-	-	-	-	0.4
JU	0.4	0.4	0.3	0.2	0.8	0.4	0.1	-	-	-	-	-	0.4
BA	-	-	-	-	0.7	-	-	-	-	-	-	-	-
CU	-	-	-	-	0.8	-	-	-	-	-	0.5	-	0.6
RU	0.1	-	0.2	0.3	0.7	0.1	-	-	-	0.5	-	-	0.4
EX	-	-	-	0.4	-	-	-	-	-	-	-	-	-
NP	0.3	0.3	0.4	0.3	0.7	0.3	0.4	0.4	-	0.6	0.4	-	-

A.2 Influence of the resolution on position prediction

A.2.1 Grid resolution of 0.01 meters

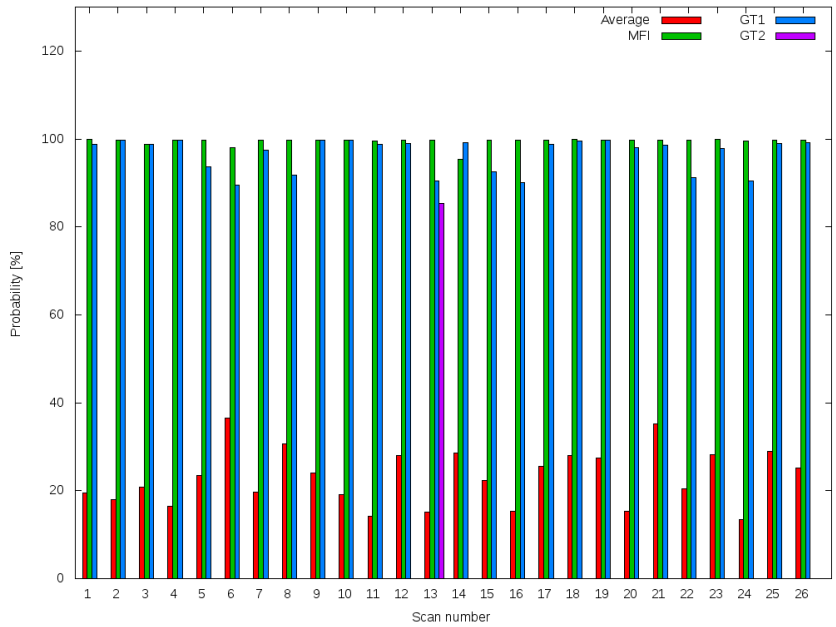


Figure A.1: Probability to find the object *mouse* at different positions (grid resolutions of 0.01 meters).

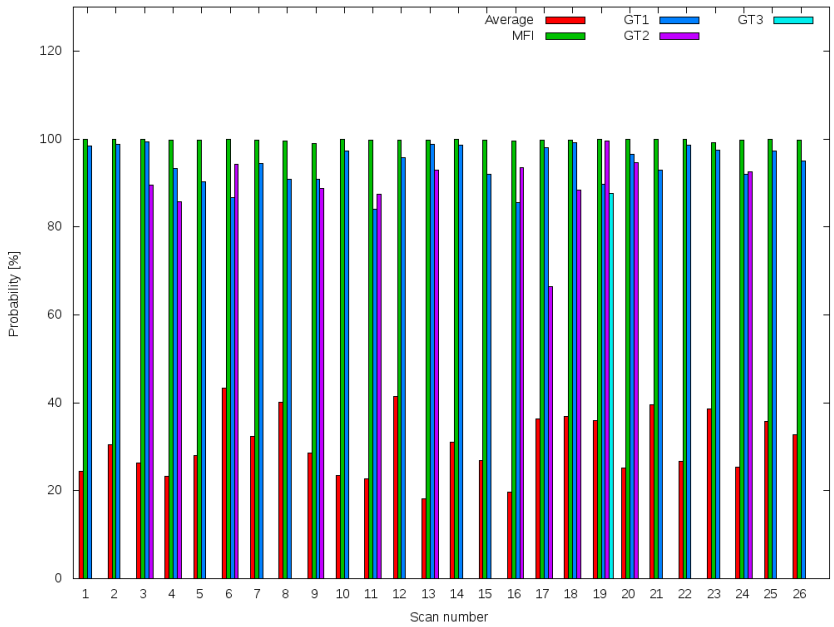


Figure A.2: Probability to find the object *monitor* at different positions (grid resolutions of 0.01 meters).

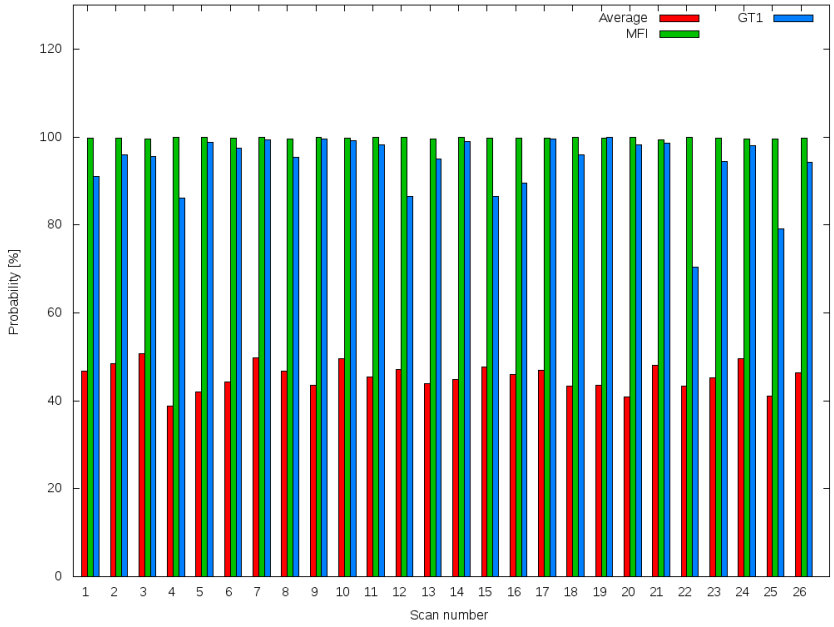


Figure A.3: Probability to find the object *table* at different positions (grid resolutions of 0.01 meters).

A.2.2 Grid resolution of 0.1 meters

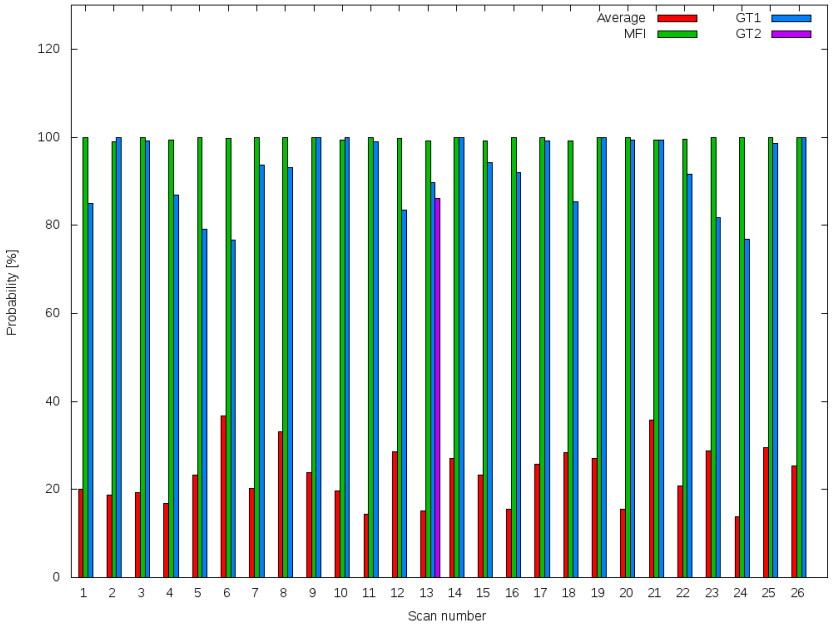


Figure A.4: Probability to find the object *mouse* at different positions (grid resolutions of 0.1 meters).

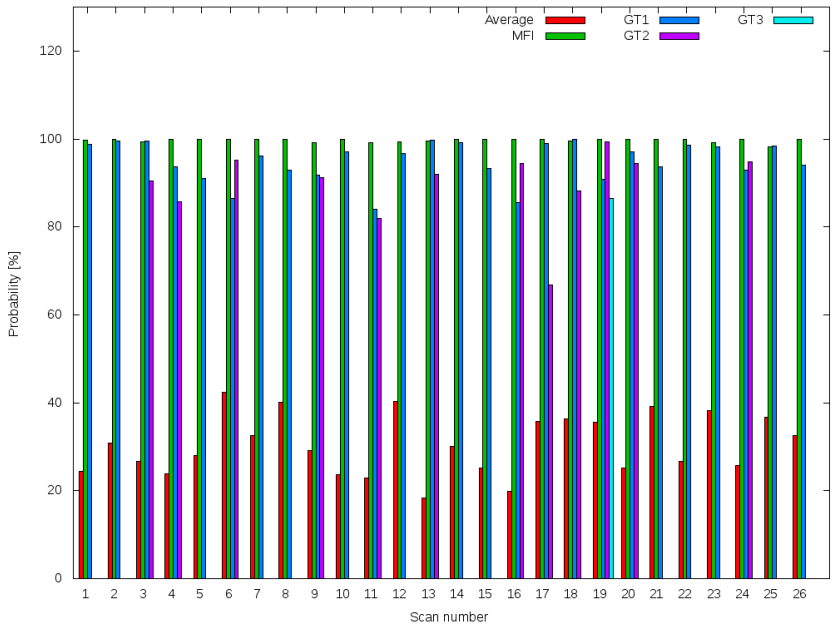


Figure A.5: Probability to find the object *monitor* at different positions (grid resolutions of 0.1 meters).

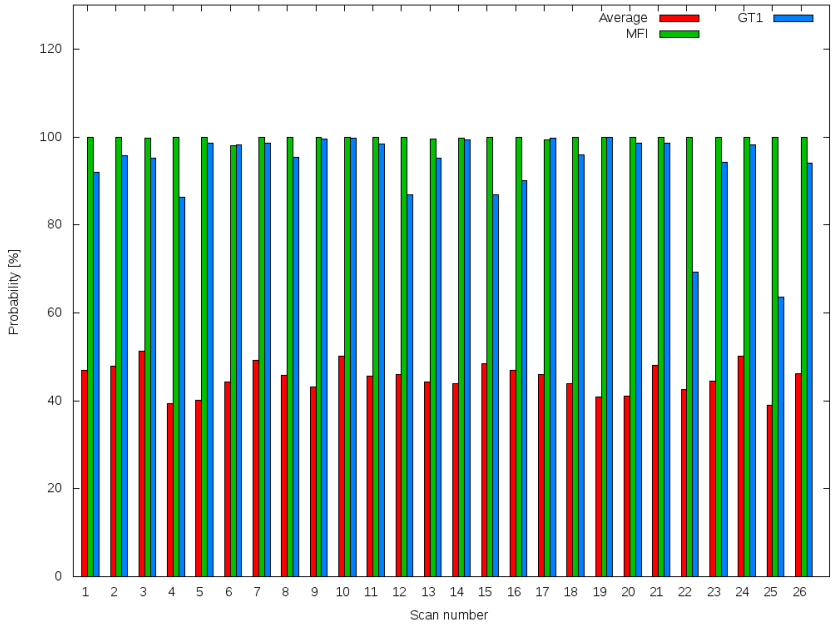


Figure A.6: Probability to find the object *table* at different positions (grid resolutions of 0.01 meters).

A.2.3 Grid resolution of 0.5 meters

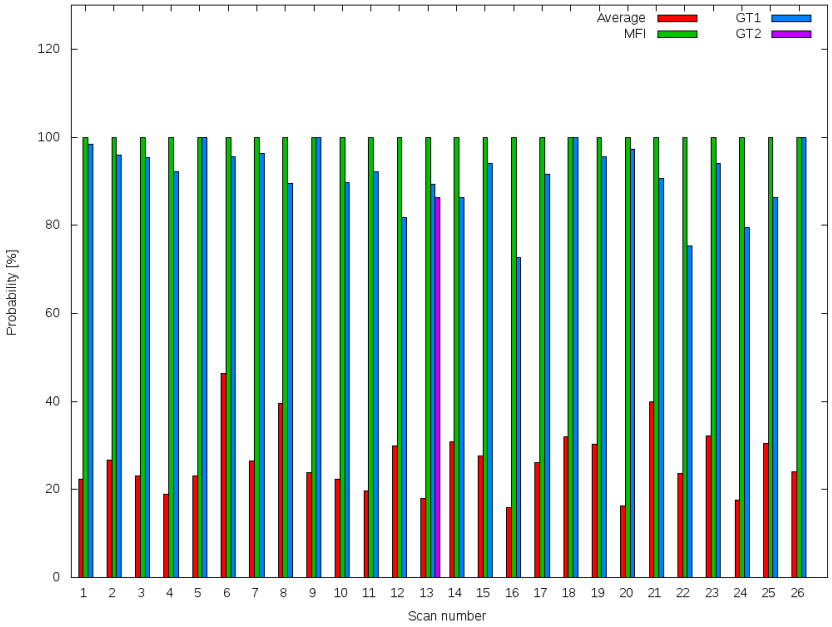


Figure A.7: Probability to find the object *mouse* at different positions (grid resolutions of 0.5 meters).

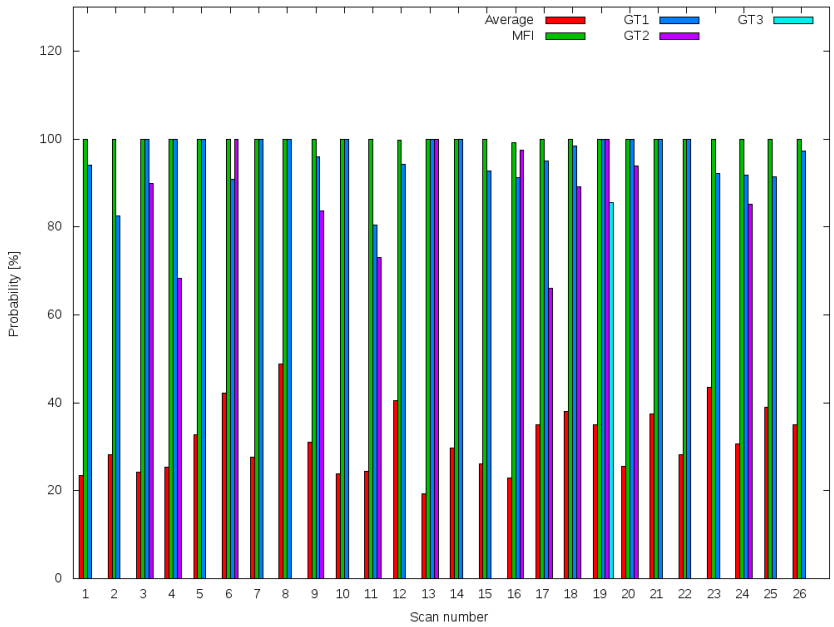


Figure A.8: Probability to find the object *monitor* at different positions (grid resolutions of 0.5 meters).

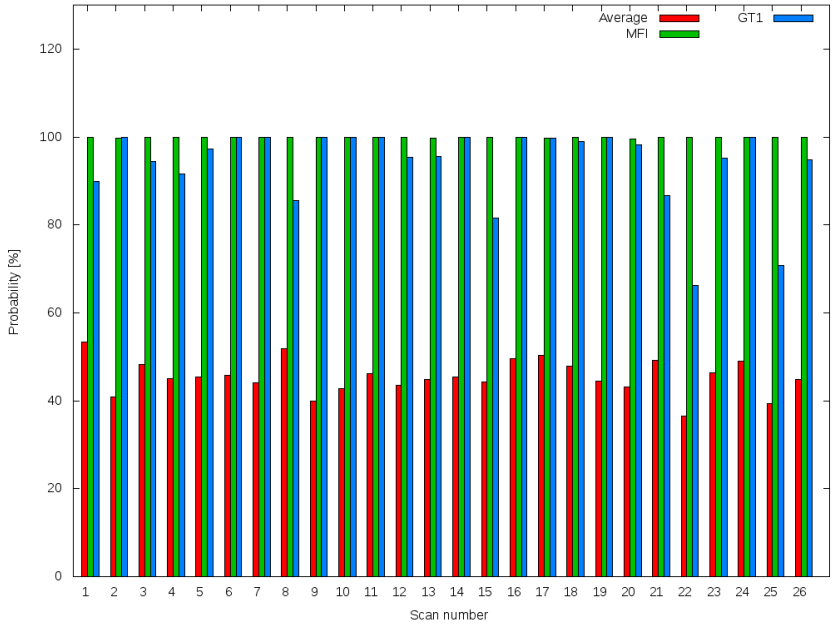


Figure A.9: Probability to find the object *table* at different positions (grid resolutions of 0.5 meters).

Acronyms

2D	two-dimensional.
3D	three-dimensional.
AI	Artificial Intelligence.
CoG	Center of Gravity.
CRF	Conditional Random Fields.
DFKI-RIC	German Research Center for Artificial Intelligence - Robotics Innovation Center.
DIA	Dia Diagram Editor.
FI	Field Intensity.
GMM	Gaussian Mixture Models.
HRI	Human Robot Interaction.
KTH	KTH Royal Institute of Technology.
MFI	Maximum Field Intensity.
MRF	Markov Random Fields.
NWU	North-West-Up.
PQSR	Probabilistic Qualitative Spatial Relations.
QSR	Qualitative Spatial Reasoning.
RCC	Region Connection Calculus.
SPF	Spatial Potential Fields.
VFH	Vector Filed Histogram.

List of Figures

List of Tables