University of Birmingham
School of Computer Science

Thesis proposal

Marco Antonio Becerra Pedraza

Semantic Mapping of Human Activities with a Mobile Robot

Intelligent Robotics Lab

Supervisor: **Dr. Nick Hawes**

Thesis group members: **Prof. Jeremy Wyatt, Dr. Peter Hancox**

Birmingham, 26.3.2015

Contents

1 Introduction **1**
- 1.1 Introduction . . . 1
  - 1.1.1 Human activity analysis with a mobile robot . . . 1
  - 1.1.2 Answer Set Programming for Knowledge Representation and Reasoning . . . 2
- 1.2 Research Problem . . . 3
  - 1.2.1 Expected outcome . . . 3
- 1.3 Test scenario: "The library setting" . . . 3
- 1.4 Document organization . . . 4

2 Related Work **6**
- 2.1 General antecedents - Perception in AI . . . 6
- 2.2 Activity Recognition . . . 8
  - 2.2.1 Single-layered approaches . . . 9
  - 2.2.2 Hierarchical approaches . . . 11
- 2.3 Description-based activity recognition and robotics . . . 13
  - 2.3.1 Description-based activity recognition . . . 13
  - 2.3.2 Activity recognition and robotics . . . 15
- 2.4 Answer Set Programming . . . 18
  - 2.4.1 ASP as a knowledge representation language . . . 19
  - 2.4.2 ASP in activity recognition . . . 20

3 Research Problem **21**
- 3.1 Problem Description . . . 21
- 3.2 Methodology . . . 22
  - 3.2.1 Sensing and feature extraction . . . 23
  - 3.2.2 Qualitative Spatio-Temporal Representations . . . 24
  - 3.2.3 Knowledge Base . . . 25
  - 3.2.4 Inferring activities . . . 25
  - 3.2.5 Discriminate activities . . . 26
  - 3.2.6 Robot Action . . . 27

4 Experimental Approach **28**
4.1 Library Example 28
4.1.1 Experimental stages 30

5 Work Plan **33**
5.1 Work Distribution 34
5.2 Time table 39

6 Conclusions **41**

Bibliography **42**

Chapter 1

Introduction

1.1 Introduction

One of the main goals in AI is having robots working autonomously in everyday environments. In such situation, robots are expected to perceive, understand and interact with its environment. However, these kind of environments are dynamic, non-structured and non-deterministic, which makes difficult for a robot to fulfil the assigned tasks. To be able to sort these obstacles, robots need to be provided with cognitive skills.

Human cognition refers to all mental activities associated with thinking, knowing, remembering and communicating, and how the information is processed (King, 2014; Myers, 2013). In robotics, the concept is associated with systems that emulate these mental processes or those that *sense*, *plan* and *act*. More precisely, cognition can refer to those systems that can perceive, understand ... and interact with their environment, and evolve in order to achieve human-like performance in activities requiring context (situation and task) specific knowledge (Christensen et al., 2010).

Everyday environments have many valuable features that a robot needs to understand, in order to succeed while performing a task, among them are human activities. Human activities are a meaningful manifestation of human behaviour along time and space. They provide rich information about the human performing it, but also, about his/her relation with other relevant components of the environment as humans, objects and locations.

1.1.1 Human activity analysis with a mobile robot

Activity recognition is the research field that studies the automatic detection and analysis of human activities by processing the data acquired from sensors (Aggarwal and Xia, 2014). It is not restricted only to sensory data and this

can also be complemented with other sources of information, i.e. domain knowledge. In the AI context, activity recognition is closely related with the areas of perception, knowledge representation and reasoning. The problem of activity recognition has been treated from different perspectives, however, computer vision has been the most popular.

In principle, robots with appropriate sensing and processing capabilities can perform activity recognition. Moreover, they have some advantages over the use of fixed cameras or wearable devices as they are able to interact with the environment. Robots are active observers, i.e. they can change their point of view on scene and be selective in the areas of the environment that are more interesting. On the other hand, they have some disadvantages as well. They don't have omnipresence, so they are not able to sense the full environment and will loose information. Also, their sensors have constraints, the data may be noisy or blurry due to movement, erratic hardware, changing environmental conditions, etc. Finally, robots are expected to work in real-time, so online activity recognition is desirable but yet difficult to achieve.

Activities involve knowledge. They associate concepts and relations between a subject and the environment. In general, an activity recognition system should be able to handle, not only sensory data but also symbolic representations, and be able link top level symbolic concepts to low level sensory data; this is known as the *anchoring problem* (Coradeschi and Saffiotti, 2003). With this in mind, knowledge processing and reasoning is a necessity for such systems.

1.1.2 Answer Set Programming for Knowledge Representation and Reasoning

There have been proposed many ideas to handle the problems of knowledge representation and reasoning (e.g. logic programming, ontologies, bayesian networks, fuzzy logic, etc.), among them is Answer Set Programming (ASP). ASP is form of declarative programming oriented towards difficult, primary NP-hard, search problems. It establish a new paradigm of logic programming that allows concepts as negation as a failure, default knowledge and non-monotonic reasoning.

ASP main concepts were proposed since the late 1980s (Gelfond and Lifschitz, 1988) and it has been used with success in many applications. Traditionally, ASP solvers were designed as one-shot problem solvers, so they lacked of reactive capabilities and, for example, whenever new data arrived, the system needed to be restarted. This has been one of the main reasons why ASP has not been fully exploited in the field of robotics, however, in

recent years an important effort has been directed towards this direction by some groups (Andres et al., 2013; Erdem et al., 2013).

1.2 Research Problem

This project is based in the consideration of ASP as a solid option for robots in problems that require symbolic representations. The focus of this project is to study **ASP-based activity recognition with a mobile robot**. The interest lies in the spatio-temporal relation between human activities and the environment and how a robot can acquire, handle and use this knowledge.

ASP allows the manipulation of incomplete information and handling multiple sources of knowledge. The integration between observations and external knowledge appear to be a more robust approach than a single sensory based approach. Hardware adds uncertainty via noise, is constrained by its specification and the outcome data is usually difficult to process. This uncertainty cannot be eradicated in robots, but the treatment of some problems as activity recognition can be boosted by emphasizing a more cognitive approach.

1.2.1 Expected outcome

The expected outcome will be a systematic analysis of ASP-based activity recognition and its use with a mobile robotic platform.

First, the problem of activity recognition will be treated via ASP and compared with other state-of-the-art approaches.

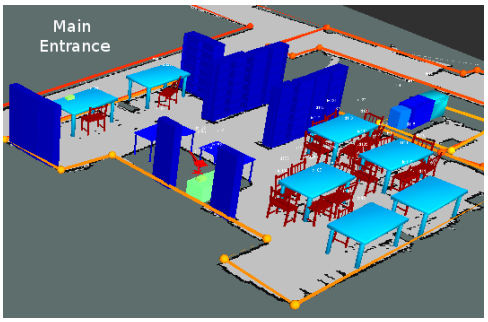
Experience acquisition will be studied via semantic mapping to progressively create a knowledge source for a robot regarding the activities occurring in a particular location and that can be used by the robot for future inferences.

Finally, we are interested in taking advantage of the activeness of the robot to improve the perception and knowledge acquisition processes in the context of activity recognition; e.g. looking for missing information, the ability to get information from the environment more efficiently, etc.

1.3 Test scenario: "The library setting"

Because human activities cover a broad range o possible situations, it becomes a necessity to bound the scope of this project and look into exemplar cases that can be used to demonstrate the ideas. With this in mind, it has

(a) Real library



(b) Semantic map of the library

Figure 1.1: The library setting

been chosen a library as a scenario to test and explain the ideas of this project.

The School of Computer Science at the University of Birmingham has a library, Fig. **??**. The physical scenario is basically a big room. It has some cabinets (where bibliographic material is stored), a reception desk, some tables with chairs and a printing area. It has only one entrance. An attendant is in charge of the book loans and retrievals, and also to assist the users. Most users use the facility to study, to consult material, to print, for team work, to use their laptops or simply as a rest area.

Because of the rules of the library and the nature of the location, the amount of activities is limited. However, non-considered activities could eventually appear, as giving a greet, using a cellphone, talking, etc. The amount of objects involved in the environment is also relatively small (books, tables, chairs, laptops, cabinets, etc.).

Our interest is to use this library setting as a test for activity recognition with a mobile robot, and moreover The problem can be analysed with simple examples via simulation to build the core of a testing system, and later use real data and eventually a mobile robot.

1.4 Document organization

The rest of the document is organized in the following manner.

Chapter 2 presents related literature to the problem and discusses the results and limitations of previous approaches. The chapter starts with a brief historic presentation of perception in Artificial Intelligence and goes towards the problem of activity recognition. Section**??** presents a summary of how the problem of activity recognition has been treated before, and particu-

larly regarding hierarchical description-based approaches, in which ASP is an alternative. In section 2.3, some previous works that integrate activity recognition and robotics are reviewed. Finally, in section 2.3, ASP is presented as an alternative for solving problems that require knowledge representation and reasoning.

Chapter 3 presents the problem (section 3.1) and the proposed methodology (section 3.2) to treat it.

Chapter 4 presents an experimental approach to evaluate the methodology proposed within this project. By using variations of the library setting environment (section 1.3), the problem can be studied gradually and the main components of an ASP-based approach can be remarked.

Chapter 5 presents a work plan for the rest of the project duration (2 years) and proposes goals and tasks to achieve them.

Finally, chapter 6 presents the conclusions final comments regarding this project and about the developed ideas.

Chapter 2

Related Work

Activity recognition is a cognitive skill that can be considered within perception. It deals with the interpretation and understanding of a scene by processing a sensory input and correlate it with domain knowledge to be able to associate it with an activity pattern.

The fundamental concepts to understand the problem of activity recognition lies in Psychology, as this is the science that studies the mind. Then, the problem that follows is to replicate this cognitive process into a machine, particularly a robot; this is the focus of study of Artificial Intelligence and Robotics. The aim of this project is to study the use of ASP in robotics for knowledge processing and reasoning; looking, in particular, to the problem of activity recognition with a mobile robot.

This chapter presents relevant related work. First the antecedents of perception in AI are traced to serve as a basis for the rest of the chapter. Then the problem of activity recognition is briefly reviewed to provide a general overview of the area and also, to locate this project in the *map* of research in the area. Finally, closely related work is presented to point the current state of the area and exhibit open trends in the area where the current project has possibilities to provide continuity.

2.1 General antecedents - Perception in AI

Perception, as a cognitive process, has been studied widely in Psychology. It refers to the process of organizing and interpreting sensory information so that it has a meaning (King, 2014). Part of the interest lies in how sensory information is processed by the brain, and which parts of it are essential. Also relevant, it is the domain knowledge, previously acquired or consulted, that the subject uses to establish a meaning to the sensory input. Together,

the sensory input and the domain knowledge are used to interpret a scene.

Sensory input is important for perception, however, not all the data is equally important to interpret a particular scene and conclusions can still be made, even with partial data. In (Heider and Simmel, 1944), an animated film was created using only moving polygons to demonstrate how the motion of abstract entities could be interpreted by human observers in meaningful ways. In (Johansson, 1973), locomotion patterns of living organisms using visual marks were studied. By this mean, the emphasis was put in the qualitative motion description of the marks rather than in the qualitative motion description of the moving body. Both works show that, even with abstract entities, the human mind is capable to create or associate concepts to them using only a limited amount of information. Also, that perception is a more complex mental process than just processing sensory data.

In Artificial Intelligence, perception has been studied principally by the computer vision community. Earlier works can be traced back to the 1960s, as part of the effort to mimic human-like intelligence using visual perception components. The main difference between computer vision and image processing has been the desire to recover the three-dimensional structure of the world from images, and to use this as a stepping stone towards full scene understanding (Winston and Horn, 1975).

One of the first works in 3D reconstruction from a single image is found in (Roberts, 1963). The developed system was able to reconstruct geometrical bodies with flat surfaces by recognizing the borders of the bodies in the scene and later analysing the shades of their visible surfaces. In (Barrow and Popplestone, 1971) object recognition was studied by decomposing an image into regions and describing the spatial relations between them, in a more qualitative, rather than the traditional quantitative pixel-based approach.

Since the early 1970s, the *block's world* was used as a test scenario for intelligent systems, particularly regarding knowledge representation, reasoning and planning. In the block's world, an initial state A and a desired state B of the environment are given. The goal is to autonomously generate a plan to transform A into B by the manipulation of the blocks. One important characteristic of the problem is that requires a symbolic description of the scene. The problem was used as a test case for the robot Shakey (Nilsson, 1984).

During the late 1980s, the concept of *active perception* emerged to emphasize the role of control during the sensing phase (Bajcsy, 1988). This is, the capacity to adapt the sensing strategy by considering the data interpretation and the goal task. It is clear that a robot is an active agent, so it is desirable to take advantage of active perception strategies. In the context of activity recognition with a mobile robot, an active perception strategy would

be to close the control loop by directing the robot actions with the former results of the activity recognition system, to improve the data collection and, eventually, improve current and future conclusions.

2.2 Activity Recognition

Activity recognition is an important research area in the context of automated perception. It has many applications as surveillance, inspection, verification, generation of automated reports, etc. The application will dictate the approach to follow and the kind on sensors that will be required.

Human activities can be classified in different ways. In (Turaga et al., 2008), two non-exclusive categories are used: actions (performed by a person) and activities (performed by many persons). One more descriptive categorization has been given in (Aggarwal and Ryoo, 2011), separating activities in four classes:

Gestures Elementary movements of a person's body part, and are the atomic components describing a meaningful motion of a person. E.g. 'stretching an arm', 'raising a leg'.

Actions Single person activities that may be composed of multiple gestures organized temporally. E.g. 'walking', 'waving'.

Interactions Human activities that involve two or more persons and/or objects. E.g. 'Two persons fighting', 'a person eating an apple'.

Group activities The activities performed by conceptual groups of multiple persons and/or objects. E.g. 'a football team playing a match', 'a group of students making an exam'.

All of these categories require to be able to sense humans with different level of detail. For example, gestures require specific algorithms (e.g. hand detection, face detection, skeleton tracing, etc.), while long distant pedestrian tracking algorithms may consider persons as *moving dots*. The type of activities to be recognized will dictate the required algorithms.

In the same fashion, activity recognition approaches can be classified in basis of different factors.

First, regarding the sensors, two approaches can be followed, distant or pervasive. The first one observes the scene from the distance as it happens with a CCTV camera or a robot. The pervasive approach relies on wearable devices to detect the activity of a person from a first person point of view.

Another possible classification of activity recognition systems focuses on how information is processed. In (Aggarwal and Ryoo, 2011) a taxonomy is proposed as shown in Fig. 2.1.

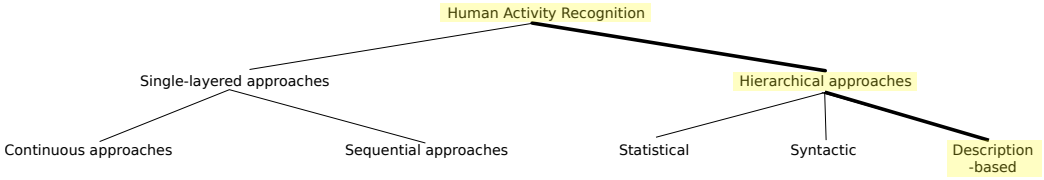


Figure 2.1: The taxonomy of research in activity recognition described in Aggarwal and Ryoo (2011). It has been highlighted the branch that follows this project.

2.2.1 Single-layered approaches

They represent activities in terms of raw sensory data[1], because of this, the activity descriptions are trained from datasets.

Single-layered approaches are suitable to recognize short-term and simple activities as gestures, movements of the body or simple interactions with objects. This is mainly because the amount of sensory data grows very easily and long-term activities would require to process larger amounts of data. Also, because activities are not always performed in the same way, even by the same individual. The shorter the activity, the more accuracy that will be attained. Finally, this approach is very sensitive to the input sensory data as this will depend on the environmental conditions (e.g. lighting, point of view).

Continuous approaches

In continuous approaches[2] the activities are recognized by analysing continuous sensory data and compare it with an activity pattern.

An activity is represented as a block of data along time where the activity was performed, and it is considered as a whole. A volume (or hyper-volume) is built by concatenating the sensor readings in time. The dimension of the data will depend on the sensing capabilities of the system; for example,

[1]The original survey (Aggarwal and Ryoo, 2011) describes single layered approaches as image-based approaches, but it leaves out the systems with other sensing capabilities (e.g. 3D sensors, sonar, GPS, etc.). However, they can be included too the activities are represented in terms of raw sensory data patterns.

[2]'Space-time approaches' in (Aggarwal and Ryoo, 2011)

a video stream would require 3 dimensions (X, Y, T) and a RGBD camera would be able to use 4 dimensions (X, Y, Z, T), etc. The sensory input is compared with the activity patterns to measure similarity. If a threshold is fulfilled, then the activity is labelled.

The advantages of this approach is that it is relatively fast and doesn't require domain knowledge. On the other hand, these methods have a big dependence of the training data and they are sensitive of the environmental conditions, different sampling rates, discontinuity of the data, the manner in which the activity is performed and the observation point of view.

There are many examples of this approach. In (Bobick and Davis, 2001) a video stream of aerobics exercises is analysed by attaching to every pixel a vector indicating the presence and recency of motion. Then, the stream was compared online with previously described activities to look for matching. In (Ke et al., 2007), volumes were built by attaching similar regions of adjacent frames. Then, the problem was transformed in an object matching problem by comparing the shapes of the volumes (sensory stream and activity patterns).

Sequential approaches

Sequential approaches represent activities as a sequence of states. A state is a vector of features observed in the scene in a specific instant. Finally, the sequence is analysed depending on the activity representation. There are two types of methodologies adopted in sequential approaches: exemplar-based and model-based.

In exemplar-based methods, activities, or a class of them, are represented as a sequence of states. Then the sensory input is compared in similarity with the patterns. An example can be found in (Darrell and Pentland, 1993), where states are built from view models. Templates of activities are from sequences of states associated with a physical change (e.g. rotation and scale). The dynamics of articulated objects in scene were recognized using the dynamic time warping algorithm (DTW) to the sequence of states.

In model based approaches, the sequence of states is compared with a set of probabilistic models of activities. The models are built assuming a temporal dependence between the states, so the transitions are modelled statistically using hidden Markov models (HMM), dynamic bayesian networks (DBN), conditional random fields (CRF), etc.

The first work to use HMM to recognize activities was (Yamato et al., 1992). They transformed a video stream into a sequence of vectors of image features. Then every vector was transformed to a symbol using vector quantization. Finally, a set of HMMs were created to model the activities, and

their parameters were optimized.

2.2.2 Hierarchical approaches

Hierarchical approaches of activity recognition refer to those where complex activities are represented in terms of simpler ones. This kind of representation organizes the activities in multiple layers of complexity and creates, by this mean, a hierarchical structure. In the lower level, atomic (or primitive) actions are represented which are indivisibles; they are usually recognized using single-layered approaches.

Hierarchical approaches are also adequate to represent activities symbolically, by taking advantage of the multi-layered organization, it is possible to describe describe semantic relations. By these means, they are less dependant to training data and they can integrate domain knowledge more easily.

Hierarchical approaches can be categorized, by the used methodology for recognition, in: statistical, syntactical and description-based.

Statistical

They are based in the hierarchical construction of statistical state-based models, such as HMMs, DBNs or CRFs.

First, the activities are defined and organized hierarchically. In the bottom level, atomic actions defined in terms of feature vectors which are obtained from sensory data. The sequence of feature vectors is analysed statistically to be able to recognize, this is performed in the same fashion as in single-layered sequential approaches. Then, the sequence of atomic actions is considered to be the input observations to recognize the next layer of activities, and the same statistical methods can be used to recognize the second layer of activities. The procedure is repeated in every layer.

Some disadvantages of the statistical approaches are their difficulty to model the temporal structure of events (e.g. A occurred 'during'/'before'/'after' B) and also, because of their sequential nature, it becomes harder to handle multiple concurrent tasks.

In (Oliver et al., 2002), the authors present layered hidden Markov models (LHMMs) for online activity recognition using data from video, sound and keyboard data. They divide their system in three layers: the first one is in charge of recognizing features from every source, the second layer trigger short events from the scene, and the last layer is in charge to recognize longer activities. The hierarchical approach adopted showed an improved performance when compared with single-layered systems. The training data

is used more efficiently and also, it becomes easier to add more detail on specific activities, i.e. to handle different granularities in the activity descriptions.

In (Liao et al., 2007a,b), the authors present studies of hierarchical DBNs and CRFs applied to location-based activity recognition. The authors use GPS data from mobile devices and domain knowledge from locations to learn and infer routines and activities. As the sensory input are the spatio-temporal traces of the person, domain knowledge from the locations is considered to add significance to the events. For example, a quick car stop can be part of an important activity as leaving the children at school or simply caused by a traffic light. First they segment spatially the traces by grouping neighbour samples that occur in the same region. They distinguish two major categories of activities: *navigation* (e.g. walking, driving a car) and *significant* (e.g. work, visit, sleep). Their model is built, in the lower layer, to classify the segmented tracks in activities (events), by looking at their location and the speed of the subject; and finally, in an upper layer, they use these activities to learn distinctive places (e.g. house, office, etc.).

Syntactic

In the syntactical approach, activities are represented symbolically as a set of production rules generating a string of atomic actions which is later recognized using parsing techniques. Atomic actions are obtained with a single-layered approach, however, in higher layers, recognition is performed symbolically. Context-free grammars (CFGs) and stochastic context-free grammars (SCFGs) are some of the techniques that have been used to recognize high level activities.

One limitation of this approach is the difficulty to handle concurrent activities, and also to consider unexpected events that are not integrated in the grammar.

An example can be found in (Ivanov and Bobick, 2000). The authors aim to recognize complex activities in sequences of video. Two layers are defined; in the lower level, atomic actions are recognized using HMMs, and in the upper one uses SCFGs. The adopted approach showed to be able to handle longer time activity constraints and also, to be more robust regarding uncertain detections in the lower level.

Description-based

This approach represent activities as a hierarchy of events, making emphasis in their spatial, temporal and logical structures.

A complex activity is modelled from de occurrence of its sub-events that satisfies certain relations. The temporal relation between sub-events is also considered in the representation, Allen's interval algebra is frequently used for this (Allen, 1983). Atomic actions are obtained from sensory data and summarized.

Now to recognize activities the problem becomes a *constraint satisfaction problem*, which is NP-hard. This approach allows a good integration of additional knowledge sources. Particularly, the

There are many possibilities to treat the problem. In (Nevatia et al., 2004; Ryoo and Aggarwal, 2006), CFGs are used to represent activities hierarchically, defining temporal relations between sub-events. In (Sridhar et al., 2010), relevant features from the scene are extracted and their behaviour is represented using qualitative spatio-temporal relations (QSTR), then patterns of activities are learnt using Markov chains.

2.3 Description-based activity recognition and robotics

In this project, the chosen approach to follow is a description-based methodology. This section presents relevant related work in this line, and sin the Answer Set Programming (ASP) paradigm for Also, here are presented the precedents of activity recognition with mobile robots.

2.3.1 Description-based activity recognition

As mentioned in section 2.2.2, description-based approaches represent activities hierarchically by decomposing complex activities in sub-events. The representation should also make emphasis in the spatial, temporal and logical structures. The recognition is performed by obtaining features from scene (spatial, temporal, logical) and creating a scene description as a *list* of facts, then the problem becomes a constraint satisfaction problem, to find the best activity match for these set of observations.

Representation of activities

An activity is represented as a set of *facts* that needs to be fulfilled with the observations. This is important, because the facts can be used as logical predicates. These facts act as constraints between the activity patterns and help to discriminate between them, some of them may be more relevant than the others.

The execution of an activity depends on the subject, and even a particular subject doesn't execute the same activity in the same way. This is the reason why activities are usually defined in qualitative terms, i.e. in a symbolic more human-like manner. Quantitative descriptions of activities are still interesting, however, they are restricted to specific domains as rehabilitation or sports.

Regarding space, Qualitative Spatial Representations (QSR) are a set of calculus which allow a machine to represent and reason about spatial geometrical entities (Cohn and Renz, 2007), e.g. lines, dots, regions, etc. They are usually combined with a temporal representation to represent the dynamics of behaviours.

Time can be represented as an instant t or as an interval (t_1, t_2). For the instants, simple temporal logic can be used to represent these kind of statements. Intervals have been typically treated with two approaches (Fisher, 2008): Interval temporal logics (Moszkowski, 1983) and Allen's interval algebra (Allen, 1983).

Allen's interval algebra is a calculus for temporal reasoning. It defines 13 possible relations between intervals, and provides a composition table that can be used for reasoning about temporal descriptions of events 2.2.

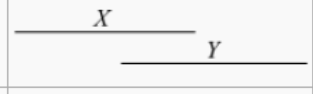
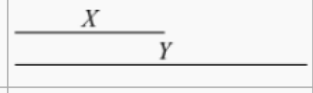
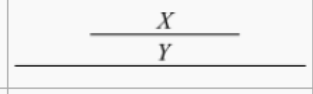
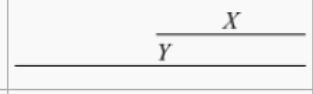
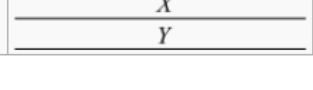
Relation	Illustration	Interpretation
$X < Y$ $Y > X$		X takes place before Y
X **m** Y Y **mi** X		X meets Y (*i* stands for ***inverse***)
X **o** Y Y **oi** X		X overlaps with Y
X **s** Y Y **si** X		X starts Y
X **d** Y Y **di** X		X during Y
X **f** Y Y **fi** X		X finishes Y
$X = Y$		X is equal to Y

Figure 2.2: Temporal relations of events in Allen's interval algebra

In (Sridhar, 2010), activities are learnt, in an unsupervised fashion, and recognized from video sequences by reasoning under qualitative spatio-temporal representations (QSTR). Objects positions and their trajectories are extracted from scenes and represented in QSTR. Activities are learnt using a Markov Chain Monte Carlo (MCMC) procedure to find the maximum a posteriori probability (MAP) of candidate interpretations. This work shows an example of unsupervised learning of activities, using simulated and real examples. The qualitative approach (QSTR) is robust to changes in the execution of actions and to sensory errors. Finally, the categorization of activities showed to be reliable in learning functional object categories which provides semantic information from the scene. On the other hand, some limitations of this work are a fixed point of view and a posterior analysis. First, the analysis is performed only in short video sequences, and in short activities. Also, the search space grows very easily as the scene becomes more complex.

In (Young and Hawes, 2013, 2014), QSTR are applied to the analysis of multi-agent behaviour in the RoboCup simulation league, particularly in estimating future behaviours. Positions, trajectories and orientation of the agents in the scene are described using region connection calculus (RCC), qualitative trajectory calculus (QTC) and the Star calculus respectively. Other agent's behaviours are learnt by using a HMM, which is fed with the current observations and a window of previous ones. As not all the data is relevant, this is first filtered. This work presents a study of activity prediction. The results show that qualitative representations are more easy to treat general cases of activities and require less training data. Some drawbacks are that the system posses global information from the environment, and the application domain is restricted.

2.3.2 Activity recognition and robotics

A robot is an ideal system to perform activity recognition. This is, indeed, a desirable skill for an autonomous robot that will share an environment with people. The robot needs to be *aware* of the surrounding humans.

Some important challenges to consider with a robot are:

sensing Sensing data is usually corrupted because of hardware limitations, presence of statistical noise, discretization by the digitalization process, unstable or moving sensors, unpredictable environment conditions, etc. The collected data is restricted by the location of robot (there is no omni-presence) and as a consequence the robot will only gather information from the visible parts of the environment, loosing the rest of

it.

storage and processing Sensory data can grow very easily and its storage and processing becomes a challenge. Domain knowledge is usually restricted to the field of application, as general knowledge sources are too big to be portable and are only available by online consult. The algorithms' complexity is usually big in the required problems to solve (e.g. pattern recognition, logic programming, etc.).

time Time constraints are relevant in robotic systems as, for many interesting applications, the data cannot be post-processed, i.e. real-time response is required.

During the early years of robotics, much of the effort in the field was put in designing reliable motion planning and control techniques, e.g. (Brooks, 1985; Moravec, 1983). Meanwhile, new advances were also made in fields as computer vision (e.g. optical flow, visual tracking, etc.), knowledge representation and reasoning (e.g. qualitative reasoning, frame languages), and machine learning (e.g. HMMs, decision trees, etc.). It was until the late 80s and early 90s that robots started facing realistic and non-controlled environments.

One of the first works in activity recognition with a mobile robot can be found in (Bonasso et al., 1996; Kortenkamp et al., 1996). The authors used a monochromatic stereo-vision system mounted on top of a mobile robot for gesture recognition. They implemented an active approach by dividing the scene in cubic volumes. Volumes with similar motion vectors are merged and they are chained to a known human model to produce a linked representation of the human. The angles of the linked representation are compared to a set of previously defined gestures, to label the execution of the gesture. This work only uses a static representation of gestures within a single layer, and was tested in a scenario where a human points regions of interest to a robot, and the robot gives a response. The authors point the the consideration of the temporal dimension, group activity recognition and integration with speech recognition as interesting research directions.

As mobile robots started to be used in everyday environments, human robot interaction became more relevant. For example, in (Burgard et al., 1999) the authors focused in the problem of motion planning in human environments (a museum). However, they point the relevance of human-robot interaction, a robot is not an isolated agent, and humans provide important information about the environment and to be able to complete the task (e.g. giving a tour). In future work, the same group studied the problem of human tracking from a mobile platform using particle filters (Schulz et al.,

2001, 2003) and motion behaviour recognition using the EM algorithm (Bennewitz et al., 2002, 2004). The authors point interesting research open areas in group tracking (instead of individual tracking), particularly the necessity of a flexible approach to handle individual and group tracks of humans and objects at the same time. Also, as motion patterns have been learnt in a particular environment, an interesting problem is to be able to reuse this knowledge in different scenarios where people show similar behaviours, i.e. a portable gait recognition system.

Human activity recognition also plays a central role in the problem of human-robotic cooperation. In this context, to achieve cooperation, a robot needs to be aware of its human partner and integrate itself to the common task in a non-obstructive way. With this in mind, a robot needs to be able to observe and also to communicate with his human partner. This requires a sensory approach of activity recognition, but also a high level treatment of the problem, to integrate the context of speech to the task execution. In (Lallee et al., 2010) the authors present an apprentice robotic system that uses speech and gesture recognition to learn new tasks from a human demonstrator. A Spoken Language Programming system (SLP) was developed (Dominey et al., 2007) to map sentences to actions, to allow verbal commands for the robot. The task for the robot is to assist a human to build a table by passing and holding material. SLP enables the system to extract semantic features from a spoken sentence: action, objects, agents. These are mapped to a set of atomic actions for their system to be able to execute the task. While executing the actions or interacting with the demonstrator, the robot visually follow the execution of the action in order to anticipate future actions or to learn new ones. Progressive benchmarking is used by the robot to learn and anticipate actions and interactions, so the robot can eventually gain confidence and take the initiative of the execution of an action. This approach has enabled a robot with defined primitive actions to assist a human demonstrator in the execution of a complex task, to learn new and complex tasks, and eventually to take the initiative in the execution of subtasks that are necessary to reach the final goal. Human interaction provides robustness for the robot understanding, first by speech recognition, but also by visual scene analysis.

Another work with a similar approach has been described in (Karg and Kirsch, 2012, 2013a,b; Karg et al., 2011), here the goal is to perform a scene diagnosis and to detect abnormal situations on it. An expectations framework has been proposed to create internal representations of *normality* for the environment. In accordance with the authors, expectations should be probabilistic and adaptable; their approach considers to build them by merging information from different sources. The approach relies on motion tracking

data and a semantic map of the environment. With this, the authors are able to segment occurring actions and to maintain probabilistic representations of activities, which are used to detect feasible future states and in accordance with a normality metric, to detect current abnormal states. This system has been tested using the TUM kitchen dataset (Tenorth et al., 2009) and simulation data. The authors express their interest in extending the framework to express expectations in a more probabilistic fashion. Also relevant, is to test the system in different scenarios (e.g. a new kitchen), where a semantic map is available and motion tracks of objects can be obtained; and to be able to segment activities properly and to detect abnormal states.

In (Ramirez-Amaro et al., 2014), the aim is to recognize activities by trying to minimize sensory observations (visual) and compensating this with semantic information. The goal is to show that with a simple sensory approach and with enough semantic information, high level activities can be inferred and that this approach is more suitable for a robotics context, mainly because of the requirement for online functionality. The authors track the motion of hands and objects from a visual input. The state of hands and the interaction with the objects is converted to a symbolic representation. They train a decision tree to generate semantic rules, which are used by a reasoning engine to generate a model of human behaviour. They explain the relevance of action segmentation, which is the problem of properly generating and grouping the atomic actions from the sensory data.

2.4 Answer Set Programming

Answer set programming (ASP) is form of declarative programming oriented towards difficult, primary NP-hard, search problems. As an outgrowth of research on the use of non-monotonic reasoning in knowledge representation, it is particularly useful in knowledge intensive applications (Lifschitz, 2008).

In declarative programming, in stead of coding the method to solve a problem, the idea is to describe the problem and leave the computer to find the solution. ASP has its roots in deductive databases, logic programming (with negation), logic-based knowledge representation, non-monotonic reasoning and constraint solving (satisfiability testing).

The basic idea in ASP is to express a problem in a logical format so that the models of its representation provide solutions to the original problem. The resulting models are referred as *answer sets* (Gebser, 2013).

A rule is expressed in ASP as:

$$L_0 \, or \ldots or \, L_k \longleftarrow \; L_{k+1}, \ldots, L_m, \; not \, L_{m+1}, \ldots, \; not \, L_n,$$

each L_i is a literal in the sense of classical logic. The above rule means that if $L_{k+1}, \ldots, L_m$ are true and if $L_{m+1}, \ldots, L_n$ can be assumed to be false, then at least one $L_0, \ldots, L_k$ must be true (Gelfond and Lifschitz, 1988). The symbol *not* is called *negation as failure*.

Monotonicity refers to the property of a logic programming system that, when more rules are added, it won't produce a reduction in the set of conclusions of the system. Non-monotonicity allows to a conclusion reduction when more rules are added (Poole, 2010). This concept is important in systems with incoming knowledge, in dynamic and non deterministic scenarios. Also, allows the assumption of truth states, or belief states and a posterior revision of them when more rules are known. It is clear, that this is a desired property, in a logic system, to handle uncertain and incomplete information.

Negation as a failure symbol *not* L_i it is often read as "it is not believed that L_i is true". However, this does not imply that L_i is believed to be false, *not* L_i is a statement about belief (Gelfond, 2014).

2.4.1 ASP as a knowledge representation language

ASP is well suited for modelling problems in the area of Knowledge Representation and Reasoning involving incomplete, inconsistent and changing information (Schaub, 2013). Some of its properties, in this context, are (Maximova-Todorova, 2003):

Restricted monotonicity ASP can behave monotonically which addition of literals about certain predicates.

Language independence The results of a program are not dependant on the ASP solver.

Sort-ignorable The sorts can be ignored through language tolerance.

Knowledge extension Knowledge can be extended by *filtering*, i.e. updating the belief state (Amir and Russell, 2003).

ASP has been particularly applied to reasoning satisfiability problems. However, other problems can be treated too by ASP, as: model enumeration, intersection or union, as well as multi-criteria and multi-objective optimization. Formally, ASP allows for solving all search problems in NP and NP^{NP} in a uniform way.

ASP implementations

ASP implementations work n two steps:

1. A **grounder** builds an intermittent representation of the problem files by generating all possible values of the variables.

2. A **solver** that reads the grounded file and generates the answer sets (solutions).

Since the inception of the concept in the 1980s (Gelfond and Lifschitz, 1988) many implementations of ASP have been created. The majority of them uses the syntax of the language *Lparse*, also called *AnsProlog**.

Some of the most popular ASP frameworks are *potassco* (Gebser et al., 2011), *DVL* (Gebser et al., 2011) and *Smodels* (Niemelä et al., 2000). Potassco is the only current framework that includes a module specifically designed for robotics: *ROSoClingo* (Andres et al., 2013). ROSoClingo is a module built for the standard robotics framework ROS (Quigley et al., 2009) that allows connection with the *clingo* ASP-solver and grounder and extends the traditional ASP paradigm by enabling the possibility to handle incoming data, e.g. robot observations.

2.4.2 ASP in activity recognition

ASP, as a declarative problem solving technique, can be used to model problems with a dynamic or sequential representations. Activities dynamic entities, and time and sequencing needs to be considered. In this case, ASP allows to work with existing formalisms as the Situation Calculus, Fluent Calculus and Event Calculus.

An example of this, can be found in (Partonia, 2015). The authors use action languages (Gelfond and Lifschitz, 1998), which enable to describe dynamic systems and consider the consequences of events. In this work, the action language C_{TAID} is used jointly with ASP for activity recognition to model and obtain an explanation on the execution of the activity. Activities are considered as a sequence of states and a probability value is attached to the recognized activities as a parameter of confidence. The focus was, not only to show which activity occurred, but also how it was performed and how certain is the occurrence of the labelled activity. The authors present their methodology and experiments about it.

Chapter 3

Research Problem

The last chapter presented the state of the art in activity recognition, in general and within the perspective of robotics. While the subject has been studied extensively in the recent years, the problem still provide a fertile research field to test different approaches. Activity recognition is particularly relevant for autonomous robots. It provides valuable knowledge from the environment necessary tu understand a situation, interact with people and learn from their actions.

Also, ASP was presented as a methodology to treat problems of knowledge representation and reasoning. Despite the main concepts of ASP were stated in the late 1980s, this approach of problem solving haven't been completely exploited, particularly in robotics, and still remains as a very active research area within the field of Logic Programming.

We are interested in exploring the possibilities of ASP in the context of autonomous robotics. In general, as a tool to handle problems that require knowledge representation and reasoning, and in particular, related to the problem of activity recognition.

This chapter presents the main problem to be studied in this project: activity recognition with a mobile robot with an ASP-based approach. It also introduces the structure of a framework to build a solution to the problem, putting special emphasis in the usage of ASP. This framework can be built and integrated with current state of the art hardware and software tools.

3.1 Problem Description

The subject of study in this project is **ASP-based activity recognition with a mobile robot**. The interest lies in the spatio-temporal relations that exist between activities and the environment, and how these can be

obtained, handled and used by a mobile robot.

The problem of activity recognition is relevant for robots, particularly for those that share their environment with humans. At the same time, ASP offers an interesting and novel approach for problems that require knowledge representation and reasoning. These three parts: *Activity Recognition* as a problem, *Robotics* as a system platform, and *ASP* as a set of techniques to treat problems that require knowledge representation and reasoning, they have been studied widely separately, but their integration still remains as an open field with particular challenges and potential solutions. This joint integration is the focus for this project.

ASP has not been fully exploited in robotic applications. In a traditional logic programming approach for problem solving, a set of facts and rules is collected first to perform later an analysis *a posteriori* (e.g. making questions). But in robotics, new data is arriving continuously, and the knowledge that the robot has about the environment will need to be updated as soon as possible; some rules may not be valid any more, and also there is the need to handle unknown and uncertain statements. Logic programming allows updates by adding and/or erasing premises and resetting the whole system, but in some ASP extensions this can be done online, e.g. ROSoClingo **??**.

The target application is to be able to build a system, on top of a robot, that can be able to observe actively and annotate the ongoing activities in a location (e.g. a library). By using a robot, human perception is partially emulated, however, this will drive to situations with incomplete information, because most of the times, the robot won't be able to sense all the environment. It is an hypothesis for this project that the lack of sensory information can be complemented with a stronger cognitive approach, in this case, by considering domain knowledge.

The results of the system (recognized activities) can be used in different ways, but particularly as semantic knowledge, which has a meaning, a context and can be used for reasoning. In this particular problem, activities have a spatial and temporal significance, that is the reason why they can be stored in a semantic map. This, can eventually be useful for a robot to have a qualitative description of the environment, to augment its navigation capabilities and task planning, and also, to bridge the gap in human-robot interaction (Kostavelis and Gasteratos, 2015).

3.2 Methodology

This section presents the proposed approach to tackle the problem.

First, the target platform is an autonomous mobile robot. In a general

fashion, a control system for a robot can be simplified as a perception-action loop. An activity recognition system fits in by providing interesting features from the environment (activities) that a robot can use to improve its performance.

The figure 3.1 presents the structure of the overall system. The principal input are observations from the environment in the form of recognized features, e.g. humans, objects, etc. Observations of the world are needed, in order to have a starting point to process. Additionally to this, symbolic information will also be used in the form of a semantic map, domain knowledge (e.g. ontologies), previous experiences and finally, the activity representations. The output of the system is a set of activities and features found on the environment, within a degree of confidence.

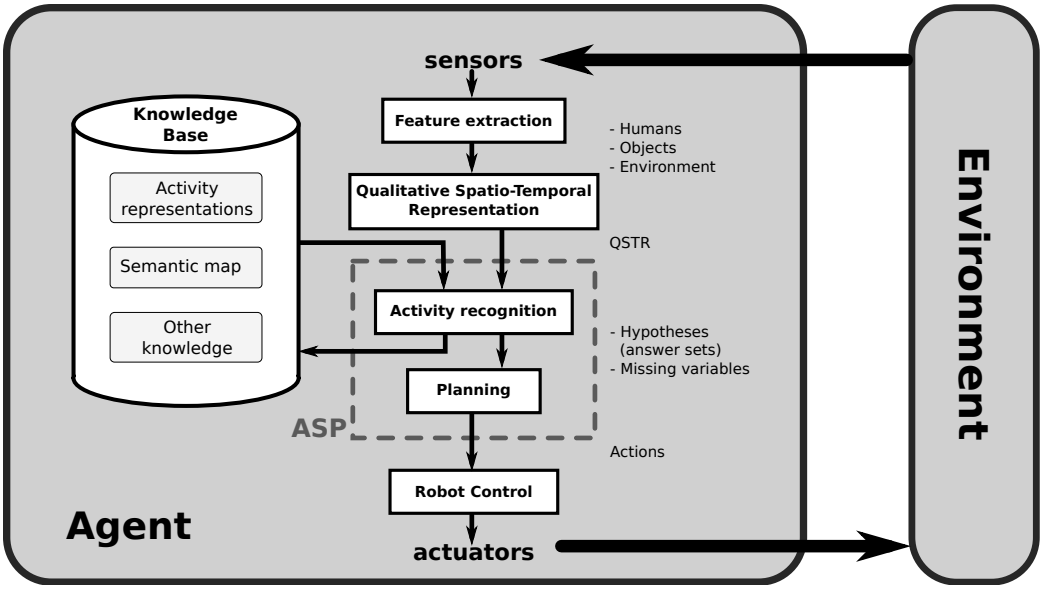


Figure 3.1: Model for the ASP-based activity recognition system with a robot.

3.2.1 Sensing and feature extraction

The end target for this project are autonomous mobile robots, with this in mind, all the sensing is in charge of the robot. No building sensing (e.g. CCTV) or portable devices (e.g. cellphone, laptop) are allowed, this is the goal. We are looking forward to use state-of-the-art sensing techniques, however, completely reliable sensing is not assured and improving sensing algorithms is beyond the scope of this project, the use of simulations and

environment marks is considered for experimental purposes.

The features to be sensed will depend on the type of activities to be recognized. By using the categorization of activities mentioned in section 3.1, we are interested in the mid-level activities. This considers single human activities and interactions with other humans and/or objects, and excludes gestures and group activities. So, human and object sensing are needed. In the same fashion, the aim is to provide a robot with some understanding of activities within a spatio-temporal conceptualization by building a semantic map, so location and mapping are also a requirement, i.e. environment sensing.

All these give us a path to decompose the scene in three different categories within a 3D space 3.2:

humans detection, recognition, kinematic description, etc.

objects detection, recognition, kinematic description, etc.

environment localization, grid and topological maps, etc.

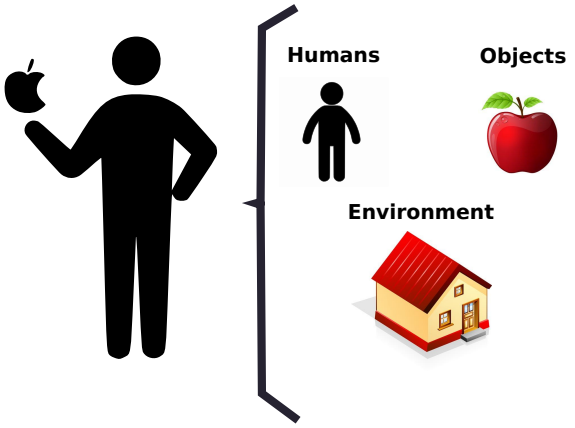


Figure 3.2: Sensing will be targeting relevant features from humans, objects and from the environment.

3.2.2 Qualitative Spatio-Temporal Representations

QSTR were introduced briefly mentioned in chapter 2. They provide formal models to represent and reason about geometrical entities in space and time. Space is handled with Qualitative Spatial Relations. Time is usually modelled in events using Allen's interval algebra (Allen, 1983).

A qualitative approach is desired as it proofs to be more robust to dynamic and non-deterministic environments. It also avoids to depend on training data or parameters which are specific to a particular location, and drives the data towards a symbolic representation. Finally, because it is a more human-alike approach, desired for a better human-robot interaction.

The idea here is to convert quantitative observations to qualitative, by representing the spatio-temporal relations between objects and/or persons, and eventually, with the environment too (e.g. regions of interests).

While it could be a temptation to build all possible relations between the entities of interest in a scene, this has proved to be an non-efficient approach (Sridhar et al., 2010), mostly because the combinatorial explosion of the amount of possible relations. The relations to be handled have to be chosen in a more *intelligent* fashion.

3.2.3 Knowledge Base

Additionally to the observations, a knowledge is required, which contains symbolic information that will be used to process the observations. First, and most importantly, the description of activities are required. This also will include a description from the environment (i.e. a semantic map) and other domain knowledge that can potentially be used to infer new knowledge.

An important question here is about the origin of this knowledge, and in particular of the activities. The simplest approach is to define the activity representations manually. For some specific activities this can be enough, however, in general other strategies are required. Activities can be learnt, from a human demonstrator or by analysing data from repeated situations. Finally the representation of activities can be consulted in larger knowledge databases using some existing tools, e.g. wikiHow (Herrick), OpenCyc (Cycorp), KnowRob (Tenorth and Beetz, 2009), RoboEarth (Zweigle et al., 2009). Here the focus is to share knowledge between robotic system instead of learning it every time as this will take considerable computational effort.

3.2.4 Inferring activities

At this point, the system has a set of qualitative observations in time and space of interesting features. Along time, these observations should be updated or thrown to have a compact set of valid premises.

Now, the problem is to map the observations into an activity representation. As mentioned before, ASP provides a the techniques to solve difficult combinatorial problems. and in particular, search problem as satisfiability

problems (SAT). This is how, the observations become a set of constraints to find definitions (of activities) that satisfy them.

It should be noticed that the observations do not necessarily fulfil all the required parameters in the definition. However, here is where ASP shows its usability by being able to provide candidate solutions, even when there are unknown parameters.

At the end we have a solution, or many possible solutions to the problem of correspondence between the observations and the defined activities.

3.2.5 Discriminate activities

If there is unique correspondence, then this can be taken as solution for the problem[1]. However, it is expected that this will barely be the case. But, there are many possibilities available to find a solution.

First, additional knowledge sources can be considered. These are, a semantic map of the environment and domain knowledge, which can be represented in ASP.

Secondly, a learning strategy can be implemented to find patterns of *frequently occurring activities* in particular locations, by specific persons or with related objects. This is, to maintain a knowledge base of experiences.

Finally, the described system can take advantage of the fact that the robot is active. The robot can interact with the environment and with persons, and it can choose points of interest in the scene.

For example, a robot may be watching a person in the library with an unknown object in the hands. The robot can create a belief, by context and by experience, that the object is a book. However, he can also go closer to that person to have a better look of the object. Finally, interaction is also possible, and the robot can simply ask the person which object is he/she holding.

One important observation should be made in the previous example, and this is, the fact that the robot should figure where to look, or what information is missing to improve its conclusions. This is an advantage, but also a necessity as it is very unlikely that a robot can have complete coverage of a scene. In the example, the required knowledge is the identification of the object, which can be helpful to discriminate between to activities, e.g. eating and studying. Knowing which parts of the missing knowledge will help to improve the conclusions, or which ones are more important, gives a path to plan an active strategy for the robot to fulfil those perception holes.

[1]Nevertheless, it should the noted that, even though the ASP premises have a logical meaning, it will be very difficult to guarantee complete certainty.

3.2.6 Robot Action

Up to this point, the robot has created an internal model of a scene, regarding activities. This model should be used, and could be incomplete. The forward step is to close the loop and use activeness of the robot to improve the process of activity recognition. This can be done by implementing a plan that takes in consideration priorities as looking for missing information, surface coverage of a location, amount of collected information, points of interest, etc.

Chapter 4

Experimental Approach

Once presented the problem and a proposed methodology to treat it, some approach is required to test the feasibility of the ideas. With this in mind, and given the practical interest of the problem, an experimental approach is proposed, using the library example (section 1.3) as a testing case to implement and analyse the proposed ideas. The library setting is flexible enough to be stated with different degrees of complexity and putting special emphasis on critical areas (e.g. temporal analysis, incomplete information, etc.).

4.1 Library Example

The library setting, presented in section 1.3, provides a good testing environment. It has some desired qualities for experimentation, principally, a bounded space and a compact set of activities.

Let first introduce a simplified version of the library setting. In Figure 4.1 is presented a *linear* library. It has five connected and consecutive regions: (A) main entrance, (B) printing area, (C) reception, (D) bookshelf and (E) common area.

In this simplified world a person and a robot can move linearly to any region, and they don't obstruct each other. An activity is performed by a person by visiting regions in a proper order and spending *enough* time on each of them, these intervals are considered within the activity representation. The challenge here is to recognize the activities of the person. The robot have an active role by tracking and following the human through the regions he visits and labelling this within its internal world model.

The following activities can be defined:

print B

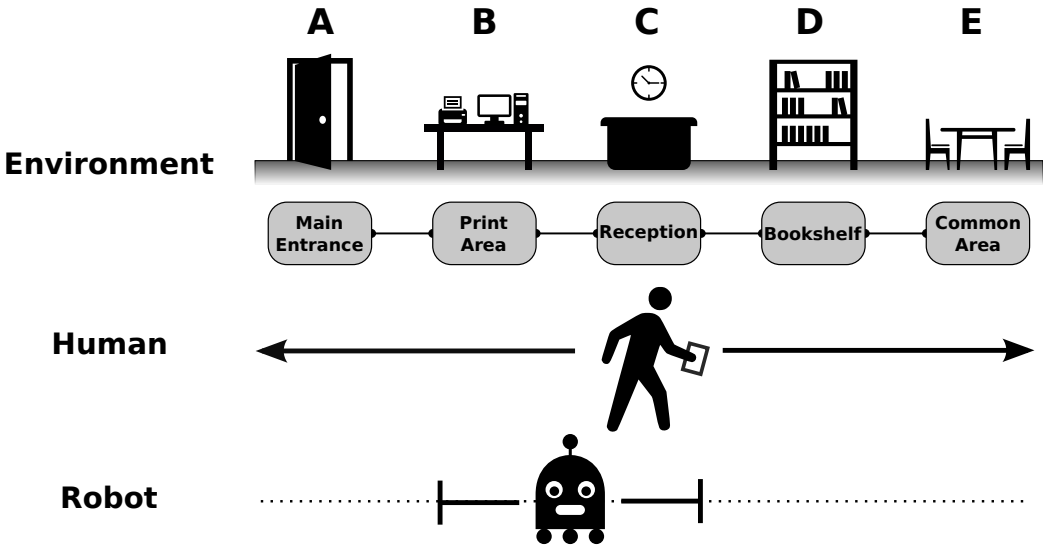


Figure 4.1: Simplified library setting

study $D \rightarrow E \rightarrow D$

bookLoan $D \rightarrow C$

bookRetrieval C

requestAssistance C

Now the problem is, given a particular state of the world, to recognize the ongoing activity. The representations can complete by considering all the occurring states for an activity, but in reality, a robot can realize about the activity in when an intermediate state is occurring. For example, the robot may be watching someone standing in front of the Bookshelf (region D), at this point there are two possible activities to consider: *bookLoan* or *study*, because D is an ambiguous state. In this case, the future state will decide the activity.

Now, some extensions can be induced to this simplified world.

- A robot can have limited sensors, so it will only be able to sense two regions at the same time. So, some activeness is required. For example, being observed a person going towards the common area, the robot can move to confirm the position of the person.

- Handling multiple subjects in the scene, complicates the problem. The robot won't be able to observe all of them, but may try to maximize

the observations and after labelling some individuals may go back and sense the rest of the subjects.

- Object recognition can be induced to disambiguate activities. For example, two activities with the same region description can have an additional parameter regarding the presence of an object. In these cases, a robot can go an confirm the presence or absence of the object to discriminate its hypotheses.

4.1.1 Experimental stages

Four different stages of the library setting are presented that enables progressive analysis and testing of the ideas within this project. These stages goes from simplistic simulations towards a real robotic platform performing activity recognition, which is the ultimate goal of the research in this project.

Terminal Simulation

An implementation of the library setting in as a *terminal* simulation. The activity representations are defined explicitly and the world states can be defined explicitly as well, or generated with another program.

Pros:

- Simplicity
- Flexible to rapidly implement different cases
- Simple agents (humans and robot) can be modelled.
- Controllable parameters
- Repeatability

Cons:

- Non realistic data

MORSE Simulation

MORSE is a simulation environment designed explicitly for robotics which allows compatibility of between the simulation environment and real robotic platforms. An implementation a 3D robotic simulator would enable to test more realistic cases which also can be repeated to test different approaches.

Robots, human avatars and sensors are available to be controlled and to retrieve data.

Pros:

- Robotic oriented environment
- More realistic representations of activities
- Controllable parameters
- Repeatability
- Straight forward software integration

Cons:

- Non realistic data
- Integration takes time

Dataset Analysis

Datasets for activity recognition are available (see (Liu et al., 2011; Tenorth et al., 2009)). They provide a standardized test bed for different techniques. While it would difficult to find datasets that take in consideration a robotic platform, it is worth to test individual parts of the overall system to compare with other approaches.

Pros:

- Standard data test bed and benchmarking
- Results available for comparison
- Repeatability

Cons:

- Analysis *a posteriori*
- No robot present, so no active behaviour would be possible
- Data relies on specific conditions

Robot

The end goal of this project is to provide robots with the capacity to recognize human activities and understand the relation between them and the environment. This goal far beyond the scope of this project. However, experimentation with real robotic systems is an important step in that direction as real environment conditions are difficult to replicate, and also because the application of these king of systems is oriented mostly in this direction.

Pros:

- Real conditions
- New problems can emerge from non considered situations

Cons:

- Difficult repeatability
- Dynamic and non deterministic environment
- Difficult implementation and slow experimentation
- Noisy and Uncertain data

Chapter 5

Work Plan

The main goal of this project is to present a concise study of the integration of ASP and robotics to treat the problem of activity recognition. By this mean, a work plan should be submitted as a guide line to direct future activities and to measure the progress in the the project. In the following section a plan is presented for the next 2 years. The proposed work plan has been created considering three classes of goals: project goals, research goals and academic responsibilities.

First *project goals* are those which moves forward the research project. As the direction of research is now clear (ASP-based activity recognition with a mobile robot), and the literature review is considerably up to date, these goals are in the sense of the implementation, testing and analysis of the proposed methodology, to find the paths that conduce to better results. At the beginning these tasks will be simple and look to proof the feasibility of the proposed ideas. In the final stages, the tasks will be driving the project towards more realistic cases and/or towards specific problems that proof to be relevant.

Research goals are those which serve for research training purposes and those enable the participation of the current project in activities within the research community, e.g. conferences, research school meetings, presentations, academic writing, etc.

Finally, *academic responsibilities* are conformed with those tasks that are mandatory for PhD students within an academic program.

5.1 Work Distribution

T01 - Terminal Simulation

Objectives

- Start the implementation of an activity recognition system using the 2D library setting.
- Explore simple representations of activities.
- Get familiar with ASP tools (Clingo, ROSoClingo).

Tasks

T01-A: The implementation of a simple (terminal) activity recognition system using ASP tools (potassco).

T01-B: Test different representations of activities based on the interaction between spaces (regions) and humans (positions).

T01-C: Test sequential representation of activities, i.e. a human visiting different locations.

T01-D: Consider objects in scene and within the activity representations.

T02 - MORSE Simulation

Objectives

- Implement a more realistic 3D simulation of the library setting.
- Integrate ROS, Potassco and MORSE in a system.
- Extend the representations of activities to a 3D environment; i.e. start considering geometrical entities as regions, trajectories, etc.

Tasks

T02-A: Create the library setting in MORSE and attach one robot to the environment.

T02-B: Integrate ROS and Potassco (via ROSoClingo) with the MORSE simulation.

T02-C: Create interfaces, in ROS, to extract and describe important spatial entities as regions, trajectories, etc.

T02-D: Extend the representation of activities for a 3D environment.

T03 - Qualitative description

Objectives

- Implement qualitative spatial representations between the entities on scene.
- Implement qualitative temporal representations between events.
- Enable the representations of activities to use QSTR.

Tasks

T03-A: Implement RCC and QTC, use regions in the map, human bounding boxes, human trajectories, objects bounding boxes, etc.

T03-B: Implement Allen's Interval Algebra between events.

T03-C: Enable and expand the representation of activities to integrate QSTR[1].

T04 - Knowledge Base

Objectives

- Integrate a knowledge base from different sources, e.g. activity representations, semantic maps, ontologies, online knowledge sources, etc.

Tasks

T04-A: Implement a Knowledge Base from different sources with the capacity to be gradually enlarged and consulted via ASP. The minimum integration would consider the activity representations and the semantic maps.

T04-B: Implement the proper consult mechanisms via ASP.

[1]The desired goal would be to have only qualitative representations. However, a proper balance between qualitative and quantitative approaches is an important topic to study more carefully.

T04-C: Implement a *reporting* module to analyse the conclusions of the activity recognition system (ASP) to **report missing data**. Is important for the system to be aware of which data is missing and also, which part of it is more relevant to get more confident results.

T04-D: Explore the capacity to generate new rules from experience.s

T05 - Semantic Mapping

Objectives

- Implement semantic mapping of activities within the system.
- Spatial and temporal analysis.

Tasks

T05-A: Extract relevant features from the environment, e.g. objects, humans and entities from the environment (regions, corridors, doors, etc.). First, at the simulation stage this can be done by explicitly getting this information, then later this can be done using markers, and finally by using non-invasive sensing methods.

T05-B: Implement the capacity to make maps of activities. Static maps first and then dynamic ones.

T05-C: Integrate pattern recognition techniques to the generation of maps.

T05-D: Integrate the semantic maps into the Knowledge Base and implement mechanisms to generate ASP rules from it.

T06 - Active Perception

Objectives

- Restrict the perception of the robot and implement active strategies (planning, control) to improve the activity recognition process with the mobile robot.

Tasks

T06-A: Implement plans of action for the robot, e.g. maximize surface coverage, maximize human sensing data collection, look for specific data in the environment (missing data), etc.

T06-B: Design experiments (simulated scene) and compare the designed active strategies.

T07 - Real data collection and testing of the ASP-based activity recognition system

Objectives

- Collect real data, and or use standard datasets for activity recognition system.
- Use this data to compare the proposed ASP-based activity recognition with other state-of-the-art approaches.

Tasks

T07-A: Get real data to analyse. This can be done by using standard datasets (Liu et al., 2011; Tenorth et al., 2009), by using collected data from the project STRANDS (Hawes, 2013) or by collecting it from a controlled environment.

T07-B: Integrate a ROS based system to be able to use the collected data, particularly, proper object recognition and human sensing techniques.

T07-C: Analysis and comparison of the results.

T08 - Robot testing

Objectives

- Test the activity recognition system with a robot in a real environment and with an active participation of it.

Tasks

T08-A: Design experiments to run.

T08-B: Integrate system within a robotic platform.

T08-C: Run experiments.

T08-D: Analysis.

T09 - Reporting progress and results

Objectives

- Report the advances within the project.
- Research training in writing and oral presentation.

Tasks

T09-A: Write RSMG reports and prepare thesis group meetings.

T09-B: Participation in activities in the University to present the project.

T09-C: Write thesis.

T10 - Publishing, Conferences and Workshops

Objectives

- Present results to the research community.
- Submit work for external judgement and receive feedback.
- Interact with other researchers in the field.
- Research training in academic writing, oral presentation.

Tasks

T10-A: Prepare and submit work in top conferences in the area (*Goal: Get work accepted for two conferences.*).

- ICRA - International Conference on Robotics and Automation
- IROS - International Conference on Intelligent Robots and Systems
- AAAI - National Conference on Artificial Intelligence
- IJCAI - International Joint Conference on Artificial Intelligence
- RSS - Robotics Science and Systems
- SMC - IEEE International Conference on Systems, Man, and Cybernetics
- LPNMR - Logic Programming and Non-monotonic Reasoning

T10-B: Prepare and submit work for a top journal in the area (*Goal: Get one article accepted for publication*).

- International Journal of Robotics Research (IJRR)
- IEEE Transactions on Robotics (TRO)
- IEEE Robotics and Automation Magazine (RAM)
- Autonomous Robots (AURO)
- Robotics and Autonomous Systems
- Journal of Intelligent Robotic Systems

5.2 Time table

The proposed time table is shown in Fig. 5.1.

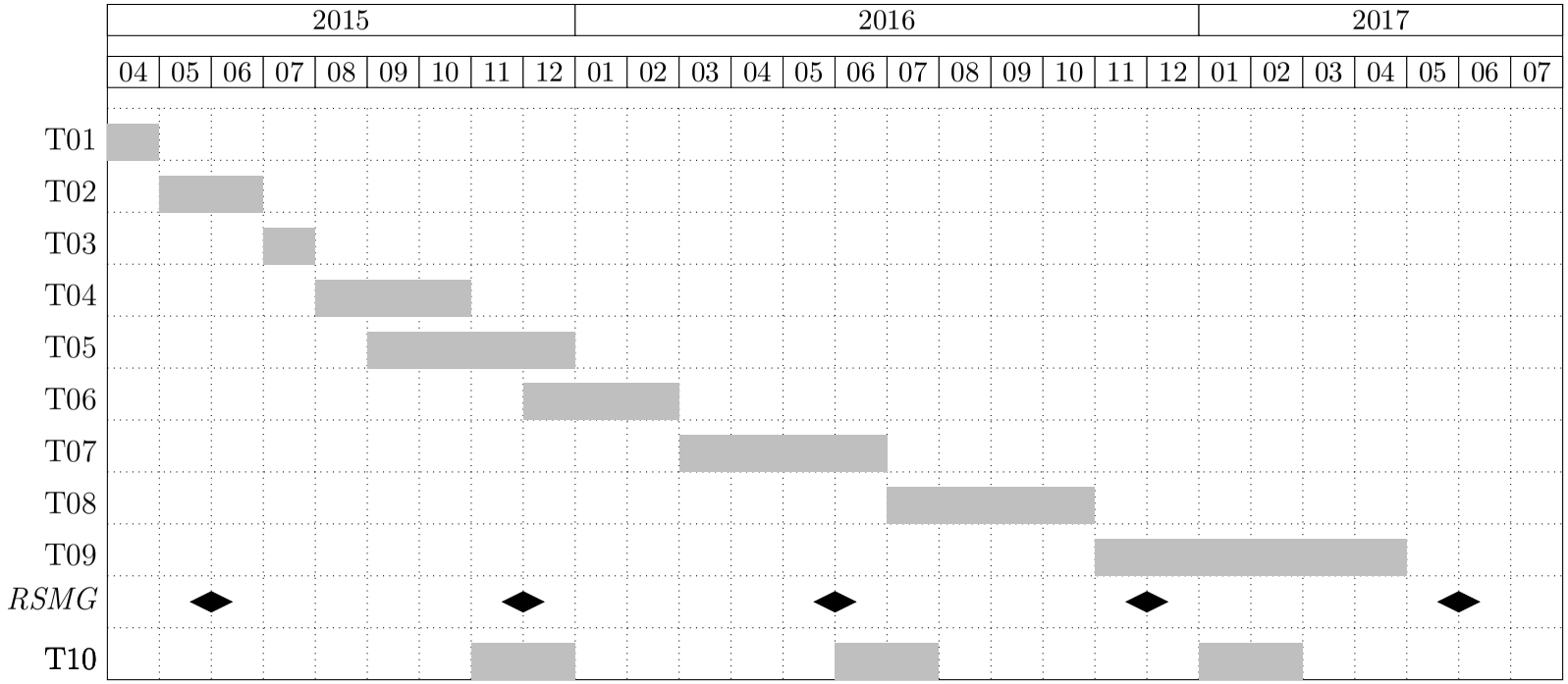


Figure 5.1: Long-term time table

Chapter 6

Conclusions

This report has presented a research project proposal about activity recognition with a mobile robot using ASP.

The problem of activity recognition is difficult and extensive. It can be treated with different approaches. Even more, activities, as languages, are evolving and cultural entities. However, the proper understanding of activities is an important skill to understand human environments. This justifies the studies of the problem and the goal to provide machines with activity recognition capabilities.

This project makes emphasis in the use of a robot, which has some evident limitations. Sensors are constrained to defined ranges and the thrown data usually presents noise. Actuators also present inaccuracies and time-delayed responses. Processing capabilities are also limited in a robot, so the amount of data collected with a robot should be minimized and processed efficiently whenever is possible. This is, however, the current state in robotics and this is considered by considering sensor models,

ASP is considered here as an interesting alternative for knowledge representation and reasoning purposes that has not been fully exploited in the context of robotics and activity recognition. ASP provides robust and well supported tools for declarative problem solving. However, it should also be used with reserve as it has been conceived to model combinatorial NP-hard problems, so the complexity of large systems can be a problem, but not in this project as the targeted environment for testing (library setting) is limited.

The experimental approach presented in this report goes from simple cases towards more realistic ones. The methodology will be evaluated by self and external comparison, and targeted running systems with a robotic platform.

Finally, as work plan has been proposed with specific goals and tasks to achieve it. Research and academic goals are considered.

Bibliography

JK Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

J.K. Aggarwal and Lu Xia. Human activity recognition from 3d data: A review. *Pattern Recognition Letters*, 48(0):70 – 80, 2014. Celebrating the life and work of Maria Petrou.

James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, November 1983.

Eyal Amir and Stuart J. Russell. Logical filtering. In Georg Gottlob and Toby Walsh, editors, *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pages 75–82, San Francisco, 2003. Morgan Kaufmann.

Benjamin Andres, Philipp Obermeier, Orkunt Sabuncu, Torsten Schaub, and David Rajaratnam. ROSoclingo: A ROS package for ASP-based robot control. *CoRR*, abs/1307.7398, 2013.

R. Bajcsy. Active perception. *IEEE Journal on Computer Vision*, 76(8): 996–1006, August 1988.

H. Barrow and R. Popplestone. Relational descriptions in picture processing. In *Machine Intelligence 6*, page 377, 1971.

Maren Bennewitz, Wolfram Burgard, and Sebastian Thrun. Learning motion patterns of persons for mobile service robots. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation, ICRA 2002, May 11-15, 2002, Washington, DC, USA*, pages 3601–3606, 2002.

Maren Bennewitz, Julio Pastrana, and Wolfram Burgard. Active localization of people with a mobile robot based on learned motion behaviors. In *Workshop on Selforganization of Adaptive Behavior (SOAVE)*, 2004.

A.F. Bobick and J.W. Davis. The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(3):257–267, Mar 2001.

R. Peter Bonasso, Eric Huber, and David Kortenkamp. Recognizing and interpreting gestures within the context of an intelligent robot control architecture. In *Technical Report, Metrica Inc. Robotics and Automation Group, NASA Johnson Space*, 1996.

Rodney A. Brooks. A robust layered control system for a mobile robot. Technical Report AIM-864, AI Lab, MIT, September 1985.

W. Burgard, A. B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, October 1999.

H. I. Christensen, Geert-Jan M. Kruijff, and Jeremy L. Wyatt, editors. *Cognitive systems*. Springer, Berlin London, 2010.

A G Cohn and J Renz. Qualitative Spatial Representation and Reasoning. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 551–596. Elsevier, Oxford, 2007.

S. Coradeschi and A. Saffiotti. An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2-3):85–96, 2003. Special issue on perceptual anchoring. Online at http://www.aass.oru.se/Agora/RAS02/.

Cycorp. Opencyc. http://www.cyc.com/platform/opencyc. Last visited in February 2015.

T. Darrell and A. Pentland. Space-time gestures. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on*, pages 335–340, Jun 1993.

P.F. Dominey, A. Mallet, and E. Yoshida. Real-time cooperative behavior acquisition by a humanoid apprentice. In *Humanoid Robots, 2007 7th IEEE-RAS International Conference on*, pages 270–275, Nov 2007.

Esra Erdem, Volkan Patoglu, and Peter Schüller. Levels of integration between low-level reasoning and task planning. *CoRR*, abs/1307.7461, 2013.

Michael Fisher. Temporal representation and reasoning. In Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, *Handbook of Knowledge Representation*, pages 513–550. Elsevier, Amsterdam, 2008.

M. Gebser, R. Kaminski, B. Kaufmann, M. Ostrowski, T. Schaub, and M. Schneider. Potassco: The Potsdam answer set solving collection. *AI Communications*, 24(2):107–124, 2011.

Martin Gebser. *Answer set solving in practice*. Morgan & Claypool Publishers, San Rafael, 2013.

M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *5th Conference on Logic Programming*, pages 1070–1080. Seattle, 1988.

Michael Gelfond. *Knowledge representation, reasoning, and the design of intelligent agents : the answer-set programming approach*. Cambridge University Press, New York, NY, 2014.

Michael Gelfond and Vladimir Lifschitz. Action languages. *Electron. Trans. Artif. Intell.*, 2:193–210, 1998.

Nick Hawes. Project proposal. STRANDS - Spatial-Temporal Representations and Activities For Cognitive Control in Long-Term Scenarios, 2013.

Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, pages 243–259, 1944.

Jack Herrick. wikihow. http://www.wikihow.com. Last visited in February 2015.

Y.A. Ivanov and A.F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):852–872, Aug 2000.

Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, 1973.

Michael Karg and Alexandra Kirsch. Acquisition and Use of Transferable, Spatio-Temporal Plan Representations for Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

Michael Karg and Alexandra Kirsch. An Expectations Framework for Domestic Robot Assistants. In *Conference on Advances in Cognitive Systems*, 2013a.

Michael Karg and Alexandra Kirsch. Simultaneous Plan Recognition and Monitoring (SPRAM) for Robot Assistants. In *Proceedings of Human Robot Collaboration Workshop at Robotics Science and Systems Conference (RSS) 2013*, 2013b.

Michael Karg, Martin Sachenbacher, and Alexandra Kirsch. Towards expectation-based failure recognition for human robot interaction. In *22nd International Workshop on Principles of Diagnosis, Special Track on Open Problem Descriptions*, 2011.

Yan Ke, Rahul Sukthankar, and Martial Hebert. Spatio-temporal shape and flow correlation for action recognition. In *CVPR*. IEEE Computer Society, 2007.

Laura King. *The science of psychology : an appreciative view*. McGraw-Hill Education, New York, NY, 2014.

David Kortenkamp, Eric Huber, and R. Peter Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*, AAAI'96, pages 915–921. AAAI Press, 1996.

Ioannis Kostavelis and Antonios Gasteratos. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66(0):86 – 103, 2015.

S. Lallee, E. Yoshida, A. Mallet, F. Nori, L. Natale, G. Metta, F. Warneken, and P.F. Dominey. Human-robot cooperation based on interaction learning. In Olivier Sigaud and Jan Peters, editors, *From Motor Learning to Interaction Learning in Robots*, volume 264 of *Studies in Computational Intelligence*, pages 491–536. Springer Berlin Heidelberg, 2010.

Lin Liao, Dieter Fox, and Henry Kautz. Hierarchical conditional random fields for gps-based activity recognition. In Sebastian Thrun, Rodney Brooks, and Hugh Durrant-Whyte, editors, *Robotics Research*, volume 28 of *Springer Tracts in Advanced Robotics*, pages 487–506. Springer Berlin Heidelberg, 2007a.

Lin Liao, Donald J. Patterson, Dieter Fox, and Henry Kautz. Learning and inferring transportation routines. *Artif. Intell.*, 171(5-6):311–331, April 2007b.

Vladimir Lifschitz. What is answer set programming? In Dieter Fox and Carla P. Gomes, editors, *AAAI*, pages 1594–1597. AAAI Press, 2008.

Haowei Liu, Rogerio Feris, and Ming-Ting Sun. Benchmarking datasets for human activity recognition. In Thomas B. Moeslund, Adrian Hilton, Volker Krger, and Leonid Sigal, editors, *Visual Analysis of Humans*, pages 411–427. Springer London, 2011.

Yana Maximova-Todorova. Representing commonsense knowledge using Answer Set Programming. Master's thesis, Universidad de las Américas Puebla, Puebla, Mexico, 2003.

H.P. Moravec. The stanford cart and the cmu rover. *Proceedings of the IEEE*, 71(7):872–884, July 1983.

Benjamin Charles Moszkowski. *Reasoning About Digital Circuits*. PhD thesis, Stanford, CA, USA, 1983. AAI8329756.

David Myers. *Psychology*. Worth Publishers, New York, 2013.

R. Nevatia, J. Hobbs, and B. Bolles. An ontology for video event representation. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pages 119–119, June 2004.

Ilkka Niemelä, Patrik Simons, and Tommi Syrjänen. Smodels: A system for answer set programming. *CoRR*, cs.AI/0003033, 2000.

Nils Nilsson. Shakey the robot. Tech Note 323, AI Center, SRI International, 1984.

N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pages 3–8, 2002.

Saeed Partonia. Reasoning about human activity. Master's thesis, UmeåUniversity, 2015.

David Poole. *Artificial intelligence foundations of computational agents*. Cambridge University Press, New York, 2010.

Morgan Quigley, Ken Conley, Brian P. Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

Karinne Ramirez-Amaro, Michael Beetz, and Gordon Cheng. Automatic Segmentation and Recognition of Human Activities from Observation based on Semantic Reasoning . In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sept 2014.

Lawrence Gilman Roberts. *Machine Perception of Three-Dimensional Solids*. PhD thesis, Massachusetts Institute of Technology, June 1963.

M.S. Ryoo and J.K. Aggarwal. Recognition of composite human activities through context-free grammar based representation. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1709–1718, 2006.

Torsten Schaub. Answer set programming: Boolean constraint solving for knowledge representation and reasoning. In Christian Schulte, editor, *CP*, volume 8124 of *Lecture Notes in Computer Science*, pages 3–4. Springer, 2013.

D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 2, pages 1665 – 1670 vol.2, 2001.

Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers. People tracking with mobile robots using sample-based joint probabilistic data association filters. *I. J. Robotic Res.*, 22(2):99–116, 2003.

Muralikrishna Sridhar. *Unsupervised Learning of Event Classes from Video*. PhD thesis, University of Leeds, 2010.

Muralikrishna Sridhar, Anthony G. Cohn, and David C. Hogg. Unsupervised learning of event classes from video. In *Proc. AAAI*, pages 1631–1638. AAAI Press, 2010.

M. Tenorth and M. Beetz. Knowrob, knowledge processing for autonomous personal robots. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 4261–4266, Oct 2009.

M. Tenorth, J. Bandouch, and M. Beetz. The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1089–1096, Sept 2009.

P. K. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Trans. Circuits and Systems for Video Technology*, 18(11):1473–1488, November 2008.

Patrick Henry Winston and Berthold Horn. *The psychology of computer vision*. McGraw-Hill computer science series. McGraw-Hill, New York, 1975. Includes index.

J. Yamato, Jun Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pages 379–385, Jun 1992.

Jay Young and Nick Hawes. Predicting situated behaviour using sequences of abstract spatial relations. In *Proceedings of the AAAI 2013 Fall Symposium How Should Intelligence be Abstracted in AI Research: MDPs, Symbolic Representations, Artificial Neural Networks, or _____?*, 2013.

Jay Young and Nick Hawes. Effects of training data variation and temporal representation in a qsr-based action prediction system. In *AAAI Spring Symposium 2014 on Qualitative Representations for Robots*, Stanford University in Palo Alto, California, US, March 2014.

Oliver Zweigle, René van de Molengraft, Raffaello d'Andrea, and Kai Häussermann. Roboearth: Connecting robots worldwide. In *Proceedings of the 2Nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, ICIS '09, pages 184–191, New York, NY, USA, 2009. ACM.