SPENCER: A Socially Aware Service Robot for Passenger Guidance and Help in Busy Airports

Rudolph Triebel[1], Kai Arras[2], Rachid Alami[3], Lucas Beyer[4], Stefan Breuers[4], Raja Chatila[5], Mohamed Chetouani[5], Daniel Cremers[1], Vanessa Evers[6], Michelangelo Fiore[3], Hayley Hung[7], Omar A. Islas Ramírez[5], Michiel Joosse[6], Harmish Khambhaita[3], Tomasz Kucner[8], Bastian Leibe[4], Achim J. Lilienthal[8], Timm Linder[2], Manja Lohse[6], Martin Magnusson[8], Billy Okal[2], Luigi Palmieri[2], Umer Rafi[4], Marieke van Rooij[9], and Lu Zhang[6,7]

Abstract We present an ample description of a socially compliant mobile robotic platform, which is developed in the EU-funded project SPENCER. The purpose of this robot is to assist, inform and guide passengers in large and busy airports. One particular aim is to bring travellers of connecting flights conveniently and efficiently from their arrival gate to the passport control. The uniqueness of the project stems from the strong demand of service robots for this application with a large potential impact for the aviation industry on one side, and on the other side from the scientific advancements in social robotics, brought forward and achieved in SPENCER. The main contributions of SPENCER are novel methods to perceive, learn, and model human social behavior and to use this knowledge to plan appropriate actions in real-time for mobile platforms. In this paper, we describe how the project advances the fields of detection and tracking of individuals and groups, recognition of human social relations and activities, normative human behavior learning, socially-aware task and motion planning, learning socially annotated maps, and conducting empirical experiments to assess socio-psychological effects of normative robot behaviors.

1 Introduction

The immensely growing passenger volume in air traffic worldwide poses an enormous challenge for all air carriers and airport operators. With the increasing number

[1]Dep. of Computer Science, TU Munich, Germany *[triebel,cremers]@in.tum.de* · [2]Social Robotics Lab, University of Freiburg, Germany *[arras,linder,okal,palmieri]@cs.uni-freiburg.de* · [3]LAAS-CNRS: Laboratory for Analysis and Architecture of Systems, Toulouse, France *[ralami, mfiore, harmish]@laas.fr*, · [4]RWTH Aachen, Germany *[beyer, breuers, leibe]@vision.rwth-aachen.de* · [5]ISIR-CNRS: Institute for Intelligent Systems and Robotics, Paris, France *[chatila, chetouani, islas]@isir.upmc.fr* · [6]University of Twente, The Netherlands *[v.evers, m.p.joosse, m.lohse]@utwente.nl* · [7]Delft University of Technology, The Netherlands *[h.hung,l.zhang]@tudelft.nl* · [8]Örebro University, Sweden *[tomasz.kucner, achim.lilienthal, martin.magnusson]@oru.se* · [9]University of Amsterdam, The Netherlands *m.m.j.w.vanrooij@uva.nl*

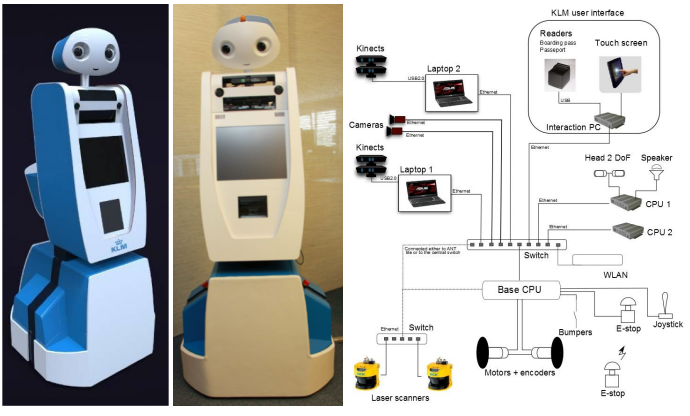
of passengers arriving and departing at an airport, the probability of delays and missed connection flights grows accordingly. Furthermore, busy hubs such as the airport of Amsterdam Schiphol are particularly challenging for the growing numbers of first-time air passengers, people with little knowledge of foreign languages or those who need any kind of special attendance. For them and for others, finding a fast and efficient way from an arrival gate to a departure gate for connection can be very difficult, especially if the first, incoming flight was delayed. For air carriers such as the Dutch KLM, missed connecting flights often result in additional cost for rebooking and baggage reloading, while for the passengers it means further delays and the inconveniences associated with them.

This is the main motivation for the launch of the EU-funded project SPENCER, which we present in this paper. In SPENCER, we develop a mobile robotic platform that efficiently guides oversea passengers at Schiphol airport from their arrival gate to the passport control point for further, inner-European connections, the so-called "Schengen barrier". The project is unique in at least two major aspects: First, it addresses a highly relevant business case with a large potential impact for the entire aviation industry, motivated by a growing need for passenger assistance and the decrease of missed connecting flights. And second, in contrast to earlier tour-guide robot systems (e.g. Burgard et al, 2000; Siegwart et al, 2003), it addresses topics in *social robotics* by developing new methods to perceive, learn and model human social behavior and to use this knowledge to plan appropriate actions in real-time for a mobile robotic platform. In doing so, SPENCER generates novel scientific contributions in the fields of

- detection, tracking and multi-person analysis of individuals and groups of people,
- recognition of human social relations, social hierarchies and social activities,
- normative human behavior learning and modeling,
- socially-aware task, motion and interaction planning,
- learning socially annotated maps in highly dynamic environments,
- empirically evaluating socio-psychological effects of normative robot behaviors.

In SPENCER, we address these problems jointly and in a multi-disciplinary project team, which enables us to exploit synergies between social science and robot engineering for the implementation of an effective cognitive system that operates robustly and safely among humans. In this paper, we present first encouraging results in all mentioned fields, as well as the insights gained from integrating all relevant system components onto the same common platform.

The paper is organized as follows: First, we present an overall view on the system regarding the platform design and the system architecture. Then, we show results of our socially aware localization and mapping module. In Sec. 4 we describe our people and group tracking component, a major building block for social analysis tools. Sec. 5 introduces the human-aware task and motion planning module of SPENCER. Then, we develop important tools to analyse human social behavior and discuss the two main approaches we pursue to implement social behavior on the robot. Finally, Sec. 8 briefly describes the integrated system and concludes the paper.



(a) (b) (c)

Fig. 1 a) and b) Design view and actual appearance of the robot platform. **c)** System architecture.

2 Platform Design and System Architecture

A key element of a socially acting and interacting robot is its physical appearance, because even if the robot's behavior fully complies with socially normative rules, it is of little use if the platform itself appears unfriendly or even threatening. Therefore, a human- or animal-like appearance is often chosen for robots that operate in human environments. However, a completely antropomorphic design has the disadvantage that it implicitly raises expectations regarding certain cognitive capabilites of the platform, which cannot be accomplished with current systems. This can lead to disappointments or to refusal of the system. To avoid this, we decided to use a human-like but abstract appearance, which combines friendliness with believability. The result is a human-size platform (see Fig. 1(a) and 1(b)), where the body resembles the functionality of an information desk, and the head serves as a device for a comprehensible but simplified non-verbal communication (e.g. nodding or orientation towards spokesperson). For physical interaction with the user, the platform has a touchscreen and a boarding pass reader. The sensors consist of two SICK LMS 500 2D laser scanners covering 360° range in total at 0.65m height, two front and two rear RGB-D cameras, and a stereo camera system at shoulder height. A schematic view of the architecture is given in Fig. 1(c). We use the Robot Operating System (ROS, see `www.ros.org`) as a middleware for the software components.

3 SLAM and Socially Annotated Mapping

Airports are very dynamic environments, and this poses a big challenge for the localisation and mapping module. Often, large parts of the range sensors' field of view

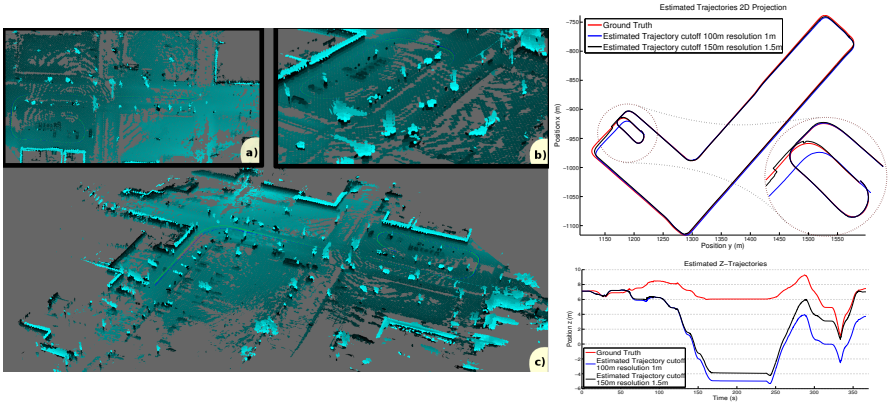
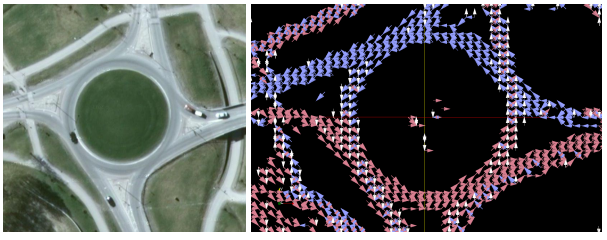


Fig. 2 Mapping and tracking results on the FORD data set. Left: Maps produced by the system while tracking a) top view, b) zoomed view of the start in point, c) overview. The ellipsoids represent height-coded scaled covariance matrices in each map cell from a map at 1 m resolution. Right: trajectory plots, at the top x-y trajectory for the 100 and 150 m cutoff settings, bottom estimated z position over time. Note the zoomed-in detail and the re-entry into a previously mapped area.

are occluded by people or semi-static objects such as carts or trolleys. When these large semi-static obstacles are placed close to walls they can cause major problems in measuring the true distance to the walls. To build consistent maps in environments with high dynamics, we recently introduced the Normal Distributions Transform Occupancy Map (NDT-OM) (Saarinen et al, 2013) and the NDT-OM Fusion algorithm (Stoyanov et al, 2013). We have also developed a data structure called the Conditional Transition Map (CTMap) to model typical motion patterns. Here, we present a novel extension of the CTmap, the Temporal CTMap, which can additionally represent motion speeds. CTMaps are very useful for "social" motion planning, as they enable to plan paths that interfere less likely with the flow of passengers.

3.1 Normal Distributions Transform Occupancy Map

NDT-OM (Saarinen et al, 2013) combines two established mapping approaches: Normal Distribution Transform (NDT) maps (Biber and Straßer, 2003; Magnusson et al, 2007) and occupancy grid maps (Moravec and Elfes, 1985). It has been shown that the NDT-OM Fusion algorithm (Stoyanov et al, 2013) produces consistent maps in large-scale dynamic environments in real time, and it can handle dynamic changes and provide a set of multi-resolution maps. For map building, the vehicle pose is tracked using a frame-to-model registration, and the sensor data are fused into the NDT-OM, by updating distributions with newly obtained and aligned points. By using submap indexing the system can represent large-scale environments at combined registration and fusion times between $100\,ms$ and $2\,s$. Evaluations on the public FORD data set (Pandey et al, 2011) yield absolute trajectory errors (ATE) of



(a) (b)

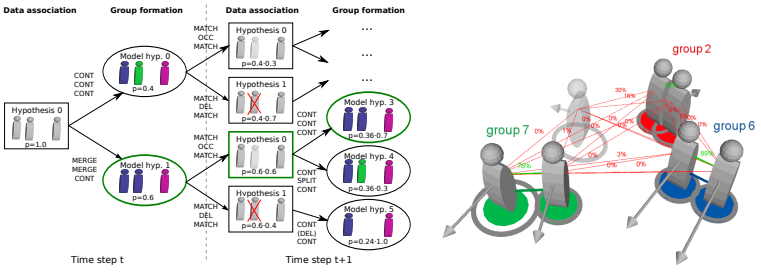
Fig. 3 Visualization of CTMap using data from a roundabout. (a) Overhead view of the enivronment. (b) Pattern of movement on the roundabout, extracted with CTMap, using a cell size of 2×2 m. As a simple denoising step we have removed edges with less than 10 exit events. For clarity, the entry directions are not shown. The colors refer to the orientation of the vectors.

$1.7m$ after $1.5km$ (see Fig. 2). Further evaluations on a ten-hour data set in a large industrial environment resulted in ATEs of under $0.1\,m$ and update rates of $510\,Hz$.

3.2 Conditional Transition Maps

NDT-OM can compactly represent dynamic environments, but for social interaction we also need to distinguish directions of motion. For that, we have developed the Conditional Transition Map (CTMap, Kucner et al, 2013), a grid-based representation that models transitions of dynamic objects in the environment. For each cell $\mathbf{x}$, CTMap learns the probability distribution of an object leaving to each neighboring cell, given the cell from which it entered into $\mathbf{x}$. Based on these learned patterns, motion directions can then be predicted, which is a very important feature for socially aware navigation. We evaluated the CTMap approach on data from a Velodyne-HDL64 3D laser scanner that was placed at the center of a roundabout during rush-hour (see Fig. 3(a)). The obtained CTMap after $1.5h$ of observation is shown in Fig. 3(b). The arrows show the *most likely* exit directions from each cell. They are distributed along highly dynamic areas and closely correspond to the shape of the roads. We also see that the map is able to capture correct motion patterns of pedestrians on the sidewalks.

As an extension to CTMap, we introduce here the Temporal CTMap. In addition to the set of conditional probabilities of exit directions stored for each entry direction of a cell, the T-CTMap stores a bivariate normal distribution to model the dependencies between entry and exit times. This allows us to not only learn the average motion directions and speeds, but also the *variations* of speed. Thus, in contrast to Pomerleau et al (2014), who average velocities of neighboring points over consecutive frames, the T-CTMap represents a complete distribution of velocities.



(a) (b)

Fig. 4 a) In our multi-model MHT approach, group formation hypotheses are interleaved between regular data association hypotheses. **b)** A social network graph, based on the output of a probabilistic SVM trained on coherent motion indicator features (relative velocity, orientation and distance).

4 People and Group Tracking

Another crucial component for a socially compliant robot is a reliable detection and tracking of humans in the environment. As described in Sec. 2, our robot uses 2D laser and RGB-D sensors, and each has benefits and drawbacks. While 2D laser data is more robust against illumination changes and provides a large field of view, it is sparse and has no appearance information. Therefore, we use multiple detection and tracking algorithms that operate on different sensors, as described next.

4.1 2D Range-based Detection and Tracking

To detect people from 2D laser data, we first segment the data points using agglomerative hierarchical clustering. Then we compute 17 different features for each segment and apply a boosted classifier that was previously trained on 9535 frames of hand-labelled data. The resulting detections are tracked using a multi-hypothesis tracker (MHT), which generates hypotheses by considering all feasible assignments between measurements and tracks, all possible interpretations of measurements as new tracks or errors, and all tracks as being matched, occluded or deleted (see Arras et al, 2008). Each hypothesis represents one possible set of assignments between measurements and track labels. Given a parent hypothesis and new detections, the MHT generates a number of assignment sets, where each produces a new child hypothesis branching off from the parent. To prune the exponentially growing hypothesis tree, a probability is computed recursively for each hypothesis using the measurement likelihood, the assignment set probability and the probability of the parent hypothesis. We use multi-parent k-best branching according to Murty (1968) and N-scan back pruning (Cox and Hingorani, 1996). A Kalman filter with a constant-velocity motion model then predicts the state of tracked people.

We extend this MHT approach in Luber and Arras (2013) for the detection and learning of socio-spatial relations and to track social groupings. To do this, layers

(a) (b) (c) (d)

Fig. 5 **a)** Person- and group-tracking experiments during a SPENCER integration meeting. The robot tracks and guides a group of people to the other end of a corridor. **b)** Group affiliations are displayed as green lines connecting the group members. The group is tracked robustly even if individuals are occluded temporarily. **c)** The groundHOG detector most likely detects persons in the distance. **d)** People near the robot, often partly visible, are detected by the upperbody detector.

with group formation hypotheses are interleaved with regular data association hypotheses (see Fig. 4(a)), each leading to a social network graph (see Fig. 4(b)). We reason about social groupings recursively to achieve real-time tracking performance. The resulting group information can be fed back into person-level tracking to predict human motion from intra-group constraints and to aid data association with track-specific occlusion probabilities. This leads to an improved occlusion handling and a better trade-off between false negative and false positive tracks. In experiments on large outdoor data sets, we obtain an improved person tracking by a significant reduction of track identifier switches (TIS) and false negative tracks. In Linder and Arras (2014), we extend this to RGB-D data, and we show that the approach can track groups with varying sizes over long distances with few TIS. Some results of the combined people and group detection and tracking method are shown in Fig. 5.

4.2 Tracking Based on RGB-D Data

For close-range, appearance-based people detection and tracking we developed a real-time RGB-D based multi-person tracker (Jafari et al, 2014), which aims at making maximal use of the depth information from the RGB-D sensors to speed up computation. It classifies the observed 3D points into *object candidates*, *ground*, and *fixed structures*, e.g. walls. *Ground* points are used to estimate the ground plane, and *object candidates* are passed to an efficient upper-body detector (Mitzel and Leibe, 2012), which uses a learned normalized-depth template to find head-shoulder regions. It operates on depth only and is thus limited to the depth range of the RGB-D sensors, i.e. up to 5 meters. To obtain also far-range detections for pedestrians, we combine the upper-body detector with a full-body HOG based detector. This second detector runs efficiently on the GPU and uses the estimated ground plane to restrict the search for geometrically valid object regions (Sudowe and Leibe, 2011). Finally, we use the estimated camera motion, the ground plane and the detections from both detectors for tracking based on Leibe et al (2008) (see Fig. 5(c) and (d)).

5 Human-aware Task and Motion Planning

In SPENCER, there are three main components responsible for planning actions, interactions and the motion of the platform: the supervision system, the task and action planner, and the motion planning module. All three operate human-aware, e.g. by aiming for legibility of the paths and collaborative planning, as detailed next.

5.1 The Supervision System

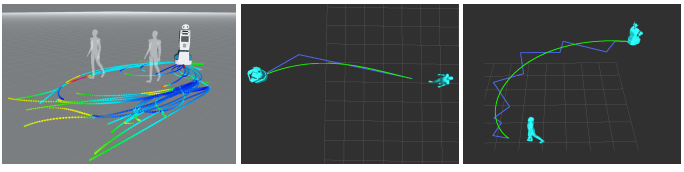
The supervision system (SUP) interacts with the user and generates and executes action plans. For interaction, we use the devices 'lights', 'head', 'screen', and 'microphone' and provide three interaction modes: Engaging with potential users before guiding, giving information to guided users, and asking other people to clear the passage. The SUP also receives safety-critical information, e.g. about planning failures or potential dangers for humans, and reacts accordingly. Using the work of Fiore et al (2014), the SUP was built and sucessfully tested in a simplified scenario.

5.2 Action Planning with Human Collaboration

Action planning and execution alone is not sufficient for a socially aware robot, because it also needs to consider actions performed by the users. For example, while guiding, the robot has to deal with situations where some members of the guided group purposely don't follow the robot. Therefore, we represent the human's intention as a hidden variable and formulate the problem as a Mixed Observability Markov Decision Process (MOMDP, Ong et al, 2009), where in contrast to standard POMDPs some state components are fully observable and others only partially. MOMDPs can be solved much more efficiently than general POMDPs. For cooperation with humans in different tasks we associate to each task a collaboration planner (CP) represented as a MOMDP. To reduce complexity we use a simplified state space, focusing on the intention estimation problem, and let the SUP adapt the MOMDP plans to the current situation. When executing a cooperative action with a human, the SUP gathers observations about the human and updates the corresponding CP, resulting in a high-level action adapted to the situation. In our system, we use a CP for the guiding action and tested it successfully with a single person following the robot. For groups, we currently regard the "most cooperative" behavior, i.e. we consider the group as following as long as a single member follows the robot.

5.3 Socially Compliant Motion Planning

The motion planning module is the system component for which the benefit of complying with social rules is most obvious. Whereas standard planning algorithms



(a) (b) (c)

Fig. 6 a) An example tree generated by the RRT* motion planner on IRL cost maps, when a single relation is in the scene. Red branches are high cost actions, low cost actions are displayed in blue. **b) and c)** Learning to approach a person using IRL. The light blue line is the result of the discretized position and the green line is the smoothed path used by the planner.

mainly aim to find shortest feasible paths, social motion planning trades the shortest path off with the cost of breaking social rules, e.g. when crossing through a group of people instead of deviating it. Therefore, our motion planner extends standard kinodynamic planning in the following ways. First, for global planning we use a human-aware cost map that ensures a path around the detected people, which humans consider as safe. Second, our planning algorithm produces *legible* paths by avoiding abrupt motion changes in presence of dynamic obstacles and by anticipating future collisions and adapting the velocity accordingly. The improved legibility of the produced paths has been experimentally validated in a user study with a robot platform similar to the SPENCER robot (see Kruse et al (2014)).

As a further extension to standard motion planning, we investigate RRT*-based planning (Karaman and Frazzoli, 2010) using low-level vehicle constraints in combination with high-level socially compliant cost maps. Our planner uses a novel extent function for differential-drive robots, which improves the smoothness of the paths and overcomes some limitations of other existing control laws (see Palmieri and Arras (2014)). To reduce planning time, we use a learning approach based on a nonlinear parametric model that infers the distance metric for selecting the nearest vertex in RRT*. Results of our improved RRT* planner using a cost map learned with inverse reinforcement learning (IRL, see Sec. 7.2) are shown in Fig. 6(a).

6 Perception of Human Social Attributes

We have shown how information about social human relations is obtained from basic cues such as tracked groups, and how social rules are used to perform human-aware actions and motions. However, for a deeper analysis and recognition of social relations and attributes, more detailed information must be extracted from the sensor data. Therefore, in SPENCER we develop tools for automatic estimation of body postures, classification of human attributes such as gender and age, estimation of head poses, spokesperson detection, and the classification of important objects in the environment. For the latter three, we present details in the following.

6.1 Head Pose Estimation

An important cue for human social interactions is the head orientation. Groups of people can often be recognized as either standing in a circular formation facing towards the centre, or walking next to each other while looking into the same direction. This suggests that the head orientation can be used to support tasks such as group detection and tracking. To estimate the head orientation, we classify a given upper-body detection as looking left, right, front, back or being a false-positive. Our approach computes a feature covariance matrix of the image's Lab colors and applies a Difference of oriented Gaussians (DooG) filter. The result is split into a regular, overlapping grid and a kernel-SVM is trained on a Riemannian approximation to the geodesic distance between covariance matrices in each cell of the grid. We have evaluated various such approximations, which can trade off computational speed for accuracy, with either an accuracy of up to 93.5% or a two orders of magnitude faster computation than the current state of the art (see Tosato et al, 2013).

6.2 Spokesperson Detection

Another key element of analysing social behaviour is the detection of a *spokesperson*, i.e. a group member who is available for interaction and can make decisions on behalf of the group. Examples include parents in a family and teachers in a school class. For the guiding scenario in SPENCER, determining a spokesperson is particularly useful, because other group members will more likely follow the robot when the spokesperson does. Thus, even if some members are not tracked due to occlusions, the robot can still guide the group as long as the spokesperson is following.

To determine a spokesperson, one can use heuristics such as people's height (this excludes children as a spokesperson) or their position relative to the robot. Another approach is to use people's speech patterns to determine dominance in multi-party meetings (see Hung et al, 2011). However, audio-related cues can not be extracted reliably in airports. Cristani et al (2012) use body behavior and gestures to classify a video of four participants having a conversation into intervals of speech or non-speech. The method achieves 72% accuracy, but the setting is static. However, in an airport people usually move. Also, from our investigations on the same data the movements associated with speech are much shorter-lived than the gesture itself, i.e. different metrics to quantify gesturing are needed. Furthermore, gestures can indicate both speaking and "active listening" behavior. In further experiments with three different implementations of speaker detection using the above data and recordings from speed datings (Veenstra and Hung, 2011), we found that gesturing alone is not a good indication for speech (up to half of the observed speech was not accompanied by strong gesturing), and that the relationship between gesturing and speaking is person-specific. We are therefore investigating the relation of gestures and the length of the subsequent speech period for a more reliable speaker detection. Meanwhile, we use the above mentioned heuristics to determine the spokesperson.

6.3 Efficient Object Classification using Online Learning

Apart from people and their attributes, the robot must also be aware of relevant objects in the environment. In an airport, these include moving objects such as carts and trolleys, which can be dangerous for the robot. However, instead of employing standard offline learning from previously obtained training data, we develop online learning methods for object classification. Particularly, we focus on *autonomous learning* methods, which have the two major advantages that they are adaptive to new situations, i.e. they can incorporate new information by updating their learned models, and they require less user interaction by selectively choosing the data that is particularly useful for training. Based on the work of Triebel et al (2013, 2014), we developed in Mund et al (2015) an efficient online multi-class classifier, that generates less label queries but better classification results than previous methods. This is particularly useful for classifying and learning many different objects online and with only little user interaction, as it is given for the application in SPENCER.

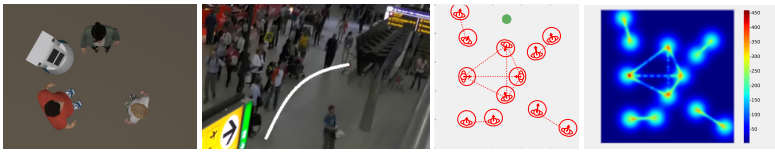
7 Analysis and Learning of Socially Normative Behaviors

So far, we have shown cues to analyse human social behavior, and how social rules can be used to perform a socially compliant robot behavior, particularly during path planning. But how can we obtain these social rules? In principle, there are two different approaches. Either the rules are provided manually by human experts and converted into machine-understandable representations, or they are learned automatically from sensor observations. In SPENCER, we pursue both approaches: High-level, complex rules are established using empirical user studies, and low-level rules are learned automatically from demonstrations. Here, we give two examples.

7.1 User Studies and Contextual Analysis

Airport environments are naturally populated by people from many different cultures. Thus, many different social rules may be required here. One example we investigate is *proxemics* (Hall, 1966), i.e. the distance the robot should keep from a group when interacting. We consider this in the exemplified scenario of a robot approaching a small group of people. The results of an online survey (N=181), which was distributed to people in China, the U.S.A. and Argentina (see Fig. 7(a)), show that participants prefer a robot that stays out of their intimate space zone just like a human would be expected to do (Joosse et al, 2014). However, Chinese participants accepted closer approaches than people from the U.S.A. and Argentinia. This suggests a culturally dependent application of social rules also for SPENCER.

Furthermore, we conducted a contextual analysis at Schiphol Airport to analyze human behavior and to identify observable social rules that the SPENCER robot



(a) (b) (c) (d)

Fig. 7 a) Results of a survey distributed to Chinese, Argentinian and U.S. participants convey cultural different preferences for human-robot spacing. **b)** Context analysis at Schiphol Airport showing that passengers keep a distance from information monitors. Socially normative behavior here means to not pass in front of the passengers. **c)** Example of a social navigation setup. The robot needs to move efficiently from the bottom to the goal (green circle), with minimal disturbance for the people and social groupings indicated by dotted lines. **d)** A costmap learned with IRL for the setup. Areas around people have high cost, but also the 'social' links between individuals.

must be aware of (Joosse et al, 2015). From video data collected during two consecutive days, we established several typical, highly relevant human behaviors. For example, one such behavior is that groups of people tend to walk in pairs or triads behind each other. Another one is the typical avoidance of areas close to information monitors (see Fig. 7(b)). These findings have direct implications both for the perception and the planning module of the system, because they potentially lead to a more reliable group tracking and to a more socially appropriate motion of the robot.

7.2 Behavior Learning via Inverse Reinforcement Learning

Inverse Reinforcement Learning (IRL Abbeel and Ng, 2004) aims at recovering an objective function that encodes a given behavior from an input reward signal. This is more robust than policy search, because rewards are better generalizable and more succinct (see Vasquez et al, 2014). We use Bayesian IRL (Michini and How, 2012) to learn a distribution over the rewards and select the best reward as the MAP estimate. For experiments we use a custom-made pedestrian simulator based on models from computational social sciences to perform behavior tests with arbitrarily large crowds, because testing on the real robot with large crowds is too costly. Fig. 7(c) shows a typical social navigation setup in a crowded environment. The learned costmap using IRL is shown in Fig. 7(d). Such a costmap is then used by the RRT-based motion planner (see Sec. 5.3) to find the desired path for the setup.

Furthermore, we aim at learning relevant social norms when approaching a person. These norms involve a comfortable speed, an appropriate approaching direction and social relations within groups if the person is in a group. Currently, however, we focus on approaching only one person. Again we use IRL, and in particular Gaussian Process IRL (Levine et al, 2011) to learn a policy from a set of demonstrations given by an expert. In our MDP formulation the states are given by distance and orientation in a human-centered frame, and actions are those performed by the motion planner. Two paths learned from 11 demonstrations are shown in Fig. 6(b) and 6(c).

8 System Integration and Conclusion

All presented system components are developed independently and simultaneously. However, to also achieve a steady progress of the entire system, all components are integrated and attuned to each other in regular meetings every six months. As a result, the platform in its current state already combines the map representation presented in Sec. 3, the laser-based people and group tracker (Sec. 4), and the task and motion planner (Sec. 5). Experiments with the complete system have shown that the robot is able to approach and engage with a person, receive a goal position and guide the person or a group to the goal while keeping track of the following person(s). If a failure of cooperation is detected when the person does not follow any more, it stops and waits for re-engagement. Encouraged by these results, a first deployment of the platform at the Schiphol airport is planned for the near future.

References

Abbeel P, Ng AY (2004) Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the Twenty-first International Conference on Machine Learning (ICML), ACM

Arras KO, Grzonka S, Luber M, Burgard W (2008) Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities. In: Proc. IEEE Intern. Conf. on Robotics and Automation (ICRA)

Biber P, Straßer W (2003) The normal distributions transform: a new approach to laser scan matching. In: IROS, IEEE, pp 2743–2748

Burgard W, Cremers A, Fox D, Hähnel D, Lakemeyer G, Schulz D, Steiner W, Thrun S (2000) Experiences with an interactive museum tour-guide robot. Artificial Intelligence 114(1-2):3–55

Cox I, Hingorani S (1996) An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. IEEE Trans Pattern Anal Mach Intell (PAMI) 18(2):138–150

Cristani M, Pesarin A, Vinciarelli A, Crocco M, Murino V (2012) Look at who's talking: Voice activity detection by automated gesture analysis. In: Constr. Ambient Intell., Springer, pp 72–80

Fiore M, Clodic A, Alami R (2014) On Planning and Task achievement Modalities for Human-Robot Collaboration. In: The Intern. Symp. on Experimental Robotics

Hall ET (1966) The Hidden Dimension. Anchor Books New York

Hung H, Huang Y, Friedland G, Gatica-Perez D (2011) Estimating dominance in multi-party meetings using speaker diarization. Tr on Audio, Speech, and Language Processing 19(4):847–860

Jafari OH, Mitzel D, Leibe B (2014) Real-Time RGB-D based People Detection and Tracking for Mobile Robots and Head-Worn Cameras. In: Int. Conf. on Robotics and Automation (ICRA)

Joosse M, Poppe R, Lohse M, Evers V (2014) Cultural Differences in how an Engagement-Seeking Robot should Approach a Group of People. In: Proc. Intern. Conf. on Collaboration Across Boundaries: Culture, Distance & Technology (CABS)

Joosse MP, Lohse M, Evers V (2015) How a guide robot should behave at an airport insights based on observing passengers. Technical Report TR-CTIT-15-01, Enschede

Karaman S, Frazzoli E (2010) Incremental Sampling-based Algorithms for Optimal Motion Planning. In: Proc. of Robotics: Science and Systems (RSS)

Kruse T, Khambhaita H, Alami R, Kirsch A (2014) Evaluating Directional Cost Models in Navigation. In: ACM/IEEE Intern. Conf. on Human-Robot Interaction (HRI)

Kucner T, Saarinen J, Magnusson M, Lilienthal AJ (2013) Conditional Transition Maps: Learning Motion Patterns in Dynamic Environments. In: Proc. IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS), pp. 1196 - 1201.

Leibe B, Schindler K, Van Gool L (2008) Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles. IEEE Trans PAMI 30(10)

Levine S, Popovic Z, Koltun V (2011) Nonlinear Inverse Reinforcement Learning with Gaussian Processes. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ (eds) Adv. in Neural Information Processing Systems 24, pp 19–27

Linder T, Arras K (2014) Multi-Model Hypothesis Tracking of Groups of People in RGB-D Data. In: Proc. IEEE Int. Conf. on Information Fusion (FUSION), pp 1–7

Luber M, Arras KO (2013) Multi-hypothesis social grouping and tracking for mobile robots. In: Robotics: Science and Systems (RSS'13), Berlin, Germany

Magnusson M, Lilienthal A, Duckett T (2007) Scan registration for autonomous mining vehicles using 3D-NDT. Journal of Field Robotics 24(10):803–827

Michini B, How JP (2012) Improving the efficiency of bayesian inverse reinforcement learning. In: Proc. IEEE Intern. Conf. on Robotics and Automation (ICRA), St. Paul, Minnesota, USA

Mitzel D, Leibe B (2012) Close-Range Human Detection for Head-Mounted Cameras. In: British Machine Vision Conference

Moravec H, Elfes A (1985) High resolution maps from wide angle sonar. In: Proc. IEEE Intern. Conf. Robotics and Automation, pp 116–121

Mund D, Triebel R, Cremers D (2015) Active online confidence boosting for efficient object classification. In: Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), to appear

Murty KG (1968) An algorithm for ranking all the assignments in order of increasing cost. Operations Research 16

Ong SC, Png SW, Hsu D, Lee WS (2009) POMDPs for robotic tasks with mixed observability. In: Proc. of Robotics: Science and Systems (RSS)

Palmieri L, Arras K (2014) POSQ: A New RRT Extend Function for Efficient and Smooth Mobile Robot Motion Planning. In: Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS)

Pandey G, McBride JR, Eustice RM (2011) Ford campus vision and lidar data set. International Journal of Robotics Research 30(13):1543–1552

Pomerleau F, Krüsi P, Colas F, Furgale P, Siegwart R (2014) Long-term 3D map maintenance in dynamic environments. In: IEEE International Conference on Robotics & Automation (ICRA)

Saarinen J, Andreasson H, Stoyanov T, Lilienthal A (2013) 3D Normal Distributions Transform Occupancy Maps: An Efficient Representation for Mapping in Dynamic Environments. Intern J of Robotics Research (IJRR) pp 1627–1644

Siegwart R, Arras KO, Bouabdallah S, Burnier D, Froidevaux G, Greppin X, Jensen B, Lorotte A, Mayor L, Meisser M, Philippsen R, Piguet R, Ramel G, Terrien G, Tomatis N (2003) Robox at Expo.02: A large-scale installation of personal robots. RAS 42(3-4):203–222

Stoyanov T, Saarinen J, Andreasson H, Lilienthal A (2013) Normal Distributions Transform Occupancy Map Fusion: Simultaneous Mapping and Tracking in Large Scale Dynamic Environments. In: Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS), pp. 4702 - 4708.

Sudowe P, Leibe B (2011) Efficient Use of Geometric Constraints for Sliding-Window Object Detection in Video. In: International Conference on Computer Vision Systems (ICVS)

Tosato D, Spera M, Cristani M, Vittorio M (2013) Characterizing humans on riemannian manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence

Triebel R, Grimmett H, Paul R, Posner I (2013) Driven learning for driving: How introspection improves semantic mapping. In: Proc of Intern. Symposium on Robotics Research (ISRR)

Triebel R, Stühmer J, Souiai M, Cremers D (2014) Active online learning for interactive segmentation using sparse gaussian processes. In: German Conference on Pattern Recognition (GCPR)

Vasquez D, Okal B, Arras KO (2014) Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison. In: IROS, Chicago, USA

Veenstra A, Hung H (2011) Do they like me? Using video cues to predict desires during speed-dates. In: Intern. Conf. on Computer Vision, Workshops, pp 838–845