Human-Robot Interaction in Public and Smart Spaces

Dylan F. Glas, Koji Kamei, Takayuki Kanda, Takahiro Miyashita, and Norihiro Hagita

ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan 619-0288
{dylan,kamei,kanda,miyasita,hagita}@atr.jp
http://www.irc.atr.jp/en/

Abstract. We have developed a "network robot system" framework with the objective of enabling practical deployment of social robots to provide real-world services in everyday social environments. This framework addresses practical issues in social human-robot interaction by integrating ambient intelligence systems, networked data stores, human supervisors, and centralized planning. All of the elements of the system have been developed and tested in public and commercial spaces such as shopping malls, resulting in a flexible robot control architecture based on practical, real-world requirements. We describe several elements of the system and demonstrate examples of its use in five years of real-world field deployments and research. Finally, we present the Ubiquitous Network Robot Platform (UNR-PF), an internationally-standardized high-level architecture for service robots based on our framework.

Keywords: network robot systems, human-robot interaction, ambient intelligence, cloud robotics, ubiquitous computing.

1 Social Robots in the Real World

Among the many fields of robotics, perhaps no application captures the public imagination like social robots – robots which are a part of our everyday lives, taking the role of social peers, interacting with people conversationally using speech, gaze, and gesture. Yet social robotics is also one of the youngest fields in robotics. Research has been progressing steadily in fields like computer vision, speech processing, and motion planning for safety, and a number of inspiring studies have demonstrated that social robots can be used as museum guides [4, 31], as receptionists for assisting visitors [11], and as peer-tutors in schools [16]. However, today's reality is that most demonstrations of social robotics are still confined to laboratory trials.

Many of the reasons for the continued absence of robots in our everyday social spaces stem from the fact that the noise and unpredictability of the real world and the complexity of social interactions provide enormous challenges for recognition and planning. In a busy shopping mall, school, or train station, a robot needs to localize itself in a world where products and furniture are

S. Mohammed et al. (eds.), *Intelligent Assistive Robots*,
Springer Tracts in Advanced Robotics 106, DOI: 10.1007/978-3-319-12922-8_9

constantly moved around; recognize and track people moving through changing lighting conditions in crowded, unstructured spaces; determine the intentions of people around it; and ultimately to interact with people, identify who it is talking to, perform speech recognition for natural language amid significant noise, and in the end provide a useful service.

With advances in algorithms and improved hardware, many of these recognition and planning problems may be solved in the near (or not-so-near) future. However, by augmenting the limited on-board capabilities of standalone robots using ambient intelligence systems and networked resources, it is possible to accelerate the development of social robots despite the limitations of today's technology.

Fig. 1. Shopping arcade and laser range finders

By embedding sensor systems in the environment as shown in Figure 1, we are able to use networked resources to provide recognition, planning, and coordination between robots, extending their capabilities far beyond the limitations of standalone robots. Combining these ambient intelligence systems with networked data sources and human operators available in the network to occasionally assist with difficult recognition and planning tasks, we have been able to demonstrate prototype robot systems to provide a variety of services in long-term deployments in real shopping malls and other social spaces.

The work presented in this chapter spans about five years of research and field studies aimed towards deploying social robots in public and commercial spaces. We will present an overview of the systems we have developed, as well as some of the practical problems we have faced in the process.

In a world where humans are becoming more and more connected through wireless networks and mobile devices, it is our belief that the "network robot system" approach [25], using environmental sensor systems, is the inevitable future direction of robot system architectures. We hope that the contributions of our work provide a framework for future systems to provide ubiquitous coverage

Fig. 2. Example scenes using our Network Robot System framework. (a) helping a customer with shopping, (b) collaboration between heterogeneous robots.

of social robot services, and ultimately to enhance the quality, enjoyment, and humanity of our lives.

This framework has been successfully used by our research group (Fig. 2 (a)) and in collaboration with others (Fig. 2 (b)) in several field deployments, and we will present many examples of robot services we have implemented in the field using this framework.

Most importantly, the lessons learned throughout our development process have been distilled into a generalizable architecture called the Ubiquitous Network Robot Platform (UNR-PF), which is recognized by international standards organizations. These standards describe an architecture which will connect robots with sensor networks and mobile devices, and enable high-level robot services to be developed independently and executed on a wide variety of robotic platforms, including those based on popular software frameworks such as ROS and RT-Middleware.

2 Human Tracking

A cornerstone of much of our research is a system we have developed for simultaneously tracking the position and body orientation of large numbers of pedestrians, in order to support the spatial perception of robots for navigational interactions. Our technique combines data from a network of laser range finders mounted at torso height. In the tracking algorithm, an individual particle filter is created to estimate the position and velocity of each human over time, and a parametric shape model representing the person's cross-sectional contour is fit to the observed data at each step.

This section will provide an overview of our tracking system. Details of the algorithms used and measures of performance can be found in [9]. This tracking system has been used throughout six years of experiments and field trials, and it has been made into a commercial product called ATRacker[1] which has been used both for robotic and non-robotic applications.

[1] ATRacker is sold by ATR-Promotions. http://www.atr-p.com/HumanTracker.html

2.1 Related Work

Much of the human-tracking research to date has been based on leg tracking, for both mobile robotics [27, 37] and environmental monitoring [2, 5, 39]. This has historically been motivated in part by the fact that many robots use laser sensors for obstacle avoidance, and for that reason already have laser sensors mounted near the ground. However, their visibility is often limited by those same obstacles, making floor-level sensors a good choice for on-board robot systems but less so for wide-area environment monitoring in cluttered spaces.

In our work, the laser sensors constitute an essential part of a ubiquitous sensor network used exclusively for human tracking in real environments. For this reason, it is important for the sensors to be mounted higher, above furniture and ground clutter. Thus the sensors in our system are mounted on poles at a height of 85-90 cm, where the arms and torso can be clearly observed, as shown in Figure 3:

Fig. 3. The ATRacker system features Hokuyo laser range finders mounted on portable poles which can easily be placed in public spaces

Placing the sensors at this height does have the drawback that small children cannot be tracked. This may be acceptable in scenarios where adults are the primary targets of services, provided that the robots are capable of using on-board sensors to detect nearby children for safety reasons.

2.2 Tracking

State Model. The state vector tracked by the particle filter consists of four variables: x, y, v, and θ. The variables x and y represent the position of the human being tracked. Although the speed v, and direction θ of motion could be calculated *a posteriori* from the position data, these variables are included in the state and updated at every step to enable the person's position to be projected forward through time for more accurate tracking. These variables are used in the motion model, described below.

Motion Model. At every update of the particle filter, each particle is propagated according to a motion model. The purpose of this motion model is to approximate the probability of a state $\mathbf{x}_t$ based on the previous state $\mathbf{x}_{t-1}$. To capture the balance between randomness and predictability in human motion, a Gaussian noise component is added to each particle's v and θ values. We then propagate the (x, y) motion linearly according to the resultant v and θ values of the particle.

Likelihood Model. The purpose of the likelihood model is to approximate the value of $p(z_t|\mathbf{x}_t^{[m]})$ based on sensor measurements. In this case, the measurement vector z is an array of raw sensor range data. An effective likelihood model must provide a robust likelihood estimate in spite of noisy sensor data, partial and full occlusions, and the irregular and varying shapes of human bodies.

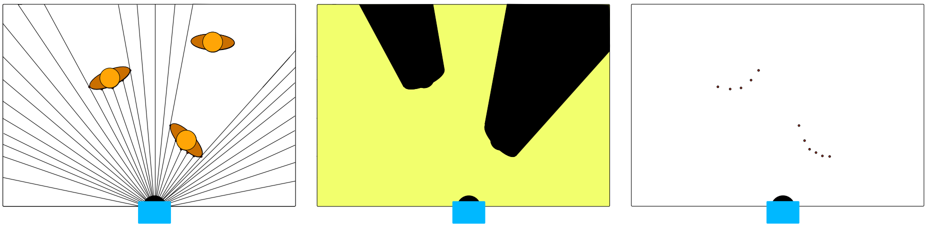


Fig. 4. A typical single-sensor laser scan. (Left) The positions of humans relative to the scanner can be seen. (Center) Occupancy information. (Right) Edge information.

Laser scan data provides two qualitatively distinct types of information useful for estimating human positions: *occupancy information*, indicating whether a certain point is occupied or empty, and *edge information*, indicating a contour which may correspond with the edge of a detected object. Fig. 4 illustrates the distinction between these two kinds of information.

To determine likelihood values from the raw sensor data, it is first necessary to create a background model. Our system uses an adaptive background model which is updated over time to determine the best estimate of the true background distance. Occupancy likelihood is then determined by dividing the world into three regions: "open", "shadow", and "unobservable". The "unobservable" region is beyond the background model for that sensor, and thus can contribute no information. The "open" region has been observed by the sensor to be unoccupied, and the remaining space is considered "shadow". Note also that every "shadow" region lies behind an "edge".

The likelihood model used to compute $p(z_t|\mathbf{x}_t^{[m]})$ is expressed in Eq. 1 and 2 and includes components reflecting both occupancy and edge information.

$$p(z_t|\mathbf{x}_t^{[m]}) = \frac{1}{n_{sensors}} \sum_{i=1}^{n_{sensors}} p_i(z_t|\mathbf{x}_t^{[m]}) - p_{collocation} \tag{1}$$

$$p_i(z_t|\mathbf{x}_t^{[m]}) = \begin{cases} p_{shadow} + p_{edge}(z_t|\mathbf{x}_t^{[m]}) & \text{in a shadow region} \\ p_{open} & \text{in an open region} \end{cases} \tag{2}$$

For a point in a shadow region (strictly speaking, we consider only those regions wide enough to contain a human), the likelihood in Eq. 2 is calculated as the sum of a constant value p_{shadow} and a likelihood $p_{edge}(z_t|\mathbf{x}_t^{[m]})$, calculated as a normal distribution centered upon a point located one approximate human radius behind the observed edge. (In our calculations a value of 25cm was used.) This reflects the fact that people are highly likely to be found just behind an observed edge, yet can plausibly exist anywhere in a shadow region (*e.g.* the occluded person in Fig. 4).

For a point in an open region (or in a shadow region too narrow to contain a human), the likelihood is theoretically zero, but for reasons described below is set to a small but nonzero constant value p_{open}. In this case, edge information is irrelevant.

Finally, in Eq. 1, these likelihood values are averaged across all $n_{sensors}$ sensors for which the proposed point lies within the sensor's "open" or "shadow" range, *i.e.* not "unobservable" to that sensor. To prevent two particle filters from tracking the same human, a value $p_{collocation}$ is subtracted from this result. Its value is calculated as a sum of normal distributions surrounding each of the other humans, based on the list of human positions from the previous time step.

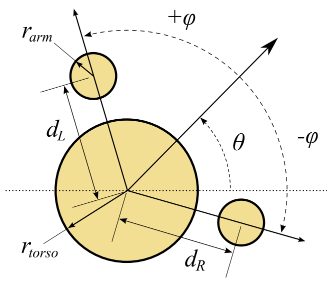
Body Shape Model. After each update of the particle filter, we update a shape model to find a best fit to the torso and arm positions of each pedestrian. Although variations in clothing, backpacks, objects being carried, and individual body size and walking style make it difficult to develop a precise, yet generalizable, model, we found a simple three-circle model to be effective for tracking body orientation.

Our model is illustrated in Figure 5. A central, large circle represents the person's torso, and two smaller circles represent the arms. This model has six parameters which can be varied to best match a subject's cross-sectional body contour.

The update procedure for fitting the shape model to observed data is based on fitting the observed contour of the body to an empirically-derived probability distribution. The details of this procedure are presented in [9]. After these parameters have been adjusted, it is possible to estimate the direction in which a person is facing, even if they are standing still.

2.3 Tracking Results

The tracking accuracy of this system is highly dependent upon several factors, including the placement geometry and number of sensors and the degree of



Parameter	Description
θ	Body orientation
φ	Arm separation angle $\varphi_L = \theta + \varphi$ for left arm $\varphi_R = \theta - \varphi$ for right arm
d_L	Distance of left arm from body
d_R	Distance of right arm from body
r_{arm}	Arm radius
r_{torso}	Torso radius

Fig. 5. Our three-circle model, with the six variable parameters indicated

crowding in the space. Our measurements in laboratory tests have shown accuracy as good as 4.6 cm ± 2.7 cm [9], and users in the field have reported accuracy ranging from 6 cm to 15 cm. Body direction tracking was found to be 8.2 ± 13.8 degrees in our laboratory studies, with occasional 180 deg reversals due to the nearly-symmetrical shape of the human body.

Shape Model Matching. To illustrate the quality of the shape model matching, Figure 6 shows raw data from five frames taken during the course of a single stride, overlaid with the model-based estimates for those time frames. Note that the swinging of the arms is clearly visible from the data, and that the model follows this movement closely.

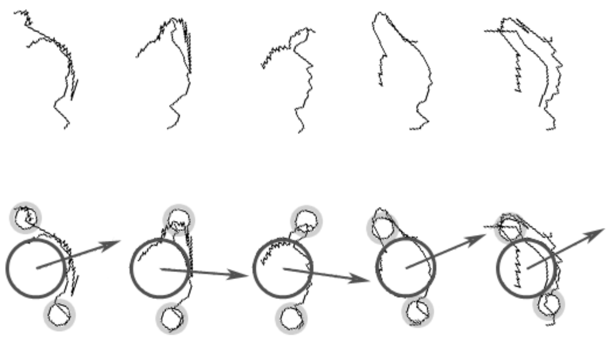


Fig. 6. Example of arm and torso movement during a single stride. Top: Five frames of raw data from laser scanners taken at 320ms intervals. Bottom: Corresponding human shape model positions for each frame.

3 Abstracting and Anticipating Trajectories

Once a high-precision pedestrian tracking system is available in an environment, the next step is to extract socially meaningful knowledge from the data. This section will present an overview of a series of techniques for abstraction of people's trajectories and a service framework for using these techniques to enhance the services provided by mobile social robots. The details of our algorithms and evaluation results for this system can be found in [15].

For a robot providing services to people in public spaces such as shopping malls, it is important to distinguish potential customers, such as window shoppers, from other people, such as busy commuters. The framework presented here enables a designer to make a robot proactively approach customers who exhibit some target local behavior, e.g. walking idly or stopping. This technique was used in a field trial to offer shop recommendations to visitors [28].

The techniques proposed in this section also enable information about the use of space and people's typical global behaviors to be automatically extracted from pedestrian trajectory data. This information enables the robot to anticipate spatial areas in which people are likely to perform the target behaviors, as well as anticipating the probable local behaviors of specific individuals a few seconds in the future. If slow-moving robots can anticipate a person's future behavior, they can start moving early to approach potential customers [26].

3.1 Abstraction Techniques

We use a series of three abstraction techniques for people's trajectories: local behavior, use of space, and global behavior. We define the term **local behavior** to refer to basic human motion primitives, such as walking, running, going straight, and so on. The observation of these local behaviors can then reveal information about the **use of space**, that is, general trends in people's behavior in different areas of the environment. Finally, for more insight into the structure of people's behaviors, we look at **global behavior,** that is, overall trajectory patterns composed of several local behaviors in sequence, such as "entering through the north entrance, walking across a street, and stopping at a shop." Global behaviors are highly dependent on the specific environment.

In addition, since timing is highly critical for social interactions, we also focus on the problem of anticipating the motion and behavior of customers, to determine where the robot should move and which customers the robot should approach. For example, if a robot is designed to invite customers to a shop, it should approach people who are walking slowly and possibly window-shopping. To approach those customers, two anticipation techniques are presented: **location-based anticipation**, based on aggregate behavior patterns observed in the environment, and **behavior-based anticipation**, based on anticipating the specific behavior of an individual person.

3.2 Robot Application Scenario

We conducted our experiments at Universal CityWalk Osaka, a popular entertainment and shopping arcade located by the entrance to Universal Studios Japan, a major theme park. We operated the robots within a 20 m subsection of the arcade, with shops selling clothing and accessories on one side and an open balcony on the other. Within this space, our objective was to enable robots to approach people and offer shop-recommendation services or entertain people. Since many people walking through the space were in a hurry and not interested in talking with the robot, one of our goals was to avoid those busy people and target instead people who appeared to be leisurely window-shopping.

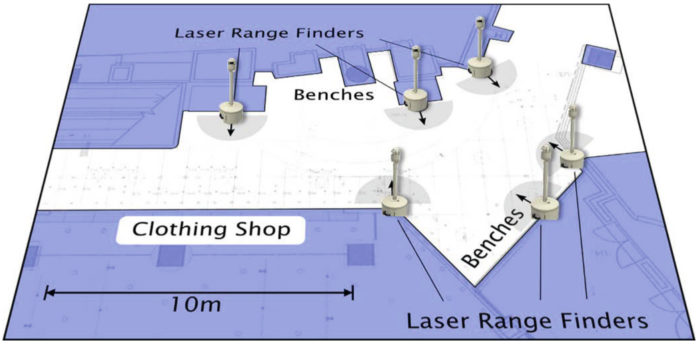


Fig. 7. Placement of laser range finders at the Universal CityWalk Osaka site

The motion of people through this area was monitored using the tracking platform presented in Section 2. Six SICK LMS-200 laser range finders were mounted around the perimeter of the trial area at a height of 85 cm (Figure 7).

3.3 System Design

Figure 8 shows how the sensor information is used to assist with providing robot services. In our framework, data from the position tracking system is abstracted into local behaviors and global behaviors in the recognition system. This information is then used to anticipate the robot's optimal position and generate a roaming path for the robot or to approach a specific individual.

Data Collection. Pedestrian motion data was first collected for a week in the shopping-arcade environment, from 11am-7pm each day, including 5 weekdays and 2 weekend days. We chose this time schedule because the shops open at 11am, and the number of visitors drops after 7pm, after the theme park closes in the evening.

In this environment, the major flow consisted of customers crossing the space from the left to the upper right or vice versa, generally taking about 20 seconds

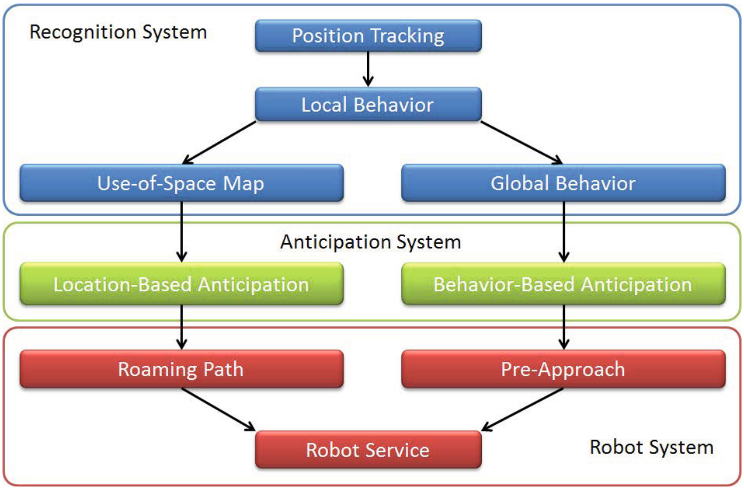


Fig. 8. Service framework

to go through. We removed trajectories shorter than 10 seconds, in order to avoid noise from false detections in the position tracking system. In all, we gathered 21,817 visitor trajectories.[2]

3.4 Local Behavior

As defined earlier, "local behaviors" represent basic human motion primitives. We began our analysis with a classification system which uses SVM (support vector machine) to categorize trajectories based on their velocity, direction, and shape features.

Trajectories were first normalized by rotating them to fit their starting points to the origin and its longest direction to the x axis. A set of 32 features was then extracted describing the shape and size of each trajectory. These features included x-coordinates, y-coordinates, and tangent angles of points sampled along each third of the trajectory, along with min, max, and average x and y values for the overall trajectory shape and a number of angles calculated within the trajectory.

A subset of the trajectories were then manually labeled as belonging to a specific "style" category, describing the trajectory's shape, (walking straight, turning left, turning right, etc.), and a "speed" category, describing the walking speed. This was performed for 5-second and 2-second trajectories, and around 200 trajectories were used for each category.

[2] In this study, we obtained approval from shopping mall administrators for this recording under the condition that the information collected would be carefully managed and only used for research purposes. The experimental protocol was reviewed and approved by our institutional review board.

We then combined these data sets and aggregated the detailed behavior classes into the following four local behavior categories: *fast-walk* (walking quickly in one direction), *idle-walk* (walking more slowly in one direction), *wandering* (turning in one direction or the other, or making a U-turn), and *stop*. Figure 9 shows examples of these local behaviors.

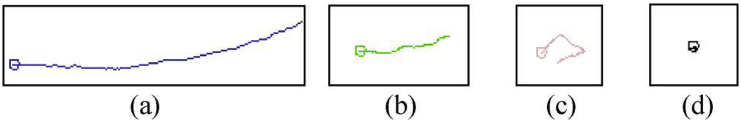


Fig. 9. Example trajectories for local behaviors: (a) fast-walk, (b) idle-walk, (c) wandering, (d) stop

We define the position P_t^n of visitor n at time t to include the x-y coordinates (x, y) as well as Boolean variables indicating the presence or absence of local behavioral primitives $P_{fast-walk}$, $P_{idle-walk}$, $P_{wandering}$, and P_{stop}.

Each trajectory has a sequence of local behaviors represented by these Boolean variables at each time step t. These values are computed by sending the most recent 2-second and 5-second trajectory segments to the SVM classifiers at each time step.

Analysis of Accumulated Trajectories. We then performed an analysis of the local behaviors to obtain a higher-level understanding of the use of space and people's global behaviors. This analysis constitutes the foundation for the robot's ability to anticipate people's local behaviors.

Use of Space (Map). The first analysis task was to identify how the space was used, and how the use of space changed over time. We applied the ISODATA clustering method [1] to achieve this. First, we partitioned the time into one-hour segments categorized as weekday or weekend. We then partitioned the space into a 25cm grid, mapping the environment into 2360 grid elements, and we clustered together the elements with similar local behavior frequencies.

Figure 10 shows a visualized output of the analysis for 40 spatial partitions and 4 temporal partitions of the space. The partitions are color-coded according to the dominant local behavioral primitive in each area.

In some areas, the use of space was very clearly observed to change as a function of time. For example, in Figure 10 (a), *busy-walk* is the dominant primitive in most of the space during weekdays in the daytime, whereas on weekends and evenings, *idle-walk* is more common. The map also provides insight into the spatial distribution of these behaviors, wherein it is clear that people *stop* at the rest spaces and the bench, or slow down in front of a map of the shopping arcade. Customers sometimes slowed down, stopped, and looked at this map.

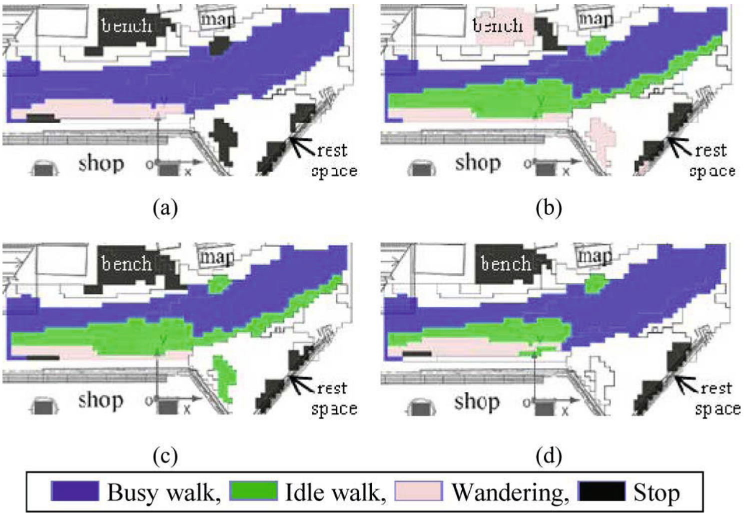


Fig. 10. Analysis of the use of space. (a)Weekday 11am-5pm, weekend 12-1pm (b)Weekday 5-6pm (c)Weekday 6-7pm (d)Weekend 11am-12pm, 1-7pm

The statistical analysis clearly revealed this phenomenon as defining a distinct behavioral space.

The areas where the *wandering* primitive was dominant are colored with pink (or very light gray). All maps in Figure 10 show the space immediately in front of the shop as having this property.

To summarize, we have demonstrated that through this analysis technique, we can separate space into semantically meaningful areas such as the corridor, the space in front of the shop, the area in front of the map, and the rest space. It also reveals how usage patterns change over time, such as the change of dynamics in the space in front of the shop.

3.5 Global Behavior

Global behaviors represent the overall spatiotemporal sequence of local behaviors performed by a pedestrian across a full trajectory. Here we will introduce a method of extracting clusters of global behaviors which represent the typical overall behavior patterns of people in the space.

State Chain Models. We analyzed trajectories based on the *state chain* model illustrated in Figure 11. That is, we converted P_t^n, which is represented in x-y coordinates, to a sequence of states, $S^i = \{s_{t0}^i, s_{t1}^i, \ldots\}$ based on spatial partitioning. s_t^i is defined as $s_t^i = \{n \in N | p_t^i \in A_n\}$ where A_n is the partition the point in trajectory p belongs to. In the example in Figure 11, the trajectory starting from partition 1, stayed in partition 1 for 3 time steps, then entered briefly into

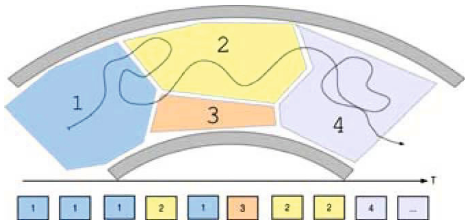


Fig. 11. State chain model

partition 2, and moved back to the partition 1 . . . , which is represented as the sequence of states 1, 1, 1, 2, 1, . . .

Distance between trajectories. We calculate the distance between two state chains, S^i and S^j, by using a dial pulse (DP) matching method (widely used in many research domains, e.g. [24]), which is identical to the comparison of strings known as the Levenshtein distance. Figure 12 illustrates this trajectory comparison technique. Here, we set the distance between partitions as the distance between the centers of the partitions. The cost for "insert" and "delete" operations is calculated as this partition distance plus a constant parameter, which represents the tradeoff cost between time and space.

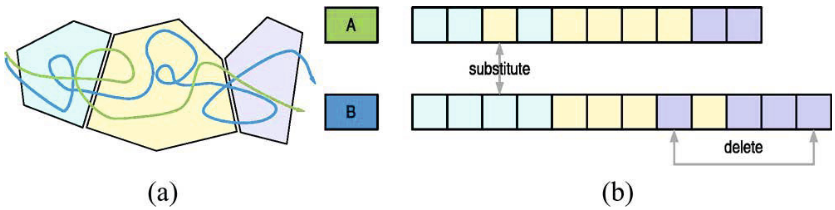


Fig. 12. Comparison of trajectories based on DP matching. (a) Two trajectories. (b) Comparison of state chains of trajectories.

The trajectories are segmented into 500 ms time steps, and they are compared with each other based on the physical distance between them at each time step. To this is added a cost function, based on "insert" and "delete" operation costs in the DP matching, where we defined the cost of a single insertion or deletion to be 1.0 m.

Clustering and Visualization. We then grouped trajectories using k-means clustering to identify typical visiting patterns. The distance between trajectories was provided from DP matching method mentioned above.

Figure 13 shows a visualization of the global behaviors at k=6. For this visualization, we separated the space into 50 similarly-sized partitions by the k-means method [19], although the actual computation used 2360 partitions. In this visualization, each area is colored according to its dominant local behavior primitive, and transitions between adjacent areas are shown as arrows. For example, blue represents *fast-walk*, and green represents *idle-walk*. Solid colors indicate a frequency of occurrence of at least one standard deviation above average, and lighter tints represent weaker dominance, down to white if the frequency is at least one standard deviation below average.

The transitions between adjacent areas are computed for each pair of adjacent areas by counting the transitions in the state chains of the trajectories that belong to each global behavior. Frequent transitions between adjacent areas are shown by arrows.

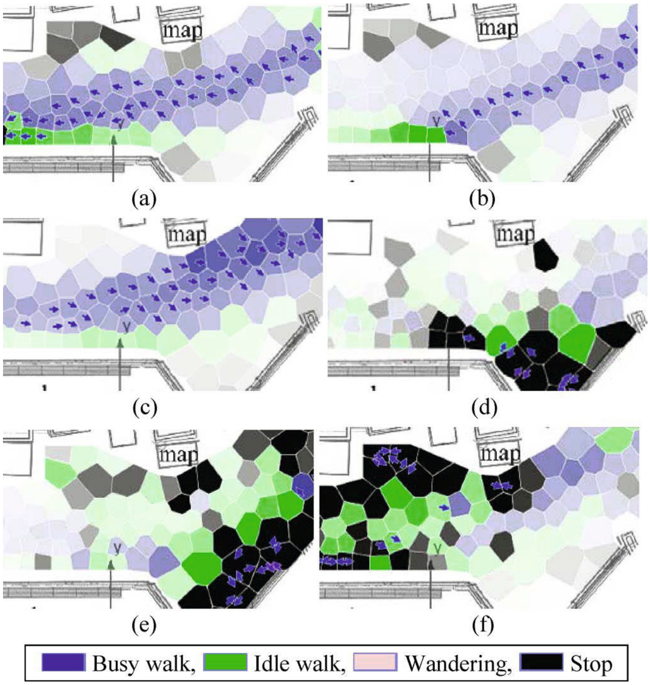


Fig. 13. Six typical patterns of global behavior. (a) From right to left. (b) From right and stop at the shop. (c) From left to right. (d) Rest at the rest space. (e) Around the rest space and right. (f) Around the shop and bench.

We can interpret about six typical global behaviors from Figure 13:

1. Pass through from right to left (7768 people)
2. Come from the right, and stop at the shop (6104 people)
3. Pass through from left to right (7123 people)

4. Rest at the rest space (213 people)
5. Around the rest space and right (275 people)
6. Around the shop and bench (334 people)

In this space, the train station was located to the left and the theme park to the right, so it is possible to interpret the meaning of several of these global behaviors. For example, people often came to stop at the shop while returning from the theme park, but not when coming from the train station.

In summary, this analysis technique has enabled us to extract typical global behavior patterns. These results show that most people simply pass through this space while a smaller number of people stop around the rest space or the map area. People tend to stop at the shop more often when they come from the right, a result which makes intuitive sense, as the shopping arcade is designed mainly to attract people coming back from the theme park.

3.6 Anticipation of Behaviors

Robots differ from other computing systems in that they are mobile, and it takes some time for a robot to reach a person in need of its service. Thus, the ability to anticipate people's actions is important, as it enables the robot to proactively pre-position itself so it can provide service in a timely manner.

We assume here that the robot's service is targeted towards people who are performing some particular local behavior, such as *stop* or *idle-walk*. The robot system uses the results of the analysis about the use of space and global behavioral primitives to anticipate the occurrence of this "target behavior". At the same time, the robot system tries to avoid people who are performing particular local behaviors, such as *fast-walk*, which we refer to as "non-target behavior". To anticipate local behaviors, we use two mechanisms: location-based anticipation and behavior-based anticipation.

Location-Based Anticipation. The robot uses the use-of-space information shown in Figure 10 to estimate the locations in which people will be statistically likely to perform the target behavior and unlikely to perform non-target behaviors. Figure 14 shows an example anticipation map. The darker areas represent areas where the system anticipates both a high likelihood of the target behavior and a low likelihood of the non-target behavior.

The robot roams through this high-likelihood area looking for people. At each time slice t, the system updates the roaming path, $\boldsymbol{P_x}$, to maximize the roaming value calculated from candidates of all possible straight-line paths from 1m to 5m in length on the 25cm-grid.

Usage Example. In one scenario, the robot's task might be to invite people to visit a particular shop. In this case, selecting *idle-walk* as the target behavior and *fast-walk* as the non-target behavior might be appropriate, since the robot wants to attract people who have time and would be likely to visit the store. Figure 14 (a) is the anticipation map for this scenario, calculated for the behavior

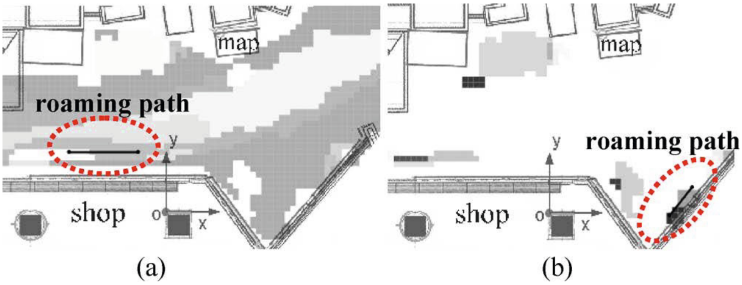


Fig. 14. Example of anticipation map. (a) Weekday 11am-5pm, idle-walk. (b) Weekday 11am-5pm, stop.

patterns observed on weekdays between 11am and 5pm. Several areas away from the center of the corridor are colored, and the roaming path is set in front of the shop. Note that the best path in this case is slightly below the line shown in the figure, but this area is very close to the boundary of the observed map. The robot's final path was translated about 50cm away from the edge for safety reasons.

In a different scenario, the robot's task might be to entertain idle visitors who are taking a break or waiting for friends. Particularly because this shopping arcade was situated near a theme park, this is quite a reasonable expectation. In this case, it would be more appropriate to select *stop* as the target behavior and *fast-walk* as the non-target behavior. Figure 14 (b) is the anticipation map for this second scenario. In this case, only a few areas are colored. The roaming path is set to the bottom-right area.

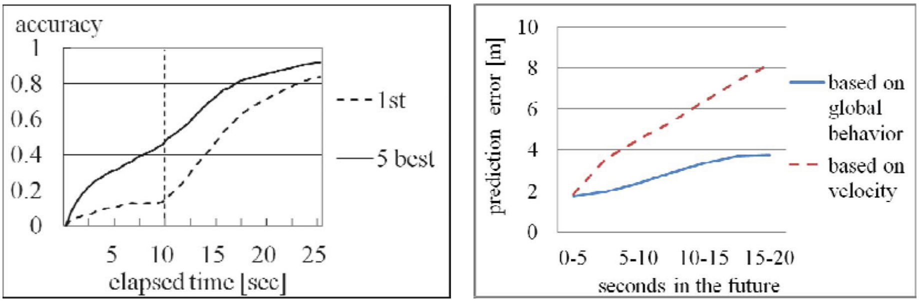
Note that since the roaming path was automatically calculated based on the anticipation map, no additional knowledge about the space was provided by designers.

3.7 Behavior-Based Anticipation

The second technique used for anticipating local behaviors is to estimate the global behaviors of people currently being observed, and then to use that information to predict their expected local behaviors a few seconds in the future.

For this analysis, we used only trajectories from our dataset that were at least 20 seconds in length, resulting in a set of 11,063 trajectories. We clustered these trajectories into 300 global behavior patterns, and for each cluster we identified a representative trajectory at the center of that cluster.

To predict the global behavior of a new trajectory which has been observed for T seconds, the system compares the new trajectory with the first T seconds of the center trajectories of each of the 300 clusters, using the same DP matching technique applied earlier for deriving the global behaviors. The cluster with the minimum distance from the new trajectory is considered to be the best-fit global behavior for that trajectory.



(a) (b)

Fig. 15. (a) Global behavior prediction accuracy. (b) Prediction error for position.

Figure 15 (a) shows the prediction accuracy for observed trajectories from 0 to 25 seconds in length. We used 6 of the 7 days of data to create the prediction model, and tested its ability to predict the remaining one day of the accumulated data. The prediction is counted to be successful if the predicted global behavior matches the classification result after observing the whole length of the trajectory.

The result labeled "1st" shows the case where the best-fit global behavior at time T was correct, and the result labeled "5 best" shows the result if we define success to mean that correct global behavior falls within the top 5 results.

We found that prediction accuracy increases with time as more information is available, and performance levels off after 20 seconds. Since there are 300 global behaviors, we believe that a success rate after 10 seconds of 45% and after 15 seconds of 71% for "5 best" represents fairly good performance.

Likewise, we computed the ability of our technique to predict a person's future position based on an average of the 5 best global behaviors. Figure 15 (b) compares our method with position prediction based on simple projection of the person's velocity over the last second. As the velocity method cannot account for motions like following the shape of the corridor, our method is about twice as accurate.

3.8 Conclusion

Here we have reported a series of abstraction techniques for retrieving information about people's behavior from their trajectories. Based on a set of over twenty thousand trajectories accumulated using robust tracking with multiple laser range finders, we were able to determine statistical patterns of local behaviors and use clustering to reveal typical global behavior patterns in the environment. These results enable us not only to identify human behavior and target robot services appropriately, but also to anticipate people's future behavior, enabling robots to more effectively approach fast-walking pedestrians in dynamic environments.

4 The Network Robot System

Aside from these frameworks for robust pedestrian tracking and trajectory analysis, a number of other components are necessary for deploying social robot teams in real public environments. In this section, we will present our overall robot control framework, which provides coordination between robots operating in the same environment, manages the assignment and scheduling of robot services, allows human operators to assist the robots in difficult recognition tasks, and enables structured knowledge sharing between system elements.

The framework we have developed is based on a "Network Robot System" (NRS) design philosophy, in which the robots themselves are merely the visible component of a network which integrates environmental sensor systems, central planning servers, cloud-based knowledge resources, and human users and supervisors. These elements will be summarized here. A detailed report of the requirements and components used to build this system, together with empirical results from field experiments, can be found in [10].

Although the studies we have conducted with this framework focus on tasks such as guiding customers in a shopping mall, our intention is to share a general approach which can be useful in service robot deployment scenarios like those explored by other groups, e.g. trash collection [21], pedestrian guidance [6], and assisting people in hospitals [22], supermarkets [36], and offices [34].

4.1 System Overview

The high-level elements of our system are shown in Fig. 16. These include a **sensing framework**, several **information registries**, a **coordination module** for navigational coordination and path planning, a system for **service allocation**, and support from a **human supervisor**. Table 1 summarizes these elements.

In this section we will present a general description of each module of our system as well as specific instances of these modules from our implementations.

4.2 Sensing Framework

The sensing framework used in this system has already been described. Its role is to perform precise tracking, trajectory analysis, and behavior anticipation for pedestrians in the environment.

Another important function is identification of individuals, as it enables personalization of robot services. In our work, we have used techniques such as RFID [29, 18], visual face recognition using OKAO Vision[3] software, and Wi-Fi-based identification using smartphones for this task. By combining these results

[3] OKAO Vision, OMRON Corporation, `http://www.omron.com/r_d/coretech/vision/okao.html`

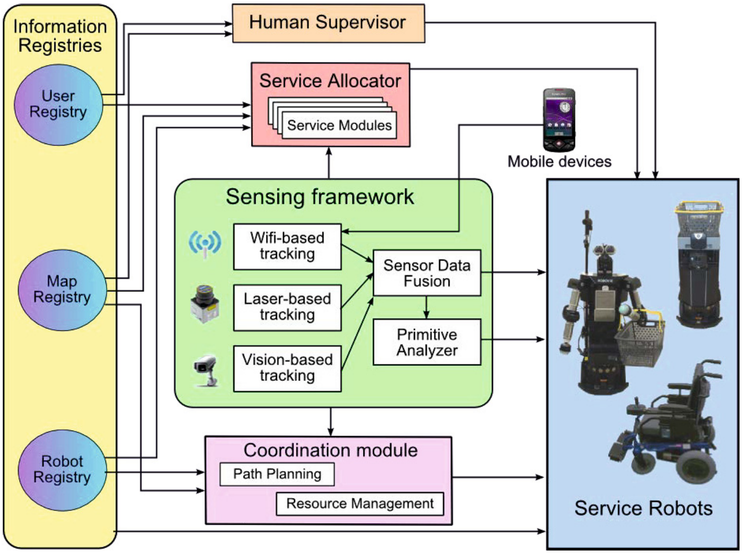


Fig. 16. Overall architecture of our Network Robot System framework

with the high-precision position information available from the position tracking systems, we can locate known individuals with high precision.

In addition to these functions, the human tracking system is often used to assist with robot localization [7], because maps of public spaces often change dramatically from day to day or during the day, as shown in Figure 17. In such environments, fixed sensors provide a better absolute reference than changing features such as product displays.

In fact, we have found precise localization to be quite important for human-robot interaction, as inconsistencies between the coordinate frames of different robots can cause a number of coordination problems. In the example shown in Fig. 18 (left), robots R1 and R2 have slight localization errors, perceiving their own poses to be R1' and R2'. Both robots believe they have identified separate people in need of help, but they have actually detected the same person, resulting in multiple robots offering services to the same person, as shown in the photo. In Fig. 18 (right) one robot has mistaken another for a pedestrian and is trying to initiate a social interaction with it. Since our robots are humanoid in form, they can be mistakenly detected as people by some sensor systems. This has resulted in robots offering services to each other, as shown in the photo. These two problems were common in our early field trials.

Fig. 17. Example of an area in a shopping mall where features change from day to day. Top: photos on two different days. Bottom: laser scan maps of the area on two different days.

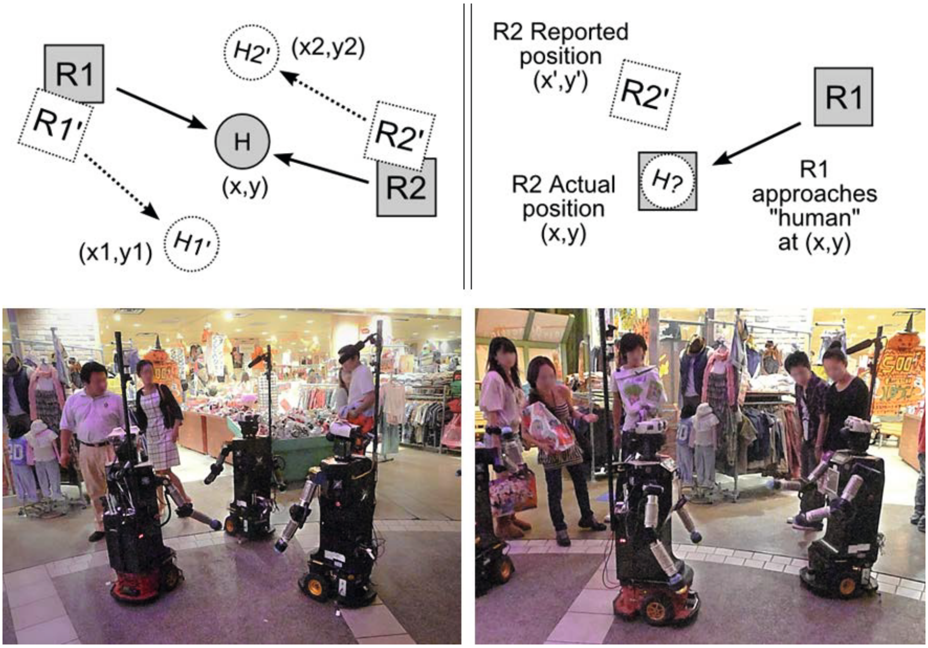


Fig. 18. Social robot failures that can occur due to localization problems. Left: Example of multiple robots approaching the same person. Right: Example of a robot mistaking another robot for a human and talking to that robot instead of an actual customer.

Table 1. Key system elements and functions

System Element	Functionality
Sensing framework	Robust tracking of people Recognize and anticipate people's behavior Assist in robot localization Identify individuals
Information registries	Store information about robots Store information about environments Store information about customers
Path planning and spatial allocation	Coordinate robot paths to avoid conflicts and deadlock Ensure smooth locomotion near people Provide paths based on robot type
Service allocation	Coordinate robot services Enable users to request services Assign services based on anticipated need
Support from a human operator	Support for recognition Direct control of robot Ability to control multiple robots

4.3 Information Registries

Roughly speaking, three categories of information are needed to support a network robot system, summarized in Table 2.

The *Robot Registry (RR)* includes information about the capabilities of each robot, which can be used by a central planner for path planning within each robot's mobility constraints and appropriate allocation of robots to perform services. This information is used by the planner to allocate robots to services appropriate to their capabilities, e.g. assigning a cart robot to a baggage-carrying task.

The *User Registry (UR)* holds personal information about customers (or other service recipients) and is necessary for applications where personalized services are to be provided. For example, in some of our field experiments, customers provided their shopping list information via smartphones. Such information is stored in the UR together with the customer's name and known device ID.

The *Map Registry (MR)* includes navigation and safety maps of an environment, to be used for localization and path planning. In our implementation, navigation maps are generated through offline SLAM using laser scan and odometry data recorded from robots, and safety zone maps are generated by hand.

The safety zone maps are necessary because public environments often contain dangerous areas that a robot cannot detect with its sensors. Transparent and reflective objects such as glass doors and mirrors, or drop-offs such as downward steps, can be difficult or impossible for robots to detect with laser range finders or cameras. Fig. 19 shows some examples of obstacles that are difficult for robots

Fig. 19. Dangerous areas in a shopping mall. Left: glass walls. Center: movable tables and shelves where only legs are visible to ground-level LRF's. Right: movable clothing rack where only the center pole is visible to ground-level LRF's.

Table 2. Summary of information registries

Registry Name	Data Provided
Robot Registry (RR)	Services offered
	Navigable terrain types
	Maximum clearance
	Maximum speed
	Ownership
Map Registry (MR)	Localization map
	Dangerous areas
	Traversability map
User Registry (UR)	Customer name
	Mobile device ID
	Application content
	Personalization data

to detect with ground-level laser range finders (a typical way for robots to detect obstacles). Maps of these invisible obstacles are stored in the MR.

4.4 Coordination Module

The *coordination module* addresses the needs of path planning and of spatial coordination between robots. Access to limited spatial resources is actively managed so as to avoid deadlock between robots which are competing to occupy a critical space, such as advertising robots crowding near an entrance or several robots with low battery attempting to access a charging station. When a robot requests use of a limited spatial resource, the coordination module will grant permission only if the space is not already in use.

Dynamic path planning is also provided, both to avoid collisions between robots and to ensure smooth movement among pedestrians, based on tracking data from the sensing framework. An example is shown in Figure 20.

Fig. 20. Coordinating multiple robots: (a) Robots are given non-conflicting paths to reach their respective customers, (b) Each robot provides its service

The coordination module can also generate socially-meaningful paths, based on the robot's current service task. For example, approach paths should be computed to approach people from a frontal direction rather than the side or back [32, 26]. A robot can also communicate intention through its locomotion, as illustrated by the example of "friendly patrolling" [12].

The path planner also considers traversability constraints. Some robots can traverse uneven surfaces, slopes, or small steps, while others cannot. These differences are reflected in a traversability map for each robot type, stored in the MR, and are used in path planning.

4.5 Service Allocation

The *service allocator* is the central planning mechanism which assigns services to robots and monitors the execution of those services. It handles service requests, identifies service opportunities, handles reservations for future services, and coordinates service allocation across multiple robots by considering the priorities of services and the capabilities and physical locations of the robots in its allocation algorithm.

In our framework, each service to be provided by robots is comprised of several *service tasks*, which are execution units managed by the server. The server contains logic determining which service tasks should be executed, under which conditions, by which types of robots. Once the server assigns a service task to a robot, the robot itself handles the details of service task execution, reporting back to the server upon success or failure.

On-demand services. For on-demand services, customers need a means to request services directly from the system, e.g. using a smartphone, or from the robots directly. In either case, the requests are sent to the service allocator. Service requests can be for immediate service or reservations for future services.

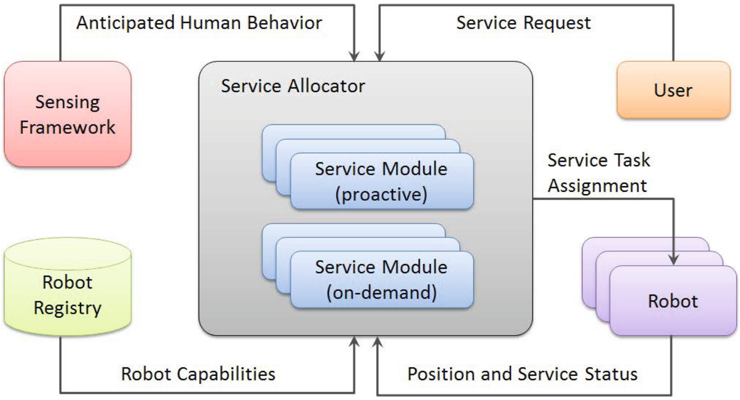


Fig. 21. Basic flow of service task allocation

Proactive services. For "proactive," or "targeted" services, such as giving directions, recommending shops, or advertising services, the robots will approach unknown people to offer their services. In such cases, the service allocator must identify *opportunities* for providing services, rather than responding to requests, and allocation logic must be developed to assign robots to services based on anticipation of who will need or want the service. To do this, it uses the statistical model of pedestrian behavior provided by the *primitive analyzer*, described in Section 3 to target and avoid people performing specific behaviors.

For example, in the case of robots advertising for a shop, the system could be configured to target customers who are exhibiting "stop" or "idle walk" behavior, in the spatial region in front of the shop, and to avoid customers who are performing "busy walk" primitive.

The model of global behaviors can be used to predict customers who are likely to perform this behavior in this area several seconds before they arrive, which gives the system time to allocate a robot and for that robot to move into the appropriate area. The robot is then sent to approach the person and give information about the shop.

4.6 Support from a Human Operator

While robots today have greater capabilities for sensor recognition, dynamic planning, error detection, and error recovery than ever before, they are still far from ready to be deployed autonomously alongside humans in unstructured,

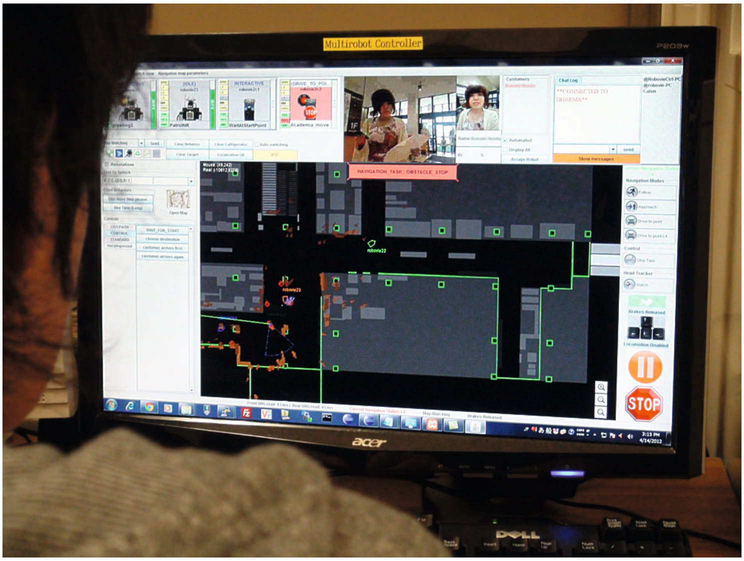


Fig. 22. Operator using a teleoperation console to supervise four robots

public environments. For the foreseeable future, we expect that there will always be a human supervisor present behind the scenes to monitor and assist robots at some level.

In our system, supervisors typically use an interface such as that shown in Fig. 22 to assist the robot's recognition of sensor inputs, e.g. speech recognition or person identification, and to monitor for and correct sensing errors, e.g. identifying dangerous situations or correcting localization. The robot performs its own speech, gesture, and motion planning autonomously, and the role of the human is only to provide occasional sensor inputs.

In rare cases, an operator will need to control the robot directly to handle "uncovered" situations, such as an unexpected question from a customer, or replanning the robot's path to avoid unmodeled obstacles. In these cases, the robot cannot respond autonomously, so the human controls the robot directly.

Finally, some mechanism is needed to enable one operator to manage multiple robots. Techniques such as *proactive timing control* [8] and *conversation fillers* [30] can help improve performance of semi-autonomous robot teams in social interactions.

In the future, as the technology for speech recognition and robot localization improves, the responsibilities of the operator will most likely shift towards handling only rare, exceptional situations and monitoring robots for safety and quality of service. Such a high-level supervisory operator could potentially manage large teams of robots, much like a supervisor at a modern call center.

4.7 Standardization

An open standard for the architecture for robotic services presented here has been formalized and is currently the focus of much standardization work and prototype development. It is known as the Ubiquitous Network Robot Platform (UNR-PF), and its details are introduced in [14].

The common platform architecture was discussed in the International Telecommunication Union Telecommunication Standardization Sector (ITU-T), study group 16 (SG16) as an application of ubiquitous sensor network (USN). The proposal was accepted as a standardization work item in 2011 and was issued as recommendation F.747.3 in March 2013.

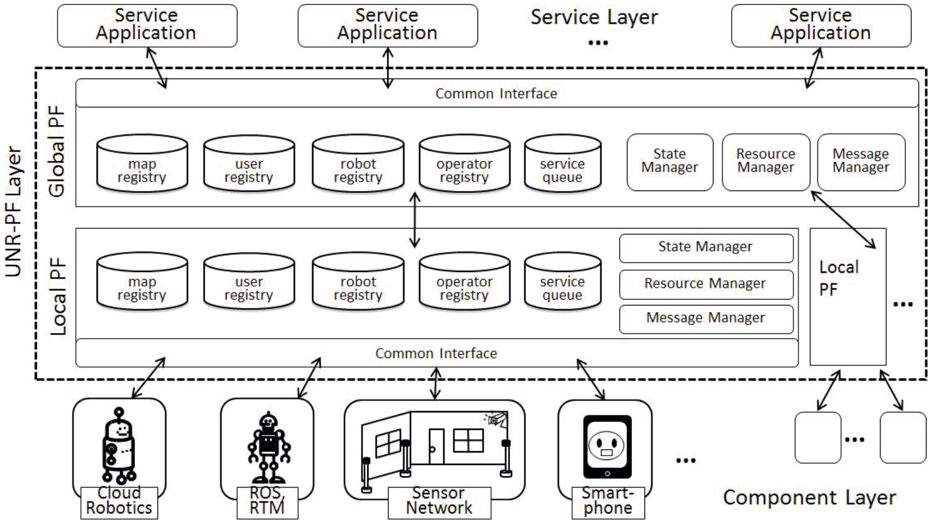


Fig. 23. Overview of the UNR-PF architecture

Figure 23 shows the UNR-PF architecture.

Some extensions in the open platform include an Operator Registry, so that robot operators with varying skills can be accommodated and assigned by the system as human supervisors for specific tasks, and a separation of the platform into local platforms (LPF) and global platforms (GPF), to accommodate deployments in multiple locations.

The standardization of common interfaces between service applications and robotic functional components, that includes robots and sensing framework, is treated in OMG as the Robotic Interaction Service (RoIS) Framework specification, which was issued in Feburary 2013 as RoIS 1.0 specification.

5 Field Deployments

We have deployed the systems described here in several locations, for both short- and long-term field studies as well as experiments for research and many months of development and testing. In this section, we will summarize the kinds of services that were provided in these field trials and how they were supported by the Network Robot System.

Robots. For most of these trials, we used various models of "Robovie," an interactive humanoid robot designed for communication using speech and human-like gesture [17](Figure 24). Robovie has a head with three degrees of freedom (DOF), two arms with four DOF each, and a wheeled differential-drive mobile base. Its height and weight are 120 cm and 40 kg. On its head it has two CCD cameras as eyes and a speaker for a mouth. It is equipped with basic computation resources, and it communicates with the NRS via wireless LAN. We used corpus-based speech synthesis [20] for generating speech. For some applications, we used a shopping-cart robot capable of carrying a shopping basket, shown in Figure 24, with no head or arms but the same speech synthesis and locomotion capabilities.

Fig. 24. Robots used in these studies, left to right: Robovie II, Robovie IIs, shopping-cart robot, Robovie-R3

Environments. Different versions of our system have been used in a variety of locations, including an elderly care home, elementary schools, a train station, and a science museum. However, the field deployments described here were mainly conducted in two locations. The first was the Universal CityWalk Osaka shopping arcade described in Section 3, and the second was APiTA Town Keihanna, a shopping mall near our laboratory which includes a large entry atrium, several aisles between shops with open storefronts, and a supermarket.

5.1 Targeted Services

In many scenarios, we have used the results of the primitive analysis presented in Section 3 to target robot services to people based on the context of their spatial or behavior primitives, rather than responding to explicit requests.

Route Guidance. Figure 25 illustrates an application in which the robot gives people directions. Near a large intersection in a shopping center, a robot is waiting to offer route guidance to customers. A woman stops in front of a map of the mall (Fig. 25 (a)). While looking at the map, she is approached and offered help by the robot: "Are you looking for a particular shop?" (Fig. 25 (b)). The robot then answers any of her questions by giving directions or accompanying her to a destination. This scenario illustrates the need to recognize and anticipate people's needs, and the ability to allocate robot services accordingly.

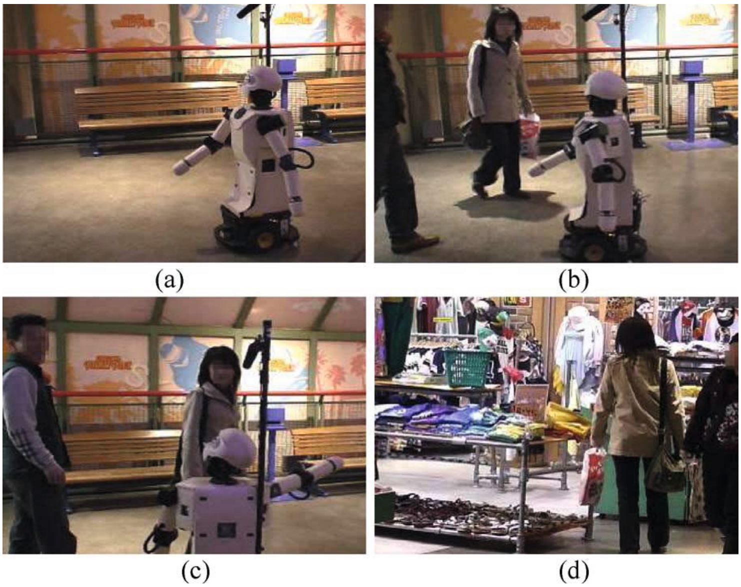
Fig. 25. An example of a *targeted service*: (a) The sensing framework detects a woman stopping in front of a map, then (b) The robot approaches her to offer information

Entertainment Application. In other trials we have had robots provide entertainment by chatting with people. As mentioned earlier, the shopping arcade is next to an amusement park, so it is a reasonable for the robot to be entertaining people who have free time. We think that such an entertainment service would be reasonable for a robot in other environments as well, as robots today are still an exciting novelty.

In this case the robot approached people and chatted about the attractions in the amusement park. For example, the robot said, "Hi, I'm Robovie. Yesterday, I saw the Terminator at Universal Studios. What a strong robot! I want to be cool like the Terminator. 'I'll be back...' ". We set the target local behavior as *stop*, and non-target as *fast-walk*, in order to serve people who are idle. Overall, people seemed to enjoy seeing a robot that approached them and spoke.

Shop Recommendations. The second example is one in which the robot recommends and invites the customer to visit a shop. In the shopping arcade, attracting people's attention to shops and products is an important task. We believe that this is also a reasonable service to expect from a robot, as the novelty of robots makes them very effective in attracting people's attention. The contents the robot provided were simple; for example, the robot said, "Hello, I'm Robovie. Do you enjoy shopping? I'd like to recommend this shop, where they sell clothes by the kilogram!" Whenever it mentioned a shop, it pointed the direction of the shop with a reference term "this" or "that" [33].

In the demonstration, many people were interested in the robot and listened to its invitations. Figure 26 shows an impressive example where the robot approached a couple who were performing *idle-walk*. When the robot pointed to the shop and gave its recommendation (Figure 26 (c)), they smiled with surprise to see a robot performing a real business task. After the robot mentioned the shop, the woman walked directly to the shop and entered it (Figure 26 (d)). Observing such behavior indicates that such an invitation task can be a promising application. As indicated above, the robot was able to attract people's attention and redirect their interests to shops and products.



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

Fig. 26. A robot successfully inviting a person to a shop

5.2 Personalized Services

When personal information for individual customers is registered in the system, the NRS enables personalized services to be provided. In one demonstration we showed a scenario in which a registered customer uses her mobile phone at home to request a robot to help her on her next shopping trip. She enters her shopping list verbally, using a mobile application which connects to the NRS (Fig. 27 (a)). Upon her arrival at the mall (Fig. 27 (b)), a robot comes to greet her, identifies her via her mobile phone, using wifi fingerprinting, and accompanies her through the supermarket carrying her basket (Fig. 27 (c)). Based on their location in the supermarket, the robot can remind her of items on her shopping list. This scenario illustrates how the personal information can be used in services. It also shows the need for a means of personal identification of the service recipient.

Fig. 27. An example of *personalized service*: (a) requesting a robot from her mobile phone, (b) detecting the customer's arrival, and (c) shopping with the customer

5.3 Shopping Assistance

In a medium-term study, shown in Figure 28, Iwamura et al. used the NRS to provide robotic shopping assistance in a supermarket to 24 senior citizens over a set of four shopping trips each [13]. In that study, humanoid and non-humanoid robots were used to accompany the users through the supermarket, conversing with them during the shopping trip. The dialogue was semi-autonomous, with the timing chosen by the teleoperator and the contents chosen by an autonomous system.

Fig. 28. Robot assisting elderly shoppers in a supermarket

5.4 Mobility Assistance

Although in most of our field trials the objective has been to develop robots which are highly autonomous, requiring a minimum level of intervention by an operator, autonomy is not always the objective. In one project, we demonstrated the possibility of using the NRS to support a robotic wheelchair, as shown in Figure 29. In this scenario, a user reserves a wheelchair using a smartphone. The system can detect when the user arrives at the shopping mall using the smartphone's GPS localization, and the robot goes out to meet the user. Once the user is seated in the wheelchair, they can drive the wheelchair themselves using a joystick, taking advantage of the wheelchair's on-board safety system, or they can request to go to a specific destination, using the path planning functionality of the NRS.

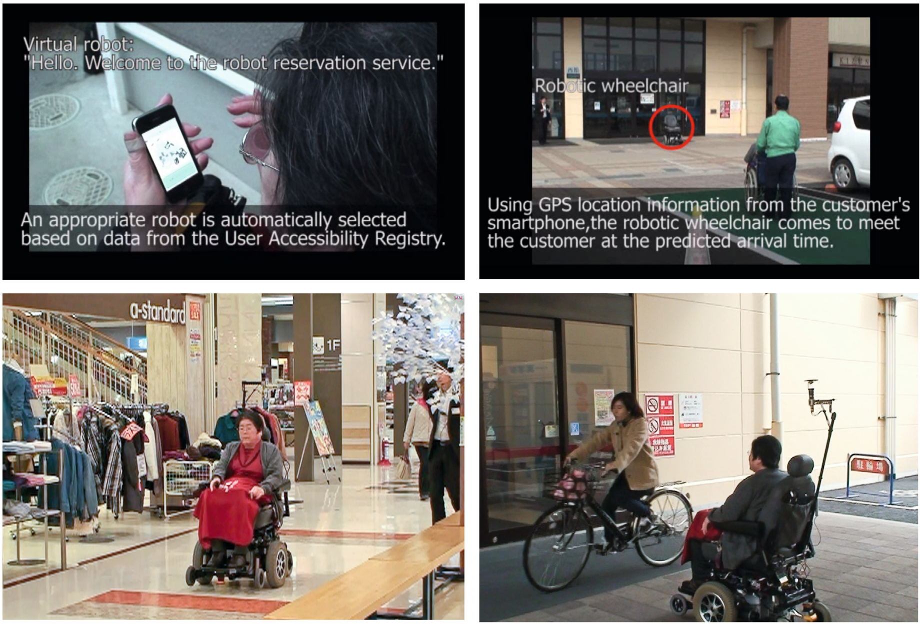


Fig. 29. Robotic wheelchair supported by the Network Robot System. (Top) User reserves the wheelchair robot ahead of time, and robot goes to meet the user as she arrives. (Bottom) Robot guides user safely through shopping mall, avoiding dynamic obstacles.

This scenario is motivated by the fact that many elderly wheelchair users are not comfortable going out alone in their wheelchair, due to safety concerns. This makes them dependent on family members whenever they want to go out. However, by using a robotic wheelchair supported by the NRS, they can take advantage of built-in safety systems, automatic navigation, and most importantly, they can take comfort in knowing that a remote operator is available if they have any trouble. In this scenario, the operator is not merely a replacement for technologies which are not yet available, but is an essential part of the service.

5.5 Cross-Platform Collaboration

One strength of our NRS framework is the ability to easily integrate new robots, new sensors, or new services to the system. Using simple protocols to connect to the planning, coordination, and localization servers, third-party robots can be integrated easily into the NRS without deep or complex software integration, enabling collaboration between very different kinds of robots.

We have conducted collaborative work with the DustCart robot from the EU Dustbot project (Fig. 30), and with Honda's ASIMO robot (Fig. 2). In these demonstrations, Robovie talked with visitors and initiated a collaborative task, wherein the other robot performed some physical task, e.g. serving a drink or carrying baggage, while Robovie continued talking to the visitor, offering chatting or verbal instruction.

Fig. 30. Collaboration between Robovie and DustCart, supported by the NRS

In each case, only 1-2 weeks of implementation and testing were necessary to integrate the new robots with the NRS platform and prepare a collaborative robot demonstration. We have also conducted other NRS demonstrations with robots such as Mitsubishi's Wakamaru and Toshiba's ApriPoco.

Discussion. Further details about people's responses to robots were examined in more detail in succeeding studies, e.g. a study of social behavior in approaching humans [26] and integration of different capabilities of robots [28], which are based on the techniques and service frameworks reported in this paper.

Practical Considerations. Aside from the target functionalities presented in this paper, our experiences have shown a number of practical benefits provided by the modular design of the NRS framework.

When robots experience hardware problems – hard drive crashes, electrical failures, etc., the modularity of the NRS framework makes it easy to swap a backup robot into the system, enabling the experiment or demonstration to continue in a nearly seamless way. Rather than statically specifying services, paths, etc. within individual robots, the NRS dynamically allocates paths and services, which minimizes the settings or code that need to be modified when the composition of the robot team changes.

We have even replaced robots with different robot models – in one field trial we had some hardware problems with a Robovie-R3 robot, and we were able to seamlessly replace it with a Robovie-R2 (a robot with a very different design) for an important demonstration. This was possible because differences in hardware components and internal implementations of gestures and poses are hidden beneath the abstraction layer of "service tasks," enabling the different robots to operate interchangeably within the network robot system.

The addition of new sensor types is also facilitated by our modular design. We have developed a new version of our human-tracking system using RGBD sensors [3]. Although developed by an independent team, such alternate sensor systems can be seamlessly used with our robots if they support the data protocols in our NRS framework design. This flexibility has been extremely helpful in managing the complexity of heterogeneous robot deployments in multiple environments with different sensor systems.

Finally, the use of the Map Registry makes it possible to easily switch environments. We often move robots between our lab and various field trial environments, and the local NRS at each environment enables the robots to automatically make use of the latest navigation maps and receive path planning and service allocation for that environment.

5.6 Conclusions

We have presented a framework for a Network Robot System, in which mobile robots, planning servers, and sensors embedded in an environment are integrated to provide robot services to people in social contexts. The requirements for this framework, motivated by our experiences in several years of field trials, primarily include the need for recognition and anticipation of people's behavior, identification of individuals, coordination of services and navigation paths between robots, and supervision by a human operator.

We presented a field experiment showcasing the capabilities of this framework by providing services with four robots in a shopping mall, and our results showed that not only was the technical framework successful in supporting the robot services, but that people who used the robots responded in a positive way, with a great majority indicating that they would like to use services like these in the future. This underscores the worth of conducting research and field experiments to investigate and develop social robot services in real-world environments, and we submit that the NRS approach is an effective and practical way to make such robot deployments a reality.

6 Cloud Networked Robotics

Although the NRS architecture has proven to be tremendously useful even with today's technology, the future promises many exciting possibilities for extending this approach.

In particular, the trend towards using cloud-based resources is becoming a popular new direction in robotics, and we believe it represents the natural next step in evolution of the NRS. The technologies of web services and service-oriented architecture (SOA), which form the technical foundation of cloud computing, have also been applied to robotic technologies in three ways.

6.1 Cloud Computing

One of the performance bottlenecks in robotics has always been recognition. Common sensory recognition tasks such as mapping and face and speech recognition require heavy processing and have long been the focus of optimization efforts. Tasks like these are excellent candidates for moving the processing load to networked servers, and eventually remote server farms.

The field of social robotics presents a wide variety of difficult processing tasks - recognition of social relationships and situations, recognition of human intention, and decisions as to how to best interact with people to provide effective services are complex tasks which could perhaps best be relegated to the cloud.

6.2 Sharing Data

Great benefits could be recognized from establishing shared data stores enabling social behavior generation or recognition. Online data exchanges such as RoboEarth[4] are already being developed as a means to provide cloud-based knowledge support for networked robots [38]. We have already conducted a proof-of-concept study demonstrating the integration of the UNR Platform and RoboEarth, combining RoboEarth's cloud-based support for robot recognition and action scripts with the platform-independent service execution architecture provided by the UNR-PF [35].

On the other hand, some knowledge stored in the information registries could be highly proprietary to the owners of a NRS, whereas other knowledge, such as map data, might be shareable or even outsourced to external services. Careful consideration will need to be given to levels of privacy and ownership of information when, for example, one organization licenses robots to multiple businesses.

Recently, several notions of cloud computing have been introduced into robotics that are known as cloud-enabled or cloud robotics.

[4] RoboEarth. `http://www.roboearth.org`

6.3 Service Continuity

Another approach utilizes robotic resources as a cloud to solve the issue of continuous support in robotic services. Since robotic services and robotic components are considered services in SOA, they can cooperate with each other if they are organized appropriately. Du et al. [40] introduced the concept of Robot as a Service and the framework of a Robot Cloud Center. Quintas et al. [23] proposed a service robotic system in which a group of robots and a smart-room share acquired knowledge over an SOA. The above projects rely on both *de facto* and *de jure* standards in the fields of networks, web service, knowledge representation for utilizing the technologies in SOA, and cloud computing. To realize cloud networked robotics, common protocols for robotic services must also be standardized for integration.

7 Conclusion

In this work, we have explored many aspects of the problem of how to enable the deployment of social robots in real-world environments. The essential points that have been presented are as follows:

- The use of external sensors to provide robust, high-precision tracking of people in the environment to assist robots in navigational interactions.
- Development of empirical models of social behavior in a space, to enable anticipation of people's future behavior.
- The use of a human operator to assist a team of robots with difficult recognition problems and unexpected situations.
- A framework for robot coordination, service allocation, and knowledge sharing to support the operation of heterogeneous teams of service robots.
- A set of global standards for a robot service architecture enabling independent application development for a networked ecology of of robots, sensors, data resources, and mobile devices.

We have demonstrated stable implementations of each of the systems presented in this work through a series of robot experiments and long-term deployments in several field environments over a period of about five years. The effectiveness of these systems has been demonstrated in the field, and the field experiences have contributed directly to the direction of this research. All of the systems presented here are still in use as framework elements supporting a variety of research in human-robot interaction.

Thus, this work provides a successful, concrete example of a coherent collection of systems for human tracking, behavior analysis and anticipation, supervisory teleoperation, interaction design, and robot service coordination, which together enable the practical deployment of teams of social robots to provide services in real-world environments.

It is our sincere hope that through our global standardization work for RoIS and the UNR-PF, we can share with the world the practical knowhow gained through our years of experimentation, and that this contribution will help to bootstrap and accelerate the development of commercial applications for service robots as we enter the fledgling years of a new cloud robotics era.

Finally, beyond the engineering solutions and system design, the development of this system has provided a deeper and more significant contribution - it has enabled us to obtain an early glimpse of what the world will be like when robots eventually work in real social environments.

Without actual, real-world deployment of robots, modeling and laboratory experiments can only take us so far – we can only speculate about how people may theoretically interact with robots and use their services in the future. But the development and deployment of this system has provided a way to truly observe and study real human-robot interactions in social environments, opening the door to a wide variety of academic studies and commercial innovations.

Acknowledgments. We would like to thank Dr. Masahiro Shiomi and Dr. Satoru Satake for their contributions in designing and building this framework, Dr. Satoshi Koizumi for his help in the organization and realization of the field trials, and the many students, interns, engineers, researchers, and assistants who have contributed directly and indirectly to the development of the Network Robot System over the years. We would also like to thank the management and staff of Universal CityWalk Osaka and APiTA Town Keihanna, without whose cooperation this work would not have been possible.

This work was partially supported by the Japanese Ministry of Internal Affairs and Communications, the Grant-in Aid for Scientific Research (A), Kakenhi, under Grant 21680022, the Program for Promoting the Establishment of Strategic Research Centers, and the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Ball, G., Hall, D.: Isodata, a novel method of data analysis and pattern classification. Technical report, DTIC Document (1965)

[2] Brooks, A., Williams, S.: Tracking people with networks of heterogeneous sensors. In: Australasian Conference on Robotics and Automation (ACRA 2003), Brisbane QLD, Australia, December 1-3, pp. 1–7 (2003)

[3] Brščić, D., Kanda, T., Ikeda, T., Miyashita, T.: Person tracking in large public spaces using 3-d range sensors. IEEE Transactions on Human-Machine Systems PP(99), 1–13 (2013)

[4] Burgard, W., Cremers, A.B., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., Steiner, W., Thrun, S.: The interactive museum tour-guide robot. In: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence, Madison, Wisconsin, United States, pp. 11–18. American Association for Artificial Intelligence (1998)

[5] Cui, J., Zhao, H., Shibasaki, R.: Fusion of detection and matching based approaches for laser based multiple people tracking. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, USA, pp. 642–649 (2006)
[6] Garrell, A., Sanfeliu, A.: Model validation: Robot behavior in people guidance mission using dtm model and estimation of human motion behavior. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5836–5841 (2010)
[7] Glas, D.F., Kanda, T., Ishiguro, H., Hagita, N.: Simultaneous people tracking and localization for social robots using external laser range finders. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, St. Louis, MO, USA, pp. 846–853. IEEE Press (2009)
[8] Glas, D.F., Kanda, T., Ishiguro, H., Hagita, N.: Teleoperation of multiple social robots. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 42(3), 530–544 (2012)
[9] Glas, D.F., Miyashita, T., Ishiguro, H., Hagita, N.: Laser-based tracking of human position and orientation using parametric shape modeling. Advanced Robotics 23(4), 405–428 (2009)
[10] Glas, D.F., Satake, S., Ferreri, F., Kanda, T., Ishiguro, H., Hagita, N.: The network robot system: Enabling social human-robot interaction in public spaces. Journal of Human-Robot Interaction 1(2), 5–32 (2012)
[11] Gockley, R., Forlizzi, J., Simmons, R.: Interactions with a moody robot. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, Salt Lake City, Utah, USA, pp. 186–193. ACM (2006)
[12] Hayashi, K., Shiomi, M., Kanda, T., Hagita, N.: Friendly patrolling: A model of natural encounters. In: Robotics: Science and Systems VII, Los Angeles, CA. MIT Press (MA) (2012)
[13] Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N.: Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets?. In: Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, pp. 449–456. ACM (2011)
[14] Kamei, K., Nishio, S., Hagita, N., Sato, M.: Cloud networked robotics. IEEE Network 26(3), 28–34 (2012)
[15] Kanda, T., Glas, D.F., Shiomi, M., Hagita, N.: Abstracting people's trajectories for social robots to proactively approach customers. IEEE Transactions on Robotics 25(6), 1382–1396 (2009)
[16] Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive robots as social partners and peer tutors for children: A field trial. Human-Computer Interaction 19(1), 61–84 (2004)
[17] Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Development and evaluation of interactive humanoid robots. Proceedings of the IEEE 92(11), 1839–1850 (2004)
[18] Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., Hagita, N.: An affective guide robot in a shopping mall. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, La Jolla, California, USA, pp. 173–180. ACM (2009)
[19] Kanda, T., Shiomi, M., Perrin, L., Nomura, T., Ishiguro, H., Hagita, N.: Analysis of people trajectories with ubiquitous sensors in a science museum. In: 2007 IEEE International Conference on Robotics and Automation, pp. 4846–4853. IEEE (2007)
[20] Kawai, H., Toda, T., Ni, J., Tsuzaki, M., Tokuda, K.: Ximera: A new tts from atr based on corpus-based technologies. In: Fifth ISCA Workshop on Speech Synthesis (2004)

[21] Mazzolai, B., Mattoli, V., Laschi, C., Salvini, P., Ferri, G., Ciaravella, G., Dario, P.: Networked and cooperating robots for urban hygiene: The eu funded dustbot project. In: Proceedings of the 5th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI (2008)
[22] Mutlu, B., Forlizzi, J.: Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, Amsterdam, The Netherlands, pp. 287–294. ACM (2008)
[23] Quintas, J., Menezes, P., Dias, J.: Cloud robotics: Towards context aware robotic networks. In: Proc. of the 16th IASTED International Conference on Robotics, Pittsburgh, USA
[24] Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 26(1), 43–49 (1978)
[25] Sanfeliu, A., Hagita, N., Saffiotti, A.: Special issue: Network robot systems. Robotics and Autonomous Systems 56(10), 791–791 (2008)
[26] Satake, S., Kanda, T., Glas, D.F., Imai, M., Ishiguro, H., Hagita, N.: How to approach humans?: Strategies for social robots to initiate interaction. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, La Jolla, California, USA, pp. 109–116. ACM (2009)
[27] Schulz, D., Burgard, W., Fox, D., Cremens, A.B.: People tracking with mobile robots using sample-based joint probabilistic data association filters. International Journal of Robotics Research (IJRR) 22(2), 99–116 (2003)
[28] Shiomi, M., Kanda, T., Glas, D.F., Satake, S., Ishiguro, H., Hagita, N.: Field trial of networked social robots in a shopping mall. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, St. Louis, MO, USA, pp. 2846–2853. IEEE Press (2009)
[29] Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N.: Interactive humanoid robots for a science museum. IEEE Intelligent Systems 22(2), 25–32 (2007)
[30] Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., Hagita, N.: How quickly should communication robots respond? In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, Amsterdam, The Netherlands, pp. 153–160. ACM (2008)
[31] Siegwart, R., Arras, K.O., 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.: Robox at expo.02: A large-scale installation of personal robots. Robotics and Autonomous Systems 42(3), 203–222 (2003)
[32] Sisbot, E.A., Alami, R., Simeon, T., Dautenhahn, K., Walters, M., Woods, S.: Navigation in the presence of humans. In: 2005 5th IEEE-RAS International Conference on Humanoid Robots, pp. 181–188 (2005)
[33] Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N., Anzai, Y.: Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. Connection Science 18(4), 379–402 (2006)
[34] Takayama, L., Marder-Eppstein, E., Harris, H., Beer, J.M.: Assisted driving of a mobile remote presence system: System design and controlled user evaluation. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 1883–1889 (2011)
[35] Tenorth, M., Kamei, K., Satake, S., Miyashita, T., Hagita, N.: Building knowledge-enabled cloud robotics applications using the ubiquitous network robot platform. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013), pp. 5716–5721 (2013)

[36] Tomizawa, T., Ohya, A., Yuta, S.: Remote shopping robot system - development of a hand mechanism for grasping fresh foods in a supermarket. In: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4953–4958 (2006)
[37] Villagrasa, A.M.: People tracking for a personal robot. Master's thesis, Royal Institute of Technology, Stockholm, Sweden (August 2005)
[38] Waibel, M., Beetz, M., Civera, J., D'Andrea, R., Elfring, J., Galvez-Lopez, D., Haussermann, K., Janssen, R., Montiel, J., Perzylo, A.: Roboearth. IEEE Robotics and Automation Magazine 18(2), 69–82 (2011)
[39] Zhao, H., Shibasaki, R.: A novel system for tracking pedestrians using multiple singlerow laser-range scanners. IEEE Transactions on Systems, Man, and Cybernetics 35(2), 283–291 (2005)
[40] Zhihui, D., Weiqiang, Y., Yinong, C., Xin, S., Xiaoying, W., Chen, X.: Design of a robot cloud center. In: 2011 10th International Symposium on Autonomous Decentralized Systems (ISADS), pp. 269–275 (2011)