Abstracting People's Trajectories for Social Robots to Proactively Approach Customers

Takayuki Kanda, Dylan F. Glas, Masahiro Shiomi, and Norihiro Hagita
IEEE Transactions on Robotics, Vol. 25, No. 6, pp. 1382-1396, Dec. 2009

Abstracting People's Trajectories for Social Robots to Proactively Approach Customers

Takayuki Kanda, Dylan F. Glas, Masahiro Shiomi, and Norihiro Hagita

Abstract**—**For a robot providing services to people in a public space such as a shopping mall, it is important to distinguish potential customers, such as window-shoppers, from other people, such as busy commuters. In this paper, we present a series of abstraction techniques for people's trajectories, and we present a service framework for using these techniques in a social robot, which enables a designer to make the robot proactively approach customers by only providing information about target local behavior. We placed a ubiquitous sensor network consisting of six laser range finders in a shopping arcade. The system tracks people's positions as well as their local behaviors such as fast walking, idle walking, wandering, or stopping. We accumulated people's trajectories for a week, applying a clustering technique to the accumulated trajectories to extract information about the use of space and people's typical global behaviors. This information enables the robot to target its services to people who are walking idly or stopping. The robot anticipates both the areas in which people are likely to perform these behaviors, and also the probable local behaviors of individuals a few seconds in the future. In a field experiment we demonstrate that this service framework enables the robot to serve people efficiently.

Index Terms **— Networked robot, Ubiquitous robot, Behavior anticipation, Social human-robot interaction**

I. Introduction

We believe that the robot can be a powerful device for bridging the gap between the digital and physical worlds. Since robots are mobile and embodied, they are well-suited for presenting digital information in the physical world. Previous studies have demonstrated that social robots can be used as museum guides [3, 4], as receptionists for assisting visitors [5], and as peer-tutors in schools [6].

On the other hand, robots have only weak sensing capabilities, which limited these robots to waiting for visitors to initiate interactions. Since we aim to realize a robot that proactively provides services in public spaces, it needs reliable observations of the positions and motion of people. However, a robot using onboard sensors can usually recognize people only within a few meters, and its sensing is not robust. To overcome these limitations, we use a "network robot system" approach [7], in which a robot is supported by a ubiquitous sensor network which observes and interprets information about people. Such an approach combines the stability and wide-area sensing capability of a ubiquitous sensor network with the intuitive presentation capabilities of the robot.

This paper describes a service framework for a network robot system, in which a mobile humanoid robot proactively approaches customers to provide information. It consists of 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 and behavior-based anticipation. The detection of local behaviors and analysis of the use of space can be valuable in anticipating where behaviors are statistically likely to occur, i.e. location-based anticipation; however, an analysis of global behavior patterns is far more powerful for predicting *individual* behavior, i.e. behavior-based anticipation. As people using the space have a variety of goals, an understanding of global behavior is essential in enabling the robot to anticipate the future behaviors of individuals.

Moreover, one of the notable features of the service framework is that a designer needs only to specify a target local behavior in order to make a robot proactively approach customers. The effectiveness of the service framework is demonstrated with a field trial with two examples of applications: one is for entertainment, and another is to invite customers to a shop.

Manuscript received February 17, 2009. This research was in part supported by the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government, and in part supported by Grant-in Aid for Scientific Research (A), KAKENHI (21680022). This paper is an extended version of conference paper [1, 2], with integrated technical details, and additional discussions and explanations.

T. Kanda, D. F. Glas, M. Shiomi, and N. Hagita are with ATR Intelligent Robotics and Communication Labs., Kyoto, Japan (corresponding author's phone: +81-774-95-1424; fax: +81-774-95-1408; e-mail: kanda@atr.jp).

Table 1: Related studies that concerns position, place, and the context with positions

	Recognition				Service		
	Local behavior	The use of space	Global behavior	Anticipation	Frame work	Human input required for context design?	domain
Reality Mining [15]	✔					---	Personal (city)
Liao et al.[16],	✔					---	Personal (city)
Subramanya et al.[17]	✔					---	Personal (city)
Suzuki et al. [21]	✔	✔				---	Public (shop)
Shao et al. [10]	✔	✔				---	Public (station)
Nurmi et al. [22]		✔				---	Personal (city)
Aipperspach et al. [23]		✔				---	Personal (home)
Activity zone [24]	✔	*1				Required	Personal (home)
Museum wearable [28]		*1	*1			Required	Public (museum)
Pre-destination [29]			✔	✔		---	Personal (car)
Bennewitz et al. [36]				✔		---	Public (corridor)
This study	✔	✔	✔	✔	✔	Not required	Public (mall)

II. Related Works

This section provides a survey of previous studies regarding these three concepts: local behavior, use of space, and global behavior. Table 1 provides a summary of this survey.

A. *Position and Local Behaviors*

People's positions and trajectories have frequently been studied in robotics and computer vision (for example, [8, 9, 10]). In ubiquitous computing, positioning devices are often used, such as GPS, or the signal strength of radio (GSM, WiFi, Bluetooth, RFID, and power line) [11, 12, 13, 14].

Ubiquitous computing technology is increasingly being used to identify people’s local behavior as well. For example, Eagle and Pentland developed a Bluetooth-based device attached to a mobile phone that enables the analysis of activities such as being at home, at the office, or elsewhere [15]. Liao *et al.* also used locations obtained via GPS with a relational Markov model to discriminate location-based activities such as being at home, at the office, and out dining [16]. Subramanya *et al.* included motion states (such as stop, walk, run) and velocity into a model to estimate people’s low-level activity and spatial context [17].

These techniques all used wearable or mobile personal devices. Our focus is on applications in an anonymous public space, so we chose a method independent of such devices. We measure walking motion using laser range finders, sensors often used in robotics due to their precision, simplicity, and non-invasiveness. A number of techniques exist for tracking people using multiple laser range finders [9, 10, and 18].

B. *The Use of Space*

Humans’ spatial behavior has attracted scientific interest for a long time. In the 1970’s and 1980’s, a technique named “space syntax" was developed to analyze town-level use of space with pre-defined logic [19]. People’s route choice and a form of trail were modeled as “active walker models” [20].

Such early studies required labor-intensive effort to collect data, which limit them to reveal only broad patterns; however, recent sensing technologies enable us to automatically accumulate large amount of trajectories with precise accuracy. Previous studies revealed that trajectories enable the identification of pausing points [21] and traffic paths [10, 21].

Information on the general use of space has also been retrieved. Nurmi *et al.* applied a spectral clustering method for identifying meaningful places [22]. Aipperspach *et al.* applied clustering to UWB sensor data to identify typical places in the home [23]. Koile *et al.* conducted a clustering of spaces with a focus on the relationships between velocity and positions, which enabled a partitioning of space into “activity zones.” For example, places for walking, working, and resting were separated [24]. Our work involves partitioning space in a similar manner, but based on position and local behavior. In addition, we also consider how the distribution of these zones varies as a function of time.

C. *Global Behavior*

Models of human walking have been developed for transportation engineering and architectural design. These models are usually concerned with how environmental information affects people’s behavior, such as a line of sight toward environmental structures [25] and movement of individuals in a crowd [26]. Positioning techniques could contribute to these models by providing automated, accurate position information.

In previous studies, positioning techniques have been used for categorizing people, and estimating people’s goals and intentions [27]. In a museum context, Sparacino developed the “museum wearable,” where people were classified into three visiting patterns. Depending upon the pattern, the system adjusted the way it presented information [28]. This is a good example of the use of global behavior; however, the places and the model of global behaviors were carefully prepared by a human designer.

In contrast, we have applied a clustering technique to identify typical visiting patterns in a museum without providing any environmental information [1]. One of the novel points of our current work is that the designer of the system provides information only about the *target local behavior*, with no knowledge about the structure of the space or of people's global behaviors. In addition to the previous work, this paper provides a method of online estimation of global behavior, which is indispensable for providing services.

The online estimation of global behaviors is difficult as, by definition, any global behavior being observed in real time is unfinished and thus not completely observable. Thus, it is necessary to estimate the true global behavior from a limited data set. Krumm *et al.* developed a technique they call "Predestination", which enables someone's driving destination to be estimated [29]. Liao *et al.* developed a technique for a person wearing GPS to infer her destination, transportation mode, and anomalous behavior [30].

While personal history of previous destinations was an important part of those studies, our anticipation technique for the shopping arcade assumes zero knowledge of a given person's individual history. Our technique is predicated on our observations of tens of thousands of people and the expectation that a new person's global behavior will be similar to those previously observed.

The concept of behavior anticipation is not without precedent in robotics. For example, Hoffman *et al.* demonstrated the value of anticipatory action in human-robot collaboration [31]. However, our use of global behaviors is a unique approach to behavior anticipation in this field.

D. Human-Robot Interaction

In the field of human-robot interaction, there have been several studies about mobile robots that provide services to people. For example, Dautenhahn *et al.* studied the appropriate behavior of a robot when it approaches a person, and found that the robot should approach people from the side but not the front [32]. Gockley *et al.* developed a natural way for a robot to follow a person [33]. Michalowski *et al.* observed how people approach a robot, and changed the robot's behavior according to their approaching style [34]. Yamaoka *et al.* established a model for a robot to appropriately position itself to effectively explain exhibits [35]. Bennewitz *et al.* developed a technique for predicting trajectories of persons for avoiding persons around it [36]. The need for this is apparently due to a lack of observation capability, which is solved in our study by having laser range finders distributed in the environment.

III. Recognition System

Figure 1 shows the service framework presented in this paper. This section explains the details of the recognition system.

A. Position

We conducted our experiments in a popular entertainment and shopping arcade located by the entrance to Universal Studios Japan, a major theme park. We operated the robot within a 20 m section of the arcade, with shops selling clothing and accessories on one side and an open balcony on the other. The motion of people through this area was monitored using a ubiquitous sensor network consisting of six SICK LMS-200 laser range finders mounted around the perimeter of the trial area at a height of 85 cm (Figure 2 and 3).

A particle filtering technique was used to track people's trajectories through this space. The location of each person in the scan area was calculated based on the combined torso-level scan data from the laser range finders.

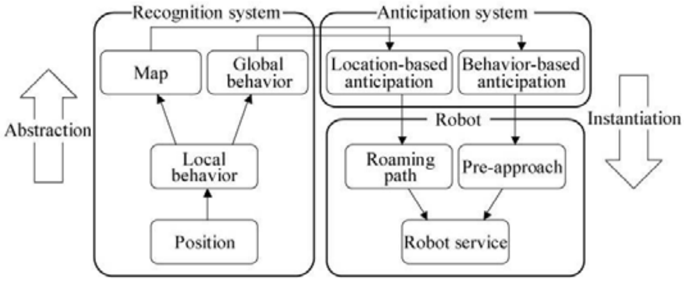


Figure 1. Service framework

Figure 2. The shopping arcade and laser range finders.

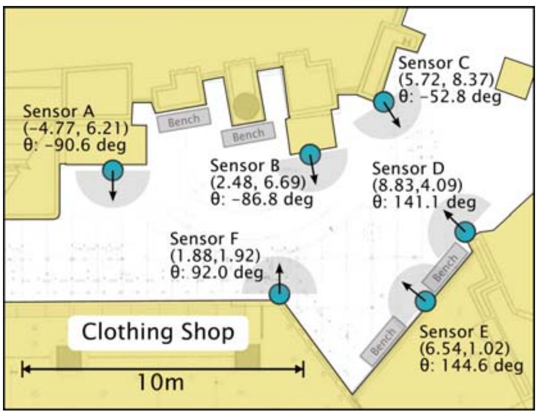


Figure 3. Placement of laser range finders

In our tracking algorithm, a background model is first computed for each sensor, by analyzing hundreds of scan frames to filter out noise and moving objects. Points detected in front of this background scan are grouped into segments, and segments within a certain size range persisting over several scans are registered as human detections.

Each person is then tracked with a particle filter, using a linear motion model with random perturbations. Likelihood is evaluated based on the potential occupancy of each particle's position (i.e. humans cannot occupy spaces which have been observed to be empty) as well as its proximity to observed points. By computing a weighted average across all of the particles, x-y position is calculated at a frequency of approximately 37 Hz. This tracking technique provides quite stable and reliable position data, with a position accuracy measured to be +/- 6 cm for our environment. Further details on this algorithm are presented in [18].

To illustrate the robustness of the system in our field environment, we analyzed two sets of data from one of the days in the middle of our experimental data set. For this analysis we considered only trajectories of at least 5 seconds in length, and each data set contained 100 trajectories. The first set was taken starting at 11:30am, a time when very few people were passing through the area, and the other was taken at 5pm, when the area was more crowded. The morning data set lasted 42 minutes, with an average trajectory length of 17.1 seconds, and the evening data set lasted 12 minutes, with an average trajectory length of 18.0 seconds.

For each of these data sets, the entry and exit times of each person passing through the space were identified manually by inspection of the raw laser scan data (this enabled more exact estimation of people's positions and entry times than inspection of video data). Any tracking errors during this period were also recorded, *e.g.* if a person entered the space and was not tracked, if two people were mistakenly switched with each other in mid-trajectory, or if a trajectory was lost in the middle of the space and reacquired with a different ID.

In fact, our system successfully tracked all people passing through the space in both cases. No tracking errors occurred, and no people entered the space without being tracked. However, the system did have some difficulty distinguishing couples walking close together. Couples were sometimes initially misinterpreted as a single person, but after a few seconds, the system always correctly identified them as two people. Since this phenomenon results in a short time lag before the system begins tracking the second person, we calculated the system's tracking success rate as the ratio between the total amount of time people were successfully tracked to the total amount of time people were present in the area. This ratio is presented in Figure 4 for the two 100-trajectory datasets.

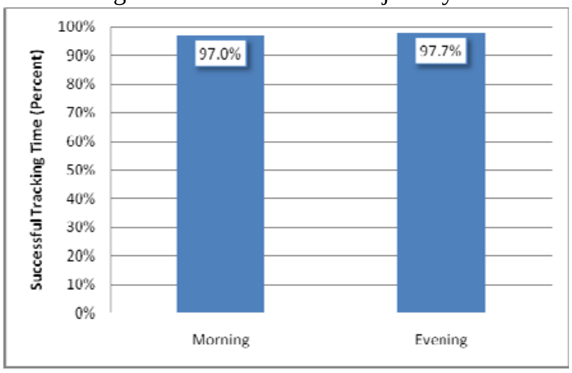


Figure 4. Tracking success rate for two datasets.

Based on this analysis, we consider our tracking system to be highly robust, particularly in terms of maintaining continuity of trajectories from beginning to end, an important requirement for the analysis we present in this paper.

B. *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.

Specifically, the following features were used for the SVM to classify the local behaviors:

(i) The end point of the normalized trajectory

Normalization refers to a rotation of the trajectory to fit its starting point to the origin and its longest direction to the x axis (Figure 5 (a) and (b)). Then, three points were sampled from the normalized trajectory: at times N/3, 2N/3, and N seconds, where N represents the duration of the trajectory. At each point, four dimensions of features were retrieved: x-coordinate, y-coordinate, arc tangent of this x-y position, and the distance of this x-y position from origin. Overall, 12 dimensions of features were retrieved.

(ii) The size of rectangle that covers the normalized trajectory

We retrieved the max value, min, and average value of x-coordinate and y-coordinate among all of the points sampled per 100ms in the N seconds of the trajectory. Overall, 6 dimensions of features were retrieved.

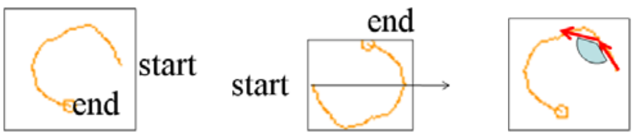
(iii) The angles of the trajectory

As shown in the Figure 5 (c), we calculated a sub-angle in a trajectory. For this calculation, the trajectory was separated into three sub-trajectories, at time "0 to N/3", "N/3 to 2N/3", and "2N/3 to N" seconds. For each sub-trajectory, the angle between start and end point was calculated. In addition, we also calculate the maximum angle as well as deviation of the angles among each sub-trajectory, within a sliding 500ms-window for each 100 ms from the start to end of the sub-trajectory. Overall, 9 dimensions of features were retrieved.

(iv) The velocity

For each 100ms interval, an immediate "sub-velocity" was calculated. The average, min, max, and variance of the sub-velocities were used as features. In addition, travel efficiency was computed by calculating the overall velocity from the start point to the end point, and dividing this by the sum of all sub-velocities. (It is nearly 1.0 if the trajectory moves straight, and nearly 0.0 if it only oscillated at the same point). Overall, 5 dimensions of features were retrieved.

In total there were 32 features. All of the features are float values and scaled within the range of 0 to 1. The SVM for the *Style* category uses all of the 32 features of (i) to (iv), while the SVM for the *Speed* category uses the features of (i), (ii), and (iv), the SVM for the *Short-term style* category uses the features of (i), (ii), and (iii), and the SVM for *Short-term speed* category use the features of (i) and (iv). Our SVM was implemented using LIBSVM [37]. The one-against-one method was used for multi-class classification [38]. For all SVM's, an RBF Kernel (Gaussian Radial Basis Function Kernel) was used.



(a) Original trajectory (b) Normalized trajectory (c) Sub-angle

Figure 5. Feature vector for calculating motion primitives.

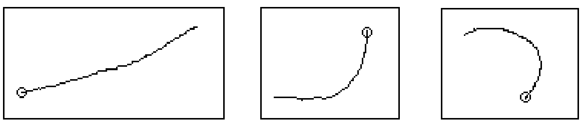
To include a wide variety of movement types, we initially defined the following four categories. Each category has about 200 samples for learning, consisting of 2- or 5-second trajectory segments. We selected typical trajectory segments that fit with the concept of each class of the categories, labeled them by hand, and put them into the training data set. We did not include trajectories which were ambiguous between classes.

(a) *Style*

This category consists of the following six classes. It requires 5.1 seconds of trajectory data for classification.

- straight (Figure 6 (a))
- left turn (Figure 6 (b))
- right turn (Figure 6 (c))
- wandering (Figure 6 (d))
- U-turn (Figure 6 (e))
- not walking (Figure 6 (f))

We labeled 226 trajectories and tested the system with the leaving-one-out method, a cross-validation-method where each of the data elements is tested by using the remaining elements for training; i.e. we created 226 subsets, each of which has one unique trajectory for testing and the remaining 225 for training, and averaged the classification accuracy of the 226 subsets. It classified with 84.5% accuracy on average. The confusion matrix shows relatively-frequent confusion between U-turn and wandering, recognizing U-turn with 79.4% accuracy and wandering with 76.6% accuracy.



(a) Straight (b) Left turn (c) Right turn
(d) Wandering (e) U-turn (f) Stop

Figure 6: Examples of *Style* category

(b) *Speed*

This category consists of the following five classes. It requires 4.9 seconds of trajectory data for classification.

- running
- fast-walk (Figure 7 (a))
- idle-walk (Figure 7 (b))
- stop : short stop is observed in a trajectory while some movements are also observed
- wait : only stopping, but no motion observed

In the the labeling, we judged the difference between "idle walk" and "fast walk" based on the speed of the trajectory. The difference between “stop” and “wait” is defined by whether the trajectory remains stopped for the full duration or not.

We labeled 166 trajectories and tested the system with the leaving-one-out method; it classified with 92.8 % accuracy on average. The confusion matrix shows frequent confusion between stop and wait, recognizing stop with 66.7 % accuracy.

(c) *Short-term style*

This category is similar to (a) *Style*, but to enable faster recognition we reduced the duration required for the classification. It requires 2.1 seconds of trajectory data for classification.

- straight
- left turn
- right turn
- U-turn
- not walking

We labeled 150 trajectories and tested the system with the leaving-one-out method; it classified with 93.3 % accuracy on average. There was no particular confusion in the confusion matrix.

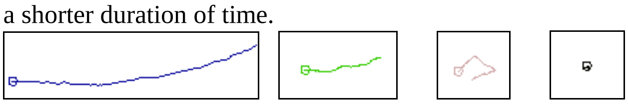
(d) *Short-term speed*

This category is similar to (b) *Speed*, but to enable faster recognition we reduced the duration required for the classification. It requires 2.2 seconds of trajectory data for classification.

- running
- fast-walk
- idle-walk
- stop

We labeled 159 trajectories and tested the system with the leaving-one-out method; it classified with 95.6 % accuracy on average. There was no particular confusion in the confusion matrix.

Note that each category requires different length of trajectories, which is the result of our minimization of the time to recognize each of these categories. For example, (c) *Short-term style* requires 2.1 seconds while (a) *Style* requires 5.1 seconds, since *Style* has a category of wandering which is confused with U-turn more if the duration is smaller than 5.1 seconds. Short-term style does not have the category of wandering, which make the system easily categorize even with a shorter duration of time.



(a) Fast-walk (b) Idle-walk (c)Wandering (d) Stop

Figure 7. Example trajectories for local behaviors

In the subsequent analysis, we merged several local behavior classes for simplicity. Within “Style”, the classes *left-turn*, *right-turn*, and *U-turn* were all merged into the *wandering* category. Within “Speed”, we merged *stop* and *wait* into the *stop* category. We also merged classes for short-duration and 5-second behavior. Thus, we reduced the set to the following four local behaviors: *fast-walk*, *idle-walk*, *wandering*, and *stop*. Figure 7 shows examples of these local behaviors. 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 [1]indicating the presence or absence of local behavioral primitives $P_{fast-walk}$, $P_{idle-walk}$, $P_{wandering}$, P_{stop}.

Each trajectory has a sequence of local behaviors represented by these Boolean variables at each time step *t*. The system split a segment of trajectory from the time step *t* to the past for the required length of each classifier, and sent it to the classifier. They remain undetermined if *t* is smaller than the minimum required time of SVMs, i.e. 2.1 seconds.

IV. Analysis of Accumulated Trajectories

Based on the position and local behavior data thus obtained, an analysis was performed 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.

A. Data Collection

Human motion data was 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 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]

B. 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 [39] 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.

The local behaviors represented by the Boolean variables are all mapped into the histogram prepared for each grid elements. Each grid element contains histogram data of local behaviors: $H_{fast-walk}(i,t)$, $H_{idle-walk}(i,t)$, $H_{wandering}(i,t)$, and $H_{stop}(i,t)$, where $H_x(i,t)$ denotes the number of occurrences of local behavior x at time slice t within grid element i, which is normalized for each local behavior x. Specifically, we normalized each histogram $H_x(i,t)$ to have a mean value of 0.0 and a standard deviation of 1.0.

To make the data set more manageable, we first combined time slices based on their similarity. The difference between time slices t_1 and t_2 is defined as:

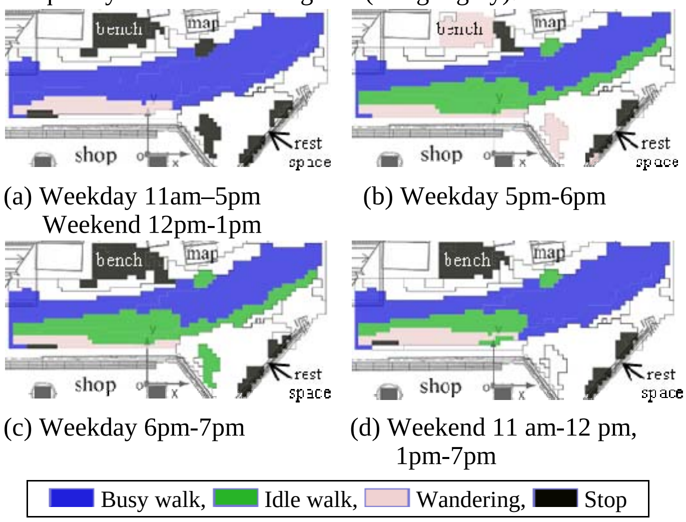
$$\sum_i \sum_x |H_x(i,t_1) - H_x(i,t_2)| \quad (1)$$

We then combined spatial grid cells where the distance was smallest and the grid was spatially connected. The distance between grid cells i and j is defined as:

$$\sum_t \sum_x |H_x(i,t) - H_x(j,t)| \quad (2)$$

As is usual for this type of explorative clustering, we arbitrarily set the number of partitions to help us intuitively understand the phenomena occurring in the environment. We chose to use 40 spatial partitions and 4 temporal partitions. Figure 8 shows a visualized output of the analysis. The partitions are color-coded according to the dominant local behavioral primitive in each area. Blue (medium gray on monochrome printouts) represents the areas where the *fast-walk* behavior occurred more frequently than any other local behaviors. Thus, people tend to pass directly through this area, which can be thought of as "corridor" space.

The areas where the *idle-walk* primitive occurred most frequently are colored with green (or light gray).



(a) Weekday 11am–5pm Weekend 12pm-1pm
(b) Weekday 5pm-6pm
(c) Weekday 6pm-7pm
(d) Weekend 11 am-12 pm, 1pm-7pm

Figure 8. Analysis of the use of space

In some areas, the use of space was very clearly observed to change as a function of time. The lower left area is in front of a shop. When the shopping arcade was busy in the evening, as in Figure 8 (b), with people coming back from the theme park, many people were observed to slow down in front of the shop, and the "corridor" space changed into "in front of shop" space with *idle-walk* becoming dominant (photo: Figure 9 (a)); however, when there were not so many people, such as midday during the week as in Figure 8 (a), these areas disappeared and became similar to other "corridor" space. The lower right side of the map represents the side of the corridor, where people tend to walk slowly when the arcade is busy (Figure 8 (b) and

[1] These Boolean variables allow each state to have a combination of fast-walk, idle, wander and stop. One 4-state variable might be appropriate depending on the purpose. For this study, our intention was to provide a local behavior classifier as capable as possible.

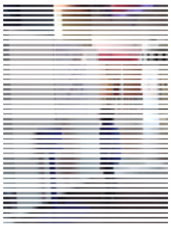
[2] In this study, we obtained approval from shopping mall administrators for this recording under the condition that the information collected would be carefully managed and only used for research purposes. The experimental protocol was reviewed and approved by our institutional review board.

(c)); these areas also disappeared and became similar to other "corridor" space (Figure 8 (a) and (d)).

(a) Idle-walk in front of a shop

(b) stop at a bench

(c) stop at rest space



(d) a map

Figure 9. Examples of the actual use of the space

The areas where the *stop* primitive was most frequent are colored with dark brown (or dark gray). In Figure 8, these areas can mainly be found in the upper center (photo: Figure 9 (b)) and the bottom right (photo: Figure 9 (c)). These areas contain benches, and can be considered "rest space".

In the upper center area, below the word 'map', there is a small space where *stop* is the dominant primitive in Figure 8 (a) whereas *idle-walk* is dominant in (b) through (d). A map of the shopping arcade is placed on that wall. Customers sometimes slowed down, stopped, and looked at this map (Figure 9 (d)). 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 8 show the space immediately in front of the shop as having this property. The areas where none of the primitives were dominant, such as the bottom-right space, are colored white. These areas were not used so much.

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.

C. Global Behavior

Based on the accumulated trajectories, we analyzed how people visited the shopping mall. In this section we introduce a method of extracting typical global behaviors.

1) Preparation: State chain models

We analyzed trajectories based on the *state chain* model illustrated in Figure 10. That is, we converted P_t^n, represented in x-y coordinates, to a sequence of states, $S^i = \{s^i_{t0}, s^i_{t1}, \cdots\}$ based on spatial partitioning. s^i_t is defined as, $s_t^i = \{n \in N \mid p_t^i \in A_n\}$, where A_n is the partition the point in trajectory p belongs to. In the example in Figure 10, the trajectory starting from partition 1, stayed in partition 1 for 3 time steps, then entered briefly into partition 2, and moved back to the partition 1 …, which is represented as the sequence of states 1, 1, 1, 2, 1, …

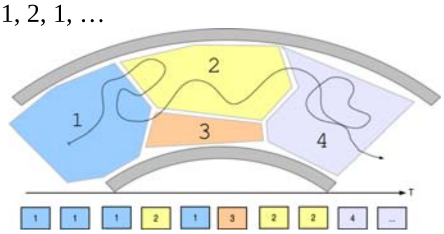
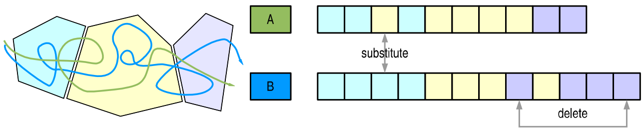


Figure 10: State chain model

2) Distance between trajectories

We calculate the distance between two state chains, $\boldsymbol{S^i}$ and $\boldsymbol{S^j}$, by using a DP matching method (widely used in many research domains, e.g. [40]), which is identical to the comparison of strings known as the Levenshtein distance. Figure 11 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.



(a) two trajectories (b) comparison of state chains of trajectories

Figure 11: Comparison of trajectories based on DP matching

For the DP matching, we again partitioned the space into a 25cm grid (2360 grid elements), to easily compare trajectories. The DP matching method was chosen for its simplicity and the fact that it does not require particular tuning of parameters. Since global behaviors naturally emerge through the interactions between people and their environment, we believe that it is best to minimize the number of parameters that need to be adjusted manually, keeping the process simple and generalizable.

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.

In addition, this state-chain representation reduces calculation cost. For example, we compared calculation cost based on raw trajectory $\boldsymbol{P^i}$ and state chain $\boldsymbol{S^i}$ for retrieving global behavior with a k-means clustering method from 28 trajectories. The state-chain method costs 0.53 sec while the raw-trajectory-based method costs 9.56 sec. Thus, using the state chain is eighteen times faster. We cached the calculation of distance between partitions in the state-chain-based method (that is, *insert*, *delete*, and *substitute* costs in DP matching), which also greatly improved the calculation speed.

3) Clustering and Visualization

We classified trajectories with a k-means method to identify typical visiting patterns. The distance between trajectories was provided from DP matching method mentioned above. We separated the space into 50 similarly-sized partitions by the k-means method [1] for this visualization, although the actual computation used 2360 partitions. We did not use these 2360 partitions or the result of analysis of the space shown in Figure 8 for the purpose of this visualization, since we are interested only in the transition pattern. K-means clustering of the space is one method which can provide similarly-sized polygonal spatial divisions distributed over the map with an arbitrary resolution, which are useful features for the visualization of global behavior.

Figure 12 shows a visualization of the global behaviors at k=6. 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. An arrow is drawn from partition i to j when ($N_{ij} - N_{ji}$) is larger than a threshold (here, set as 0.1) where N_{ij} indicates a transition from i to j.

Of course, we can analyze behavior patterns at any k value; a larger number k will result in more detailed separation of visiting patterns.

We can interpret about six typical global behaviors from Figure 12:

(a) Pass through from right to left (7768 people)
   This pattern represents one of the major flows of people, who are coming back from the theme park (on the right) on their way to the train station (on the left). In this pattern, most of the areas are colored blue because the most frequent primitive in those areas was *fast-walk*. In front of the shop, there are some areas colored green, which represent spaces where people slow down to look at the shop.

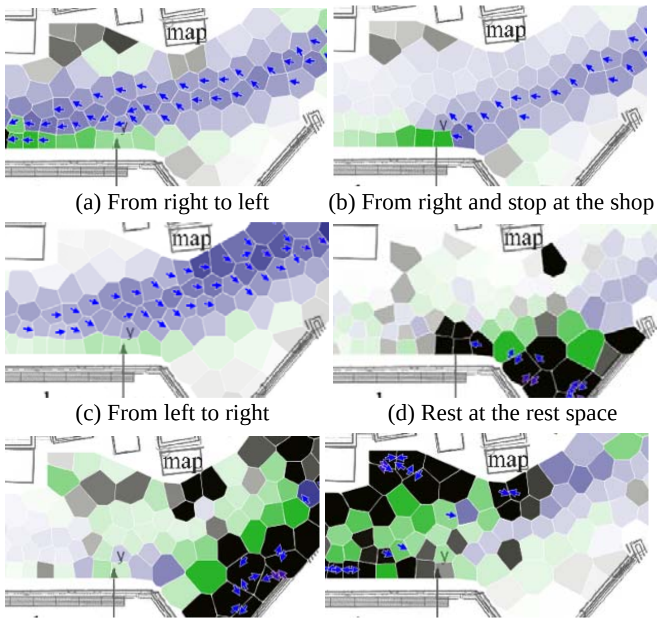
(b) Come from the right, and stop at the shop (6104 people)
   This pattern is similar to the pattern (a); but people either stop at the shop or go through the shop to go to the left area, as trajectories mostly disappeared at the shop.

(c) Pass through from left to right (7123 people)
   This is also a major pattern, where people are coming from the train station and going in the direction of the theme park. In contrast to the patterns in (a) and (b), people rarely stopped or slowed down in front of the shop.

(d) Rest at the rest space (213 people)
   In this pattern, people mostly spent time in the bottom right rest space (Figure 8 (c)) where benches were placed.

(e) Around the rest space and right (275 people)
   Similar to the pattern in (d), but people moved around the right area more, and not around the shop area. Some people also stopped in front of the map or the upper rest area.

(f) Around the shop and bench (334 people)
   People mainly came from the left side, walking slowly, and stopped in front of the shop as well as in front of the map.



(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

Busy walk, Idle walk, Wandering, Stop

Figure 12. Six typical patterns of global behavior

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.

V. Anticipation System

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.

A. Location-Based Anticipation

As shown in Figure 8, the system has use-of-space information about the frequency of the local behaviors associated with spatial and temporal partitions. The robot uses this information to estimate the locations in which people will be statistically likely to perform the target behavior. In addition, we assume that a moving robot would attract people's attention more than a robot standing still, which makes it easier for the robot to initiate interaction; thus, the system provides a path for the robot to roam around such locations, rather than choosing a single point at which to wait.

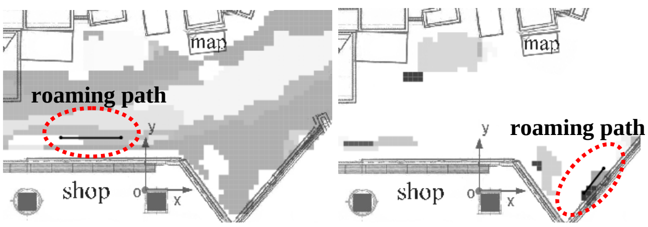
Figure 13 shows an example anticipation map. The darker areas represent areas where the system anticipates both a high likelihood of the target behavior and a low likelihood of the non-target behavior. In the graph, areas where the likelihood of the non-target behavior is higher than the likelihood of the target behavior are shown in white.

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

$$roaming\_value(\vec{P}_x, t) = \sum_{i \in P_x} (H_{target}(i,t) - H_{non-target}(i,t)) \qquad (3)$$

where $H_{target}(i,t)$ represents the histogram of the target behavior at the point grid *i* at time slice *t* (see IV B for the calculation to retrieve the histogram).

After finding the best path, the system modifies it according to safety considerations; the robot is constrained to operate within a safety buffer of two grid elements from the outside of observed area (these areas are too close to a wall for the robot to pass through), so the points of the path are translated to the nearest points within the safe area. The black line in Figure 13 represents its automatically-generated roaming path.



(a) Weekday11-17, *idle-walk* (b) Weekday 11-17, *stop*

Figure 13. Example of anticipation map

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 13 (a) is the anticipation map for this scenario, calculated for the behavior 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 13 (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.

B. 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.

To ensure prediction accuracy, we used a large number of clusters for the global behavior analysis. We clustered the human motion data collected earlier into 300 global behavior patterns. For this analysis, since we are interested in behaviors several seconds in the future, we only used trajectories observed for a sufficient amount of time. We filtered out trajectories less than 20 seconds long, leaving 11,063 trajectories for analysis.

Next, 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 trajectory 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.

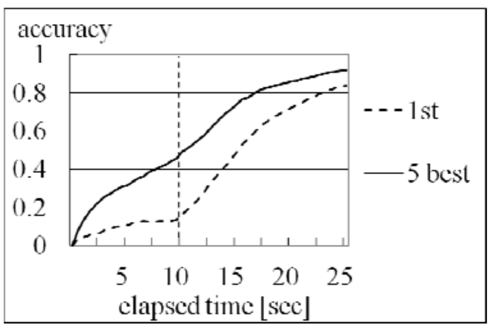


Figure 14. Accuracy of the prediction of global behavior

Figure 14 shows the prediction accuracy for observed trajectories from 0 to 25 seconds in length. Here, 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 with the one the trajectory belongs to, i.e. the classification result after observing the whole length of the trajectory. The accuracy accounts for only trajectories of total length greater than 20 seconds, as we filtered out shorter trajectories for calculating global behaviors. The result labeled “1st" represents the case where the best-fit global behavior at time *T* was the correct one (the cluster the trajectory finally fit with at completion). The result labeled “5 best” is the result if we define success to mean that correct global behavior falls within the top 5 results. 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.

After the most likely global behaviors are selected, the person's future position and local behavior are predicted based on an “expectation map." An expectation map is a data structure prepared *a priori* for each global behavior. For each 500-ms time step along the trajectories, a 25-cm grid representation of the observed space is added to the map. Each element of this grid contains likelihood values for each of the four local behaviors to occur in that location at any time *after* that time step. These likelihood values are empirically derived from the original observed trajectories falling within the chosen global behavior cluster, and they represent the average frequency of the occurrence of each local behavior after that time step. We used the 5-Best result to create an expectation map for the person by combining expectation maps from each of the 5-Best global behaviors.

Figure 15 shows expectation maps for various time increments. The solid circles represent the positions of people walking through the space, with the person of interest outlined in red. The expectation map for that person’s estimated global behavior is shown, where the area colored blue represents the area where *fast-walk* is expected, and the green area represents the area where *idle-walk* is expected. The three figures in the top row show the trajectory for person 1, who was first observed at time t_1. The first figure shows time $t_1 + 5$ sec, where the expected local behaviors can be seen tracing a path through the corridor, heading toward the upper right. In fact, this course was correctly predicted, and the person followed that general path. The second line is the trajectory for person 2, first observed at time t_2. Here, since the person walked slowly, it predicted the course to the left with *idle-walk* behavior. At time t_2+15, it started to predict the possibility of *stop* at the shop, which finally came to be true at time t_2+22. (See *multimedia attachment in IEEE Xplore* for more dynamic examples of successful prediction)

We measured the accuracy of position prediction for four time windows: 0-5, 5-10, 10-15, and 15-20 seconds in the future. Predictions were begun after a trajectory had been observed for 10 seconds, as the estimation of global behavior is not stable until then. We again used 6 days of data from the accumulated trajectories to predict the data of the remaining day. Our method predicts the future position as the center-of-mass of the expectation map. Figure 16 compares our method with position prediction based on the velocity over the last second. As the velocity method cannot account for motions like following the shape of the corridor, our method performs about twice as accurately.

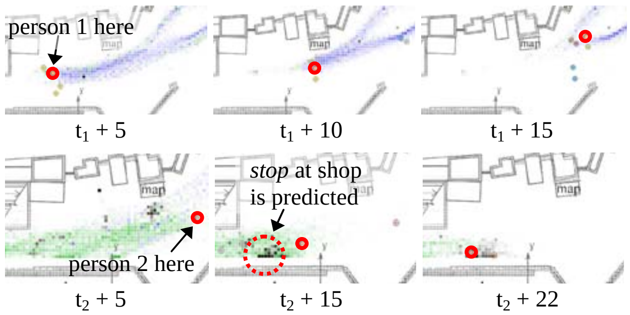


Figure 15. Example of prediction of future behaviors

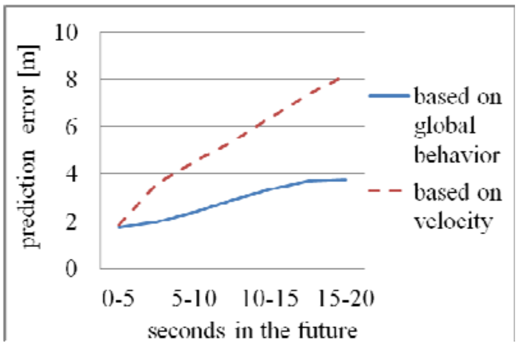


Figure 16. Prediction accuracy for position

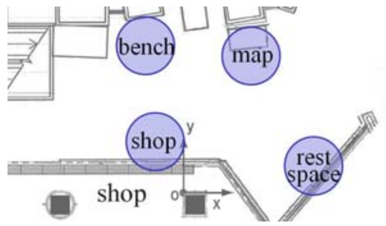


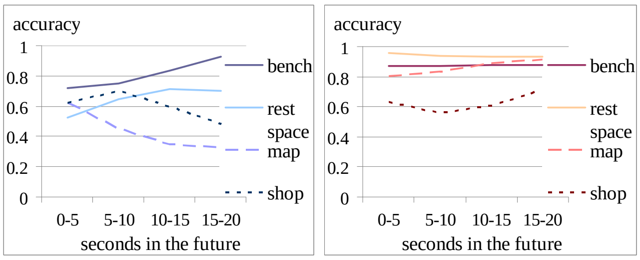
Figure 17. Places used for the measuring the performance

We then measured the correctness of the system’s predictions of the future positions and local behaviors for each person, evaluated in four places (indicated by three-meter circles in Figure 17) where qualitatively distinct behaviors were observed in the use-of-space analysis. For each place, at each moment, the system predicted whether the person would exhibit each of the local behaviors at that place for forecast windows of 0-5, 5-10, 10-15, and 15-20 seconds.

Figures 18 and 19 show the system’s prediction performance. In each figure, the left graph shows the accuracy of the prediction for the case where the target local behavior occurred at each place, and the right graph show the accuracy of the prediction where the behavior did not occur. We define the occurrence of the local behavior as the case where the person appeared at the place in the predicted 5-second window (*e.g.* between 5 sec and 10 sec), and performed the target local behavior more than other local behaviors. The accuracy value used for each person is the average across all predictions made for that person, and the value shown in the graph is the average across all people.

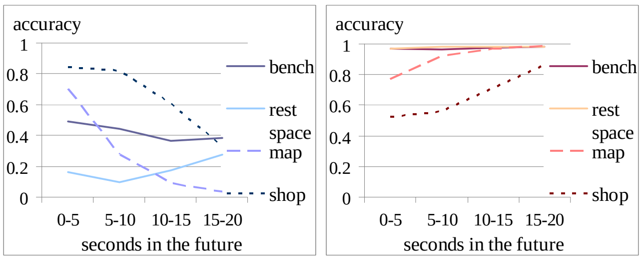
Figure 18 shows that the prediction was fairly accurate for the *stop* behavior, particularly at the bench and the rest space. Prediction was 92% accurate at the bench even for 15-20 seconds in the future, while non-occurrence was predicted with 88% accuracy. This good performance was due to the fact that people who stay in these areas often stay for a long time. Results were more marginal at the map and shop, with 62% accuracy for occurrence and 63% for non-occurrence predicted at the shop for 0-5 seconds in the future. For 15-20 seconds in the future, the performance is still marginal, with 48% accuracy for occurrence and 71% for non-occurrence predicted at the shop.

In contrast, as Figure 19 shows, the system predicted *idle-walk* with high accuracy 0-5 seconds ahead at the map and the shop. Even for 15-20 seconds ahead, the system was able to predict 33% of the occurrences at the shop as well as 86% of the non-occurrences, which we consider to be a good result, as it is rather difficult to predict walking behavior in the future. The prediction of occurrence was not successful at the rest space, as the system mostly predicted non-occurrence, since *idle-walk* rarely happened there.



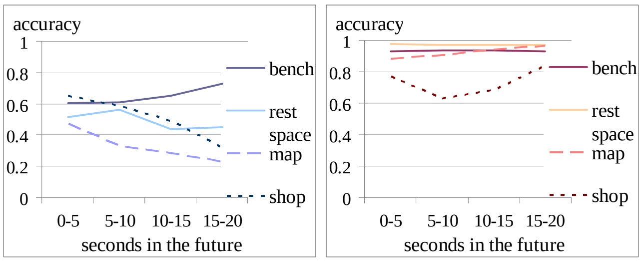
Behavior occurred at the place Behavior did not occur at the place

Figure 18. Prediction accuracy for *stop* behavior



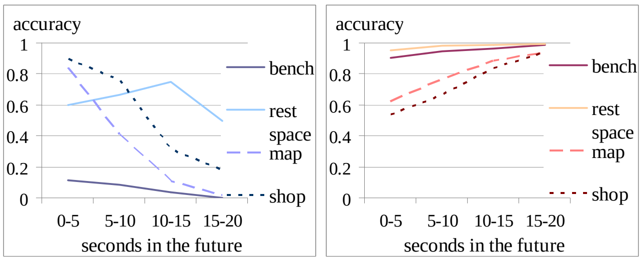
Behavior occurred at the place Behavior did not occur at the place

Figure 19. Prediction accuracy for *idle-walk* behavior



Behavior occurred at the place Behavior did not occur at the place

Figure 20. Prediction accuracy for *wandering* behavior



Behavior occurred at the place Behavior did not occur at the place

Figure 21. Prediction accuracy for *fast-walk* behavior

Regarding the remaining two behaviors, for *wandering* (Figure 20), the system predicted over 50% of occurrences and 85% of non-occurrences for 0-5 seconds ahead at all four places. For the 15-20 second window, it predicted 73% of occurrences and 93% of non-occurrences at the bench but not so well for the map and shop. It predicted *fast-walk* (Figure 21) at map and shop well until 10 seconds; for example, it predicted 86% of occurrences and 60% of non-occurrences at the shop for 5-10 seconds in the future, though it does not predict the future well beyond 10 seconds.

We believe these anticipation results are useful for the robot. The robot is designed to wait for people in areas where it anticipates frequent occurrence of the target behavior. Behavior-based anticipation performs particularly well in areas where the anticipated behaviors occur often, such as *stop* near the benches and rest space, and *idle-walk* in the corridor in front of the map and shop. As these are the areas predicted by the location-based anticipation method, the two anticipation techniques complement each other nicely.

VI. Service from a Social Robot

In this section, we show examples where a social robot provides services using our system. A human designer defines the contents of the service as well as the context in which the robot should provide the service. Here, the notable point is that the designer only specifies the target local behavior, such as “stopping”. The robot system then automatically computes the information about space and global behavior so that the robot can efficiently wait for people in promising areas, and then proactively approach people who are anticipated to perform the target local behavior.

For these services a robot has an advantage over cellular phones or other mobile devices, in that people do not need to carry any hardware; however, there is the additional challenge that robots need to approach the person quickly enough to start the service. For this purpose, anticipation plays an important role.

A. Robot Hardware

“Robovie” is an interactive humanoid robot characterized by its human-like physical expressions and its various sensors [41] (Figure 22). Robovie has a head, two arms, a body, and a wheeled mobile base. Its height and weight are 120 cm and 40 kg. The robot has the following degrees of freedom (DOFs): two for the wheels, three for its neck, and four for each arm. 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 sensor network via wireless LAN. We used a corpus-based speech synthesis [42] for generating speech.

B. Entertainment Application

The first example of an application that we would like to discuss is an entertainment robot, which interacts with people in the form of chatting. 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. In addition, 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.

The chat was about the attractions in the amusement park. For example, the robot says, “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.

We conducted a field trial to investigate the effectiveness of the system. Figure 23 is a scene where the robot is approaching a person who is “stopping”. Based on the anticipation mechanism and its current position, the robot set its roaming path near the bench and waited for a person to approach. When the robot predicted that a detected person would probably do the *stop* behavior, the robot began positioning itself near her general area (pre-approach) (Figure 23 (a)). When she came in front of the shop, she stopped (partly, we assume, because she was intending to stop regardless of the robot, and partly because she noticed the robot approaching her). Once she stopped, the robot approached her directly, and they had a chat (Figure 23 (b)). This is a typical pattern illustrating how people and the robot started to interact. Overall, people seemed to enjoy seeing a robot that approached them and spoke.

To evaluate the performance, we compared the situation with the developed system “with anticipation”, and “without anticipation”, and measured how much the anticipation mechanism improved the efficiency. In the “without anticipation” condition, the robot simply approached the nearest person who is doing the *stop* behavior. We measured the performance for one hour in total for each condition. We prepared several time slots and counter-balanced the order.

Figure 24 shows the number of people to whom the robot provided services. Due to the novelty of the robot, people often initiated interactions on their own; in such cases, the anticipation mechanism is irrelevant. Thus, we classified the robot’s interactions into two categories. The first case, “robot-initiated”, is the situation where the robot initiated the service by approaching the person and entering into conversation distance. Thus, the number of “robot initiated” services indicates how the robot’s anticipation system improved the efficiency of the service. The second case, “person-initiated”, is the situation where the person approached the robot while it was talking to someone else. Figure 25 shows one of such scenes. In this scene, when the robot was talking with the girls, a child came from the left. When the girls left, the child stood in front of the robot to start talking with it.

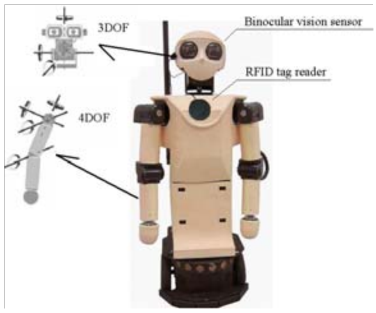
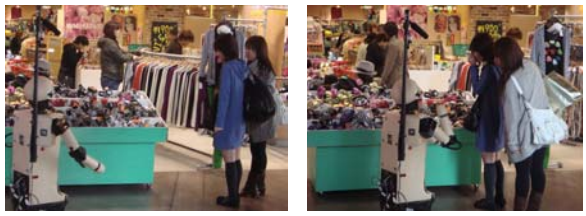


Figure 22. Robovie



(a) (b)

Figure 23. A robot approaching a person to chat

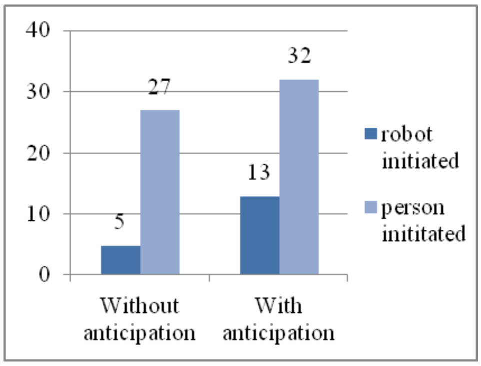


Figure 24. The number of services provided

(a) (b) (c)

Figure 25. A child initiates an interaction with the robot

The results in Figure 24 indicate that the number of “robot-initiated” services in “with anticipation” is much larger than “without anticipation.” In other words, anticipating enables the robot to provide the service more efficiently. Due to the novelty factor of the robot, the number of “person-initiated” services is quite large. We believe that in the future when robots are no longer so novel to people, there will be less person-initiated interaction, and the results concerning anticipation will become much more significant.

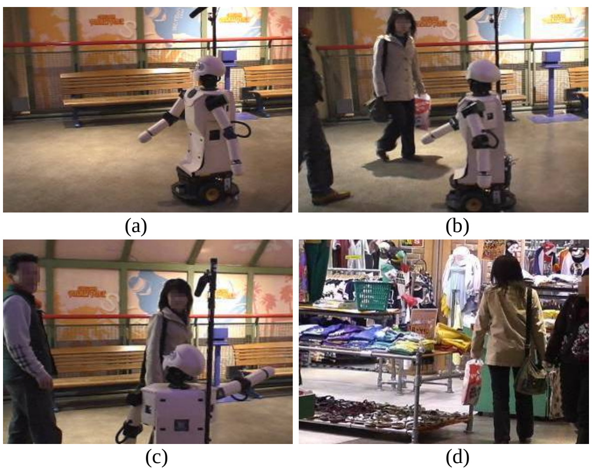
C. Invitation Application

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" [43].

We chose *idle-walk* as the target local behavior, because people who are walking slowly might be window-shopping. We set the non-target local behavior as *fast-walk*, so as not to bother people who seem uninterested in shopping. We used anticipation and the pre-approach function for the *idle-walk* behavior; when the robot predicted a person's future behavior as *idle-walk*, it moved towards that person's location.

We ran a field trial with the invitation robot in the shopping arcade as well. Just as in the entertainment application, the robot modified its behavior in accordance with the anticipation mechanism; the robot roamed around in front of the shop, where *idle-walk* was anticipated to be most likely, and approached people who were window-shopping.

In the demonstration, many people were interested in the robot and listened to its invitations. Figure 26 (see *multimedia attachment in IEEE Xplore* as well) 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)

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

VII. Discussion

A. Does the presence of the robot affect global behavior?

Our model is based on data recorded without having a robot in the environment. Thus, the system tried to predict people's behavior independent of the presence of the robot. However, as a robot is still a novel object, some people were attracted by the robot, slowed down, approached the robot, and even talked to the robot. In this case, the prediction cannot be correct, since such the behaviors are not in the model.

For the application shown in this paper, this had a positive effect on the robot's ability to provide the service. Even when the prediction from the robot was incorrect, as the robot approached, sometimes the person was nevertheless attracted by the presence of the robot, and stopped, which enabled the robot to provide its service.

For a different possible application such as a delivery task where the robot tries to avoid people in *idle-walk* and *stop*, however, this would affect the robot's ability negatively, as the robot's presence might attract a busy person to stop, and as a result the robot's route would be blocked. Thus, it will be useful to create a behavior model incorporating the effects of the robot.

B. To what extent is accuracy of positioning required?

In this study, we used a robust and accurate positioning technique with laser range finders; however, the whole approach does not depend on the positioning algorithm. In our previous work [1], we reported the analysis of global behavior where tracking was performed with RFID tags and readers, which provides people's position with 2.8 m error in an 80 x 40 m space. Like that example, our method is applicable for trajectories obtained through a different positioning technique. On the other hand, the classification of local behavior is based on some details of the position data. Thus, better positioning techniques will provide a better performance in local behavior classification.

One important characteristic of our positioning technique is robustness in terms of the continuity of the trajectory. Our method of analysis of global behavior requires that the whole length of the trajectories be observable. Thus, our method can be used with any tracking system that provides robust continuity of trajectories, even if it provides less positioning accuracy, e.g. our example with RFID tags and readers, but might be not feasible using a method without robustness in tracking.

C. Other possibilities of services with robots

Since we intended to highlight the connection between the robot and the infrastructure with ubiquitous sensors, we focused on the beginning part of the service (finding a person, approaching, and initiating conversation), and show two simple examples of services such a robot could provide. These services are appropriate under the situation where a robot is novel to people. Even such simple services provide enough value to people who are eager to experience an interaction with the robot.

As a future scenario, we can extend the service by having a designer in the context design. For example, many people stopped in front of the map, which can be seen in the analysis of the use of space; after discovering this fact, we can design a robot to provide guidance services for a person who is standing in front of the map.

D. Other Possible Applications

We believe that the infrastructure shown in the paper can be useful for other systems, *e.g.* ubiquitous computing applications. One possible direction is to apply it to ambient intelligent environments, in which facilities (robots, display, music, illumination, etc.) are proactively controlled according to the types of users. For instance, an electronic poster could anticipate who is likely to stop nearby, and change its advertisement content in advance to something targeted to that person.

Another possibility is to combine it with mobile devices. Although GPS and WiFi have been used for locating people, laser range finders can provide more accurate positioning. The information provided by the infrastructure developed here could also complement other location-based services. For instance, if a user with a mobile device providing pedestrian navigation information entered a space with this infrastructure available, the device could then present additional information appropriate to that user's anticipated global behavior.

E. Privacy Concerns

Systems operating in public spaces should be carefully designed to protect the privacy of people. In our application, the system does not identify individuals (e.g. names), and it finishes tracking people when they leave the environment. We believe that this is a privacy-safe application. When the system is scaled up (e.g. extended to cover a large area, or associated with personal information), privacy should be more carefully considered.

VIII. CONCLUSION

We reported a series of abstraction techniques for retrieving information about people's behavior from their trajectories. Based on robust tracking with multiple laser range finders, more than ten thousand trajectories have been accumulated. Clustering techniques revealed how they used the space as well as their global behavior in the environment. Our service framework includes an anticipation system: it utilizes abstracted information to send a robot to provide services to people who are exhibiting a pre-defined local behavior associated with a particular service. It is notable that designers need to only specify target local behavior to use the anticipation system.

Results from our field trial demonstrated the effectiveness of the service framework, and also indicated that entertainment and invitation are promising applications for the robot. People appeared excited about the presence of the robot, enjoyed interacting with it, and sometimes followed its invitations. The service framework developed here enables the robot to provide such services in a real shopping arcade. Further details about people's response to the robot were examined in more detail in succeeding studies, e.g. a study of social behavior in approaching humans [44] and integration of different capabilities of robots [45], which are based on the techniques and service frameworks reported in this paper.

ACKNOWLEDGEMENT

We wish to thank Dr. Miyashita, Mr. Nishio, and Mr. Nohara for their help.

REFERENCES

[1] T. Kanda, et al, Analysis of People Trajectories with Ubiquitous Sensors in a Science Museum, *In Proc. Int. Conf. on Robotics and Automation (ICRA2007)*, pp. 4846-4853, 2007.

[2] T. Kanda, D. F. Glas, M. Shiomi, H. Ishiguro and N. Hagita, Who will be the customer?: A social robot that anticipates people's behavior from their trajectories, *Tenth International Conference on Ubiquitous Computing (UbiComp 2008)*, pp.380-389, 2008.

[3] W. Burgard, A.B. Cremers, D. Fox, D. Hänel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, The interactive museum tour-guide robot, *Proc. of National Conference on Artificial Intelligence*, pp. 11-18, 1998.

[4] R. Siegwart *et al.*, Robox at Expo.02: A Large Scale Installation of Personal Robots, *Robotics and Autonomous Systems*, 42(3), pp. 203-222, 2003.

[5] R. Gockley, J. Forlizzi, and R. Simmons, Interactions with a Moody Robot, *ACM/IEEE International Conference on Human-Robot Interaction (HRI2006),* pp. 186-193, 2006.

[6] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial, *Human Computer Interaction*, 19(1-2), pp. 61-84, 2004.

[7] A. Sanfeliu, N. Hagita, and A. Saffiotti, Special Issue: Network Robot Systems, *Robotics and Autonomous Systems*, 2008.

[8] C. Hua, H. Wu, Q. Chen, T. Wada. A General Framework for Tracking People, *In Proc. Int. Conf. Automatic Face and Gesture Recognition (FG2006)*, pp. 511-516, 2006.

[9] X. Shao, H. Zhao, K. Nakamura, R. Shibasaki, R. Zhang, Z. Liu, Analyzing Pedestrians' Walking Patterns Using Single-Row Laser Range Scanners, *IEEE Int. Conf. on Systems, Man and Cybernetics (SMC '06)*, pp. 1202-1207, 2006.

[10] X. Shao, H. Zhao, K. Nakamura, K. Katabira, R. Shibasaki, Y. Nakagawa, Detection and tracking of multiple pedestrians by using laser range scanners, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2007)*, pp. 2174-2179, 2007

[11] J. Letchner, et al., A., Large-Scale Localization from Wireless Signal Strength, *In Proc. AAAI-05*, 2005.

[12] A. Madhavapeddy, and A. Tse, A study of Bluetooth propagation using accurate indoor location mapping, *In Proc. Ubicomp2005,* pp. 105-122, 2005.

[13] S. N. Patel, et al., PowerLine Positioning: A Practical Sub-Room-Level Indoor Location System for Domestic Use. *In Proc. Ubicomp 2006,* pp. 441-458, 2006.

[14] D. Schulz, D. Fox, and J. Hightower, People Tracking with Anonymous and ID-sensors Using Rao-Blackwellised Particle Filters, *In Proc. IJCAI-03*, 2003.

[15] N. Eagle, and A. Pentland, Reality Mining: Sensing Complex Social Systems, *Personal and Ubiquitous Computing,* Vol. 10, No. 4, pp. 255-268, 2006.

[16] L. Liao, et al., Location-Based Activity Recognition using Relational Markov Networks, *In Proc. IJCAI-05,* 2005.

[17] A. Subramanya, A. Raj, J. Bilmes, D. Fox, Recognizing Activities and Spatial Context Using Wearable Sensors, *In Proc. Uncertainty in Artificial Intelligence*, 2006.

[18] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, Laser-Based Tracking of Human Position and Orientation Using Parametric Shape Modeling, *Advanced Robotics,* vol. 23, pp. 405-428, 2009.

[19] B. Hillier, J. Hanson, *The social logic of space,* Cambridge University Press, 1984

[20] D. Helbing, F. Schweitzer, J. Keltsch, P. Molnár, Active walker model for the formation of human and animal trail systems, *Physical review E,*56(3), pp. 2527-2539, 1997.

[21] N. Suzuki, et al., Learning Motion Patterns and Anomaly Detection by Human Trajectory Analysis, *In Proc. Int. Conf. Systems, Man and Cybernetics (SMC2007)*, 2007.

[22] P. Nurmi, J. Koolwaaij, Identifying meaningful locations, *In Proc. Mobiquitous 2006,* pp. 1-8, 2006.

[23] R. Aipperspach, T. Rattenbury, A. Woodruff, J. Canny, A Quantitative Method for Revealing and Comparing Places in the Home, *In Proc. Ubicomp 2006,* pp. 1-18, 2006.

[24] K. Koile, et al., Activity Zones for Context-Aware Computing, *In Proc. Ubicomp 2003*, pp. 90-106, 2003.

[25] A. Turner, A. Penn, Encoding natural movement as an agent-based system: an investigation into human pedestrian behavior in the built environment, *Environment and Planning B: Planning and Design,* 29, pp. 473-490, 2002.

[26] G. Antoninia, M. Bierlaireb, and M. Weber, Discrete choice models of pedestrian walking behavior, *Transportation Research Part B,* 40(8), pp. 667-687, 2006.

[27] X. Chai, and Q. Yang, Multiple-Goal Recognition From Low-level Signals, *In Proc. AAAI-05,* pp. 3-8, 2005.

[28] F. Sparacino, The Museum Wearable, *In Proc. Museums and the Web (MW2002),* 2002.

[29] J. Krumm, and E. Horvitz, Predestination: Inferring Destinations from Partial Trajectories, *In Proc. Ubicomp2006,* pp. 243-260, 2006.

[30] L. Liao, D. Patterson, D. Fox, H. Kautz, Learning and Inferring Transportation Routines, *Artificial Intelligence,* 171(5-6), pp. 311-331, 2007.

[31] G. Hoffman, C. Breazeal, Effects of Anticipatory Action on Human-Robot Teamwork, *In Proc. Int. Conf. on Human-Robot Interaction (HRI2007),* pp. 1-8, 2007.

[32] K. Dautenhahn, et al., How may I serve you?: a robot companion approaching a seated person in a helping context, *In Proc. Int. Conf. on Human-robot interaction (HRI2006),* pp. 172 – 179, 2006.

[33] R. Gockley, J. Forlizzi, R.G. Simmons, Natural person-following behavior for social robots, *ACM/IEEE Int. Conf. on Human-robot interaction (HRI2007),* pp. 17-24, 2007.

[34] M. P. Michalowski, S. Sabanovic R. Simmons, A Spatial Model of Engagement for a Social Robot, *AMC'06*, 2006.

[35] F. Yamaoka, et al., How Close? A Model of Proximity Control for Information-presenting Robots, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008),* 2008.

[36] M. Bennewitz, W. Burgard, G. Cielniak, S. Thrun, Learning Motion Patterns of People for Compliant Robot Motion, *The Int. Journal of Robotics Research,* 24(1), pp. 31-48, 2005.

[37] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[38] Chih-Wei Hsu and Chih-Jen Lin, A Comparison of Methods for Multiclass Support Vector Machines, *IEEE TRANSACTIONS ON NEURAL NETWORKS,* Vol. 13, No. 2, pp. 415-425, 2002.

[39] G.H. Ball, D.J. Hall, Isodata: a novel method of data analysis and pattern classification, *Technical Report AD 699616,* Stanford Research Institute, Menlo Park, CA, 1965.

[40] H. Sakoe, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 26, pp. 43-49, 1978.

[41] T. Kanda, H. Ishiguro, M. Imai, T. Ono, Development and Evaluation of Interactive Humanoid Robots, *Proceedings of the IEEE*, Vol.92, No.11, pp. 1839-1850, 2004.

[42] H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda, XIMERA: A New TTS from ATR Based on Corpus-Based Technologies, *Proc. of Fifth ISCA Workshop on Speech Synthesis (SSW5),* pp. 179-184, 2004.

[43] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model, *connection science*, 18(4), pp. 379-402, 2006.

[44] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, N. Hagita, How to Approach Humans?-Strategies for Social Robots to Initiate Interaction, *4th ACM/IEEE International Conference on Human-Robot Interaction(HRI2009)*, pp.109-116, 2009.

[45] M. Shiomi, T. Kanda, D. F. Glas, S. Satake, H. Ishiguro, N. Hagita, Field Trial of Networked Social Robots in a Shopping Mall, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2009),* 2009.

Takayuki Kanda (M'04) received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively.

From 2000 to 2003, he was an Intern Researcher at ATR Media Information Science Laboratories, and he is currently a Senior Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots.

Dr. Kanda is a member of the ACM, the Robotics Society of Japan, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence.

Dylan F. Glas received two S.B. degrees, one in aerospace engineering and one in earth, atmospheric, and planetary science, from MIT, Cambridge, Massachusetts, USA, in 1997, and he received his M.Eng. in aerospace engineering in 2000, also from MIT.

He is currently a Researcher at the Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan. From 1998-2000 he worked in the Tangible Media Group at the MIT Media Lab. His research interests include networked robot systems, teleoperation for social robots, human-machine interaction, ubiquitous sensing, and artificial intelligence.

Masahiro Shiomi received his M.S. and Ph. D. degrees in engineering from Osaka University, Osaka, Japan, in 2004, and 2007, respectively. He was an Intern Researcher at ATR Intelligent Robotics and Communication Laboratories, and he is currently a Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human-robot interaction, and network robots.

Norihiro Hagita (M'85–SM'99) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Keio University, Tsuruoka City, Japan, in 1976, 1978, and 1986, respectively.

During 1978–2001, he was with the Nippon Telegraph and Telephone Corporation (NTT). He then joined the Advanced Telecommunications Research Institute International (ATR) to establish the ATR Media Information Science Laboratories and the ATR Intelligent Robotics and Communication Laboratories, in 2001 and 2002, respectively. His current research interests include communication robots, network robot systems, interaction media, and pattern recognition.

Dr. Hagita is a Fellow of the Institute of Electronics, Information, and Communication Engineers, Japan. He is a member of the Robotics Society of Japan, the Information Processing Society of Japan, and The Japanese Society for Artificial Intelligence. He is a Co-Chair for the IEEE Technical Committee on Networked Robots.