Laser-Based Tracking of Human Position and Orientation Using Parametric Shape Modeling

Dylan F. Glas Takahiro Miyashita Hiroshi Ishiguro Norihiro Hagita

ATR Intelligent Robotics and Communication Laboratories
2-2-2 Hikaridai, Keihanna Science City
Kyoto, Japan
{*dylan, miyasita, hagita*}*@atr.jp, ishiguro@ams.eng.osaka-u.ac.jp*

Abstract

Robots designed to interact socially with people require reliable estimates of human position and motion. Additional pose data such as body orientation may enable a robot to interact more effectively by providing a basis for inferring contextual social information such as people's intentions and relationships. To this end, we have developed a system for simultaneously tracking the position and body orientation of many people, using a network of laser range finders mounted at torso height. An individual particle filter is used to track the position and velocity of each human, and a parametric shape model representing the person's cross-sectional contour is fit to the observed data at each step. We demonstrate the system's tracking accuracy quantitatively in laboratory trials, and we present results from a field experiment observing subjects walking through the lobby of a building. The results show that our method can closely track torso and arm movements even with noisy and incomplete sensor data, and we present examples of social information observable from this orientation and positioning information that may be useful for social robots.

keywords: people tracking, particle filtering, motion analysis, human-robot interaction, laser-based tracking

1 INTRODUCTION

A new class of service robots is emerging, one in which social interaction is a fundamental aspect of a robot's performance. Experimental field trials have demonstrated the possibility of robots acting as museum guides [1], receptionists [2], classroom assistants [3], guides in shopping centers, and other social roles in everyday life. As the natural-language and gestural communication capabilities of these robots improve, people's expectations of the robots' interaction skills will commensurately increase, and these robots will need to be responsive not only to speech, but to subtle cues of nonverbal communication as well.

Movement and positioning, for example, contain implicit information about a person's intentions, social relationships, mood, and status. A person's walking speed, trajectory, proximity to other people, and facing direction all provide information which can contribute to an understanding of social context.

Such knowledge could be used by service and communication robots to identify people who have lost their way or are in need of help, to stay out of the way of people in a hurry, to identify group

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

leaders for guidance or sales applications, to understand when the robot is the center of attention and when it is being ignored, to identify booths in an exhibition or exhibits in a museum that a person has missed, and for many other purposes.

Although a robot's on-board sensors can be used for some of these tasks, ubiquitous sensor networks can monitor larger areas and are subject to fewer size, power, and bandwidth restrictions. In many of our experiments and field trials, laser range finders are used for tracking people's positions as they are easier to install and less obtrusive than floor sensors, require far less processing than video tracking systems, and have a much higher precision and faster response time than RFID tracking or GPS.

To use these resources effectively, one goal of our research is to extract as much information as possible from this laser scan data. If nuances of a person's movement, such as the direction in which they are facing, can be extracted from the same laser scan data already used to determine their location, then information which is potentially useful for understanding social context will have been gained at no additional hardware cost.

In this paper we present an algorithm we have developed for tracking people using a parametric shape model which includes arm positions and facing direction in addition to basic position tracking. The algorithms used in this system are described, and quantitative results of a laboratory experiment to characterize the system's tracking accuracy are presented. A second experiment was conducted in the entrance lobby of an office building, to demonstrate the system's performance with multiple subjects in natural walking situations. Qualitative results from that experiment are presented, illustrating the system's effectiveness in tracking many people simultaneously and suggesting types of social information that can be observed in the tracking results. Finally, considerations concerning performance tuning and real-time operation of the system are discussed.

2 RELATED WORK

Human tracking itself is not a new field, and many aspects of the problem have already been explored extensively. Like many of its predecessors, our system tracks people by using particle filters to estimate their position and velocity. Particle filters are a well-known tool in the robotics community and have often been used in conjunction with laser scan data for the purposes of robot localization and mapping [4, 5] as well as human tracking. A general overview of applications of particle filters in robotics can be found in [6].

Much of the human-tracking research to date has been based on leg tracking, for both mobile robotics [7, 8] and environmental monitoring [9, 10, 11]. 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 at a height of 85-90 cm, where the arms and torso can be clearly observed.

Although less common than leg-tracking, torso-level tracking is not without precedent in research. For example, Fod *et al.* created a system using a Kalman filter to track people's trajectories with waist-height laser scanners [12], and Almeida *et al.* developed a real-time torso-level laser-based human tracking system utilizing particle filters in [13]. These systems, however, were focused specifically on

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

position tracking, whereas our work is concerned with observing body orientation and pose in addition to position.

3 POSITION TRACKING

Our algorithm was developed to track both human position and orientation. The strategy of this algorithm is to first estimate each person's position using a particle filter, and then to fit a shape model, representing the person's body orientation and arm positions, to the observed contour data.

Our initial approach to this problem had been to calculate both position and orientation using the particle filter. This resulted in an unacceptably slow system for our real-time applications. However, we observed that a majority of the computation time for each particle was being spent on orientation calculations.

In fact, the edge-based calculations used for orientation are not particularly well-suited for use in a particle filter. For position, calculations are efficient because their likelihood distributions are stable over time (regions are clearly defined and change slowly), relatively smooth in space, and easy to calculate from raw sensor data. Edge-based likelihood distributions are more complex to calculate, not stable over time (the number and placement of detected points can change rapidly between frames with a great deal of randomness), nor are they smooth in space, as the best-fit orientation can change wildly over even small variations of the assumed center position. It is thus difficult to obtain a meaningful average orientation value over a scattered set of particles.

In our technique, the orientation calculations are highly dependent upon position, but the position calculations do not depend on orientation. Thus the orientation calculations can be removed from the particle filter and performed after the position estimate is evaluated. Having done so, at each time step we need only calculate orientation once for each particle filter, rather than once for each particle. In addition, by removing variables from the particle filter we are able to reduce its dimensionality, consequently reducing the number of particles necessary for accurate tracking. By separating the calculations into a two-step process, we are thus able to dramatically increase real-time performance. More details on this topic can be found in Sec. 7.2.

3.1 Detection and Association

A common problem in tracking is the association between detected features and objects being tracked. In our algorithm, each person is tracked by a single particle filter. Doing so enables these feature-object associations to be handled implicitly by the particle filters, which follow the detected features over time. Thus explicit feature-object associations only need to be made when creating new particle filters for previously untracked humans.

To identify new humans, the raw data is segmented at every time step, to extract continuous segments of foreground data roughly corresponding to expected human widths. Clusters of these patterns are grouped together and flagged as human candidates. Candidates coinciding with humans already being tracked are removed from the list, and those remaining are propagated to the next time step, where they are merged with the candidates detected during that step. If a human candidate survives beyond a threshold number of time steps, it is considered to be a valid detection, and a new particle filter is assigned to that location, initialized with the position and velocity of the human candidate it replaces.

The removal process is much simpler than the addition process. When the particles within a filter

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

spread out beyond a defined dispersion threshold, or when their average likelihood value goes below a defined probability threshold, that particle filter is assumed to no longer be tracking a human, and it is removed.

3.2 Particle Filtering

A key component of our tracking algorithm is the particle filter, the basic principles of which will be very briefly explained here. For a more in-depth explanation, [14] provides a thorough treatment of particle filters and many other state estimation techniques.

Particle filtering is a method of estimating the state $\mathbf{x}_t$ of a system by using a cloud of "particles", each of which represents a hypothesis about that state. The following four-step procedure is performed at each iteration of a particle filter.

1. **Update** The state of each particle is updated by applying an internal *motion model*, reflecting the dynamics of the system, to the previous state estimate. The motion model used in our work is described in Section 3.4.

2. **Assign Weights** Particles are assigned weights representing their relative likelihoods according to a *likelihood model*. The likelihood model provides an approximation of the conditional probability $p(z_t|\mathbf{x}_t^{[m]})$ for particle $m, (m = 1..M)$ and measurement vector z_t taken at time step t of the particle filter. Our likelihood model is described in Section 3.5.

3. **Estimate State** An estimate of the state is then calculated, generally as a weighted average of the states of the particles.

4. **Resample** Particles are removed or propagated based on their weights to produce a new set of particles which more accurately reflects the true state of the system. Several resampling techniques exist; our system uses the sampling importance resampling technique [15].

In this way, the cloud of particles converges on the most likely state and follows it over time.

3.3 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.

3.4 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 $\mathbf{x}_{t-1}$.

As has been observed in [16], the modeling of human motion presents difficulty because it is neither Brownian in nature, nor can it be modeled as a smooth linear function, since people may stop or change direction abruptly. Thus, as a compromise between the two, a Gaussian noise component is added to each particle's v and θ values to capture the randomness of human motion. We then propagate the (x, y) motion linearly according to the resultant v and θ values of the particle.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

3.5 Likelihood Model

The purpose of the likelihood model is to approximate the value of $p(z_t|\mathbf{x}_t^{[m]})$. In this case, the measurement vector z is an array of raw sensor range measurements. 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.

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. 1 illustrates the distinction between these two kinds of information.

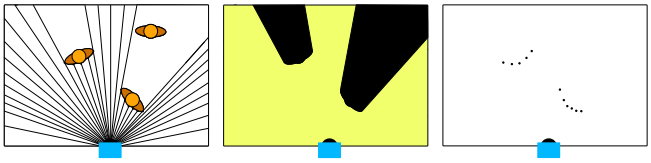


Figure 1: A typical single-sensor laser scan. (Left) The positions of humans relative to the scanner can be seen. (Center) Occupancy information. (Right) Edge information.

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. 1).

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

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

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.

3.5.1 Error Tolerance

In an ideal system, the "open" regions could be assigned a likelihood value of zero. However, in real systems there are many possible sources of error, such as calibration errors (the exact position and angle of each sensor may not be properly calibrated, leading to imperfect alignment of shadow regions), measurement errors (some textures of clothing cause noisy sensor readings and thus apparent gaps in people's bodies), timing synchronization errors (sensor data feeds are sent in real-time over a network and may arrive asynchronously, causing old data to be mixed with new), and hardware or transmission errors (which produce occasional bursts of sensor noise). The binary discretization of space into "open" and "shadow" regions is thus a slightly imperfect representation of reality. Consequently, we set the likelihood of "open" regions to a small but nonzero value p_{open}. This adds a small amount of resilience to the system, allowing particles to survive outside of the shadow regions for a short time in order to provide smoother performance with respect to such sources of error. This does not destabilize the particle filter since the likelihoods of these particles are substantially lower, and particles lying outside of the shadow regions for too long will naturally be culled in the resampling process.

4 ORIENTATION ESTIMATION

Our algorithm for calculating a person's orientation uses a parametric shape model, which we describe in Section 4.1. An angular array representation, presented in in Section 4.2, is used to store laser scanner data as a set of edge distances. As a tool for our calculations, an empirical distribution of expected distances for such an array, relative to the person's forward-facing direction, was generated based on laboratory motion-capture data. We describe the derivation of this distribution in Section 4.3.

The computation itself consists of first determining a rough estimate of body orientation, described in Section 4.4, based on the observed contour shape and the empirical distance distribution mentioned above. The second step, explained in Section 4.5, is to determine the individual arm angles, based on this rough estimate. The arm angles are then used to generate a refined estimate of orientation. Finally, Section 4.6, presents a technique for reducing accidental 180-degree reversals by considering motion direction and velocity.

4.1 Theoretical Shape Model

Large variations in cross-sectional contour shape were observed between subjects. This is due in part to individual differences in body shape, and also to differences in height. For example, arm motion is more pronounced for taller subjects, and their arms sometimes disappear if their hands briefly swing out of the scan plane.

Clothing also affects contour shape. For example, a loose shirt or a heavy coat can make a person's torso appear unusually large or asymmetrical, as can a backpack or purse.

Taking these factors into consideration, the amount of variation between subjects makes it difficult to develop a precise, yet generalizable, model. Thus a simple three-circle model was used for determining body orientation.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

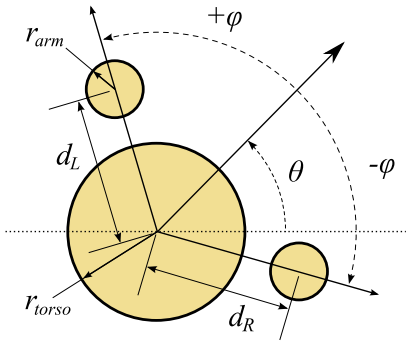


Figure 2: Our three-circle model, with the six variable parameters indicated.

Table 1: Model Parameters

Parameter	**Description**
θ	Average direction of 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

Our model is illustrated in Figure 2. 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 parameters describing the state of this model are summarized in Table 1. The two parameters of primary interest to us are θ and φ. The other parameters are held constant for this application, although they can be estimated from the data if necessary.

We have designated θ to represent the angle midway between the two arms. When a subject is standing still, this coincides with the direction of torso orientation. While the subject is walking, the swinging of the arms and torso cause θ to oscillate around the direction of motion.

The parameter φ represents the angle of separation between each arm and the center angle designated by θ. This tends not to vary far from 90 degrees, as the arms swing in alternate directions during walking.

4.2 Radial Data Representation

For these calculations, we need a way to represent 2D edge data in a consistent way for analysis. To achieve this, the information contained in these points is mapped to an angular array of distances. Distance values from the body center to the detected edge points are stored in an array of bins which

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

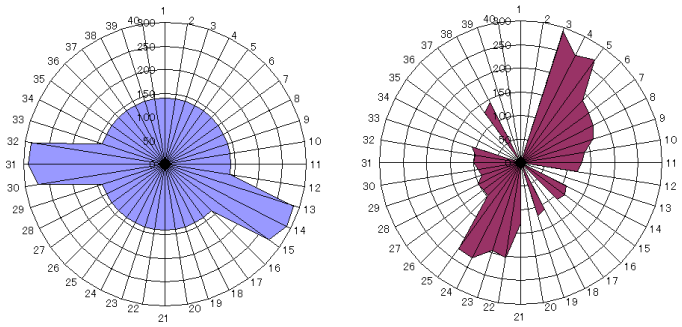


Figure 3: Examples of populated radial arrays. Left: Radial array reflecting an ideal human shape model. Right: Radial array populated with observed sensor data.

represent an angular discretization of the space surrounding the estimated human position. For each angular division, the distance to the furthest observed data point within 50 cm of the estimated human position is stored in that bin. Fig. 3 illustrates such array representations of both the ideal shape model and a set of actual shape data. A linear representation of such an array is shown in Fig. 5a.

4.3 Empirical Distance Distribution Model

A predictive distribution of radial distances is also needed for these calculations. An empirically-derived predictive distribution function representing average expected distance values as a function of angular deviation from θ was constructed from the laser scan and motion capture data gathered in the laboratory trials described in Sec. 5. This distribution function is shown in Fig. 4(a).

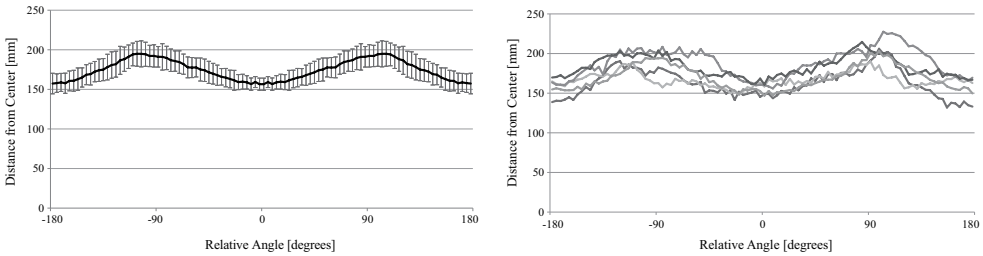
Two minutes of laser scan and motion capture data were recorded for each of five subjects. Each subject's angle at each time step was computed using the motion capture system, and a radial accumulator with 100 divisions (3.6 degrees each) was populated with the laser scan data for that time step, oriented relative to that angle. This distance data was collected over approximately 4500 time steps and averaged to determine an expected distance distribution function for each subject. These distribution functions are shown in Fig. 4(b).

Next, the data distributions were averaged between subjects. The resultant function was still somewhat noisy and asymmetrical. Making the assumption that this distribution should be symmetrical (and if there is a physiological reason for the asymmetry, to eliminate any bias based on handedness) the mirror images of the subjects' data distributions were also included in the average. Fig. 4(a) shows the standard deviation error bars for this combined distribution. The resultant distribution was then smoothed using a sliding 3-point window to reduce remaining noise. Finally, a constant offset was subtracted from the filter and it was normalized, steps which do not alter its effectiveness as a convolution filter.

4.4 First-Pass Theta Determination

The strategy for the first approximation of theta involves two radial arrays. The first is populated with actual observed distance of data points from the body center, with the angular divisions corresponding to absolute angles. The second array holds the expected distribution of distances derived in

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754



(a) Predictive arm distribution filter (b) Raw data used to derive filter

Figure 4: Predictive arm distribution filter showing standard deviation error, and raw data used to derive the filter

Section 4.3, where the angular divisions represent angles relative to θ, the person's forward direction. By convolving these arrays with each other, we can compute a goodness-of-fit function between the predicted distribution and the observed distribution, as a function of θ. The maximum point of that function is the point where the observed data best fits with the expectation model, and is thus a good first-pass estimate for θ.

To begin, we need to construct an approximate model of the actual shape profile, beginning with the radial array shown in Fig. 5a. There will nearly always be angular divisions in the radial array with no points in them. Since we have no knowledge of the actual distances of these points, we set those bins to the average value across all occupied bins, to produce a model with no gaps, as shown in Fig. 5b. (This same array will be normalized and used later as a probability distribution function for arm positions, as explained below.)

This distribution, shown in Fig. 5c, is convolved with the data array shown in Fig. 5b to generate a function representing the goodness-of-fit between the observed data and the predicted data distribution. The maximum point of the resultant distribution indicates the θ value which gives the best match between the empirical distribution and the observed data.

One challenge in this determination of θ lies in the near-symmetry of the human body. Although the expected value of the arm angles is less than 180 degrees, the observed distribution and its 180-degree mirror image overlap significantly. Thus, particularly with noisy and incomplete data, it is possible that the best-fit angle is actually rotated 180 degrees from the true θ direction. To stabilize this variable, the secondary maximum in the θ likelihood function is designated as a second θ candidate. The angular distance from the previous θ estimate to the two new θ candidates is compared and the nearest neighbor selected as the first-order θ approximation. Correction of these reversals is discussed in Sec. 4.6.

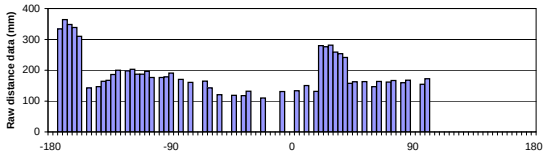
4.5 Second-Pass Theta Determination

Using this rough θ estimate, the next step is to determine the arm angles φ_L and φ_R, which will be used for determining the final θ estimate. For this step, it is necessary to derive a probability distribution function (PDF) for the arm positions from the observed data.

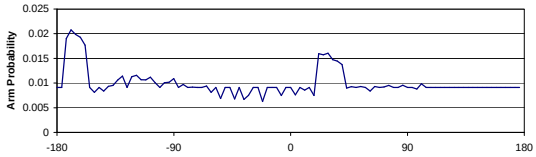
For this purpose, the shape profile model derived in the previous step can be used as a rough approximation of the arm position PDF, as it exhibits many of the essential features of such a dis-

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

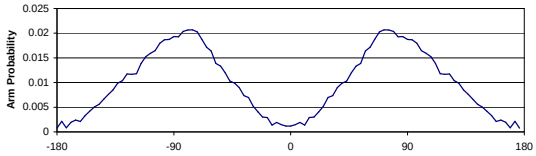
(a) Maximum observed distance values in raw data array



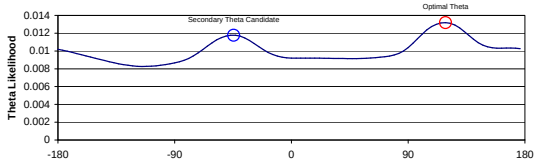
(b) Interpolated shape profile / arm angle probability distribution



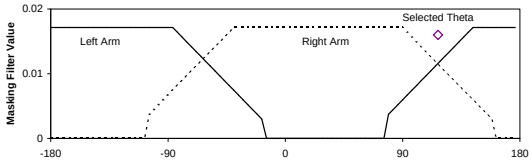
(c) Empirically-derived theta-centric distance distribution



(d) Result of convolution with distance distribution



(e) Masking functions for left and right arms



(f) Probability functions for left and right arms

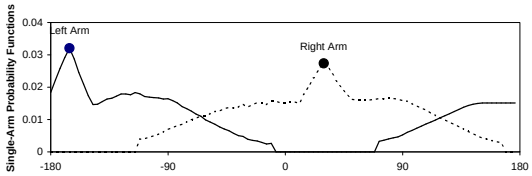


Figure 5: Intermediate steps in arm angle determination.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

tribution. For example, the presence of distant points indicates a high likelihood that an arm is in that direction. Likewise, the presence of closer-than-average points indicates a low likelihood of an arm being in that direction. Several points observed in a row give a higher-confidence estimate than a single point, high or low, and points with no data provide no information about the presence or absence of an arm. All of these features are found in both a theoretical PDF for arm distribution as well as the data array derived above. Thus, by normalizing that array, we obtain a rough approximation of that PDF.

The arm probability distribution in the radial array is then masked into two 180-degree regions by using trapezoidal masking filters on either side of the selected θ direction as shown in Fig. 5e (trapezoidal rather than rectangular masks were used for stability). These masks are multiplied with the data array from Fig. 5b to generate the two probability distributions shown in Fig. 5f. The peaks of these distributions are used as estimations of the left and right arm angles φ_L and φ_R, respectively. A refined estimate of θ is then calculated as the midpoint between these angles.

Note that at this point, if desired, the shape profile can be revisited to calculate parameters such as d_L, d_R, r_{arm}, and r_{torso}. However, this step is not necessary if θ is the only parameter of interest.

4.6 Correction of Reversals

One of the greatest difficulties in determining the person's facing direction lies in resolving the 180-degree ambiguity between forward and backward orientation. The human shape is nearly symmetrical, and even by eye it is sometimes quite difficult to discern front and back from laser scan data alone.

To resolve this ambiguity, we utilize the assumption that motion direction generally tends to coincide with the forward orientation direction. We verified this assumption quantitatively using the data recorded in the trials described in Sec. 5.

By running the basic human-tracking program without any reversal correction, we generated a dataset of human positions and orientations. Reversals (defined as periods in which the directional error was greater than 90 degrees) were identified by comparing these results with the ground-truth data from the motion-capture system. A velocity distribution was then computed for each set of data points. The results of this analysis are illustrated in Fig. 6.

An examination of this velocity distribution reveals that retrograde motion at low velocities is common, probably due to a combination of actual motion, noise, and tracking lag of the particle filter; however, higher retrograde velocities (above 500 mm/s) are almost nonexistent. Thus, any retrograde motion larger than a threshold speed of 500 mm/s is interpreted as a reversal and corrected. A time-averaged velocity estimate is used to minimize the influence of noise.

5 LABORATORY PERFORMANCE ANALYSIS

We performed an experiment in our laboratory to verify the accuracy of the human tracking system, and to gather empirical data to refine the reversal-detection and theta-approximation functions in our tracking algorithm.

5.1 Setup and Procedure

We used a Vicon motion-capture system to measure the accuracy of our laser tracking system. The Vicon system uses several infrared cameras to track reflective markers with an accuracy of 1 mm at a frequency of 60 Hz. Four SICK LMS-200 laser scanners were used, each set to a maximum range

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

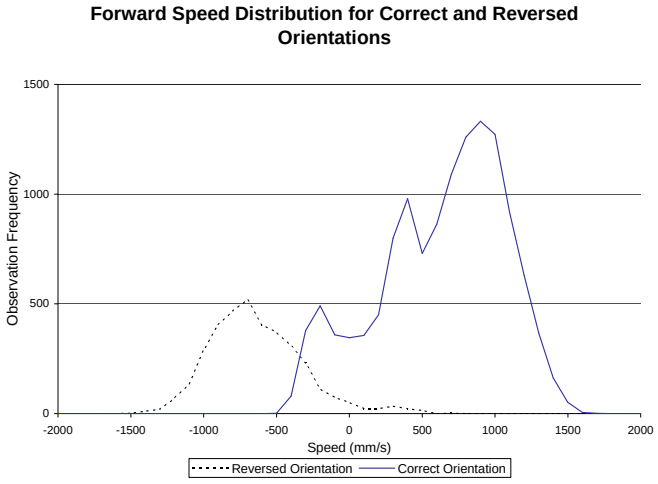


Figure 6: Distribution of forward velocity. The forward component of the velocity vector was calculated for each time-step, and a frequency histogram was computed using bin sizes of 100 mm/s. Nearly every observation with a forward velocity component below -500 mm/s was the result of a reversed direction estimate.

of 8 m, a distance resolution of 10mm, an angular range of 180 degrees, an angular resolution of 0.5 degrees, and a scan frequency of 37.5 Hz.

The space used for our experiment was a four-meter-square area with the four laser scanners situated outside the center of each edge of the square. The scan plane for each laser scanner was located at a height of 85 cm from the ground. Additionally, numbered markers were placed on the floor, as depicted in Fig. 7.

Five subjects were instructed to walk a series of patterns within the square. First, they stood in the center of the square and turned in a circle, stopping at each of the four cardinal directions for two seconds. Second, they walked figure-eight patterns, touching each of the numbered markers in order, twice. Third, they walked in a circular path inside the square, twice clockwise and twice counterclockwise. Finally, they walked randomly within the square until a total of two minutes had elapsed.

Each subject wore four reflective markers for the Vicon system. One marker was placed on the outside of each wrist, one on the subject's sternum, and one in the middle of the subject's back.

Raw data from each of the laser scanners was recorded, and the human tracking algorithm was executed offline.

5.2 Results

To compare the motion-capture data with the laser-tracking data, the midpoint between each subject's sternum and back markers was used as an estimate of the subject's body center. The absolute positional error (in the x,y plane) and absolute angular error between the laser-tracking data and the motion-capture data were then calculated for every time step in the laser-based tracking data.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

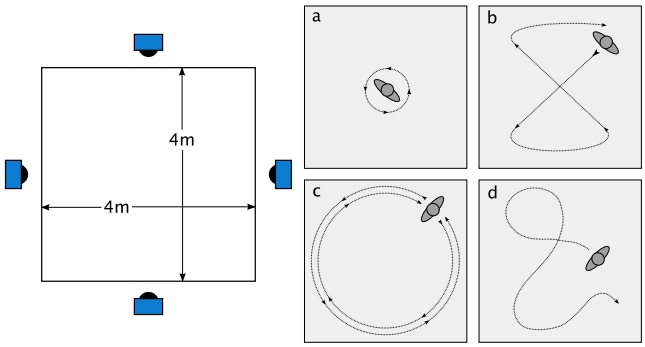


Figure 7: Floor layout for our laser tracking validation tests. Subjects were observed by four laser range finders while walking several patterns within a 4m by 4m square.

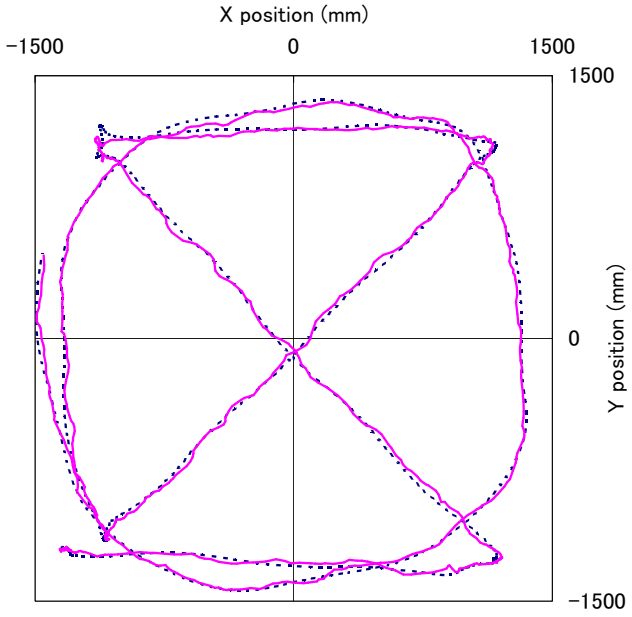


Figure 8: Tracking example from walking data. The dashed line represents ground-truth data from the motion capture system, and the solid line represents laser tracking data.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

The average positional error over all five subjects was 4.6 cm ± 2.7 cm, and the average angular error was 8.2 ± 13.8 degrees. During the 10 minutes of tracking, there were 9 brief 180-degree reversal errors. One of these lasted for 2.2 seconds, and all others were automatically corrected within 0.2 seconds. The average error with those intervals are excluded from the data was 7.4 ± 7.9 degrees.

6 NATURAL WALKING EXPERIMENT

Although the trials in our motion capture room provided useful data for verifying the system's accuracy, it is difficult to simulate natural human walking motion in such a restricted space.

To verify that the system could also work with natural walking data, we ran several trials in an open lobby, roughly 19 meters long and 8 meters wide. Experimental subjects were instructed to walk through the area several times under a number of different conditions, *e.g.* individually, in groups, wandering aimlessly, walking purposefully, making U-turns, and stopping to ask for directions.

Raw data from a network of six laser range finders monitoring this area was recorded for each trial, which we processed offline to determine human positions.

6.1 Setup and Procedure

The area of interest in our experimental environment was a space within the lobby roughly 19 meters long and 8 meters wide. We used six SICK LMS-200 laser scanners, set to scan an angular area of 180 degrees at a resolution of 0.5 degrees, covering a radial distance of 8 meters with a nominal system error of ± 20 mm, providing readings of 361 data points every 26 ms. These were placed around the periphery of the experimental area such that every point within the area of interest would be covered by at least two sensors, to minimize occlusions.

The sensors were mounted at a uniform height of 90cm, slightly above waist-level for most subjects. Tables, benches, and a small mobile robot were also placed within the walking area, but all of these were below 90cm and thus not visible to the laser scanners.

Twelve adults participated as subjects in this experiment, although at any given time only a subset of the group was walking within the sensor area. Six trials were conducted, and a total of 172 minutes of raw sensor data was collected.

6.2 Results

Two aspects of the results of this experiment will be considered here. The first is the accuracy of our method in tracking the subjects' motions, and the second is the ability to interpret this data in terms of actual body language and behavior.

6.2.1 Tracking Individuals

Quantifying the accuracy of this tracking technique is challenging due to the lack of more precise measurement techniques to establish a ground truth for evaluation. A side-by-side visual comparison of the raw data with the model-based estimate is perhaps the most effective indicator of the tracking accuracy.

Figure 9 shows raw data from five frames taken during the course of a single stride, and compares them with the model-based estimates for those time frames. Note that the swinging of the arm is clearly visible from the data, and that the model follows this movement closely.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

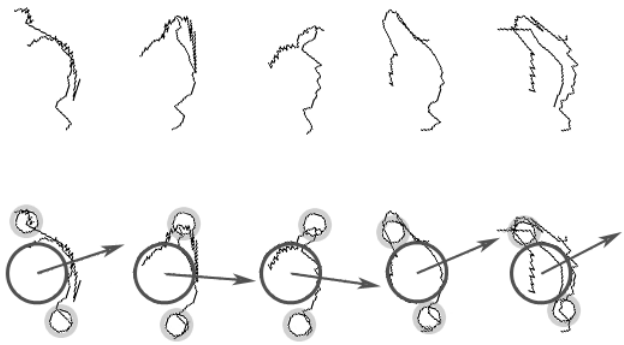


Figure 9: Example of arm and torso movement during a single stride. Top: Five frames of raw data from laser scanners taken at 320ms intervals. Bottom: Corresponding human shape model positions for each frame.

Another indicator of the tracking accuracy of our technique is the resolution of movement that is visible over time. Figure 10 shows a sample path walked by one of the subjects during our experiment. The variations in θ due to the swinging of the arms and torso with each stride are quite clearly visible, with little noise present. The more subtle change in angle as the subject walks around a curving path is also quite clearly visible from the data.

These tracking results were then visually compared with video recorded during the experiment. The subjects' arm-swinging motions were observed to match with the data. The subject's torso rotations were not as exaggerated as the variations of θ in our model, which suggests the possibility that modeling the motion of the arms during walking may offer a better estimate of torso orientation.

Interestingly, our tracking results for one trial indicated an asymmetry of motion, with one arm moving much more than the other. Inspection of the video revealed that this was not a tracking error at all, but an idiosyncrasy of the subject's walking style, an observation which suggests the possibility of using the information in this model for identifying individuals or making inferences about personality or mood.

6.2.2 Observing Interactions

In addition to the model's tracking accuracy, it is important to consider what information can be observed regarding groups of people in social situations.

Figure 11 shows three scenes from our experiment. In the top scene, two subjects are seen walking together. The model correctly shows that they are walking side-by-side, facing slightly towards each other. It is possible that the relative directions in which people face while walking together might include information about their social relationship.

In the center scene, one of the two subjects is asking a third subject for directions. The model clearly shows the social situation, in which Subjects A and B are focusing their attention on subject C. Subject A is standing back at a respectful distance, which seems to imply that A and B are not part of the same group, or perhaps that their relationship is very formal.

The bottom scene illustrates the tracking of a group of subjects. Again, the group dynamic is

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

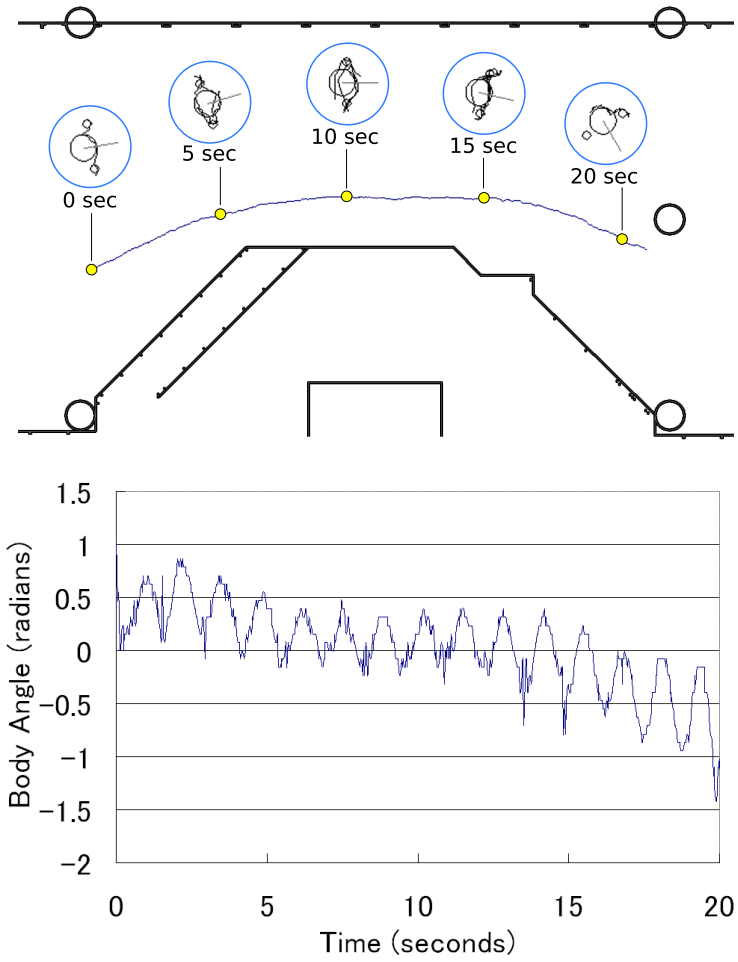


Figure 10: Body angle tracking during 20 seconds of walking motion. Top: An overview of the walking path in our lobby experiment shows the subject's walking path, as well as close-up views of the subject's body position at several points along the path. Bottom: Observed body angle variations (in room-centric coordinates). Periodic oscillations due to natural arm-swinging motion during each stride are clearly visible.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

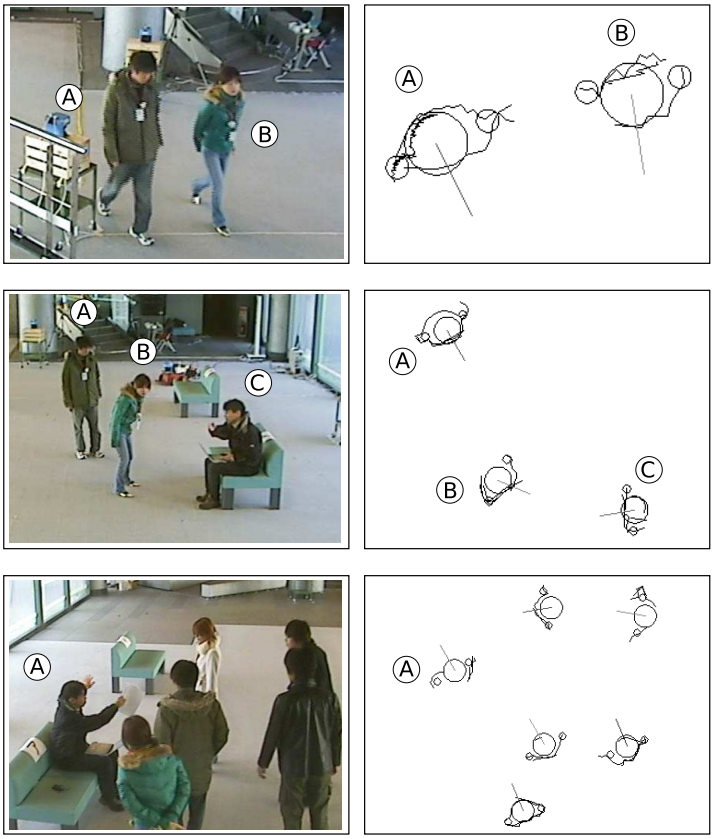


Figure 11: Scenes from the experiment.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

apparent, in that all of the subjects are listening to instructions from Subject A. (Note that the model is unable to correctly determine the direction of Subject A because he is sitting and holding his arms in an unusual position.)

All three of these examples illustrate information that could not have been determined from location alone, and they suggest many possible types of social information that may be observable from this data.

7 DISCUSSION

7.1 Performance Tuning

Many variables affect the performance of the system in terms of operating speed, position and angle accuracy, smoothness of motion, and false or missed detections. By reducing the velocity noise added during the motion model updates, for example, higher positional accuracy and smoother trajectories can be attained, but the particle filter becomes less able to follow trajectories that change abruptly.

The number of particles is another variable. If a large number of particles are used, the particle cloud's trajectory stabilizes and becomes smoother, but this comes at the cost of an increased reaction delay and increased computation time. Our algorithm uses the technique of KLD-sampling [17] to adapt the number of particles based on the density of their distribution, down to a fixed minimum limit.

7.2 Real-time Operation

Although the results presented in this paper were generated offline, this tracking software has primarily been developed for use with real-time data streams. Using this software in a real-time system raises the critical issue of processing speed. If the time required to process the data for one time step exceeds the sampling interval of the sensors, then data will be lost and tracking accuracy will begin to decrease.

Here we present a performance analysis using a Windows XP system with a 2.4 GHz Intel Pentium 4 processor and 1 GB of RAM. The tracking software was implemented in Java and executing using a Java 6 Virtual Machine. The tracking analysis was performed on a 4.5-minute data sample from a shopping center, during which between 3 and 18 people were tracked simultaneously. Six sensors were used in this experiment, with a frequency of 37.5 Hz, i.e. a sampling interval of 26.6 ms. A minimum of 50 particles was used for each person.

To illustrate the importance of the speed improvement gained by performing the orientation calculations separately from the particle filter, Fig. 12 compares our system's performance against an algorithm in which orientation calculations are integrated with the position calculations within the particle filter.

This performance comparison illustrates two key points. The first point is that even with the relatively slow Pentium 4 machine used here, it can be expected that 10-12 people can be tracked without any loss of data, *i.e.*, the tracking calculations can be completed within one data update cycle. With 18 people, incoming data would be dropped, but every second data frame would still be processed.

The second point is that, as stated in Sec. 3, the orientation computations are not very well-suited for integration with the particle filter. Fig. 12 shows that the integrated algorithm requires about four times the computation time of the two-step algorithm. In other words, the improved efficiency of the two-step algorithm enables four times as many people to be tracked at once. To address the question of

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

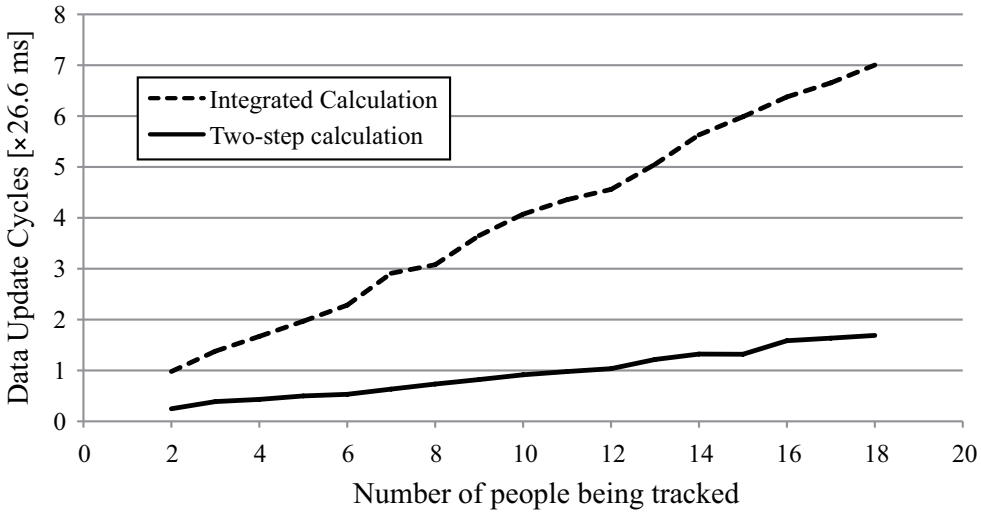


Figure 12: Variation of average computation time with number of humans being tracked. These results indicate that 10-12 people can be tracked before computation time exceeds the sensor sampling interval of 26.6 ms.

whether the choice of algorithm affects tracking accuracy, we repeated the analysis from Sec. 5.2 using the integrated algorithm. Results were substantially worse than with the two-step algorithm. First, many more reversals were observed with the integrated algorithm. Even correcting for the reversals, the average angular error was still 25.2 degrees, as opposed to 7.4 degrees for the two-step algorithm. This was most likely due to the issues stated in Sec 3, such as the non-smooth likelihood model and high sensitivity to position error.

7.3 Future Work

The next step in this research is to use the generated position and orientation data to improve robotic applications. Techniques should be developed for analyzing a person's trajectory through a given environment to learn about that person's intentions. Information about the directions in which people in a group are facing and their relative standing or walking positions may be helpful in identifying social rank within that group. Trajectory and orientation data might be useful in identifying people in a crowd who are interested in talking with the robot, or who have lost their way and need guidance.

Another possible area for future research is the addition of anatomically-based physical dynamics. Rather than simply modeling motion using a geometric circular model, incorporation of arm swinging and stride motion into the model could provide much more stable and accurate results. Currently, the system is able to extract a person's torso direction, which has been observed to oscillate from left to right while walking. A more detailed dynamic model could incorporate walking speed and rhythm to determine an even better estimate of the person's direction of attention.

Finally, the integration of this system with other tracking technologies, such as a leg-based laser tracking system, could provide a very robust estimate of a person's pose and enable the interpretation of more subtle expressions of body language.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

8 CONCLUSIONS

We have developed a system in which a network of laser range finders is used for tracking the positions and orientations of people.

Comparison with results from a motion capture system verified the position accuracy to be 4.6 cm $\pm$ 2.7 cm and the orientation accuracy of to be 7.4 $\pm$ 7.9 degrees (excluding 180 degrees reversals). The system is expected to perform without performance degradation while tracking 10-12 people in real time on a Pentium 4 Windows PC.

This human tracking system has already been used extensively for providing ground-truth data and tracking humans in several experiments and field trials. The system is also actively being used as a platform for extracting useful social information from human movement for social robotics applications.

References

[1] W. Burgard *et al.*, The interactive museum tour-guide robot, in *Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, pp. 11–18, 1998.

[2] R. Gockley *et al.*, Designing robots for long-term social interaction, in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2199 – 2204, IEEE, 2005.

[3] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, Human Computer Interaction **19**, 61 (2004).

[4] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, Monte carlo localization for mobile robots, in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 1999)*, pp. 1322–1328, ICRA, 1999.

[5] M. Montemerlo, S. Thrun, and W. Whittaker, Conditional particle filters for simultaneous mobile robot localization and people-tracking, in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 695–701, 2002.

[6] S. Thrun, Particle filters in robotics, in *Uncertainty in Artificial Intelligence (UAI)*, pp. 511–518, 2002.

[7] D. Schulz, W. Burgard, D. Fox, and A. B. Cremens, International Journal of Robotics Research (IJRR) **22**, 99 (2003).

[8] A. M. Villagrasa, People tracking for a personal robot, Master's thesis, Royal Institute of Technology, 2005.

[9] A. Brooks and S. Williams, Tracking people with networks of heterogeneous sensors, in *Australasian Conference on Robotics and Automation (ACRA)*, pp. 1–7, Brisbane QLD, Australia, 2003.

[10] J. Cui, H. Zhao, and R. Shibasaki, Fusion of detection and matching based approaches for laser based multiple people tracking, in *Proc. IEEE CVPR* Vol. 1, pp. 642–649, 2006.

[11] H. Zhao and R. Shibasaki, IEEE Transactions on Systems, Man, and Cybernetics **35**, 283 (2005).

[12] A. Fod, A. Howard, and M. J. Mataric, Laser-based people tracking, in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3024–3029, 2002.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754

[13] A. Almeida, J. Almeida, and R. Araujo, Real-time tracking of moving objects using particle filters, in *Proc. of the IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 1327–1332, Dubrovnik, Croatia, 2005.

[14] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics* (MIT Press, 2005).

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, Radar and Signal Processing, IEE Proceedings-F **140**, 107 (1993).

[16] A. Bruce and G. Gordon, Better motion prediction for people-tracking, in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, LA, USA, 2004.

[17] D. Fox, KLD-Sampling: Adaptive Particle Filters, in *Advances in Neural Information Processing Systems (NIPS) 14*, pp. 713–720, MIT Press, 2001.

This is an electronic version of an article published in Advanced Robotics [23]4, pp. 405-428, 2009. Advanced Robotics is available online at : DOI: 10.1163/156855309X408754