2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Multi-hypothesized Oscillation Models Employing Floor Sensors for Tracking People

Takuya Murakita and Hiroshi Ishiguro

Abstract— **One of the critical research issues for the future robot is the avoidance of collisions with people while it moves among them. Therefore, people-tracking and motion prediction are important. People are tracked more readily and successfully by using distributed sensor networks than robot's local sensors. Floor sensor networks, in particular, are resistant to changes in lighting conditions and other environmental disturbances. However, the problem with the system is that a person walking is observed as if it were being done in "rabbit hops": The signal is nonlinear and even nonholonomic. We tackled the problem by assuming that human walking is regular in terms of walking rhythms. Then the signal was modeled by an oscillation model based on four walking parameters: cycles, phases, directions, and strides. To adapt the parameters to irregular walking, the model was multi-hypothesized based upon a particle filtering algorithm. Experimental results showed that the multi-hypothesized oscillation models showed more than 80% tracking accuracy for four walking patterns: straight walking, stop-and-go, turn, and winding walking. Further, the models were superior to the nearest neighbor filter with regard to the performance of data association for two persons.**

I. Introduction

ROBOTS are expected to work in spaces where a number of people walk, as shown in figure 1. In order for the robot to move in such a space efficiently, the avoidance of collisions with people is an essential function. The robot has to know people's locations to realize this avoidance; so a people-tracking technology is important.

So far, the people-tracking function for a mobile robot has been performed by using its own local sensors, such as a laser range finder [1]. Although such a sensing technology, based on the first-person point of view, is important in terms of the robot's autonomy, sensor networks, such as those that [2] propose, can track people more readily and successfully. This is because sensors are fixed in a global coordinate system, and the sensors' locations are known.

One of the most effective sensor networks is a multi-camera system [3,4]. A camera is a very common sensor for tracking people; however, it tends to be susceptible to changes in lighting conditions and visual occlusions. Visual occlusions are particularly difficult because they require elaborated occlusion reasoning, for example [4], for tracking multiple persons. On the other hand, floor sensor networks [7–9] can avoid these difficulties because they detect human weight directly.

One may think that the locations of people can be readily obtained by using floor sensors. This is true in the case of a robot because it generates a continuous signal, as shown in figure 2. The difference in the positions of the robot over consecutive time steps is small enough; therefore, it is readily tracked by the nearest neighbor filter [10]: One of the simplest tracker. However, a person uses bipedal walking, which results in a discrete signal, as shown in figure 3. Thus, the difference in the positions of the footprints over consecutive time steps is not negligible. We call these "rabbit hop" signals because the footprints jump rapidly with the alternating steps

Figure 1. Robots that work with people.

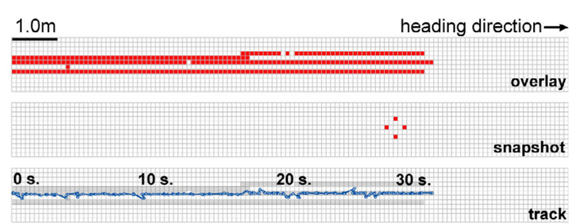


Figure 2. Overlay, snapshot, and track of the continuous signal generated by a robot moving on four wheels.

Figure 3. A rabbit hop signal generated by a person.

Corresponding author: T. Murakita, Osaka University Graduate School of Engineering, 2-1 Yamada-oka, Suita, Osaka 565-0874, Japan. TEL & FAX: +81-6-6879-4180. e-mail: murakita@ed.ams.eng.osaka-u.ac.jp
H. Ishiguro, Osaka University Graduate School of Engineering, 2-1 Yamada-oka, Suita, Osaka 565-0874, Japan. TEL & FAX: +81-6-6879-4180. e-mail: ishiguro@ams.eng.osaka-u.ac.jp

of the person. The dynamics of the signal is nonlinear and even nonholonomic. Therefore, the Kalman filter is not applicable. Using the nearest neighbor filter is one thought; however, it leads to poor data association. Therefore the point of this paper is to propose a human-tracking model that has a tight validation area for robust data association. The validation area is the extent of a distribution of particles around which observations are associated to the filter.

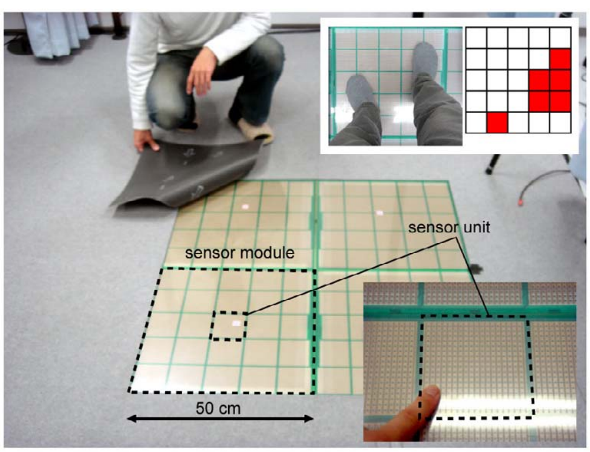
A real-time four-state switching model [8] has already been proposed by us for the rabbit hop signal. However, it employed an old type of floor sensor network; its spatio-temporal resolution was coarse, and observations were sometimes missed because of the small aperture of the sensor. Further, the data association of multiple persons was not considered to be sufficient.

To propose a new human-tracking model, human steps were first assumed to be regular in terms of walking rhythms. Then the signal was modeled by an oscillation model based on four walking parameters: cycles, phases, directions, and strides. To adapt the parameters to irregular walking, the model was multi-hypothesized based on a particle filtering algorithm. The following sections describe the model and how to multi-hypothesize it.

Our IT (Information Technology) society is evolving into an IRT (Information and Robot Technology) society. In the IRT society, the Internet will be extensively used by integrating it with robot and sensor network technologies. To achieve the IRT society, the Network Robot Forum, where major electronic and robotic enterprises are working together, is taking the initiative. Our work reported in this paper will contribute to these activities.

I. FLOOR SENSOR NETWORK

Figure 4 shows the floor sensor system employed in this paper. It was developed by Vstone Corporation (Osaka, Japan). The system is composed of 50-cm by 50-cm sensor modules that are connected serially based on a multi-hop communication protocol. Each sensor module has 25 sensor units that are 10 cm by 10 cm in size and output binary pressure data. Furthermore, each sensor unit has 289 binary electrical contacts, a certain number of which give the sensor unit a pressure threshold. One hundred sensor modules were installed in our laboratory, as shown in figure 5. The square composing the meshed region represents a sensor unit. The total data of the 2,500 sensor units is obtained 15 times per second through the RS-232 interface of a personal computer.



Total detecting area:	25.0 m^2
Spatial resolution:	100 mm × 100 mm.
Temporal resolution:	15.0 frames/ s.
Pressure resolution:	1 bit
Saturation pressure:	0.2−0.25 kgf/ cm^2

Figure 4. Overview of the floor sensor system.

II. MULTI-HYPOTHESIZED OSCILLATION MODELS

A. Assumptions

Floor sensors can be activated by arbitrary objects: Tables, chairs, luggage, robots, people, and so forth. Those objects have their own spatio-temporal characteristics in sensor signals. Therefore, they can be tracked by applying different tracking models. However, only the tracking of people is discussed in this paper.

People's postures are variable; standing, walking, running, lying, sitting and so forth. Although all of them should be successfully tracked, in this paper the types of postures are confined to standing or walking, which are common postures when a person communicates with a robot.

Tracking models for a person depend on sensor resolutions. The tracking model was developed assuming the sensor specifications shown in figure 4. The spatio–temporal resolution of the sensor seems to be somewhat coarse; however, a smaller amount of information has an advantage in terms of production, communication, and computational cost. In view of communication theory, high-resolution sensors can imitate low resolution sensors. Therefore, it is reasonable to develop tracking models with low-resolution sensors for versatility.

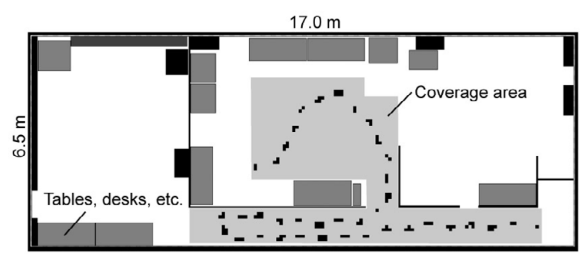


Figure 5. Layout of the sensor units.

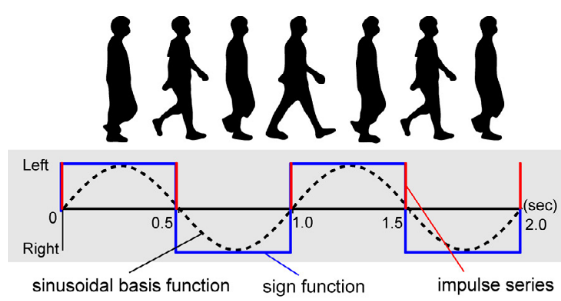


Figure 6. Representation of a person's walk using a sinusoidal basis function.

B. The oscillation model

Human walking is first assumed to be periodic in terms of walking rhythms. Then the rhythm is represented by a sinusoidal curve as shown in figure 6. The curve, fixing magnitude, is characterized by cycle T and phase θ,

$$f(t) = \sin\left(\frac{2\pi}{T}t + \theta\right) \cdot \tag{1}$$

By taking the sign of the function, a sign function $\beta(t)$ is obtained,

$$\beta(t) = \mathrm{sgn}\left[\sin\left(\frac{2\pi}{T}t + \theta\right)\right] \cdot \tag{2}$$

The difference of the sign over consecutive time steps gives an impulse series that represents changes of the signs,

$$\alpha(t) = \left(\frac{\beta(t) - \beta(t-1)}{2}\right)^2 \cdot \tag{3}$$

Then, the following function represents the rabbit hop signal, and this dynamics is called an oscillation model:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \alpha(t)\begin{bmatrix} L\cos\Phi \\ L\sin\Phi \end{bmatrix}, \tag{4}$$

where $\boldsymbol{x}_t$ represents the two-dimensional positions of a foot on the floor at time step t. L and Φ are the stride and heading direction, respectively. The second term of the right side becomes nonzero only when the impulse is triggered, which results in the rabbit hop dynamics. This model can represent variable signals by altering the values of the four parameters. Figure 7 shows the signals produced by the oscillation model.

C. Multi-hypothesizing the oscillation model

The oscillation model assumed regular walking. The next step is to adjust the parameters to irregular walking. This is performed by multi-hypothesizing the oscillation model. Figure 8 illustrates the dynamics of an aggregation of 1,000 oscillation models, each one of which has different values for the parameters. The left column shows the dynamics only when the strides were multi-hypothesized by using Gaussian noise. The initial positions of the models were given to form a Gaussian noise that has 10-cm standard deviation. Similarly, the middle and right columns show the dynamics only when the cycles and directions were multi-hypothesized, respectively. Phases were not multi-hypothesized because they obey cycles and time. The reason for this is given in the next subsection.

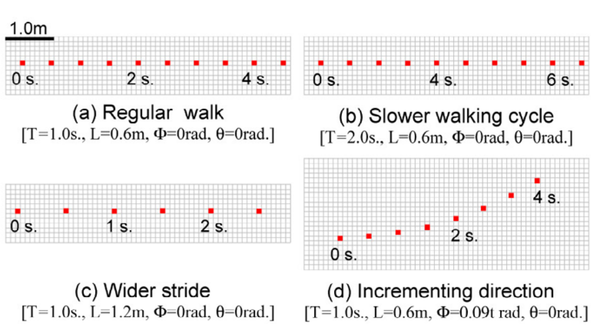


Figure 7. Variable walking signals demonstrated by the oscillation model.

Note that the parameters of oscillation models were not updated. Each parameter keeps a constant value given initially. To track dynamic human walking, parameters must be validated and updated every time step. The next section describes how parameters are updated based on a particle filter.

D. Implementation on a particle filter

The oscillation model shows entirely nonlinear dynamics; therefore, the Kalman filter, even the Extended Kalman filter, cannot be easily applied to update the parameters. Therefore, a particle filter was employed. The particle filter is a numerical solving method for posterior probability distributions. As described in [11], particle filters, bootstrap filters, the CONDENSATION algorithm, and the Monte Carlo filter are synonymous. They are Bayesian filters that repeat prediction and observation to calculate posterior probability distributions numerically. A particle filter can be viewed as a multi-hypothesized tracker if the particle is considered to maintain a hypothesis of states. The state, observation, prediction model, and observation model were implemented based on this concept.

The state was a four-dimensional vector that characterizes the oscillation model,

$$\boldsymbol{x}_t = \begin{bmatrix} T_t & \Phi_t & L_t & \theta_t \end{bmatrix}^T \cdot \tag{5}$$

Note that the parameters are time varying. The observation was a two-dimensional vector that points to the location of a foot on the floor,

$$\boldsymbol{z}_t = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^T \cdot \tag{6}$$

The state transition is essentially equal to equation 4. However, it should be built onto an observation model for theoretical justice. The prediction model was a common diffusion model,

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \boldsymbol{w}_t\,, \tag{7}$$

where $\boldsymbol{w}_t$ is a noise, each entry of which is drawn from the following Gaussians,

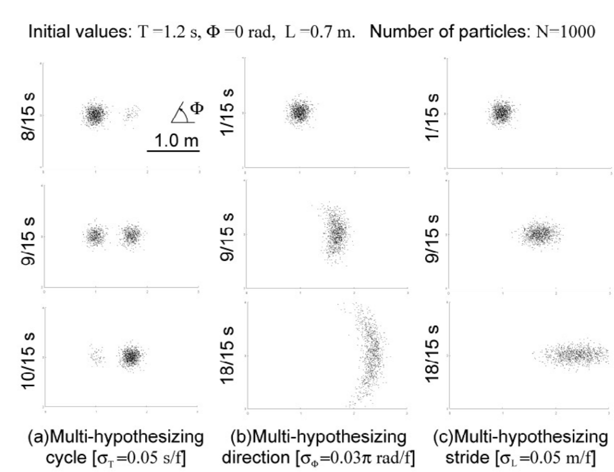


Figure 8. Multi-hypothesized 1000 oscillation models each one of which was given a different set of parameter values.

$$\boldsymbol{w}_t = \begin{bmatrix} w_T \\ w_\Phi \\ w_L \\ w_\theta \end{bmatrix} \approx \begin{bmatrix} N_T(0,\sigma_T) \\ N_\Phi(0,\sigma_\Phi) \\ N_L(0,\sigma_L) \\ w_{\theta,t} \end{bmatrix}, \tag{8}$$

w_T and w_Φ were imposed as bounding conditions given a priori:

$$\begin{aligned} w_T^{Lower} &= 0.5\ s,\ w_T^{Upper} = 4\ s, \\ w_L^{Lower} &= 0.2\ m,\ w_L^{Upper} = 1 m. \end{aligned} \tag{9}$$

$w_{\theta,t}$ is special. It represents compensation noise and depends on the cycle and time. It was chosen to satisfy the following continuity condition of the sinusoidal curve specified by equation 1:

$$w_{\theta,t}(T_{t+1}) = 2\pi\left(\frac{1}{T_t} - \frac{1}{T_{t+1}}\right)t \cdot \tag{10}$$

Standard deviations of equation 8 were tuned manually to maximize the tracking performance of 50 walking samples:

$$\begin{aligned} \sigma_T &= 0.33\text{s/frame.} \\ \sigma_\Phi &= 0.13\pi \text{ rad/frame.} \\ \sigma_L &= 0.05\text{m/frame.} \end{aligned} \tag{11}$$

Note that the frame rate was fixed at 15 fps.

The observation model was a conventional Gaussian mixture,

$$p(z_t \mid \boldsymbol{x}_t) = \exp\left\{-\frac{1}{2}\sum_{j=1}^{M}\left(\boldsymbol{x}_t^- - z_{t,j}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_t^- - z_{t,j}\right)\right\}, \tag{12}$$

where $\boldsymbol{x}_t^-$ is calculated by equation 4 based on the predicted state. M is the total number of active sensor units, and $z_j, j = 1,\ldots,M$ represent the position vectors of these activated sensor units. The covariance matrix was chosen to be $\boldsymbol{\Sigma} = 0.07^2 I$, considering a typical foot size, and the unit was m^2. A threshold was applied to equation 11 to make data association robust. Any deviation exceeding $0.3\ m$ has a likelihood of zero. Figure 9 illustrates how particles fit human walking based on the particle filter. It is interesting to note that the particles form a time varying bimodal distribution, each peak of which corresponds to a foot.

E. Model switching

Although the multi-hypothesized oscillation models adapt to irregular walking, they often failed to track a standing person because it is not periodic motion. Therefore, a standing model was introduced. It was a conventional two-dimensional Gaussian tracker as used in our former research [8]. The state vector and observation vector are in the same two-dimensional space that represents the locations on the floor:

$$\boldsymbol{x}_t = [x_1 \quad x_2]^T, \tag{13}$$

$$z_t = [z_1 \quad z_2]^T. \tag{14}$$

The prediction model was a conventional diffusion model with Gaussian noise,

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{x}_t + \boldsymbol{w}_t \\ \boldsymbol{w}_t &\approx N(0,\Sigma) \end{aligned}, \tag{15}$$

where the covariance matrix was set to be $\boldsymbol{\Sigma} = 0.1^2 I$, and the unit was m^2. The observation model was a Gaussian mixture,

$$p(z_t \mid \boldsymbol{x}_t) = \exp\left\{-\frac{1}{2}\sum_j\left(\boldsymbol{x}_t - z_{t,j}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_t - z_{t,j}\right)\right\}. \tag{16}$$

The multi-hypothesized oscillation models transferred to the standing model if the mean of the cycles exceeded 3 sec. Conversely, the standing model transferred to the multi-hypothesized oscillation model if the location moved more than 30 cm within a time step.

F. Initialization

Activated sensor units obtained every time step were associated with existing tracks. A track means a particle filter being executed. How they were associated is mentioned in the following subsection. If an activated sensor unit was not associated with any tracks, or there were no tracks, it was considered to be evoked by a new person. Those units were

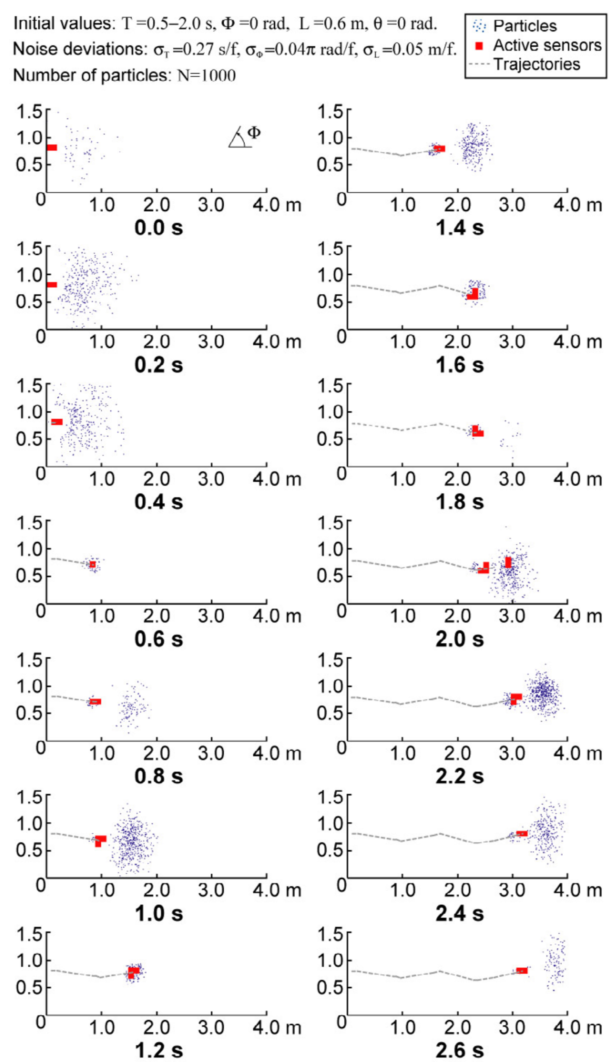


Figure 9. Tracking a person with multi-hypothesized oscillation models.

first segmented into circles (16 cm in radius assuming a typical foot size) by using a conventional clustering method. Then the Gaussian tracker mentioned in the preceding subsection was created, and the segmented sensor units were assigned to the tracker.

G. Data association

Given an observation; an activated sensor unit, the summation of all likelihoods of particles was calculated for each existing track. Then the observation was associated with the track that showed the highest summation. This method imitates the Bayes decision rule.

III. Experimental results

A. Tracking a person with variable walking

If people walked by maintaining sufficient space between each other, the nearest neighbor filter, which associates nearer observations on a first-come basis, would track the people by making the validation area (the area to search observations) wide. However, if people came close to each other, the validation areas of the respective filters would overlap, and the filters would fail to associate observations correctly. Therefore the point is how to tighten the validation area. The multi-hypothesized oscillation models enable this by using a time varying bimodal distribution as illustrated in figure 9.

It is natural that a tighter validation area makes tracking more difficult. Therefore, we examined whether the model is reliable for four common types of walking: straight walking, stop-and-go, turns, and winding walking as shown in figure 10. For straight walking, test subjects were instructed to walk at their usual speeds. For stop-and-go, the time during which the subject was stopped was about three seconds. For turns, the test subjects were instructed to turn within two seconds. For winding walking, two pylons were installed at a three-meter interval, and the test subjects wound around them. For each type, 10 persons (eight men and two women 21—32 years old, mean23.2, S.D.4.4) performed 5 walks, resulting in 50 walking samples.

The rates of successful tracks from out of 50 walks for each walking pattern are illustrated in figure 11. The operation time calculated by a Pentium 4 2.8-GHz processor is also illustrated. Note that standard deviations of walking parameters were tuned in advance, and the number of particles was varied because it greatly affects the operation time; real-time processing is essential for human-robot interaction

B. Tracking two persons and data association

As mentioned before, a situation where people come close to each other makes data association difficult. Because the evaluation of the tracking performance under such a situation is complex, the situation was reduced to data association of two persons. The performance was examined about three common approach patterns: facing, parallel, and crossover approach. For facing and parallel approach, the gap of persons was 50cm.

Ten sets of walks were performed for each approach pattern. If the tracking was successful, the tracking system would draw two tracks, each of which corresponds to each person as shown in figure 12. However, if the tracker failed, the two tracks would split, merge, or counterchange; the failure would disturb the tracks, and we could recognize the occurrence of mis-tracking. If tracks were never disturbed from initialization throughout termination, the trial was defined successful. A multi-hypothesized oscillation model and a conventional Gaussian model as described by equations 12 through 15 were compared in terms of data association. Figure 13 shows the success rates of the ten trials.

IV. Discussion

A. Tracking performance

According to figure 11, the straight walk showed the highest tracking performance. That is reasonable because it highly satisfies the assumption of the oscillation model, where the values of the parameters were converged as shown in figure 14. The other three types of walking were inhibited by a stop or a change of direction. However, by introducing the stop model and the multi-hypothesization, they showed more than 80% tracking accuracy.

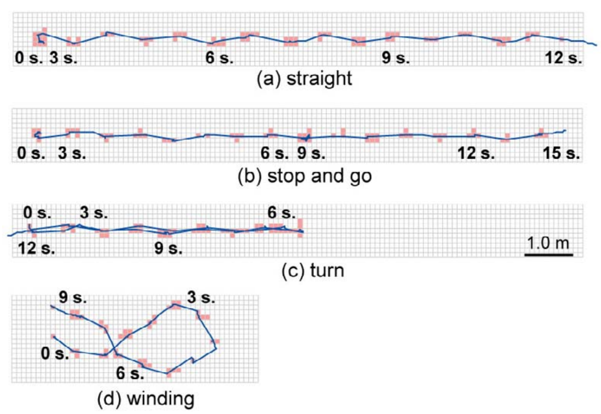


Figure 10. Four common types of walking.

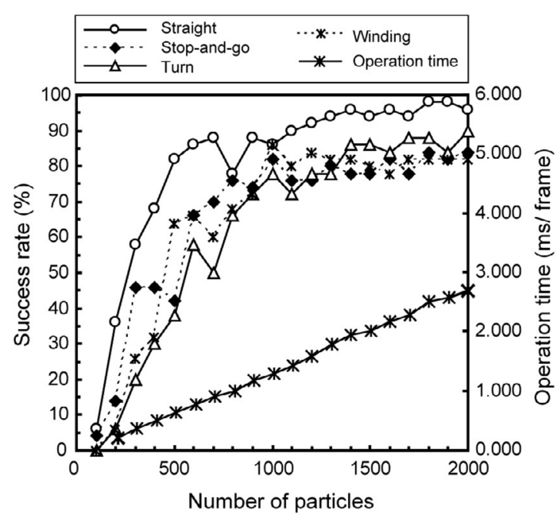


Figure 11. Success rates and operation time vs. number of particles.

Figure 15 shows two types of typical mis-tracking. One is an initialization failure. At 0 sec., two trackers are initialized. That indicates that both feet of the same person got onto the floor sensors simultaneously. That is very rare; however, the floor sensors sometimes missed an initial load applied by a person because they were protected by carpets to prevent physical damage. The other mis-tracking was most common: inappropriate distribution of particles. At time 6 sec., track 2 missed a new signal because there was no particle whose likelihood became greater than zero, and track 3 assigned the missed signal. This type of mis-tracking can be recovered based on the proximity of the distance of the tracks.

The difference between the Gaussian tracker and multi -hypothesized oscillation models was clear in the crossover approach. Basically, the Gaussian tracker is equivalent to the nearest neighbor filter in terms of data association if the number of particles is large enough; therefore it associates nearer observations with itself. That causes an association error as shown in figure 16a. On the other hand, multi-hypothesized oscillation models keep a proper distribution of particles, predicting the next position of a foot as shown in figure 16b. However, if there were more than two persons, the recovery may lose the consistency of the person-observation correspondence.

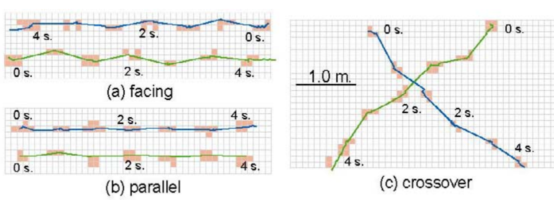


Figure 12. Three common types of approach patterns.

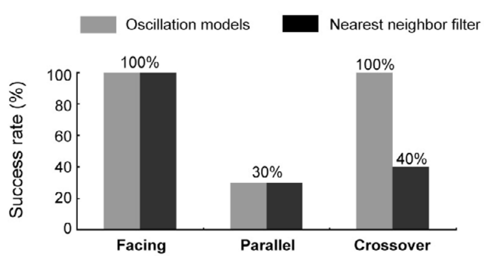


Figure 13. Success rates vs. approach patterns.

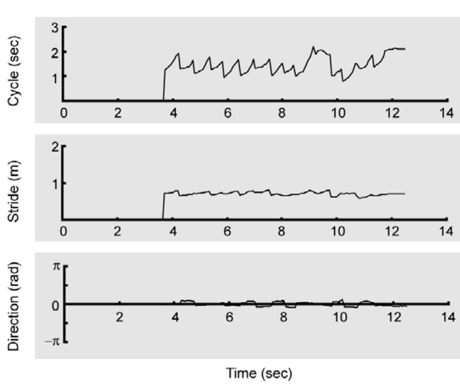


Figure 14. Conversion of the values of parameters.

According to figure 13, the facing approach was successful both with the Gaussian tracker and multi -hypothesized oscillation models. This is because particles properly captured signals. However, the parallel approach showed poorer performance. The reason for this was that the gap between the two persons was too close when the trackers were initialized; observations to each person were clustered together, and only one tracker was initialized. However, whenever initialization was successful, tracking was successful.

Thus, multi-hypothesized oscillation models show higher tracking performance compared with the Gaussian tracker. In terms of computational cost, the multi-hypothesized oscillation model can track more than 10 persons in real-time, by using a common personal computer. In fact, it tracked five persons simultaneously, as shown in figure 17, although it was not always successful. Based on the above discussion, it is concluded that our tracking system is reliable if people keep more than 50-cm distances and do not enter a tracking area simultaneously, which may be a common tracking situation.

B. Future topics of research

We have to profess that several important things have not been evaluated yet because experiments were concentrated on showing the fundamental performance of the tracking model. Therefore, for example, the tracking performance of children, elderly people, or people with wheelchair is not evaluated. Further, people can sit on a bench, carry luggage, or walk with children. Such postures would disturb the pattern of sensor activation and limit the application of out system. We think these difficulties will be solved by data fusion with vision sensors. Developing more robust tracking system employing floor sensors, video cameras, and infrared sensors is one of our next topics of research.

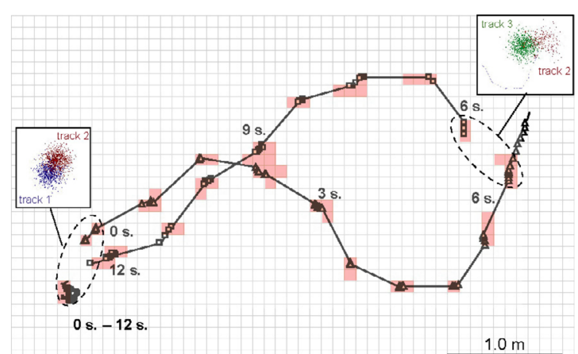


Figure 15. Two typical types of mis-tracking.

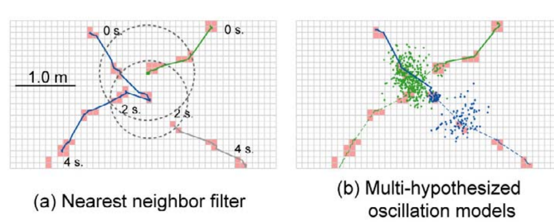


Figure 16. Data association of a crossover approach.

Although this paper focuses on people tracking, our tracker can also track a robot with the Gaussian tracker shown in figure 17. So far, this kind of tracking has been performed by a local sensing technique, such as robot vision or a laser range finder mounted on the robot's body. However, the local sensing technique is not robust. Using a floor sensor network is more pragmatic and robust because sensor modules can be readily calibrated. One may claim that the floor sensor network costs too much; however, we believe emerging technologies will solve this problem. Now that a robust people tracker has been developed, the next topic of research will be to develop a robot that is aware of, can work with, or can communicate with people by sensing their locations with a floor sensor network.

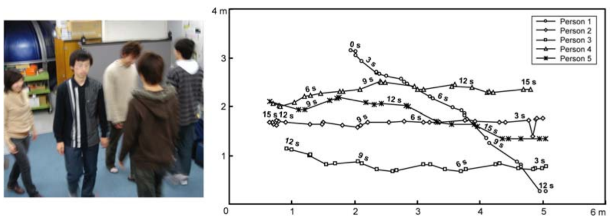


Figure 16 Tracking five persons simultaneously.

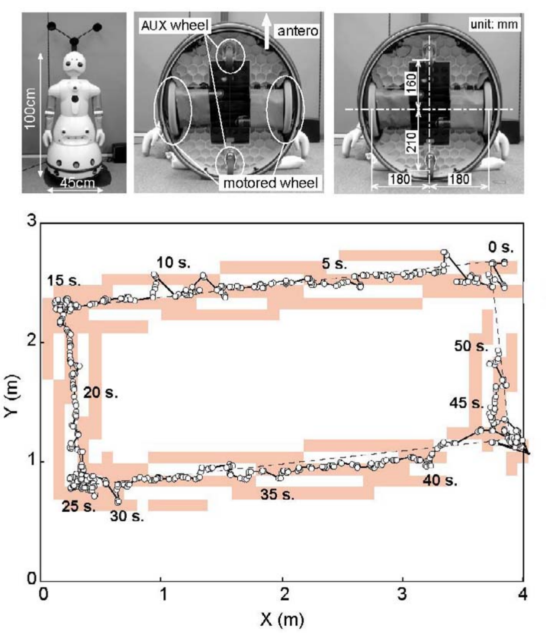


Figure 17 Tracking a four-wheeled mobile robot.

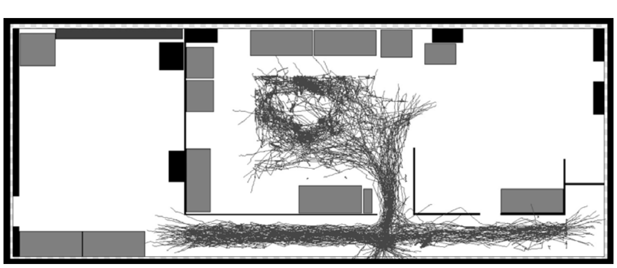


Figure 18 A history of 300 walking paths.

Further, a history of people's walking paths, as shown in figure 18, is interesting. People avoid obstacles such as walls, desks, chairs, and so forth. That implicitly teaches a robot the environmental structure. If a robot selects arbitrary paths, it will navigate successfully. Further, a robot will learn and predict human behaviors by employing the histories. That will enable the robot to communicate with people smoothly.

V. Conclusion

This paper focused on developing a real-time human-tracking model employing a floor sensor network. Human walking looks like a "rabbit hop" when it is captured using floor sensors. The signal is nonlinear and even nonholonomic; therefore, the Kalman filter is not applicable. This problem was first tackled by assuming human walking to be regular, and the signal was modeled with four walking parameters: cycles, phases, directions, and strides. The model was then multi-hypothesized to adjust the parameters to irregular walking. Experimental results showed that the multi-hypothesized oscillation models exhibited more than 80% tracking accuracy for four walking patterns: straight walking, stop-and-go, turns, and winding walking. Further, the models were superior to a conventional nearest neighbor filter with regard to data association for two persons. From the results, it is concluded that our tracking system is reliable if people keep more than a 50-cm space between them and do not enter the tracking area simultaneously, which may be a common tracking situation.

References

[1] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots," *Artificial Intelligence*, vol.128, no.1–2, pp.99–141, 2000.

[2] H. Ishiguro, "Distributed vision system: A perceptual information infrastructure for robot navigation," *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pp.36–41, Nagoya, Japan, Aug. 1997.

[3] T. Sogo, H. Ishiguro, and M. Trivedi, "Real-Time Target Localization and Tracking by N-Ocular Stereo," *Proceedings of IEEE Workshop on Omnidirectional Vision*, pp.153–160, Jun. 2000.

[4] A. Mittal and L.S. Davis, "M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene Using Region-Based Stereo," *Proceedings of European Conference on Computer Vision*, pp.18–33, Copenhagen, Denmark, 2002.

[5] R. Orr and G. Abowd, "The smart floor: A mechanism for natural user identification and tracking," *Proceedings of the 2000 Conference on Human Factors in Computing Systems*, The Hague, Netherlands, April 2000.

[6] T. Murakita, T. Ikeda, and H. Ishiguro, "Human Tracking using Floor Sensors based on the Markov Chain Monte Carlo Method," *7th. International Conference on Pattern Recognition (ICPR)*, pp.917–920, Aug. 2004.

[7] T. Mori, Y. Suemasu, H. Noguchi and T. Sato, "Multiple People Tracking by Integrating Distributed Floor Pressure Sensors and RFID System," *Proceedings of IEEE International Conference on System Man and Cybernetics (SMC)*, vol.6, no.6, pp.5271–5278, 2004.

[8] I.J. Cox, "A review of statistical data association techniques for motion correspondence," *International Journal of Computer Vision*, vol.10, no.1, pp.53–66, 1993.

[9] A. Doucet, N. de Freitas, and N. Gordon, Sequential Monte Carlo Methods in Practice, Springer, 2001.