Finding a Person with a Wi-Fi Device in a Crowd of Pedestrians

Masahiro Shiomi,[*1] Kaoru Kurumizawa,[*1, 2] Takayuki Kanda,[*1] Hiroshi Ishiguro[*2,1] and Norihiro Hagita[*1]

[1]*ATR-IRC, Kyoto, Japan*

[2]*Osaka Univ., Osaka, Japan*

Masahiro Shiomi received a Ph.D. in engineering from Osaka University in 2007. He is currently a researcher at the Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto. His research interests include human-robot interaction, interactive humanoid robots, and networked robots .

Kaoru Kurumizawa received a master's degree in engineering from Osaka University in 2011 and is currently working at Toyota. His research interests include networked robots, human-robot interfaces, and field trials.

Takayuki Kanda received a Ph.D. in computer science from Kyoto University in 2003. He is currently a senior researcher at ATR-IRC. His research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots.

Hiroshi Ishiguro received his D. Eng. degree from Osaka University in 1991. He is now a professor in the Department of Adaptive Machine Systems, Osaka University, and a group leader in the ATR Hiroshi Ishiguro Laboratories.

Norihiro Hagita received a Ph.D. in electrical engineering from Keio University in 1986. He is now a director of ATR-IRC. His research interests include communication robots, networked robot systems, interaction media, and pattern recognition.

Finding a Person with a Wi-Fi Device in a Crowd of Pedestrians

This paper proposes a technique to find and track persons carrying Wi-Fi devices. Our method integrates Wi-Fi signal strength information with LRF-based people-tracking results and estimates device holders from the statistical similarity of a time series of observed Wi-Fi signal strengths and predicted signal strengths based on the positions provided by people-tracking. Since human bodies significantly hinder radio transmission, our algorithm considers the Wi-Fi device location of the holding person. Experiment result revealed that our developed algorithm successfully found the device's holder with 87.2% successful identification rate in a real environment.

Keywords: personal identification, sensor fusion, Wi-Fi positioning

1 Introduction

For a social robot, one basic interaction among people is greetings [1]. For instance, a person recognizes a friend at 10-20 meters, approaches him, and addresses him from a few meters: "Hello, Mr. Yamada." We expect similar person-identification and greeting capabilities from robots (Fig. 1). In fact, previous studies have revealed that people appreciate such name-identifying greeting from them [2]. We expect that greetings will be frequently used by robots that serve as shopkeepers, porters, office co-workers, etc.

For a robot to greet a person from a distance, it needs to simultaneously localize and identify that person from a distance. However, no research satisfies accurate localization, robust identification, and user convenience. These requirements are important factors to realize useful applications in a real environment. Much past research has suggested techniques to realize infrastructure that accurately locates people. Methods employing laser range finders (LRFs) [3], camera vision with gaits sensing [4], ultra-wideband (UWB) [5], and Wifi based localization [6-7], techniques are well-known for their effectiveness in precisely positioning people. But these methods are

unsatisfactory for identifying individually tracked persons. The LRF method has difficulty identifying individuals, vision recognition is quite sensitive to changes in the amount of light, occlusion and a point of view of camera, and the UWB method requires that users be equipped with extra unusual devices.

In this paper, as one solution to realize physical services for targeted users in a real environment, we propose a method that positions and identifies individuals by integrating two kinds of information: signal strength from Wi-Fi devices measured at an access point (AP) and the positions of people accurately identified by position infrastructure. Note that we use multiple LRFs as position infrastructure.

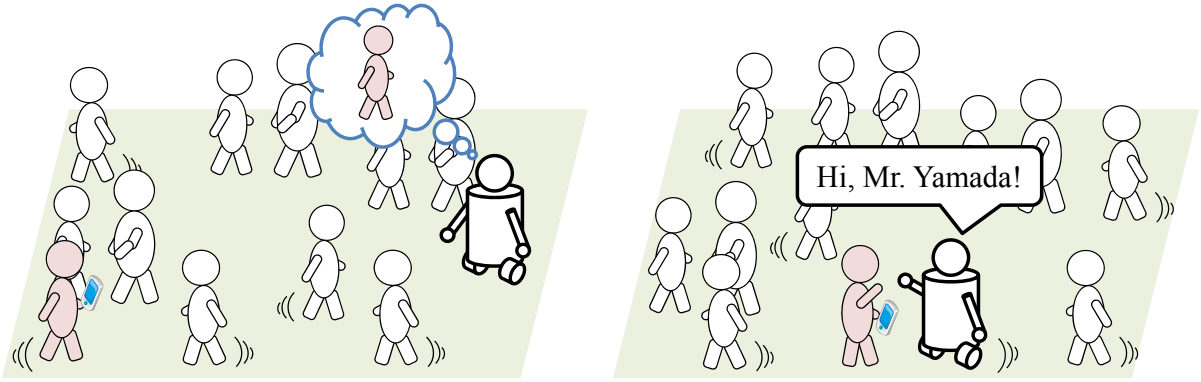


Fig. 1 Finding a Wi-Fi device holder in a flow of pedestrians

2 Proposed Method

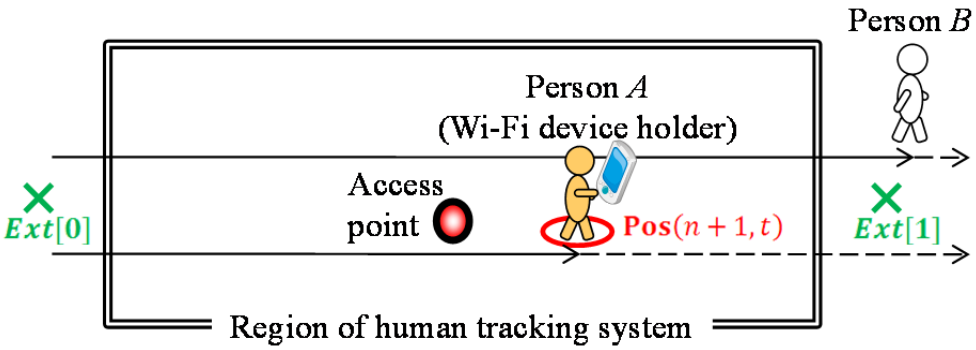
2.1 Overview of algorithm

Our method compares the time series of the observed and predicted signal strengths to find a Wi-Fi device holder (hereinafter holder). Fig. 2 shows an example of the transition of the observed and predicted signal strengths when two pedestrians are passing through an environment. By comparing the similarity between the predicted and observed signal strengths, we can tell which person is likely a holder.

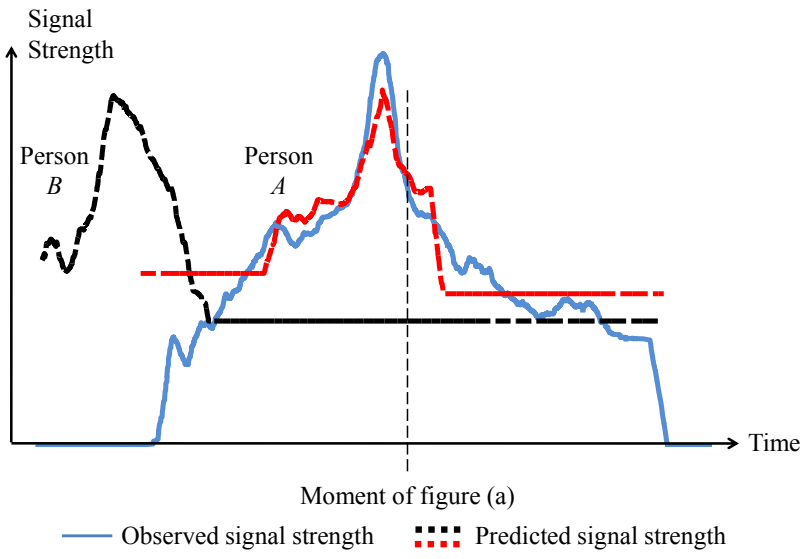
To predict signal strength, we used fingerprinting approaches [6-7]. These research works enabled systems to localize Wi-fi devices by gathering Wi-fi signal

strength with particle filter-based algorithm [6] or signal propagation modelling [7]. We extended such fingerprinting approach and included the influence of the device orientation, which expresses where the Wi-Fi device is held relative to the body. We also consider the shielding effect from human bodies [8].

Figure 3 shows an overview of the system. Positions are tracked with LRFs (Section 2.2). The tracking results are combined with the hypothecation of the device orientation (Section 2.3) to predict the signal strength for pairs of people and relative device orientation. For the prediction, we measured the signal strength distribution considering the position and device orientation (Section 2.4). The predicted signal strength is compared with the observed signal strength to identify a holder (Section 2.5).



(a)



(b)

Fig. 2 Comparison of observed/predicted signal strength

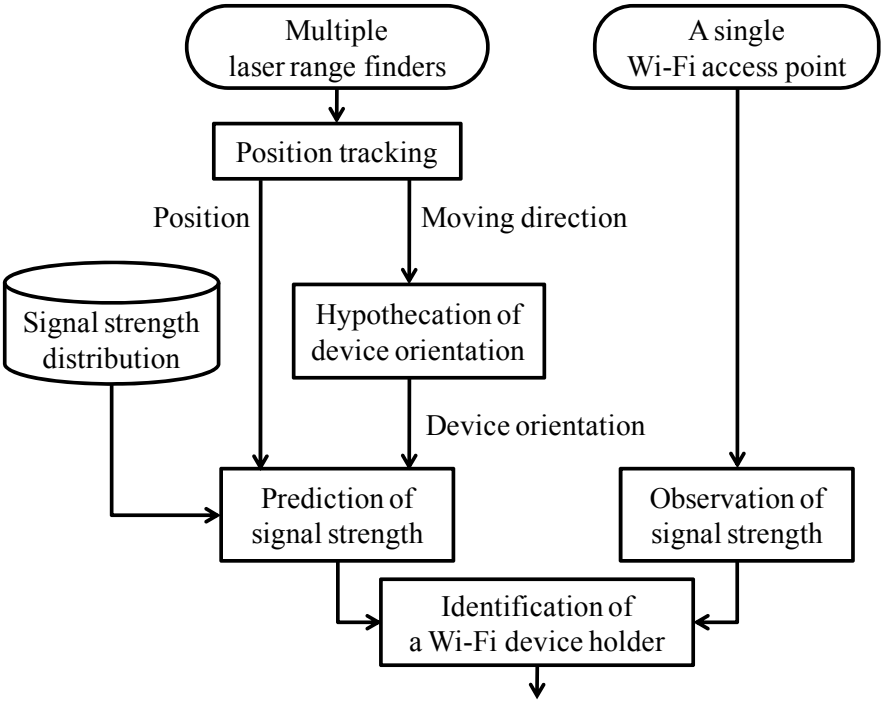


Fig. 3 Overview of proposed algorithm

2.2 Position tracking

We use an algorithm developed by Glas et al. [3] for tracking people with LRFs that perform shape matching at torso-level with a particle filter method. The position error with this algorithm averaged 6 cm, the frequency is 30 msec.

$$\boldsymbol{P}_n(t) = \begin{cases} \mathbf{Pos}(n,t) \ \ (T^n_{\text{first}} \le t \le T^n_{\text{last}}) \\ \text{NearestNeighbor}_{Ext\in\{\boldsymbol{Ext}[\mathbf{0}],\boldsymbol{Ext}[\mathbf{1}]\}}(Ext, \mathbf{Pos}(n, T^n_{\text{first}})) \ (t < T^n_{\text{first}}) \\ \text{NearestNeighbor}_{Ext\in\{\boldsymbol{Ext}[\mathbf{0}],\boldsymbol{Ext}[\mathbf{1}]\}}(Ext, \mathbf{Pos}(n, T^n_{\text{first}})) \ \ (t > T^n_{\text{last}}) \end{cases}, \quad (1)$$

where $P_n(t)$ is the x-y position of person *n* at time *t*, **Pos** (*n,* t) gives the position vector of person *n* at time *t*, T^n_{first} and T^n_{last} mean the times when person *n* appears and disappears, NearestNeighbor (*v,u*) gives vector *v*, which is nearest *u*, and *Ext* is a position vector that expresses the system's external regions. We set group *Ext* in the east

and west exits of the system, because in our experimental environment people mainly exited from the measured area to the west or the east.

Thus, the lower two substitutions indicate that $\boldsymbol{P}_n(t)$ takes an external position vector that is the nearest one to the first/latest position of person *n*:

$$\theta_n(t) = \begin{cases} \mathrm{Dir}(n,t) & (\text{for } T^n_{\mathrm{first}} \le t \le T^n_{\mathrm{last}}) \\ \mathrm{Dir}(n, T^n_{\mathrm{first}}) & (\text{for } t < T^n_{\mathrm{first}}) \\ \mathrm{Dir}(n, T^n_{\mathrm{last}}) & (\text{for } t > T^n_{\mathrm{last}}) \end{cases}, \quad (2)$$

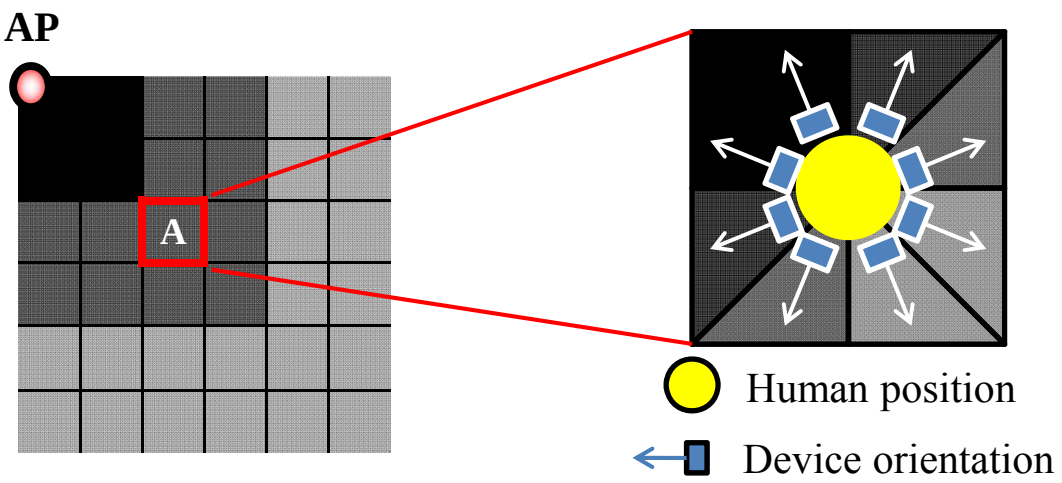
where θ_n is the history of person *n*'s moving direction and Dir(*n, t*) gives an angle of his/her moving direction at time *t*.

2.3 Hypothecation of device orientation

To estimate the device orientation, we designed our system to maintain eight hypotheses, each of which expresses a direction relative to the person's body, where the device might be held. Fig. 4 illustrates an example distribution of the signal strength. The Wi-Fi signal strength is measured in relation to the locations on an x-y grid in an environment; in this paper, we used 50-cm square as a grid. Fig. 4(b) takes grid A from Fig. 4(a) and shows how the angle segments differentiate the signal strength. In the upper-left segment of Fig. 4(b), if a device is held toward the AP, larger signal strength is observed. We note the signal strength are 10~20% different due to the direction of the device, even if the positions are the same. Moreover, the system ignores shielding by other people who are between a holder and the AP.

We define the hypotheses of the device orientation as *RelativeDeviceOrientation* [k], where *k* means the number of candidates of the relative device orientation. We depicted the derivation of device orientation $\varphi_n(k,t)$, which is an angle in the radian of person *n* at time *t* with the *k*-th hypothesis:

$$\varphi_n(k,t) = RelativeDeviceOrientation\,[k] + \theta_n(t). \quad (3)$$



(a) Predicted signal strength on each grid (b) Predicted signal strength on each segment

Fig. 4 Examples of signal strength distribution. Color strength on each grid represents predicted signal strength.

The system always keeps *k*-th hypotheses of the device orientation per one person during calculation. In other words, the system calculates *k*-th effect sizes for one person in every time steps. Different from positions and orientations, the system did not narrow the candidates of RelativeDeviceOrientatoin because we assumed immediately changes of device orientation due to Wifi device holders' actions, such as picking up the device from their pockets and insert it to other pocket. It is difficult to track the device orientation accurately by using signal strengths only.

2.4 Prediction of signal strength

2.4.1 Preparation

To predict the signal strength, we recorded it beforehand with a synchronized position and device orientation; i.e, our system records tuples of the form (*pos, orientation, SS*). From the data set, the signal strength at a specific position and device orientation can be

extracted as SS (*pos, orientation*). The system uses the averaged values in each grid and the angle segment as the predicted signal strength:

$$\underset{\{pos|Grid(\boldsymbol{P})\},\{orientation|Segment(\varphi)\},}{\mu SS(\boldsymbol{P},\varphi)} = \text{Average}\,(SS(pos, orientation)) \tag{4}$$

where *pos* is a subset of Grid (P) that gives a grid including position P and the orientation is a subset of *Segment* (φ) that gives an angle segment including angle φ. The system predicts a standard deviation of the predicted signal strength at each grid and angle segment:

$$\underset{\{pos|Grid(\boldsymbol{P})\},\{orientation|Segment(\varphi)\}.}{\sigma SS(\boldsymbol{P},\varphi)} = \text{SDev}(\text{SS}(pos, orientation)) \tag{5}$$

The value of (5), which can be regarded as a normal variation range, is used to investigate whether the difference between the predicted and observed signal strength is within the bound of the standard deviation.

2.4.2 Computing predicted signal strength

Our system predicts the signal strength from the position and device orientation. This prediction is done for each pair of person *n* and hypothesis *k* with the signal strength distribution. The predicted signal strength at time *t* for pair (*n, k*) and the standard deviation are given as

$$PrdSS_n(k,t) = \mu SS\big(\boldsymbol{P}_n(t), \varphi_n(k,t)\big). \tag{6}$$

2.4.3 Observation of signal strength

Our system observes the time-series signal strength at an AP. We defined an expression of the observed signal strength as

$$\widehat{ObsSS}(d,t) = \begin{cases} ObsSS(d,t) & \left(t^d_{\text{first}} \le t \le t^d_{\text{last}}\right) \\ 0 & \left(t < t^d_{\text{first}} \text{ or } t > t^d_{\text{last}}\right) \end{cases}, \quad (7)$$

where $\widehat{ObsSS}(d,t)$ expresses the observed signal strength of device *d* at time *t* and t^d_{first} and t^d_{last} denote the times when the communication between *d* and AP starts and ends.

2.5 Identification of a holder

2.5.1 Proposed method to compare time-series data

To measure the strength of the relationship between two variables in a statistical population, we chose Residual Mean Square (RMS), which is sometimes described as the Effect Size. It is often used for such purposes and is calculated by the following expression:

$$RMS = \frac{1}{m}\sqrt{\sum_{t=1}^{m} r(t)^2}, \quad (8)$$

where *r*(*t*) is a residual value at time *t* between two kinds of time-series data. In this paper, we apply (8) to compare the time-series transitions and set *r* as the difference between the predicted and observed signal strengths at time *t*:

$$r_n(k,d,t) = PrdSS_n(k,t) - \widehat{ObsSS}(d,t). \quad (9)$$

We used the absolute value of the residual instead of the squared value.

Consequently,

$$PrdErr_n(k,d,t) = \frac{|r_n(k,d,t)|}{Prd\sigma_n(k,t)} \quad (10)$$

shows the prediction error that indicates the mismatch in the standard normal distribution between the prediction and the observation because we assume that noise of the observed signal strength is generated based on normal distribution. If the prediction is theoretically correct, meaning when (*n*,*k*) is a correct pair of an actual holder and a real device orientation,

$$PrdErr_n(k,d,t) \sim N(0,1). \tag{11}$$

Finally, the following expression indicates the evaluation value:

$$Err_n(k,d) = \frac{1}{m}\sqrt{\sum_{t=t_{\min}}^{t_{max}} (\ PrdErr_n(k,d,t)\ )^2}, \tag{12}$$

where

$$t_{min} = \min\left(T^n_{\text{first}}, t^d_{\text{first}}\right). \tag{13}$$

$$t_{max} = \max\left(T^n_{\text{last}}, t^d_{\text{last}}\right). \tag{14}$$

Here, $Err_n(k,d)$ indicates a mismatch over the time between the prediction for each human and the actual observation. The smaller value of $Err_n(k,d)$, given by pair (n,k), the more likely that person n is a holder and that hypothesis k is the actual *relative device orientation*.

2.5.2 Judging whether the holder is in the environment

When the holder has not yet been observed by the tracking system, the value of the effect size is often high since the predicted signal strength does not really match the observed signal strength. In such situations, to avoid misidentifying persons as holders, we set a threshold. Here, we define (n^*_d, k^*_d) as a pair of a person and a device orientation hypothesis identified as the holder of *x* and the actual device orientation:

$$(n^*_d, k^*_d) = \begin{cases} (\hat{n}, \hat{k}) & \left(\text{if } minimumErr_d(\hat{n}, \hat{k}) \le TH\right) \\ (\text{nobody}, \text{null}) & (\text{otherwise}) \end{cases}, \tag{15}$$

where

$$minimumErr_d(\hat{n}, \hat{k}) = \underset{\{n|N\},\{k|K\}}{\operatorname{argmin}} Err_n(k,d). \tag{16}$$

The above formulation means a conditional branch. If the minimum effect size among all pedestrian candidates does not exceed certain threshold *TH*, the pair argument is identified as the holder of Wi-Fi device *d* and his/her device orientation.

Otherwise, the system judges that the holder is not among the candidates because the large effect size means that the signal strengths predicted for the detected people are quite different from the observed one. *TH* denotes a threshold for rejecting a misidentification of a person who is not carrying the device, which typically happens when the owner has not arrived within the tracking system. We empirically set threshold TH 1.28 to improve the performance in our settings.

3 Evaluation

3.1 Data collection

We collected data at a shopping mall (Fig. 5). Most people pass directly through the arcade, and some visit the shops or rest on chairs. We recorded the position information and the wireless signal strengths. To record the signal strength, we only arranged a single AP; we used a laptop computer with a FreeBSD operating system as the AP and used a conventional Wi-Fi device (iPhone, Apple). The observed signal strength was updated every 10 msec.

We randomly recruited two men from the web. We asked them to freely walk through the area's corridor. They started from one base point and moved to the other (Fig. 6). To investigate the shielding effect of the holder's body, we specified where they should hold the device in every trial among four ways: front (coat pocket), behind (hip pocket), left, and right of their bodies (pockets of pants). We conducted 30 tests for each way of holding the Wi-Fi device (120 tests). On average, 9.5 people, including our participants, existed in the environment.

Fig. 5 Shopping arcade

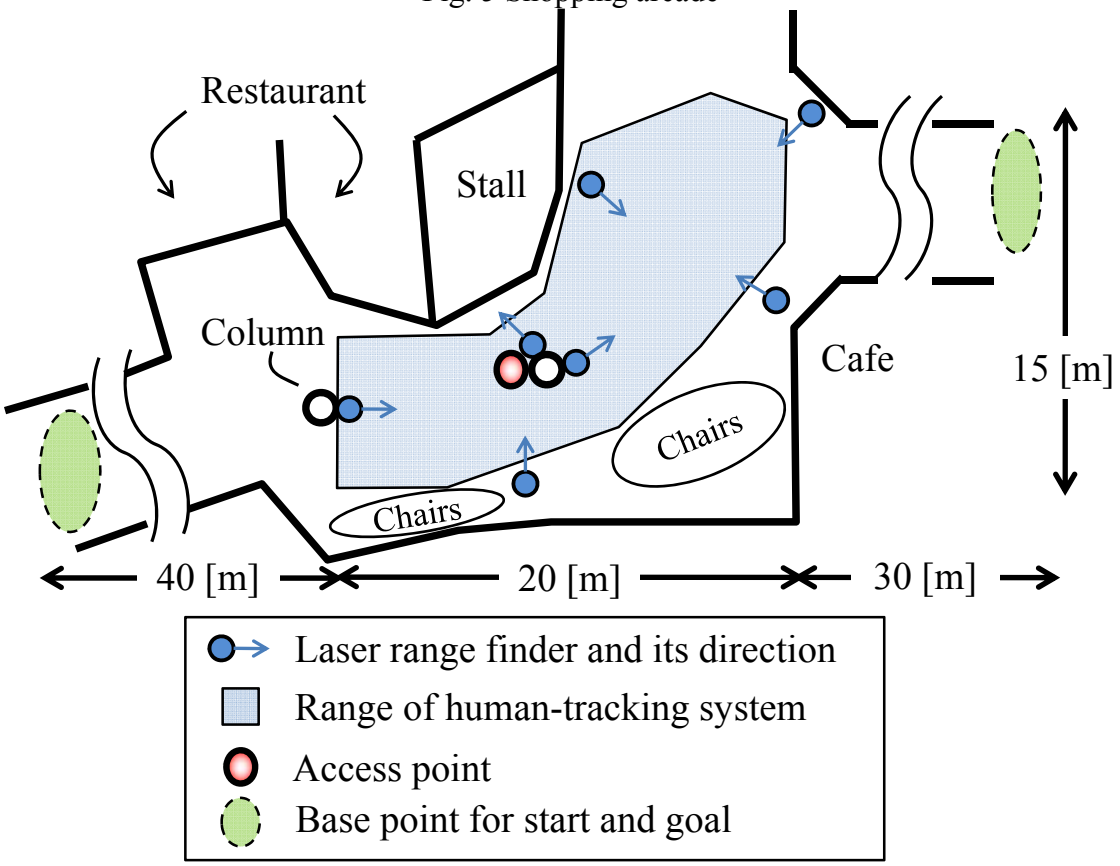


Fig. 6 Environment and sensor arrangement

3.2 Performance evaluation

In evaluations, we used leave-one-out cross-validation. We defined the critical times in the participant movements in Fig. 7(a). The t_0 is the minimum value of $t_0(n)$ (n=1~120),

the timing when a Wi-Fi device started to communicate with AP at n-th test. The t_1 is the time when the participant entered the range of the human-tracking system. The t_2 is the maximum value of t_2 (n) (n=1~120), the timing when the Wi-Fi device loses communication at n-th test. The t_1 is used as a standard point in below figures.

3.2.1 Successful identification rate

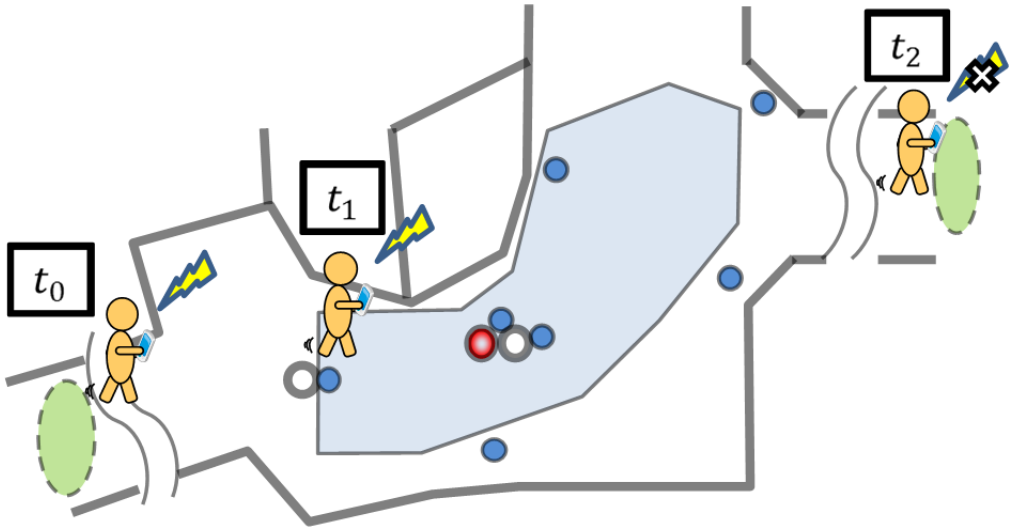
Figure 7(b) shows the rate at t_2, which is the final calculation step. The definition of success means that the system correctly identified holder at each time step. To investigate whether our system can accurately identify holder more than alternative method, we used all length of data set. At this time, the rate is 87.2%, meaning that the proposed method can locate the holders with 5.8 cm accuracy (the ground truth of the human-tracking system) at 87.2%. We note that there is no false negative; the wifi devices correctly connected to AP in all cases.

Identification failure was caused by two reasons: another person walking closely to the participant and tracking system error. The former indicates that it is difficult to identify the participant when someone else walks nearby for a long time; the person has a similar effect size resulting from movements similar to the participant. The average distance from the participant to the person was around 1.3 m. As for the latter reason, sometimes the human tracking system mistakenly assigns ID numbers to the participant and others. When a participant enters the area and a person leaves the area near him, the human tracking system may assign the same ID number of the disappearing person to him. Therefore, the system cannot correctly calculate the effect size.

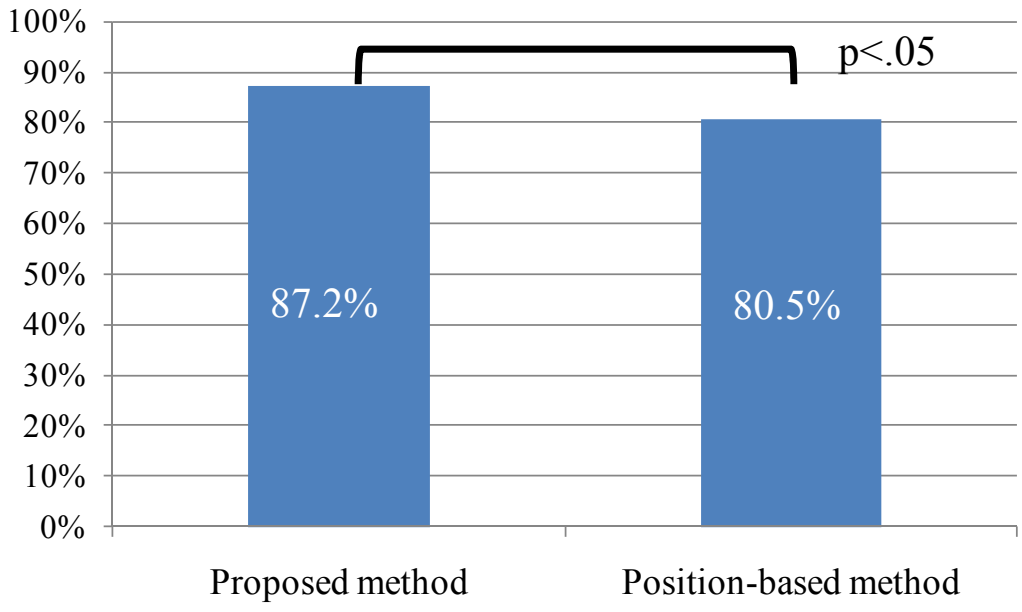
In Fig. 7(c), we show the time-series transition of the rate. At the initial second, the rate of the proposed method is 77.5%. The identification results did not change after the participant leaves the area (in average they leave the area about 30 sec).

3.2.2 Comparison with position-based method

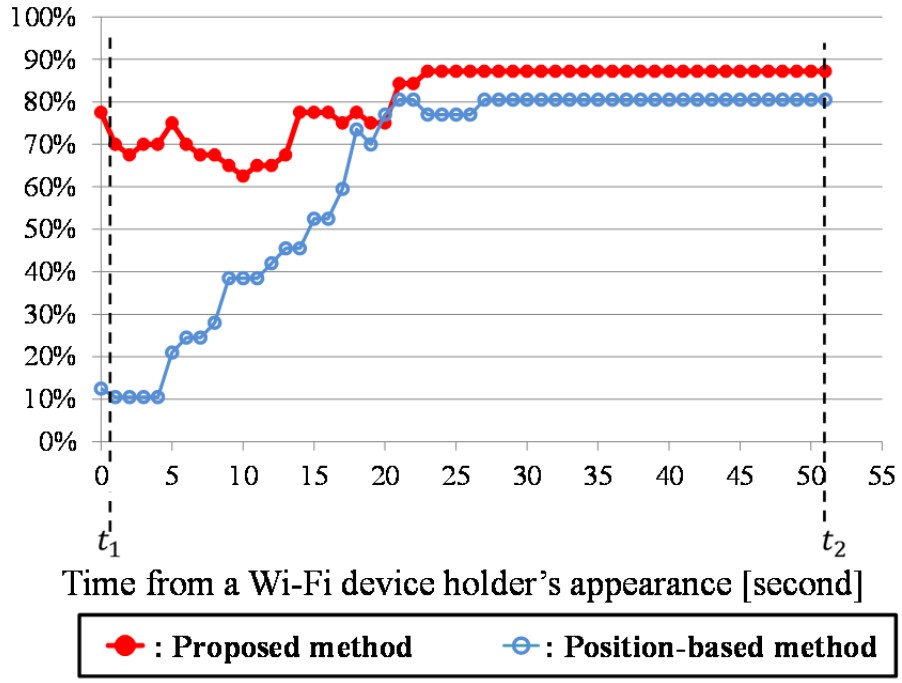
We compared the results with an alternative method that does not use our model: a position-based method. This method only uses positions to predict signal strength without device orientations. The rate at t_2 by the position-based method is also shown in Fig. 7(b). It is 80.5% and showed a significant difference against the proposed method from a chi-square test ($p = .035$, $\chi 2\ (1) = 4.45$, $\phi = .20$).



(a) Movement of a holder and definitions of times



(b) Successful identification rate at final calculation step



(c) Transition of successful identification

Fig. 7 Successful identification rate

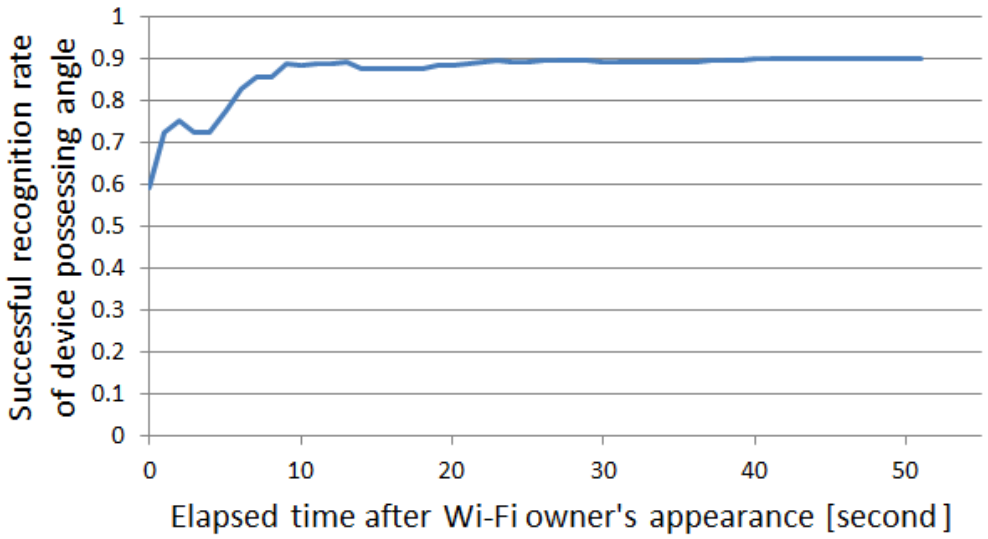


Fig. 8 Recognition rate of *relative device orientation*

3.2.3 Success rate of relative device orientation

Figure 8 shows the transition of the success rate of the device orientation. At time 0, i.e., t1, it was 59.3%; however, it gradually increased to 90.0%. This is because the system obtains more information as the participant approaches the AP.

3.2.4 False-positive rate

We also measured the false-positive rate. In this section, we only describe the false-positive rate when the participants did not enter the human-tracking system's range. Thus, false-positive means that the system misidentified a person even if the holder did not enter to the area.

The rate (Fig. 9) is 0% at time t_0, but it gradually increases. Consequently, false-positives reached 70.0% at t_1; but the rate immediately decreased after t_1. We note that the number of data set around t_0 in this figure is smaller than the number of data set around t_1, because the length of each data set is different. But the average difference between t_0 and t_1 in all data set is a few seconds; therefore it would not be critical problem to show the trend.

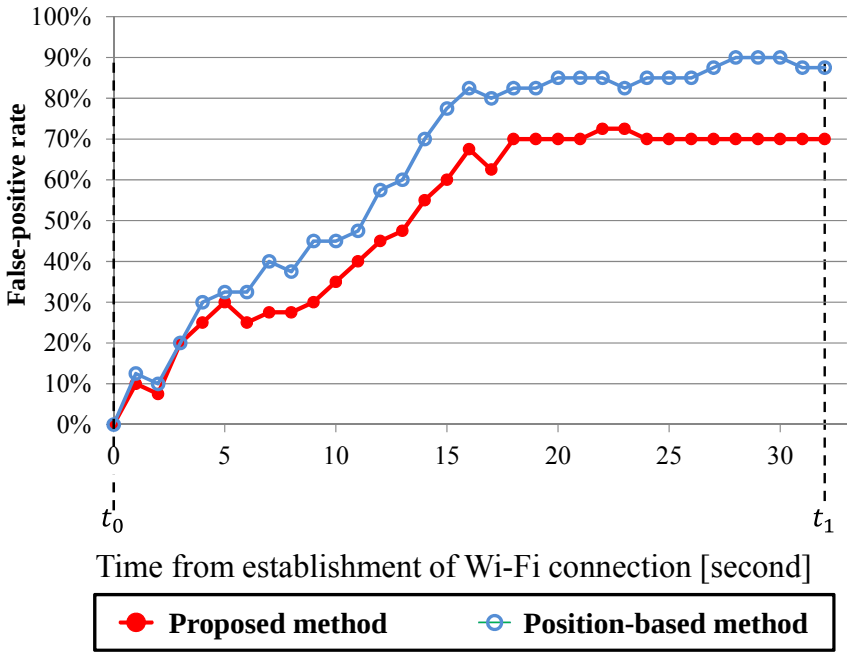


Fig. 9 False-positive rate

(a) Robot greeting a holder

(b) Robovie-II

Fig. 10 Using our developed system

4 Example Use of Developed System

We investigated the contribution of our developed system by a social robot by developing a simple application in combination with a robot to demonstrate one possible use of the system. The robot approaches and greets a specific customer who is carrying a Wi-Fi device (Fig. 10(a)). We used Robovie-II, which uses a Pioneer 3-DX as a lower mobile base (Fig. 10(b)). The system first tried to find a holder (i.e., the robot was waiting until the system find the target), and then the robot approaches this person and guides him or her.

We conducted an evaluation where a robot approaches a Wi-Fi device (iPhone, but a different device from the data collection) holder who is walking in the environment; the holder walks from a base point to another point. We again randomly recruited two men from the web and conducted 30 trials with each participant. In the evaluation, always more than five pedestrians existed.

The robot successfully approached the holder 78.3% of the time: 47 of 60 trials. Within the 21.7% failure rate, 11.7% resulted from misidentification by the proposed system which causes wrong approach behaviour of the robot, and the other 10.0% resulted from bystanders who blocked the robot. The average elapsed time was 10.3 sec (S.D. 2.7, min: 5.7 sec, max: 13.4 sec.).

5 Conclusion

We developed a method for tracking and identifying holders with positioning accuracy of 6 cm at a success rate of 87.2%.One feature of proposed method is its prediction of signal strength by incorporating a hypothetical model of device orientation. Evaluation results show that the device orientation model improves the successful identification rate by 6.7% and that it can correctly identify holders much earlier than a model that does not consider device orientation.

Acknowledgements: We thank the administrative staff at the Asia and Pacific Trade Center for their cooperation. This research was supported in part by the Strategic Information and Communications R&D Promotion Programme (SCOPE), Ministry of Internal Affairs and Communications (132107010), and the Ministry of Internal Affairs and Communications of Japan.

References:

[1] Kendon A., Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press; 1990.

[2] Kanda T., Shiomi M., Miyashita Z., Ishiguro H., and Hagita N. A Communication Robot in a Shopping Mall. IEEE Transactions on Robotics. 2010; 26: 897-913.

[3] Glas D. F., Miyashita, T., Ishiguro H., and Hagita N. Laser-based tracking of human position and orientation using parametric shape modelling. Advanced Robotics. 2009; 23: 405-428.

[4] H. Iwama, D. Muramatsu, Y. Makihara, Y. Yagi, "Gait-based Person Authentication System for Criminal Investigation", In Proc. of the 7th Int. Workshop on Robust Computer Vision (IWRCV2013), 2013.

[5] Chung W. C. and Ha D. S. An accurate ultra wideband (UWB) ranging for precision asset location. In Proc. UWBST. 2003:389-393.

[6] Sajid Siddiqi, Gaurav S. Sukhatme and Andrew Howard, "Experiments in Monte-Carlo Localization using WiFi Signal Strength", Proceedings of the International Conference on Advanced Robotics, pages , Coimbra, Portugal, Jul 2003

[7] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '00), vol. 2, pp. 775–784, 2000.

[8] Ryckaert J., DeDoncker P., Donnay S., and Delehoye A. Channel model for wireless communication around the human body. Electron. Lett. 2004; 40-9: 543-544.