Estimating Person Positions Using a Camera and Wireless Devices in a Space with Temporary Shielding

Y. Makimura, D. Hayashi, T. Kobiki, S. Sakemoto and M. Nishiyama

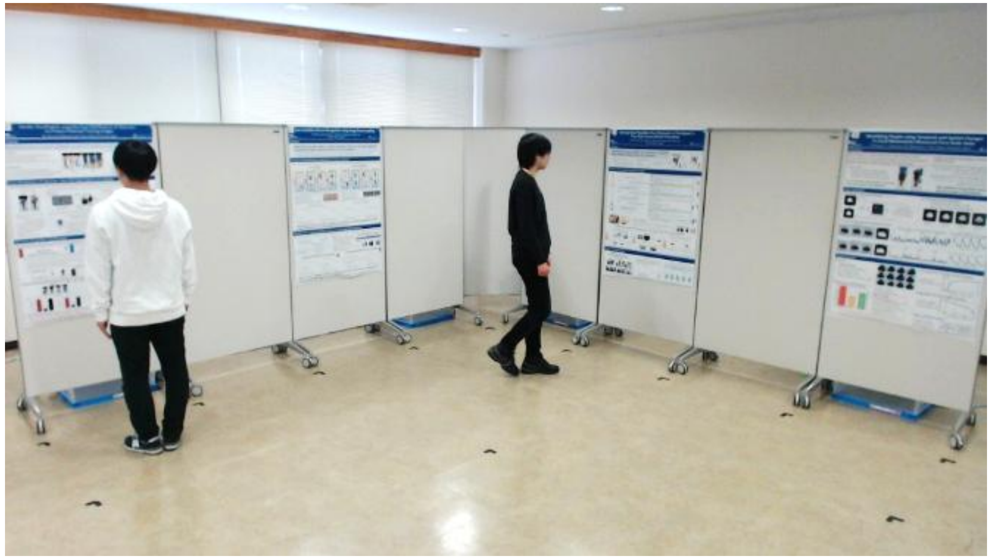
Tottori University, Japan

Introduction

There is a demand for technologies to estimate a person's position in public spaces.

Cameras that can image large spaces with high spatial resolution are commonly used as sensors to estimate an individual person's position.

Camera



Without temporary shielding

With temporary shielding

However, the camera's field of view may sometimes be blocked unavoidably by objects that have been placed temporarily.

Wireless devices as sensors

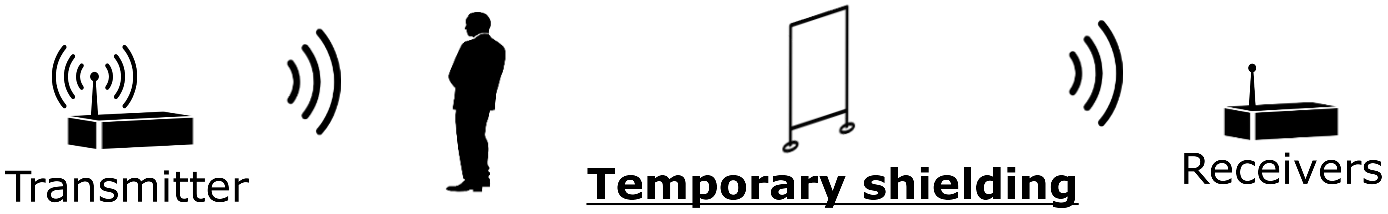
We consider whether there is a sensor configuration that can estimate a person's position stably even when temporary shielding is placed.

Research has actively explored wireless sensing devices that emit radio waves with longer wavelengths than visible light and can penetrate shielding.

Existing methods using wireless devices

- Person position estimation (Youssef et al., 2007; Booranawong et al., 2019)
- Body shape and posture estimation (Xue et al., 2021; Wang et al., 2023)

This enables reliable person position estimation using wireless devices, even when temporary shielding is placed between the transmitter and receiver.

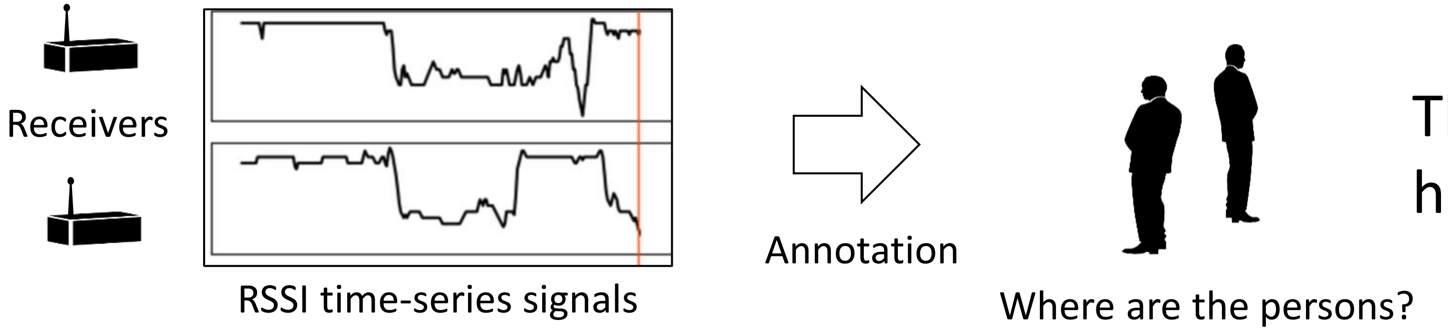


Challenges in wireless sensing

When using wireless devices as sensors, it is necessary to consider how to collect large numbers of training samples because the existing methods are based on machine learning and deep learning techniques.

However, radio signals such as RSSI that are acquired from wireless devices cannot be understood intuitively by humans because the space in this case is invisible.

RSSI: Received Signal Strength Indicator

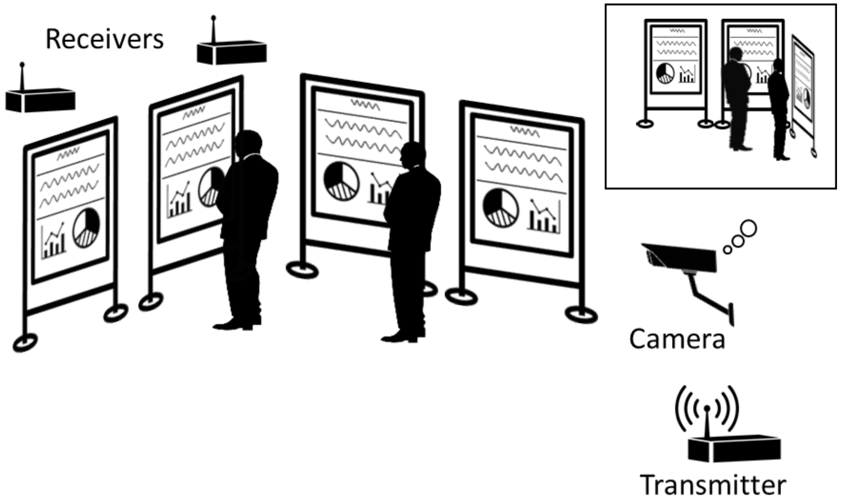


This makes it difficult to determine how to perform the annotations.

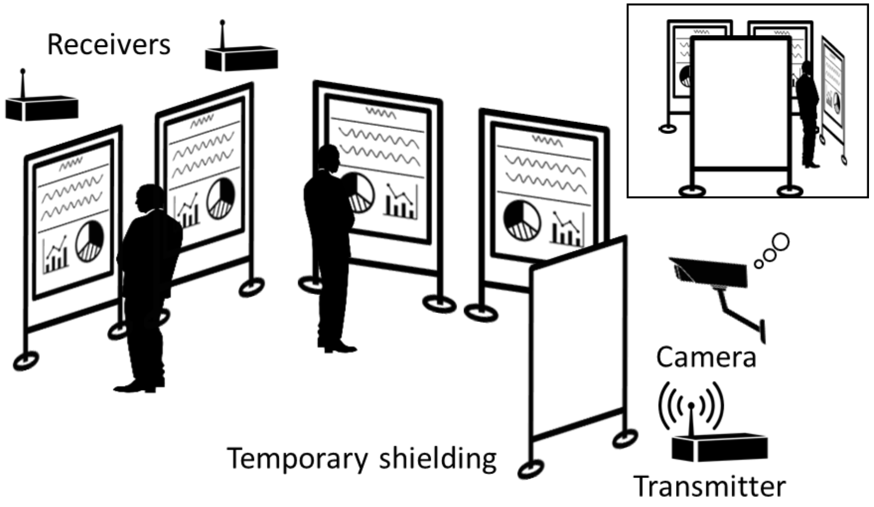
We must consider how the training samples, which consist of pairs of radio wave strength and person position data, are annotated.

Research objective

We propose a method using a camera and wireless devices that can estimate a person's position within a space even when temporary shielding is present.



Without temporary shielding
(Training phase)



With temporary shielding
(Inference phase)

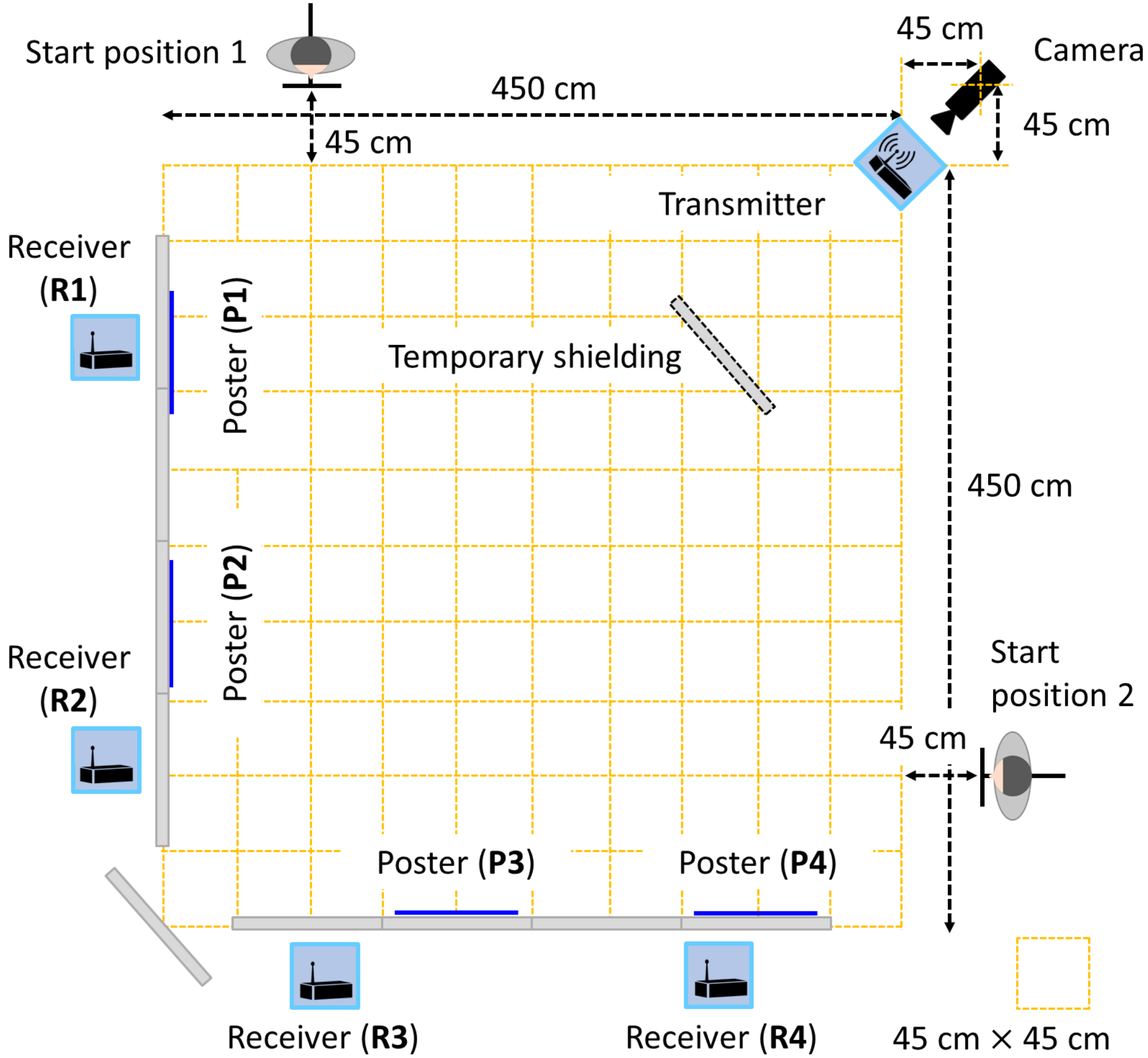
We use Wi-Fi and RSSI time-series signals, a type of wireless technology that offers the advantage of easy accessibility.

Poster Panel Area (PPA) dataset

We originally collected the PPA Dataset by considering a poster exhibition room.

Visitors may stop in front of a poster of interest and read its contents, or they may simply glance at the poster and leave if they are uninterested.

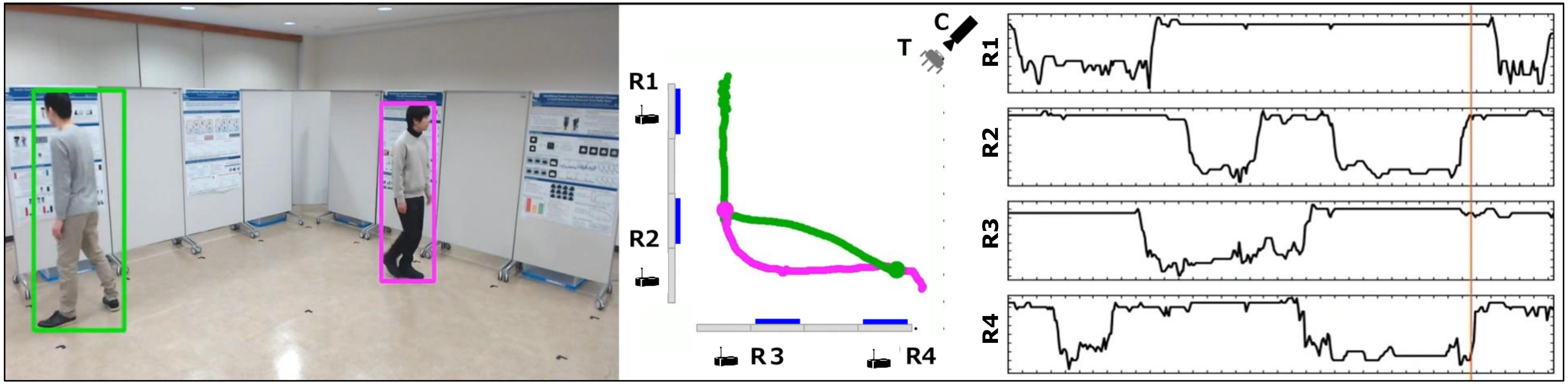
In a typical poster space, each poster has a staff member present, but for the PPA dataset, we assumed a situation in which posters are displayed but no staff member is present, e.g., during breaks.



Examples of Poster Panel Area (PPA) dataset

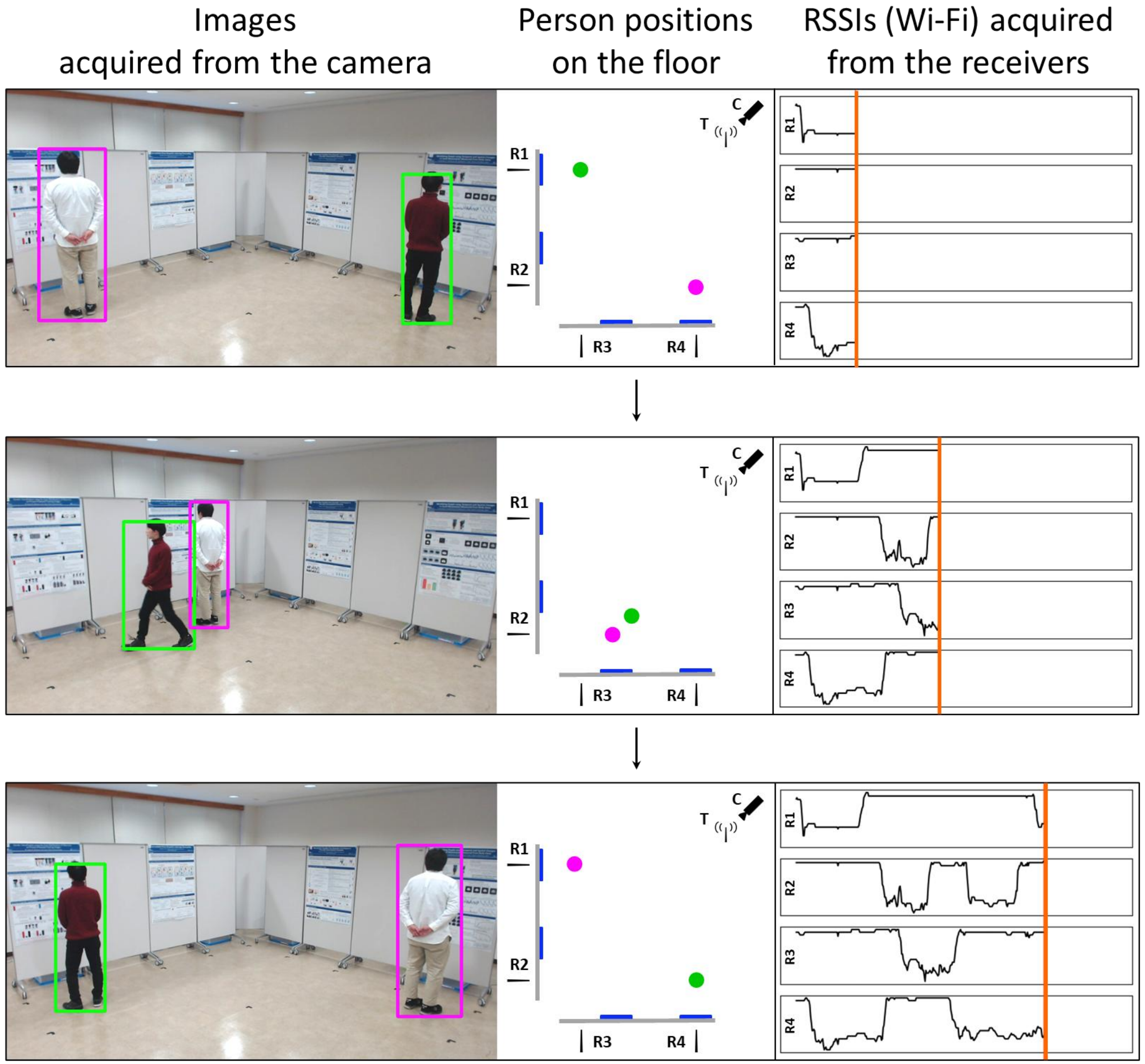
Each sequence shows the image acquired from the camera, a bird's-eye view showing each person's position, and a graph of the Wi-Fi receivers' RSSI time-series signals.

Images	Person positions	RSSIs (Wi-Fi) acquired
Acquired from the camera	On the floor	From the receivers

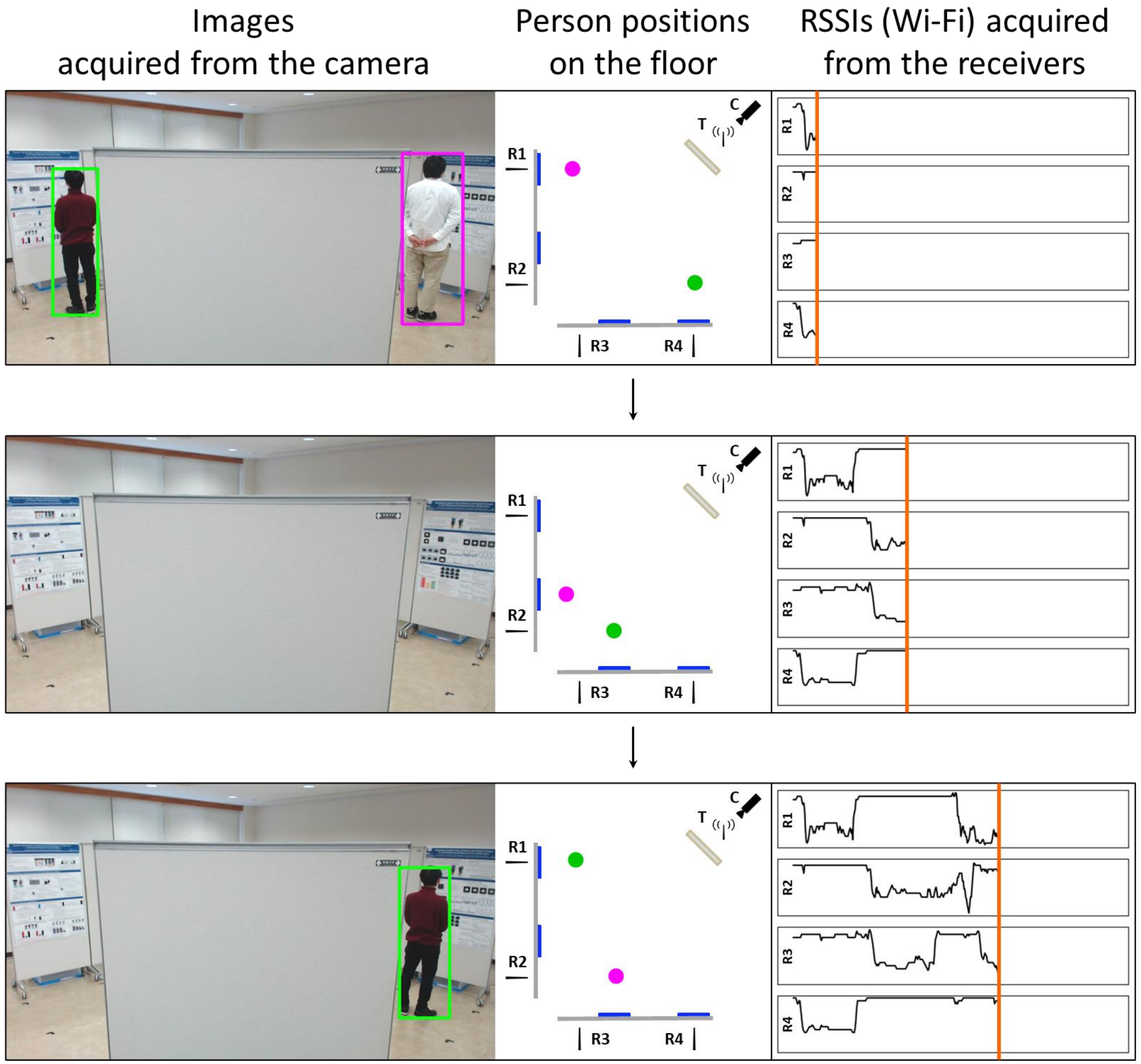


- The color of each bounding box corresponds to the color of that person's position in the overhead view.
- The bird's-eye view shows each person's position within the floor coordinate system.
- In the RSSI time-series graph, the horizontal axis represents time and the vertical axis represents the RSSI values acquired from the Wi-Fi receivers.

Difference between w/o and w/ temporary shielding



Without temporary shielding (Training phase)



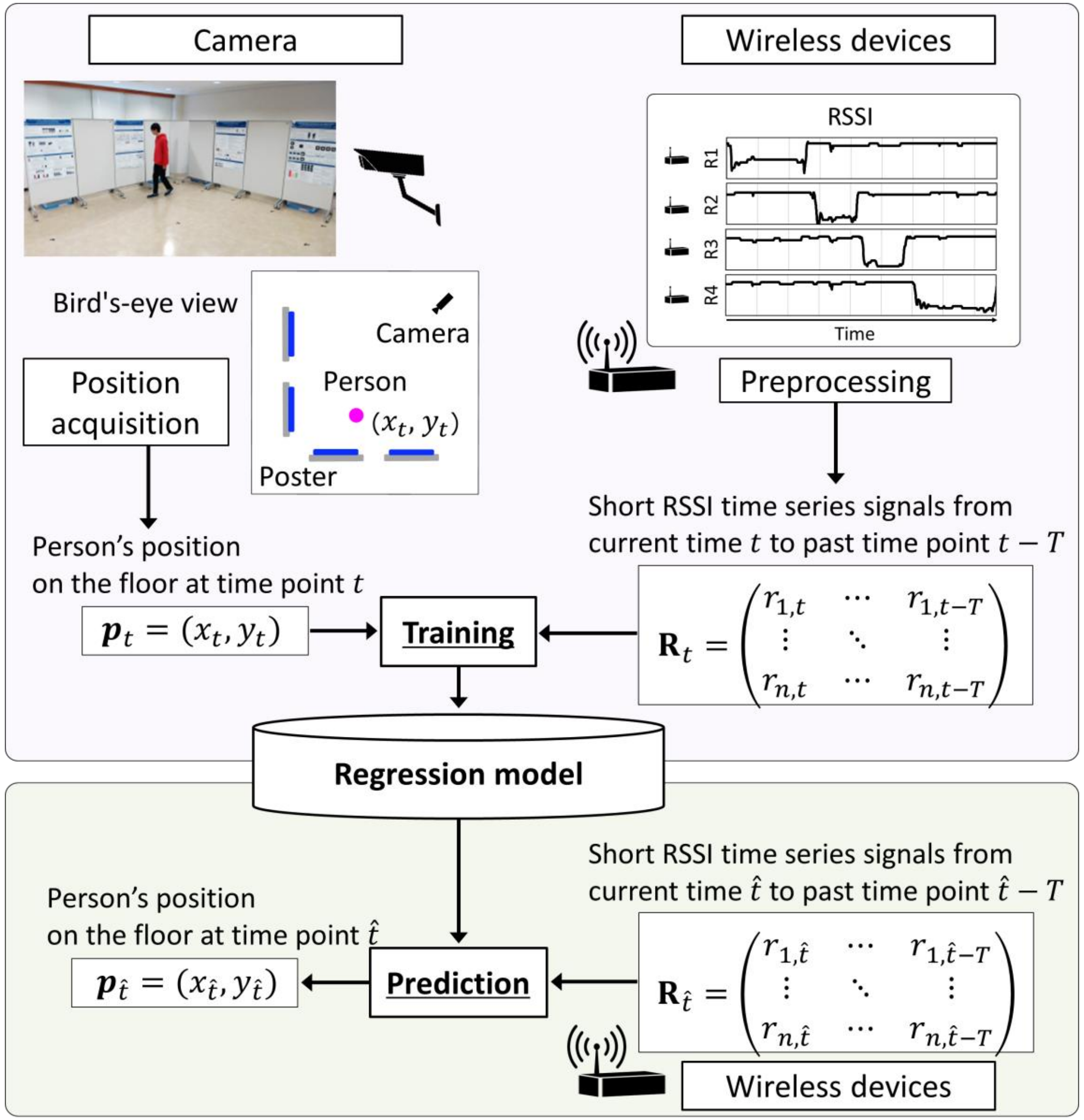
With temporary shielding (Inference phase)

Proposed method

Our method estimates each person's position using a regression model that only uses the RSSI time-series signal during temporary shielding.

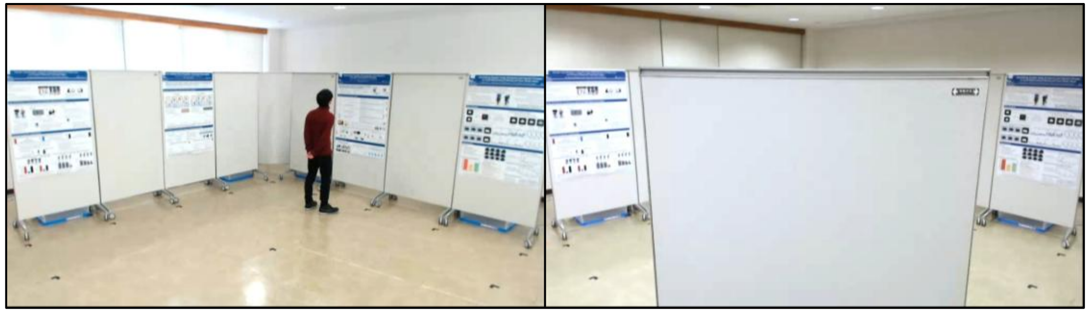
When no shielding is present, our method collects training samples composed of pairs of person positions and RSSI time-series signals and trains a regression model.

When temporary shielding occurs, our method estimates the person's position using only RSSI time-series signals and the trained regression model.

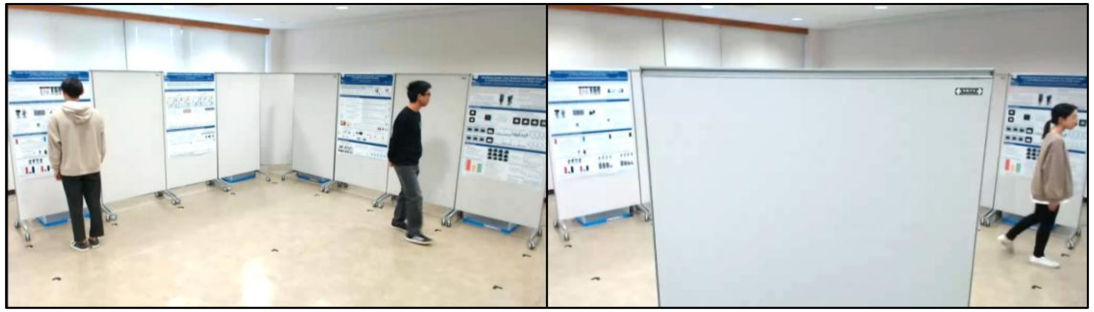


Experiments

We evaluated the accuracy of person position estimation using the PPA dataset when one or two persons were walking in the space.

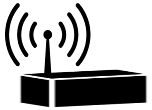


One person



Two persons

Camera



Wi-Fi

We applied leave-one-person-out cross-validation.

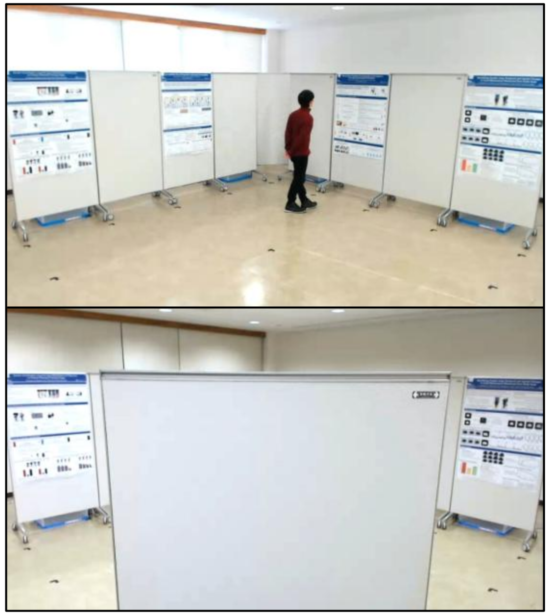
We measured the error between the predicted and ground-truth positions at each time point and compared the errors across regression models.

To investigate the impact of preprocessing, including noise and peak removal, we compared performance with and without preprocessing of the RSSI time-series signals.

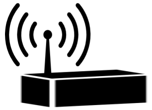
Results (One person)

Error (in cm) in person position estimation when one person was walking in the space

Regression Model	One person	
	W/O Preprocessing	W/ Preprocessing
LR: Linear Regression	27.0±7.8	26.6±7.6
MLP: Multilayer Perceptron regression	18.5±6.7	16.2±4.7
RF: Random Forest regression	20.4±10.2	**12.4±3.8**
XGB: XGBoost regression	20.5±9.2	13.1±3.8
*k*NN: k-Nearest Neighbor regression	17.9±5.9	15.9±5.2



Camera



Wi-Fi

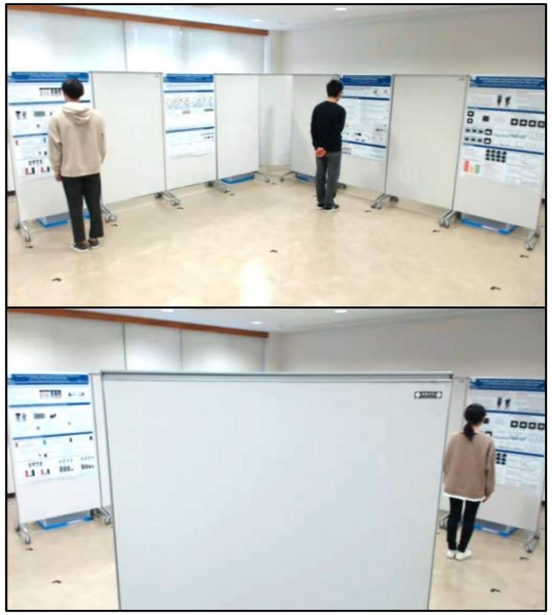
Among the regression models, MLP, RF, XGB, and *k*NN had reduced errors when compared with LR.

The error was 12.4±3.8 cm for our method using RF and preprocessing when only one person was walking within the space.

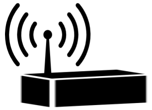
Results (Two persons)

Error (in cm) in person position estimation when two persons were walking in the space

Regression Model	Two persons	
	W/O Preprocessing	W/ Preprocessing
LR: Linear Regression	61.7±11.2	64.1±10.1
MLP: Multilayer Perceptron regression	43.0±11.3	38.1±8.6
RF: Random Forest regression	48.2±14.5	36.6±11.9
XGB: XGBoost regression	51.1±18.3	40.5±11.8
*k*NN: k-Nearest Neighbor regression	22.0±9.6	**18.7±6.6**



Camera



Wi-Fi

For all regression models, the error increased when compared with the case where only one person was walking, as described earlier.

Only kNN showed a small increase in the error, with a mean error of 18.7±6.6 cm, whereas LR, MLP, RF, and XGB all showed significant increases.

Results

Error (in cm) in person position estimation

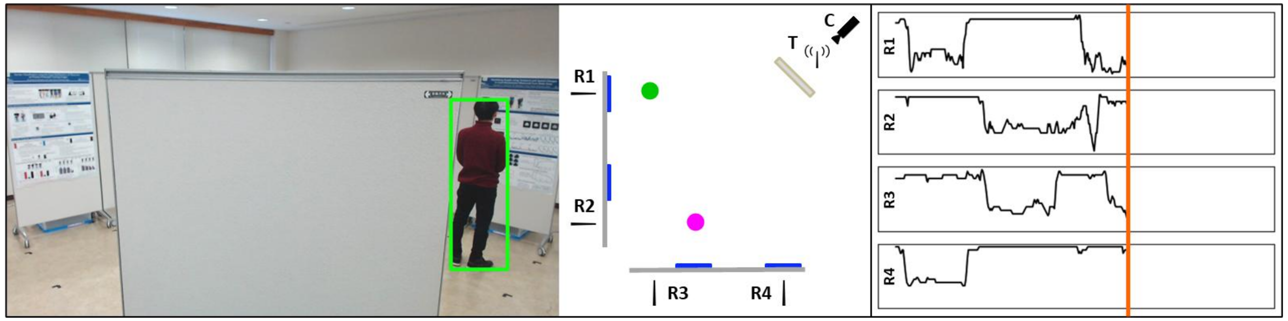
Regression Model	One person		Two persons	
	W/O Preprocessing	W/ Preprocessing	W/O Preprocessing	W/ Preprocessing
LR	27.0±7.8	26.6±7.6	61.7±11.2	64.1±10.1
MLP	18.5±6.7	16.2±4.7	43.0±11.3	38.1±8.6
RF	20.4±10.2	**12.4±3.8**	48.2±14.5	36.6±11.9
XGB	20.5±9.2	13.1±3.8	51.1±18.3	40.5±11.8
*k*NN	17.9±5.9	15.9±5.2	22.0±9.6	**18.7±6.6**

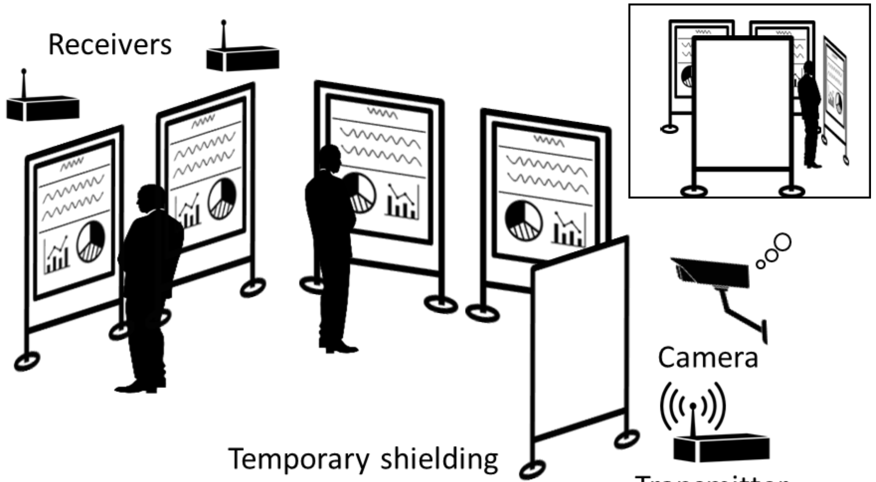
The error decreased in each case when our preprocessing step was applied when compared with the case where no preprocessing was performed.

Because the RF result for one person and the *k*NN result for two persons are both smaller than the average human foot length (about 22–27 cm), the proposed method achieves practically meaningful positioning accuracy.

Conclusions

We have proposed a method that uses a camera and wireless devices to collect training samples of the RSSI time-series signal and a person's position automatically when no temporary shielding is present and to estimate a person's position accurately using a regression model from the RSSI time-series signals when temporary shielding exists.





Future work:

- Develop a method to estimate a person's position that is independent of the number of people walking within the shielded space simultaneously.
- Expand the evaluation when more people are present and when walking patterns change.