User Equipment Tracking for a Millimeter Wave System Using Vision and RSSI

SHOICHIRO MIHARA, TAKAHIDE MURAKAMI, AKIRA YAMAGUCHI (Member, IEEE), AND HIROYUKI SHINBO

1XR Division, KDDI Research, Inc., Fujimino 3568502, Saitama, Japan
2Wireless Technology Division, KDDI Research, Inc., Fujimino 3568502, Saitama, Japan

CORRESPONDING AUTHOR: T. MURAKAMI (e-mail: tk-murakami@kddi.com)

This work was supported by the Ministry of Internal Affairs and Communications in Japan (JPIJ000254).

ABSTRACT In a mobile millimeter wave (mmWave) communication system, blockages cause disconnections or serious degradation of communications. Several techniques have been proposed to control radio links between multiple base stations based on blockage prediction using camera images to avoid these problems. However, blockage prediction requires continuously determining the position of user equipment (UE) with decimeter precision, which is difficult when sensors and resources on the UE side are not available, and there are many moving objects around the UE. To resolve this problem, we propose a UE tracking method that uses a received signal strength indicator (RSSI) and RGB-D camera images from the base station. The proposed method consists of a combination of visual tracking and reidentification of the UE using radio information and camera images. We use the synchronization of the RSSI variation with the occlusion on the image by movements of the UE and objects for reidentification. We evaluated the proposed method experimentally in an outdoor environment by simulating a communication area formed by mmWave band base stations. The proposed method achieved an 11.5% improvement in tracking accuracy compared with conventional visual tracking.

INDEX TERMS Image analysis, millimeter wave communication, object tracking, received signal strength indicator.

I. INTRODUCTION

WE ASSUME that by approximately 2025, user requirements for the performance of fifth-generation mobile communication systems (5G systems) will increase. We call the 5G system in approximately 2025 an “advanced 5G system” [1] that will satisfy the quality requirements of each communication flow to meet the various applications. For example, an automated robot requires all of the following: high-reliability, low-latency, and high-capacity communications, such as data transfers from a 4K camera and other sensors and control signals. To meet these strict requirements, millimeter wave (mmWave) band signals will be used in advanced 5G systems, in addition to the 6 GHz and lower bands that have been primarily used for mobile communication systems. However, for mmWave bands, moving objects around user equipment (UE) can lead to shadowing because mmWaves have the propagation characteristics of greater attenuation than microwaves [2]. Shadowing may block a radio link between a base station (BS) and the UE and cause disconnections or serious degradation of communications. The current 5G system uses a mechanism called handover that switches the UE connection to another BS when the received signal strength drops for a certain period of time. However, when the received signal strength suddenly drops because of mmWave blockage, handover does not work, and the radio link may be cut off. Many mmWave blockages may occur in congested outdoor places, such as downtown areas. In [3], modeling of the blockage, including self-blockage by UE holder and blockage by objects around the UE, was presented along with countermeasures for blocking. In [4] and [5], the UE design and beamforming codebooks for mitigating blockage by hand was introduced. In [6], a method was introduced to use the UE position obtained by image processing that combines information from the camera and the UE for beamforming control. Other alternative methods include coordinated...
multipoint (CoMP) and cooperative beamforming [7]. With these methods, efficient operation is achieved with accurate blockage predictions.

We have been attempting to predict blockages using an RGB-D camera in an outdoor environment with many objects, such as pedestrians and vehicles. The distinctive characteristics of our proposed method is only the use of BS side information to avoid transmitting data from the UE and the use of the RGB-D camera alongside the BS as a wide-area sensor for grasping the objects in the field. The RGB-D camera is attached to the BS, and the goal of this research is to control mmWave radio links over BSs based on a wide-area sensor for grasping the objects in the field. The use of the RGB-D camera alongside the BS as a system for transmitting the abovementioned information was sensitive to the UE position, which required decimeter accuracy for precise blockage prediction due to the short wavelength of mmWaves. Using sensors on the UE side, such as a global positioning system (GPS) and cameras, it is possible to continuously acquire the UE position. In addition, if other information, for example, velocity, acceleration, visual characteristics of the UE holder, markers associated with the UE holder, etc., is provided from the UE, it is also effective for the acquisition of the UE position [6]. However, a system for transmitting the abovementioned information from the UE to the BS must be constructed, which places a heavy load on the communication links.

Therefore, it is necessary to continuously locate the UE position using sensors and computational resources on the BS side only. First, we must identify the object holding the UE (i.e., the UE holder) and then track the UE holder. To achieve this, we proposed a method of estimating the UE position triggered by mmWave blockage [8] that uses RGB-D camera images and radio information acquired on the BS side. The position of the UE can be estimated with an accuracy of approximately 0.6 m when the UE is occluded by moving objects several times. This method can be used for initial identification of the UE holder. After initial identification, visual tracking can effectively and continuously locate the UE holder. If the UE is clearly visible in the image, tracking the UE holder is not challenging. However, in a crowded environment with many moving objects, the UE holder can frequently be occluded by other objects, and visual tracking may be interrupted, or another object may be tracked by mistake. In addition, assuming the coverage of the BS using mmWave, the shooting range is wider than that of general cameras, and objects in the distance are smaller in size in the camera image. As a result, visual tracking may not extract the features necessary to identify the same object and, therefore, may not track the object correctly.

In this paper, we propose a tracking method for UE using a combination of visual tracking and reidentification using radio information and images. We use temporal synchronization between the variations of the received signal strength indicator (RSSI) of mmWaves and the occlusions in the RGB-D camera images to reidentify the UE holder after occlusions occur. The proposed method can compensate for the drawback that visual tracking tends to fail when the UE is occluded and allows the UE to be located continuously, even in a crowded environment. The proposed method uses a metric that combines RSSI and the occlusion in the RGB-D camera image, which is the same as the UE position estimation in [8]. This has the advantage of simplifying the system configuration because a common metric can be used for both UE position estimation and tracking. In addition, as reported in [8], we use only the information from the BS side, which is different from the prior work [6]. The advantage of this process is that the proposed method is not limited by the sensing ability or the processing capability of the UE.

The contributions of this paper are summarized as follows:

1) We propose a method for tracking UEs that is robust with regard to occlusion and can be used in crowded areas and for distant UE.
2) We propose a UE reidentification method that combines radio information and images to solve the problem of a visual tracker losing its target during occlusion.
3) We evaluate the proposed method using an outdoor experiment that simulates a communication environment with mmWave BSs. The proposed method achieves an 11.5% improvement in tracking accuracy compared with conventional visual tracking.
4) We provide a comprehensive parameter validation of the proposed method, an investigation of mmWave radio information, and a discussion of their influence on the proposed method.

The remainder of this paper is organized as follows. Section II presents related work, and Section III describes the proposed method. Section IV evaluates the validity of the proposed method and presents several discussions through experiments, and we present conclusions in Section V.

II. RELATED WORK

First, previous studies on blockage prediction for mmWave are introduced. Blockage prediction is important for the application of the proposed method. Then, the techniques of object tracking are reviewed.

A. MMWAVE BLOCKAGE PREDICTION

For advanced 5G systems, several methods, as listed in Table 1, have been proposed to predict the blocking effects on mmWaves using camera images [9], [10], [11], [12]. They utilized deep learning models based on convolutional neural networks. The blockage effects of moving objects on the RSSI of mmWaves can be predicted. However, these methods assume that the UE position is known, and deep learning models are trained on images in which the UE position is fixed. Therefore, these methods cannot be applied to

1564  VOLUME 3, 2022

MIHARA et al.: USER EQUIPMENT TRACKING FOR A MILLIMETER WAVE SYSTEM USING VISION AND RSSI
TABLE 1. Summary of related research on mmWave blockage prediction.

| References | Method | Complexity | Remarks |
|------------|--------|------------|---------|
| [9], [10], [11], [12], [13] | Machine learning with vision | Low: No need for preprocessing | UE position is assumed to be known |
| Machine learning with vision and beamforming control information | Middle: Need for preprocessing to associate between vision and beamforming control information | Difficult to apply to complex environments because of the coarse UE position obtained from beamforming information |
| [35] | Machine learning with top view images | High: Need for preprocessing to generate top view images by using position information of the UE and objects | Applicable to moving UEs and crowded environments |

*Complexity is addressed in terms of preprocessing for machine learning. Other aspects, such as machine learning algorithms, are comparable.

TABLE 2. Summary of related research on UE tracking.

| References | Method | Remarks |
|------------|--------|---------|
|           | Radio-tracking | Limited accuracy (10 m >) |
| [15]      | Radio triangulation | Limited accuracy (10 m >) |
| [16], [17] | GPS with correction information | Accuracy degradation in areas without open sky |
| [18], [19] | OTDOA | Limited accuracy (10 m >) |
| [20]      | E-CID | Limited accuracy (10 m >) |
| [21]      | Fingerprints | Limited accuracy (10 m >) |
| [22]      | mmWave radar | Limited angular resolution |
| [23], [24] | Proposed | Tracking with vision and RSSI | Applicable to moving UEs at distance |

outdoor scenes where many objects are moving, and the positions of the UEs are usually unknown. In the most recent study, a method of predicting mmWave blockage without explicitly estimating the position of the UE has been proposed [13]. A deep learning model is trained using beamforming control information and images to predict mmWave blockage by a vehicle. The position of the UE is represented as a latent feature that is extracted from the beam information. However, beamforming information provides only coarse UE directions, which makes it difficult to accurately predict blockages in complex environments where there are many smaller objects, such as pedestrians. Another possible method is assumed to be predictions based on blockage occurrence models. For example, a prediction based on the Markov processes, such as the four-state Markov process model [14], could be considered. The probability of the occurrence of a blockage can be obtained from the probability of the transition to other states if the current state is estimated. However, in a crowded environment, multiple blockages by a variety of different objects may occur concurrently. To deal with such a situation, a large number of states may be defined. As a result, modeling is expected to be difficult since the state transition branches become complex and enormous. Therefore, machine learning-based methods have an advantage in predicting the different situations between the terminal and surrounding objects.

B. OBJECT TRACKING

Techniques for tracking objects using radio information have conventionally been used in a variety of fields as summarized in Table 2. One of the traditional techniques is radio-tracking [15], where a radio transmitter is attached to the tracking target, and the position is measured remotely. When using VHF radio waves, the position is estimated by triangulation using multiple receiving stations [16]. PHS may also be used [17]. However, these methods have errors on the order of tens of meters. In other cases, a GPS receiver is attached to the tracking target, and the location information is transmitted wirelessly. GPS is widely used for positioning in outdoor environments. However, a GPS error typically exceeds 5 m. Even when using correction information, the error is still approximately 1 m [18], [19]. In addition, the error is even higher in urban areas where exposure to the celestial sphere is limited. In mobile communication systems, several techniques for positioning and tracking UEs have been developed, including the observed time difference of arrival (OTDOA) [20], enhanced cell ID (E-CID) [21], and fingerprints [22]. However, all of these techniques have error margins of tens of meters [18]. A mmWave radar is a device that is dedicated to detecting and tracking objects by radio information. By analyzing the point cloud measured by mmWave radar, several objects can be tracked simultaneously [23], [24]. However, in a crowded outdoor environment, object tracking is difficult because of insufficient angular resolution [25].

Recently, visual tracking based on image recognition has been extensively studied. Benefiting from the high representational capability of the deep neural networks, tracking performance has improved. The standard approach to tracking is tracking-by-detection [26], where object
detection is performed, and detections are associated between camera frames based on the similarity metric. DeepSORT [27] uses an overlap of detections and appearance features to perform an association using the Hungarian method. Because of its better balance of efficiency and tracking performance, DeepSORT is widely used in real-world applications, such as person tracking using surveillance cameras. In the latest research, methods using a Siamese network with two inputs have been proposed [28], [29]. The Siamese network can be trained to measure the similarity of feature maps from input images. A one-shot tracking approach has also been studied [30] where object detection and association are output at once. These tracking methods use deep neural networks to achieve high tracking performance. However, the problem of tracking failure due to occlusions has not been completely solved. In addition, visual tracking faces difficulties when applied to the nano-area [1], which is a communication area of up to 50 m formed by mmWave band BSs. The nano-area is wider than the shooting range of general cameras (e.g., surveillance cameras and robot-mounted cameras) targeted by visual tracking. When a camera is installed at the BS, the resolution of the UE on the captured image can be coarse, which is a difficult condition for visual tracking due to the lack of fine-grained features of the object. This problem is difficult to overcome even if training is performed on a dataset that includes distant objects because it does not compensate for the lack of missing features.

A method for tracking UE by combining images and information on the UE has been proposed [6] as described in Section I. However, transferring the information from the UE to the BS results in a heavy load on the communication links. Issues also need to be considered in applying the visual tracking to the nano-area as mentioned above.

Methods for tracking objects by combining images and radio information have been studied. Most are sensor fusion techniques that use a camera and mmWave radar [31], [32], [33]. Images from a camera and point clouds from mmWave radar are aligned, and detection and tracking based on the respective data are extrinsically merged. The limitation of these methods is that they require a particular device for mmWave radar. The proposed method uses mmWave radio information that is used in a mobile communication system. This radio information can be obtained without any additional device in the base station environment we assume. In addition, the proposed method uses the intrinsic combination of camera images and radio information for tracking, which is distinct from these sensor fusion methods [31], [32], [33]. A UE localization estimation method using camera images and RSSI was proposed [34]. The information used for location estimation is similar to the method proposed by the authors in [8], while the estimation method is different. In [34], the UE location is estimated by estimating the Fresnel ellipse. In order to estimate the Fresnel ellipse, a number of samples of RSSI variations due to blockages are required. This means that it takes time to estimate the UE position, which is an issue for application to the tracking of moving UE.

III. PROPOSED METHOD

In this section, we elaborate on a system model and a proposed method for tracking the UE using a combination of visual tracking and reidentification by radio information and images.

A. SYSTEM MODEL

The system model is shown in Fig. 1. Multiple BSs are placed in an area with a radius of approximately 50 m. An RGB-D camera is placed alongside the antenna of the radio unit (RU). The camera images and the RSSI are collected at the server for UE tracking. At the server, the position of the UE holder is first identified by the method proposed in [8]. Next, the identified UE is tracked by the method proposed in this paper. Then, objects around the tracked UE holder, such as pedestrians and vehicles, are recognized from the image, and the positional relationship is estimated between the UE holder and the objects. Based on this positional relationship, blockages are predicted using such methods as machine learning [35]. Tracking of the UE holder enables continuous blockage prediction. Continuous blockage prediction ensures that the BS switching and beam control as described in Section I can be performed by the server at the distributed unit (DU) at the appropriate times, which is expected to secure wireless connectivity in mmWave. Throughout the processes of UE holder estimation, UE tracking, and blockage prediction, the holder is assumed to have a single UE, and the case of the holder with multiple UEs is not included in this paper. We focus on the blockage caused by objects surrounding the UE, aiming at blockage prediction over a wide area, such as nano-area, using a camera alongside the BS toward management of cooperative beamforming [7]. Self-blockage by the UE holder [4], [5] is also an important issue in mmWave communication. However, additional challenges are assumed to be faced in realizing self-blockage prediction with a camera at the BS side, particularly due to the limited
resolution of distant objects in the camera image, which is subject to future work.

In terms of the complexity of blockage prediction as shown in Table 1, the machine learning algorithm is comparable to previous studies [9], [10], [11], [12], [13], except for the preprocessing. Specifically, the proposed method estimates object location information from camera images, whereas previous studies input camera images directly input to machine learning. As a result of the preprocessing, we expect the proposed method to be effective in improving the accuracy of blockage prediction since the location of objects around the UE can be accurately determined. This paper focuses on UE tracking, and the integration of the proposed method with blockage prediction and the control of wireless communications as described above will be the subject of future work.

B. OVERVIEW OF THE PROPOSED METHOD

The proposed method tracks a UE using the BS side information only. By combining visual tracking and reidentification using radio information and images, the UE is continuously tracked even in a crowded environment. While we have a line of sight to the UE, we can use visual tracking. When the UE is occluded by other objects, we use a combination of radio information and camera images from the BS to reidentify the UE holder. Visual tracking is corrected based on the reidentification results. This reidentification process enhances the tracking performance of the UE in both a crowded environment and where it is far away from the camera. This is a unique part of the proposed method.

To reidentify the UE, we focus on the fact that the variation in radio information when a blockage occurs is temporally synchronized with occlusions on the image [8]. The UE holder is reidentified by matching the temporal variations of the RSSI and the quantified values of the occlusions on the camera images. Because mmWaves propagate solely by line-of-sight paths, the RSSI drops rapidly when the line-of-sight path is blocked by an object and recovers as soon as the blockage is removed. Visible light captured by the camera is instantly occluded when it is blocked by an object and vice versa. Therefore, at the position of the UE on the image, the timing of occlusion/disocclusion coincides with the timing when the RSSI decreases/recoveres. We combine the radio information with the camera images using this coincidence for reidentification. The reidentification is accomplished by integrating the same metrics into the visual tracking as in [8].

The essential component of the proposed method is a uniquely defined method for quantifying the occlusion in an image. We use the object region extracted from the image and the depth information to calculate the intensity of the occlusion. In this paper, this quantified value is called the occlusion intensity. We calculate the occlusion intensity as a continuous value by approximating the propagation path of mmWave with a simple distribution to make it comparable to the RSSI. We do not consider interested in the temporal variation not the accurate absolute value. The occlusion intensity on the image and the RSSI are used to reidentify the UE. Specifically, we search for the UE trajectory where these temporal variations are consistent. We estimate several possible trajectories that the UE could have traveled and calculate the occlusion intensity for each of them. The trajectory with the highest correlation between the measured RSSI and the occlusion intensity is associated with the UE. The object at the end of the associated trajectory can be reidentified as the UE holder. Finally, after validating reidentification in terms of geometry and radio information, a correction to visual tracking is performed.

The process flow is shown in Fig. 2. The proposed method consists of nine processing blocks, and solid line blocks and dashed line blocks describe the proposed unique methods and existing methods, respectively. The inputs are an image sequence taken by an RGB-D camera and the RSSI measured at the BS. Each step in the process is described as follows.

(a) Object detection is processed for each input RGB image.
(b) Visual tracking is processed for the detected objects.
(c) Occlusions and disocclusions to the UE are detected from the temporal variation of the RSSI. When occlusion and disocclusion are detected, the processes for tracking correction are started. Objects detected simultaneously with the disocclusion are candidates for the UE holder.
(d) For each candidate, the trajectory during the occlusion is estimated by interpolation.
(e) The moving object segments are extracted from the RGB images.
(f) We then calculate the occlusion intensity for each candidate with the depth images, RGB images, and the estimated trajectory.
(g) The similarity is calculated between the temporal variations in the RSSI and the occlusion intensity. This similarity is regarded as the metric for reidentification. In this paper, we refer to this as the radio metric.
(h) Based on the radio metric, the UE holder is reidentified. We correct the visual tracking after validating the reidentification result by thresholding for the radio metric.
(i) The geometric distance is also calculated by movement prediction and used for validating the reidentification.

Processes (e), (f), and (g) perform calculations for the radio metric the same as those presented in [8]. Since a common metric can be used for estimating and tracking the UE holder, the system can be simply structured. The time required for the reidentification is longer than the duration of the occurrence of the blockage because it acquires the data necessary for the calculation of the radio metric. Since the reidentification is initiated by the occurrence of a blockage, real-time performance corresponding to the frame rate of the camera is not required.
C. PROBLEM DEFINITION
The problem considered in this paper is defined as tracking the UE holder using the time series data of the RSSI and camera vision. We let the RSSI measured at time $t$ be $s_j^t$, $j \in J$, where $J$ is the set of UEs, and the bounding boxes of the objects in the image be $B_t = \{b_i^t\}$, $i \in I_t$, where $I_t$ is the index set of the bounding boxes at time $t$. The bounding box is represented in image coordinates as $b_i^t = (x_i^t, y_i^t, w_i^t, h_i^t)$, where $x_i^t$ and $y_i^t$ are the coordinates of the top-left corner of the bounding box, and $w_i^t$ and $h_i^t$ are the width and height of the bounding box, respectively. The UE tracking problem is defined as finding the index series $a_j^t \in I_t$ to which UE $j$ corresponds for time $t \geq 0$, given the above time series data. We assume that index $a_0^t$ at the initial time is known. We omit UE index $j$ in the following because each UE can be tracked independently.

D. DETAILS OF EACH PROCESSING BLOCK
In this section, we describe the details of each process in the proposed method.

1) VISUAL TRACKING
In visual tracking, an object is first identified in block (a) in Fig. 2, and the bounding box is obtained. Next, the object is tracked based on the bounding box in block (b) in Fig. 2. If the UE is not occluded by another object, the UE is detected and tracked by camera vision. While the UE is occluded, the result of visual tracking is tentative. We let $t_o$ be the time of occlusion and $t_e$ be the time of disocclusion. At time $t_e$, we correct the result of visual tracking according to reidentification based on the combination of radio information and images. We determine $t_o$ and $t_e$ by occlusion detection as described in the next section.

2) OCCLUSION DETECTION
Occlusion to the UE is detected by analyzing the temporal variation in the RSSI. The blockage of mmWave synchronizes with the occlusion on the image. At the moment of occlusion, the RSSI drops rapidly. Conversely, at the moment of disocclusion, the RSSI recovers. Therefore, by finding a sudden variation in the RSSI, we can detect occlusion and disocclusion to the UE. The proposed method analyzes the temporal variation in the RSSI using the first-order derivative operation. A decision threshold is applied to the derivative value of the RSSI to detect occlusion and disocclusion. Specifically, the times of occlusion and disocclusion, $t_o$ and $t_e$, are determined by detecting the consecutive RSSI derivatives across the threshold. We let $s'(t)$ be the derivative of the RSSI obtained by the first-order derivative operation and let $\beta_\text{th}$ be the threshold. Then, $t_o$ and $t_e$ are determined as the times that satisfy the following conditions:

$$s'(t_o + \varepsilon_o) > -\beta_\text{th} > s'(t_o + \varepsilon_o + \eta),$$  \hspace{1cm} (1)

$$s'(t_e - \varepsilon_e) < \beta_\text{th} < s'(t_e - \varepsilon_e + \eta),$$  \hspace{1cm} (2)

where $\varepsilon_o$ and $\varepsilon_e$ are margins, and $\eta$ is a sampling interval. If a series of occlusions occur in a short period of time, they are merged into a single occlusion. For example, we let the times of two consecutive occlusions be $(t_1^o, t_1^e)$ and $(t_2^o, t_2^e)$, and if $t_2^s < t_1^e$, then $(t_1^o, t_2^s)$ is the detected period of the occlusion.

3) TRAJECTORY INTERPOLATION
When the UE is occluded, the UE is often not detected correctly. Therefore, the trajectory traveled by the UE during the occlusion is approximated by interpolation. We assume that the position of the UE before the occlusion is known by a visual tracker. The UE position $b_i^0$ at time $t_o$ is obtained. After disocclusion, the UE can be detected. The bounding
boxes $B_{te}$ detected at time $t_e$ are candidates for the UE holder. We correct the visual tracking by reidentifying the UE holder from these candidates using the combination of images and radio information. The UE moves from position $b_{te}^s$ to one of the positions of candidates $B_t = \{ b_{te} \}$. The trajectory $T^i$ for each candidate is approximated by interpolation. In this paper, we use linear interpolation, assuming that the UE has a constant speed during occlusion. Specifically, the trajectory $T^i$ is calculated as follows:

$$T^i = \left\{ b_{te}^{hs} + \frac{k\eta}{t_e - t_s} (b_{te}^s - b_{te}^{hs}) \right\} \quad k \in K,$$

where $K = \{ 0, 1, \ldots, (t_e - t_s)/\eta \}$.

Linear interpolation is effective in situations where pedestrians and vehicles are passing on the road, which is assumed to be one of the typical use cases. In such situations, the duration of the blockage is less than a few seconds, and the assumption of constant velocity linear motion is reasonable in most cases, except for situations of sudden directional changes or acceleration/deceleration. On the other hand, in situations where the duration of the blockage is long, for example, when a pedestrian passes behind a parked vehicle, the assumption is not valid. For such a situation, nonlinear interpolation, such as the Kalman filter or machine learning, are expected to be applied. Since a detailed study of algorithms and accuracy is required for application, it is outside the scope of this paper.

4) OCCLUSION INTENSITY CALCULATION

The occlusion intensity described in Section III-A is key to reidentifying the UE holder. In this section, the definition of occlusion intensity is described.

Fig. 3 shows the simplified geometry of mmWave propagation with a blockage and camera projection. A schematic diagram of the definition of occlusion intensity is shown. The degree of occlusion by objects in the image is quantified by simplifying mmWave propagation. We assume that an RGB-D camera is installed near the antenna of a BS and captures images, similar to that presented by Ito et al. [9]. The process flow is shown in Fig. 2. The proposed method consists of nine processing blocks, and solid line blocks and a mmWave propagates with a spread in a path centered on a line of sight connecting the transmitter and receiver antennas. The region that obstructs the radio propagation is represented by an ellipsoid (i.e., the Fresnel zone). When this ellipsoid is projected onto the image plane in the manner of camera geometry, it can be approximated as a circular region with the intensity centered on the position of the UE. We refer to this region as the propagation distribution. Because mmWaves have a linear propagation characteristic (i.e., they have a similar propagation characteristic to that of visible light), we assume that mmWaves are blocked in accordance with the overlap between the propagation distribution and moving objects on the image plane. The moving object segments on the image can be extracted by background subtraction. The occlusion intensity is defined as the negative integral of the part of the propagation distribution that is occluded by the object segments (i.e., the shaded area in Fig. 3). The sign inversion aligns the sign of the variances in the occlusion intensity and the RSSI.

The specific method for calculating the occlusion intensity is described below. We assume that the UE is at a point $P_i$ in a space corresponding to a point $P_i(i \in I_i)$ in an image, and the point $P_i$ is inside the bounding box $b_i$. Because where the UE is located in the bounding box is unknown, the point $P_i$ is assumed to be the center of the bounding box. We thus consider calculating the occlusion intensity $h'$ when moving objects are blocking the mmWave arriving from point $P_i$. The propagation distribution $G$ can be approximated by a Gaussian distribution of variance $\sigma$ centered at point $P_i'$, where $r$ is the distance from point $P_i$ on the image plane:

$$G(r, \sigma) \approx \frac{1}{\sqrt{2\pi \sigma}} \exp\left( -\frac{r^2}{2\sigma^2} \right).$$

Next, object segments in the image are extracted. We let $O$ be a set of pixels belonging to objects. The portion of the extracted segment $O$ that is in front of point $P_i$ can affect mmWave blockage. Therefore, we define a blocking region function $Q'(q) = \{ q \in \{ |b_i| \} \}$ that equals 1 at the pixels with depths smaller than the bounding box’s depth $\tilde{d}$ and $-1$ at the pixels with depths larger than $\tilde{d}$, where $\langle \cdot \rangle$ denotes a set of pixels inside the bounding box. The depth of $\tilde{d}$ is calculated as the mean depth of segments $O \cap \{ |b_i| \}$ at time $t < t_s$ and $t_e < t$. At time $t_s \leq t \leq t_e$, $\tilde{d}$ is interpolated between time $t_s$ and $t_e$. For pixels not belonging to the object segments $O$, we set $Q'(q) = 0$:

$$Q'(q) = \begin{cases} -1 + 2 \cdot \left[ d_q < \tilde{d}_i \right], & q \in O \\ 0, & q \notin O, \end{cases}$$

where \([ \cdot ]\) denotes Iverson’s notation, and $d_q$ is the depth of pixel $q$.

Based on the definition presented above, the occlusion intensity $h'$ can be calculated as follows, where $\theta$ is the angle in the polar coordinates with point $P_i$ as the origin and $q_{i,r,\theta}$ is the index of the pixel determined by $i$, $r$, and $\theta$:

$$h' = -\sum_{r,\theta} G(r, \sigma) Q'(q_{i,r,\theta}).$$
5) TRACKING CORRECTION

Visual tracking is corrected based on the reidentification of the UE holder by comparing the temporal variations in the occlusion intensity calculated from the image sequence and the RSSI of the UE. While an object passes across the line of sight between the BS and the UE, the sequence of occlusion intensity calculated at the UE position has a similar temporal variation to the sequence of the RSSI. Therefore, we can obtain the likelihood for each candidate $b_{te}$ at time $t_e$ to hold the UE by evaluating the similarity between the temporal variations in occlusion intensity $h_i$ for the trajectory $T_i$ and the RSSI $s_t$. In this paper, we calculate a zero-mean normalized crosscorrelation as an indicator of the similarity. The RSSI $s_t$ and the occlusion intensity $h_i$ are smoothed by a moving average with window size $\phi_w$. The calculated value is denoted as the radio metric $L^i$, which represents the likelihood that the UE is held by the candidate $b_{te}$:

$$L^i = \frac{\sum_i (\Phi(h_i) - \bar{h}) (\Phi(s_t) - \bar{s})}{\sqrt{\sum_i (\Phi(h_i) - \bar{h})^2 \sqrt{\sum_i (\Phi(s_t) - \bar{s})^2}}}$$

(7)

where $\Phi$ denotes the smoothing function.

As described in Section III-C4, point $P_i$ is set to the center of the bounding box. If the UE is located off center, the occlusion intensity and RSSI are misaligned temporally. To compensate for this misalignment, the RSSI is shifted within a range of $\pm \delta$ when calculating the similarity, and the maximum similarity is used as the radio metric for each candidate. Finally, the UE holder is reidentified through the following association process:

$$a_{te} = \arg \max_i L^i.$$  

(8)

The proposed method determines whether to perform the tracking correction by reidentification based on the radio metric validation and geometric validation described in the following. If the tracking correction is not executed as a result of the validations, then we try it again at the next time step of $t_e + 1$. If the number of iterations exceeds $n_{th}$, the iteration is terminated, and the tentative tracking of visual tracking becomes fixed.

a) Radio metric validation: The calculation of the radio metric involves the approximation of trajectories by interpolation, which is not always correct. Therefore, it is necessary to validate whether the tracking could be corrected using a radio metric. We measure the validity of the radio metric based on its own value. The radio metric for the UE holder should be a relatively large value. If all radio metrics calculated for candidates are small, we can conclude that the calculation of occlusion intensity for the UE holder failed, and we suspend the tracking correction. Specifically, when the maximum value of the radio metric $L^i$ for candidates $B_{te}$ is smaller than the threshold $l_{th}$, we stop correcting the tracking.

b) Geometric validation: The reidentification process includes the elimination of candidates based on movement prediction in the same manner as conventional visual tracking. We predict the position $\hat{b}_{te}$ of the UE holder at time $t_e$ using the tracking history up to time $t_e$ before the occlusion. We eliminate the UE holder candidates $b_{te}$ whose distances from the predicted position are larger than the threshold $g_{th}$.

IV. EXPERIMENT AND EVALUATION

To demonstrate the effectiveness of the proposed method, we performed experiments in an outdoor environment simulating the nano-area [1], which is a communication area of up to 50 m formed by mmWave band BSs, and we evaluated the tracking results using measured data.

A. EXPERIMENTAL ENVIRONMENT

The experimental environment is shown in Fig. 4. Variations in the height of the BSs were designed to simulate real-world environments. We constructed a 10 m high BS (high BS) to simulate the eaves of a building or the roof of a low-rise residential building and a 4 m high BS (low BS) to simulate a utility pole or a traffic light. The UE was placed opposite each BS at three different horizontal distances of approximately 30 m, 40 m, and 50 m. The UE was connected to a low BS for a distance of 30 m and to a high BS for other distances. For the transmitter and receiver, we used WiGig devices that operated in the 60 GHz band to easily implement the experimental system. Table 3 shows the specifications of the devices. A beamforming function
of the WiGig device can automatically adjust the directional characteristics. The control procedure for beamforming is in accordance with the WiGig standard specification, where the directional characteristics of the antenna are searched horizontally to set the beam in a direction that provides a high RSSI. Actually, in an environment with clear visibility, we confirmed that the directional characteristics were selected to correspond to the direction of the line of sight. The UE was mounted on a pushcart and moved along line segments A, B, and C as shown in Fig. 4(a). The height of the UE was 1.2 m. The person who pushed the cart was the UE holder, who was accompanied behind by another person. Two persons acted as blocking objects and moved between the UE and the BS. At each BS, the RSSI of the UE was measured at 16 samples per second. A stereo camera was installed as an RGB-D camera next to the antenna of the BSs and captured grayscale and depth images at 10 fps. The image resolution was 1600 × 1200, the baseline length of the stereo camera was 0.25 m, and the depth resolution was 1.93 m at a distance of 50 m. The number of pixels of the person on the image was approximately 20 × 70. The experiment was conducted during the daytime on a sunny or cloudy day.

B. RESULT OF TRACKING CORRECTION

To show that the proposed method can correct visual tracking, we apply the proposed method to the measured experimental data. The parameters of the proposed method are set as follows: \( \beta_{th} = 3.0, \varepsilon_s = 1.5, \varepsilon_e = 1.5, \sigma = 10, \phi_w = 0.3, \delta = 0.1, \eta_{th} = 10, \delta_{th} = 0.5, \) and \( g_{th} = 100. \) The effect of these parameters on the tracking performance is discussed in Section IV-D1. We use YOLOv4 [36] as the detector and DeepSORT [27] as the visual tracker. Both methods achieve a good balance of processing speed and performance and are de facto standards that are commonly used in practice. The detector trained by the MS COCO dataset [37] is fine-tuned by the portion of measured data that is not used in the evaluation, and the precision and recall of the detector are 0.93 and 0.93, respectively. The performance of object detection is sufficient for the experimental environment. With regard to the training of the appearance descriptor of DeepSORT, no dataset was found that included distant objects suitable for the assumptions made in this paper, and thus a widely used MARS dataset [38] was used. For segment extraction, the dynamic background subtraction method [39] was used. This method is robust to changes in ambient light.

An example of the results of the proposed method is shown in Fig. 5. The UE and the blocking objects moved along line segment B in Fig. 4(a). The red solid box in Fig. 5(a) shows the tracking result of the proposed method. From top to bottom, each figure in Fig. 5(a) corresponds to the time indicated by the black triangle in the RSSI graph in Fig. 5(b). The position of the UE is indicated by a white dot. The tracking result is correctly associated with the UE holder. To demonstrate the problem with the conventional method, the results of naive visual tracking for the UE are shown as yellow dashed boxes, which shows that conventional visual tracking switches to another object after occlusion occurs.
The graphs in Fig. 5(b) show the measured RSSI and the internal outputs of the proposed method, where the time when the UE is blocked is shown as vertical red lines. The graph of the RSSI indicates that the blockage caused two rapid drops in the RSSI. In the graph of the output of the first-order derivative operation, the graph is across the horizontal dashed lines, which indicate the threshold \( \pm f_{th} \), around the time of the blockages. Each occlusion is correctly detected. The occlusion intensity calculated for each of the candidates, which are labeled (A) to (D) in Fig. 5(a), is shown at the bottom of Fig. 5(b). The radio metric (i.e., normalized cross-correlation with the RSSI) is shown on the bottom right of each graph. The graph of the candidate (B) who holds the UE fluctuates in synchronization with the RSSI, and the radio metric is the largest among the candidates. As a result, the UE holder is correctly reidentified. The processing time for correcting a tracked object with the proposed method is approximately 300 ms at one time on a PC with an Intel Core i7-7700K CPU, 32 GB of memory, GeForce RTX 2800 Ti GPU, and the Ubuntu OS. The processing time for conventional image-based tracking is approximately 40 ms at one time.

C. EVALUATION

1) TRACKING ACCURACY

To show that the proposed method improves tracking performance, we statistically evaluated the proposed method experimentally. The evaluation result is shown in Table 4. To evaluate the effectiveness of the proposed method, we compared the proposed method and conventional visual tracking for which we used DeepSORT [27]. Because DeepSORT is also used in the proposed method, this comparison evaluates the effectiveness of the tracking correction of the proposed method, which is the difference in these methods. Whether a target is successfully tracked between two times \( t_1 \) and \( t_2 \) is determined by the indices of \( d_{t_1} \) and \( d_{t_2} \). If these indices correspond to the identical object, the target is successfully tracked. An evaluation score \( f(t_1, t_2) \) is assigned to this definition as follows:

\[
f(t_1, t_2) = \begin{cases} 
1, & \text{if } M(b^	ext{gt}_{t_1}) = M(b^g_{t_2}) \\
0, & \text{otherwise,}
\end{cases}
\]  

(9)

where \( M(b^g_i) \) is a ground truth label identifying the bounding box \( b^g_i \). The ground truth is manually annotated. We calculated the tracking accuracy \( E_{\text{acc}} \) by averaging the evaluation score \( f(t_0, t_2) \) over all trials and converting it as a percentage. In addition, to measure the effect of the radio metric and geometric validation of the proposed method, we calculated the execution rate of the tracking correction and the tracking accuracy when the correction was executed, which are denoted as \( R_{\text{cor}} \) and \( E_{\text{acc}} \), respectively. The evaluation score \( f(t_1, t_2) \) is comparable to those of other well-known metrics that describe the top-1 accuracy used in the field of person reidentification and the robustness included in the performance evaluation protocol of the VOT Challenge [40].

A total of 154 trials were conducted, and each trial included consecutive blockages by two persons. The parameters of the proposed method were set in two patterns: one with strict geometric validation (SGV in the table) and the other with relaxed geometric validation (RGV in the table). Because of parameter tuning, SGV and RGV have different thresholds \( l_{th} \) for radio metric validation. SGV has more relaxed radio metric validation than RGV. The specific values and detailed understanding of each parameter setting are described in the discussion in Section IV-D1.

Table 4 shows that the tracking accuracy \( E_{\text{acc}} \) with the proposed method is higher than that of conventional visual tracking. Comparing the results for all trials, the proposed method improves tracking accuracy by 11.5% compared to the conventional method. In all patterns, except for the RGV setting and the 30 m distance, the proposed method achieved better tracking accuracy than the conventional method. The results thus show the effectiveness of tracking correction by reidentification using radio information and images. Comparing the results of SGV and RGV, the tracking accuracy is higher for SGV than RGV. In terms of distance, the tracking accuracy of the RGV setting deteriorates as the distance becomes shorter because we ignore the effect of the distance to the radio metric; the parameter \( \sigma \) is set to be a constant for the basic investigation. The value of \( \sigma \) was favorable for the more distant objects. The parameter \( \sigma \) of the propagation distribution \( G \) needs to be appropriately adjusted depending on the distance but is set to a constant value in this experiment. Because the RGV setting is highly dependent on the radio metric (see Section IV-D1 for details), the effect of the distance variation becomes stronger. Considering the distance between the UE and BS is left for future research.

Regarding other measures of the correction execution rate \( R_{\text{cor}} \) and the tracking accuracy with tracking correction \( E^c_{\text{acc}} \), it can be argued that the improved tracking performance for distant objects is the result of the proposed method that successfully detects the occurrence of a blockage and performs the appropriate correction. These measures are also higher for SGV, which means that SGV effectively prevents incorrect corrections of tracking. In the experiment, the movement

| Method | Dist. [m] | \( E_{\text{acc}} \) [\%] | \( E^c_{\text{acc}} \) [\%] | \( R_{\text{cor}} \) [%] |
|--------|----------|----------------|----------------|----------------|
| DeepSORT | 30 | 84.0 | - | - |
| | 40 | 69.6 | - | - |
| | 50 | 62.9 | - | - |
| | All | 77.4 | - | - |
| DeepSORT w/ours (SGV) | 30 | 88.0 | 100.0 | 48.0 |
| | 40 | 80.3 | 26.5 | 54.9 |
| | 50 | 77.9 | 100.0 | 72.5 |
| | All | 88.9 | 92.4 | 61.4 |
| DeepSORT w/ours (RGV) | 30 | 72.0 | 70.0 | 40.0 |
| | 40 | 75.2 | 79.3 | 43.4 |
| | 50 | 93.4 | 100.0 | 52.3 |
| | All | 81.4 | 84.5 | 48.5 |
prediction was accurate because the UE moved at a nearly constant speed. Therefore, SGV allowed us to appropriately cancel the incorrect correction. This setting is suitable for an environment where objects move along in a constant flow. Conversely, the RGV setting can be applied to environments with more chaotic movement because it does not assume that the movement prediction is accurate. For RGV, the improvement in tracking accuracy shows the contribution of reidentification using radio information and images.

2) CORRELATION BETWEEN OCCLUSION INTENSITY AND RSSI
To show that the occlusion intensity appropriately quantifies occlusions on the image, we analyzed the temporal correlation between occlusion intensity and the RSSI. The proposed method uses the fact that the occlusion to the UE on the image and the RSSI variation are synchronized. The occlusion intensity is an essential component in observing this temporal coincidence. Table 5 shows the results of the correlation analysis. Using the same trials as in Section IV-C1 and the same parameter settings as in Section IV-B, we averaged the correlations calculated for the real UE holders and the non-holders. Additionally, average correlations were calculated in the same manner for trials in which a correction to the visual tracking was performed. For implementation reasons, correlations below zero were rounded up to zero. Table 5 shows that the correlation is higher for the UE holders than for the non-holders, which means that the occlusion intensity calculated for the UE holder has a similar temporal variation as the RSSI. The occlusion intensity is thus shown to appropriately capture the temporal synchronization between the occlusion on the image and the RSSI.

D. DISCUSSION
1) VALIDATION OF THE PARAMETERS
The proposed method has some parameters that must be set manually; thus, we examined the influence of these parameters on tracking performance. Because it is impractical to examine all possible combinations of parameter values, we use a validation curve, which is used in the machine learning community. The parameters are fixed except for one, and the tracking accuracy is examined by incrementally changing the value of only one parameter. Fig. 6 shows a plot of the validation curve for each parameter. The parameters of the proposed method and the test set for each parameter are listed in Table 6. The fixed value (hereafter referred to as the default value) for each parameter is in bold.

![FIGURE 6. Validation curves for each parameter of the proposed method.](image)

**TABLE 5. Correlation between the occlusion intensity and RSSI.**

|                | UE holder | Non-holder |
|----------------|-----------|------------|
| All trials     | 0.423     | 0.110      |
| Trials w/correction | 0.557   | 0.115      |
thresholding by geometric validity is 1.18. This result can be interpreted that the geometric validation narrowed down the candidates to one bounding box in most cases. Fig. 7 shows the validation curve of the threshold $l_{th}$ when we set the fixed value of $g_{th}$ to 30. When the threshold $l_{th}$ is decreased, the tracking accuracy is not considerably degraded, which means that the geometric validation narrowed down the candidates to the correct ones. The parameter setting of the SGV described in Section IV-C is a pattern that uses 0.3 as the threshold $l_{th}$ for the radio metric validation in this case.

Time shift $\delta$ for the RSSI is best for 0.2 as shown in Fig. 6(g). The UE is located on a pushcart, which is not at the center of the bounding box of the UE holder. The blockage timing between the UE and the center of the bounding box differed by 0.1 s to 0.2 s in the experiment. When $\delta$ is larger than 0.3, tracking accuracy decreases. This is because the temporal difference between blockages of the UE holder and the accompanying person is less than 0.3 s. Thus, distinguishing which object’s blockage is synchronized with the RSSI drop becomes impossible.

Tracking performance is best if the threshold of occlusion detection, $\beta_{th}$, is 3 as shown in Fig. 6(d). Tracking performance decreases when the time margins $\epsilon_s$ and $\epsilon_e$ are smaller than 1.0 as shown in Figs. 6(e) and (f). The effect of noise increases relatively when the number of samples is small. Fig. 6(a) shows the impact of the smoothing parameter $\phi_w$ on tracking accuracy for a possible range of $\phi_w$ in the environment of the experiment. It shows that the parameter $\phi_w$ does not have a strong effect on tracking performance.

Fig. 6(h) shows that tracking performance is best if the variance of propagation distribution $\sigma$ is 5. Actually, $\sigma$ should be adjusted by the distance between the transmitter and receiver antennas and the position of the blocking object. The introduction of a more rigorous blockage model is a subject for future research.

Finally, we discuss the dependency of these parameters on the environmental conditions. The optimal value for the geometric validity threshold $g_{th}$ depends on the motion state of the UE. If the environment is chaotic, where the UE moves in a complex way with strong acceleration/deceleration and path changes, the error in the movement prediction increases. The geometric validity threshold $g_{th}$ should be larger than the probable movement prediction error. The range of the time shift $\delta$ depends on the velocity and the size of the UE holder, which determines the temporal difference in the blockages between the UE and the center of the UE holder.

The optimal values for the smoothing parameter $\phi_w$ and the time margins $\epsilon_s$ and $\epsilon_e$ depend on the magnitude of the noise in the measured data. These values could be relatively environment independent. In terms of the height of the BS, we did not find marked differences in the optimal parameters between the high and low BSs. The tracking accuracies for the high and low BSs with the settings of default values are 80.8% and 80.2%, respectively.

2) INVESTIGATION OF MMWAVE RADIO INFORMATION

We investigated the mmWave radio information measured in the experiment and discuss the influence of its characteristics on the proposed method. We classified the measured RSSI into typical patterns of sharp, rectangular, and unrecognizable by the shape of the waveform. Fig. 8 shows examples of the respective waveforms. Both datasets were acquired when two persons were blocking the UE. The characteristics of temporal variation are different. This diversity derives from a combination of the specifications of the wireless devices used in the experiment and the influence of the UE movement. The wireless device uses beamforming to control their directional characteristics. When the UE moves, the directional characteristics are selected to follow the UE. The RSSI fluctuates depending on the selected directional characteristics. In this paper, we do not consider the effect of beamforming control on the RSSI. Therefore, the RSSI variation due to beamforming and UE movement is considered to be a disturbance in the proposed method. Suppressing

| PARAMETER | FUNCTION | UNIT | TEST SET |
|-----------|----------|------|----------|
| $\phi_w$  | Smoothing| s    | 0.1, 0.3, 0.5... 1.9, 2.1 |
| $l_{th}$  | Threshold of radio metric validity | - | 0.0, 0.05... 0.60... 0.95 |
| $g_{th}$  | Threshold of geometric validity | px | 10, 20, 30, 40, 50, 100, 200, 300 |
| $\beta_{th}$ | Threshold of occlusion detection | - | 1, 2, 3, 4, 5, 6 |
| $\epsilon_s$ | Time margin before occlusion | s | 0.8, 0.9..., 1.5..., 2.0 |
| $\epsilon_e$ | Time margin after occlusion | s | 0.8, 0.9..., 1.5..., 2.0 |
| $\delta$  | Range of time shift for RSSI | s | 0.1, 0.2, 0.3, 0.4, 0.5 |
| $\sigma$  | Variance of propagation distribution | - | 2.5, 5, 10, 15, 20, 25 |
The negative effects of RSSI fluctuations is left to future research. The use of other radio information is also one of the issues to be considered in future research. The beamforming information described above can be used to determine the approximate UE location. The WiGig device used in the experiment has a predefined table that maps directional characteristics to beam IDs. Beam IDs from 1 to 62 are assigned uniformly to azimuths from $-45$ to $45$ degrees. Fig. 9 shows the beam ID selected at high BS when the UE moves along line segments A, B, and C in Fig. 4 using pseudocolors along the UE’s trajectory. The UE holder carried an RTK-GNSS receiver, and the RTK-GNSS coordinate system and the image coordinate system were calibrated by the Perspective-n-Point method [41]. Positioning information is projected to the image to plot the UE’s trajectory. Fig. 9 shows that the beam ID approximately indicates the horizontal direction of the UE. To use the beamforming information in the proposed method, further examination of the beamforming of the advanced 5G system [42] is required because the information is different from the WiGig system.

3) APPLICATION TO BLOCKAGE PREDICTION

This section discusses the effectiveness and challenges of the proposed method when applied to BS control based on mmWave blockage prediction as described in Section I. We aim to ensure reliable wireless connectivity by predicting mmWave blockage and using this information to switch the BS to which the UE is connected in advance [1]. Continuous tracking of the UE holder enables continuous blockage prediction. Continuous blockage prediction is expected to ensure wireless connectivity for mmWave since BS switching and/or beam control can be performed at the appropriate instant. An alternative method includes CoMP. It ensures continuous communication without handover even if blockage occurs by establishing wireless links simultaneously with multiple BSs and UE. However, this leads to excessive link establishment when no blockage occurs, which is an issue in terms of frequency utilization efficiency. Additionally, cooperative beamforming has been studied as another method [7]. In combination with blockage prediction, it enables efficient communication by concentrating power on the links with a low probability of occurrence of blockage. In this method, the behavior is equivalent to that of CoMP when the accuracy of blockage prediction is low. This means that the accuracy of blockage prediction is closely related to the efficiency of communication.

To accurately predict mmWave blockage, it is necessary to identify the UE location with decimeter accuracy. However, this study developed a method to track the UE holder not a method to estimate the detailed location of the UE; therefore, the accuracy of the position is insufficient. On the other hand, if we must only predict the line-of-sight blockage roughly, it is sufficient to know which object is holding the UE. Also, even if the UE holder is mistakenly tracked, it will not be a problem if the location is not markedly wrong. For example, mmWave blockage is likely to be correctly predicted even if a person who walks with the UE holder in a group is mistakenly tracked as the UE holder. Depending on the positional accuracy required by mmWave blockage prediction, groupwise tracking is also an issue that should be considered in the future.

To apply the proposed method in a real communication environment, it is necessary to tackle nighttime and poor weather conditions. By using LiDAR or far-infrared cameras instead of the RGB-D camera used in this study, sensing can be performed without being affected by the time of day or weather conditions. However, to apply the proposed method to these sensor data, it is necessary to re-examine the method according to the sensor characteristics, data structure, and spatial resolution, which are different from an RGB-D camera.

V. CONCLUSION

For an advanced 5G system, we proposed a method for tracking UEs using radio information and camera images that can be obtained at the BS. The proposed method reidentified the UE using the synchronization between the RSSI drops of mmWave and the occlusions of the UE on the image. Based on experiments performed in an outdoor environment that simulated a real mmWave communication area, the proposed method provided higher tracking accuracy than conventional visual tracking methods. We achieved an 11.5% improvement in tracking accuracy when occlusions occurred. In the future, improvements to the proposed UE tracking method will be explored to address resilience to UE movement and crowded
environments by using wireless information obtained at the BS, such as beamforming. Methods using sensors other than RGB-D cameras will also be considered to cope with the different times of the day and weather conditions. Finally, to ensure wireless connectivity in 5G advanced systems using mmWave, we will study a blockage prediction method using UE location information obtained with the proposed tracking method, and control methods for wireless communication at the BS incorporated with the blockage prediction method.

REFERENCES

[1] T. Murakami et al., “Research project to realize various high-reliability communications in advanced 5G network,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Online, 2020, pp. 1–8.

[2] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, “Modeling and analyzing millimeter wave cellular systems,” IEEE Trans. Commun., vol. 65, no. 1, pp. 403–430, Jan. 2017.

[3] V. Raghavan et al., “Statistical blockage modeling and robustness of beamforming in millimeter-wave systems,” IEEE Trans. Microw. Theory Tech., vol. 67, no. 7, pp. 3010–3024, Jul. 2019, doi: 10.1109/MTT.2019.2899946.

[4] V. Raghavan, M.-L. Chi, M. A. Tassoudji, O. H. Koymen, and J. Li, “Antenna placement and performance tradeoffs with hand blockage in millimeter wave systems,” IEEE Trans. Commun., vol. 67, no. 4, pp. 3082–3096, Apr. 2019, doi: 10.1109/TCOMM.2019.2891669.

[5] V. Raghavan, R. A. Motos, M. A. Tassoudji, T.-C. Ou, O. H. Koymen, and J. Li, “Mitigating hand blockage with non-directional beamforming codebooks,” Apr. 2021, arXiv:2104.06472.

[6] J. H. Ryu, V. Raghavan, and J. Li, “Using image processing to assist with beamforming,” U. S. Patent 10,601,488, Mar. 24, 2020.

[7] H. Iimori, G. T. F. de Abreu, O. Taghihizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, “Stochastic learning robust beamforming for millimeter wave systems with path blockage,” IEEE Wireless Commun. Lett., vol. 9, no. 9, pp. 1557–1561, Sep. 2020.

[8] S. Mihara, S. Ito, T. Murakami, and H. Shinbo, “Positioning for user equipment of a mmWave system using RSSI and stereo camera images,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Nanjing, China, Mar. 2021, pp. 1–7, doi: 10.1109/WCNC49053.2021.9417559.

[9] S. Ito, S. Mihara, T. Murakami, and H. Shinbo, “Feasibility study on package placement with machine learning in outdoor mm-Wave environment,” IEICE Tech. Rep., vol. 120, no. 89, pp. 7–12, 2020.

[10] H. Okamoto, T. Nishio, M. Morikura, K. Yamamoto, D. Murayama, and K. Nakashima, “Simulation model and location accuracy for observed time difference of arrival (OTDOA) positioning technique in third generation system,” in Proc. IEEE Student Conf. Res. Develop. (SCoReD), Kuala Lumpur, Malaysia, Dec. 2010, pp. 63–66, doi: 10.1109/SCORED.2010.5703972.

[11] G. Celik, H. Celebi, and G. Tuna, “A novel RSRP-based E-CID positioning for LTE networks,” in Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Valencia, Spain, Jun. 2017, pp. 1689–1692, doi: 10.1109/IWCMC.2017.7968638.

[12] B. Li, I. J. Quader, and A. G. Dempster, “On outdoor positioning with Wi-Fi,” J. Global Position. Syst., vol. 1, no. 13, pp. 18–26, 2008.

[13] P. Zhao et al., “miD: Tracking and identifying people with millimeter wave radar,” in Proc. 15th Int. Conf. Distrib. Comput. Sens. Syst. (DCOSS), Santorini Island, Greece, May 2019, pp. 33–40, doi: 10.1109/DCOSS.2019.00028.

[14] J. Pegoraro and M. Rossi, “Real-time people tracking and identification from sparse mm-Wave radar point-clouds,” IEEE Access, vol. 9, pp. 78504–78520, 2021.

[15] M. Kishida, K. Ohguchi, and M. Shono, “79 GHz-band high-resolution millimeter-wave radar,” Fujitsu Sci. Technol. J., vol. 51, no. 4, pp. 55–59, 2015.

[16] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, “Deep learning in video multi-object tracking: A survey,” Neurocomputing, vol. 381, pp. 61–88, Mar. 2020.

[17] N. Wojke, A. Bewley, and D. Paulus, “Simple online and real-time tracking with a deep association metric,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.

[18] S. Lee and E. Kim, “Multiple object tracking via feature pyramid siamese networks,” IEEE Access, vol. 7, pp. 8181–8194, 2019, doi: 10.1109/ACCESS.2018.2889442.

[19] O. Urbann, O. Bredtmann, M. Otten, J.-P. Richter, T. Bauer, and D. Zibriczky, “Online and real-time tracking in a surveillance scenario,” Jun. 2021, arXiv:2106.01153.

[20] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, “Fairmot: On the fairness of detection and re-identification in multiple object tracking,” Int. J. Comput. Vis., vol. 129, no. 11, pp. 3069–3087, 2021.

[21] S.-H. Bae, “Online multi-object tracking with visual and radar features,” IEEE Access, vol. 8, pp. 90324–90339, 2020, doi: 10.1109/ACCESS.2020.2994000.

[22] F. O. A. et al., “Vision and radar multi tracking for pedestrian by fusion of millimeter wave radar and vision,” in Proc. IEEE Radar Conf (RadarConf), Atlanta, GA, USA, May 2021, pp. 1–6, doi: 10.1109/RadarConf2147009.2021.9455185.

[23] X. Guo, J. D. Ju, G. Gao, and W. Wang, “Pedestrian detection based on fusion of millimeter wave radar and vision,” in Proc. Int. Conf. Artif. Intell. Pattern Recognit. (ICIP), Beijing, China, 2018, pp. 38–42, doi: 10.1109/ICIP.2018.8628868.

[24] T. Sunami, S. Iihara, Y. Koda, T. Nishio, and K. Yamamoto, “Computer vision-assisted single-antenna and single-anchor RSSI localization harnessing dynamic blockage events,” Dec. 2021, arXiv:2107.04770.

[25] T. Murakami, K. Yoshikawa, A. Yamaguchi, and H. Shinbo, “Blockage prediction in an outdoor mm-Wave environment by machine learning employing a top view image,” presented at the IEEE Annu. Int. Conf. Personal, Indoor Mobile Radio Commun. (PIMRC), Sep. 2022.

[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 2020, arXiv:2004.10934.

[27] T.-Y. Lin et al., “Microsoft coco: Common objects in context,” in Proc. Eur. Conf. Comput. Vis., 2014, pp. 740–755.

[28] L. Zheng et al., “Mars: A video benchmark for large-scale person re-identification,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 686–684.

[29] R. Watanabe, J. Chen, T. Konno, and S. Naito, “Accurate background subtraction using dynamic object presence probability in sports scenes,” in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), Milan, Italy, Jan. 2021, pp. 2521–2528, doi: 10.1109/ICPR48066.2021.9412754.

[30] M. Kristan et al., “The eighth visual object tracking VOT2020 challenge results,” in Proc. Eur. Conf. Comput. Vis., Cham, Switzerland, 2020, pp. 547–601, doi: 10.1007/978-3-030-68238-5_39.
[41] E. Marchand, H. Uchiyama, and F. Spindler, “Pose estimation for augmented reality: A hands-on survey,” IEEE Trans. Vis. Comput. Graphics, vol. 22, no. 12, pp. 2633–2651, Dec. 2016, doi: 10.1109/TVCG.2015.2513408.

[42] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A tutorial on beam management for 3GPP NR at mmWave frequencies,” IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 173–196, 1st Quart., 2019, doi: 10.1109/COMST.2018.2869411.

SHOICHIRO MIHARA was born in Takarazuka, Japan, in 1990. He received the B.S. and M.S. degrees in engineering science from the Osaka University, Osaka, Japan, in 2012 and 2014, respectively. He joined KDDI Corporation in 2014, and since 2017, he has been a Research Engineer with KDDI Research, Inc. His research interests include augmented reality, image recognition, and human-robot interaction. He was a recipient of the IPSJ Yamashita SIG Research Award in 2018 and the Best Paper Award of the IEEE Wireless Communications and Networking Conference in 2021.

TAKAHIDE MURAKAMI received the B.E., M.E., and Ph.D. degrees in communication engineering from Tohoku University, Sendai, Japan, in 2002, 2004, and 2007, respectively. He joined KDDI Corporation in 2007 and has been with KDDI Research, Inc. (formerly KDDI R&D Laboratories, Inc.). His research interests include wireless communications and radio access network.

AKIRA YAMAGUCHI (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Waseda University in 1989 and 1991, respectively. He joined KDDI Corporation (currently KDDI Corporation) in 1991, where he conducted research into mobile access networks and ubiquitous networks. He was on secondment to Advanced Telecommunications Research Institute International from 2005 to 2008. He is currently a Research Manager of Advanced Radio Application Laboratory, KDDI Research, Inc. His research interests include radio access network and space communications.

HIROYUKI SHINBO received the B.S. degree in electro information communication and the M.S. degree from the Graduate School of Information and System, University of Electro-Communications, Tokyo, Japan, in 1997 and 1999, respectively. He joined KDD Corporation (currently KDDI Corporation) in 1999 and works with KDD R&D Laboratory (currently KDDI Research, Inc.). He was seconded to work in Advanced Telecommunications Research Institute International from 2013 to 2016. He is currently a Senior Manager of Advanced Radio Application Laboratory, KDDI Research, Inc. His research interests include beyond 5G/6G systems (especially radio access network), TCP/IP, flying base stations, network operation systems, and space communications.