Handover Management for mmWave Networks With Proactive Performance Prediction Using Camera Images and Deep Reinforcement Learning

Yusuke Koda, *Student Member, IEEE*, Kota Nakashima, *Student Member, IEEE*,
Koji Yamamoto, *Member, IEEE*, Takayuki Nishio, *Member, IEEE*, and Masahiro Morikura, *Member, IEEE*

***Abstract*—For millimeter-wave networks, this paper presents a paradigm shift for leveraging time-consecutive camera images in handover decision problems. While making handover decisions, it is important to predict future long-term performance—e.g., the cumulative sum of time-varying data rates—proactively to avoid making myopic decisions. However, this study experimentally notices that a time-variation in the received powers is not necessarily informative for proactively predicting the rapid degradation of data rates caused by moving obstacles. To overcome this challenge, this study proposes a proactive framework wherein handover timings are optimized while obstacle-caused data rate degradations are predicted before the degradations occur. The key idea is to expand a state space to involve time-consecutive camera images, which comprises informative features for predicting such data rate degradations. To overcome the difficulty in handling the large dimensionality of the expanded state space, we use a deep reinforcement learning for deciding the handover timings. The evaluations performed based on the experimentally obtained camera images and received powers demonstrate that the expanded state space facilitates (i) the prediction of obstacle-caused data rate degradations from 500 ms before the degradations occur and (ii) superior performance to a handover framework without the state space expansion.**

***Index Terms*—Millimeter-wave communication, deep reinforcement learning, handover management, proactive prediction, camera image.**

I. Introduction

Millimeter-wave (mmWave) communications are expected to play an important role in next-generation wireless networks, such as fifth-generation mobile networks or wireless local area networks [1]–[4]. The exploitation of wider spectrum bands in the mmWave band facilitates multi-gigabit data transmission and thereby supports communication services, such as ultra-high-definition televisions [2], virtual reality (VR) [5], or augmented reality (AR) [6] that require the multi-gigabit data transmission.

However, designing robust millimeter networks is quite challenging owing to the high frequency of the mmWave bands. The distinct feature of mmWave communication is the use of directional antennas to compensate for high path loss in mmWave bands. The directional antennas can be implemented by embedding many small antenna elements designed for mmWave in a limited physical space in mobile terminals as well as mmWave base stations (BSs). However, the antenna directivity makes mmWave communication links vulnerable to link blockage caused by moving obstacles. The link blockage suddenly penalizes the mmWave link budget by 20–30 dB in the case of data transmission comprising the use of directional antennas [7], [8]. The sudden and damaging degradation in the received power causes frequent interruptions within a transmission of streamed data, which is a crucial problem for VR/AR applications.

To overcome the blockage problem and provide reliable mmWave communications, a handover between multiple BSs is envisioned as a promising scheme [9]–[13]. By performing handovers at appropriate times, the decreased link budget can be compensated with another BSs. In next-generation cellular networks, an increasing number of mmWave BSs will be deployed to ensure a line-of-sight (LOS) path between a mobile terminal and one of the deployed BSs; hence, designing a decision problem concerning when and to which BS a handover should be triggered, which is referred to as a *handover decision problem*, is an important research direction.

In a handover decision problem, it is important to predict a future long-term performance, e.g., the time-average or cumulative sum of the data rates prior to performing a handover in order to avoid making myopic decisions [14]–[19]. This is because a handover involves a service disruption caused by procedures that are necessary for changing association and for data forwarding to a BS to which handover is performed [20]. Performing handovers based on a short-term performance, i.e., making myopic decisions, results in frequent handovers that may cause the overall long-term performances to deviate [16], [17]. Thus, a future long-term performance in both the currently associated BS and the candidate BSs should be predicted prior to triggering a handover, and a handover decision rule should be formed such that the predicted performance is maximized.

In addition to the avoidance of redundant handovers, predicting future long-term performance is beneficial to avoid

Manuscript received June 26, 2019; revised November 2, 2019; accepted December 11, 2019. Date of publication December 24, 2019; date of current version June 9, 2020. This work was supported in part by JSPS KAKENHI Grant Numbers JP17H03266, JP18H01442, and KDDI Foundation. The associate editor coordinating the review of this article and approving it for publication was T. Chen. *(Corresponding author: Yusuke Koda.)*

The authors are with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: koda@imc.cce.i.kyoto-u.ac.jp; nakashima@imc.cce.i.kyoto-u.ac.jp; kyamamot@i.kyoto-u.ac.jp; nishio@i.kyoto-u.ac.jp; morikura@i.kyoto-u.ac.jp).

Digital Object Identifier 10.1109/TCCN.2019.2961655

a lower data rate situation, particularly in mmWave communications, what is a main topic of this study. Due to moving obstacles, mmWave links experience faster data rate variation compared to microwave links. Given such constraints, the data rate provided by current BS may be lower than the rates provided by another BS before a handover execution is performed if the handover occurs after data rate variation. As a result, severe loss of the data rate takes place. By predicting future data rates within a longer time horizon, handover is performed to avoid data rate loss, so *proactive handover* is beneficial rather than detrimental.

However, it is still challenging to predict the future long-term performance in mmWave links proactively under the condition that moving obstacles cause the rapid variation of received powers or data rates. This is because the sudden variation exhibits little prior indications in the radio frequency (RF) signal domain such as received power samples and channel state information.[1] Thus, to predict the rapid variations in data rates or received powers proactively, we should utilize other information domains that provide more informative features for predicting such variations.

To address this challenge, this study develops a proactive framework wherein future data rate degradations caused by moving obstacles are predicted from several hundreds of milliseconds before the degradation occurs and the handover timings are optimized based on the predicted values. The key idea is to leverage the time consecutive camera images[2] and to use deep reinforcement learning (RL). Time consecutive camera images comprise information of the spatiotemporal dynamics of moving obstacles, which exhibits informative features for predicting the future obstacle-caused degradation of data rates in mmWave links. The optimization of the handover timings while predicting such future degradations based on camera images is a new challenge. We incorporate the usage of camera images into the RL-based handover frameworks (discussed in detail in the following section) by expanding the state space such that the state involves camera images. Moreover, by using a deep RL [27], we overcome the difficulty in handling the large dimensionality of the state space incurred by the state space expansion.

The most closely related work was presented at the IEEE CCNC 2020 [28], while the contributions of this paper are different from those in [28]. As discussed later in detail, the main contribution of this study is the presentation of proactive prediction in handover decision problems by leveraging camera images. Meanwhile, [28] addressed the issue of how to compensate for a blind spot of a single camera while applying the framework proposed in this paper and proposed a multi-camera operation. Thus, the contribution of [28] is to demonstrate the feasibility of incorporating the multi-camera operation into the framework proposed in this paper.

The contributions of this paper are summarized as follows:

- We highlight that the variation in the received powers before blockage events is not necessarily informative in predicting future data rate degradation in mmWave links. To confirm this, we obtained experimentally a received power time series that exhibits the variation and predicted the cumulative sum of future data rates with the RL method based on the state of received power obtained.
- Based on the following two ideas, we propose a proactive framework wherein handover timing is optimized while the degradation in data rate caused by obstacles is predicted within hundreds of milliseconds before degradation. The first idea is to expand the states such that the states comprise time-consecutive camera images, which provide informative features for predicting degradations, i.e., spatiotemporal dynamics of moving obstacles. The second idea is to leverage deep RL to overcome the computational complexity of learning the optimal handover policy incurred by the expanded state.

The rest of this paper is organized as follows. Section III presents an experimental evaluation of the received-power-based prediction of the cumulative sum of the future data rates in a handover decision problem. Section IV presents our image-based handover framework, which leverages time-consecutive camera images in a handover decision problem. Finally, Section V presents concluding remarks.

It should be noted that Sections III and IV are related to each other. The former provides a baseline for the framework without camera images to be compared with the proposed image-based handover framework, and the latter details the image-based handover framework. In concrete, in Section III, the problem of a received power-based handover framework summarized in the first contribution is highlighted. This received power-based handover framework is referred to as baseline without camera images, and compared to the proposed image-based handover framework in Section IV. In Section IV, focusing on the highlighted problem, we propose the image-based handover framework presented in the second contribution. Subsequently, we discuss the difference between the handover policies learned with and without camera images by comparing our image based-handover framework with the received power-based handover framework.

II. Related Works

A. Handover Decision Problems

In many studies, handover decision-making problems or cell selection problems in heterogeneous microwave networks or millimeter wave networks were formulated with the objective of maximizing the future long-term performance [14]–[19], [29]. The authors of [14], [19], [29]

[1] With regard to the degradation of the received powers, there is a slight fluctuation in the received powers within 100 ms prior to the degradation, which is known as diffraction effects [8], [21]–[23]. Hence, by analyzing the time-series of the received powers, we can predict the degradation from at most 100 ms before the occurrence [24]. Nonetheless, it is worthwhile utilizing the camera image domain for the two reasons. First, as experimentally confirmed in this paper, based only on the variation, the degradation cannot be necessarily predicted in a proactive manner. Second, the degradation should be predicted earlier because the service interruption incurred by a handover could be several hundreds of milliseconds long [25].

[2] We used depth images pixels of which are used to measure the distance between the obstacles and the camera [26]. Depth images allow us to obtain geometric relations between components within the scene. In the following discussion, we consider that the depth images are available to a network controller.

TABLE I
COMPARISON OF HANDOVER-RELATED PREVIOUS WORKS

	[14], [19], [29]	[16]–[18]	[15], [30], [31]	[32]	[33]–[35]	[36], [37]	**This paper**
Optimization of handover timings	Yes	Yes	Yes	Yes	No	No	**Yes**
Frequency band	Microwave	Microwave	mmWave	mmWave	mmWave	mmWave	**mmWave**
Usage of camera images	No	No	No	No	Yes	Yes	**Yes**
Proactive blockage prediction	No	No	Yes	Yes	Yes	Yes	**Yes**
Approach	DP	RL	DP	RL	Heuristic	SL	**RL**

designed the optimal cell selection problem in heterogeneous wireless networks with the objective of maximizing the weighted sum of the network bandwidth and network delay via the Markov decision process (MDP) models or optimal control models. The optimal strategies are provided via dynamic programming (DP) techniques. In [15], optimal cell selection in mmWave networks was proposed to maximize the long-term throughput or total received data in a mobile terminal using a similar approach. The authors of [16]–[18], [38] applied an RL algorithm to learn the optimal cell selection with the objective of maximizing the long-term quality of experiences or channel capacities, wherein an optimal strategy of cell selection can be learned without prior knowledge of the transition probability of the channel states or received powers. However, in the aforementioned studies, a decision process was considered wherein a decision maker makes a decision based on a current network state such as the channel information, received power, or network bandwidth. These studies did not detail the challenge of predicting the future long-term performance in mmWave links under the condition of moving obstacles causing blockage effects and received powers at a station (STA) or the BSs and the data rates in the mmWave links undergoing rapid degradation.

Other works have addressed handover decision-making problems in mmWave networks by using user mobility information or pedestrian mobility information [30]–[32]. User mobility information facilitates the prediction of future data rates in mmWave links with blockage effects that occur when users are entering areas blocked by static obstacles [30], [31]. However, the proactive prediction of the data rate degradations caused by moving obstacles is not addressed. In our previous work [32], we addressed handover decision problems based on the positions and velocities of a moving pedestrian. However, the proposal is not applicable to handover decision problems wherein more pedestrians cause blockage effects because of the challenge of capturing the spatial features of each pedestrian such as their height or shape. In contrast, our current proposal uses camera images that comprise spatial information, thereby capturing the spatial features of moving obstacles.

B. Camera Image-Based Frameworks in mmWave Networks

The authors of [33]–[35] have conceptualized a camera-assisted proactive handover system for mmWave networks. The camera images are employed to predict the occurrence of blockage effects caused by pedestrians approaching a LOS path between a BS and an STA. The experiments conducted in these works demonstrated that using camera images, a handover can be triggered several seconds before blockage the occurrence of the blockage effects. However, the methods embedded in the experiments are focused on predicting the timings at which blockage effects occur, and they do not quantitatively predict the future data rate degradation caused by pedestrians. As discussed in the previous section, the optimal handover requires a prediction of the future long-term performances; hence, the aforementioned methods cannot provide the optimal solution to handover decision problems.

Motivated by the issue, a novel method for quantitatively predicting a future received power value in mmWave communications was proposed in [37]. The method predicts a received power value from several hundreds of milliseconds before the value is observed. In this method, camera images are mapped, via a supervised learning (SL) technique, to a future received power value that is obtained several hundreds of milliseconds after the camera images are obtained. However, the prediction method in [37] is not specific to handover decision problems. While the method in [37] can be used to predict a future data rate at a certain time period, the optimization of the handover timings requires a different prediction, i.e., the prediction of the expected cumulative sum of future long-term data rates as confirmed in the previous studies discussed in Section II-A. Thus, the method in [37] cannot be necessarily adopted directly in handover decision problems. Table I summarizes the main aspects of the previous works related to this paper.

III. RECEIVED POWER-BASED HANDOVER FRAMEWORK

The main objective of this section is to highlight that the future degradation of data rates in mmWave links caused by moving obstacles cannot necessarily be predicted based only on a variation in received powers. To illustrate this point, we perform a prediction of the cumulative sum of the future data rates using RL with the state information of the experimentally obtained received powers. We will refer to the received power-based handover framework as a baseline without camera images, to be compared with the proposed image-based handover framework in Section IV-B4. First, we provide an overview of the RL. Then, we present the decision process considered in this experiment. Finally, we provide an experimental study of the prediction based on the received powers.

A. Overview of RL

General RL algorithms are performed over an MDP. An MDP consists of the following four elements: a state space $\mathcal{S}$, an action space $\mathcal{A}$, a reward function $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$, and transition probabilities $q : \mathcal{S} \times \mathcal{A} \to \Omega(\mathcal{S})$, where $\Omega(\mathcal{S})$ denotes the collection of the probability distribution over $\mathcal{S}$. At each decision epoch $t \in \mathbb{N}$, a decision maker observes the state information $s_t \in \mathcal{S}$. Subsequently, the decision maker selects an action on the basis of the *policy* $\pi : \mathcal{S} \to \mathcal{A}(s_t)$, where $A(s_t) \subseteq \mathcal{A}$ denotes the set of possible actions when the state s_t is observed. Given the current state s_t and selected action $a_t \in \mathcal{A}(s_t)$, the state transitions to $s_{t+1} \in \mathcal{S}$ at the next decision epoch $t + 1$ according to the transition probability $q(s_{t+1}, s_t, a_t)$; thereafter, the decision maker is given a reward $r(s_{t+1}, a_t, s_t)$.

The objective of the decision maker is to determine the optimal policy $\pi^\star$ that maximizes the total expected discounted reward. The optimal policy satisfies the following condition:

$$\mathbb{E}\left[\sum_{t'=0}^{\infty} \gamma^{t'} r(s_{t+t'+1}, \pi^\star(s_{t+t'}), s_{t+t'}) \,\middle|\, s_t = s\right] \geq \mathbb{E}\left[\sum_{t'=0}^{\infty} \gamma^{t'} r(s_{t+t'+1}, \pi(s_{t+t'}), s_{t+t'}) \,\middle|\, s_t = s\right], \tag{1}$$

$\forall s \in S$ and $\forall \pi$, where $\gamma \in [0, 1)$ represents the discount factor. In the MDP wherein $\mathcal{S}$ and $\mathcal{A}$ are both countable non-empty sets, there exists at least an optimal policy [39].

To obtain the optimal policy in an MDP, it is sufficient to obtain the optimal action-value function $Q^\star : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. The optimal action-value function is defined as follows:

$$Q^\star(s, a) := \mathbb{E}_{s'}\big[r(s', a, s) + \gamma V^\star(s') \mid s, a\big], \quad s \in \mathcal{S}, a \in \mathcal{A}(s), \tag{2}$$

where $\mathbb{E}_{s'}[\ \cdot \mid s, a]$ denotes the expectation operator under the transition probability $q(s', s, a)$ and $V^\star(s)$ denotes the left-hand side in (1). This is attributed to the fact that the optimal action-value function is related to the optimal policy as follows [39]:

$$\pi^\star(s) = \underset{a \in \mathcal{A}(s)}{\arg\max}\, Q^\star(s, a). \tag{3}$$

In other words, the policy that selects the action that maximizes $Q^\star(s, a)$ is optimal. In this study, the optimal action-value function is learned using deep RL [27].

B. States, Actions, Rewards, and State Transition Rules

We present the decision process considered in this experiment by detailing the states, actions, rewards, and state transition rules. In the process, a network controller makes handover decisions in the mmWave networks based on the received power values. We consider a mmWave network wherein multiple mmWave BSs and an STA are deployed. There exist obstacles that block the LOS path between the STA and the BS associated with the STA. We also consider the decision of whether a handover should be triggered with respect to the time length of service disruption. The communication between the BS and STA can be disrupted because of the necessary procedures for the association, which involves beam alignment and for data forwarding to a BS to which a handover is performed [25], [40], [41]. We define the duration for which the communication is disrupted as the service disruption time T_{dis}.

It should be noted that in many existing studies [14], [15], [19], [30], [31], the handover decision process was formulated as an MDP, although it was assumed that the interval between the decision epochs was several seconds long, which is longer than a realistic service disruption time of several tens or hundreds of milliseconds [40]. Hence, the service disruption occurs within an interval between the successive decision epochs. However, the assumption of the large interval is not suitable for predicting the blockage effects that moving obstacles cause within several hundred milliseconds [23]. Hence, we reformulate the problem wherein an interval between the successive decision epochs is shorter than several tens or hundreds of milliseconds, and several decision epochs could be within a service disruption.

1) States: For the network controller to detect blockage effects based on received powers, we design the states such that they include the received power values. Let the number of time-consecutive received power values used in making handover decisions be denoted by *N*. We set the state space as follows:

$$\mathcal{S}_{\text{rp}} := \underbrace{\mathcal{P} \times \cdots \times \mathcal{P}}_{N} \times \mathcal{J} \times \mathcal{C}. \tag{4}$$

In (4), $\mathcal{P} \subseteq \mathbb{R}^J$ denotes the set of all possible received powers observed at all BSs, $\mathcal{J} := \{1, \ldots, J\}$ denotes the set of the BS indices, and $\mathcal{C} := \{\, c \mid c \in \mathbb{Z}, 0 \leq c \leq \lfloor T_{\text{dis}}/\tau \rfloor \,\}$ denotes the set of the remaining decision epochs until the service disruption time ends, where J denotes the number of the deployed BSs, $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{R}$ denotes the floor function, and τ denotes the interval between the successive decision epochs.

Let $s_t = (p_t, p_{t-1}, \ldots, p_{t-N+1}, j_t, c_t) \in \mathcal{S}_{\text{rp}}$ denote the state at the decision epoch *t*. The element $p_{t-k} \in \mathcal{P}$ for $k \in \{0, 1, \ldots, N-1\}$ is set as the received power observed at the decision epoch $t - k$. The element $j_t \in \mathcal{J}$ is set as the index of the BS associated with the STA. The element $c_t \in \mathcal{C}$ is set as the number of remaining decision epochs that the network controller experiences until the service interruption ends. When the decision epoch is not within the service disruption time, c_t is set as zero.

2) Actions: We let the set of possible actions $\mathcal{A}(s_t)$ be as follows:

$$\mathcal{A}(s_t) := \begin{cases} \mathcal{J}, & c_t = 0; \\ \{j_t\}, & c_t \neq 0. \end{cases} \tag{5}$$

In other words, the controller selects one of the BSs when the decision epoch is not within the service disruption time; otherwise, the controller selects only the index of the BS to which a handover is performed.

3) Reward: We set the reward as a performance metric in the link provided by the BS that is currently associated with the STA with the exception that when the next decision epoch $t + 1$ is within the service disruption duration, we set the reward as zero as follows:

$$r(s_{t+1}, a_t, s_t) := \begin{cases} R_{j_{t+1}, t+1}, & c_{t+1} = 0; \\ 0, & c_{t+1} \neq 0. \end{cases} \tag{6}$$

In (6), $R_{j_{t+1}, t+1}$ denotes the performance metric in the link provided by BS j_{t+1} at $t + 1$. In the performance evaluation, we set $R_{j_{t+1}}$ as the achievable data rate provided by BS j_{t+1} as discussed in Section III-C.

4) State Transition: The state transition to the next state is as follows. Let the state at epoch $t+1$ be $s_{t+1} = (p_{t+1}, p_t, \ldots, p_{t-N+2}, j_{t+1}, c_{t+1}) \in \mathcal{S}_{\mathrm{rp}}$. Evidently, the received power values $(p_{t+1}, p_t, \ldots, p_{t-N+2})$ at $t+1$ are updated by concatenating the received power values at p_{t+1} with the current values $(p_t, p_{t-1}, \ldots, p_{t-N+1})$ and removing the oldest value p_{t-N+1}. Based on the definition of the state, the term j_{t+1} is determined as follows:

$$j_{t+1} = a_t. \tag{7}$$

The term c_{t+1} is determined as follows:

$$c_{t+1} = \begin{cases} c_t - 1; & c_t \neq 0, \\ \lfloor T_{\mathrm{dis}}/\tau \rfloor; & c_t = 0, a_t \neq j_t, \\ 0; & c_t = 0, a_t = j_t. \end{cases} \tag{8}$$

It should be noted that without knowing the transition probabilities, we learn the optimal action-value function using deep RL [27]. To learn the optimal policy, we only require transition samples (s_t, a_t, r_t, s_{t+1}) that can be obtained while making decisions in the learning procedure.

We detail an example of the temporal transition of the decision process. We consider that at the decision epoch t, $s_t = (p_t, p_{t-1}, \ldots, p_{t-N+1}, 1, 0)$, i.e., the received power values $(p_t, p_{t-1}, \ldots, p_{t-N+1})$ are available, the STA is associated with BS 1, and the decision epoch is not within the service disruption time. If the controller selects action $a_t \neq 1$, i.e., a handover is performed, then the state transitions to $s_{t+1} = (p_{t+1}, p_t, \ldots, p_{t-N+2}, a_t, \lfloor T_{\mathrm{dis}}/\tau \rfloor)$. The controller is subsequently given a reward of zero because $c_t = \lfloor T_{\mathrm{dis}}/\tau \rfloor \neq 0$ (see (6)). In this case, until the service disruption time ends, the controller selects action a_t, is given a reward of zero, and decreases the last element of the state by one. Conversely, if the controller selects action $a_t = 1$, i.e., the handover is not performed, then the state transitions to the state $s_{t+1} = (p_{t+1}, p_t, \ldots, p_{t-N+2}, 1, 0)$ and the controller is then given the reward $R_{1,t+1}$.

C. Experimental Evaluation

1) Evaluated Scenario: As shown in Fig. 1, two BSs and an STA are deployed in an indoor room whose length, width, and height are 4.87 m, 5.34 m, and 2.57 m, respectively. The size of the room corresponds to that of the room where the measurement in the next section was conducted. The two BSs are operated over the 60 GHz channel. The STA is initially associated with the BS that observes a higher received power as compared to that of the counterpart when there are no obstacles within the deployed area. We term the BS that is initially associated with the STA as BS 1 and the other as BS 2. BS 2 is a candidate BS for a case in which the link between BS 1 and the STA is blocked by obstacles. In the following discussion, we detail the considered scenario including coverage area, STA mobility, channel characteristics, beamforming, initial access procedure, and beam tracking.

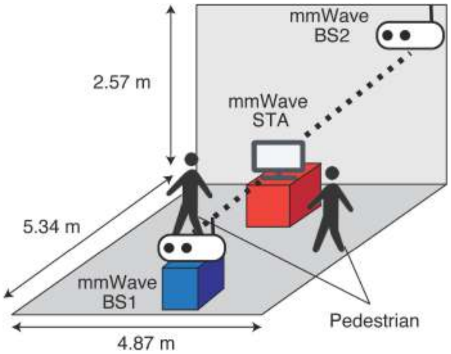


Fig. 1. Experimented scenario of mmWave links.

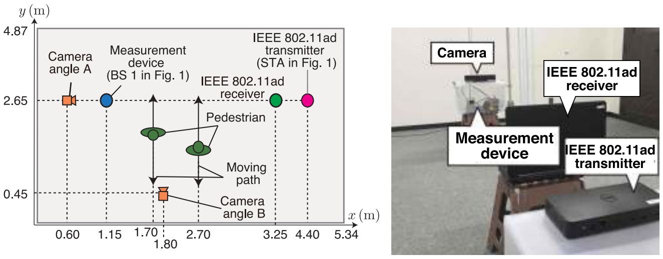


Fig. 2. Top view of the measurement environment (left) and measurement setup showing the mmWave transmitter, measurement device, and camera placed at the position A (right). The measurement device and mmWave transmitter correspond to BS 1 and the STA in Fig. 1, respectively.

Coverage area: A BS covers an entire room in Fig. 1 at least in a LOS condition, which is examined as follows. We determine whether a BS can cover a certain position or not by examining whether a commercially available IEEE 802.11ad equipped transmitter and receiver can associate with each other. In this examination, we validated that the transmitter can associate with the receiver placed 10 m apart from the transmitter in a LOS condition. Because the maximum distance between two positions is 7.67 m in the room, we can say that a BS covers the room in Fig. 1 at least in a LOS condition, which is sufficient for performing the evaluation.

STA mobility: In this evaluation, the received power varies only because of moving obstacles, not because of the STA movements to focus on the sudden variation of the data rate caused by moving obstacles. To obtain such received power samples, the experiment is arranged such that the position of the STA is static, and a handover is performed to compensate the degraded data rate caused by the obstacles rather than to support the STA mobility. The evaluation in such a scenario is sufficient for the two objectives of this study: highlights that the such sudden variations of the data rate cannot be predicted from a received power time series (provided in Section III) and demonstrates the feasibility of the proactive prediction on such variations achieved by camera images (provided in Section IV). This scenario of the static STA is reasonable for certain realistic application such as the transmission of streaming data to a wireless monitor in an office, discussed in Section IV-B1 in detail.

Channel characteristics between BS 1 and STA: The channel between BS 1 and STA is based on the measurement in the next section, where the overall characteristics are similar to "dynamic 60 GHz radio channel" [23] in terms of the variation

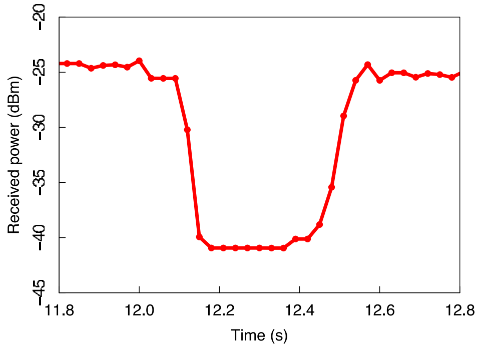


Fig. 3. An example of the variation of the received power in a blockage effect.

TABLE II
ESTIMATED DISTRIBUTION PARAMETERS CHARACTERIZING BLOCKAGE EVENTS OBSERVED IN MEASUREMENT

Values	Distribution	Estimated Distribution Parameters
$t_{\mathrm{decay,5\,dB}}$	Gaussian [43]	mean: 0.059, standard deviation: 0.034
$t_{\mathrm{rise,5\,dB}}$	Log Normal [43]	log mean: −3.01, log standard deviation: 0.195
A_{mean}	Gaussian [43]	mean: 14.2, standard deviation: 2.08
t_{D}	Weibull [43]	scale: 0.553, shape: 4.08
t_{LOS}	Weibull [42]	scale: 2.31, shape: 1.51

of the received powers only due to the moving obstacles. An example of the variation of the received power in a blockage effect is shown in Fig. 3, and the overall characteristics are as follows. In a blockage effect, the received power decreases by approximately 15 dB within 50–200 ms. Subsequently, the received power remains constant for 200–300 ms, and then it recovers to the original value within 50–200 ms.

Blockage distributions: Because the channel between BS 1 and STA is based on the measurement provided in the next section, the distributions characterizing the blockage events also depend on the measurement. To quantitatively characterize the blockage events in the measurement, in Table II, we provide the estimated distribution parameters of the following five variables, which are essential in characterizing blockage events [23], [42]. In Table II, $t_{\mathrm{decay,5\,dB}}$ denotes the duration in which signal attenuation level increases from 0 dB to 5 dB, $t_{\mathrm{rise,5\,dB}}$ denotes the duration in which signal attenuation level decreases from 5 dB to 0 dB, A_{mean} is the mean signal attenuation, and t_{D} denotes the duration of the blockage event. The value t_{LOS} is the duration wherein a LOS condition sustained. The definitions of the values follow the works of [23] and [42]. The choice of the distribution functions is based on [23] or [42], and the parameters of the distributions are determined by the maximum-likelihood estimation. The blockage events occurred within the 21% time-length relative to the whole measurement time.

Channel characteristics between BS 2 and STA: Meanwhile, the mmWave channel between BS 2 and STA is static, and it is assumed that BS 2 is free of blockages. The assumption is reasonable given that a network controller is likely to perform a handover to a BS that is not blocked by obstacles. In the following discussion, it is considered that BS 2 is at a position where pedestrians cannot block the path between the STA and BS 2, and the received power at BS 2 is constant over time. Since the focus of the evaluation is on a blockage effects between STA and BS 1, we detail the link between the STA and BS 1 in the following discussion.

Beamforming: The STA and BS 1 communicate with each other with directional antennas. Because channel characteristics between BS 1 and STA are based on the measurement, the antenna gain is also attributed to the measurement equipment. In the next section, we detail antenna gains of a transmitter and measurement device, which corresponds to the STA and BS 1, respectively.

Initial access procedure: Prior to the evaluation, we established the beam of STA based on an initial access procedure termed as iterative beam search method [41], which is used in the IEEE 802.11ad standard. This initial access procedure is because we used the a commercially available IEEE 802.11ad equipped transmitter as the STA in the measurement. Meanwhile, we established the beam of BS 1 manually such that the beams of BS 1 and the STA point towards each other. We discuss the procedure of establishing the beam directions of the STA and BS 1 in detail in the next section.

Beam tracking: In the next section, we conduct the measurement such that the STA and BS 1 do not perform a beam tracking. The aim is to eliminate the variation of the received powers due to beam tracking whose mechanisms depend on manufacturers and thereby to focus only on the sudden variations of the received powers due to moving obstacles. In the next section, we detail how the measurement is conducted such that the beam tracking is not performed.

2) Measurement Setup: We set up an IEEE 802.11ad equipped transmitter/receiver, a measurement device, and a camera as shown in Fig. 2. The transmitter and a measurement device correspond to the STA and BS 1, respectively. The transmitter and measurement device is place at the height of 0.70 and 0.85 m, respectively. The camera is placed at either position (0.60, 2.65) and (1.80, 0.45) and at the heights of 1.50 m and 1.25 m, respectively to obtain the a dataset with two different camera angles. The angle from the former position is termed angle A while the latter is termed angle B. The measurement device is equipped with a horn antenna with directivity gain of 24 dBi and the half-power beam width (HPBW) of 11 degree while the transmitter is equipped with an array antenna with size of 16, directivity gain of approximately 8 dBi, and HPBW of approximately 15 degree [44].

The beam directions of the measurement device and transmitter are established from the following procedure to configure the beams of the measurement device and transmitter such that the beams point towards each other. First, the transmitter and receiver in Fig. 2 performed, in a LOS condition, an iterative beam search wherein the beam pair was searched with a two-stage beam scanning [41] such that the receiver can benefit from the maximum received power. Through the procedure, the beam of the transmitter is configured such that the beam points towards the receiver. Subsequently, we placed the measurement device behind the receiver such that the

TABLE III
EXPERIMENTAL EQUIPMENT AND PARAMETERS

IEEE 802.11ad transmitter	Dell Wireless Dock D5000
IEEE 802.11ad receiver	Dell Latitude E5540
Spectrum analyzer	Tektronix RSA306
Down-converter	Sivers IMA FC2221V
Antenna	Sivers IMA Horn antenna, 24 dBi
Depth camera	Microsoft Kinect for Windows (Model:1656)
Channel	60.48 GHz
Sampling frequency	56 MHz
Transmit antenna gain	10 dBi [44]
Receive antenna gain	24 dBi
Measurement bandwidth W	40 MHz, 20 MHz

TABLE IV
PARAMETERS ASSOCIATED WITH RL

Discount factor, γ	0.99
Number of obtained received powers T	16860
Number of received powers used for learning T'	13500
Number of iterations of learning and evaluation	1000
Exploration rate ϵ	1–0.01 (Decreased by 0.01 per iteration)
Number of received power values in state N	2
Interval between successive decision epochs τ	30 ms
Noise power spectral density σ^2	−173 dBm/Hz
Received power at BS 2, $p_{2,t}$	−129 dBm (const.)
Batch size [27]	32
Frequency of updating the target network [27]	10000

horn antenna attached with the measurement device faced for the transmitter. Since the beam of the transmitter also points towards the measurement device, the beams of the transmitter and measurement device point towards each other.

We conduct the measurements as in [22] and obtain the received powers and camera images. The mmWave transmitter transmits signals at the carrier frequency of 60.48 GHz to the receiver, and subsequently, the measurement device behind the receiver measures the power of a part of the signals [22]. The transmitted signals are considered as uplink signals from the STA to BS 1. In this environment, two pedestrians walk along the moving path in Fig. 2 and obstruct the path between the transmitter and measurement device. Tables III summarizes the experimental equipment and parameters associated with the experiment.

We conduct the measurement such that the beam tracking between the measurement device and transmitter is not performed. This eliminates the variation of the received powers due to beam tracking whose mechanisms depend on manufacturers, and we can thereby focus only on the sudden variations of the received powers due to moving obstacles. The details are as follows: The measurement device was located behind the receiver, and the pedestrians traveled between the receiver and measurement device indicated in Fig. 2. This arrangement prevented the receiver and transmitter from performing beam tracking because the received power at the receiver was not altered even when the LOS path between the receiver and measurement device was blocked. In this situation, the beam direction of the transmitter is almost fixed. Consequently, the beam directions in the transmitter and measurement device were also fixed, wherein the beam tracking between them was not performed.

It should be noted that the camera images obtained in this experiment are not used in this evaluation but are used in the next section. This evaluation provides the baseline that does not utilize camera images and only uses the time series of received powers to decide handover timings.

3) Procedure of Performing Decision Process: We divide the received powers into two parts, and the individual parts are used for the learning and performance evaluation, respectively. Let the obtained received power values be denoted by $(p_{1,t})_{t\in\mathcal{T}}$, where $p_{1,t}$ denotes the received power obtained at the *t*th observation, and $\mathcal{T} = \{1, 2, \ldots, T\}$ denotes the set of the time indices. We divide $\mathcal{T}$ into the following two subsets: $\mathcal{T}_1 = \{1, 2, \ldots, T'\}$ and $\mathcal{T}_2 = \{T'+1, T'+2, \ldots, T\}$, where $1 < T' < T$. We use $(p_{1,t})_{t\in\mathcal{T}_1}$ to learn the optimal action-value function and $(p_{1,t})_{t\in\mathcal{T}_2}$ to evaluate the learned policy. In the following discussion, we denote p_t as the received power values observed at BS 1 and at BS 2, i.e., $(p_{1,t}, p_{2,t})$, where $p_{2,t}$ is the received power value observed at BS 2 and is constant $\forall t \in \mathcal{T}$.

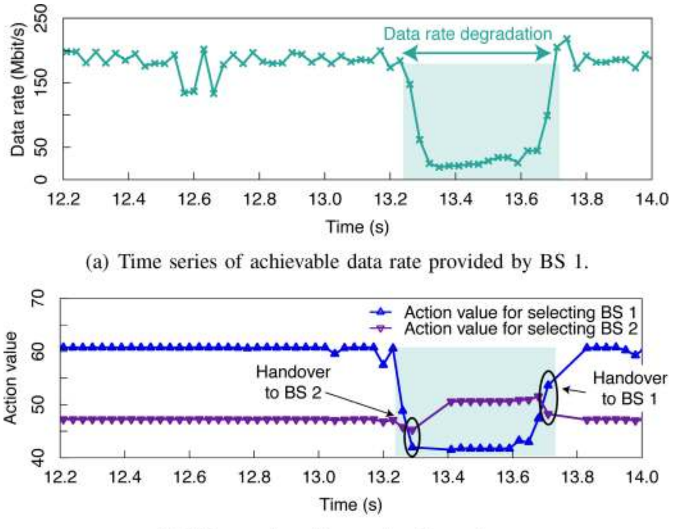
We simulate the decision process in the learning procedure using $(p_t)_{t\in\mathcal{T}_1}$. The decision epoch is set as the time step at which a received power value is obtained. The decision process starts at the time step at which $p_{1,N}$ is observed. The STA is initially associated with BS 1 and the time at which the process starts is not within a service disruption time, i.e., $j_1 = 1$ and $c_1 = 0$. Thus, the state s_N is set as $(p_N, p_{N-1}, \ldots, p_1, 1, 0)$. The action a_N is selected according to a heuristic ϵ-greedy policy [27]; the next state s_{N+1} is then set such that it includes the images $(p_{N+1}, p_N, \ldots, p_2)$, j_{N+1}, and c_{N+1}, where j_{N+1} and c_{N+1} are determined based on a_N as shown in (7) and (8). The procedure is iterated, and it then ends when the state includes the received power values $p_{T'-1}$.

The performance metric $R_{j,t+1}$ for $j \in \mathcal{J}$ in (6) is set as the achievable data rate provided by BS *j*, which is associated with the STA and is calculated as follows. The metric $R_{j,t+1}$ is calculated by the Shannon capacity formula with the received power value $p_{j,t+1}$ as follows:

$$R_{j,t+1} = W \log_2\Big(1 + \frac{p_{j,t+1}}{\sigma^2 W}\Big),$$

where σ^2 denotes the noise power spectral density. It should be noted that the metric at $R_{2,t+1}$ is set as a constant value based on the assumption that the received power at BS 2 is a constant over time. It also should be noted that the reward is re-scaled by multiplying 10^{-8} to range from 0 to 10 in performing deep RL.

We evaluate the performance of the learned policy in the following step termed as performance test. In the performance test, we simulate a decision process using the same procedure as the learning procedure with the exception that we use $(p_t)_{t\in\mathcal{T}_2}$, and the action is selected according to a greedy policy [39]. We calculate the time average of the reward as a performance metric of the learned policy. We then iterate the learning and evaluation using the same data set. We evaluate



(a) Time series of achievable data rate provided by BS 1.

(b) Time series of learned action values.

Fig. 4. Comparison between time series of achievable data rate provided by BS 1 and that of learned action values. The action value, i.e., the estimation of the cumulative sum of the future data rates, decreases after the data rate decreases, which indicates that the degradation in the data rate provided by BS 1 cannot be predicted in a proactive manner.

the policy that achieves the highest average reward throughout the iterations.

It should be noted that the handover policy is learned via deep RL [27] with a neural network (NN) that is different from that shown in Fig. 6 (discussed later). We simplify the NN architecture because the input of the NN in this scenario comprises several elements—the four elements in the evaluation. We replace the combination of the convolutional neural network (CNN) and long short-term memory (LSTM) in Fig. 6 with a fully connected multi-layer with eight hidden units and 32 output units, where the two layers are activated using rectified linear units [45]. The weights of the NN are optimized by Adam optimizer with the learning rate of 1.0×10^{-3} and the decaying rate parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The parameters associated with the deep RL are summarized in Table IV.

4) Results: In this experiment, it was shown that the obstacle-caused degradation in the data rates could not be predicted in a proactive manner by analyzing the learned action-value function in Fig. 4. Fig. 4 shows the time series of the achievable data rate provided by BS 1 and the corresponding learned action values. The data rate provided by BS 1 is degraded from approximately 13.25 s to 13.70 s because a pedestrian walks between the measurement device and transmitter. First, we observe in Fig. 4 (a) that the data rate oscillates within approximately 90 ms before the degradation occurs, and thus, the time-variation in the received powers is successfully observed before the degradation, as confirmed in other propagation experiments [8], [21]–[23]. However, the action value decreases sharply after the degradation in the data rate provided by BS 1. As the action value is defined as the expected sum of the future performance, we can conclude that the obstacle-caused degradation of data rates in a mmWave link cannot necessarily be predicted proactively based only on the variation in the received powers.

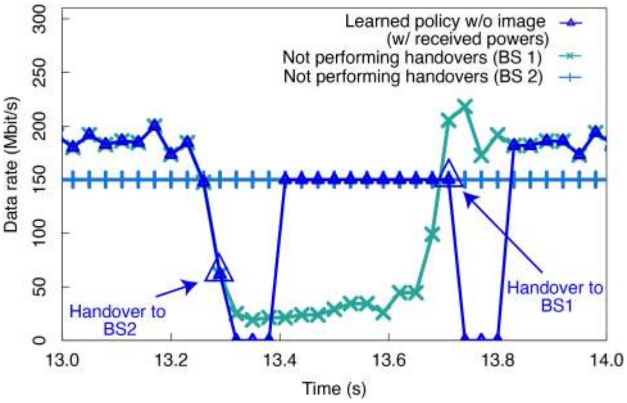


Fig. 5. Example of handover timing when $T_{\mathrm{dis}} = 0.09$ s. Handovers are performed after the variation in the data rate provided by BS 1.

Owing to the characteristics of the action-value functions, handovers are performed after the variation in the data rate provided by BS 1. Fig. 5 shows an example of the time-varying data rate provided by our image-based handover framework when the service disruption time $T_{\mathrm{dis}} = 0.09$ s. It can be observed that a handover is performed after the variation in the data rate, the degradation is experienced within approximately 60 ms. If the handover is performed earlier, we can prevent the occurrence of the degradation in the data rate provided by BS 1 and enhance the time-average of the data rates.

It should be noted that, if we observe the received powers to have a short time-resolution, e.g., one millisecond, we could predict the degradation in the data rate provided by BS 1 from based on the time-variation in the received powers that occurred before the degradation. However, as the time-variation occurred within approximately 90 ms, the degradation cannot be predicted from several hundreds of milliseconds before the degradation. This example motivated us to develop a framework using other state information that exhibits more informative features for predicting even such degradation in the data rates in a proactive manner.

IV. Image-Based Handover Framework

This section details a proactive framework wherein the handover timings are decided while the future degradation in the data rates is predicted in a proactive manner. First, to enable the proactive prediction, we expand the state information such that the state includes time-consecutive camera images. Using the time-consecutive camera images, we can capture the spatiotemporal dynamics of obstacles that are informative for predicting the degradation. We then demonstrate that with the expansion of the state space, the degradation can be predicted from several hundreds of milliseconds in advance and confirm that a performance gain is realized owing to the proactive prediction.

A. State Space Expansion for Proactive Prediction

For the network controller to leverage camera images for making handover decisions, we expand the state space in the previous section such that the state includes consecutive camera images. Let the number of time-consecutive camera images used in making handover decisions be denoted by N. We set the state space as follows:

$$\mathcal{S}_{\mathrm{img}} := \mathcal{S}_{\mathrm{rp}} \times \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{N}, \tag{9}$$

where $\mathcal{X}$ denotes the set of all possible images. It should be noted that we consider the same actions, rewards, and state transition rules as (5)–(8), respectively, to obtain a fair comparison of the performances achieved with the state space $\mathcal{S}_{\mathrm{rp}}$ and $\mathcal{S}_{\mathrm{img}}$, respectively.

The state design enables an RL to predict the future data rate degradations in mmWave links caused by moving obstacles and facilitates the maximization of the expected cumulative sum of the future data rates as in (1). This is because the state involving time-consecutive camera images reflects the spatiotemporal dynamics of the moving obstacles—for example, the dynamics of the obstacles approaching a LOS path—thus, reflecting the behavior of the data rates provided by deployed BSs at the future decision epochs $t+1, t+2, \ldots$, which may comprise the decision epochs, in which one of the BSs is blocked. We demonstrate that the novel state design allows us to predict the degradation in the data rates caused by moving obstacles from several hundreds of milliseconds before the degradation occurs in the following section.

In the evaluation, we do not use the time-series of the received power values in the state information because the information of the received power values may be redundant in the case of the existence of consecutive camera images. This is because the camera images $x_t, x_{t-1}, \ldots, x_{t-N+1}$, reflect the spatial features at the decision epochs $t, t-1, \ldots, t-N+1$, which may also be informative for capturing the received power values $p_t, p_{t-1}, \ldots, p_{t-N+1}$ because the received power values are heavily dependent on the spatial features, such as the distance between a transmitter and receiver and the positions and shapes of obstacles that obstruct the path between the receiver and transmitter [21]. Thus, in this evaluation, we consider the state space as follows:

$$\hat{\mathcal{S}}_{\mathrm{img}} := \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{N} \times \mathcal{J} \times \mathcal{C}.$$

That is, we omit the sets of received power values $\mathcal{P}$ from the state space $\mathcal{S}_{\mathrm{img}}$.

B. Experimental Evaluation

We evaluate the image-based handover framework discussed above. The objective of this evaluation is to verify the feasibility of the proactive prediction on data rate variation caused by moving obstacles if a camera is available for performing the prediction. Hence, further issues incurred by introducing cameras, such as costs for camera installments, were left aside, and we perform the evaluation focusing on the objective. In the next section, we detail the evaluated scenario, and we provide realistic scenarios where the results from this feasibility study can be applied possibly without any additional costs for camera installments.

1) Evaluated Scenario: We consider the scenario as discussed in Section III-C1 with regard to the deployment of the BSs and STA, channel, initial access procedure, and the coverage area of the BSs. In the scenario, a camera monitors the two pedestrians that walk between BS 1 and the STA. As we have assumed that BS 2 is free of blockages, we do not perform the proactive prediction in the performance of the link between BS 2 and the STA.

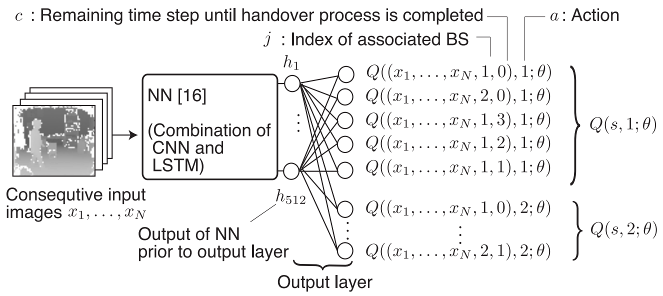

c : Remaining time step until handover process is completed
a: Action
j : Index of associated BS
h_1
NN [16]
(Combination of CNN and LSTM)
$Q((x_1, \ldots, x_N, 1, 0), 1; \theta)$
$Q((x_1, \ldots, x_N, 2, 0), 1; \theta)$
$Q((x_1, \ldots, x_N, 1, 3), 1; \theta)$
$Q((x_1, \ldots, x_N, 1, 2), 1; \theta)$
$Q((x_1, \ldots, x_N, 1, 1), 1; \theta)$
$Q(s, 1; \theta)$
Consequtive input images $x_1, \ldots, x_N$
h_{512}
$Q((x_1, \ldots, x_N, 1, 0), 2; \theta)$
$Q((x_1, \ldots, x_N, 2, 1), 2; \theta)$
$Q(s, 2; \theta)$
Output of NN prior to output layer
Output layer

Fig. 6. NN architecture for approximating optimal action-value function $Q^{\star}(s, a)$ defined in (2) for $\mathcal{C} = \{0, 1, 2, 3\}$ and $\mathcal{J} = \{1, 2\}$. With the exception of the output layer, the architecture herein is identical to that used in [37]. The architecture is a combination of a CNN, which deals with images, and an LSTM, which deals with sequential inputs [45].

This experiment is performed by fixing measurement device and transmitter positions and changing camera angles motivated by the objective of this evaluation. The objective of this experiment is to validate the feasibility of the proactive prediction achieved by introducing cameras in the two basic angles. Hence, it is beyond the scope of this study to perform experiments in various configurations such as in terms of the parameters irrelevant to cameras.

There are some realistic scenarios to which the results from this feasibility study can be applied. In this feasibility study, it is examined that we can at least perform the proactive prediction with camera images if an STA and BSs are static, and the order of the distances from the STA and BSs are several meters. Hence, we can expect that the results are also applied to, for example, a video streaming to static wireless monitors in an office, where the STA and BSs are also static and the order of the distances from an STA to BSs are several meters.

Moreover, in such realistic scenarios, additional costs are not necessarily incurred when we can utilize pre-installed cameras. In concrete, given the aforementioned scenarios transmitting streaming data, we can utilize pre-installed surveillance cameras monitoring the entire office. In this case, the results from this feasibility study can be applied possibly without any additional costs for camera installments.

Concerning a real implementation, we evaluate the image-based handover framework in the two totally different camera angles shown in Fig. 2 (left). The camera angles affect how the obtained images represent the movement of pedestrians, and thus, they may also affect the feasibility of the proactive prediction more strongly compared to other parameters irrelevant of cameras such as the distance of the STA and BS or their heights. Hence, concerning the camera angles may be important for a real implementation, and we perform the evaluation in the two basic camera angles that are orthogonal to each other.

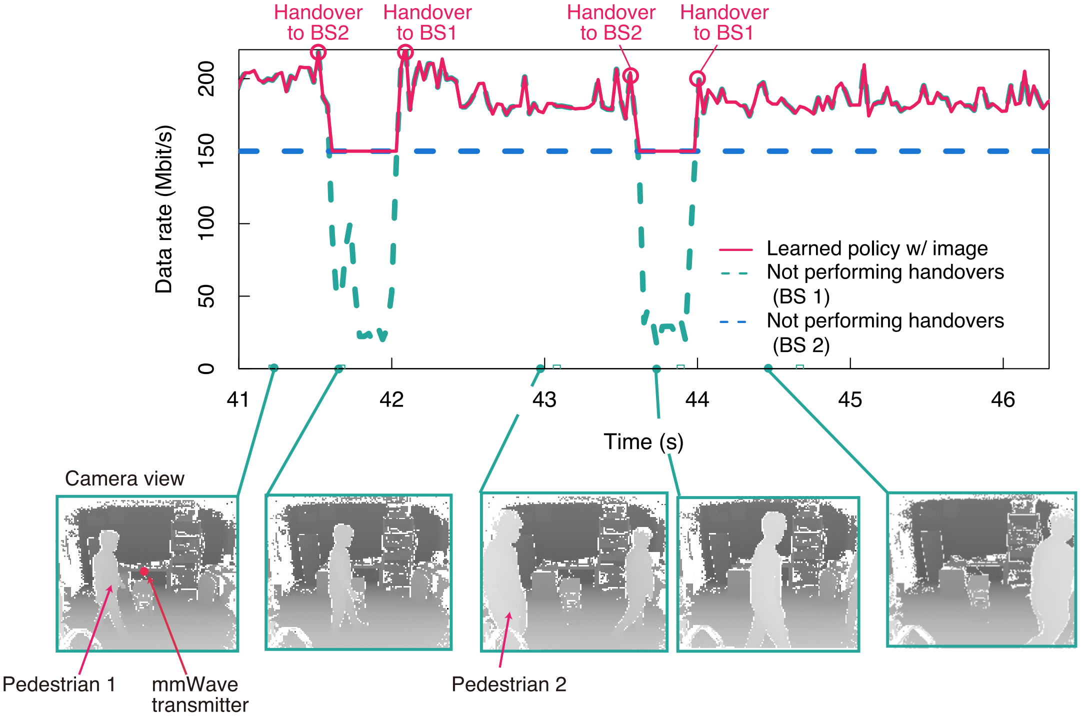


Fig. 7. Time series of the achievable data rates under the condition of the service disruption time of $T_{\text{dis}} = 0$ s and the corresponding camera images.

2) Procedure of Performing Decision Process: We perform the decision process in the image-based handover framework using a procedure similar to that used in the previous evaluation with the exception that the state includes consecutive camera images obtained in the experiment. Let x_t denote the camera image (that contains 40 × 40 pixel values in the experiment and obtained with the frame rate of 30 frame per second) obtained at the same time when the received power value $p_{1,t}$ is obtained. From the state definition in (10), we replace the received power values $p_t, p_{t-1}, \dots, p_{t-N+1}$ in the state s_t in the previous evaluation with the time consecutive images $x_t, x_{t-1}, \dots, x_{t-N+1}$. We learn the optimal policy using deep RL with an NN that is specifically used for handling the time consecutive camera images as discussed in the following section. The parameter associated with the deep RL is set as shown in Table IV.

3) Neural Network Architecture: In the deep RL, an NN is trained such that the NN is a good approximation of the optimal action-value function $Q^*(s,a)$ in (2) [27]. We focus on the NN architecture designed to perform deep RL in the decision process discussed in the previous subsection.[3]

We design the NN architecture such that the NN has separate outputs for each possible combination of $j \in \mathcal{J}$, $c \in \mathcal{C}$, and $a \in \mathcal{A}$, as shown in Fig. 6. The design allows us to divide the parameters into two parts: the parameters associated with the camera images and those associated with the other low-dimensional observations j, c, and a. Let $Q(s,a;\theta)$ be the NN, where $s \in \hat{\mathcal{S}}_{\text{img}}$, $(x_1, \dots, x_N) \in \mathcal{X}^N$, and θ denote the parameters of the NN. In the architecture, the NN is expressed as follows:

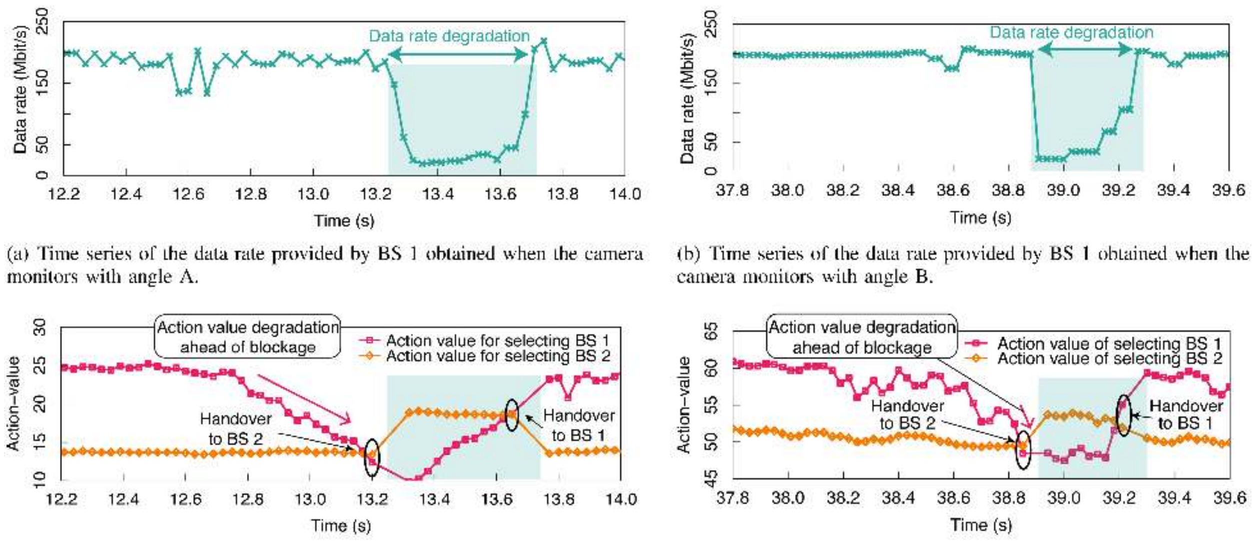
$$Q((x_1, \dots, x_N, j, c), a; \theta) = \sum_{k=1}^{512} \theta_{j,c,a,k} h_k, \tag{10}$$

where $h_1, \dots, h_{512}$ denote the output values of the layer prior to the output layer and $\theta_{j,c,a,1}, \dots, \theta_{j,c,a,512}$ denote the parameters in the output layer corresponding to the combination of j, c, and a. The parameters used to obtain the output values $h_1, \dots, h_{512}$ are associated with the camera images, and the parameters in the output layer, $\theta_{j,c,a,1}, \dots, \theta_{j,c,a,512}$ are associated with the low-dimensional observations j, c, and a.

The motivation for the architecture is that it is necessary to use the observations j and c for handover control. In our problem setting, the state s consists of N consecutive images $(x_1, \dots, x_N)$ with thousands of elements and (j, c) with only two elements. If we let the input of the NN be $(x_1, \dots, x_N, j, c)$, and thereby, process the camera images $(x_1, \dots, x_N)$ and (j, c) using the same parameters, the variation in (j, c) does not significantly impact the NN output values. This is because NNs generally estimate the feature representations of overall inputs; thus, they do not propagate the variation in one or two elements in the inputs to the output [45]. Hence, the controller can ignore the variation in (j, c) while making a handover decision.

It should be noted that we employ the NN architecture used in [37] with the exception of the output layer. The architecture is reported to facilitate the prediction of a future data rate in an mmWave link based on camera images. Hence, it is expected that the architecture also facilitates the learning of the optimal action-value function, which is the cumulative sum of the performance data rates in our problem setting.

[3] The NN is trained using the method discussed in [27]. For details of the training, [27] may be referred to.



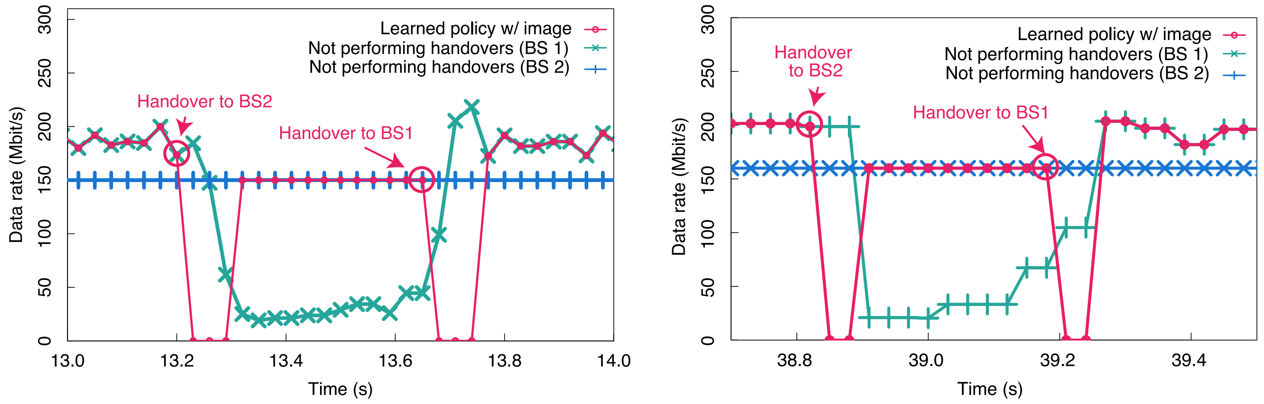
(a) Time series of the data rate provided by BS 1 obtained when the camera monitors with angle A.

(b) Time series of the data rate provided by BS 1 obtained when the camera monitors with angle B.

(c) Time series of action value in proposed framework when the camera monitors with angle A.

(d) Time series of action value in proposed framework when the camera monitors with angle B.

Fig. 8. Comparison between time series of the data rate provided by BS 1 and that of the learned action-value function. The action value in the proposed image-based framework decreases several milliseconds before the performance degradation at BS 1, which indicates that the proposed framework successfully predicts the future performance degradation in advance (the action value is defined as the expected cumulative discounted sum of the future performance).



(a) Portion of the time series of the data rate and hanover timing learned when the camera monitors with angle A.

(b) Portion of the time series of the data rate and hanover timing learned when the camera monitors with angle B.

Fig. 9. Example of handover timing. The proposed image-based framework performed handovers before the variation in the channel at BS 1 and STA with either camera angle A or angle B.

4) Results: We confirm that the deep RL successfully maximizes the time-average of the achievable data rate in the mmWave links in the state design in (10). Fig. 7 shows an example of a time series of the data rate in the case wherein $T_{\text{dis}} = 0$ s. The pedestrians walk in front of the mmWave transmitter at approximately 41.5 s and 43.9 s. At the same time, the data rate provided by BS 1 is degraded from approximately 200 Mbit/s to 30 Mbit/s. Our framework successfully selects the BS that provides a higher data rate than the counterpart at each decision epoch and thereby maximizes the overall data rate.

It should be noted that in Fig. 7, the intervals between the two successive handovers, i.e., the handover from BS 1 to BS 2 and that from BS 2 to BS 1 are according to the durations wherein the blockage effects sustained. This results can be interpreted that our image-based handover framework can form a handover strategy while predicting such durations wherein the blockage effects sustain in an end-to-end manner. In this regard, we've achieved the prediction of such durations implicitly in learning the handover policy.

We show that the proposed framework predicts a future data rate degradation from several hundreds milliseconds before the degradation occurs by analyzing the learned action-value function shown in Fig. 8. Fig. 8 shows the learned action value at each decision epoch before and after the blockage effect in Fig. 5. We can see that the action value begins to decrease from approximately 500 ms (in camera angle A) or 200 ms (in camera angle B) before the actual degradation in the data rate provided by BS 1. As the action value is defined as the expected sum of the future data rates, we can consider that our image-based framework successfully predicts the future performance degradation several hundred milliseconds before the blockage effects occur.

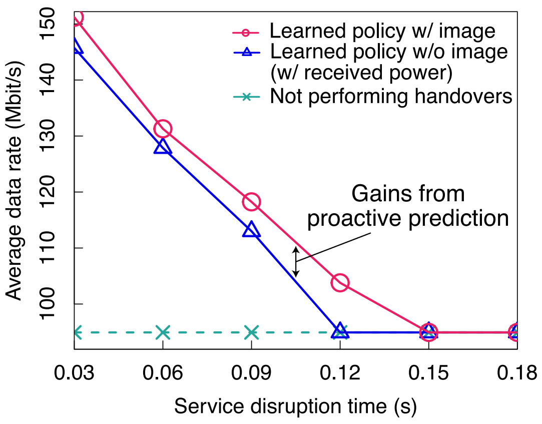


Fig. 10. Performance comparison between proposed image-based framework and received power-based framework under various service disruption times T_{dis} when the camera in angle A.

Owing to the proactive prediction, our image-based handover framework triggers a handover in a proactive manner. Fig. 9 (a) shows an example of a time-varying data rate provided by our image-based handover framework with camera angle A. The plotted duration corresponds to that in Fig. 5. Our proposed framework is different from the received-power-based prediction presented in the previous section in Fig. 5 and successfully triggers handovers prior to the variation in the data rate provided by BS 1. Fig. 9(b) shows an example of a time-varying data rate provided by the proposed image-based handover framework with the camera angle B. Similarly, with a different angle, the image-based framework triggers handovers prior to the degradation of the data rate provided by BS 1.

It should be noted that results show the feasibility of the proactive prediction even with the image time-series with 40 × 40 pixels and a frame rate of 30 frames per second. Hence, to accomplish the proactive prediction, it is sufficient to leverage such qualities of image videos, which cannot be obtained only with sophisticated cameras exampled by Kinect sensors but also with commercial products of smart phones [46] or surveillance cameras [47].

We compare the proposed image-based handover framework with a handover framework that does not leverage images, i.e., the received power-based handover framework. Fig. 10 shows the average data rate achieved by the two handover policies over the duration of specific events 200 ms before and after a blockage.[4] As blockage event, the one depicted in Figs. 5 and 9(a) has been chosen since handover policies with and without camera images exhibit different behavior according to the aforementioned figures. The choice of 200 ms is attributed to the fact that the two handover policies exhibited different a behavior from at most 200 ms before and after the blockage event. From Fig. 10, we can see that the handover policy learned with images achieves a higher or equal data rate as compared to the policy learned without images.

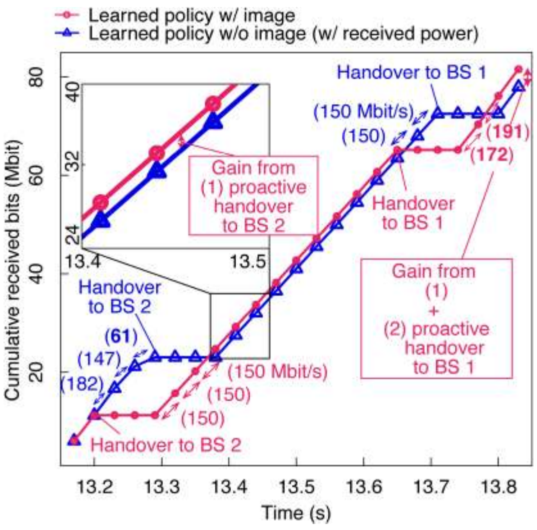


Fig. 11. Comparison of cumulative received bits between our image-based framework and the received-power-based framework.

A realistic scenario where we can benefit from the gain is exemplified by combining Agile-Link [41] as the beam search method and make-before-break [49] as the handover procedure. In such a scenario, the service interruption subjected by a beam alignment is under 1 ms with a 128 size array, and by the other handover procedure would be tens of milliseconds. This leads to an overall service interruption time T_{dis} of several tens milliseconds. Recalling that there is a gain from the proactive handover when T_{dis} is in the order of tens milliseconds in Fig. 10, the system benefits from the gain in such a scenario.

To illustrate how the proactive handover led to the performance gain provided in Fig. 10, we show the cumulative received bits in the proposed image-based handover framework and in the received power-based framework. Fig. 11 shows the amount of cumulative received bits from the time 200 ms before the blockage event. The horizontal axis corresponds to that either in Figs. 5 and 9(a). After a handover to BS 2 is performed in the image-based handover framework, the amount of cumulative data bits is temporarily lower than that in the received power-based handover framework. Meanwhile, the amount of cumulative received bits in the image-based handover framework is larger than that in the power-based framework by 1.7 Mbit from instant 13.4 s. These results confirm the benefits of proactive handover in the long run, and the increase of the received bits can be interpreted as the gain from proactively performing a handover to BS 2. Similarly, the amount of cumulative bits in the image-based handover framework is larger than that in received power-based framework by 3.6 Mbit from instant 13.8 s. This can be attributed to the fact the image-based framework benefits earlier from a recovering data rate in BS 1 while the STA remains to be associated with BS 2 in the received power-based framework. The increase in received bits can also be interpreted as a gain from proactively performing handover to BS 1.

We analyze the computation time required for making a handover decision in the context of an example. Fig. 12 shows

[4] In detail, the blockage event is defined as the event where the received power is 3 dB below from the one observed in a LOS condition, which is according to [48].

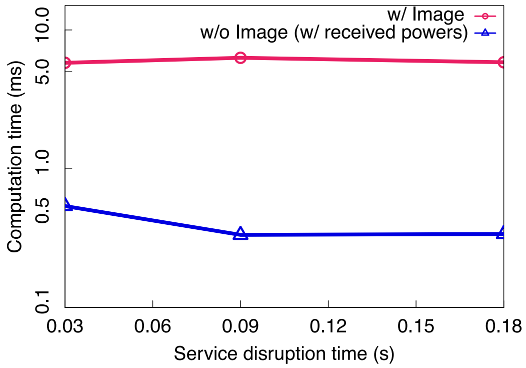


Fig. 12. Computation time for making handover decisions.

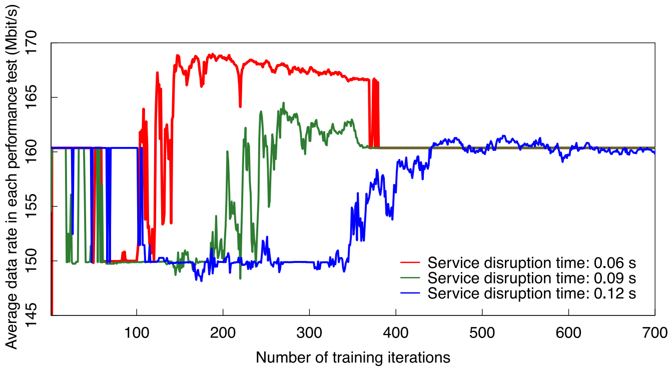


Fig. 13. Average data rate in each performance test corresponding to a training iteration.

an example of the computation time for making a handover decision. The computation time is defined as the time for calculating the action-value from an input of images and is measured with a GeForce GTX 1080 Ti GPU. The received power-based handover achieved the smaller computation time because of the lower dimensionality of the input. Meanwhile, in the proposed image-based handover framework, the computation time was still in the order of several milliseconds. The computation time is sufficiently shorter than the required handover interval, i.e., an interval between the two successive handovers, and is reported as 750 ms in mmWave 5G systems [50]. Thus, a shorter computation time relative to the required handover interval is possible. To this regard, the computation time incurred by the large-dimensionality of images can be overcome.

We investigate the convergence property of the training procedure in Fig. 13. Fig. 13 shows the learning curve, i.e., the average data rate in the performance test corresponding to each training step. We obtained the following trends as the training steps are iterated: the performance enhancement, achievement of the maximum performance, and convergence to the degraded performance These results shows that the training procedure does not converge to maximum performance. This is attributed to the fact that the training process is not stable, which is commonly reported in deep RL [27], [39] when a non-linear approximator for the action-value function is used. These results motivate us to design an improved algorithm that converges to a maximum performance; however, seeking for the better convergence property is beyond the scope of this study.

Nevertheless, a practical solution can be employed to benefit from the results in this study. The solution named, "off-policy evaluation framework" [51], keeps track of the best performing policy. As we are evaluating the policy that achieves maximum average data rates among the learned policies, we can benefit from the results in this study by designing the algorithm such that the off-policy evaluation framework is performed.

V. Conclusion

We presented a new paradigm for leveraging time-consecutive camera images in handover decision problems for realizing the proactive prediction of future long-term performances. We first experimentally noted that the obstacle-caused data rate degradation in mmWave links cannot necessarily be predicted in a proactive manner based only on the time-variation of the received powers before the degradation. To solve this problem, we proposed the expansion of the state space in order for the state information to comprise consecutive camera images, which comprise informative features for proactively predicting long-term data rates in mmWave links. To overcome the difficulty of the higher dimensionality of the expanded state space, we use deep RL for predicting the cumulative sum of the future data rates and deciding handover timings based on the predicted values. By performing deep RL using the state information of experimentally obtained camera images, we confirmed that the state expansion allows the prediction of future obstacle-caused data rate degradation from approximately 500 ms before the degradation occurs. We also evaluated the time-average of the data rates over approximately two minutes and revealed that the proposed expansion of the state space resulted in a performance gain.

References

[1] K. Sakaguchi *et al.*, "Millimeter-wave wireless LAN and its extension toward 5G heterogeneous networks," *IEICE Trans. Commun.*, vol. E98-B, no. 10, pp. 1932–1947, Oct. 2015.

[2] Y. Niu, Y. Li, D. Fin, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.

[3] C. Dehos, J. L. Gonzàlez, A. D. Domenica, D. Kténas, and L. Dussopt, "Millimeter wave access and backhauling: The solution to the exponential data traffic increase in 5G mobile communication systems?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 88–95, Sep. 2014.

[4] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 168–178, Jan. 2015.

[5] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[6] "IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU, Geneva, Switzerland, ITU-Recommendation M2083, pp. 1–19, Sep. 2015.

[7] K. Haneda, "Channel models and beamforming at millimeter-wave frequency bands," *IEICE Trans. Commun.*, vol. E98-B, no. 5, pp. 755–772, May 2015.

[8] G. R. MacCartney and T. S. Rappaport, "A flexible millimeter-wave channel sounder with absolute timing," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1402–1418, Jun. 2017.

[9] X. Zhang, S. Zhou, X. Wang, D. Zhu, and M. Lei, "Improving network throughput in 60 GHz WLANs via multi-AP diversity," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 4803–4807.

[10] Y. Oguma, R. Arai, T. Nishio, K. Yamamoto, and M. Morikura, "Implementation and evaluation of reactive base station selection for human blockage in mmWave communications," in *Proc. APCC*, Kyoto, Japan, Oct. 2015, pp. 1–6.

[11] M. Umehira, G. Saito, S. Takeda, T. Miyajima, and K. Kagoshima, "Feasibility of RSSI based access network detection for multi-band WLAN using 2.4/5 GHz and 60 GHz," in *Proc. WPMC*, Sydney, NSW, Australia, Sep. 2014, pp. 1–6.

[12] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.

[13] Y. Sun, G. Feng, S. Qin, Y. C. Liang, and T.-S. P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 6, pp. 1456–1468, Jun. 2018.

[14] E. S. Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1243–1254, Mar. 2008.

[15] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An MDP model for optimal handover decisions in mmWave cellular networks," in *Proc. EUCNC*, Athens, Greece, Jun. 2016, pp. 100–105.

[16] C. Dhahri and T. Ohtsuki, "Adaptive *Q*-learning cell selection method for open-access femtocell networks: Multi-user case," *IEICE Trans. Commun.*, vol. 97, no. 8, pp. 1679–1688, Aug. 2014.

[17] C. Dhahri and T. Ohtsuki, "*Q*-learning cell selection for femtocell networks: Single and multi-user case," in *Proc. IEEE GLOBECOM*, Anaheim, CA, USA, Dec. 2012, pp. 4975–4980.

[18] H. Tabrizi, G. Farhadi, and J. M. Cioffi, "Dynamic handoff decision in heterogeneous wireless systems: *Q*-learning approach," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 3217–3222.

[19] B.-J. Chang and J.-F. Chen, "Cross-layer-based adaptive vertical handoff with predictive RSS in heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3679–3692, Nov. 2008.

[20] Q. Liu and C. W. Chen, "Smart downlink scheduling for multimedia streaming over LTE networks with hard handoff," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1815–1829, Nov. 2015.

[21] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter wave human blockage at 73 GHz with a simple double knife edge diffraction model and extension for directional antennas," in *Proc. IEEE VTC-Fall*, Montreal, QC, Canada, Sep. 2016, pp. 1–5.

[22] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Measurement method of temporal attenuation by human body in off-the-shelf 60 GHz WLAN with HMM-based transmission state estimation," *Hindawi Wireless Commun. Mobile Commun.*, vol. 2018, Apr. 2018, Art. no. 7846936.

[23] M. Jacob, C. Mnianke, and T. Kürner, "A dynamic 60 GHz radio channel model for system level simulations with MAC protocols for IEEE 802.11ad," in *Proc. IEEE ISCE*, Brunswick, Germany, Jun. 2010, pp. 1–5.

[24] O. Kaltiokallio, H. Yiğitler, and R. Jäntti, "A three-state received signal strength model for device-free localization," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9226–9240, Oct. 2017.

[25] C.-R. Lin, Y.-J. Chen, and L.-C. Wang, "Handoff delay analysis in SDN-enabled mobile networks: A network calculus approach," in *Proc. IEEE VTC-Fall*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.

[26] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. NIPS*, Montreal, QC, Canada, Dec. 2014, pp. 1–9.

[27] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 529, pp. 529–533, Feb. 2015.

[28] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Cooperative sensing in deep RL-based image-to-decision proactive handover for mmWave networks," in *Proc. IEEE CCNC*, Las Vegas, NV, USA, Jun. 2019, pp. 1–6.

[29] N. W. Sung, N.-T. Pham, T. Huynh, and W.-J. Hwang, "Predictive association control for frequent handover avoidance in femtocell networks," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 924–927, May 2013.

[30] S. Zang *et al.*, "Mobility handover optimization in millimeter wave heterogeneous networks," in *Proc. IEEE ISCIT*, Cairns, QLD, Australia, Sep. 2017, pp. 1–6.

[31] S. Zang, W. Bao, P. L. Yeoh, B. Vucetic, and Y. Li, "Managing vertical handovers in millimeter wave heterogeneous networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1629–1644, Feb. 2019.

[32] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Reinforcement learning based predictive handover for pedestrian-aware mmWave networks," in *Proc. IEEE INFOCOM Workshops*, Honolulu, HI, USA, Apr. 2018, pp. 1–6.

[33] Y. Oguma, R. Arai, T. Nishio, K. Yamamoto, and M. Morikura, "Proactive base station selection based on human blockage prediction using RGB-D cameras for mmWave communications," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[34] Y. Oguma, T. Nishio, K. Yamamoto, and M. Morikura, "Proactive handover based on human blockage prediction using RGB-D cameras for mmWave communications," *IEICE Trans. Commun.*, vol. E99-B, no. 8, pp. 1734–1744, Oct. 2016.

[35] T. Nishio, R. Arai, K. Yamamoto, and M. Morikura, "Proactive traffic control based on human blockage prediction using RGB-D cameras for millimeter-wave communications," in *Proc. IEEE CCNC*, Las Vegas, NV, USA, Jan. 2015, pp. 152–153.

[36] H. Okamoto, T. Nishio, M. Morikura, and K. Yamamoto, "Machine-learning-based throughput estimation using images for mmWave communications," in *Proc. IEEE VTC-Spring*, Sydney, NSW, Australia, Jun. 2017, pp. 1–6.

[37] T. Nishio *et al.*, "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, Nov. 2019.

[38] X. Tan, X. Luan, Y. Cheng, A. Liu, and J. Wu, "Cell selection in two-tier femtocell networks using *Q*-learning algorithm," in *Proc. ICACT*, Pyeongchang, South Korea, Feb. 2014, pp. 1036–1040.

[39] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[40] W. Jiao, P. Jiang, and Y. Ma, "Fast handover scheme for real-time applications in mobile WiMAX," in *Proc. IEEE ICC*, Glasgow, U.K., Jun. 2017, pp. 6038–6042.

[41] H. Hassanieh, O. Abari, M. Rodriguez, M. A. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Proc. ACM SIGCOMM*, Budapest, Hungary, Aug. 2018, pp. 432–445.

[42] B. Peng, S. Rey, D. M. Rose, S. Hahn, and T. Kuerner, "Statistical characteristics study of human blockage effect in future indoor millimeter and sub-millimeter wave wireless communications," in *Proc. IEEE VTC-Spring*, Porto, Portugal, Jun. 2018, pp. 1–5.

[43] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Standard 802.11ad-2012, 2012.

[44] T. Nische, G. Bielsa, A. Loch, and J. Widmer, "Boon and bane of 60 GHz networks: Practical insights into beamforming, interference, and frame level operation," in *Proc. ACM CoNEXT*, Heidelberg, Germany, Dec. 2015, pp. 1–6.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[46] P. Ji, H.-M. Tsai, C. Wang, and F. Liu, "Vehicular visible light communications with LED taillight and rolling shutter camera," in *Proc. IEEE VTC-Spring*, Seoul, South Korea, May 2014, pp. 1–6.

[47] R. Hill, C. S. Madden, A. van den Hengel, H. Detmold, and A. R. Dick, "Measuring latency for video surveillance systems," in *Proc. IEEE DICTA*, Melbourne, VIC, Australia, Dec. 2009, pp. 89–95.

[48] T. Choi *et al.*, "Measurement based directional modeling of dynamic human body shadowing at 28 GHz," in *Proc. IEEE GLOBECOM*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[49] L. C. Gimenez, P.-H. Michaelsen, K. I. Pedersen, T. E. Kolding, and H. C. Nguyen, "Towards zero data interruption time with enhanced synchronous handover," in *Proc. IEEE VTC-Spring*, Sydney, NSW, Australia, May 2017, pp. 1–6.

[50] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff rates for millimeter-wave 5G systems," in *Proc. IEEE VTC-Spring*, Seoul, South Korea, May 2014, pp. 1–5.

[51] G. Theocharous, P. S. Thomas, and M. Ghavamzadeh, "Personalized ad recommendation systems for life-time value optimization with guarantees," in *Proc. IJCAI*, Buenos Aires, Republica Argentina, Jul. 2015, pp. 1806–1812.

Yusuke Koda (Student Member, IEEE) received the B.E. degree in electrical and electronic engineering from Kyoto University in 2016, and the M.E. degree from the Graduate School of Informatics, Kyoto University in 2018, where he is currently pursuing the Ph.D. degree. In 2019, he visited the Centre for Wireless Communications, University of Oulu, Finland, to conduct collaborative research. He received the VTS Japan Young Researcher's Encouragement Award in 2017. He was a Recipient of the Nokia Foundation Centennial Scholarship in 2019. He is a member of ACM and IEICE.

Kota Nakashima (Student Member, IEEE) received the B.E. degree in electrical and electronic engineering from Kyoto University in 2018, where he is currently pursuing the M.E. degree with the Graduate School of Informatics. He received the VTS Japan Young Researcher's Encouragement Award in 2018.

Koji Yamamoto (Member, IEEE) received the B.E. degree in electrical and electronic engineering and the master's and Ph.D. degrees in informatics from Kyoto University in 2002, 2004, and 2005, respectively. From 2004 to 2005, he was a Research Fellow of the Japan Society for the Promotion of Science. Since 2005, he has been with the Graduate School of Informatics, Kyoto University, where he is currently an Associate Professor. From 2008 to 2009, he was a Visiting Researcher with Wireless@KTH, Royal Institute of Technology, Sweden. He was a Tutorial Lecturer in ICC 2019. His research interests include radio resource management, game theory, and machine learning. He received the PIMRC 2004 Best Student Paper Award in 2004, the Ericsson Young Scientist Award in 2006, the Young Researcher's Award, the Paper Award, SUEMATSU-Yasuharu Award from the IEICE of Japan in 2008, 2011, and 2016, respectively, and IEEE Kansai Section GOLD Award in 2012. He serves as an Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and the *Journal of Communications and Information Networks*, the Track Co-Chair of APCC 2017, CCNC 2018, APCC 2018, and CCNC 2019, and the Vice Co-Chair of IEEE ComSoc APB CCC. He is a Senior Member of IEICE and the Operations Research Society of Japan.

Takayuki Nishio (Member, IEEE) received the B.E. degree in electrical and electronic engineering from Kyoto University in 2010, and the M.I. and Ph.D. degrees in communications and computer engineering from the Graduate School of Informatics, Kyoto University, Kyoto, Japan, in 2012 and 2013, respectively. From 2012 to 2013, he was a Research Fellow (DC1) of the Japan Society for the Promotion of Science. Since 2013, he has been an Assistant Professor in communications and computer engineering with the Graduate School of Informatics, Kyoto University. From 2016 to 2017, he was a Visiting Researcher with Wireless Information Network Laboratory, Rutgers University, USA. His current research interests include mmWave networks, wireless local area networks, application of machine learning, and sensor fusion in wireless communications. He received the IEEE Kansai Section Student Award in 2011, the Young Researcher's Award from the IEICE of Japan in 2016, and the Funai Information Technology Award for Young Researchers in 2016. He is a member of ACM and IEICE.

Masahiro Morikura (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electronics engineering from Kyoto University, Kyoto, Japan, in 1979, 1981, and 1991, respectively. He joined NTT in 1981, where he was engaged in the research and development of TDMA equipment for satellite communications. From 1988 to 1989, he was a Guest Scientist with the Communications Research Centre, Canada. From 1997 to 2002, he was active in the standardization of the IEEE 802.11a-based wireless LAN. He is currently a Professor with the Graduate School of Informatics, Kyoto University. His current research interests include WLANs and M2M wireless systems. He received the Paper Award, the Achievement Award from IEICE in 2000 and 2006, respectively, the Education, Culture, Sports, Science, and Technology Minister Award in 2007, the Maejima Award in 2008, and the Medal of Honor with Purple Ribbon from Japan's Cabinet Office in 2015.