Mesh-Clustering-Based Radio Maps Construction for Autonomous Distributed Networks

Keita Katagiri† and Takeo Fujii†
† Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications
1–5–1 Chofugaoka, Chofu, Tokyo 182–8585, Japan
Email: {katagiri, fujii}@awcc.uec.ac.jp

***Abstract*—We have proposed a method of the radio map construction using clustering algorithm in our conventional work. The method enables us to accurately predict the radio environment while reducing the registered data size. However, this clustering algorithm has been only applied to the wireless system with fixed transmitter location. Thus, this paper considers the radio maps construction based on the clustering for the autonomous distributed networks that both transmitter and receiver dynamically move. The proposed method classifies the similar average received signal power samples using k-means++. The emulation results clarify that the proposed method can estimate the radio environment with high accuracy while reducing the registered data size compared to the conventional radio map.**

***Index Terms*—Radio map, clustering, radio propagation, autonomous distributed networks**

I. Introduction

Nowadays, a radio map attracts attention as an enabler for accurately estimating the radio environment characteristics [1]–[4]. The radio map stores statistical information of the radio environment, such as an available spectrum and an average received signal power, in each location. Distributed mobile terminals observe the radio environment and report the measured datasets to a cloud server. Subsequently, the cloud server creates the radio map by statistically processing these datasets in each mesh, which is calculated based on the latitude and longitude. This construction method is well known as the *crowdsourcing* [5], [6]. The method allows us to construct the radio map with low observation costs and short times.

Here, it is necessary to use the smaller mesh size than a correlation distance of the shadowing component to generate the precise radio map. For instance, this correlation distance is empirically around 20 [m] in the urban area [7]. Thus, the 10m-scale mesh is often utilized in the radio map construction [8]–[10]. However, the registered data size may be huge according to a communication range since the cloud server manages the radio environment information in each mesh. The trade-off between the mesh size and estimation accuracy of the radio environment is an important topic for the radio map construction.

We have proposed the clustering-based radio map construction methods [11], [12] to solve the trade-off by considering the spatial correlation of the shadowing [7]. In these methods, similar received signal power samples are unified into a representative value based on clustering algorithms, such as the k-means++ [13]. Our results have clarified that the clustering algorithms enable us to accurately predict the average received signal power in each mesh while reducing the registered data size of the cloud server. However, these works have only considered the wireless system with fixed transmitter location.

Meanwhile, autonomous distributed networks, such as vehicle-to-vehicle (V2V) communications [14], [15] and device-to-device communication [16], are actively discussed. Since both transmitter and receiver dynamically move in these systems, the communication reliability may be poorer than the wireless system with fixed transmitter location. If we create radio maps even in the autonomous distributed networks, it is necessary to accumulate radio environment information in each transmission / reception position [17]. As a result, the registered data size is very enormous.

Motivated by this problem, we apply a clustering algorithm to the construction of the radio maps for autonomous distributed networks. The proposed method classifies the similar average received signal power samples in each transmission / reception position using the k-means++. Subsequently, a representative value of the average received signal power is calculated in each cluster. The emulation results clarify that the proposed method can estimate the radio environment with high accuracy while reducing the registered data size compared to the conventional radio map.

The remainder of this paper is organized as follows. Sect. II describes the overview of the radio maps for the autonomous distributed networks. Then, the proposed method is explained in Sect. III. After Sect. IV shows the measurement datasets, the emulation results are described in Sect. V. Finally, we conclude this paper in Sect. VI.

II. Radio Maps for autonomous distributed networks

A. System Model

Fig. 1 presents the overview of the conventional radio maps for the autonomous distributed networks. We consider that S transmitters and D receivers exist and dynamically move in the communication area. Each transmitter sends the signal, including own transmitter ID and transmission position to D receivers. After each receiver records the received signal power, reception position, center frequency, and receiver ID in addition to the transmitted information, these datasets are uploaded to the cloud server. The cloud server divides the communication area into two-dimensional meshes and

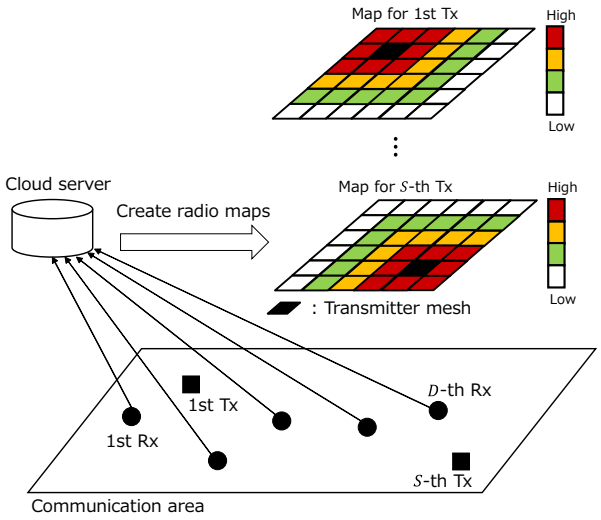


Fig. 1. Overview of the radio maps for autonomous distributed networks.

Tx mesh code	Rx mesh code	Average received signal power [dBm]
xxxx-xx-xx-xx	aaaa-aa-aa-aa	-90.0123456789
yyyy-yy-yy-yy	bbbb-bb-bb-bb	-84.2345677890
zzzz-zz-zz-zz	cccc-cc-cc-cc	-78.3456789012
⋮	⋮	⋮

(a) Conventional radio map.

Mesh table

Tx mesh code	Rx mesh code	Cluster label
xxxx-xx-xx-xx	aaaa-aa-aa-aa	0
yyyy-yy-yy-yy	bbbb-bb-bb-bb	1
zzzz-zz-zz-zz	cccc-cc-cc-cc	2
⋮	⋮	⋮

Clustering table

Cluster label	Representative value [dBm]
0	-90.0
1	-84.0
2	-78.0
⋮	⋮

(b) Proposed method.

Fig. 2. Registration contents for radio map.

statistically processes the reported datasets in each m-th transmission / reception mesh pair ($m = 1, \cdots, M$), here, M is the number of transmission / reception mesh pairs. As the statistical data, the cloud server constructs the radio map in each transmission mesh; that is, S maps are stored. Here, we assume that the radio map expresses the average received signal power in each reception mesh.

A transmitter accesses to the cloud server for obtaining the radio map that corresponds to own transmission position. The radio map enables the transmitter to accurately predict the path loss and shadowing in each location, and appropriately design the communication parameters, such as the modulation format.

B. Problem and Solution

The registration contents for the radio map are illustrated in Fig. 2. As shown in Fig. 2(a), the conventional radio map stores the average received signal power in each transmission / reception mesh. Here, each mesh is identified using the mesh code; that is, the cloud server assigns the unique string format in each mesh based on the latitude and longitude information [12]. However, the registered data size may be huge since the number of meshes increases according to the range of the communication area and mesh size.

In order to solve this problem, the proposed method unifies the similar average values to the representative value using the clustering algorithm. Subsequently, the two tables; the mesh table and the clustering table, are linked using the clustering label to manage the transmission / reception mesh codes and the representative value. The proposed method can reduce the registered data size if there are many similar average values.

III. Proposed Method

A. Selection of Clustering Algorithm

It is necessary to choose the something clustering algorithm for unifying the similar average values. Here, it is well known that the spatial correlation of the shadowing component exponentially decays against the sum of the movement distance for both transmitter and receiver even in the autonomous distributed networks [18]. In other words, the average received signal power may be similar in vicinity meshes. Motivated by this viewpoint, we use the k-means++ [19] for the clustering algorithm. This method can easily classify the input data with high accuracy if several data are collectively distributed in the vicinity area.

Although the Gaussian mixture model (GMM) is often utilized as the soft clustering, this method assumes that the input data follows the Gaussian distribution. In the real environment, the average received signal power may not follow the log-normal distribution owing to the site-specific fluctuation of the radio propagation. Thus, we do not use GMM.

B. Definition of Input Data

To utilize the k-means++, an input data vector for the s-th transmission mesh ($s = 1, \cdots, S$) and the d-th reception mesh ($d = 1, \cdots, D$) of the m-th mesh pair is defined as follows,

$$\boldsymbol{z}_{s,d,m} = (x_{s,m}, y_{s,m}, x_{d,m}, y_{d,m}, \bar{p}_{s,d,m}), \tag{1}$$

where $(x_{s,m}, y_{s,m})$ and $(x_{d,m}, y_{d,m})$ are the coordinate values of the s-th transmission mesh and the d-th reception mesh of the m-th mesh pair, respectively. $\bar{p}_{s,d,m}$ [dBm] is the average received signal power in the d-th reception mesh based on the s-th transmission mesh for the m-th mesh pair. Here, each coordinate value is calculated by the mesh code [12].

The accuracy of the clustering may be poor since the scale between the coordinate value and received signal power is different. Hence, we standardize each input data and multiply each data by weight before the clustering is performed. As the standardization procedures, the mean and standard deviation are first calculated for each input data. Then, the cloud server subtracts the mean from each value and divides by the standard deviation.

C. k-means++

k-means++ is a non-hierarchical clustering method that is improved the k-means considering the initial cluster placement. This method classifies the input data so that the following evaluation function J is minimized:

$$J = \sum_{k=1}^{K} \sum_{m=1}^{M} u_{mk} d_{\mathrm{e}}(\boldsymbol{z}_{s,d,m}, \boldsymbol{\mu}_k), \tag{2}$$

TABLE I
MEASUREMENT PARAMETERS

Communication standard	ARIB STD-T109 [20]
Center frequency [MHz]	760
Transmission power [dBm]	19.2
Modulation format	QPSK/OFDM
The number of subcarriers	52
Communication header [byte]	61
Payload length [byte]	77
Noise floor [dBm]	-96.0

where K is the number of clusters, u_{mk} is 1 if $\boldsymbol{z}_{s,d,m}$ belongs to the k-th cluster; otherwise, 0. $\boldsymbol{\mu}_k$ is the centroid vector of the k-th cluster and $d_{\mathrm{e}}(\boldsymbol{z}_{s,d,m}, \boldsymbol{\mu}_k)$ is an euclidean distance between $\boldsymbol{z}_{s,d,m}$ and $\boldsymbol{\mu}_k$.

The k-means++ determines each initial centroid as follows:

a). The data $\boldsymbol{z}_{s,d,m}$ is randomly selected as the first centroid $\boldsymbol{\mu}_1$ from the M mesh pairs.

b). $\boldsymbol{z}'_{s,d,m}$ is determined as another centroid $\boldsymbol{\mu}_k$ from the M mesh pairs using the following probability,

$$\frac{\min\limits_{1\le k\le K} d_{\mathrm{e}}(\boldsymbol{z}'_{s,d,m}, \boldsymbol{\mu}_k)}{\sum_{m=1}^{M} \min\limits_{1\le j\le K} d_{\mathrm{e}}(\boldsymbol{z}_{s,d,m}, \boldsymbol{\mu}_j)}. \tag{3}$$

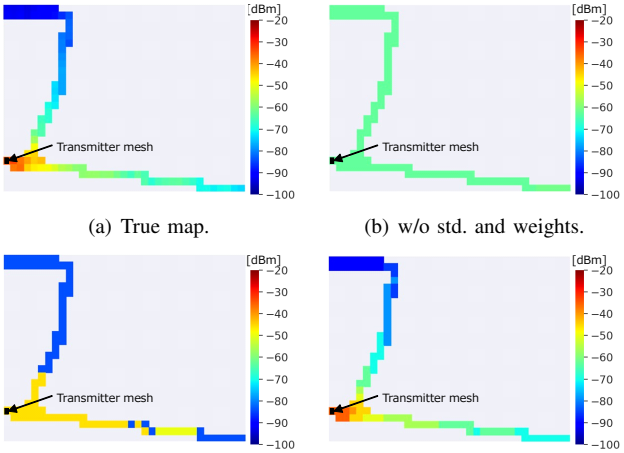
c). b). is repeated while K centroids are chosen.

The cloud server classifies the average received signal power samples into K clusters based on Eq. (2). Then, the representative value is calculated for each cluster by averaging the average received signal power samples having the same cluster label. The cloud server registers the transmission / reception mesh codes, cluster label k, and the representative value.

IV. MEASUREMENT DATASETS

We evaluate the effectiveness of the proposed method by using the measurement datasets of the V2V communications in the real environment [17]. The measurement was performed in Chofu City and Mitaka City, typical suburban areas in Tokyo, Japan over three days in January 2017. The three vehicles that implemented an on-board unit traveled on the red line shown in Fig. 3. The orange dotted line is the evaluation area to show the example of the radio maps. In the measurement, the received signal power, transmission / reception position, and IDs of both transmitter and receiver were observed in each location. Here, we obtained the position information and observation time by using the Garmin GPS 18x, a USB-connected device. The accuracy of the GPS is 95 [%] within 15 [m]. The communication standard is IEEE 802.11p-based method called *ARIB STD-T109*, which is developed for 700 MHz band intelligent transport systems by the Association Industries and Businesses (ARIB) [20]. Table I is the measurement parameters. After the measurement, we statistically processed the 2839076 datasets using MySQL 5.7 and constructed radio maps for each 10m transmission / reception mesh.

Fig. 3. The measurement route.



(a) True map. (b) w/o std. and weights.

(c) w/ std. and w/o weights. (d) w/ std. and weights.

Fig. 4. Examples of radio maps.

V. EMULATION RESULTS

This section describes the emulation results. The weights for $(x_{s,m}, y_{s,m})$, $(x_{d,m}, y_{d,m})$, and $\bar{p}_{s,d,m}$ are 0.05, 0.05, and 0.9 in k-means++, respectively.

A. Example of Radio Maps

Fig. 4 presents an example of the radio maps in the evaluation area. In these maps, each color mesh is the average received signal power when the transmitter is located in the black mesh corresponds to the red square shown in Fig. 5. Fig. 4(a) expresses the average received signal power in each mesh without the clustering. The constructed radio maps using k-means++ are depicted in Figs. 4(b), 4(c), and 4(d). Here, the notation 'std.' means the standardization for each input data.

Fig. 5. The evaluation area.

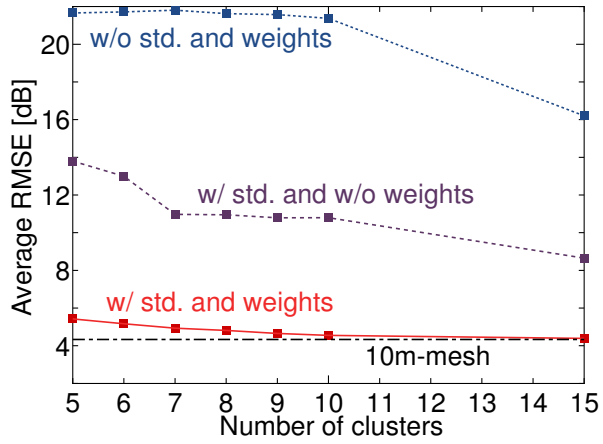


Fig. 6. The average RMSE characteristics.

These maps clarify that the k-means++ w/ std. and weights can accurately classify the similar average values compared with others. Meanwhile, Figs. 4(b) and 4(c) construct the radio maps with poor accuracy since the centroid may not be appropriately calculated.

B. Estimation Accuracy

Next, the root mean squared error (RMSE) is derived using the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{s,d}} \sum_{i=1}^{N_{s,d}} (P_{s,d,i} - \bar{P}_{s,d})^2} \quad [\mathrm{dB}], \tag{4}$$

where $P_{s,d,i}$ [dBm] is the i-th instantaneous received signal power in the s-th transmission mesh and d-th reception mesh, $\bar{P}_{s,d}$ [dBm] and $N_{s,d}$ are the estimated average received signal power and the number of datasets in the s-th transmission mesh and d-th reception mesh, respectively. Note that $\bar{P}_{s,d}$ corresponds to the representative value in the cluster to which the s-th transmission mesh and d-th reception mesh belong. We divided the 2839076 datasets into the three groups and calculated the average RMSE based on the cross-validation.

TABLE II
REGISTRATION CONTENTS IN THE PROPOSED METHOD

(a) Mesh Table

Item	Type	Size [byte]
10m mesh code of a Tx mesh	text	11
10m mesh code of a Rx mesh	text	11
Cluster label k	int	4
Total data size per mesh		26

(b) Clustering Table

Item	Type	Size [byte]
Cluster label k	int	4
Average received signal power	double	8
Total data size per cluster		12

(c) Single Table

Item	Type	Size [byte]
1st code of a Tx mesh	text	5
1st code of a Rx mesh	text	5
Total data size		10

TABLE III
REGISTRATION CONTENTS IN THE CONVENTIONAL RADIO MAP

(a) Mesh Table

Item	Type	Size [byte]
10m mesh code of a Tx mesh	text	11
10m mesh code of a Rx mesh	text	11
Average received signal power	double	8
Total data size per mesh		30

(b) Single Table

Item	Type	Size [byte]
1st code of a Tx mesh	text	5
1st code of a Rx mesh	text	5
Total data size		10

The RMSE characteristics are shown in Fig. 6. Here, the black dashed line means the estimation accuracy of the conventional radio map; that is, the clustering is not performed. It can be found that the RMSE of the k-means++ w/ std. and weights is superior to the others. The others methods cannot accurately predict the average received signal power because of the inaccurate clustering.

C. The Registered Data Size

Finally, this subsection describes the registered data size of the radio map. The registration contents of the proposed method and conventional radio map are shown in Tables II and III, respectively. The proposed method accumulates the cluster label k in each mesh as shown in Table II(a). The 10m mesh code can be expressed as the 16 [byte] text type based on reference [21]. Note that the data size of the 10m mesh code per mesh is 11 [byte] because 1st mesh code is the same value in Fig. 3. Hence, the 1st mesh code of each transmission and reception is registered only 5 [byte] as shown in Table II(c). To link the cluster label k and average received signal power in the k-th cluster, we use Table II(b) as the

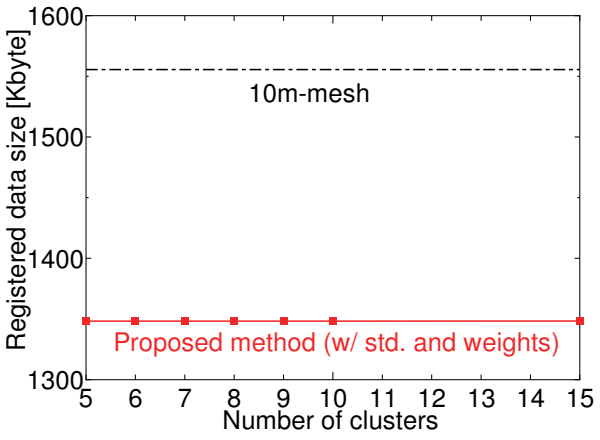


Fig. 7. The registered data size.

clustering table. The proposed method creates K clusters with different average received signal power; thus, Table II(b) is registered for K clusters. Meanwhile, the conventional radio map accumulates the average received signal power with each 10 m transmission / reception mesh as shown in Table III(a). The 1st mesh code of each transmission and reception is registered in Table III(b).

The registered data size of the proposed method R_1 is represented as follows,

$$R_1 = (M \times 26) + (K \times 12) + 10 \quad [\text{byte}]. \tag{5}$$

Additionally, the registered data size of the conventional radio map R_2 is represented as follows,

$$R_2 = M \times 30 + 10 \quad [\text{byte}]. \tag{6}$$

As the calculation procedures, we first divide the 2839076 datasets into three groups. Subsequently, the two groups are utilized as the statistical data to calculate the registered data size. The average M is about 53000.

Fig. 7 shows the registered data size versus the number of clusters. We can confirm that the proposed method can notably reduce the registered data size compared with the conventional radio map. The detailed reduction rate of the registered data size is about 13.33 [%]. This result implies that the similar average received signal power samples can be accurately classified using the k-means++ because of the high spatial correlation for the shadowing component.

VI. Conclusion

We have proposed the mesh-clustering-based radio maps construction method for the autonomous distributed networks. The proposed method classifies the similar average received signal power samples for each transmission / reception mesh using the k-means++. The emulation results have clarified that the proposed method can accurately predict the location-dependent radio environment while notably reducing the registered data size compared to the conventional radio map. Our method will contribute the realization of the practical radio map even in the autonomous distributed networks.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 18H01439, 18KK0109, 19J23352.

References

[1] X. Wang, M. Umehira, B. Han, P. Li, Y. Gu, and C. Wu, "Online incentive mechanism for crowdsourced radio environment map construction," in *Proc. IEEE ICC 2019*, May 2019.

[2] Y. Ye and B. Wang, "Rmapcs: Radio map construction from crowdsourced samples for indoor localization," *IEEE Access*, vol. 6, pp. 24224-24238, April 2018.

[3] B. Huang, Z. Xu, B. Jia, and G. Mao, "An online radio map update scheme for WiFi fingerprint-based localization," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6909–6918, Aug. 2019.

[4] X. Mo, Y. Huang, and J. Xu, "Radio-map-based robust positioning optimization for UAV-Enabled wireless power transfer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 179–183, Feb. 2020.

[5] S. H. Jung and D. Han, "Automated construction and maintenance of Wi-Fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, Dec. 2017.

[6] X. Ying, S. Roy, and R. Poovendran, "Pricing mechanism for quality-based radio mapping via crowdsourcing," in *Proc. 2016 IEEE GLOBECOM*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[7] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.

[8] K. Suto, S. Bannai, K. Sato, K. Inage, K. Adachi, and T. Fujii, "Image-driven spatial interpolation with deep learning for radio map construction," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1222–1226, Mar. 2021.

[9] K. Sato, K. Suto, K. Inage, K. Adachi, and T. Fujii, "Space-frequency-interpolated radio map," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 714–725, Jan. 2021.

[10] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected UAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975–1989, Nov. 2020.

[11] K. Katagiri, K. Sato, K. Inage, and T. Fujii, "Experimental verification of shadowing classification for radio map," in *Proc. IEEE VTC2020-Fall*, Victoria, BC, Canada, Dec. 2020, pp. 1–7.

[12] R. Hasegawa, K. Katagiri, K. Sato, and T. Fujii, "Low storage, but highly accurate measurement-based spectrum database via mesh clustering," *IEICE Trans. Commun.*, vol. E101-B, no. 10, pp. 2152-2161, Oct. 2018.

[13] R. M. Esteves, T. Hacker, and C. Rong, "Cluster analysis for the cloud: Parallel competitive fitness and parallel K-means++ for large dataset analysis," in *Proc. 4th IEEE Int. Conf. Cloud Comput. Technol. Sci. Proc.*, Taipei, Taiwan, Dec. 2012, pp. 177–184.

[14] M. Sepulcre and J. Gozalvez, "Heterogeneous V2V communications in multi-link and multi-RAT vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 20, no. 1, pp. 162–173, Sep. 2019.

[15] C. B. Lehocine, E. G. Strom, and F. Brannstrom, "Hybrid combining of directional antennas for periodic broadcast V2V communication," *IEEE Trans. Intell. Transp. Syst.*, early access, doi: 10.1109/TITS.2020.3033094.

[16] X. Yuan, H. Tian, and B. Fan, "Mobility-aware joint resource allocation and power allocation for D2D communication," in *2019 IEEE WCNC*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.

[17] K. Katagiri, K. Sato, and T. Fujii, "Crowdsourcing-assisted radio environment database for V2V communications," *Sensors*, vol. 18, no. 4, 1183, Apr. 2018.

[18] Z. Wang, E. K. Tameh, and A. R. Nix, "Joint shadowing process in urban peer-to-peer radio channels," *IEEE Trans. Veh. Tech.*, vol. 57, no. 1, pp. 52–64, Jan. 2008.

[19] Y. Uesugi, K. Katagiri, K. Sato, K. Inage, and T. Fujii, "Clustering of signal power distribution toward low storage crowdsourced spectrum database," in *Proc. IEEE VTC2019-Fall*, Sept. 2019.

[20] ARIB STD-T109: 2013. 700MHz band intelligent transport systems version 1.2. Association of Radio Industries and Businesses.

[21] K. Sato, M. Kitamura, K. Inage, and T. Fujii, "Measurement-based spectrum database for flexible spectrum management," *IEICE Trans. Commun.*, vol. E98-B, no. 10, pp. 2004-2013, Oct. 2015.