ARTICLE IN PRESS

The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express xxx (xxxx) xxx

www.elsevier.com/locate/icte

Universal gated recurrent unit-based 3D localization method for ultra-wideband systems

Doan Tan Anh Nguyen[a], Jingon Joung[a,*], Xin Kang[b]

[a] *Department of Electronic and Electrical Engineering, Chung-Ang University, Seoul 06974, Republic of Korea*
[b] *Center for Intelligent Networking and Communications (CINC), University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China*

Received 6 April 2021; accepted 28 June 2021
Available online xxxx

Abstract

In this study, a universal gated recurrent unit (u-GRU)-based location estimation (LE) method is proposed to obtain the location of an ultra-wideband (UWB) transmitter. The proposed u-GRU-LE system consists of a GRU-based classifier and nine localizers for nine channel models (CMs). The classifier first predicts the CM, and then, the proper localizer is selected according to the predicted CM to estimate the location of the transmitter. Rigorous simulations are executed with various CMs. From the results, it is verified that the proposed universal GRU-based 3D localization method generally performs well irrespective of the channel environments.

Keywords: 3D localization; Deep learning; Gated recurrent unit (GRU); Ultra-wideband (UWB) system

1. Introduction

The demand for accurate indoor localization systems is growing as they are an important part of the Internet of Things. There is a wide range of applications for indoor localization systems, such as user tracking in buildings or unmanned vehicle control inside warehouses. In [1,2], the authors proposed a WiFi-based indoor positioning system, where [2] employs a sensor fusion method. The ultra-wideband (UWB) technology is mostly used for indoor localization because of robustness against multipath effect and interference, high definition, and the ability to penetrate various materials. In [3], a convolutional neural network (CNN)-based method is proposed to estimate the distance between a UWB transmitter (Tx) and a receiver (Rx). In [4], a more complex CNN model is proposed to find the two-dimensional location of Tx from the signal captured in three Rx's. Three-dimensional (3D) moving object is tracked by using the measurements in open areas, such as time difference, frequency difference, and angle of arrival [5]. In [6], a gated recurrent unit (GRU)-based model is designed to estimate the 3D location of Tx in a confined space, which is called a GRU-based location estimation (GRU-LE) method.

In [6], the GRU-LE method needs to train a deep learning model for every specific environment. If the environment changes, the model has to be retrained with the new data of the new environment. Thus in this study, the GRU-LE method in [6] is extended to operate in various environments without the retraining. To the best of our knowledge, there has been lack of research on the environment detection/classification based on UWB signals. To this end, the channel environment is classified by using the proposed GRU-based classifier, and the Tx location is then estimated by using a proper GRU-LE that is trained for the classified channel model. The proposed method is called a universal GRU-LE (u-GRU-LE) method. Numerical results verify that the proposed u-GRU-LE localizes the UWB Tx well irrespective of the channel models.

2. System and signal models

Consider a confined cubic area with the volume $D_1 \times D_2 \times D_3$, where D_1, D_2, and D_3 denote the width, depth, and height of the area, as illustrated in Fig. 1. Inside the area, a UWB Tx emits a signal, and P Rx's at the eight corners (i.e., $P = 8$) capture the signal. The locations of Tx and Rx_p are (x, y, z)

* Corresponding author.
E-mail addresses: tananh0404@cau.ac.kr (D.T.A. Nguyen), jgjoung@cau.ac.kr (J. Joung), kangxin@uestc.edu.cn (X. Kang).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2021.06.006

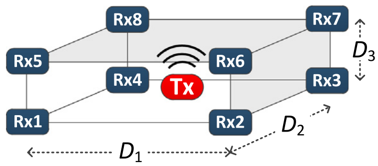

Fig. 1. 3D environment model with one Tx and $P = 8$ Rx's. The location of Tx is estimated from the received signals from P Rx's.

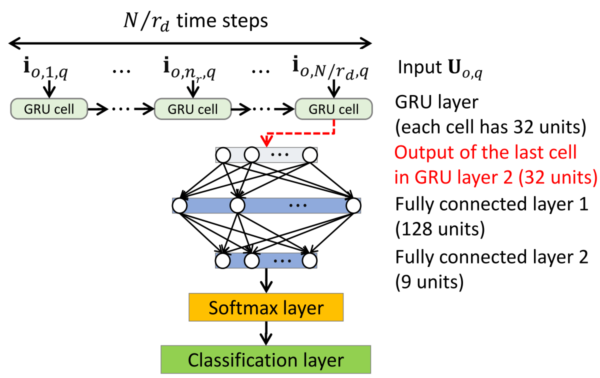

Fig. 2. Proposed GRU classifier architecture. From the P received UWB signals, CM is classified to CM-o.

Table 1
Channel model notation and environments [7].

Channel model	Environment
CM-1	Residential, LoS
CM-2	Residential, NLoS
CM-3	Office, LoS
CM-4	Office, NLoS
CM-5	Outdoor (sub-urban), LoS
CM-6	Outdoor (sub-urban), NLoS
CM-7	Industrial, LoS
CM-8	Industrial, NLoS
CM-9	Open Outdoor (Snow-covered, farm)

and (x_p, y_p, z_p), respectively, where $p \in \{1, \ldots, 8\}$. The Tx is randomly located inside the area, i.e., $x \sim \mathcal{U}[0, D_1]$, $y \sim \mathcal{U}[0, D_2]$, and $z \sim \mathcal{U}[0, D_3]$, where $\mathcal{U}[a, b]$ indicates the uniform distribution within interval $[a, b]$. The aim of the considered localization system is to estimate the location of Tx inside the area from the received signals.

For this research, consider nine UWB channel models (CMs) in Table 1 that follow the IEEE 802.15.4a standard [7]. Here, residential, office, outdoor, and industrial environments modeled with line-of-sight (LoS) or non-line-of-sight (NLoS). Assuming the localization system operates in CM-o with $o \in \{1, \ldots, 9\}$, the received signal of Rx$_p$, denoted by $r_{o,p}(t)$, is modeled as follows [7]: $r_{o,p}(t) = h_{o,p}(t) * s(t) + n_{o,p}(t)$, where $h_{o,p}(t)$ is the impulse response of the channel between Tx and Rx$_p$, $s(t)$ denotes the transmitted signal at time t, $n_{o,p}(t)$ denotes an additive white Gaussian noise at Rx$_p$, and $*$ represents the convolution operation. Here, the channel is modeled as follows [7]: $h_{o,p}(t) = \sum_{l=0}^{L_p-1} \sum_{k=0}^{K_l-1} a_{k,l} \exp(j\phi_{k,l})\delta(t - T_l - \tau_{k,l})$, where L_p represents the total number of clusters; K_l represents the total number of multipath components of the lth cluster; $a_{k,l}$ denotes the tap weight of the kth multipath component of the lth cluster; $\phi_{k,l}$ denotes the uniformly distributed phase of the kth multipath component of the lth cluster; $\delta(\cdot)$ is the delta function; T_l is the delay for the lth cluster; and $\tau_{k,l}$ is the intra-cluster delay for the kth multipath component of the lth cluster. The signal after matched filtering is denoted by $y_{o,p}(t)$ as $y_{o,p}(t) = s(T_d - t) * r_{o,p}(t)$, where T_d is the delay for causality of the signals.

3. Proposed method

In this section, we describe the designs of the GRU-based classifier and localizer, as well as the generation of the input training samples.

3.1. Input generation

The input generation method follows the method in [6]. To generate Q training samples, the signal $y_{o,p,q}(t)$ is captured in CM-o at Rx$_p$, where $o \in \{1, \ldots, 9\}$, $p \in \{1, \ldots, P\}$, and $q \in \{1, \ldots, Q\}$. These signals are then sampled with sampling frequency f_s to a complex-valued vector $\mathbf{y}_{o,p,q} \in \mathbb{C}^{N\times 1}$, whose nth element is denoted by $y_{o,p,q}[n]$, where $n \in \{1, \ldots, N\}$. Here, N is designed to be 3600. A real-valued standardized vector $\mathbf{v}_{o,p,q} \in \mathbb{R}^{N\times 1}$ is generated with the nth element defined as follows:

$$v_{o,p,q}[n] = \frac{\left|y_{o,p,q}[n]\right| - \mu_{o,p,q}}{\sigma_{o,p,q}}, \tag{1}$$

where $\mu_{o,p,q}$ and $\sigma_{o,p,q}$ are the mean and standard deviation of $\left|y_{o,p,q}[n]\right|$. To reduce the computational complexity and noise effect, $\mathbf{v}_{o,p,q}$ is down-sampled to $\mathbf{u}_{o,p,q} \in \mathbb{R}^{N/r_d\times 1}$ with down-sampling rate r_d with the n_rth element defined as follows:

$$u_{o,p,q}[n_r] = \max_{n\in\{(n_r-1)r_d+1,\ldots,n_r r_d\}} v_{o,p,q}[n], \tag{2}$$

where $n_r \in \{1, \ldots, N/r_d\}$. All the down-sampled signals at P Rx's are then combined to construct the input training matrix $\mathbf{U}_{o,q}$ as follows:

$$\mathbf{U}_{o,q} = \left[\mathbf{u}_{o,1,q}^T \quad \cdots \quad \mathbf{u}_{o,P,q}^T\right]^T \in \mathbb{R}^{P\times(N/r_d)}. \tag{3}$$

Here, $\mathbf{U}_{o,q}$ can be divided time step-wise into many column vectors, i.e., $\mathbf{U}_{o,q} = [\mathbf{i}_{o,1,q} \cdots \mathbf{i}_{o,N/r_d,q}]$. The n_rth column vector of $\mathbf{U}_{o,q}$, denoted by $\mathbf{i}_{o,n_r,q}$, is the input vector at time step n_r in GRU, whereas the pth row vector is the pth feature for the GRU learning network.

3.2. Classifier architecture

The proposed GRU-based classifier of the channel models is illustrated in Fig. 2. The classifier is a deep learning model that is composed of (i) an input layer, (ii) a GRU layer, (iii) two fully connected (FC) layers, (iv) a softmax layer, and (v) a classification layer.

The GRU layer consists of N/r_d GRU cells, which is the same as the number of input time steps. Denote the input and output vectors of a GRU cell in the lth GRU layer at time step

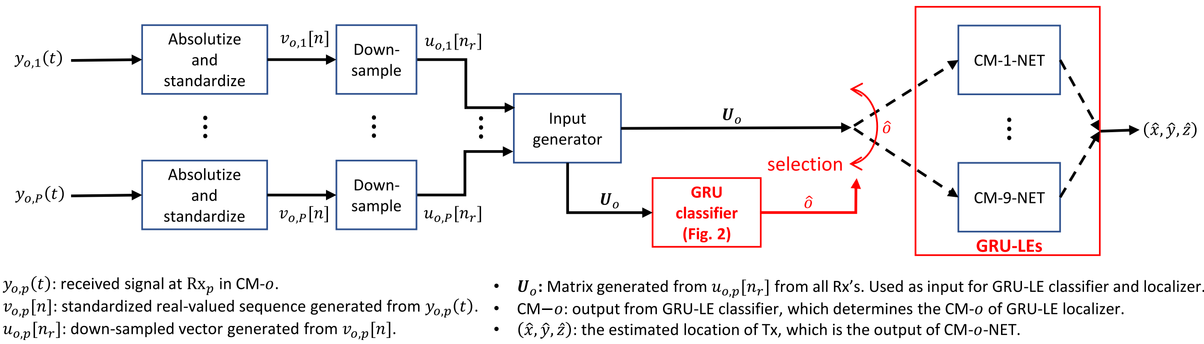

Fig. 3. Proposed universal gated recurrent unit(u-GRU)-based location estimation (u-GRU-LE) system for the nine channel models (CMs).

t by $\mathbf{g}_{l,t} \in \mathbb{R}^{I_g \times 1}$ and $\mathbf{c}_{l,t} \in \mathbb{R}^{C_g \times 1}$, respectively. The operation inside the cell is then modeled as follows:

$$\mathbf{f}_{l,t} = \sigma\left(\mathbf{W}_f \mathbf{g}_{l,t} + \mathbf{R}_f \mathbf{c}_{l,t-1} + \mathbf{b}_f\right) \tag{4}$$

$$\mathbf{m}_{l,t} = \sigma\left(\mathbf{W}_m \mathbf{g}_{l,t} + \mathbf{R}_m \mathbf{c}_{l,t-1} + \mathbf{b}_m\right) \tag{5}$$

$$\tilde{\mathbf{c}}_{l,t} = \tanh\left(\mathbf{W}_{\tilde{c}} \mathbf{g}_{l,t} + \mathbf{R}_{\tilde{c}}\left(\mathbf{m}_{l,t} \odot \mathbf{c}_{l,t-1}\right) + \mathbf{b}_{\tilde{c}}\right) \tag{6}$$

$$\mathbf{c}_{l,t} = \left(1 - \mathbf{f}_{l,t}\right) \odot \mathbf{c}_{l,t-1} + \mathbf{f}_{l,t} \odot \tilde{\mathbf{c}}_{l,t} \tag{7}$$

where $\mathbf{f}_{l,t}$, $\mathbf{m}_{l,t}$, and $\tilde{\mathbf{c}}_{l,t} \in \mathbb{R}^{C_g \times 1}$ denote the update gate, reset gate, and candidate activation of GRU cell in the lth GRU layer at time step t; $\mathbf{W}_f$, $\mathbf{W}_m$, and $\mathbf{W}_{\tilde{c}} \in \mathbb{R}^{C_g \times I_g}$ describe the input weights for $\mathbf{f}_{l,t}$, $\mathbf{m}_{l,t}$, and $\tilde{\mathbf{c}}_{l,t}$, respectively; $\mathbf{R}_f$, $\mathbf{R}_m$, and $\mathbf{R}_{\tilde{c}} \in \mathbb{R}^{C_g \times C_g}$ indicate the recurrent weights for $\mathbf{f}_{l,t}$, $\mathbf{m}_{l,t}$, and $\tilde{\mathbf{c}}_{l,t}$, respectively; $\mathbf{b}_f$, $\mathbf{b}_m$, $\mathbf{b}_{\tilde{c}} \in \mathbb{R}^{C_g \times 1}$ indicate the biases for $\mathbf{f}_{l,t}$, $\mathbf{m}_{l,t}$, and $\tilde{\mathbf{c}}_{l,t}$, respectively; $\sigma(\cdot)$ and $\tanh(\cdot)$ denote the sigma function and the tanh function, respectively; and $\odot$ denotes the element-wise multiplication operation.

A GRU cell can encode both its input and output of the cell in the previous time step. The length of the GRU cell output is considered as the number of units of the GRU layer. Here, the number of units at the GRU layer is 32. The output of the last cell will be the input for the first FC layer. The number of units for two FC layers are 128 and nine units, respectively. The classifier output is the estimated class of CM of the input signals, which is denoted by $\hat{o}$. The parameters and layers of the GRU classifier are numerically designed such that the highest accuracy of the classifier is obtained in the shortest training time.

3.3. Localizer architecture

After the CM of the input training signals is predicted at the proposed GRU classifier, a localizer model, i.e., CM-NET is selected accordingly among CM-o-NET's. The CM-o-NET is trained by the training data set generated from the received signals under the given CM-o. Each CM-o-NET follows the GRU-LE model in [6, Fig. 2] that consists of (i) an input layer, (ii) two GRU layers with 64 and 32 units, respectively, (iii) three FC layers with 256, 32, and three units, respectively, and (iv) a regression layer. Here, the second GRU layer takes the input from the output of the first GRU layer. Having two GRU layers on top of each other helps the GRU-LE model to learn different levels of abstractions of input signals over time. The regression layer calculates the loss between the output of the GRU-LE model and the ground truth location of Tx that is modeled as

$$L_{Tx} = \frac{1}{2Q}\sum_{q=1}^{Q}\left[\left(x_q - \hat{x}_q\right)^2 + \left(y_q - \hat{y}_q\right)^2 + \left(z_q - \hat{z}_q\right)^2\right], \tag{8}$$

where $\left(x_q, y_q, z_q\right)$ and $\left(\hat{x}_q, \hat{y}_q, \hat{z}_q\right)$ represent the ground truth and the estimation of Tx location for the qth training sample, respectively. The GRU-LE model training process tries to minimize the loss L_{Tx} in (8).

3.4. Structure of proposed u-GRU-LE system

Fig. 3 illustrates the proposed u-GRU-LE system. The UWB signal, $y_{o,p}(t)$, is received at Rx p and collected at the u-GRU-LE system, where $p \in \{1, \ldots, P\}$. Each signal is normalized to $v_{o,p}[n]$ in (1), which is then down-sampled to $u_{o,p}[n_r]$ in (2). A training matrix, $\mathbf{U}_o$, is structured by stacking P row vectors as shown in (3). The proposed GRU classifier in Fig. 2 predicts the class of CM, i.e., $\hat{o} \in \{1, \ldots, 9\}$, from $\mathbf{U}_o$. Once the class of CM is determined, the training matrix $\mathbf{U}_o$ is transferred to the localizer of the corresponding CM-$\hat{o}$, which is denoted by CM-$\hat{o}$-NET. The location of the Tx, i.e., $(\hat{x}, \hat{y}, \hat{z})$, is then estimated through the CM-$\hat{o}$-NET.

4. Simulation result

In this section, we compare the performance of the proposed u-GRU-LE method with the conventional GRU-LE method under various environments. As the performance metric, we use the root mean-squared-error (RMSE) between the estimated and ground truth of the 3D location of Tx that is defined as

$$RMSE \triangleq \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\left(x_t - \hat{x}_t\right)^2 + \left(y_t - \hat{y}_t\right)^2 + \left(z_t - \hat{z}_t\right)^2\right]}.$$

In a cubic area, eight Rx's are located at eight corners, i.e., $P = 8$, as illustrated in Section 2. The simulation is performed in `MATLAB 2021a`. The sampling frequency is fixed as $f_s = 24$ GHz. The down-sampling rate r_d is set to 30 as it provides the best performance for the u-GRU-LE method, which means the input training matrix size for both GRU-based models is 8×120. With $180,000$ training samples, the classifier is trained using the adaptive moment estimation

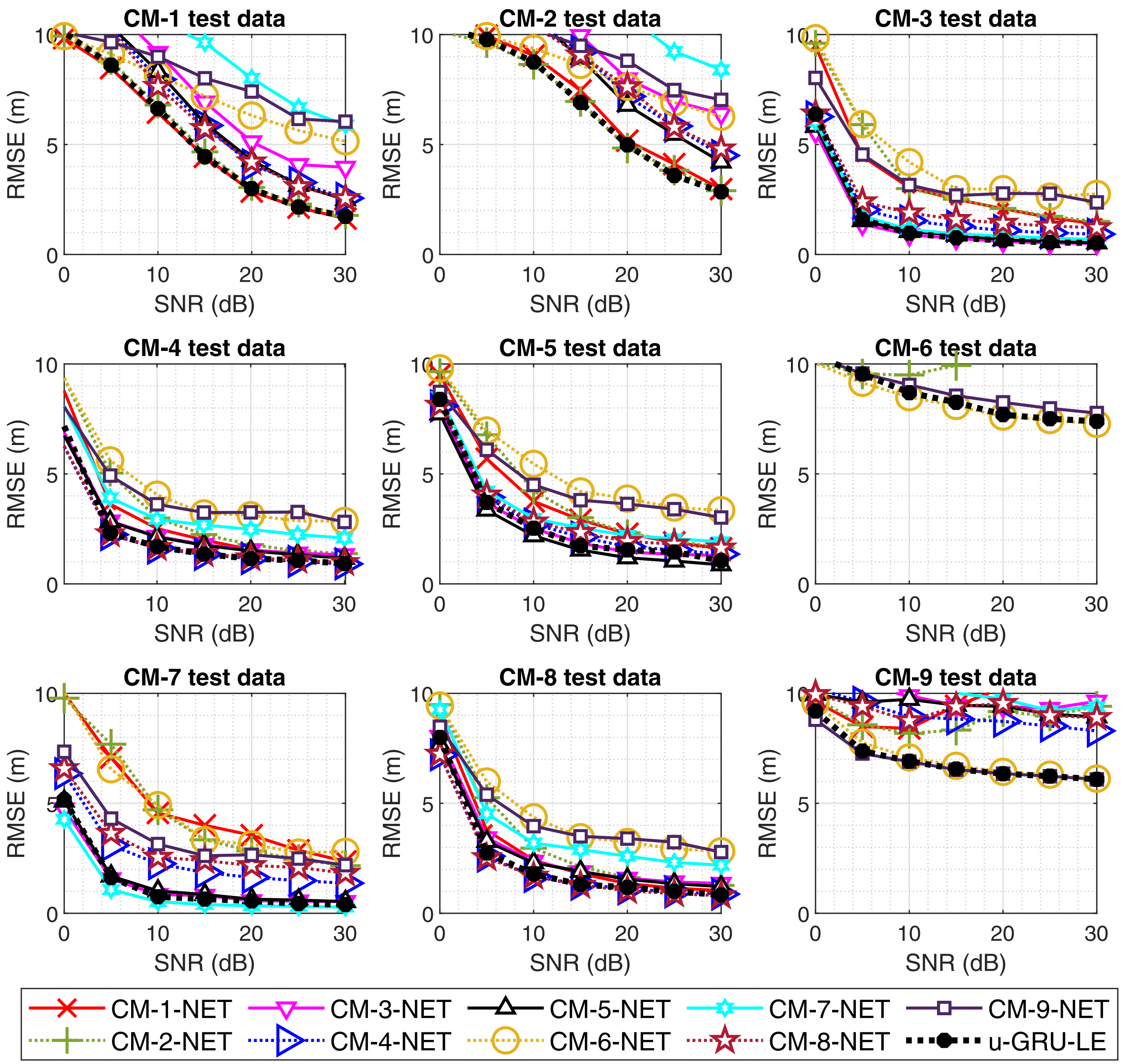

Fig. 4. RMSE performances (y-axis) across SNR (x-axis). The CM-o-NET and u-GRU-LE are tested with various CMs.

(ADAM) algorithm. The number of epochs is 20, with 200 samples per mini-batches. The initial learn rate is 0.02, which drops to 90 percent for every two epochs, and the gradient threshold is 1. It takes approximately 5 minutes to train the classifier.

In Fig. 4 RMSE of CM-o-NET and u-GRU-LE methods across SNRs under various CMs, i.e., RMSE for various test data is given. The RMSE decreases as the SNR increases. As expected, CM-o-NET operates well for the test data generated under CM-o environment. The performance of CM-o-NET, however, severely deteriorates if it is tested with data generated under CM-$\bar{o}$, where $o \neq \bar{o}$. On the other hand, the proposed u-GRU-LE operates well irrespective of the CMs (especially, for CMs with LoS). From the results, it is verified that the designed GRU classifier is effective to determine the CM, even though its accuracy is insufficient, e.g., approximately 90 % (75 %) when SNR is greater than 20 dB (10 dB), based on simulation results omitted in this paper.

5. Conclusion

In this study, we have extended the conventional GRU-based location estimation (GRU-LE) method, which is trained and operates for a specific channel model (CM), to a universal localization method operating well various CMs. To this end, we design a GRU-based classifier that effectively classifies the

CMs as numerically verified. The proposed universal GRU-LE (u-GRU-LE) that employs the designed GRU classifier can localize a UWB transmitter regardless of the types of CM if the potential CMs are sufficiently trained. Since the GRU network of the proposed u-GRU-LE does not need to be retrained according to the type of CMs when it is deployed in a different environment, the dynamic operation of the UWB localization system is possible at the cost of marginal computational complexity increase for the GRU classifier.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was supported by the National Research Foundation (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A2C1084168 & 2021R1A4A200 1316).

References

[1] N. Park, S. Jung, D. Han, An adaptive hybrid filter for practical wifi-based positioning systems, ICT Express 1 (2) (2015) 82–85.

[2] D. Han, S. Jung, S. Lee, A sensor fusion method for wi-fi-based indoor positioning, ICT Express 2 (2) (2016) 71–74.

[3] J. Joung, S. Jung, S. Chung, E.-R. Jeong, CNN-Based tx–rx distance estimation for UWB system localisation, Electron. Lett. 55 (17) (2019) 938–940.

[4] D.T.A. Nguyen, H.-G. Lee, E.-R. Jeong, H.L. Lee, J. Joung, Deep learning-based localization for UWB systems, Elect. 9 (10) (2020) 1712–1729.

[5] H. Seo, H. Kim, J. Kang, I. Jeong, W. Ahn, S. Kim, 3D Moving target tracking with measurement fusion of tdoa/fdoa/aoa, ICT Express 5 (2) (2019) 115–119.

[6] D.T.A. Nguyen, J. Joung, X. Kang, Deep gated recurrent unit-based 3D localization for UWB systems, IEEE Access 9 (2021) 68798–69798.

[7] A.F. Molisch, K. Balakrishnan, D. Cassioli, C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, K. Siwiak, IEEE 802.15.4a Channel Model - Final Report, IEEE 802.15 WPAN Low Rate Alternative PHY Task Group, Tech. Rep. IEEE 802.15-04-0662-02-004a, 2004.