IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Machine learning for practical localization system using multiview CSI

MINSEUK KIM[1], DONGSOO HAN[2], AND JUNEKOO KEVIN RHEE.[1]
[1]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea
[2]School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Korea

Corresponding author: Junekoo Kevin Rhee (e-mail: rhee.jk@ kaist.edu).

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-2018-0-01402) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

ABSTRACT Various machine learning techniques on indoor localization using radio signals are being rapidly developed to achieve a sub-meter accuracy under noisy and complex environments. A fingerprint database using channel state information (CSI) extracted from a radio packet based on an orthogonal frequency diversity multiplexing (OFDM) channel can provide enough information to localize a transmitter device with a neural network (NN) based machine learning technique. In this paper, we concern about the more practical use of the localization system using machine learning. We introduce a novel design of a signal preprocessing method for NN fingerprinting. To deal with the real building environment with corridors where certain signals cannot arrive at the receiver, our preprocessing with nonnegative matrix factorization (NMF) recovers multiview CSI of the original signal and complete the sparse CSI matrix, which enables robust and practical localization. The recovered CSI is then applied to variational inference-based machine learning that finds informative corridor views among multiview CSI. Our proposed system significantly outperforms other existing machine learning-based systems and shows a localization accuracy of 89 cm, while it still maintains the reliable accuracy even with 30 % sparse network. It is the first time to consider how to design a practical localization system in an empirical building environment.

INDEX TERMS indoor localization, channel state information (CSI), multiview data, real building environment, non-line of sight (NLoS), practical localization, network sparsity, nonnegative matrix factorization (NMF), machine learning, variational inference, joint optimization.

I. INTRODUCTION

A precise analysis of radio signals gives us an opportunity to find the accurate location of the radio device. The relative location information between transmitter and receiver devices derived from the signal characteristic can be applied for a variety of purposes in indoor areas, such as retail and healthcare by providing the client device locations. IEEE 802.11 technology provides data rates of 1 Gbps or more over wireless local area networks (WLANs), but noise and fading cause the uncertainties of the radio signal. Therefore, for practical use in indoor areas, the radio signal should be precisely analyzed to construct a reliable localization system. Compared to the global positioning system (GPS), which leads to about 5 m outdoor localization accuracy [1], indoor localization requires more accurate positioning of the transmitter devices. Such a localization system locates a transmitter based on signal information received from multiple receivers. The system collects all the signal information as a high-dimensional data and tries to find a more accurate location through geometric data analysis.

The localization systems can be built on several IEEE 802.11 protocols such as WiFi, Bluetooth, and ZigBee. Each protocol according to its standard specification results in different localization efficiency in different environments. Recently, a popular choice for a localization parameter is the channel state information (CSI) of WiFi [2]. On a radio channel of IEEE 802.11 a/b/g/n/ac using orthogonal frequency division multiplexing (OFDM), the CSIs of 64 (in a case of 20 MHz bandwidth) subcarriers are extracted at the receiver side. The CSI of each subcarrier contains not only amplitude which represents the signal strength, but also phase which represents a signal communication time. Since the subcarriers with different communication frequencies have different phases, their CSIs can be com-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

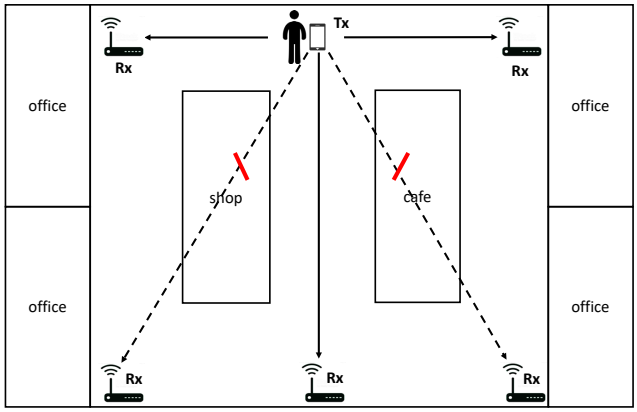


FIGURE 1: Example of a real building environment for localization

bined into high-dimensional data. Besides, the arrangement of multiple receiver antennas provides additional CSIs with different phases, with which the localization accuracy can be improved.

By analyzing the CSI phase, multiple signal classification (MUSIC) [3] finds time of flights (ToFs) and angle of arrivals (AoAs) of the received signal. Even if the signal consists of multiple propagation paths, the pseudospectral analysis of MUSIC can theoretically find multiple peaks to get the aforementioned information on the paths. However, in a real environment, multi-path fading significantly interferes with the desired dominant path information. Besides, the noise of the dominant path itself also causes fluctuation of the CSI phase. Such high phase ambiguity from the noise and fading makes the MUSIC hard to find true transmitter location. In the experiment, the maximum CSI phase error of the same transmitter-receiver location pair was about $\pi/2$, which causes several meters of localization accuracy.

Recently, machine learning techniques have been utilized to suppress the phase ambiguity and to design a more accurate localization system. For example, by using CSIs from multiple receivers as training input to a neural network (NN), phase errors can be mitigated when the NN trains the training location of a transmitter. From the perspective of supervised discriminative learning, the localization system trains for either classification, clustering, or regression to predict a test location. The localization accuracy depends on the preprocessing and organizing of the CSI training data and the training model. Obviously, the accuracy of the transmitter localization can be improved by increasing the number of receivers and their antennas.

In this paper, we focus on the practicality of the localization system. We assure reliable localization accuracy even in a real environment with several corridors as in Fig. 1. In such an environment, the wall blocks a certain line of sight (LoS) communication path of a signal and the remaining non-line of sight (NLoS) paths make the received CSI non-informative or useless for localization. If the CSI is non-informative, it corrupts the learning and degrades the localization accuracy. Besides, the signal may not even arrive at the receiver side in this complex environment. Therefore, we propose a system to handle the following two problematic cases while ensuring localization accuracy.

- Case1: corridor wall blocks LoS path, resulting in non-informative CSI data
- Case2: certain signals are lost, resulting in sparse CSI data

The proposed system consists of three steps as follows.

1) CSI preprocessing using nonnegative matrix factorization
2) Latent sampling using variational inference
3) Regression (localization) assisted by view classification

On the preprocessing step, we introduce a nonnegative matrix factorization (NMF) method to recover the lost elements of the sparse CSI matrix. As well as matrix recovery, NMF trains to extract a specific number of principal components, which represent the dominant features of the CSI. Thus, NMF can serve as both data smoothing and performance improvement of localization. To apply NMF for a test inference, we also propose a novel recovery method using the coefficient matrix obtained from training. After the CSI preprocessing, a concept of variational inference [4] is applied to our machine learning for localization. Our learning technique refers to [5], which introduced view-selective deep learning using multiview CSI data consisting of CSI views of multiple receivers. In the variational deep network, the noise and fading of multiview CSIs can be efficiently mitigated by sampling latent feature vectors consisting of multiple variables of independent Gaussian processes (GPs). Using the generated latent feature vectors, we then find the desired transmitter location as the result of the improved regression assisted by the dominant view classification.

We construct the localization system which is robust for its practical use in the real building environment. Our contribution allows for the reliable localization of the transmitter device when certain CSIs are coming from NLoS paths or do not exist. Compared to [4], we recovered the lost CSI elements by suggesting a novel signal preprocessing method using NMF. In addition, we efficiently exclude the non-informative CSIs from NLoS paths through advanced view-selective deep learning with an additional trade-off parameter. In the building experiment with three corridors, our system shows under 1 m localization accuracy, which outperforms the best-known system by 29.9 %. Besides, even if there is 30 % sparsity on the radio communication network, our novel preprocessing successfully recovers the lost CSIs and assures 1.11 m of localization accuracy, which still improves the best-known system with no sparsity.

II. RELATED WORK

Localization researches using multidimensional CSI have been explosively increased. They reported more accurate localization accuracy with complex-valued CSI data which

provides information on multiple subcarriers of multiple receiver antennas. The classical algorithm, MUSIC aims to estimate a latent steering vector that comes from a signal of multi-antenna CSIs using pseudospectral analysis. The MUSIC and other modified MUSIC algorithms [6], [7] tried to find the absolute target device location from which the dominant path generated. However, it is required to be focused on the CSI uncertainties. Even if the MUSIC enables us to find the separated information of paths including the dominant path of the signal, radio noise, hardware offset, and multi-path fading tangle the whole signal. In a practical experiment environment, CSI values across the subcarriers highly fluctuate, and due to various noise and offsets, the MUSIC marks the transmitter location to a completely different location which is away from the true location. Manual cabling [8] or bi-directional measurement [9] to resolve such problems takes too much effort to construct the localization system with the off-the-shelf devices. Therefore, many machine learning-based localization systems are being arisen. The supervised machine learning can mitigate the noise and offsets with NN fingerprinting and find invisible features generated at training locations. The systems can discriminate the test CSI data to predict the location labels. For more efficient learning, obtained raw CSIs are applied in advance to various kinds of preprocessing or calibration such as inverse fast Fourier transform (IFFT) [10], linear sanitization [11], [12], kernel density estimation (KDE) [13] or principal component analysis (PCA) [14].

Start from simple deep NN (DNN) [14], several restricted Boltzmann machine (RBM) based localization systems have been introduced [15]–[17]. The authors applied the concept of Bayes classifier [18]. After CSI data reconstruction using NN fingerprinting, the likelihood of each training location was calculated for the reconstructed test data. The test location was determined by a weighted average of posterior, which is obtained by the likelihood and the Bayes' rule. The weakness of RBM is that it takes too much computation time to calculate the likelihood for every training location. On the other hand, by considering over 10 packets as a single batch, convolutional NN (CNN) classified CSI images to the training location labels [19]–[23]. However, for real-time localization, it is better to be processed packet-by-packet for the CSI test inference. Support vector regression (SVR) based system was introduced in [24] for device-free localization which [25], [26] also considered. There are other various learning techniques for localization, such as autoencoder [27], cluster-mapping (C-Map) [28], kNN [29], multi-layer perceptron (MLP), [30], canonical correlation analysis (CCA) [31], and visibility graph (VG) [32].

The articles tried to prove the reliability of their localization systems in indoor experiments. However, the systems did not consider the NLoS signals, in other words, non-informative signals. There may exist corridors so that wall blocks the LoS signal. To let our system train only for the informative LoS signals and obtain better localization accuracy, we refer to [5] using variational inference [4], which was first used for variational autoencoder (VAE) [33], along with reparameterization scheme [34]. From a perspective of a discriminative model, authors in [35] utilized the variational inference for deep network classification, where improves SVM, stand-alone DNN, and other recent learning techniques [34]. Especially for multiview data such as in our case where the data comes from multiple receivers, we can mitigate the CSI uncertainties by inducing the multivariate standard normal distribution to approximate the latent vector of the CSI.

III. CSI PRELIMINARIES

Let us suppose a localization system where a transmitter communicates with multiple receiver access points (APs) in a radio WiFi environment. A radio communication packet on IEEE 802.11 a/b/g/n/ac OFDM channel consists of information of multiple subcarriers. Since the OFDM enables multi-input multi-output (MIMO) air interface, the receiver devices with multiple antennas could obtain high-dimensional CSI for each packet. We use a CSI extraction tool using the physical layer API on desktop APs equipped with the Intel WiFi link (IWL) 5300 network interface controller (NIC) [36]. The CSI is obtained for the subcarrier set $\{s_i\}$, where $i \in \{1, \ldots, I\}$. Due to the different channel frequencies of OFDM subcarriers, CSI can be interpreted in a more sophisticated way.

More specifically, the received signal $\hat{\mathbf{H}}$ with true CSI $\mathbf{H}$ and noise $\mathbf{N}$ is represented as $\hat{\mathbf{H}} = |\mathbf{H}|e^{j2\pi\angle\mathbf{H}} + \mathbf{N}$. For the CSI of subcarrier i and receiver antenna $m \in \{1, \ldots, M\}$, its amplitude is represented as $|H_{m,i}|$ and its phase is represented as:

$$\angle H_{m,i} = (s_i - 1)(\delta_f \tau) + (m-1) f_c \frac{d \sin\theta}{c}, \quad (1)$$

where constant δ_f is the frequency difference between subcarriers, f_c is the center frequency of the channel, d is the distance between adjacent receiver antennas, and c is the speed of light. The classical aim of MUSIC is to find true ToF τ and AoA θ caused by different communication time according to different i and m [3]. But in real 802.11 communication, several offsets and noise accompany them so that measured $\angle \hat{H}_{m,i}$ is represented as:

$$\angle \hat{H}_{m,i} = \angle H_{m,i} + s_i \times (\lambda + \epsilon) + \mu_m + \beta + Z_{m,i}, \quad (2)$$

where λ and ϵ denote the subcarrier dependent packet detection delay (PDD) and the sampling frequency offset (SFO), respectively, μ_m denotes the receiver antenna dependent carrier frequency offset (CFO), and β and $Z_{m,i}$ denote the phase-locked loop (PLL) offset and the noise, respectively [8]. Here we can take the phase difference between adjacent antennas to remove τ, λ, and ϵ as:

$$\begin{aligned} \triangle\angle \hat{H}_{m,i} &= \angle \frac{\hat{H}_{m+1,i}}{\hat{H}_{m,i}} \\ &= \angle H_{m,i} + \mu_{m+1} - \mu_m + \triangle\beta + \triangle Z \\ &= f_c \frac{d \sin\theta}{c} + \mu_{m+1} - \mu_m + \triangle\beta + \triangle Z. \end{aligned} \quad (3)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

To find ToFs, many kinds of research applied the linear sanitization previously. However in the experiments, the offsets λ and ϵ are too large to be removed. We instead remove all the subcarrier dependent parameters and consider only the phase difference between adjacent antennas. We use the data feature extracted from (3) as learning input, which consists of the CSI phase difference of multiple subcarriers of multiple receiver antenna pairs. By concatenating the data from multiple receivers at distinct locations to a single input, it becomes a multiview data containing multiple probability distributions.

In addition, to exploit the normalized phase data to machine learning, we should maintain phase continuity as the subcarrier index increases, to avoid the phase unwrapping problem. Therefore we use the CSI phase difference as a form of normalized phasors which consist of in-phase and quadrature values $I_{m,i}$ and $Q_{m,i}$, respectively:

$$e^{j2\pi\triangle\angle\hat{H}_{m,i}} = I_{m,i} + jQ_{m,i}. \tag{4}$$

With V receivers, the multiview learning input $\mathbf{X} \in \mathbb{R}^{2\times\mathbf{I}\times(\mathbf{M}-\mathbf{1})\times\mathbf{V}}$ is a collection of all information $\angle\mathbf{H}$ for every sample packet.

IV. CSI PREPROCESSING: NMF

To design a practical localization system in a corridor-existing environment as in Fig. 1, let us suppose a situation that several receiver APs at distinct locations receive CSIs from a transmitter at the same time. Obviously, APs near the transmitter can receive relatively good CSI. However, APs behind a wall may not receive the CSI. We describe an example of CSI data matrix composition in (5):

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccccc} \text{AP1} & \text{AP2} & \text{AP3} & \text{AP4} & \text{AP5} \end{array} \\ \begin{bmatrix} \circ & \circ & \times & \circ & \circ \\ \times & \times & \circ & \circ & \circ \\ \circ & \circ & \circ & \times & \times \end{bmatrix} \end{array} \begin{array}{l} \text{packet at location 1,} \\ \text{packet at location 2,} \\ \text{packet at location 3,} \end{array} \tag{5}$$

where each row indicates a packet at a distinct location and each column indicates an index of AP. Neither the MUSIC nor the machine learning cannot be applied to this sparse CSI data. In order to complete such a sparse CSI data matrix with only the given information and utilize it as a learning input, we recover the empty elements through a matrix factorization (MF) method. Since the transmitter location is potentially related to all of the CSI elements, NMF of which the consisting matrix elements are all nonzero is applied. As well as PCA and singular value decomposition (SVD), NMF is one of the representative methods for matrix completion. With the coefficient matrix obtained from the NMF, we can predict the empty elements for the training CSI and even for a single vector of test CSI inference. In addition, we can extract the dominant features that have reduced dimension to properly represent the original CSI data, in other words, smooth the data as investigated in Fig. 2.

Suppose we have a CSI matrix $\mathbf{X}$ consisting of N CSIs of T subcarriers where its lost CSI elements are assumed to have zero values. Then $\mathbf{X}$ can be approximated by $\hat{\mathbf{X}}$ which is represented as a multiplication of a basis matrix $\mathbf{P}$ and a coefficient matrix $\mathbf{Q}$, $\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{PQ}$:

$$\underset{(N\times T)}{\hat{\mathbf{X}}} = \underset{(N\times L)}{\mathbf{P}} \cdot \underset{(L\times T)}{\mathbf{Q}} \tag{6}$$

Here, the dimension of column of $\mathbf{P}$ and row of $\mathbf{Q}$ is L, which indicates the number of latent factors. L is much less than N as well as T so that should be empirically decided to extract latent feature. To approximate $\hat{\mathbf{X}}$ to $\mathbf{X}$, we define a residual of squares as a cost function:

$$\begin{aligned} E^2 &= \|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \|\mathbf{X} - \mathbf{PQ}\|^2 \\ &= \Sigma_{n=1}^{N}\Sigma_{t=1}^{T}(x_{nt} - \hat{x}_{nt})^2, \end{aligned} \tag{7}$$

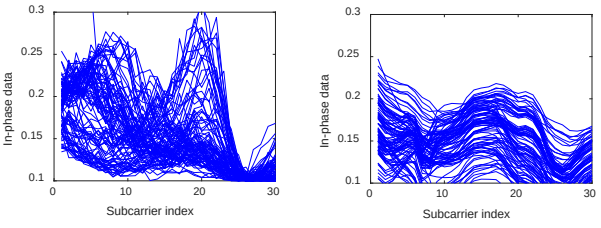
for the received nonzero CSI elements $\{x_{nt}\}$. The elements of $\mathbf{P}$ and $\mathbf{Q}$ are updated by the rules:

$$\begin{aligned} p_{nl} &\leftarrow p_{nl}\frac{(\mathbf{XQ}^\top)_{nl}}{(\mathbf{PQQ}^\top)_{nl}}, \\ q_{lt} &\leftarrow q_{lt}\frac{(\mathbf{P}^\top\mathbf{X})_{lt}}{(\mathbf{P}^\top\mathbf{PQ})_{lt}}, \end{aligned} \tag{8}$$

respectively, while being kept as nonnegative values. The cost in (7) is nonincreasing under the update rules by the proof in [37]. Even with the sparse CSI input $\mathbf{X}$, we can recursively update $\mathbf{P}$ and $\mathbf{Q}$ on an element by element basis and recover the empty elements in $\mathbf{X}$. A pseudocode of NMF for training CSI is described in Algorithm 1.

In addition, with a sparse test inference $\mathbf{x}$ which is a form of single vector, we predict a basis vector $\mathbf{p}$ of it and recover the empty elements ($\mathbf{x} \approx \hat{\mathbf{x}} = \mathbf{pQ}$):

$$\underset{(1\times T)}{\hat{\mathbf{x}}} = \underset{(1\times L)}{\mathbf{p}} \cdot \underset{(L\times T)}{\mathbf{Q}} \tag{9}$$



(a) (b)

FIGURE 2: NMF of In-phase CSI data. (a) Data before NMF. (b) Data after NMF. The data is smoothed with dominant feature extraction.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

Algorithm 1: Preprocessing: Nonnegative matrix factorization (NMF) for CSI recovery

Input: sparse training CSI matrix $\mathbf{X}$, number of latent factors L, basis matrix $\mathbf{P}$, coefficient matrix $\mathbf{Q}$, and learning rate α

Output: recovered training CSI $\hat{\mathbf{X}}$ and result of $\mathbf{Q}$ initialization;

while *epoch* $\leq$ *maxepoch* **do**
  while $n \leq N$ **do**
    while $t \leq T$ **do**
      if x_{nt} *exists* **then**
        while $l \leq L$ **do**
          update $p_{nl} \leftarrow p_{nl} \frac{(\mathbf{XQ}^\top)_{nl}}{(\mathbf{PQQ}^\top)_{nl}}$;
          update $q_{lt} \leftarrow q_{lt} \frac{(\mathbf{P}^\top\mathbf{X})_{lt}}{(\mathbf{P}^\top\mathbf{PQ})_{lt}}$;

$\hat{\mathbf{X}} = \mathbf{PQ}$;

Algorithm 2: Preprocessing: Prediction for sparse test CSI inference

Input: sparse test CSI vector $\mathbf{x}$ and coefficient matrix $\mathbf{Q}$

Output: recovered test CSI $\hat{\mathbf{x}}$

$\mathbf{x} \leftarrow \mathbf{x}[\,1, \text{nonzero}(\mathbf{x})]$; // take received CSI elements

$\mathbf{Q} \leftarrow \mathbf{Q}[\,:, \text{nonzero}(\mathbf{x})]$; // take corresponding vectors

calculate $\mathbf{p} = \mathbf{xQ}^\top(\mathbf{QQ}^\top)^{-1}$;

$\hat{\mathbf{x}} = \mathbf{pQ}$;

The coefficient matrix $\mathbf{Q}$ obtained from the NMF of training CSI is supposed to be equally applied to the test inference. In order to approximate $\hat{\mathbf{x}}$ to $\mathbf{x}$, we utilize $\mathbf{Q}$ again for an inverse transformation. The residual sum of squares can be represented as:

$$e^2 = (\mathbf{x} - \mathbf{pQ})(\mathbf{x} - \mathbf{pQ})^\top. \tag{10}$$

Then a partial derivative with respect to $\mathbf{p}$ is derived to optimize:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{p}} e^2 &= \frac{\partial}{\partial \mathbf{p}} (\mathbf{x} - \mathbf{pQ})(\mathbf{x} - \mathbf{pQ})^\top \\
&= \frac{\partial}{\partial \mathbf{p}} (\mathbf{xx}^\top - \mathbf{xQ}^\top\mathbf{p}^\top - \mathbf{pQx}^\top + \mathbf{pQQ}^\top\mathbf{p}^\top) \\
&= \frac{\partial}{\partial \mathbf{p}} (\mathbf{xx}^\top - 2\mathbf{xQ}^\top\mathbf{p}^\top + \mathbf{pQQ}^\top\mathbf{p}^\top) \\
&= \frac{\partial}{\partial \mathbf{p}} (-2\mathbf{xQ}^\top\mathbf{p}^\top + (\mathbf{pQ})(\mathbf{pQ})^\top) \\
&= -2\mathbf{xQ}^\top + 2\mathbf{pQQ}^\top \\
&= 0.
\end{aligned} \tag{11}$$

Finally $\mathbf{p}$ is obtained as:

$$\mathbf{p} = \mathbf{xQ}^\top(\mathbf{QQ}^\top)^{-1}, \tag{12}$$

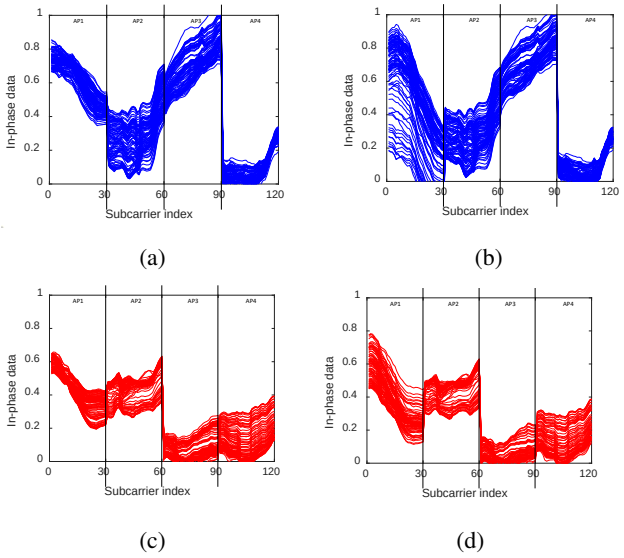


FIGURE 3: CSI recovery using NMF. (a) Original training CSI. (b) Recovered training CSI for lost data of AP1. (c) Original test CSI. (d) Recovered test CSI for lost data of AP1.

with the received nonzero CSI elements of $\mathbf{x}$ and corresponding vectors of $\mathbf{Q}$. Now we can use $\mathbf{p}$ and $\mathbf{Q}$ to predict desired $\hat{\mathbf{x}} = \mathbf{pQ}$ for the test inference. Our NMF has a strong advantage as it processes test inference on a packet by packet basis for robust real-time localization. A pseudocode of NMF for test CSI is described in Algorithm 2.

The results of NMF sparse recoveries for both 100 packets of training and test CSIs are shown in Fig. 3. In the figures, empty CSI elements lost from AP1 are recovered by referring to the remaining elements of AP2, AP3, and AP4. For the test CSIs, their elements can be recovered for a single vector inference, rather than a two-dimensional matrix. As you can see, the original and recovered data may be different for better feature extraction.

V. SYSTEM MODEL

We refer to the machine learning technique introduced in [5] which used multiview data consisting of multiple AP views to improve localization accuracy. As seen in Fig. 1, the APs may receive packets of NLoS paths from the transmitter behind a wall. However, they need not be utilized in training since the corridor wall blocks the informative LoS signals. Therefore we divide the multiview CSI into multiple corridor views to train selectively for informative views.

A. VARIATIONAL DEEP NETWORK

The variational deep network for regression basically utilizes a probabilistic model to approximate a posterior $q(\mathbf{z}|\mathbf{y})$ to an original distribution $p(\mathbf{z}|\mathbf{y})$, using a latent feature vector $\mathbf{z}$ extracted from NNs and given true location label $\mathbf{y}$. Let us suppose a localization system on a network topology consisting of K corridors. Then the CSI input $\mathbf{X}$ can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

Algorithm 3: Localization: variational deep network

Input: CSI data $\mathbf{X}$, label $\mathbf{y}$, given view label $\mathbf{u}$, number of views K, number of latent variables J, coefficient parmeter β, latent sampling NN h, and regression NN g
Output: latent feature vector $\mathbf{z}$
initialization;
while *epoch* $\leq$ *maxepoch* **do**
  while $k \leq K$ **do**
    $\mathbf{z}_k = h_k(\mathbf{X}_k)$; // hidden layers for latent sampling
    $\hat{\mathbf{y}} = g_k(\mathbf{z}_k|\mathbf{y})$; // hidden layers for regression
    loss = $\{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\} + \beta\mathbf{KL}[q(\mathbf{z}_k|\mathbf{y})||p(\mathbf{z}_k)]$;
    if $u_k == 1$ **then**
      update NNs h_k and g_k; // to minimize loss
    else
      do nothing;

Algorithm 4: Localization: view-classified regression network

Input: latent feature vector $\mathbf{z}$, label $\mathbf{y}$, given view label $\mathbf{u}$, number of views K, trade-off parameter γ, view classification NN h_Q, and regression NN h_R
Output: regression result $\hat{\mathbf{y}}$
initialization;
while *epoch* $\leq$ *maxepoch* **do**
  while $k \leq K$ **do**
    $\tilde{u}_k = \frac{u_k}{\Sigma_{i=1}^{K} u_i}$; // normalization
  $\hat{\mathbf{u}} = h_Q(\mathbf{z}|\tilde{\mathbf{u}})$; // hidden layers for view classification
  while $k \leq K$ **do**
    $\mathbf{z}'_k = \hat{u}_k \odot \mathbf{z}_k$; // reweight of latent vectors
  $\hat{\mathbf{y}} = h_R(\mathbf{z}'|\mathbf{y})$; // hidden layers for regression
  loss = $\gamma\{((y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\} + (1 - \gamma)\{\Sigma_k(\tilde{u}_k - \hat{u}_k)^2\}$;
  update NNs h_Q and h_R; // to minimize loss

divided into subsets $\{\mathbf{X}_1, \ldots, \mathbf{X}_K\}$ where each corridor view $\mathbf{X}_k$ represents the CSI collected from APs within the corridor $k = \{1, \ldots, K\}$. The mapping $\mathbf{z}_k = h_k(\mathbf{X}_k)$ of latent sampling NN extracts a latent feature vector using variational inference generated from J variables of GPs, i.e., $z^j \sim \mathcal{GP}(\mu_j, \sigma_j^2)$ with mean-field feature vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. Then from each latent vector $\mathbf{z}_k$, we obtain regression output $\hat{\mathbf{y}}$ through the mapping $\hat{\mathbf{y}} = g_k(\mathbf{z}_k|\mathbf{y})$ of regression NN. The weights and biases of NNs $h_k(\mathbf{X}_k)$ and $g_k(\mathbf{z}_k|\mathbf{y})$ are updated to jointly minimize the objectives of the variational deep network:

$$\{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\} + \beta\mathbf{KL}[q(\mathbf{z}_k|\mathbf{y})||p(\mathbf{z}_k)], \quad (13)$$

where the terms mean regression loss of two-dimensional Cartesian coordinate and Kullback–Leibler (KL) divergence, respectively. We modified the variational deep network of [5] by adding the trade-off parameter β which was introduced in β-VAE [38]. We set β as a value of 100 in our regression model for more balanced learning. We also utilize the information $\mathbf{u}$ named given view label used in [5], where u_k has a value of 1 only if the location of the sample is within corridor k and the CSI of it is considered to be informative. The pseudocode of our variational deep network is described in Algorithm 3.

B. VIEW-CLASSIFIED REGRESSION NETWORK

The latent feature vector $\{\mathbf{z}_k\}$ obtained from the section V-A is used for view-classified regression. In this section, we first classify the concatenated $\mathbf{z} = [\mathbf{z}_1, \ldots, \mathbf{z}_K]$ to $\hat{\mathbf{u}} = [\hat{u}_1, \ldots, \hat{u}_K]$ where each element $\hat{u}_k$ means the likelihood that the sample will be in corridor k. Again we use the normalized given view label as $\tilde{u_k} = \frac{u_k}{\Sigma_{i=1}^{K} u_i}$ to apply it for mapping $\hat{\mathbf{u}} = h_Q(\mathbf{z}|\tilde{\mathbf{u}})$ of view-classified NN. The obtained $\hat{\mathbf{u}}$ passed through Softmax activation is then multiplied to $\mathbf{z}_k$ in a way of element-wise reweighting to reflect the importance of view k for regression. A mapping $\hat{\mathbf{y}} = h_R(\mathbf{z}'|\mathbf{y})$ of regression NN obtains the desired output $\hat{\mathbf{y}}$, using reweighted latent feature vector $\mathbf{z}'$. The two NNs $h_Q(\mathbf{z}|\tilde{\mathbf{u}})$ and $h_R(\mathbf{z}'|\mathbf{y})$are updated to jointly minimize the losses:

$$\gamma\{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\} + (1 - \gamma)\{\Sigma_k(\tilde{u}_k - \hat{u}_k)^2\}, \quad (14)$$

where the terms mean regression and view classification losses, repectively. For more balanced learning, the trade-off parmeter γ should be adjusted. A pseudocode of view-classified regression network is described in Algorithm 4.

VI. FIELD EXPERIMENT

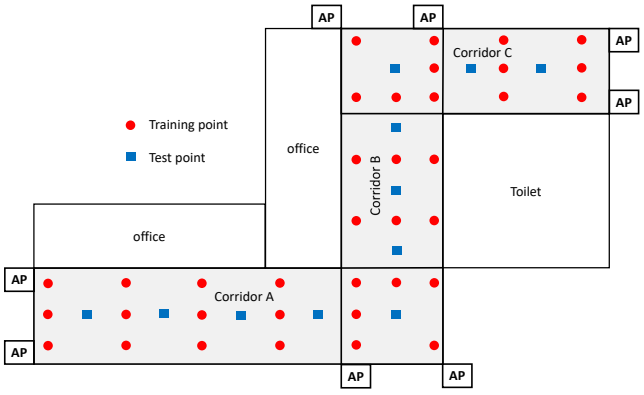


FIGURE 4: Experiment layout

We apply our localization system in a building environment as illustrated in Fig. 4. The building structure is constructed of three corridors. Totally eight receiver APs are allocated in the topology, where each corridor view consists of the CSIs received at four APs. All APs receive CSI packets

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

TABLE 1: Experiment parameters for localization system

Symbol	Parameter	Value
M	Number of receiver antannas	3
I	Number of subcarriers for 20MHz bandwidth OFDM channel	30
d	distance between adjacent antannas	3 cm
f_c	Center frequency of OFDM channel	5.18 GHz
δ_f	Frequency difference between adjacent antennas	312.5 KHz
V	receiver APs	8
N	Total size per CSI inference	2 × 30 × (3-1) × 8 = 960
T	Total number of packets for training	100 packets × 36 training locations = 3600
L	Number of latent factors for NMF	30
K	Number of corridors, i.e., views	3
J	Number of latent variables for variational deep network	120
α	Learning rate	10^{-5}
β	Trade-off coefficient between regression and KL divergence	100
γ	Trade-off coefficient between regression and view classification	0.2
$h_k^1, h_k^2, g_k^1, g_k^2$	Hidden nodes for variational deep network	1000,500,1000,500
$h_Q^1, h_Q^2, h_R^1, h_R^2$	Hidden nodes for view-classified regression network	1000,500,1000,500

from a transmitter at the same time using monitor mode, which allows the packets to be captured without having an association. For the transmitter, 36 training and 11 test points are in distributed locations. As seen in the figure, depends on the location of the transmitter, it is determined whether each corridor view is informative or not. For every transmitter location, one or two views are considered to be informative. The values of parameters for our system are specified in Table 1. In a hardware setup, we install the desktop APs equipped with M = 3-antenna IWL 5300 NIC. The adjacent antennas are apart by d = 3 cm, approximately by a half wavelength on channel 36 at f_c = 5.18 GHz. 30 of 64 OFDM subcarriers are used where the adjacent subcarriers are 312.5 KHz (= 20 MHz / 64) apart. Therefore a total of N = 960-size CSI data is collected for every packet from the transmitter. We collect the CSIs of 100 packets at each location, which means T = 3600 packets for training. For the parameters L, J, α, β, γ, and the hidden nodes, we specify the values that are optimized to obtain the best performance in the experiment.

At first, we show the impact of L, the number of latent factors for NMF in Fig. 5. As well as the regression loss which is represented as the first term in (14) and can be substitute into localization accuracy, we plot the average reconstruction loss E^2/NT, using E^2 derived in (7). With a larger L value, we can approximate a more accurate CSI input matrix so that reduce the reconstruction loss. However, too large L value cannot capture the latent feature from the N = 960 size of the input. The best regression is obtained for

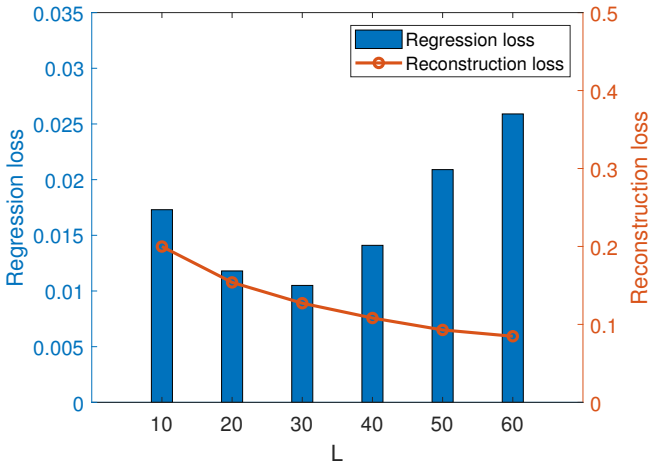


FIGURE 5: Regression and reconstruction losses versus L, number of latent factors

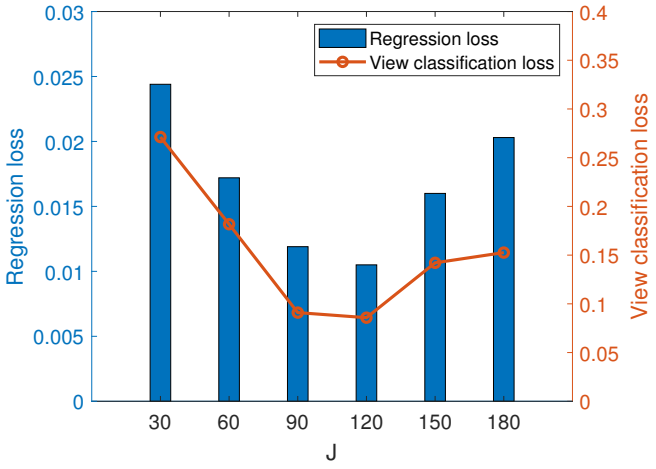


FIGURE 6: Regression and view classification losses versus J, number of latent variables for variational deep network

L = 30, which is a value of N/32. Such a high compression ratio implies that for the same receiver antenna pair m and receiver AP v, the CSI feature can be expressed in a simple way across all subcarriers.

The impact of another parameter, J which is the number of latent variables for the variational deep network is plotted in Fig. 6. We plot both the regression and the view classification losses which are derived in (14). Compare to L, the variables J of GPs should have larger value to minimize the KL divergence which can be represented as a reparameterized function of mean-field vectors [34]. The best regression is obtained for J = 120 which is a value of N/8. For too large J value, the performance becomes worse. Here, we can see our system properly works since both the regression and view classification losses change in the same tendency.

The two trade-off parameters β and γ for the variational deep network and view-classified regression network, respectively, should also be considered as important parameters to improve localization accuracy. Although the β is not

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3029598, IEEE Access

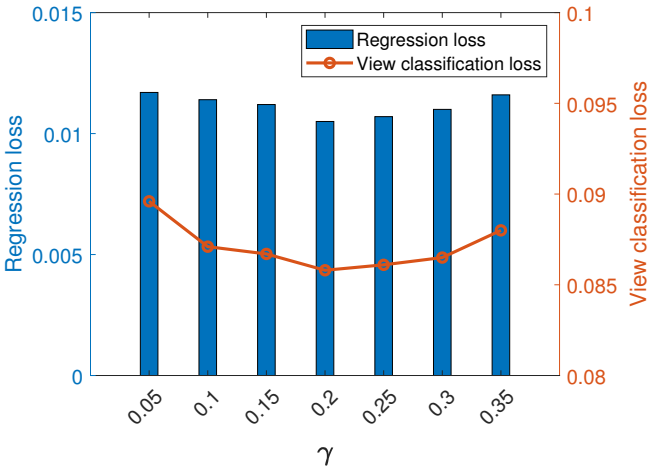


FIGURE 7: Regression and view classification losses versus γ, trade-off coefficient

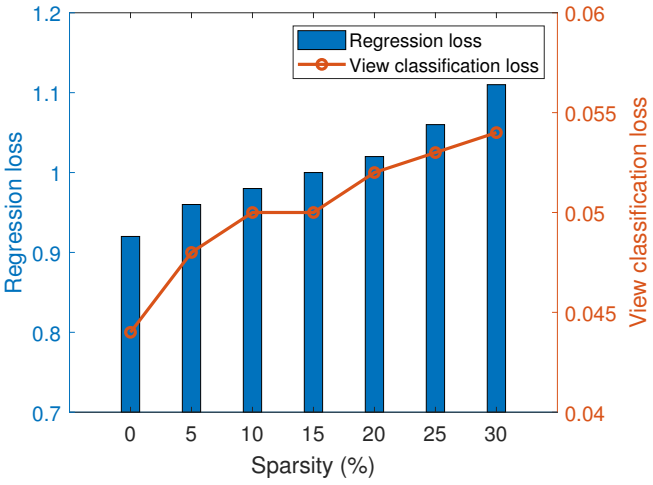


FIGURE 8: Regression and view classification losses versus data sparsity

investigated in detail, we find the best regression when it has a value of 100. For the parameter γ suggested in [5], we plot the related performances as it increases from 0.05 to 0.35 in Fig. 7. Our system obtains the best regression loss when $\gamma = 0.2$, in other words, when the system considers the view classification loss as four times important as regression loss. As γ increases to 1, it gradually ignores the view classification loss and results in poor regression performance as well. It implies that our reweighting scheme from view classification highly assists the desired regression.

In Fig. 8, we also prove that our preprocessing method using NMF properly recovers the CSI data matrix while its latent feature is preserved. In the experiment topology, a certain AP may lose to receive CSI since the AP and the transmitter are in different corridors. In such a situation, we assume that the APs lose CSIs with a probability of from 5 % to 30 %. The NMF recovers the empty elements of the CSI data matrix which is constructed together with lost CSIs.

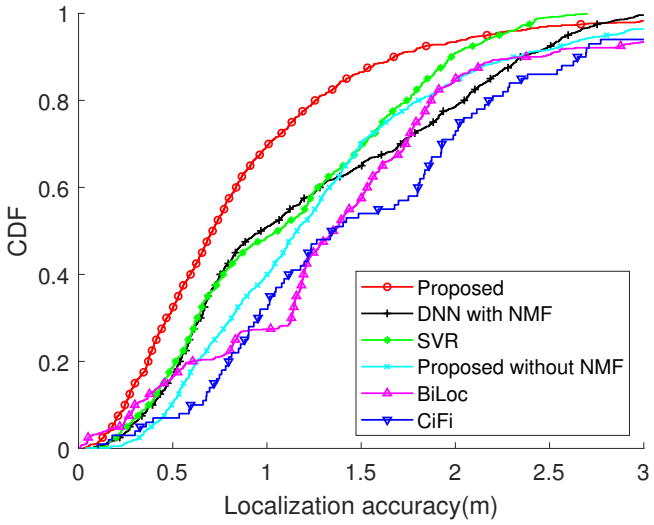


FIGURE 9: Localization accuracy CDF of the systems. The proposed system significantly outperforms other previous systems

Algorithm	Localization accuracy (m)
Proposed	0.89
DNN with NMF	1.22
SVR	1.27
Proposed without NMF	1.29
BiLoc	1.41
CiFi	1.54

TABLE 2: Localization accuracy comparison. The proposed system improves 29.9 % of the localization accuracy

Even with a sparsity of 30 %, we can assure the localization accuracy as 1.11 m. The performance has been reduced by 20 %, but it is still reliable and implies that our system is practically constructed for sparse CSIs.

In order to compare the performance of our system, the existing localization systems using various machine learning techniques are implemented. More specifically, we implement the CiFi [24] by using 10 packets as a batch to generate each CSI image and applying CNN. The Biloc [17] is implemented based on RBM and Bayes's rule. The SVR which was proposed in [19] is also implemented. Besides, we implement a system applying stand-alone DNN with NMF to see how much our preprocessing improves the performance. On the other hand, the proposed system without NMF is also implemented. In terms of localization accuracy, we show the performances of the systems in our experiment scenario in Fig. 9 and Table 2. First of all, the CiFi shows poor accuracy of 1.54 m since it is not appropriate to analyze multiview CSI data as images. To improve the accuracy, it requires more packets than the number of 100 which we collected, in order to generate large size CSI images. The BiLoc shows an accuracy of 1.41 m, a bit improved than CiFi but still unreliable. Besides, it takes too much computation

time to compute the reconstruct loss of test inference for every training location. The SVR shows 1.27 m accuracy, which is the best among the existing systems. However, the systems consider every view including non-informative views, which negatively impact the learning and thus result in poor localization performances. Therefore the systems SVR, BiLoc, and CiFi cannot take any advantages in such a real building scenario and their performances are even worse than that of stand-alone DNN with our NMF, which shows an accuracy of 1.22 m. Since our system selectively updates NNs only for the informative views and classifies the dominant view to consider, it significantly outperforms other systems. With the best parameters discussed in the above figures, we investigate the performance of our system as 0.89 m, which is 29.9 % improved compared to that of SVR. On the other hand, the NMF preprocessing of our system improves the accuracy by 31 %, from 1.29 m to 0.89 m.

VII. CONCLUSION

In this paper, we discussed the practicality of the localization system using CSI on the WiFi OFDM channel. In a real building environment with corridors, CSIs of NLoS radio signals can be lost or become useless. To recover the lost CSI information and make the localization system practical, we introduced the preprocessing method, which recovers empty elements of CSI matrix by factorizing the latent feature. Even in the case of a single test CSI where 30 % of the information was lost, we assured 1.11 m localization accuracy, which still improves the best-known system with no sparsity. In addition by considering the CSIs collected at multiple receiver APs as multiview learning data, our variational inference-based learning achieved the localization accuracy of 89 cm, by finding the dominant corridor view that assists the regression. This work lays the foundations for advanced localization technology in complex environments. Furthermore, the scheme used in our system can be widely applied in applications that aim to selectively extract features from the sparse multiview learning data. In future work, we plan to verify our system in more complex environments. For future applications, we need to figure out the trends in how the best parameters of our system change as the size of CSI input and the number of views increases.

REFERENCES

[1] H. Teng, K. Ota, A. Liu, T. Wang, and S. Zhang, "Vehicles joint uavs to acquire and analyze data for topology discovery in large-scale iot systems," *Peer-to-Peer Networking and Applications*, pp. 1–24, 2020.

[2] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.

[3] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[4] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1303–1347, 2013.

[5] M. Kim, C. Kim, D. Han, and J.-K. K. Rhee, "Multiview variational deep learning with application to scalable indoor localization," 2020.

[6] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 2015, pp. 269–282.

[7] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*, 2013, pp. 71–84.

[8] A. Tzur, O. Amrani, and A. Wool, "Direction finding of rogue wi-fi access points using an off-the-shelf mimo–ofdm receiver," *Physical Communication*, vol. 17, pp. 149–164, 2015.

[9] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, 2016, pp. 165–178.

[10] L. Chen, I. Ahriz, and D. Le Ruyet, "Csi-based probabilistic indoor position determination: An entropy solution," *IEEE Access*, vol. 7, pp. 170 048–170 061, 2019.

[11] W. Gong and J. Liu, "Roarray: Towards more robust indoor localization using sparse recovery with commodity wifi," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1380–1392, 2018.

[12] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the mona lisa: Spot localization using phy layer information," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 2012, pp. 183–196.

[13] T.-C. Tai, K. C.-J. Lin, and Y.-C. Tseng, "Toward reliable localization by unequal aoa tracking," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 444–456.

[14] X. Dang, J. Ren, Z. Hao, Y. Hei, X. Tang, and Y. Yan, "A novel indoor localization method using passive phase difference fingerprinting based on channel state information," *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, p. 1550147719844099, 2019.

[15] X. Wang, L. Gao, S. Mao, and S. Pandey, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2016.

[16] X. Li, J. Shi, and J. Zhao, "Defe: indoor localization based on channel state information feature using deep learning," in *Journal of Physics: Conference Series*, vol. 1303, no. 1. IOP Publishing, 2019, p. 012067.

[17] X. Wang, L. Gao, and S. Mao, "Biloc: Bi-modal deep learning for indoor localization with commodity 5ghz wifi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[18] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, "Passive indoor localization based on csi and naive bayes classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1566–1577, 2017.

[19] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with csi images," *IEEE Transactions on Network Science and Engineering*, 2018.

[20] S. Tewes, A. A. Ahmad, J. Kakar, U. M. Thanthrige, S. Roth, and A. Sezgin, "Ensemble-based learning in indoor localization: A hybrid approach," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–5.

[21] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[22] B. Berruet, O. Baala, A. Caminada, and V. Guillet, "Delfin: a deep learning based csi fingerprinting indoor localization in iot context," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018, pp. 1–8.

[23] P. Chen, F. Liu, S. Gao, P. Li, X. Yang, and Q. Niu, "Smartphone-based indoor fingerprinting localization using channel state information," *IEEE Access*, vol. 7, pp. 180 609–180 619, 2019.

[24] R. Zhou, J. Chen, X. Lu, and J. Wu, "Csi fingerprinting with svm regression to achieve device-free passive localization," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2017, pp. 1–9.

[25] R. Zhou, M. Hao, X. Lu, M. Tang, and Y. Fu, "Device-free localization based on csi fingerprints and deep neural networks," in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2018, pp. 1–9.

[26] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "Lifs: low human-effort, device-free localization with fine-grained subcarrier information," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 243–256.

[27] H.-C. Tsai, C.-J. Chiu, P.-H. Tseng, and K.-T. Feng, "Refined autoencoder-based csi hidden feature extraction for indoor spot localization," in *2018 IEEE 88th vehicular technology conference (VTC-Fall)*. IEEE, 2018, pp. 1–5.

[28] W. Liu, Q. Cheng, Z. Deng, X. Fu, and X. Zheng, "C-map: Hyper-resolution adaptive preprocessing system for csi amplitude-based fingerprint localization," *IEEE Access*, vol. 7, pp. 135 063–135 075, 2019.

[29] Z. Gao, Y. Gao, S. Wang, D. Li, Y. Xu, and H. Jiang, "Crisloc: Reconstructable csi fingerprintingfor indoor smartphone localization," *arXiv preprint arXiv:1910.06895*, 2019.

[30] C.-H. Hsieh, J.-Y. Chen, and B.-H. Nien, "Deep learning-based indoor localization using received signal strength and channel state information," *IEEE access*, vol. 7, pp. 33 256–33 267, 2019.

[31] T. F. Sanam and H. Godrich, "A multi-view discriminant learning approach for indoor localization using amplitude and phase features of csi," *IEEE Access*, vol. 8, pp. 59 947–59 959, 2020.

[32] Z. Wu, L. Jiang, Z. Jiang, B. Chen, K. Liu, Q. Xuan, and Y. Xiang, "Accurate indoor localization based on csi and visibility graph," *Sensors*, vol. 18, no. 8, p. 2549, 2018.

[33] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

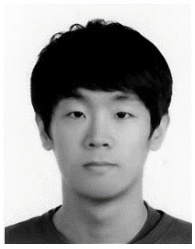
[34] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.

[35] A. G. Wilson, Z. Hu, R. R. Salakhutdinov, and E. P. Xing, "Stochastic variational deep kernel learning," in *Advances in Neural Information Processing Systems*, 2016, pp. 2586–2594.

[36] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 159–170, 2011.

[37] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[38] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.



MINSEUK KIM received the B.E. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree. His research interests include content delivery network, wireless networking, and machine learning, especially for indoor localization

DONGSOO HAN received the PhD degree in information science from the Kyoto University. He is a professor of computer science at the Korea Advanced Institute of Science and Technology (KAIST). He is also a director at the Indoor Positioning Research Center, KAIST. His research interests include indoor positioning, pervasive computing, and location-based mobile applications

JUNE-KOO KEVIN RHEE received the B.E. and M.Sc. degrees from Seoul National University, Seoul, South Korea, in 1988 and 1990, respectively, and the Ph.D. degree from the University of Michigan, Ann Arbor, MI, USA, in 1995, all in electrical engineering. He is currently a Professor of electrical engineering with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. Prior to joining KAIST, he held research and teaching positions with Princeton University (1995–1996), NEC Research Institute (1996–1998), Corning Incorporated (1998–2002), Samsung Advanced Institute of Technology (2003–2005), and the Information and Communications University (2005–2009). He holds more than 60 patents. He has authored or coauthored more than 180 technical journal and conference papers. His research interests include the areas of optical and wireless networking, quantum communication, and quantum computing. Prof. Rhee was the recipient of the Ministry Certificate of Commendation from the Korean Ministry of Science, ICT, and Planning