Transformable Fingerprinting with Deep Metric Learning Approach for Indoor Localization

Xiangsheng Zeng¹, Limin Xiao¹, Ming Zhao¹, Xibin Xu¹ and Yunzhou Li¹

Beijing National Research Center for Information Science and Technology, Dept. of Electronic Engineering, Tsinghua University, Beijing, 100084, China
Email: zengxs17@mails.tsinghua.edu.cn; xiaolm@tsinghua.edu.cn; zhaoming@tsinghua.edu.cn; xuxb@tsinghua.edu.cn; liyunzhou@tsinghua.edu.cn

Abstract. Within state-of-the-art indoor localization approaches, fingerprinting based method is more applicable and easier to integrate into most of today’s commodity Wi-Fi devices such as mobile phones and IOT devices which require low cost and computation burden. However, most fingerprinting systems intrinsically depend on fixed channel propagation environment and thus suffers huge reconstruction cost and high localization error when environment changes. In this paper, we propose a novel transformable fingerprinting localization method based on deep metric learning approaches. Our fingerprinting reconstruction method only requires some fresh measurements of CSI (Channel State Information) on a few reference points (RPs) with all the outdated CSI fingerprinting. Extensive system level simulations on Quadriga show that an average of 0.2m error reduction is achieved when our reconstruction method is applied.

1. Introduction
Indoor Wi-Fi localization has been an active research interest due to the development of Wi-Fi communication technology and tremendous deployment of the Wi-Fi network [1]. In industry, indoor localization systems (IPSs) is a key module which enables most of the localization-based services such as robotics navigation, multi-agent collaboration and home automation [2-4]. In general, Wi-Fi based indoor localization systems either utilizes channel estimation method or adopt fingerprinting matching approach. In channel estimation, information from physical layer such as angle of arrival (AoA) [5] and time of flight (ToF) [6] are extracted from propagation signal using channel parameters measurement or estimation approaches [7]. Terminal’s position is then determined by geometric computation such as trilateration. However, these methods heavily depend on the high precision of the device, such as wider communication bandwidth, various geometric antenna arrays and antenna polarization technology [8-10], which is generally not applicable with commodity Wi-Fi devices due to the hardware limitation. Another alternative for indoor Wi-Fi localization is the position fingerprinting matching method which achieves a better performance in common hardware condition.

In general, fingerprinting method is composed of offline construction stage and online inference stage. In offline stage, statistical or deterministic wireless feature is measured and extracted from the wireless propagation channels on a collection of geometric reference points (RPs) to build a radio map. This radio map is stored in database for further matching purpose. In online stage, feature is extracted on unknown position and certain matching algorithm is performed to search for the best-fitted candidates in radio map, then a proper position estimation is returned.

Received signal Strength (RSS) is commonly used in fingerprinting-based localization system [11], which is easy to acquire in most Wi-Fi devices, in spite of which, it can suffer performance...
degradation due to its sensitivity to the time varying channel and multipath effects. Unlike RSS’s weak indication of the propagation power, channel state information (CSI) from physical layer can precisely characterize the channel response on each of the transmitter-receiver pair in the level of multiple subcarriers in multiple-input, multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) communication system. Therefore, CSI can capture more abundant location-specific information such as multipath fading and small scale fading and is more suitable for fingerprinting [12].

The main defect about the fingerprinting method is the large amount of labour and time consumed in the offline construction stage. Besides, the reconstruction process must be performed repeatedly when the propagation environment alternation happens, such as the move or malfunction of the access point (APs). To cope with such issues, some researcher proposes to transform the outdated and updated fingerprinting into a common vector space using analytic method based on distance constraint [13]. However, this method introduces an inequality constraint which is not easy to determine in variable indoor environment.

Metric learning is a branch of machine learning which perform transformation on the input space into a space with desired characteristic such as aggregation and dispersion. Besides, fingerprinting reconstruction aims to find a space which keeps the main structure as outdated fingerprints where closer positions have more similar fingerprints. Therefore, our insight is that metric learning approach is suitable for the metric-constraint fingerprinting transformation tasks. Moreover, to capture the characteristic from the huge amount of data transmission in wireless channel, deep learning based metric learning approach implies a promising performance [14].

In this paper, we propose a novel transformable fingerprinting positioning method based on deep metric learning approaches. Our fingerprinting reconstruction method only requires some fresh measurements of channel state information (CSI) on a few reference points (RPs) with outdated CSI fingerprinting, which can reduce the labour and time cost in massive remeasurement and is vital for the fingerprinting based localization systems.

2. CSI Basis

Basically, in a MIMO-OFDM system, CSI is an \( N_{rx} \times N_{tx} \times N_{sc} \) Complex value matrix \( H \), where \( N_{rx}, N_{tx}, N_{sc} \) is the number of receive antennas, transmit antennas and OFDM subcarriers respectively. CSI on the \( i_{rx} \) receive antenna is of the following form:

\[
H_{i_{rx}} = \begin{bmatrix}
h_{1,1} & h_{1,2} & \cdots & h_{1,N_{sc}} \\
h_{2,1} & h_{2,2} & \cdots & h_{2,N_{sc}} \\
\vdots & \vdots & \ddots & \vdots \\
h_{N_{rx},1} & h_{N_{rx},2} & \cdots & h_{N_{rx},N_{sc}}
\end{bmatrix}
\]  

(1)

In an environment with \( L \) paths,

\[
h_{i,k} = \sum_{l=1}^{L} \alpha_l \exp[j(\phi - \frac{2\pi d \cos \theta_l + c \tau_l}{\lambda_k})]
\]

(2)

is the CSI corresponding to the \( z^{th} \) transmit antenna and \( k^{th} \) subcarrier, which is the aggregation of all \( L \) paths. For \( l^{th} \) path, \( \alpha_l \) is the complex attenuation. \( \phi \) is the initial phase of the electromagnetic wave when sent. \( d \) is the spacing between consecutive transmit antennas. \( c \) is the speed of light. \( \theta_l \) is related to the angle of departure (AoD) of the \( l^{th} \) path and \( \tau_l \) is related to its delay in second. \( \lambda_k \) is the wave length for \( k^{th} \) subcarrier.

From equation (2), channel state information (CSI) in MIMO-OFDM system can comprehensively depict the channel from the perspectives of frequency domain and spatial domain with multiple frequencies and transceiver antennas, which enable CSI to capture abundant information. CSI is the sum of the multiple paths transmitted signal propagates along, so for a specific location, CSI is distinct. Besides, CSI is mainly determined by the signal propagation geometric environment which is stable.
when the transceivers are established, and close locations tend to have similar CSIs. As a result, CSI is a better alternative for fingerprinting based indoor localization system.

3. Deep Metric Learning for Fingerprint Reconstruction

Fingerprint database needs to be reconstructed when signal propagation environment changes, where in most case the APs alter. Let \( x_o \in \mathbb{C}^N \) denote the outdated CSI fingerprint for a certain reference point’s location \( y \) and \( x_u \in \mathbb{C}^N \) is the correspondent updated CSI fingerprint when AP alters. Therefore, joint distribution \( P(x, z) \) alters where \( z \) is the location. Our insight is that for a fixed location the geometric environment is stable, so outdated fingerprint \( x_o \) still contains information about new distribution. Thus, reconstruction can be done based on the outdated fingerprint with a few fresh fingerprint measurements of the new channel environment. Formally, task of fingerprint reconstruction is to find a mapping \( f(\cdot) \in \mathbb{C}^N \rightarrow \mathbb{C}^M \) which minimize the distance in mapped space

\[
\min_f ||f(x_o) - f(x_u)||
\]

(3)

where \( x_o \) and \( x_u \) are chosen from a few reference points only. Solving equation (3) resulting in a new representation for outdated fingerprint and can be extended to all reference points. However, some properties for fingerprint need to be maintained in the mapped space. As aforementioned, close points own similar CSI fingerprint so CSI on several close reference points can act as prior information for the inference of target’s location. To make sure the transformed fingerprint maintains the same properties as CSI, we propose that \( f(\cdot) \) should satisfy:

\[
\min_f ||f(x_o) - f(x_i^o)|| - ||f(x_o) - f(x_i^f)||
\]

(4)

where \( f(x_o) \) denotes a specific location’s new fingerprint, \( f(x_i^o) \) is the new fingerprint of the location that geometrically close to \( x_i \) and \( f(x_i^f) \) denotes the new fingerprint of the location that geometrically far away from \( x_i \). Equation (4) tends to pull away the distance for farther point-pair relative to the closer point-pair in mapped space.

In our design, we propose to interpret \( f(\cdot) \) by deep neural network (DNN) mainly because DNN has been proved to be effective in data mining and pattern recognition tasks due to its ability in nonlinearity fitting. In our reconstruction tasks, not only fingerprint data’s form is of great complexity but the mapping \( f(\cdot) \) is a nonlinear transformation for fingerprint. So DNN is a best fit for our task. In the field of deep metric learning, triplet learning has been a state-of-the-art work and is fit for our \((x_o, x_i, x_i^f)\) triplet design. Formally, triplet learning is to train a neural network supervised by minimizing the triplet loss [14]:

\[
L_{\text{triplet}} = \max \left( ||f(x_o) - f(x_p)|| - ||f(x_o) - f(x_n)|| + \alpha, 0 \right)
\]

(5)

where \( x_o \) is the sample from anchor class, \( x_p \) is the positive sample chosen from the same class as \( x_o \) and \( x_n \) is the negative sample from a different class as \( x_o \). \( \alpha \) is a positive parameter denote the distance margin between \((x_o, x_p)\) pair and \((x_o, x_n)\) pair.

We design a 3-layer DNN with Rectified Linear Unit (ReLU) function as activation function. During training, triplet selection mechanism is a key factor for guiding the training process. Selecting hardest positive and negative sample can introduce the network to output a same vector for all input samples. Instead, we adopt a semi-hard triplet selection method that every triplet should satisfy:

\[
||f(x_o) - f(x_p)|| < ||f(x_o) - f(x_n)|| + \alpha,
\]

\[
||f(x_o) - f(x_p)|| > ||f(x_o) - f(x_n)||,
\]

(6)

which will steadily perform loss reduction in training process.
4. Implementation Details and Evaluation

4.1. Fingerprint Setup and Deep Neural Network Design
As aforementioned, we adopt the raw CSI as our initial fingerprinting before transformation. We regard the complex CSI values as two real number, so the input data of the DNN is a real value vector with the shape:

\[ N_{rx} \times N_{tx} \times N_{sc} \times 2. \]

To perform triplet metric learning, corresponding label is the second input to DNN. For a certain sample \( x_i \), a label \( y_i \) is taken from:

\[ Y = \{ y \in \mathbb{Z} | 1 \leq y \leq N_{cls} \}, \]

where \( N_{cls} \) is the number of categories in the dataset. The input layer is followed by two consecutive fully-connected layer with ReLU as activation function and dropout modules are added to prevent overfitting. Finally, an output layer without activation which produces mapped fingerprint embedding is added. There are two loss layers in our design called triplet loss and contrastive loss respectively. Each loss is performed on the same embedding output. For the triplet loss layer, an \( l_2 \)-normalized version for each embedding is explicitly calculated. For a fixed training batch size \( N_{batch} \), a pairwise cosine distance matrix of \( N_{batch} \times N_{batch} \) is calculated, based on which the semi-hard triplet selection mechanism is performed to train the network. In the contrastive loss layer, we also choose the cosine distance as a metric to punish the dissimilarity between outdated fingerprint embedding and updated fingerprint embedding on the same location as stated in equation (3). The proposed DNN is visualized in figure 1. For training the DNN, we setup the parameters as illustrated in table 1.

![Figure 1. Basic structure of proposed DNN](image)

In our settings, \( N_{rx} = 1, N_{tx} = 3, N_{sc} = 4, N_o = 10, N_{cls} = 9 \). Semi-hard triplet loss and contrastive loss are applied to train the network.

4.2. Simulation Setup
Our experiments are performed on the Quasi deterministic Radio Channel Generator (QuaDRiGa) simulation software [15]. We setup a channel based on the standardized channel models for LoS and NLoS indoor scenarios. The setup is visualized in figure 2. To simulate Wi-Fi channels, we setup the center frequency as 2.4GHz and signal bandwidth as 10MHz. 4 subcarrier frequencies are sampled uniformly from signal bandwidth. There are 3 APs in our setup, each with linear antenna array.
composed of 3 omnidirectional antennas. The terminals with 1 omnidirectional antenna are set inside a 20m × 20m region for DNN training, radio map construction and online inference test accordingly.

For training the transformation DNN, in the 20m×20m indoor region, we setup 9 averagely distributed 2×2 areas for simulation of the 9 reference points from different class where both outdated and updated CSI is collected. For each 2m×2m area, $N_s$ randomly distributed snapshots of the CSI are collected for training. For the radio map construction, we uniformly choose $N_r$ points which cover the whole region, where only one snapshot of outdated fingerprint is collected for the purpose of fingerprinting database construction and further evaluation. For online inference test, we randomly generate $N_t$ test points in the 20m×20m region.

### 4.3. Pipeline

#### 4.3.1. Initial radio map construction

As aforementioned, CSI is collected in each of the reference points to build a radio map. $N_s$ CSI samples on the 9 reference points are collected for preparation of the training process.

| Training Parameters                  | Setup   |
|--------------------------------------|---------|
| Initial learning rate                | 0.002   |
| Batch size for triplet loss training | 100     |
| Batch size for contrastive loss training | 40   |
| Number of reference points class     | 9       |
| Number of iterations                 | 150000  |
| Optimizer                            | Adam    |

![Figure 2. Simulation region.](image)
4.3.2. Training the DNN
When AP alters, say move toward the northwest for about \( d \) meters, as visualized in figure 2, the signal propagation environment changes to an unknown state. All fingerprints in the radio map become outdated. \( N_s \) CSI samples from the 9 reference points are then collected as a few fresh measurements of the new distribution. Each sample is assumed to be correspondent to the \( N_s \) samples collected before AP alters. To minimize equation (5) in DNN training, \( N_s \) samples which are collected before AP alters are used. Simultaneously, training the DNN to minimize the \( N_s \) correspondent samples pairs collected on outdated and updated channel are performed, namely minimizing equation (3).

4.3.3. Fingerprint reconstruction
When DNN training is done, a mapping \( \hat{f}(\cdot) \) forms. We use this mapping \( \hat{f}(\cdot) \) to transform the \( N_r \) outdated radio map and \( N_t \) updated test points fingerprint \( x \) into new radio map and new fingerprints \( \hat{x} \) respectively. For online position inference, a 3-KNN matching algorithm is performed on the new space to estimate the location of each test point.

4.4. Evaluation
In our settings, we set \( N_s, N_r, N_t, d \) as 278, 361, 200 and 2 respectively. Figure 3 shows the localization error our method produces compared to the altered and unaltered raw CSI fingerprint. Each evaluation is performed on the same set of the 200 test points. Localization results show that our method is effective in reconstruction of the CSI when propagation environment altered. Besides, compared to the raw CSI fingerprint method evaluated in unaltered case, our method on altered case even wins. This is because our proposed metric learning algorithm trained on the 9 representative locations effectively aggregate the near RPs and increase the distance between RPs that are far away, thus can enhance the discriminability compared to the raw CSI. The average localization error is reduced by 0.2m compared to the raw CSI fingerprint when environment changes. In spite of the fact that we only use naive CSI values as fingerprinting and 3-KNN as online inference method which, our average localization error achieve 1.04m, which will be further decreased when more precise techniques are applied such as CSI calibration.

![Figure 3. Localization error for altered fingerprint (the blue line), unaltered fingerprint (the yellow line) and the altered but transformed fingerprint (the orange line) which is our proposed method.](image)

5. Conclusion
In this paper, we propose a novel transformable fingerprinting localization method based on deep metric learning approaches. Our fingerprinting reconstruction method is efficient and applicable because it only requires some fresh measurements of channel state information (CSI) on a few
reference points (RPs) when propagation environment altered. Extensive system level simulations on Quadriga show that an average of 0.2m localization error reduction is achieved when our reconstruction method is applied.

6. Acknowledgments
This work is supported by the National Natural Science Foundation of China (No.61631013), Beijing National Research Center for Information Science and Technology, National Major Project (NO. 2018ZX03001006-003).

7. References
[1] Zafari F, Gkelias Z and Leung K 2019 A survey of indoor localization systems and technologies IEEE Communications Surveys & Tutorials, vol 21, no 3, pp 2568-2599
[2] Piaggio M, Sgorbissa A and Zaccaria R 2001 Autonomous navigation and localization in service mobile robotics Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, vol 4, pp 2024-2029
[3] Luo Z, Zhang Q, Ma Y, Singh M and Adib F 2019 3D backscatter localization for fine-Grained robotics 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19), pp 765-782
[4] Bianchi V, Ciampolini P and Munari I 2019 RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes IEEE Transactions on Instrumentation and Measurement, vol 68, no 2, pp 566-575
[5] Kotaru M, Joshi K, Bharadia D and Katti S 2015 SpotFi: Decimeter level localization using WiFi Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, pp 269-282
[6] Gong W and Liu J 2018 SiFi: Pushing the limit of time-based WiFi localization using a single commodity access point Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol 2, no 1
[7] Schmidt R 1986 Multiple emitter location and signal parameter estimation IEEE Transactions on Antennas and Propagation, vol 34, no 3, pp 276-280
[8] Vasisht D, Kumar S and Katabi D 2016 Decimeter-level localization with a single WiFi access point 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), pp 165-178
[9] Xiong J and Jamieson K 2013 ArrayTrack: A fine-grained indoor location system Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), pp 71-84
[10] Giorgetti G, Cidronali A, Gupta S and Manes G 2009 Single-anchor indoor localization using a switched-beam antenna IEEE Communications Letters, vol 13, no 1, pp 58-60
[11] Sadowski S and Spachos P 2018 RSSI-based indoor localization with the internet of things IEEE Access, vol 6, pp 3049-3061
[12] Li H, Zeng X, Li Y, Zhou S and Wang J 2019 Convolutional neural networks based indoor Wi-Fi localization with a novel kind of CSI images China Communications, vol 16, no 9, pp 250-260
[13] Gao Z, Gao Y, Wang S, Li D, Xu Y and Jiang H 2019 CRISLoc: Reconstructable CSI fingerprinting for indoor smartphone localization ArXiv e-prints
[14] Schroff F, Kalenichenko D and Philbin J 2015 FaceNet: A unified embedding for face recognition and clustering IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 815-823
[15] Jaeckel S, Raschkowski L, Börner K and Thiele L 2014 QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials IEEE Transactions on Antennas and Propagation, vol 62, no 6, pp 3242-3256