IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2019, accepted February 24, 2019, date of publication March 6, 2019, date of current version March 25, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2903273*

# Transferred Knowledge Aided Positioning via Global and Local Structural Consistency Constraints

**XIANSHENG GUO[1,2], (Member, IEEE), LEI WANG[1], (Student Member, IEEE), LIN LI[1], (Student Member, IEEE), AND NIRWAN ANSARI[3], (Fellow, IEEE)**

[1]Department of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Wuhu Overseas Students Pioneer Park, Wuhu 241006, China
[3]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: Xiansheng Guo (xsguo@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61371184, Grant 61671137, Grant 61771114, and Grant 61771316, and in part by the Application Foundation Projects of Science and Technology Department in Sichuan Province Grant 2018JY0242 and Grant 2018JY0218.

**ABSTRACT** The fluctuation of received signal strength (RSS) induced by changing environment is the main hindrance from practical applications of the fingerprint-based indoor positioning methods. Transfer learning can mitigate the fluctuation of RSS by transferring knowledge from a source domain (off-line RSS data) to a target domain (online RSS data). However, the existing transfer learning approaches do not fully take into account the full constraints in Global and LOcal Structural conSistency (GLOSS), thus resulting in insufficient knowledge transfer. To overcome the above drawback, we propose a Transferred knowlEdge-Aided POsiTioning (TEAPOT) approach via GLOSS constraints in this paper. TEAPOT imposes the global structural consistency by minimizing the differences between the marginal and conditional distributions of the source and target domains and maximizing the samples variance in a latent subspace. Simultaneously, it also imposes the local structural consistency by minimizing within class variance and maximizing between class variance to retain the source discriminative information and preserving the local neighborhood relationship by using manifold regularization. Furthermore, a nonlinear TEAPOT is derived to improve the ability of TEAPOT to alleviate the limitation of linear projection. Compared with the existing methods, two of our proposed TEAPOT approaches via GLOSS constraints show higher accuracy and better ability in handling out-of-sample generalization. The experimental results verify that the proposed method significantly outperforms the existing methods.

**INDEX TERMS** Indoor positioning, Wi-Fi fingerprint, transfer learning, global and local structural consistency.

## I. INTRODUCTION

Location based services (LBSs) rely on user locations to deliver context aware functionalities. The position information is vital to Internet of Things (IoT), which involves the extension of the Internet to small and low-cost "things" in actualizing smart environments in order to provide new services to the users [1]–[3]. Global positioning system (GPS) has been widely deployed in the outdoor environment, but it performs poorly in the complex indoor environment owing to non-line-of-sight transmission between the satellite and the receiving devices [4]. Hence, indoor positioning research has attracted much attention in the past two decades [5], [6].

Different wireless indoor positioning systems based on WiFi, UWB, Bluetooth, etc. have been developed recently. Among them, the received signal strength (RSS)-based WiFi indoor positioning is the most viable one because WiFi has been widely deployed. Generally, the RSS-based WiFi positioning algorithms can be classified into geometric-based [6]–[9] and fingerprint-based algorithms [2], [10]–[13]. The former measures some geometric parameters and then calculates the user's position

The associate editor coordinating the review of this manuscript and approving it for publication was Kai Yang.

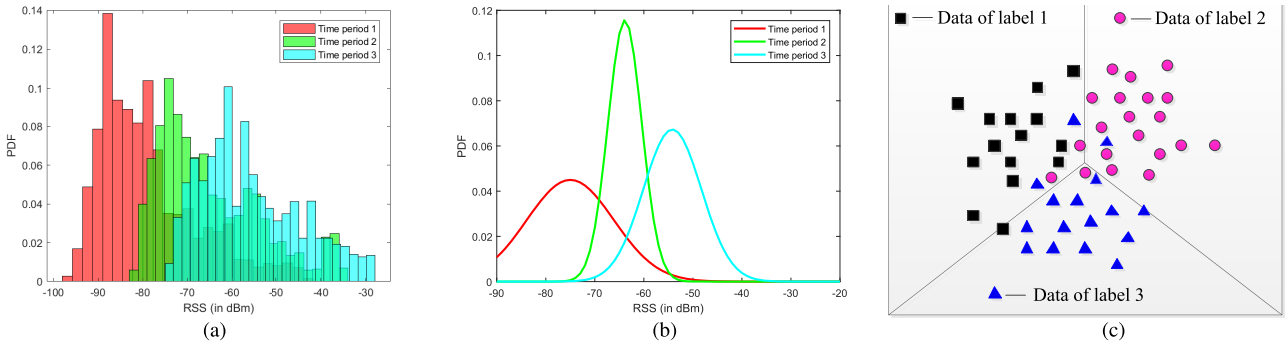


**FIGURE 1.** The study of RSS variations. (a) The distributions in different time periods at all grid points. (b) The distributions in different time periods at the same grid point. (c) The feature space of RSS fingerprints at 3 adjacent grid points.

by geometric principles. However, due to factors such as multipath fading and non-line-of-sight effects, it is hard to determine an accurate propagation model in a complex indoor scenario. Therefore, geometric-based indoor positioning algorithms lack accuracy. On the contrary, the fingerprint-based algorithms do not need to calculate accurate geometric parameters and can obtain accurate positioning without the knowledge of the layout of the indoor environment and the locations of the access points (APs) in advance.

The RSS fingerprint-based WiFi positioning approach includes two phases: offline and online phases. In the offline phase, the main work is to construct a fingerprint database. First, the RSS values of all APs are collected by a smart phone at each grid point. Then, the RSS values and the coordinate of the grid point are stored to construct the fingerprint database. In the online phase, the RSS samples collected from a target with unknown location are matched with the fingerprint database to yield a location estimate. Some existing matching approaches, such as RSS direct matching and machine learning, assume that the offline data and the online testing data satisfy the same distribution. However, RSS data usually fluctuate due to changing environment, such as short-term interference (e.g., walking by a pedestrian) and long-term interference (e.g., changes of temperature and humidity) [14]. The RSS value measured in real time may significantly deviate from the RSS value previously stored in the fingerprint database, thus resulting in the severe degeneration of the positioning performance of RSS based methods.

The fluctuation of RSS will pose statistical variations. From the global perspective, marginal and conditional distributions are different, as shown in Figs. 1(a) and 1(b). The former shows the marginal distribution of RSS values at all grid points, and the latter shows the conditional distribution of RSS values at the same grid point, both in different time periods. They show that the RSS distribution varies greatly in complex indoor environments as time changes. From the local perspective, the RSS values received at the same grid point in the same time period may exist certain discrepancy, i.e., the RSS data of different labels show poor differentiations, as depicted in Fig.1 (c). Therefore, the conventional methods show poor adaptivity to changing environment.

Transfer learning methods [15]–[18] have been proven to be an efficient strategy to mitigate the impact of RSS variations in lieu of calibration [11]. However, the existing methods show poor performance in severe RSS variations induced by changing environment. For example, Pan *et al.* [16] and Zou *et al.* [18] reduced the difference of the marginal distributions between the source and target data by constraining the global structural consistency to obtain the marginal distribution adaption. However, they neglect the local structure consistency because the RSS values in local adjacent grid points may exhibit certain ambiguity, as depicted in Fig. 1(c). Manifold regularization [15] can learn a low-dimensional embedding by constructing a weighted graph to preserve the local neighborhood relationship, but performs poorly in reducing RSS variations when the overall distribution may change, as depicted in Fig. 1(a).

To overcome the above drawback, we propose a Transferred knowlEdge Aided POsiTioning (TEAPOT) approach by leveraging the Global and LOcal Structural conSistency (GLOSS). TEAPOT learns a projection in mapping the source and target domains into a latent feature subspace by constraining both global and local structural consistency. Firstly, from the perspective of constraining global structural consistency, the difference of the distributions between the source and target domains in the latent space is minimized; meanwhile, the variance of all the RSS data in the latent space is maximized, thus making the overall distribution between the two domains as close as possible and all the RSS data in the latent space as separable as possible. Furthermore, from the perspective of constraining local structural consistency, the within class variance is minimized and the between class variance is maximized, thus making the projection of samples within the same label as close as possible and the projection of samples with different labels as far as possible, respectively; the manifold regularization is adopted to preserve the neighboring relationship between samples. Then, in the latent subspace, conventional matching or machine learning algorithm can be used to yield the label (location) estimates. The key insight of our work is that we propose a highly efficient knowledge transfer method by constraining the consistency of global and local structures to improve the

positioning accuracy in presence of RSS variation without any labeled data in the target domain. The proposed method is validated through extensive experimental results.

Our main contributions are summarized below.

1) Aiming to minimize the difference in statistic characteristics induced by the fluctuation of RSS, we derive a (linear) TEAPOT method by taking into account of the constraints in both global and local structural consistency (GLOSS) constraints. GLOSS can improve the efficiency of knowledge transfer as compared with the existing transfer learning methods.
2) A nonlinear version of TEAPOT is also derived to improve the ability of the linear TEAPOT to alleviate the limitation of linear projection.
3) Aided by the knowledge transferred by using the GLOSS constraint, the proposed TEAPOT algorithms show higher accuracy and better ability in handling out-of-sample generalization without any fingerprint /hardware calibrations.
4) We validate the performance of our proposed TEAPOT algorithms from the transductive setting and out-of-sample setting in dealing with the long term variation RSS data offered by [19].

The rest of this paper is organized as follows. In Section II, we first introduce the related works about transfer learning. In Section III, we formulate the problem of transfer learning based indoor positioning. Then, we detail our proposed two TEAPOT algorithms in Section IV. In Section V, we conduct two experiments to verify the effectiveness of the proposed methods. Finally, we draw a conclusion in Section VI.

## II. RELATED WORK

The traditional WiFi fingerprint based positioning methods can be normally classified into two categories: the deterministic-based approaches and the probabilistic-based approaches. The former are range or proximity based using the characteristics of the RSS fingerprint, to find the distance from the mobile device to APs. The widely known RADAR system [20] is addressed as one of most the representative deterministic approaches for the WiFi fingerprint indoor positioning by selecting the K-nearest neighbor fingerprint samples to estimate the location label. Other advanced deterministic-based approaches include random forests [21], neural network [22], and adaboost [13]. The basic idea of probabilistic-based approaches is to pre-store the RSS distribution with respect to each hearable AP into a fingerprint database, and then use it to infer the location estimation. The Horus system is a representative probabilistic method, which estimates the user's location by the Bayesian inference [23]. Note that both kinds of methods assume that the offline RSS samples and the online testing RSS samples follow the same distribution. However, in a complex indoor environment, the offline training data and the online testing data may likely violate the above assumption due to changing environment.

Transfer learning has been studied extensively in recent years and applied successfully in many fields [16], [24]–[27]. In indoor positioning, Sun *et al.* [15] proposed a LuMA algorithm based on manifold alignment to solve the RSS variations caused by time changing. Pan *et al.* [16] proposed a transfer learning method called transfer component analysis (TCA), which can reduce the difference between domain distributions and preserve important properties of domains by maximizing the data variance after projecting data onto the latent space. Furthermore, a semi-supervised TCA (SSTCA) method was proposed to preserve the local neighbor relationship and enhance the dependence between RSS data and their labels by using manifold regularization and HSCI criteria. Transfer kernel learning (TKL) [18] learns a domain-invariant kernel by directly matching the source and target distributions in the reproduced kernel Hilbert space, and the resultant kernel can be used as input for the support vector regression (SVR) training. In image processing, joint distribution analysis (JDA) [26] improves upon TCA by jointly adapting both the marginal and conditional distributions in a principled dimensionality reduction. Balanced distribution adaptation (BDA) [28] adaptively balances the importance of the marginal and conditional distributions. Joint geometrical and statistical alignment (JGSA) [27] learns different projections for the source and target domain data so that the geometrical and distribution shift are reduced simultaneously in the projected subspace.

However, most of the above methods cannot fully trade off the constraints in global and local structural consistency to reduce the impact of the RSS variations, thus resulting in insufficient knowledge transfer. Comparatively, our proposed TEAPOT algorithms can fully take into account of GLOSS with regard to RSS variations and do not require extra labeled data in the target domain, thus achieving out-of-sample generalization.

## III. PROBLEM FORMULATION

Assume that there are $m$ APs deployed in an indoor scenario covered by $C$ grid points. We can construct the RSS fingerprints $\boldsymbol{r}_i = \left[r_i^1, r_i^2, \cdots, r_i^m\right]^T \in \mathcal{R}^m$ for $i = 1, 2, \cdots, n_s$, from $m$ APs at all grid points by using a smart phone. $n_s$ is the number of samples, and the corresponding location labels are represented as $c_i \in \{1, 2, \cdots, C\}$. The offline RSS fingerprint and its location label are combined into the labeled data in the source domain $\mathcal{D}_s$, i.e., $\mathcal{D}_s = \{(\boldsymbol{r}_i, c_i)\}_{i=1}^{n_s}$. Assume that we can obtain some unlabeled data at other time using a smart phone with unknown location $\boldsymbol{z}$ to form the target domain, denoted as $\mathcal{D}_t = \left\{\tilde{\boldsymbol{r}}_j\right\}_{j=1}^{n_t}$.

Mathematically, we can express the source domain as $\mathcal{D}_s = \{\boldsymbol{X}_s, \boldsymbol{c}_s\}$, where $\boldsymbol{X}_s = \left[\boldsymbol{r}_1^T, \boldsymbol{r}_2^T, \cdots, \boldsymbol{r}_{n_s}^T\right]^T$ is the fingerprint set, and $\boldsymbol{c}_s = \left[c_1, c_2, \cdots, c_{n_s}\right]$ is the corresponding label set. The unlabeled target domain $\mathcal{D}_t$ is denoted as $\mathcal{D}_t = \{\boldsymbol{X}_t\}$, where $\boldsymbol{X}_t = \left[\tilde{\boldsymbol{r}}_1^T, \tilde{\boldsymbol{r}}_2^T, \cdots, \tilde{\boldsymbol{r}}_{n_t}^T\right]^T$. We assume the feature space and label spaces in source and target domains are the same, but the marginal and conditional distributions in both domains

**TABLE 1. List of Notations.**

| Notation | Definition |
|---|---|
| $\boldsymbol{X}_s$ | the source data |
| $\boldsymbol{X}_t$ | the target data |
| $\boldsymbol{X}$ | $\boldsymbol{X} = [\boldsymbol{X}_s, \boldsymbol{X}_t]$, the source and target data, denoted as $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n_s+n_t}]$ |
| $n_s$ | The number of source samples |
| $n_t$ | The number of target samples |
| $n$ | $n = n_s + n_t$, the total number of samples in source and target domains |
| $m$ | The number of APs |
| $\boldsymbol{c}_s$ | $\boldsymbol{c}_s = [c_1, c_2, \cdots, c_{n_s}]^T$, the location label vector of source domain |
| $C$ | The number of grid points |
| $k$ | The dimension of embedding data |
| $\boldsymbol{A}$ | Projection matrix $\boldsymbol{A} \in \mathcal{R}^{m\times k}$ in the linear problem, $\boldsymbol{A} \in \mathcal{R}^{n\times k}$ in the nonlinear problem |
| $\boldsymbol{Z}$ | The embedded data $\boldsymbol{Z} = \boldsymbol{A}^T\boldsymbol{X}$ in the linear problem, $\boldsymbol{Z} = \boldsymbol{A}^T\boldsymbol{K}$ in the nonlinear problem |

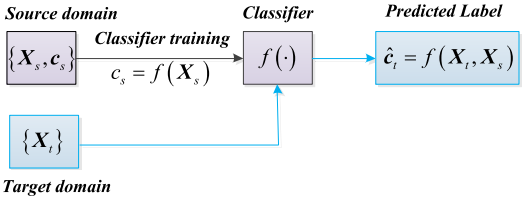


**FIGURE 2. The overview of the traditional learning based positioning framework.**

are different, i.e., $P(\boldsymbol{r}_i) \neq P(\tilde{\boldsymbol{r}}_j)$, $P(c_i|\boldsymbol{r}_i) \neq P(\tilde{c}_j|\tilde{\boldsymbol{r}}_j)$, where $\tilde{c}_j$ is the label of the target domain. To ease the understanding, we summarize some important notations in Table 1.

As we all known, the traditional learning based positioning framework directly trains a classifier $\boldsymbol{c}_s = f(\boldsymbol{X}_s)$, such as K-Nearest Neighbor (KNN) and support vector machine (SVM) in the source domain $\mathcal{D}_s$ and then does the localization $\hat{\boldsymbol{c}}_t = f(\boldsymbol{X}_t, \boldsymbol{X}_s)$ based on the trained classifier by directly using the target domain data $\mathcal{D}_t$ without considering the distribution difference between the source and target domains, as depicted in Fig. 2. Comparatively, our proposed TEAPOT positioning framework aims to determine the labels of RSS vectors in $\mathcal{D}_t$ by using the transferred knowledge from $\mathcal{D}_s$. Specifically, we first learn a projection matrix $\boldsymbol{A}$ to map $\boldsymbol{X}_s$ and $\boldsymbol{X}_t$ into the latent feature subspace (i.e., the embedding data $\boldsymbol{A}^T\boldsymbol{X}_s$, $\boldsymbol{A}^T\boldsymbol{X}_t$) by constraining the GLOSS. Then, in the latent subspace, we train a standard classifier $\boldsymbol{c}_s = f(\boldsymbol{A}^T\boldsymbol{X}_s)$, and then the location label can be estimated by $\hat{\boldsymbol{c}}_t = f(\boldsymbol{A}^T\boldsymbol{X}_t, \boldsymbol{A}^T\boldsymbol{X}_s)$, as shown in Fig. 3.

Note that our proposed TEAPOT positioning framework will show better performance in complex environments because the transferred knowledge from the source to target domain by using the GLOSS constraint can mitigate the impact of RSS variations to some extent, thus yielding higher accuracy in positioning performance.

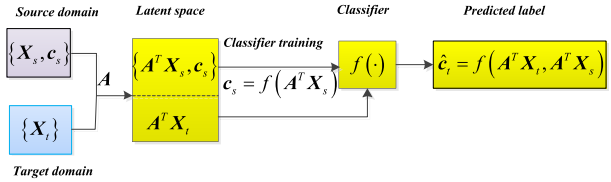


**FIGURE 3. The overview of TEAPOT positioning framework.**

## IV. PROPOSED ALGORITHM

In this section, we will detail our proposed (linear) TEAPOT first, and then derive a nonlinear TEAPOT to alleviate the limitation of linear projection in complex environments. Fig. 4 shows our proposed TEAPOT positioning framework, which includes two main parts: global structural consistency and local structural consistency. From the global perspective, reducing the difference in distributions between domains can reduce the distance between the sample means of the source and target data, and variance maximization can make all the RSS data in the latent space as separable as possible. From the local perspective, minimizing the within class variance and maximizing the between class variance can make the projection of samples within the same label as close as possible, and the projection of samples with different labels as far as possible, respectively, and manifold regularization can preserve the neighbor relationship before and after transfer. Hence, it is a good strategy to jointly constrain the global and local structural consistency in reducing the RSS variations in complex environments.

### A. GLOBAL STRUCTURAL CONSISTENCY CONSTRAINT

As depicted in Fig. 4, the distribution difference between domains and the separability of projected RSS samples are the two key problems in transferring knowledge from

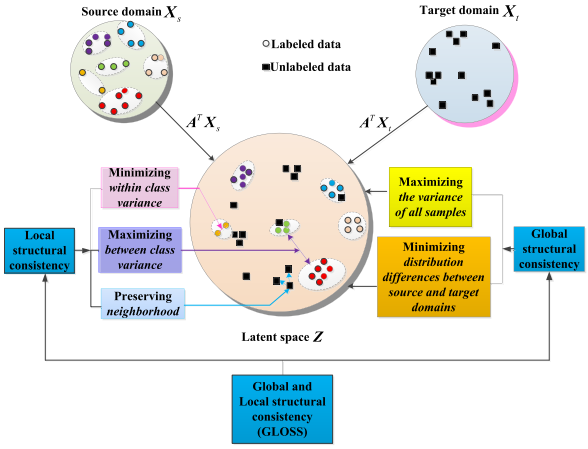


**FIGURE 4. The overview of the GLOSS constraints in our proposed TEAPOT positioning framework.**

the source to target domain. Therefore, we must consider minimizing the distribution difference and maximizing the variance of all RSS data to impose the global structural consistency.

The difference of distributions between domains normally involves the marginal and conditional distributions [16], [26]–[28]. Owing to changes in the environment, the RSS value received in most of the location areas may change from the RSS values previously stored in the fingerprint database. The marginal distribution is different from previous marginal distribution, i.e., $P(\boldsymbol{r}_i) \neq P(\tilde{\boldsymbol{r}}_j)$ (e.g. Gaussian distribution with different means). In addition, the distribution of RSS values corresponding to the same grid point has also changed, thus resulting in the conditional distribution difference, i.e. $P(c_i|\boldsymbol{r}_i) \neq P(\tilde{c}_j|\tilde{\boldsymbol{r}}_j)$. Therefore, we need to reduce the marginal and conditional distribution differences between domains simultaneously.

#### 1) MARGINAL DISTRIBUTION ADAPTATION

To measure the distance between two distributions, there exist many criteria, such as Kullback-Leibler (KL) divergence. However, most of them are parametric or need an intermediate density estimate. Maximum mean discrepancy (MMD) criteria [25] is a nonparametric and effective distribution distance estimator, which is based on the reproduced kernel hilbert space (RKHS) to estimate the distance between distributions. Therefore, we can use the MMD criterion to measure the distribution distance between domains. We first consider reducing the marginal distribution difference $P(\boldsymbol{r}_i)$ and $P(\tilde{\boldsymbol{r}}_j)$ between domains by the MMD criterion as follows

$$\left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \boldsymbol{A}^T \boldsymbol{x}_i - \frac{1}{n_t} \sum_{j=n_s+1}^{n} \boldsymbol{A}^T \boldsymbol{x}_j \right\|^2 = Tr\left( \boldsymbol{A}^T \boldsymbol{X} \boldsymbol{M}_0 \boldsymbol{X}^T \boldsymbol{A} \right), \quad (1)$$

where $Tr(\cdot)$ is the trace operator and $\boldsymbol{M}_0$ is the MMD matrix given by

$$(\boldsymbol{M}_0)_{ij} = \begin{cases} \frac{1}{n_s^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_s \\ \frac{1}{n_t^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_t \\ \frac{-1}{n_s n_t}, & otherwise \end{cases}. \quad (2)$$

The derivation process of Eq. (1) has been proven in [16] and [28]. By minimizing Eq. (1), we can reduce the marginal distribution difference between domains to mitigate the impact of the RSS variation to some extent.

#### 2) CONDITIONAL DISTRIBUTION ADAPTATION

Conditional distribution difference is an another key metric to measure the distribution difference [26]. Owing to the lack of true location labels of the samples in the target domain, we cannot directly match the conditional distribution $P(c_i|\boldsymbol{r}_i)$, $P(\tilde{c}_j|\tilde{\boldsymbol{r}}_j)$ between domains. Long *et al.* [26] minimized the class conditional distribution discrepancy between $P(\boldsymbol{r}_i|c_i)$ and $P(\tilde{\boldsymbol{r}}_j|\tilde{c}_j)$ according to the sufficient statistics when sample sizes are large. They calculated $P(\tilde{\boldsymbol{r}}_j|\tilde{c}_j)$ by iteratively updating the pseudo labels of the target data predicted by a classifier trained on the source domain. Here, we also use this strategy to minimize the conditional distribution difference as

$$\left\| \frac{1}{n_s^{(c)}} \sum_{\boldsymbol{x}_i \in \mathcal{D}_s^{(c)}} \boldsymbol{A}^T \boldsymbol{x}_i - \frac{1}{n_t^{(c)}} \sum_{\boldsymbol{x}_j \in \mathcal{D}_t^{(c)}} \boldsymbol{A}^T \boldsymbol{x}_j \right\|^2 = Tr\left( \boldsymbol{A}^T \boldsymbol{X} \boldsymbol{M}_c \boldsymbol{X}^T \boldsymbol{A} \right), \quad (3)$$

where $c \in \{1, 2, \cdots, C\}$ is the location label. $\mathcal{D}_s^{(c)}$ and $\mathcal{D}_t^{(c)}$ denote the samples belonging to the grid point $c$ in the source and target domains, respectively. $n_s^{(c)} = \left|\mathcal{D}_s^{(c)}\right|$ and $n_t^{(c)} = \left|\mathcal{D}_t^{(c)}\right|$ denote the numbers of samples belonging to $\mathcal{D}_s^{(c)}$ and $\mathcal{D}_t^{(c)}$, respectively. The MMD matrix $\boldsymbol{M}_c$ is given as follows [26]

$$(\boldsymbol{M}_c)_{ij} = \begin{cases} \frac{1}{n_s^{(c)} n_s^{(c)}}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{n_t^{(c)} n_t^{(c)}}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{n_s^{(c)} n_t^{(c)}}, & \begin{cases} \boldsymbol{x}_i \in \mathcal{D}_s^{(c)}, \boldsymbol{x}_j \in \mathcal{D}_t^{(c)} \\ \boldsymbol{x}_i \in \mathcal{D}_t^{(c)}, \boldsymbol{x}_j \in \mathcal{D}_s^{(c)} \end{cases} \\ 0, & otherwise \end{cases}. \quad (4)$$

By minimizing Eq. (3), the conditional distribution difference between domains can be reduced under the new representation $\boldsymbol{Z} = \boldsymbol{A}^T \boldsymbol{X}$. Furthermore, we can trade off the importance of the marginal and conditional distributions based on application requirements. Specifically, we use a balance factor $\mu$ to trade off the importance of the two distributions:

$$\mu \sum_{c=1}^{C} \left\| \frac{1}{n_s^{(c)}} \sum_{\boldsymbol{x}_i \in \mathcal{D}_s^{(c)}} \boldsymbol{A}^T \boldsymbol{x}_i - \frac{1}{n_t^{(c)}} \sum_{\boldsymbol{x}_j \in \mathcal{D}_t^{(c)}} \boldsymbol{A}^T \boldsymbol{x}_j \right\|^2 + (1-\mu) \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \boldsymbol{A}^T \boldsymbol{x}_i - \frac{1}{n_t} \sum_{j=n_s+1}^{n} \boldsymbol{A}^T \boldsymbol{x}_j \right\|^2 = Tr\left( \boldsymbol{A}^T \boldsymbol{X} \boldsymbol{M} \boldsymbol{X}^T \boldsymbol{A} \right), \quad (5)$$

where $\mu \in [0, 1]$, and the matrix $\boldsymbol{M}$ can be expressed as

$$\boldsymbol{M} = (1-\mu)\boldsymbol{M}_0 + \mu \sum_{c=1}^{C} \boldsymbol{M}_c. \quad (6)$$

Hence, by minimizing Eq. (5), the marginal and conditional distribution differences can be reduced simultaneously, where the importance of the two distributions is treated differently.

The RSS variations can be alleviated by reducing the marginal and conditional distribution differences to some extent. To further impose the global structural consistency, we also need to maximize the variance of all RSS data, thus making the projected RSS data as separable as possible.

#### 3) VARIANCE MAXIMIZATION

Given the projection matrix $\boldsymbol{A}$, the covariance matrix of the projected samples can be expressed as $\boldsymbol{A}^T \boldsymbol{X} \boldsymbol{H} \boldsymbol{X}^T \boldsymbol{A}$, where

$\boldsymbol{H} = \boldsymbol{I}_n - (1/n)\,\mathbf{1}\mathbf{1}^T$ is the centering matrix, $\boldsymbol{I}_n \in \mathcal{R}^{n\times n}$ is the identity matrix, and $\mathbf{1} \in \mathcal{R}^n$ is the column vector with all ones [16]. Hence, the variance maximization can be achieved as follows

$$\max_{\boldsymbol{A}} Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T\boldsymbol{A}\right). \tag{7}$$

Therefore, by minimizing Eq. (5) and maximizing Eq. (7), we can reduce the distribution difference between domains and preserve the separability of projected RSS samples to impose the global structural consistency.

### B. LOCAL STRUCTURAL CONSISTENCY

Unlike the global structural consistency which considers the overall statistics of the data on the latent space, local structural consistency imposes the relationship between the data and their labels as well as geometrical relationship when transferring. Among the local structural consistency, source discriminative information and neighborhood preservations are the two efficient strategies in transfer learning. The former can be realized by minimizing the within class variance and maximizing the between class variance, and the latter can be obtained by manifold regularization.

#### 1) SOURCE DISCRIMINATIVE INFORMATION PRESERVATION

In source discriminative information preservation, the projections of samples within the same label should be as close as possible; on the contrary, the projections of samples with different labels should be as far as possible. Minimizing the within class scatter and maximizing the between class scatter are the two efficient strategies in scatter component analysis to measure the discriminative information [29]. Therefore, we preserve the source discriminative information by using

$$\min_{\boldsymbol{A}} Tr\left(\boldsymbol{A}^T\boldsymbol{S}_w\boldsymbol{A}\right) \tag{8}$$

and

$$\max_{\boldsymbol{A}} Tr\left(\boldsymbol{A}^T\boldsymbol{S}_b\boldsymbol{A}\right), \tag{9}$$

where $\boldsymbol{S}_w$ is the within class scatter matrix and $\boldsymbol{S}_b$ is the between class scatter matrix of the source domain data, which are computed as

$$\boldsymbol{S}_w = \sum_{c=1}^{C} \boldsymbol{X}_s^{(c)}\boldsymbol{H}_s^{(c)}\left(\boldsymbol{X}_s^{(c)}\right)^T \tag{10}$$

and

$$\boldsymbol{S}_b = \sum_{c=1}^{C} n_s^{(c)}\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)^T, \tag{11}$$

respectively, where $\boldsymbol{X}_s^{(c)} \in \mathcal{R}^{m\times n_s^{(c)}}$ is the set of source samples belonging to the label $c$. $\boldsymbol{H}_s^{(c)} = \boldsymbol{I}_s^{(c)} - \frac{1}{n_s^{(c)}}\mathbf{1}_s^{(c)}\left(\mathbf{1}_s^{(c)}\right)^T$ is the centering matrix, where $\boldsymbol{I}_s^{(c)} \in \mathcal{R}^{n_s^{(c)}\times n_s^{(c)}}$ is the identity matrix, and $\mathbf{1}_s^{(c)} \in \mathcal{R}^{n_s^{(c)}}$ is the column vector with all ones. In addition, $\boldsymbol{m}_s^{(c)} = \frac{1}{n_s^{(c)}}\sum_{i=1}^{n_s^{(c)}}\boldsymbol{x}_i^{(c)}$ and $\bar{\boldsymbol{m}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{x}_i$.

#### 2) NEIGHBORHOOD PRESERVATION

Neighborhood preservation should also be considered in local structural consistency due to the local manifold structure of RSS [15]. The manifold assumption is an effective way to preserve the neighborhood relationship between samples by assuming that the data located on a manifold structure and the adjacent samples share similar labels [30]. Hence, we can use the manifold regularization to preserve the local neighborhood relationship.

Specifically, we can construct an undirected connection graph $G = (V, E)$, where $V$ is the set of vertices corresponding to all labeled samples $\boldsymbol{X}_s$ and unlabeled samples $\boldsymbol{X}_t$; $E$ is the set of edges whose element $e_{ij}$ is the edge connecting vertices $i$ and $j$. We can calculate the $(i, j)$th element $\boldsymbol{W}_{ij}$ in the affinity matrix $\boldsymbol{W} \in \mathcal{R}^{n\times n}$ as

$$\boldsymbol{W}_{ij} = \begin{cases} exp\left(-\left\|\boldsymbol{x_i} - \boldsymbol{x_j}\right\|^2/2\sigma^2\right), & (i,j) \in \mathcal{N} \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

where $(i, j) \in \mathcal{N}$ means that $\boldsymbol{x}_i$ is one of the K nearest neighbors of $\boldsymbol{x}_j$, or vice versa. $\boldsymbol{W}_{ij}$ can reveal the similarity between two samples. Afterwards, we can get the graph Laplacian matrix as

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}, \tag{13}$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal element $\boldsymbol{D}_{ii} = \sum_{j=1}^{n}\boldsymbol{W}_{ij}$. In order to ensure that neighbor relationship can be preserved before and after transferring, we can minimize the following equation:

$$\sum_{(i,j)\in\mathcal{N}} \boldsymbol{W}_{ij}\left\|\left[\boldsymbol{A}^T\boldsymbol{X}\right]_i - \left[\boldsymbol{A}^T\boldsymbol{X}\right]_j\right\|^2 = 2Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T\boldsymbol{A}\right), \tag{14}$$

where $\left[\boldsymbol{A}^T\boldsymbol{X}\right]_i$ and $\left[\boldsymbol{A}^T\boldsymbol{X}\right]_j$ are the $i$th and $j$th columns of $\left[\boldsymbol{A}^T\boldsymbol{X}\right]$, respectively, i.e., the embedded coordinates in the latent subspace.

In summary, the local structural consistency can be realized by optimizing Eqs. (8), (9), and (14) from source discriminative information and neighborhood preservation perspectives.

### C. THE JOINT OPTIMIZATION IN TEAPOT

By jointly optimizing the objective functions of GLOSS constraints, we can formulate the overall optimization function of our proposed linear TEAPOT method as Eq. (15), at the bottom of the next page, in which $\beta, \lambda$ are trade-off parameters to balance the importance of each quantity, and $\alpha$ is the regularization parameter. The term $Tr\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ controls the complexity of $\boldsymbol{A}$. In Eq. (15), the numerator consists of two maximization components and the denominator includes three minimization components. Note that Eq. (15) is invariant to different scales of $\boldsymbol{A}$, and so it can be rewritten as

$$\begin{aligned} &\max_{\boldsymbol{A}} Tr\left[\boldsymbol{A}^T\left(\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \beta\boldsymbol{S}_b\right)\boldsymbol{A}\right] \\ &\text{s.t. } Tr\left[\boldsymbol{A}^T\left(\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T + \lambda\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T + \beta\boldsymbol{S}_w + \alpha\boldsymbol{I}\right)\boldsymbol{A}\right] = 1. \end{aligned} \tag{16}$$

Denote $\boldsymbol{\Phi} = \text{diag}(\phi_1, \phi_2, \cdots, \phi_k) \in \mathcal{R}^{k\times k}$ as Lagrange multipliers, then Lagrange function of Eq. (16) is

$$\mathcal{L} = Tr\left[\boldsymbol{A}^T\left(\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \beta\boldsymbol{S}_b\right)\boldsymbol{A}\right] + Tr\left\{\left\{\left[\boldsymbol{A}^T\left(\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T + \lambda\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T + \beta\boldsymbol{S}_w + \alpha\boldsymbol{I}\right)\boldsymbol{A}\right] - \boldsymbol{I}\right\}\boldsymbol{\Phi}\right\}. \tag{17}$$

Setting $\frac{\partial\mathcal{L}}{\partial\boldsymbol{A}} = \boldsymbol{0}$, we obtain

$$\left(\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T + \beta\boldsymbol{S}_b\right)\boldsymbol{A} = \left(\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T + \lambda\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T + \beta\boldsymbol{S}_w + \alpha\boldsymbol{I}\right)\boldsymbol{A}\boldsymbol{\Phi}. \tag{18}$$

Thus, by solving the generalized eigenvalue decomposition problem of Eq. (18), we can get the optimal transformation matrix $\boldsymbol{A}$ by selecting the $k$ leading eigenvectors. After having obtained $\boldsymbol{A}$, we can obtain the updated label by using

$$\hat{\boldsymbol{c}}_t = f\left(\boldsymbol{A}^T\boldsymbol{X}_t, \boldsymbol{A}^T\boldsymbol{X}_s\right), \tag{19}$$

where $f(\cdot)$ is the classifier trained on $\left(\boldsymbol{A}^T\boldsymbol{X}_s, \boldsymbol{c}_s\right)$. The location estimate can be given by

$$\hat{\boldsymbol{z}}_t = g\left(\hat{\boldsymbol{c}}_t\right), \tag{20}$$

where $g(\cdot)$ is a mapping function which transforms a label to a 2-D coordinate of the target. The linear TEAPOT algorithm will converge until the error between two adjacent location estimates is smaller than a tolerance parameter $\epsilon$ (such as $10^{-3}$), i.e.,

$$\frac{1}{n_t}\left\|\hat{\boldsymbol{z}}_t(l) - \hat{\boldsymbol{z}}_t(l-1)\right\|_F < \epsilon, \tag{21}$$

where $\hat{\boldsymbol{z}}_t(l)$ is the location estimate at the $l$th iteration, $\|\cdot\|_F$ is the Frobenius norm.

The linear TEAPOT is summarized in Algorithm 1. The derivation of linear TEAPOT is given in Appendix VI.

**Algorithm 1 Linear TEAPOT**

**Input:** 1) Source and target domain data: $\boldsymbol{X} = [\boldsymbol{X}_s, \boldsymbol{X}_t]$;
2) Source label vector: $\boldsymbol{c}_s$;
3) Parameters: $k, \mu, \beta, \lambda, \alpha, \epsilon$;
**Output:** Estimates of the location: $\hat{\boldsymbol{z}}_t$.

$l \leftarrow 0$
Initialize the location of target $\hat{\boldsymbol{z}}_t(l) = \boldsymbol{0}$.
Construct $\boldsymbol{M}_0, \boldsymbol{S}_w, \boldsymbol{S}_b, \boldsymbol{L}$ by Eqs. (2), (10), (11), (13).
Set $\{\boldsymbol{M}_c := \boldsymbol{0}\}_{c=1}^{C}$.
**repeat**
  $l \leftarrow l + 1$
  Compute the matrix $\boldsymbol{M}$ using Eq. (6).
  Obtain $\boldsymbol{A}$ by calculating the $k$ leading eigenvectors of Eq. (18).
  Train a classifier $f(\cdot)$ on $\left\{\boldsymbol{A}^T\boldsymbol{X}_s, \boldsymbol{c}_s\right\}$.
  Update the pseudo label $\hat{\boldsymbol{c}}_t(l) = f\left(\boldsymbol{A}^T\boldsymbol{X}_t, \boldsymbol{A}^T\boldsymbol{X}_s\right)$.
  Update MMD matrices $\{\boldsymbol{M}_c\}_{c=1}^{C}$ by Eq. (4).
  Obtain the location estimate $\hat{\boldsymbol{z}}_t(l) = g\left(\hat{\boldsymbol{c}}_t(l)\right)$
**until** $\frac{1}{n_t}\left\|\hat{\boldsymbol{z}}_t(l) - \hat{\boldsymbol{z}}_t(l-1)\right\|_F < \epsilon$
**return** $\hat{\boldsymbol{z}}_t(l)$

### D. KERNELIZATION

In complex environments, linear projection may not be sufficient to find a subspace to reduce RSS variations. To address this problem, we can extend the linear TEAPOT to a nonlinear version to alleviate the limitation of linear projection.

Specifically, given a kernel mapping $\psi : \boldsymbol{x} \mapsto \psi(\boldsymbol{x})$ and a kernel matrix $\boldsymbol{K} = \psi(\boldsymbol{X})^T\psi(\boldsymbol{X}) \in \mathcal{R}^{n\times n}$, such as radial basis function (RBF) kernel [31], we utilize the Representer theorem $\boldsymbol{P} = \psi(\boldsymbol{X})\boldsymbol{A}$ [32] to formulate our method, where $\psi(\boldsymbol{X}) = [\psi(\boldsymbol{x}_1), \psi(\boldsymbol{x}_2), \cdots, \psi(\boldsymbol{x}_n)]$. Hence, the objective function becomes Eq. (24), at the bottom of this page, where all $\boldsymbol{X}$ are replaced by $\psi(\boldsymbol{X})$ and $\boldsymbol{A}$ are replaced by $\boldsymbol{P} = \psi(\boldsymbol{X})\boldsymbol{A}$. The within class scatter matrix and the between class scatter matrix are constructed

$$\boldsymbol{S}_w = \sum_{c=1}^{C}\boldsymbol{K}_s^{(c)}\boldsymbol{H}_s^{(c)}\left(\boldsymbol{K}_s^{(c)}\right)^T \tag{22}$$

and

$$\boldsymbol{S}_b = \sum_{c=1}^{C} n_s^{(c)}\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)^T, \tag{23}$$

respectively, where $\boldsymbol{K}_s^{(c)} = \psi(\boldsymbol{X})^T\psi\left(\boldsymbol{X}_s^{(c)}\right)$, $\boldsymbol{m}_s^{(c)} = \frac{1}{n_s^{(c)}}\sum_{i=1}^{n_s^{(c)}}\boldsymbol{k}_i^{(c)}$ and $\bar{\boldsymbol{m}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{k}_i$, with $\boldsymbol{k}_i = \psi(\boldsymbol{X})^T\psi(\boldsymbol{x}_i)$ and $\boldsymbol{k}_i^{(c)} = \psi(\boldsymbol{X})^T\psi\left(\boldsymbol{x}_i^{(c)}\right)$. In addition, $\boldsymbol{K}_s = \psi(\boldsymbol{X})^T\psi(\boldsymbol{X}_s)$ and $\boldsymbol{K}_t = \psi(\boldsymbol{X})^T\psi(\boldsymbol{X}_t)$. $\boldsymbol{A} \in \mathcal{R}^{n\times k}$ is the projection matrix for the nonlinear TEAPOT method. Note that the dimension of $\boldsymbol{A}$ in the nonlinear TEAPOT is different from $\boldsymbol{A} \in \mathcal{R}^{m\times k}$ in the linear TEAPOT because the number of samples $n$ is not equal to the number of APs $m$ in general. After having obtained the kernelized objective function, we can easily solve it like the aforementioned linear TEAPOT problem, i.e., $\boldsymbol{A}$ can be obtained by solving the following generalized eigenvalue decomposition problem

$$\max_{\boldsymbol{A}} \frac{Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T\boldsymbol{A} + \beta\boldsymbol{A}^T\boldsymbol{S}_b\boldsymbol{A}\right)}{Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T\boldsymbol{A} + \lambda\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T\boldsymbol{A} + \beta\boldsymbol{A}^T\boldsymbol{S}_w\boldsymbol{A} + \alpha\boldsymbol{A}^T\boldsymbol{A}\right)}, \tag{15}$$

$$\max_{\boldsymbol{A}} \frac{Tr\left(\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}^T\boldsymbol{A} + \beta\boldsymbol{A}^T\boldsymbol{S}_b\boldsymbol{A}\right)}{Tr\left(\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{M}\boldsymbol{K}^T\boldsymbol{A} + \lambda\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}^T\boldsymbol{A} + \beta\boldsymbol{A}^T\boldsymbol{S}_w\boldsymbol{A} + \alpha\boldsymbol{A}^T\boldsymbol{A}\right)}, \tag{24}$$

**Algorithm 2 Nonlinear TEAPOT**

**Input:** 1) Source and target domain data: $\boldsymbol{X} = [\boldsymbol{X}_s, \boldsymbol{X}_t]$;
2) Source label vector: $\boldsymbol{c}_s$;
3) Parameters: $k, \mu, \beta, \lambda, \alpha, \epsilon$;
4) A kernel mapping $\psi$;

**Output:** Estimates of the location: $\hat{\boldsymbol{z}}_t$.

$l \leftarrow 0$
Initialize the location of target $\hat{\boldsymbol{z}}_t(l) = \boldsymbol{0}$.
Construct $\boldsymbol{M}_0, \boldsymbol{S}_w, \boldsymbol{S}_b, \boldsymbol{L}$ by Eqs. (2), (22), (23), (13).
Compute the kernel matrix $\boldsymbol{K}_s = \psi(\boldsymbol{X})^T \psi(\boldsymbol{X}_s)$.
Compute the kernel matrix $\boldsymbol{K}_t = \psi(\boldsymbol{X})^T \psi(\boldsymbol{X}_t)$.
Set $\{\boldsymbol{M}_c := \boldsymbol{0}\}_{c=1}^{C}$.
**repeat**
  $l \leftarrow l + 1$
  Compute the matrix $\boldsymbol{M}$ using Eq. (6).
  Obtain $\boldsymbol{A}$ by calculating the $k$ leading eigenvectors of Eq. (**??**).
  Train a classifier $f(\cdot)$ on $\{\boldsymbol{A}^T\boldsymbol{K}_s, \boldsymbol{c}_s\}$.
  Update the pseudo label $\hat{\boldsymbol{c}}_t(l) = f\left(\boldsymbol{A}^T\boldsymbol{K}_t, \boldsymbol{A}^T\boldsymbol{K}_s\right)$.
  Update MMD matrices $\{\boldsymbol{M}_c\}_{c=1}^{C}$ by Eq. (4).
  Obtain the location estimate $\hat{\boldsymbol{z}}_t(l) = g\left(\hat{\boldsymbol{c}}_t(l)\right)$
**until** $\frac{1}{n_t}\left\|\hat{\boldsymbol{z}}_t(l) - \hat{\boldsymbol{z}}_t(l-1)\right\|_F < \epsilon$
**return** $\hat{\boldsymbol{z}}_t(l)$

The nonlinear TEAPOT is summarized in Algorithm 2. The derivation of the nonlinear TEAPOT is given in Appendix VI-B.

## V. EXPERIMENTS

### A. EXPERIMENTAL SETUP

To evaluate the accuracy of our method, we use the average error distance (AED) as a metric, which is defined as

$$AED = \frac{1}{N}\sum_{j=1}^{N}\left\|g\left(\hat{c}_j\right) - g\left(c_j\right)\right\|, \tag{25}$$

where $g\left(\hat{c}_j\right)$ represents the $j$-th 2-D coordinate prediction and $g\left(c_j\right)$ is the $j$-th true 2-D coordinate of the RSS sample. $N$ is the number of experimental trials.

The experiment was conducted on the second floor of the library building from Universitat Jaume I, in Spain [19], as shown in Figs. 5 and 6. The measurements were collected at 230 grid points by one person using one Android smartphone during 15 months. The area is about 308.4 $m^2$ and totally 448 APs are detected during the experiment. Specifically, we treat the first month's database as the source domain, including 1380 samples obtained by moving average of the original database. For the remaining 14 months, we treat them as 14 different target domains, respectively, each containing 3696 testing samples. Denote the data collected in the first month and the remaining 14 months by $\mathcal{D}_s$ and $\mathcal{D}_{t_i}$, $i = 2, 3, \cdots, 15$, respectively. Moreover, we randomly split $\mathcal{D}_{t_i}$ into $\mathcal{D}_{t_i}^u$ (the label information is removed in training) and $\mathcal{D}_{t_i}^o$, $\left|\mathcal{D}_{t_i}^u\right| = 1500$, $\left|\mathcal{D}_{t_i}^o\right| = 2196$. In the transductive evaluation setting, we learn a model from $\mathcal{D}_s$ and $\mathcal{D}_{t_i}^u$, and then evaluate it on $\mathcal{D}_{t_i}^u$. In the out-of-sample evaluation setting, we learn a model from $\mathcal{D}_s$ and $\mathcal{D}_{t_i}^u$, and then evaluate it on $\mathcal{D}_{t_i}^o$.


**FIGURE 5.** The interior environment of the library building [19].

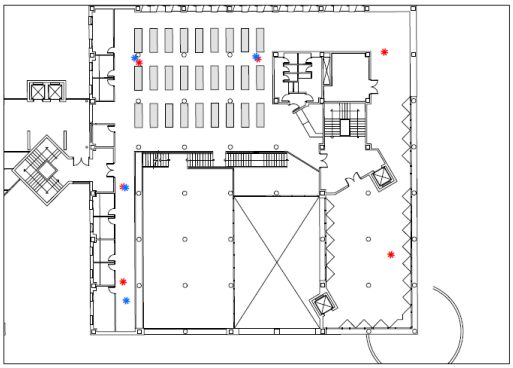

**FIGURE 6.** The layout of the library building [19].

We compare the performance of our TEAPOT methods with the following several state-of-the-art baseline methods

- K-nearest neighbor classifier (KNN)
- Principal component analysis (PCA)
- Kernel principal component analysis (KPCA)
- Transfer component analysis (TCA) [16]
- Semi-supervised transfer component analysis (SSTCA) [16]
- Joint distribution adaptation (JDA) [26]
- Joint geometrical and statistical alignment (JGSA) [27]

Among them, KNN, PCA and KPCA are traditional learning methods, while TCA, SSTCA, JDA and JGSA are state-of-the-art transfer learning approaches. Considering that training a classifier at each iteration is a time-consuming work, so we use a KNN classifier to test the performance of these methods.

The TEAPOT methods involve five parameters: $\mu, \alpha, \beta, \lambda, k$. In the following sections, we conducted an empirical analysis of parameter sensitivity to verify that TEAPOT can achieve stable performance over a wide range of parameter values. To tune parameters of TEAPOT, we sample 50 labeled data from the source domain as a validation set. Note that we evaluate the effect of the validation set on the projected subspace to select the parameters. In the comparison study,

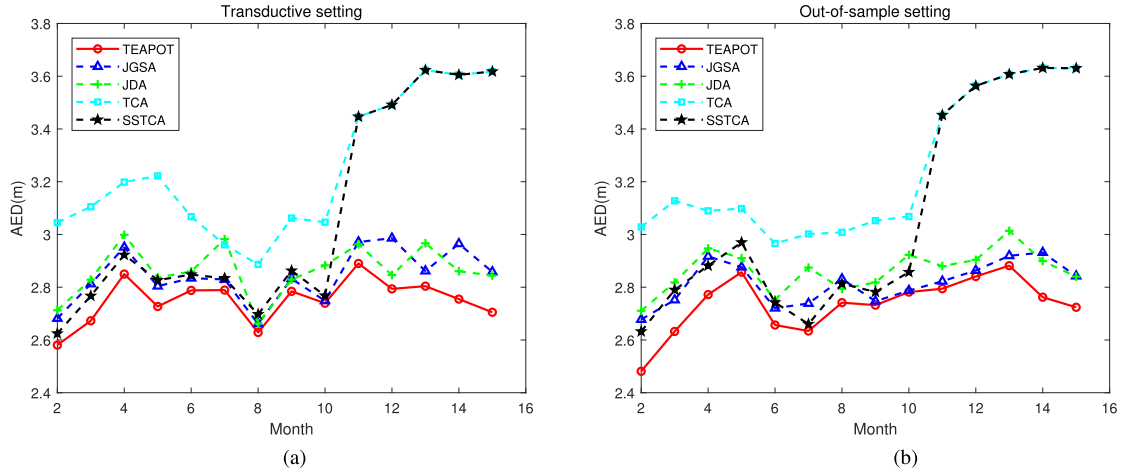


**FIGURE 7.** Comparisons with other state-of-the-art transfer learning approaches. (a) The AEDs of different positioning methods in the transductive setting. (b) The AEDs of different positioning methods in the out-of-sample setting.

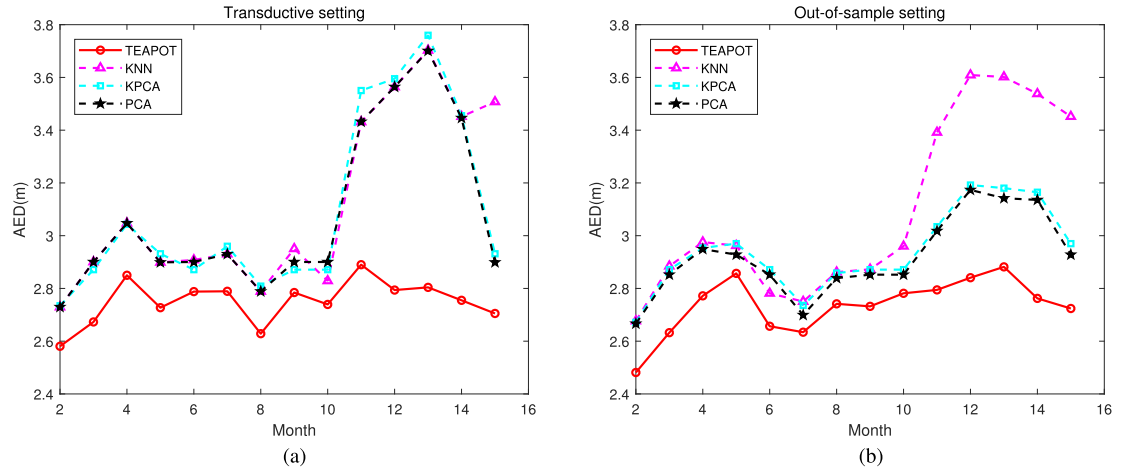


**FIGURE 8.** Comparisons with traditional learning approaches. (a) The AEDs of different positioning methods in the transductive setting. (b) The AEDs of different positioning methods in the out-of-sample setting.

we set $k = 200$, $\mu = 0.8$, $\alpha = 0.2$, $\beta = 10$ and $\lambda = 0.6$, and $\epsilon = 10^{-2}$. For the dimension of subspace $k$ and the parameter $K$ in the KNN classifier, we set $k = 200$ and $K = 10$ for all approaches. The parameter values used for the other approaches are set according to the parameter values in the corresponding papers. In addition, we adopt nonlinear TCA, SSTCA, JDA, JGSA, and TEAPOT by using the RBF kernel. Specifically, we summarize the parameter values for the comparative methods in Table 2. Note that only $k$ and $K$ are common parameters for all methods and the other parameters are specific to each method.

**TABLE 2.** The parameter values for the comparative methods.

| | |
|---|---|
| PCA,KPCA | $k = 200, K = 10$ |
| TCA/SSTCA | $k = 200, K = 10, \mu = 0.2, \gamma = 0.6, \lambda = 0.8$ |
| JDA | $k = 200, K = 10, \mu = 0.2$ |
| JGSA | $k = 200, K = 10, \mu = 1, \beta = 0.1, \alpha = 1$ |

### B. POSITIONING PERFORMANCE

The positioning results on 14 target domains are shown in Figs. 7 and 8. Fig. 7 (a) and (b) show the AEDs of TEAPOT and state-of-the-art transfer learning approaches in the transductive setting and the out-of-sample setting, respectively. We can observe that TEAPOT outperforms other methods in both settings. Note that JGSA and JDA, which impose partial constraints in GLOSS, fail to consider the problem of RSS variations in practice when transferring knowledge. Therefore, the average AEDs of JDA on the 14 target domains are obviously higher than TEAPOT. TCA and SSTCA perform better on the 2nd-10th target domains, but worse on the remaining target domains. A major limitation of TCA and SSTCA is that the difference between the conditional

**TABLE 3.** The percentages (%) of *AED* < 3*m* in the transductive seeting.

| Target domain | Transfer learning | | | | | Traditional learning | | |
|---|---|---|---|---|---|---|---|---|
| | TEAPOT | JGSA | JDA | TCA | SSTCA | KNN | PCA | KPCA |
| $\mathcal{D}_{t_2}$ | 61.93 | 58.66 | 58.93 | 55.60 | 60.13 | 58.33 | 58.33 | 58.00 |
| $\mathcal{D}_{t_3}$ | 60.53 | 57.00 | 56.67 | 54.07 | 57.73 | 56.47 | 56.47 | 57.73 |
| $\mathcal{D}_{t_4}$ | 56.20 | 53.93 | 53.80 | 51.27 | 54.67 | 52.67 | 52.67 | 52.87 |
| $\mathcal{D}_{t_5}$ | 61.27 | 57.80 | 57.47 | 52.40 | 58.27 | 57.40 | 57.40 | 57.07 |
| $\mathcal{D}_{t_6}$ | 58.94 | 57.60 | 58.33 | 53.67 | 57.73 | 56.73 | 56.73 | 56.20 |
| $\mathcal{D}_{t_7}$ | 59.20 | 57.20 | 54.13 | 55.20 | 57.53 | 55.73 | 55.73 | 55.07 |
| $\mathcal{D}_{t_8}$ | 61.67 | 60.86 | 60.67 | 57.93 | 60.00 | 58.07 | 58.07 | 57.53 |
| $\mathcal{D}_{t_9}$ | 58.67 | 57.67 | 58.46 | 54.87 | 57.87 | 54.73 | 54.73 | 54.87 |
| $\mathcal{D}_{t_{10}}$ | 59.87 | 58.80 | 56.60 | 55.93 | 58.67 | 57.60 | 57.60 | 56.60 |
| $\mathcal{D}_{t_{11}}$ | 57.40 | 55.80 | 55.00 | 49.80 | 49.13 | 48.00 | 48.00 | 46.13 |
| $\mathcal{D}_{t_{12}}$ | 56.87 | 54.80 | 55.07 | 46.40 | 46.87 | 46.07 | 46.07 | 45.67 |
| $\mathcal{D}_{t_{13}}$ | 57.60 | 55.27 | 56.20 | 46.80 | 46.80 | 43.13 | 43.13 | 42.13 |
| $\mathcal{D}_{t_{14}}$ | 58.60 | 54.60 | 57.67 | 45.13 | 45.20 | 46.67 | 46.73 | 47.00 |
| $\mathcal{D}_{t_{15}}$ | 58.73 | 57.33 | 57.27 | 43.53 | 43.13 | 44.93 | 44.87 | 44.20 |

**TABLE 4.** The percentages (%) of *AED* < 3m in the out-of-sample seeting.

| Target domain | Transfer learning | | | | | Traditional learning | | |
|---|---|---|---|---|---|---|---|---|
| | TEAPOT | JGSA | JDA | TCA | SSTCA | KNN | PCA | KPCA |
| $\mathcal{D}_{t_2}$ | 64.25 | 58.56 | 59.79 | 54.83 | 59.34 | 58.83 | 59.02 | 58.65 |
| $\mathcal{D}_{t_3}$ | 60.70 | 57.60 | 56.19 | 52.05 | 56.69 | 56.38 | 57.15 | 56.47 |
| $\mathcal{D}_{t_4}$ | 57.42 | 54.87 | 54.55 | 52.73 | 55.42 | 52.73 | 53.23 | 53.14 |
| $\mathcal{D}_{t_5}$ | 57.15 | 56.42 | 54.46 | 53.87 | 54.01 | 54.14 | 54.37 | 53.87 |
| $\mathcal{D}_{t_6}$ | 60.84 | 59.20 | 58.61 | 56.15 | 58.79 | 58.33 | 59.29 | 59.29 |
| $\mathcal{D}_{t_7}$ | 62.57 | 59.65 | 56.97 | 55.83 | 61.25 | 60.25 | 60.47 | 59.79 |
| $\mathcal{D}_{t_8}$ | 59.06 | 56.92 | 57.38 | 54.51 | 56.69 | 55.46 | 56.24 | 55.37 |
| $\mathcal{D}_{t_9}$ | 59.93 | 58.79 | 58.24 | 54.46 | 58.83 | 57.33 | 57.56 | 57.24 |
| $\mathcal{D}_{t_{10}}$ | 59.47 | 59.28 | 56.38 | 55.92 | 57.29 | 55.74 | 55.97 | 55.87 |
| $\mathcal{D}_{t_{11}}$ | 57.74 | 57.01 | 56.42 | 49.13 | 49.00 | 47.68 | 54.19 | 53.92 |
| $\mathcal{D}_{t_{12}}$ | 56.06 | 55.51 | 55.01 | 46.77 | 46.81 | 44.63 | 50.77 | 50.59 |
| $\mathcal{D}_{t_{13}}$ | 56.15 | 54.14 | 54.28 | 45.08 | 44.49 | 44.76 | 49.86 | 49.27 |
| $\mathcal{D}_{t_{14}}$ | 58.01 | 55.51 | 56.19 | 44.22 | 44.58 | 45.54 | 50.32 | 50.14 |
| $\mathcal{D}_{t_{15}}$ | 58.93 | 56.24 | 56.38 | 44.72 | 44.72 | 46.99 | 49.95 | 48.22 |

distributions is not constrained. The conditional distribution of RSS data may undergo great changes in the last few target domains. TEAPOT can overcome this limitation via the GLOSS constraint.

Figs. 8 (a) and 8 (b) show the AEDs of TEAPOT and traditional learning approaches in the transductive setting and the out-of-sample setting, respectively, showing that our TEAPOT outperforms the traditional learning approaches. The reduced dimensional space of PCA and KPCA may still contain noise, and the positioning results obtained by directly matching is not optimal. In addition, there are large distribution differences between source and target domains due to the fluctuation of RSS data. TEAPOT can reduce the fluctuation by matching the distributions.

Tables 3 and 4 show the percentages of AED less than 3m on the 14 target domains. We observe that TEAPOT is better than the other methods. Specifically, in the transductive setting, the average percentages of AED achieved by TEAPOT, JGSA, JDA, TCA, SSTCA, KNN, PCA and KPCA on the 14 target domains are 59.11%, 56.95%, 56.88%, 51.61%, 53.84%, 52.61%, 52.61%, and 52.22%, respectively. In the out-of-sample setting, the average percentages of AED achieved by TEAPOT, JGSA, JDA, TCA, SSTCA, KNN, PCA, and KPCA on the 14 target domains are 59.17%, 57.12%, 56.49%, 51.45%, 53.42%, 52.77%, 54.89%, and 54.42%, respectively. Figs. 7 and 8 and Tables 3 and 4 demonstrate that TEAPOT is more robust to changing environment than other methods because it can solve the fluctuation of RSS effectively.

### C. EFFECTIVENESS VERIFICATION

To demonstrate the effectiveness of the nonlinear TEAPOT in alleviating the limitation of linear projection in complex environments, we compare the performance of the linear TEAPOT with the nonlinear TEAPOT. Then, we further validate the effectiveness of our proposed TEAPOT methods by testing the constraints of global structural consistency and local structural consistency.

#### 1) EFFECTIVENESS OF THE NONLINEAR TEAPOT

We run the linear TEAPOT and the nonlinear TEAPOT using the RBF kernel on 14 target domains, respectively. From Figs. 9 (a) and 9 (b), we find that the nonlinear TEAPOT outperforms the linear TEAPOT in both settings. In complex indoor environments, there may not be a linear projection to obtain a latent subspace by constraining GLOSS. Nevertheless, we can extend the linear TEAPOT to the nonlinear one in a high dimensional space to effectively improve the projection.

#### 2) EFFECTIVENESS OF THE GLOBAL STRUCTURAL CONSISTENCY

We first investigate the effectiveness of the global structural consistency. Here, we run the linear TEAPOT by only imposing local structural consistency constraint, named as TEAPOT-NonGSC, on 14 target domains. Figs. 10 (a) and 10 (b) show the comparison between the linear TEAPOT and the linear TEAPOT-NonGSC on 14 target domains. From these results, we can observe that: 1) the performance of all the target domains are improved with global structural consistency, indicating that global structural consistency constraint is effective when transferring knowledge. 2) TEAPOT-NonGSC only constrains the local structural consistency by manifold regularization, and minimizing within class variance and maximizing between class variance, which slightly

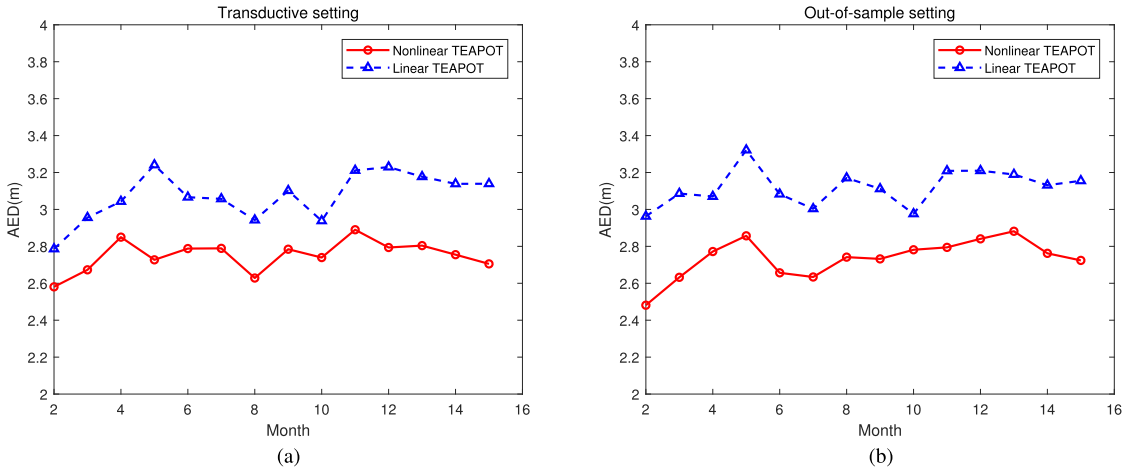


**FIGURE 9.** Effectiveness of the nonlinear TEAPOT. (a) The AEDs of linear TEAPOT and nonlinear TEAPOT in the transductive setting. (b) The AEDs of linear TEAPOT and nonlinear TEAPOT in the out-of-sample setting.

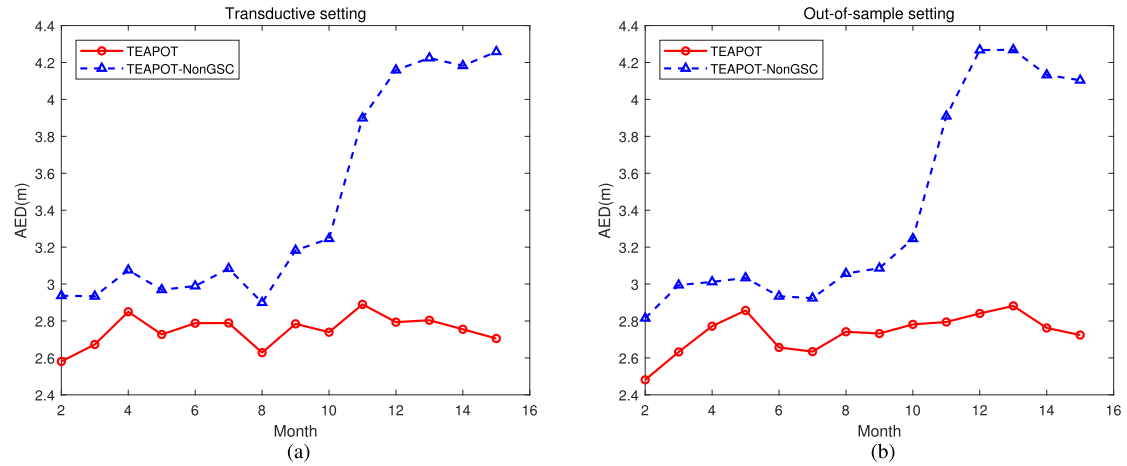


**FIGURE 10.** Effectiveness of the global structural consistency. (a) The AEDs of TEAPOT and TEAPOT-NonGSC in the transductive setting. (b) The AEDs of TEAPOT and TEAPOT-NonGSC in the out-of-sample setting.

reduces the fluctuation of RSS data, but cannot achieve satisfied positioning accuracy. 3) The conditional distribution of RSS data may undergo great changes in the last five target domains, resulting in a sharp increase in the positioning error of TEAPOT-NonGSC, while the using of global structural consistency constraint can yield better results.

#### 3) EFFECTIVENESS OF THE LOCAL STRUCTURAL CONSISTENCY

Similarly, we run the linear TEAPOT by only imposing global structural consistency constraint, called as TEAPOT-NonLSC, on 14 target domains. Figs. 11 (a) and 11 (b) show the AEDs of the linear TEAPOT and TEAPOT-NonLSC. Note that TEAPOT outperforms TEAPOT-NonLSC because TEAPOT-NonLSC ignores the local neighborhood relationship and the source discriminative information, i.e., ignoring the local structural consistency, thus resulting in insufficient knowledge transfer. While TEAPOT via GLOSS constraint can transfer knowledge with higher efficiency, and thus performs better than TEAPOT-NonLSC.

### D. PARAMETER SENSITIVITY

In this section, we test the performance of our methods with different parameters. For simplicity, we only list the results on the $\mathcal{D}_{t_2}, \mathcal{D}_{t_3}, \mathcal{D}_{t_4}$ target domains; similar trends can be found on other target domains.

Fig.12 (a) illustrates the AEDs versus different subspace dimension $k$. We observe that TEAPOT is robust when $k \in [200, 400]$. Fig.12 (b) shows the AEDs versus the balance factor $\mu$ on the target domains, from which we can observe that $\mu$ varies on different target domains, indicating that the marginal and conditional distributions have different importance. Fig.12 (c) depicts the AEDs versus different $\alpha$. When $\alpha \rightarrow 0$, the optimization problem is ill-defined.

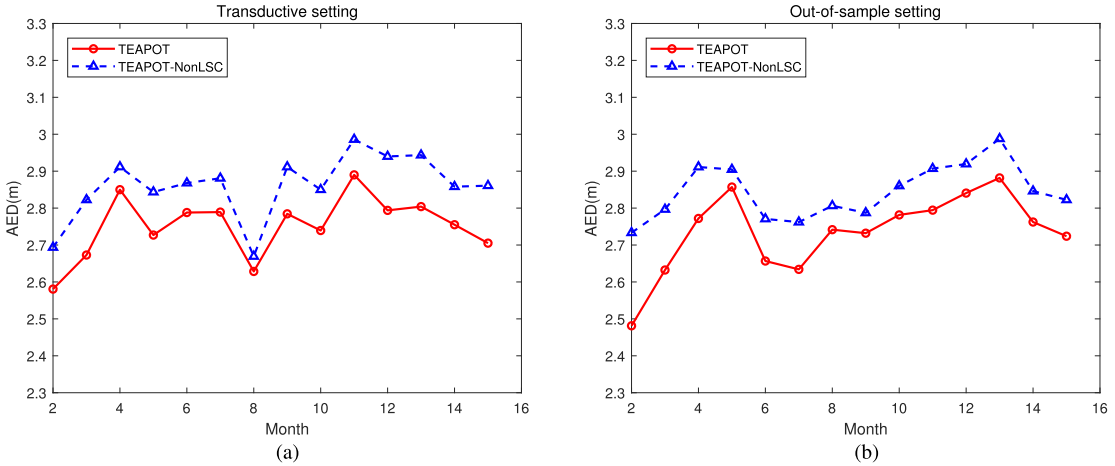


**FIGURE 11.** Effectiveness of the local structural consistency. (a) The AEDs of TEAPOT and TEAPOT-NonLSC in the transductive setting. (b) The AEDs of TEAPOT and TEAPOT-NonLSC in the out-of-sample setting.

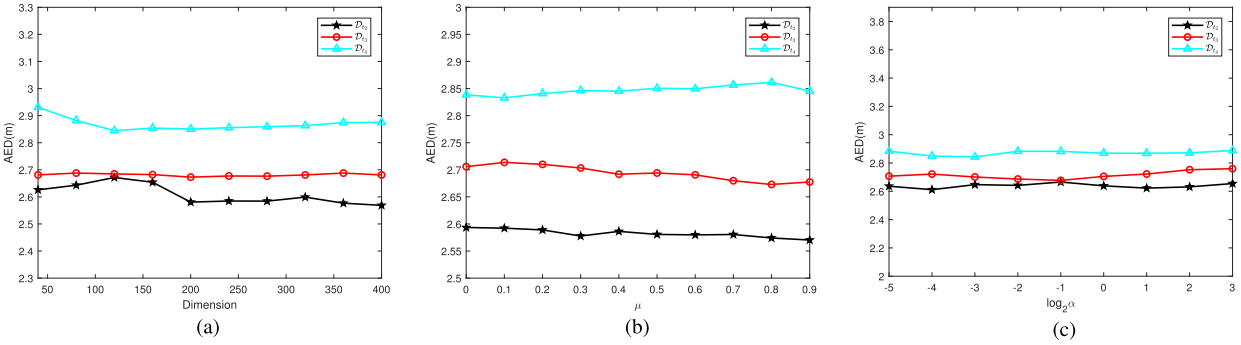


**FIGURE 12.** Parameter sensitivity study of TEAPOT on three target domains. (a) Dimension of subspace $k$. (b) Balance factor $\mu$. (c) Regularization parameter $\alpha$.

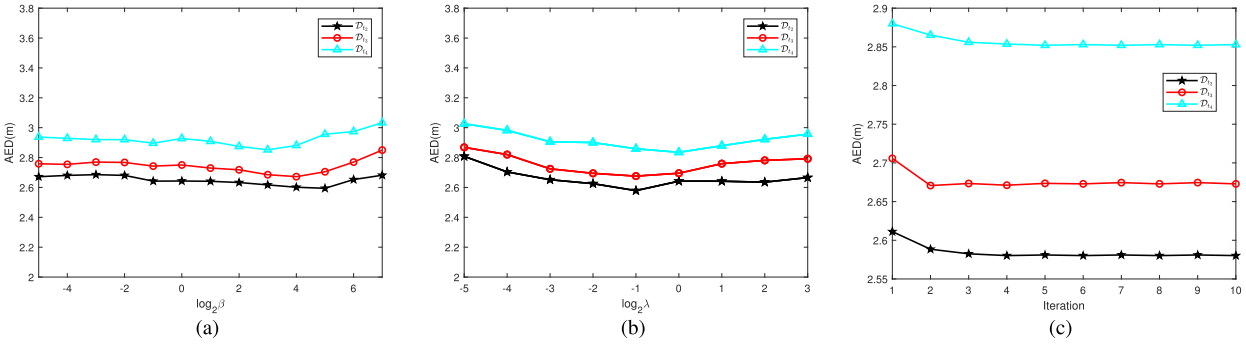


**FIGURE 13.** Parameter sensitivity study of TEAPOT on three target domains. (a) Trade-off parameter $\beta$. (b) Trade-off parameter $\lambda$. (c) Number of iteration $T$.

When $\alpha \to \infty$, the other optimization is not performed, and TEAPOT cannot transfer useful knowledge. It indicates that $\alpha \in [2^{-5}, 2^3]$ can be the optimal parameter values.

We show the sensitivity of the trade-off parameter $\beta$ with a wide range of $\beta \in [2^{-6}, 2^7]$ in Fig. 13 (a). If $\beta$ is too small, the source discriminative information is not considered. If $\beta$ is too large, the classifier may be overfit to the source domain. Hence, a large range of $\beta \in [2^{-1}, 2^4]$ can be chosen to obtain high positioning accuracy. Fig. 13 (b) shows the impact of $\lambda$. Theoretically, a large $\lambda$ makes local neighborhood preservation more important in TEAPOT, and a small $\lambda$ makes the knowledge transfer insufficient. We can observe that TEAPOT can achieve a robust performance when $\lambda \in [2^{-2}, 1]$. Fig. 13 (c) illustrates the convergence performance of our method which shows that TEAPOT achieves the optimal performance no more than $T = 10$ iterations.

## VI. CONCLUSION

In this paper, two versions of TEAPOT approaches via GLOSS constraint have been proposed. The two TEAPOT methods impose the global structural consistency by minimizing the differences between the marginal and conditional distributions of source and target domains, and maximizing the samples variance in a latent subspace. Simultaneously, they also impose the local structural consistency by minimizing within class variance and maximizing between class variance to retain the source discriminative information and preserving the local neighborhood relationship by using manifold regularization. Our methods can effectively mitigate the impact of RSS variations by transferring the knowledge with higher efficiency. Experimental results demonstrate that our proposed TEAPOT methods are very promising for fingerprint-based indoor positioning.

## APPENDIX A PROOF DETAILS IN NONLINEAR TEAPOT

### A. PROOFS OF EQS. (1) AND (3)

By expanding the left side of Eq. (1), we can obtain Eq. (A.1), shown at the bottom of this page. By using the quadratic forms of matrix to Eq. (A.1), we obtain

$$Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{M}_0\boldsymbol{X}^T\boldsymbol{A}\right), \tag{A.2}$$

where $\boldsymbol{M}_0$ is the MMD matrix given by

$$(\boldsymbol{M}_0)_{ij} = \begin{cases} \frac{1}{n_s^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_s \\ \frac{1}{n_t^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_t \\ \frac{-1}{n_s n_t}, & otherwise \end{cases}. \tag{A.3}$$

Similar derivations can be made to yield Eq. (3).

### B. PROOF OF EQ. (14)

By expanding the left side of Eq. (14), we can obtain Eq. (A.4), at the top of the next page, where $d_i = \sum_{j=1}^{n} \boldsymbol{W}_{ij}$ and $\boldsymbol{D}$ is a diagonal matrix with diagonal element $d_i$.

## APPENDIX B PROOF DETAILS IN NONLINEAR TEAPOT

In nonlinear TEAPOT, we utilize the Representer theorem $\boldsymbol{P} = \psi(\boldsymbol{X})\boldsymbol{A}$ to formulate our method. First, by replacing all $\boldsymbol{X}$ and $\boldsymbol{A}$ in Eq. (7) with $\psi(\boldsymbol{X})$ and $\boldsymbol{P}$, respectively, we obtain

$$\begin{aligned} &Tr\left[\boldsymbol{P}^T\psi(\boldsymbol{X})\boldsymbol{H}\psi(\boldsymbol{X})^T\boldsymbol{P}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{H}\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{A}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}^T\boldsymbol{A}\right]. \end{aligned} \tag{B.5}$$

By maximizing Eq. (B.5), we can achieve variance maximization to preserve the separability of projected RSS samples. Similarly, Eq. (5) can be expressed as

$$\begin{aligned} &Tr\left[\boldsymbol{P}^T\psi(\boldsymbol{X})\boldsymbol{M}\psi(\boldsymbol{X})^T\boldsymbol{P}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{M}\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{A}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{M}\boldsymbol{K}^T\boldsymbol{A}\right]. \end{aligned} \tag{B.6}$$

By minimizing Eq. (B.6), we can reduce the distribution difference between domains.

Furthermore, Eq. (14) is written as

$$\begin{aligned} &Tr\left[\boldsymbol{P}^T\psi(\boldsymbol{X})\boldsymbol{L}\psi(\boldsymbol{X})^T\boldsymbol{P}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{L}\psi(\boldsymbol{X})^T\psi(\boldsymbol{X})\boldsymbol{A}\right] \\ &\quad = Tr\left[\boldsymbol{A}^T\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}^T\boldsymbol{A}\right]. \end{aligned} \tag{B.7}$$

By minimizing Eq. (B.7), the neighborhood relationship between samples is preserved.

$$\begin{aligned} &\left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j \right\|^2 \\ &= \left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right)^T\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right) \\ &= Tr\left[\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right)^T\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right)\right] \\ &= Tr\left[\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right)\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{A}^T\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{A}^T\boldsymbol{x}_j\right)^T\right] \\ &= Tr\left[\boldsymbol{A}^T\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{x}_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{x}_j\right)\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{x}_i^T - \frac{1}{n_t}\sum_{j=n_s+1}^{n}\boldsymbol{x}_j^T\right)\boldsymbol{A}\right] \\ &= Tr\left[\boldsymbol{A}^T\left(\frac{1}{{n_s}^2}\sum_{i=1}^{n_s}\boldsymbol{x}_i\sum_{i=1}^{n_s}\boldsymbol{x}_i^T - \frac{1}{n_s n_t}\sum_{i=1}^{n_s}\boldsymbol{x}_i\sum_{j=n_s+1}^{n}\boldsymbol{x}_j^T - \frac{1}{n_s n_t}\sum_{j=n_s+1}^{n}\boldsymbol{x}_j\sum_{i=1}^{n_s}\boldsymbol{x}_i^T + \frac{1}{{n_t}^2}\sum_{j=n_s+1}^{n}\boldsymbol{x}_j\sum_{j=n_s+1}^{n}\boldsymbol{x}_j^T\right)\boldsymbol{A}\right], \end{aligned} \tag{A.1}$$

$$
\begin{aligned}
&\sum_{(i,j)\in\mathcal{N}} \boldsymbol{W}_{ij} \left\| \left[\boldsymbol{A}^T\boldsymbol{X}\right]_i - \left[\boldsymbol{A}^T\boldsymbol{X}\right]_j \right\|^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\left(\boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j\right)^T\left(\boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j\right) \\
&= Tr\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\left(\boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j\right)^T\left(\boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j\right)\right] \\
&= Tr\left[\boldsymbol{A}^T\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{x}_i{}^T - \boldsymbol{x}_j{}^T\right)\right)\boldsymbol{A}\right] \\
&= Tr\left[\boldsymbol{A}^T\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_i{}^T - \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j{}^T - \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_j\boldsymbol{x}_i{}^T + \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_j\boldsymbol{x}_j{}^T\right)\boldsymbol{A}\right] \\
&= 2Tr\left[\boldsymbol{A}^T\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_i{}^T - \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j{}^T\right)\boldsymbol{A}\right] \\
&= 2Tr\left[\boldsymbol{A}^T\left(\sum_{i=1}^{n} d_i\boldsymbol{x}_i\boldsymbol{x}_i{}^T - \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j{}^T\right)\boldsymbol{A}\right] \\
&= 2Tr\left[\boldsymbol{A}^T\left(\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^T - \boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\right)\boldsymbol{A}\right] \\
&= 2Tr\left(\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T\boldsymbol{A}\right).
\end{aligned} \tag{A.4}
$$

$$
\begin{aligned}
&Tr\left[\boldsymbol{A}^T\psi\left(\boldsymbol{X}\right)^T\sum_{c=1}^{C} n_s^{(c)}\left(\frac{1}{n_s^{(c)}}\sum_{i=1}^{n_s^{(c)}}\psi\left(\boldsymbol{x}_i^{(c)}\right) - \frac{1}{n_s}\sum_{i=1}^{n_s}\psi\left(\boldsymbol{x}_i\right)\right)\left(\frac{1}{n_s^{(c)}}\sum_{i=1}^{n_s^{(c)}}\psi\left(\boldsymbol{x}_i^{(c)}\right) - \frac{1}{n_s}\sum_{i=1}^{n_s}\psi\left(\boldsymbol{x}_i\right)\right)^T\psi\left(\boldsymbol{X}\right)\boldsymbol{A}\right] \\
&= Tr\left[\boldsymbol{A}^T\sum_{c=1}^{C} n_s^{(c)}\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)^T\boldsymbol{A}\right] \\
&= Tr\left[\boldsymbol{A}^T\boldsymbol{S}_b\boldsymbol{A}\right]
\end{aligned} \tag{B.9}
$$

Similarly, the within class variance in Eq. (8) and the between class variance in Eq. (9) can be rewritten as Eqs. (B.8) and (B.9) (at the top of this page).

$$
\begin{aligned}
&Tr\left[\boldsymbol{A}^T\psi\left(\boldsymbol{X}\right)^T\sum_{c=1}^{C}\psi\left(\boldsymbol{X}_s^{(c)}\right)\boldsymbol{H}_s^{(c)}\psi\left(\boldsymbol{X}_s^{(c)}\right)^T\psi\left(\boldsymbol{X}\right)\boldsymbol{A}\right] \\
&= Tr\left[\boldsymbol{A}^T\sum_{c=1}^{C}\boldsymbol{K}_s^{(c)}\boldsymbol{H}_s^{(c)}\left(\boldsymbol{K}_s^{(c)}\right)^T\boldsymbol{A}\right] \\
&= Tr\left[\boldsymbol{A}^T\boldsymbol{S}_w\boldsymbol{A}\right].
\end{aligned} \tag{B.8}
$$

The within class scatter matrix and the between class scatter matrix are constructed as follows:

$$
\boldsymbol{S}_w = \sum_{c=1}^{C}\boldsymbol{K}_s^{(c)}\boldsymbol{H}_s^{(c)}\left(\boldsymbol{K}_s^{(c)}\right)^T \tag{B.10}
$$

and

$$
\boldsymbol{S}_b = \sum_{c=1}^{C} n_s^{(c)}\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)\left(\boldsymbol{m}_s^{(c)} - \bar{\boldsymbol{m}}_s\right)^T, \tag{B.11}
$$

respectively, where $\boldsymbol{K}_s^{(c)} = \psi(\boldsymbol{X})^T\psi\left(\boldsymbol{X}_s^{(c)}\right)$, $\boldsymbol{m}_s^{(c)} = \frac{1}{n_s^{(c)}}\sum_{i=1}^{n_s^{(c)}}\boldsymbol{k}_i^{(c)}$ and $\bar{\boldsymbol{m}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{k}_i$, with $\boldsymbol{k}_i = \psi(\boldsymbol{X})^T\psi(\boldsymbol{x}_i)$ and $\boldsymbol{k}_i^{(c)} = \psi(\boldsymbol{X})^T\psi\left(\boldsymbol{x}_i^{(c)}\right)$.

Eq. (24) can thus be obtained by combining Eqs. (B.5-B.9).

## REFERENCES

[1] X. Sun and N. Ansari, "Dynamic resource caching in the IoT application layer for smart cities," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 606–613, Apr. 2018.
[2] X. Guo, L. Li, N. Ansari, and B. Liao, "Knowledge aided adaptive localization via global fusion profile," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1081–1089, Apr. 2018.
[3] X. Guo, L. Li, X. Feng, and N. Ansari, "Expectation maximization indoor localization utilizing supporting set for Internet of Things," *IEEE Internet Things J.*, to be published.
[4] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 13–32, 1st Quart., 2009.
[5] X. Guo, S. Shao, N. Ansari, and A. Khreishah, "Indoor localization using visible light via fusion of multiple classifiers," *IEEE Photon. J.*, vol. 9, no. 6, Dec. 2017, Art. no. 7803716.
[6] W. Zhang, Q. Yin, H. Chen, F. Gao, and N. Ansari, "Distributed angle estimation for localization in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 527–537, Feb. 2013.
[7] J. Yang and Y. Chen, "Indoor localization using improved RSS-based lateration methods," in *Proc. IEEE Global Telecommun. Conf.*, Nov./Dec. 2009, pp. 1–6.
[8] X. Guo, L. Chu, and X. Sun, "Accurate localization of multiple sources using semidefinite programming based on incomplete range matrix," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5319–5324, Jul. 2016.
[9] X. Guo, L. Chu, and N. Ansari, "Joint localization of multiple sources from incomplete noisy Euclidean distance matrix in wireless networks," *Comput. Commun.*, vol. 122, pp. 20–29, Jun. 2018.
[10] X. Guo, S. Zhu, L. Li, F. Hu, and N. Ansari, "Accurate WiFi localization by unsupervised fusion of extended candidate location set," *IEEE Internet Things J.*, to be published.
[11] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 269–280.
[12] X. Guo, L. Li, N. Ansari, and B. Liao, "Accurate WiFi localization by fusing a group of fingerprints via a global fusion profile," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7314–7325, Aug. 2018.
[13] X. Guo and N. Ansari, "Localization by fusing a group of fingerprints via multiple antennas in indoor environment," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9904–9915, Nov. 2017.
[14] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *J. Comput. Netw. Commun.*, vol. 2013, Aug. 2013, Art. no. 185138.
[15] Z. Sun, Y. Chen, J. Qi, and J. Liu, "Adaptive localization through transfer learning in indoor Wi-Fi environment," in *Proc. 7th Int. Conf. Mach. Learn. Appl.*, Dec. 2008, pp. 331–336.
[16] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[17] V. W. Zheng, E. W. Xiang, Q. Yang, and D. Shen, "Transferring localization models over time," in *Proc. 23rd Conf. Artif. Intell.*, 2008, pp. 1421–1426.
[18] H. Zou, Y. Zhou, H. Jiang, B. Huang, L. Xie, and C. Spanos, "Adaptive localization in dynamic indoor environments by transfer kernel learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
[19] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E.-S. Lohan, and J. Huerta, "Long-term Wi-Fi ngerprinting dataset and supporting material," *Data*, vol. 3, no. 3, pp. 1–17, 2018.
[20] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. Conf. Comput. Commun. 19th Annu. Joint Conf. IEEE Comput. Commun. Societies*, Mar. 2000, pp. 775–784.
[21] X. Guo, N. Ansari, L. Li, and H. Li, "Indoor localization by fusing a group of fingerprints based on random forests," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4686–4698, Dec. 2018.
[22] C. Laoudias, D. G. Eliades, P. Kemppi, C. G. Panayiotou, and M. M. Polycarpou, "Indoor localization using neural networks with location fingerprints," in *Proc. Int. Conf. Artif. Neural Netw.* Limassol, Cyprus: Springer, 2009, pp. 954–963.
[23] M. Youssef and A. Agrawala, "The horus WLAN location determination system," in *Proc. ACM MobiSys*, 2005, pp. 205–218.
[24] S. J. Pan, and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[25] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proc. 23rd Conf. Artif. Intell. AAAI*, vol. 8, 2008, pp. 677–682.
[26] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.
[27] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5150–5158.
[28] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1129–1134.
[29] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1414–1430, Jul. 2017.
[30] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds," in *Proc. AISTATS*, 2005, pp. 120–127.
[31] I. Sohn and N. Ansari, "Configuring RBF neural networks," *Electron. Lett.*, vol. 34, no. 7, pp. 684–685, Apr. 1998.
[32] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. Int. Conf. Comput. Learn. Theory*. Amsterdam, The Netherlands: Springer, 2001, pp. 416–426.

**XIANSHENG GUO** (S'07–M'11) received the B.Eng. degree from Anhui Normal University, Wuhu, China, in 2002, the M.Eng. degree from the Southwest University of Science and Technology, Mianyang, China, in 2005, and the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2008, where he is currently an Associate Professor with the Department of Electronic Engineering. From 2008 to 2009, he was a Research Associate with the Department of Electrical and Electronic Engineering, The University of Hong Kong. From 2012 to 2014, he was a Research Fellow with the Department of Electronic Engineering, Tsinghua University. From 2016 to 2017, he was a Research Scholar with the Advanced Networking Laboratory, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, USA. Since 2016, he has been a Research Fellow with Wuhu Overseas Students Pioneer Park, Anhui, China. His research interests include array signal processing, wireless localization, machine learning, information fusion, and software radio design.

**LEI WANG** (S'14) received the B.Eng. degree from Jilin University, Changchun, China, in 2018. He is currently pursuing the master's degree in information and communication engineering with the university of Electronic Science and Technology of China, Chengdu, China. His current research interests include indoor localization, machine learning, and transfer learning.

**LIN LI** (S'12) received the B.Eng. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Information and Communication Engineering. His research interests include indoor localization, machine learning, and transfer learning.

**NIRWAN ANSARI** (S'78–M'83–SM'94–F'09) received the B.S.E.E. degree *(summa cum laude)*, with a perfect GPA, from the New Jersey Institute of Technology (NJIT), in 1982, the M.S.E.E. degree from the University of Michigan, in 1983, and the Ph.D. degree from Purdue University, in 1988. He has also been a Visiting (chair) Professor at several universities. He is currently a Distinguished Professor of electrical and computer engineering with NJIT. He has recently authored *Green Mobile Networks: A Networking Perspective* (Wiley-IEEE, 2017) with T. Han and has co-authored two other books. He has also (co-)authored more than 500 technical publications, over 200 published in widely cited journals/magazines. He has also been granted 36 U.S. patents. His current research focuses on green communications and networking, cloud computing, and various aspects of broadband networks. He was elected to serve in the IEEE Communications Society (ComSoc) Board of Governors as a Member-at-Large, has chaired ComSoc Technical Committees, and has been actively organizing numerous IEEE international conferences/symposia/workshops. He has frequently been delivering keynote addresses, distinguished lectures, tutorials, and invited talks. Some of his recognitions include several Excellence in Teaching Awards, several Best Paper Awards, the NCE Excellence in Research Award, the IEEE TCGCC Distinguished Technical Achievement Recognition Award, the COMSOC AHSN TC Technical Recognition Award, the NJ Inventors Hall of Fame Inventor of the Year Award, the Thomas Alva Edison Patent Award, and the Purdue University Outstanding Electrical and Computer Engineer Award. He is a COMSOC Distinguished Lecturer. He has guest edited a number of special issues covering various emerging topics in communications and networking. He has served on the Editorial/Advisory Board of over ten journals.