# Co-Localization from Labeled and Unlabeled Data Using Graph Laplacian

**Jeffrey Junfeng Pan, Qiang Yang**
Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong
{panjf, qyang}@cse.ust.hk

## Abstract

This paper addresses the problem of recovering the locations of both mobile devices and access points from radio signals, a problem which we call *co-localization*, by exploiting both *labeled* and *unlabeled* data from mobile devices and access points. We first propose a solution using Latent Semantic Indexing to construct the relative locations of the mobile devices and access points when their absolute locations are unknown. We then propose a semi-supervised learning algorithm based on manifold to obtain the absolute locations of the devices. Both solutions are finally combined together in terms of graph Laplacian. Extensive experiments are conducted in wireless local-area networks, wireless sensor networks and radio frequency identification networks. The experimental results show that we can achieve high accuracy with much less calibration effort as compared to several previous systems.

## 1 Introduction

Accurately tracking *mobile* devices in wireless networks using radio-signal-strength (RSS) values is a useful task in robotics and activity recognition. It is also a difficult task since radio signals usually attenuate in a highly nonlinear and uncertain way in a complex environment where client devices may be moving. Existing approaches to RSS localization fall into two main categories [Ferris *et al.*, 2006]: (1) radio propagation models [Maligan *et al.*, 2005; Savvides *et al.*, 2001], which rely on the knowledge of access point locations; (2) statistical machine learning models [Nguyen *et al.*, 2005; Letchner *et al.*, 2005; Bahl and Padmanabhan, 2000], which require a large amount of costly calibration.

However, in cities and large buildings where wireless networks are set up by different network suppliers, it is not easy to ask them to share the location information of all access points for business or privacy reasons. Besides, a mobile device may also want to locate access points for obtaining stable connections or to spot them in hostile areas. In all these cases, sufficient calibration (*labeled*) data on mobile devices and access points may not always be available due to the lack of GPS coverage or costly human effort.

In this paper, we address the problem of simultaneously recovering the locations of both mobile devices and access points, a problem which we call *co-localization*, using *labeled* and *unlabeled* RSS data from both mobile devices and access points. We take two steps for solving this problem. In the first step, we assume that only *unlabeled* RSS data are given. In such case, we show that the problem can be solved by Latent Semantic Indexing (LSI) or Singular Value Decomposition (SVD) [Deerwester *et al.*, 1990], techniques that are popular in information retrieval. Consequently, the relative locations of APs and mobile device trajectory can be determined. In the second step, we assume that a small amount of *labeled* RSS data from mobile devices and access points are given. To determine the absolute locations of the devices and access points, we apply a semi-supervised algorithm with graph Laplacian and manifold learning [Chung, 1997; Belkin and Niyogi, 2003; Ham *et al.*, 2005]. Finally, we provide a unified framework for both the above unsupervised and semi-supervised solutions.

We tested our *co-localization* algorithms in different indoor environments using both static and mobile client devices. We also tested the algorithms with different hardware such as 802.11 Wireless Local Area Networks (WLAN), Wireless Sensor Networks (WSN) and Radio Frequency Identifiers (RFID). Experimental results showed that we can achieve a higher accuracy with much less calibration effort in different environments, motion patterns and with different hardware.

## 2 Related Works

Propagation-model-based approaches are widely used for location estimation due to their simplicity and efficiency [Letchner *et al.*, 2005]. These methods usually assume that access points are *labeled*, e.g., their locations are known. They estimate the distance of the mobile devices relative to some fixed access points based on signal strengths through models that predicts the signal propagation patterns [Savvides *et al.*, 2001]. Researchers have also used Bayesian models to encode the signal propagation pattern [Letchner *et al.*, 2005; Maligan *et al.*, 2005] and infer the locations using Monte Carlo methods [Thrun *et al.*, 2001]. A drawback of propagation-model-based methods is that these models may become inaccurate in a complex domain.

An alternative is to apply machine-learning-based algorithms. With these algorithms the *labels* of access points need not be known. Instead, they usually rely on models that are trained with RSS data collected on a mobile device and are *labeled* with physical locations [Letchner *et al.*, 2005; Nguyen *et al.*, 2005; Ni *et al.*, 2003; Bahl and Padmanabhan, 2000]. The training data are usually collected offline. These signal values may be noisy and nonlinear due to environmental dynamics. Therefore, sufficient data shall be collected to power algorithms for approximating the signal to location mapping functions using K-Nearest-Neighbors [Bahl and Padmanabhan, 2000], kernels [Pan *et al.*, 2005], Bayesian filters [Letchner *et al.*, 2005] and Gaussian processes [Ferris *et al.*, 2006]. A drawback of these models is that they may require much calibration effort.

A viable approach is to use both *labeled* and *unlabeled* data. For example, Bayesian frameworks can be applied to use both *labeled* and *unlabeled* access points [Letchner *et al.*, 2005] and mobile device trajectory [Chai and Yang, 2005]. Our work differs from the above in that we treat mobile devices and access points in a completely symmetric manner: we use both the *labeled* and *unlabeled* data from mobile devices and access points to recover the locations of both of them rather than locating the mobile devices only. To the best of our knowledge, this is the first such work.

# 3 Methodology

## 3.1 Problem Definition

Consider a two-dimensional *co-localization* problem. Assume that a user holds a mobile device and navigates in an indoor wireless environment $\mathcal{C} \subseteq \mathbb{R}^2$ of $n$ access points, which can periodically send out beacon signals. At some time $t_i$, the RSS values from all the $n$ access points are measured by the mobile device to form a row vector $\mathbf{s}_i = [s_{i1}\ s_{i2}\ \ldots\ s_{in}] \in \mathbb{R}^n$. A sequence of $m$ signal strength vectors form an $m \times n$ matrix $S = [\mathbf{s}'_1\ \mathbf{s}'_2\ \ldots\ \mathbf{s}'_m]'$, where "prime" is used to denote matrix transposition. Here, the locations of some access points and the mobile devices at some time $t$ are known or *labeled*, while the rest are *unlabeled*.

Our objectives are stated as follows: We wish to estimate the $m \times 2$ location matrix $P = [\mathbf{p}'_1, \mathbf{p}'_2, \ldots, \mathbf{p}'_m]'$ where $\mathbf{p}_i = [p_{i1}\ p_{i2}] \in \mathcal{C}$ is the location of the mobile device at time $t_i$ and the $n \times 2$ location matrix $Q = [\mathbf{q}'_1, \mathbf{q}'_2, \ldots, \mathbf{q}'_n]'$ where $\mathbf{q}_j = [q_{j1}\ q_{j2}] \in \mathcal{C}$ is the location of the $j$ access points. Our objectives are to determine the locations of all of the remaining access points and the trajectory of the mobile device. We call this problem *co-localization*.

**Example 1** As an example, Figure 1(a) shows an indoor 802.11 wireless LAN environment of size about $60m \times 50m$. It is equipped with $n = 5$ access points. A user with an IBM T42 notebook that is equipped with an Intel Pro/2200BG internal wireless card walks from $A$ through $B, \ldots, E$ to $F$ at time $t_A, t_B, \ldots, t_F$. $m = 6$ signal strength vectors are extracted and the $6 \times 5$ matrix $S$ is shown in Table 1. By walking from $A$ to $B, \ldots, E$ and finally to $F$ in the hallways, we collected $500$ signal strength vectors from $5$ access points. Note that the blank cells denote the missing values, which we can fill in a small default value, e.g., $-100dBm$.

Our task is to estimate the trajectory matrix $P$ of the mobile device at all times and to determine the location matrix $Q$ of the access points $AP_1, AP_2, \ldots, AP_5$.

## 3.2 SVD-based Relative Co-Localization

Given *unlabeled* data only, we can determine the relative locations of the mobile device and the access points. This problem is called relative *co-localization*. Intuitively, we may observe the following characteristics of the data (see Table 1):

Table 1: Signal Strength (unit:dBm)

| | $AP_1$ | $AP_2$ | $AP_3$ | $AP_4$ | $AP_5$ |
|---|---|---|---|---|---|
| $t_A$ | -40 | | -60 | -40 | -70 |
| $t_B$ | -50 | -60 | | -80 | |
| $t_C$ | | -40 | -70 | | |
| $t_D$ | -80 | | **-40** | -70 | |
| $t_E$ | -40 | | -70 | -40 | -80 |
| $t_F$ | -80 | | | -80 | -50 |

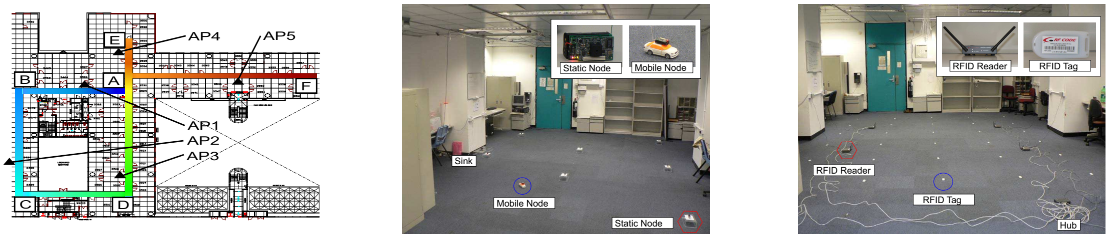
(All values are rounded for illustration)

1. Considering two *rows* of the data, the mobile device at two different time may spatially close to each other if their signal strengths are similar when received from most access points, e.g., the time $t_A$ and $t_E$.
2. Considering two *columns* of the data, two access points may be spatially close to each other if the signal strengths to the mobile device be similar most of the time, e.g., $AP_1$ and $AP_4$.
3. Considering a *single cell* $s_{ij}$ of the data, the mobile device and the $j$ access point may spatially close to each other at time $t_i$ if the signal be strong, e.g., the mobile device is close to $AP_3$ at time $t_D$.

The above observations enabled us to relate co-localization with information retrieval. Not surprisingly, the *co-localization* is closely related to the Latent Semantic Indexing (LSI) [Deerwester *et al.*, 1990]. In this view, we treat an access point as a term and a mobile device at some time as a document. The above three observed characteristics would be mapped to the similarities of document-document, term-term and document-term respectively. Estimating the positions of the mobile device and the access points corresponds to discovering the latent semantics of documents and terms in some concept space.

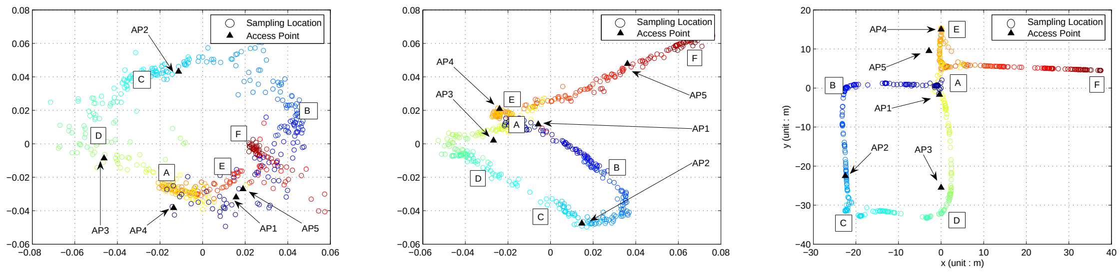
More specifically, we can estimate the relative coordinates by performing Singular Value Decomposition (SVD).

1. Transform the signal matrix $S = [s_{ij}]_{m\times n}$ to a non-negative weight matrix $A = [a_{ij}]_{m\times n}$ by a *linear* function $a_{ij} = s_{ij} - s^{min}$ where $s^{min}$ is the minimal signal strength detected, e.g., the noise level or $-100dBm$.
2. Normalize the weight matrix by $A_{\mathcal{N}} = D_1^{-1/2} A D_2^{-1/2}$. Here, $D_1$ and $D_2$ are both diagonal matrices such that $D_1 = diag(d_1^1, d_2^1, \ldots, d_m^1)$ where $d_i^1 = \sum_{j=1}^n a_{ij}$ and $D_2 = diag(d_1^2, d_2^2, \ldots, d_n^2)$ where $d_j^2 = \sum_{i=1}^m a_{ij}$.
3. Perform SVD on the normalized weight matrix by $A_{\mathcal{N}} \approx U_{m\times r} \Sigma_{r\times r} V'_{n\times r}$. The columns of $U_{m\times r} = [\mathbf{u}_1 \ldots \mathbf{u}_r]$ and $V_{n\times r} = [\mathbf{v}_1 \ldots \mathbf{v}_r]$ are the left and right singular vectors. The singular values of the diagonal matrix $\Sigma_{r\times r} = diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$ are ranked in *non-increasing* order.
4. The (latent) location matrices of the mobile device $P$ and that of the access points $Q$ can be estimated using $P = D_1^{-1/2}[\mathbf{u}_2\ \mathbf{u}_3]$ and $Q = D_2^{-1/2}[\mathbf{v}_2\ \mathbf{v}_3]$. Note that we skip the first singular vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ which mostly capture some constant since matrix $A_{\mathcal{N}}$ is not centering.




(a) WLAN Test-bed (b) WSN Test-bed (c) RFID Test-bed

Figure 1: The Wireless LAN, Wireless Sensor Network and the RFID Test-beds




(a) (Linear) SVD Co-Localization (b) (Nonlinear) SVD Co-Localization (c) Manifold-based Co-Localization

Figure 2: 802.11 Wireless LAN test in an indoor environment

As an example, after performing SVD on data in **Example 1**, we obtained the latent coordinates of the mobile device and the access points, which are shown in Figure 2(a). In this example, it is easy to see that the hallway structure is not well preserved by comparing the true location sequence shown in Figure 1(a). This is because SVD assumes a *linear* subspace, while the correlation of RSS values and distance to APs is often nonlinear [Nguyen *et al.*, 2005].

A better solution is using Kernel SVD or simply transforming signal strengths to weights by some *nonlinear* function. More specifically, we transform the signal matrix $S = [s_{ij}]_{m\times n}$ to a new weight matrix $A = [a_{ij}]_{m\times n}$ by a Gaussian kernel $a_{ij} = exp(-|s_{ij} - s^{max}|^2/2\sigma_A^2)$ where $s^{max}$ is the maximal signal strength detected, e.g., the signal strength around an access point or $-30dBm$. Figure 2(b) plots the *co-localization* result using $P$ and $Q$. Intuitively, the reconstructed hallway structure and the locations of access points are better than that shown in Figure 2(a) while referring to the ground truth illustrated in Figure 1(a).

## 3.3 Manifold-based Absolute Co-Localization

When the physical locations of some access points and the mobile device at some time are known, we can ground the unknown coordinates by exploiting the geometry of the signal distribution. More specifically, we can use manifold-based learning, which generally assumes that if two points are close in the intrinsic geometry of the marginal distribution, their conditional distributions are similar [Belkin *et al.*, 2005; Ham *et al.*, 2005]. This implies that the mobile device shall be spatially close to each other if their signal vectors are similar along some manifold structure [Patwari and Hero, 2004; Pan *et al.*, 2006]. For example, the mobile device at time $t_A$ and $t_E$ shall be spatially close to each other (Figure 1(a)) since their signal strengths are similar (Table 1).

When the manifold assumption holds, the optimal solution is give by $\mathbf{f}^* = \arg\min \Sigma_{i=1}^{l}|f_i - y_i|^2 + \gamma\mathbf{f}^T L\mathbf{f}$ [Ham *et al.*, 2005] where the first term measures the fitting error and the second term poses the smoothness along the manifold and $L$ is the graph Laplacian [Chung, 1997]. For our problem, the objective is to optimize:

$$P^* = \arg\min_{P\in\mathbb{R}^{m\times 2}} (P - Y_P)' J_P (P - Y_P) + \gamma_P P' L_P P \quad (1)$$

Here, $P$ is the coordinate matrix of the mobile device to be determined; $J_P = diag(\delta_1, \delta_2, \ldots, \delta_m)$ is an indication matrix where $\delta_i = 1$ if the coordinate of the mobile device at time $t_i$ is given and otherwise $\delta_i = 0$; $Y_P = [\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_m]'$ is an $m \times 2$ matrix supplying the calibration data where $\mathbf{y}_i$ is the given coordinate of the mobile device at time $t_i$ if $\delta_i = 1$ and otherwise the value of $\mathbf{y}_i$ can be any, e.g., $\mathbf{y}_i = [0\ 0]$; $\gamma_P$ controls the smoothness of the coordinates along the manifold; $L_P = D_P - W_P$ is the graph Laplacian; $W_P = [w_{ij}]_{m\times m}$ is the weight matrix and $w_{ij} = exp(-\|\mathbf{s}_i - \mathbf{s}_j\|^2/2\sigma_P^2)$ if $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors along the manifold and otherwise $w_{ij} = 0$; $D_P = diag(d_1, d_2, \ldots, d_m)$ and $d_i = \sum_{j=1}^{m} w_{ij}$.

Setting the derivative of Equation (1) to zero, the optimal solution is given by [Ham *et al.*, 2005]

$$P^* = (J_P + \gamma_P L_P)^{-1} J_P Y_P \quad (2)$$

Similarly, the coordinates of the access points are given by

$$Q^* = \arg\min_{Q\in\mathbb{R}^{n\times 2}} (Q - Y_Q)' J_Q (Q - Y_Q) + \gamma_Q Q' L_Q Q \quad (3)$$

and

$$Q^* = (J_Q + \gamma_Q L_Q)^{-1} J_Q Y_Q \quad (4)$$

where $L_Q = D_Q - W_Q$ is the graph Laplacian, $W_Q$ is the weight matrix and $D_Q$ is constructed from $W_Q$.

Thus, when the locations of the mobile device and the access points are partially known, we can *co-localize* them by solving Equations (2) and (4) respectively. Alternatively, we can combine them into a single equation as

$$R^* = (J + \gamma_B L_B + \gamma_C L_C)^{-1} JY \quad (5)$$

Here, $R = [P' \ Q']'$ is the coordinate matrix of the mobile device and the access points; $Y = [Y_P' \ Y_Q']'$ gives the label information; $J = \begin{bmatrix} J_P & 0 \\ 0 & J_Q \end{bmatrix}$ is the indication matrix; $L_B = \begin{bmatrix} L_P & 0 \\ 0 & 0 \end{bmatrix}$ and $L_C = \begin{bmatrix} 0 & 0 \\ 0 & L_Q \end{bmatrix}$ are the graph Laplacians.

In practice, the graph Laplacians $L_B$ and $L_C$ in Equation (5) are normalized [Belkin and Niyogi, 2003; Shi and Malik, 2000]. Figure 2(c) shows an example of the manifold-based *co-localization* when the locations of the mobile device at time $t_A, t_B, t_C, t_D, t_E, t_F$ and the access points $AP_2, AP_3, AP_4$ are known. As can be seen, the trajectory of the mobile device is well grounded when compared to the ground truth shown in Figure 1(a). However, due to the limited number of access points, their locations are estimated badly, e.g., the location of $AP_5$.

In the following, we will combine the SVD-based and the Manifold-based *co-localization* together so that we can align the mobile device and the access points to the ground truth and to each other.

### 3.4 A Unifying Framework

So far, we have formulated the unsupervised *co-localization* based on SVD and the semi-supervised *co-localization* based on the manifold assumption using Equation (5) by exploiting the correlation within the mobile device and the access points. In this section, we integrate them through a unifying theory. Essentially, performing SVD on $A_{\mathcal{N}}$ is equivalent to solving the generalized eigenvalue problem [Dhillon, 2001]

$$L_A Z = D_A Z \Lambda \quad (6)$$

where $L_A = D_A - W_A$ is the graph Laplacian [Chung, 1997], $W_A = \begin{bmatrix} 0 & A \\ A' & 0 \end{bmatrix}$ and $D_A = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$. The eigenvalues of the diagonal matrix $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{m+n})$ are ranked in *non-decreasing* order. $Z = [\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_{m+n}]$ are the eigenvectors. $[P' \ Q']' = [\mathbf{z}_2 \ \mathbf{z}_3]$. Note that we skip the first eigenvector $\mathbf{z}_1$ since the solution is trivial. Furthermore, it is interesting to see that we have $\lambda_i = 1 - \sigma_i$ where $i = 1, 2, \ldots, r$ [Dhillon, 2001]. Detailed analysis and comparison of LSI, SVD and graph Laplacian can be found in Latent Semantic Indexing [Deerwester *et al.*, 1990; Dhillon, 2001; Hendrickson, 2006].

Putting these together, our objective is to optimize:

$$R^* = \underset{R \in \mathbb{R}^{(m+n)\times 2}}{\arg\min} \ (R - Y)' J (R - Y) + \gamma R' L R \quad (7)$$

The first term measures the fitting error and the second term constrains the smoothness among the mobile device and the access points. $L = \gamma_A L_A + \gamma_B L_B + \gamma_C L_C = D - W$. The solution is given by:

$$R^* = (J + \gamma L)^{-1} JY \quad (8)$$

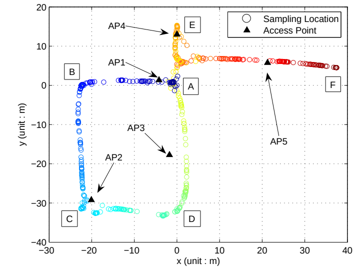



Figure 3: Co-Localization with graph embedding

In practice, the graph Laplacian $L$ is normalized. An example of applying the above *co-localization* algorithm using Equation (8) is shown in Figure 3 when the locations of the mobile device at time $t_A$, $t_B$, $t_C$, $t_D$ $t_F$ and the access point $AP_4$ are known. As can be seen, most of the locations are correctly recovered while using less calibration data than that in Figure 2(c).

## 4 Experiments

We evaluated the performance of the *co-localization* algorithm on three sets of different devices and test-beds:

**(1)** Wireless Local Area Network (WLAN): a person carrying an IBM© T42 notebook, which is equipped with an Intel© Pro/2200GB internal wireless card, walked in an indoor environment of about $60m \times 50m$ in size as shown in Figure 1(a). A total of 2000 examples are collected with sample rate $2Hz$. The ground-truth location labels are obtained by referring to landmark points such as doors, corners and dead-ends. The localization area is composed by one-dimensional hallways.

**(2)** Wireless Sensor Network (WSN): We used a number of MICA2 sensors from Crossbow© for experiments. As can be seen from Figure 1(b), 8 static nodes (AP) were placed in a room of size $5m \times 4m$. One mobile node (MD) was attached on the top of a robot that moved around freely in this domain. A total of 4000 examples are collected with sample rate $2Hz$. The ground-truth location labels of the mobile node were supported by the cameras deployed on the ceiling. The localization area is a two-dimensional plane.

**(3)**Radio Frequency Identification (RFID): We used 4 Mantis readers (AP) and 30 tags (MD) from RF Code©. They were all deployed as stationary nodes, which is shown in Figure 1(c). A total of 2000 examples were collected. The ground truth locations were marked down manually.

We summarize our three experimental setups in Table 2.

For comparison, we also run the following baseline algorithms (1) LANDMARC, a nearest-neighbor weighting based method designed for RFID localization [Ni *et al.*, 2003]; (2) Support Vector Regression (SVR), a simplified variant of a kernel-based method used for WSN localization [Nguyen *et al.*, 2005]; (3) RADAR, a K-Nearest-Neighbor method for WLAN localization [Bahl and Padmanabhan, 2000].

In each experiment, we randomly pick up 500 examples for training and the rest for testing. The training data is further split into *labeled* and *unlabeled* parts. The results shown in Figure 4 are averaged over 10 repetitions for reducing statistical variability. All results are measured in *relative error*

Table 2: The experimental setups of WLAN, WSN and RFID

| Infrastructure | AP | MD | Test-bed | Scale | Dataset Size | Motion Pattern |
|---|---|---|---|---|---|---|
| WLAN | 5 Access Points | 1 Notebook | Hallway | $60m \times 50m$ | 2000 | Mobile (robot) |
| WSN | 8 Static Nodes | 1 Mobile Node | Room | $5m \times 4m$ | 4000 | Mobile (human) |
| RFID | 4 RFID Readers | 30 RFID Tags | Room | $5m \times 4m$ | 2000 | Static |

*distances*, which are error distances in percentage while referring to the maximal error distance in each figure for easy comparison. All parameters are determined from a validation subset. LANDMARC, RADAR and SVR use the *labeled* part of training data only.

In contrary, the *co-localization* method used both *labeled* and *unlabeled* data. We will show how our algorithm benefits from the additional *unlabeled* data and reduces calibration effort. In all, we tested on two configurations for the *co-localization* method: (1) '**Co-Localization no AP**' uses partially *labeled* data from mobile devices for training, in which we tries to recover the locations of the access points; and (2) '**Co-Localization with AP**' repeats the same experiments with the locations of all access points known.

Figures 4(a), 4(b) and 4(c) show the localization error of different mobile devices by varying the number of labeled examples in a training subset which size is fixed to be 500. The three figures could be read in two directions. First, if we compare the results vertically in each figure, we can see how the *unlabeled* data help improve the result in the proposed methods. For example in Figure 4(c), most compared methods have a relative error distance of around 80% when using 50 *labeled* examples. In contrary, the proposed methods have an error of around 40% by employing additional 450 *unlabeled* examples. Secondly, if we compare the results horizontally in each figure, we can find how our methods reduce calibration effort. For example in Figure 4(a), most compared methods have a relative error distance of around 60% when all 500 examples are *labeled*. The proposed '**Co-Localization with AP**' has a similar performance when using 50 *labeled* and 450 *unlabeled* examples. We save the calibration effort.

We found that the mobility of the mobile device and the environment complexity are two main factors that affected the performance of the *co-localization* algorithm. In a static and plane-shaped test-bed (Figure 4(a)), the radio signals are less noisy and the '**Co-Localization no AP**' configuration demonstrated similar performance as RADAR, LANDMARC and SVR when the number of *labeled* examples is small. In a mobile and complex environment, as shown in (Figure 4(c)), the radio signal is more noisy and the '**Co-Localization no AP**' performed much better and more robust than the compared methods. We have also tried some other combinations of experiments that led to a similar conclusion, such as using RFIDs in a mobile scenario.

While comparing the results of '**Co-Localization no AP**' and '**Co-Localization with AP**' in Figures 4(a), 4(b) and 4(c), we can find that knowing the locations of access points is more helpful for localizing the mobile devices in a static and planar scenario (Figure 4(a)) than in a mobile and complex environment (see Figure 4(c)).

Similarly, we can see from Figures 4(d), 4(e) and 4(f) that knowing the locations of mobile devices are more helpful for localizing access points in a static and plane-shaped scenario rather than a mobile and complex environment.

## 5 Conclusion

We have developed a novel graph Laplacian approach to solve the problem of simultaneously recovering the locations of both mobile devices and access points. In our co-localization framework, we find the relative locations of mobile devices and access points by exploiting a SVD based method, and find the absolute locations using a small collection of labeled data through graph Laplacian methods. Our extensive experiments in three different configurations showed that we can achieve high performance with much less calibration effort as compared to several previous approaches. The significance of the work is that we can leverage both the knowledge of the access point locations and the mobile device trajectories to obtain more accurate localization. Indeed this is one of our future works. Besides, we would try to evaluate the performance in a large-scale and dynamic environment, e.g., in a city level and in different time. We may also vary more parameters such as number of access points and their deployment density and study the robustness.

## Acknowledgement

We thank NEC China Lab (NECLC05/06.EG01) for their generous support of this research.

## References

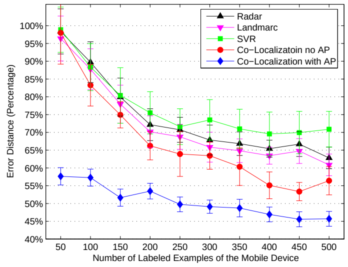
[Bahl and Padmanabhan, 2000] P. Bahl and V.N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of the Conference on Computer Communications*, volume 2, pages 775–784, Tel Aviv, Israel, March 2000.

[Belkin and Niyogi, 2003] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

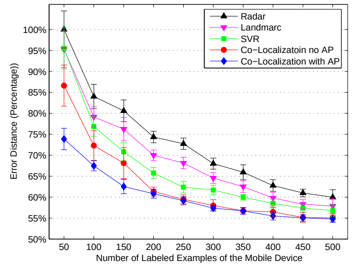
[Belkin *et al.*, 2005] Misha Belkin, Partha Niyogi, and Vikas Sindhwani. On manifold regularization. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 17–24. Society for Artificial Intelligence and Statistics, January 2005.

[Chai and Yang, 2005] X. Chai and Q. Yang. Reducing calibration effort for location estimation using unlabeled samples. In *Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications*, pages 95–104, Kauai Island, HI, USA, 2005.

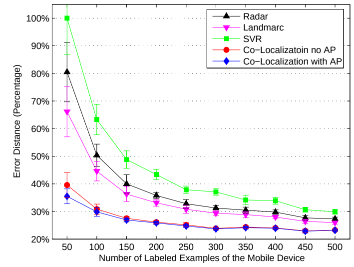
[Chung, 1997] F. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.



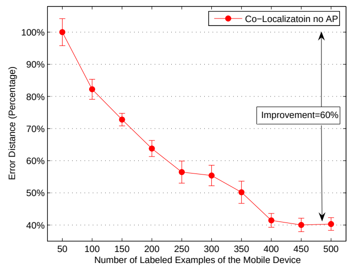

(a) RFID MD (tags)



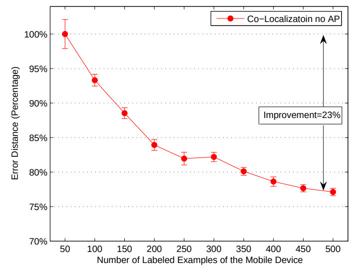

(b) WSN MD (mobile sensor node)



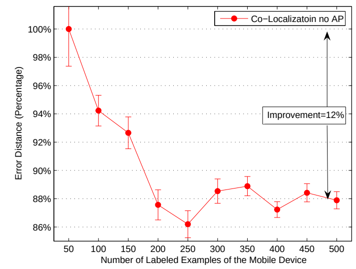

(c) WLAN MD (notebook)




(d) RFID AP (readers)




(e) WSN AP (static sensor nodes)




(f) WLAN AP (access points)

Figure 4: Experimental Results over 10 Repetitions (Mean and Std.): MD for Mobile Device; AP for Access Point


[Deerwester *et al.*, 1990] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6):391–407, 1990.

[Dhillon, 2001] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274, New York, NY, USA, 2001. ACM Press.

[Ferris *et al.*, 2006] B. Ferris, D. Hahnel, and D. Fox. Gaussian processes for signal strength-based location estimation. In *Proceedings of Robotics: Science and Systems*, Philadelphia, Pennsylvania, USA, August 2006.

[Ham *et al.*, 2005] J. Ham, D. Lee, and L. Saul. Semisupervised alignment of manifolds. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 120–127. Society for Artificial Intelligence and Statistics, January 2005.

[Hendrickson, 2006] B. Hendrickson. Latent semantic analysis and fiedler embeddings. In *Proceedings of SIAM Workshop on Text Mining*, April 2006.

[Letchner *et al.*, 2005] J. Letchner, D. Fox, and A. LaMarca. Large-scale localization from wireless signal strength. In *Proceedings of the 20th National Conference on Artificial Intelligence*, pages 15–20, Pittsburgh, USA, July 2005.

[Maligan *et al.*, 2005] D. Maligan, E. Elnahrawy, R. Martin, W. Ju, P. Krishnan, and A.S. Krishnakumar. Bayesian indoor positioning systems. In *Proceedings of the Conference on Computer Communications*, volume 2, pages 1217–1227, Miami, FL, USA, March 2005.

[Nguyen *et al.*, 2005] X. Nguyen, M. I. Jordan, and B. Sinopoli. A kernel-based learning approach to ad hoc sensor network localization. *ACM Transactions on Sensor Networks*, 1(1):134–152, 2005.

[Ni *et al.*, 2003] L.M. Ni, Y. Liu, Y.C. Lau, and A.P. Patil. LANDMARC: Indoor location sensing using active RFID. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, pages 407–416, Dallas, TX, USA, March 2003.

[Pan *et al.*, 2005] J. J. Pan, J. T. Kwok, Q. Yang, and Y. Chen. Accurate and low-cost location estimation using kernels. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 1366–1371, Edinburgh, Scotland, 2005.

[Pan *et al.*, 2006] J. J. Pan, Q. Yang, H. Chang, and D. Y. Yeung. A manifold regularization approach to calibration reduction for sensor-network based tracking. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 988–993, Boston, USA, July 2006.

[Patwari and Hero, 2004] N. Patwari and A. O. Hero. Manifold learning algorithms for localization in wireless sensor networks. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, May 2004.

[Savvides *et al.*, 2001] A. Savvides, C. Han, and M. B. Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pages 166–179, Rome, Italy, 2001.

[Shi and Malik, 2000] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[Thrun *et al.*, 2001] S. Thrun, D. Fox, W. Burgard, and Dellaert. F. Robust monte carlo localization for mobile robots. *Artificial Intelligence*, 128(1-2):99–141, 2001.