

Localization Analysis Based on Maximum Scatter Difference in Indoor WLAN Environments

Jia-Han Jan, Kun-Chou Lee

Department of Systems and Naval Mechatronics Engineering, National Cheng Kung University, Tainan
Email: jch615@hotmail.com, kclee@mail.ncku.edu.tw

Received: Jan. 7th, 2013; revised: Jan. 23rd, 2013; accepted: Feb. 3rd, 2013

Copyright © 2013 Jia-Han Jan, Kun-Chou Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: In this paper, we utilize Maximum Scatter Difference (MSD) for indoor localization over Wireless Local Area Network (WLAN) received signal data. Maximum Scatter Difference can solve the singular matrix problems by linear discriminant analysis calculates the inverse matrix of within-class scatter matrix. The objective functions of Maximum Scatter Difference are to maximize the inter-class scatter matrix and minimize the within-class scatter matrix. Therefore, Maximum Scatter Difference cannot need to construct the inverse matrix of within-class scatter matrix. It can avoid the singular matrix problems. We utilize location fingerprinting approaches for indoor localization. The approaches are divided into two parts including off-line and on-line stages. In the off-line stage, we collected received signal data are projected onto the feature space of Maximum Scatter Difference. In the on-line stage, we receive real-time signal data compared with the off-line stage that was collected data using Maximum Likelihood (ML) to estimate current location. Maximum Scatter Difference applied to indoor WLAN localization that not only avoids the singular matrix but also reduces the localization computation. From Simulation results show that this algorithm can be implemented.

Keywords: Maximum Scatter Difference; WLAN; Indoor Localization

利用最大散度差于室内无线网络定位分析

詹佳翰, 李坤洲

国立成功大学系统及船舶机电工程学系, 台南
Email: jch615@hotmail.com, kclee@mail.ncku.edu.tw

收稿日期: 2013 年 1 月 7 日; 修回日期: 2013 年 1 月 23 日; 录用日期: 2013 年 2 月 3 日

摘要: 在本研究中将最大散度差应用在无线网络室内定位上, 最大散度差主要是可以解决线性鉴别分析在计算类内散度矩阵时所遭遇的奇异性矩阵问题, 它是利用类间散度矩阵与类内散度矩阵之差作为目标函数, 由于不需要构造类内散度矩阵的反矩阵, 故可以避免掉奇异性矩阵的问题。本研究中其定位的方式是透过指纹特征比对法的概念来进行定位, 共分为两个阶段: 在离线阶段, 利用最大散度差将原始讯号数据库作前置处理; 在在线阶段, 将接收到的实时讯号数据利用最大似然法计算出目前接收讯号的所在位置。将最大散度差应用于室内无线网络进行定位, 不仅可以避免奇异性矩阵的问题并且还可以降低定位时的计算量, 从实验结果得知此应用是可以实行的。

关键词: 最大散度差; 无线网络; 室内定位

1. 引言

在图样辨识领域当中，线性鉴别分析(Linear Discriminant Analysis, LDA)^[1-3]一直是大家公认非常重要的算法之一，其被广泛应用在人脸图像辨识领域上，LDA 其主要的目的是目标是寻找一个最优的投影方向，样本数据在该方向上的投影能够让不同类别的类间散度距离拉大，并且能使相同类别内的类内散度距离缩小，如此一来便能增加类别数据的分群能力。但是 LDA 算法在应用时遇到计算类内散度矩阵的反矩阵时，容易遇到奇异性矩阵的问题而造成无法求解。为了避免此奇异性矩阵的问题，本文中引用的最大散度差(Maximum Scatter Difference, MSD)算法^[4-6]与 LDA 算法很相似，但是 MSD 和 LDA 算法的不同之处在于可以解决奇异性矩阵的问题，它是利用类间散度矩阵与类内散度矩阵之差作为目标函数，由于不需要构造类内散度矩阵的反矩阵，故可以避免计算类内散度矩阵遭遇到奇异性矩阵的问题。

在此研究中，其定位方式是引用指纹特征比对法的概念^[7,8]来进行室内定位，其中流程就跟在对比人类指纹时相同，共分为两个阶段，分别是训练(即离线阶段)和测试(即在线阶段)，离线阶段是在选定的参考位置上收集来自各个讯号来源(Access Point, AP)的强度值(dBm)储存至数据库，在训练(即离线阶段)透过最大散度差法此算法将数据库内数据作前置处理，将其特征提取出来投影至特征空间，藉此降低原始数据库的维度，减少定位时的计算量；接着测试(即在线阶段)将接收到的实时讯号同样投影至特征空间，然后透过最大似然法(Maximum Likelihood, ML)^[9,10]与原始数据投影至特征空间内数据做比对，来估算出目前接收讯号者的所在位置。此种定位方式的优点不仅能避免掉多重路径反射对于定位的影响，并且只需要由现在广泛存在的无线网络和可以接收讯号的接收机器(本研究是利用笔记型计算机)即可实现，故并不需要额外的专用设备，可以节省硬件之成本。透过实验结果得知，将最大散度差此算法应用在室内无线网络定位上，可以准确的计算出接收讯号者的所在位置，证明本研究的可行性及高效性。

2. 最大散度差(Maximum Scatter Difference, MSD)

假设我们在其中某个地点接收到的原始讯号来

源(Access Point, AP)维度大小为 D ，我们将此讯号定义为向量 $\bar{X} = [x_1, x_2, \dots, x_D]^T$ ，接着我们总共收集了 L 个地点(Receive position)，所以有 L 个类别，而每个类别中有 M_l 笔数据，故所有数据笔数为 $M = \sum_{l=1}^L M_l$ ，我们将这些 M 笔数据定义为 $\bar{X}_i, i=1, \dots, M$ ，将这些数据储存至数据库，排序方式如图 1 所示。定位主要分为训练(即离线阶段)和测试(即在线阶段)两个阶段。

在训练(离线阶段)，我们要寻找一个转换矩阵 \bar{A} ，投影到特征空间是利用此线性变换即为 $\bar{Y}_i = \bar{A}_i^T \bar{X}_i$ ，将原始资料 \bar{X} 转换为 $\bar{Y} = [y_1, y_2, \dots, y_d]^T$ ，其中 $(d \ll D)$ ，转换成最大散度差的所构造出来的特征空间。

首先计算出每个类别内的数据平均值

$$\bar{m}_l = \frac{1}{M_l} \sum_{i=1}^{M_l} \bar{X}_i, l=1, \dots, L \quad (1)$$

接着计算出全部数据的平均值

$$\bar{m} = \frac{1}{M} \sum_{l=1}^L \sum_{i=1}^{M_l} \bar{X}_i \quad (2)$$

计算不同类别之间的特征参数散布情况，类间散度矩阵定义为

$$\bar{S}_B = \sum_{l=1}^L \left[M_l (\bar{m}_l - \bar{m}) \cdot (\bar{m}_l - \bar{m})^T \right] \quad (3)$$

而计算同一类别内的特征参数散布情况，类内散度矩阵为

$$\bar{S}_W = \sum_{l=1}^L \sum_{i=1}^{M_l} \left[(\bar{X}_i - \bar{m}_l) \cdot (\bar{X}_i - \bar{m}_l)^T \right] \quad (4)$$

最大散度差判别准则其目标是寻找一个最优的投影方向，经过转换矩阵 \bar{A} 投影至特征空间时，使得

	Reference Position # 1				...	Reference Position # L			
	\bar{X}_1	\bar{X}_2	...	\bar{X}_D		\bar{X}_D
	$t = \Delta t$	$t = 2\Delta t$...	$t = N\Delta t$...	$t = \Delta t$	$t = 2\Delta t$...
AP #1	signal	signal	...	signal	...	signal	signal	...	signal
AP #2	signal	signal	...	signal	...	signal	signal	...	signal
...
AP #D	signal	signal	...	signal	...	signal	signal	...	signal

Figure 1. The structure of the database
图 1. 数据库之构造图

不同类别的类间散度最大且同一类别的类内散度最小，所以最大散度差的目标函数为

$$\arg \max_{\bar{A}} \bar{A}^T (\bar{S}_B - \bar{S}_W) \bar{A} \quad (5)$$

其目标函数最大化可转换为下式(6)广义特征值求解问题

$$(\bar{S}_B - \bar{S}_W) \bar{A} = \lambda \bar{A} \quad (6)$$

在求解目标函数与 LDA 在求解时类似，但是 LDA 在求解时必须计算 \bar{S}_W^{-1} ，而从式(6)可以看出 MSD 不需要运算 \bar{S}_W^{-1} ，因此可以避免掉奇异性矩阵的问题。接着我们求出特征值方程式的解，算出特征值 λ_i 和所对应的特征向量 \bar{A}_i ，其中 $i=1, \dots, D$ ，透过选取前 $d (d \leq D)$ 个最大的正数特征值 ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$) 所相对应的特征向量，经过转换矩阵将其投影到 MSD 特征空间，投影完后之特征空间如图 2 所示，其中选取的特征值数目对应的特征向量即是主成份的个数。

3. 最大似然法(Maximum Likelihood, ML)

接着在测试(在线阶段)中，我们同样收集各个参考地点(即 $\bar{r}_1, \bar{r}_2, \dots, \bar{r}_L$) 的 D 个讯号来源。将收集到的讯号投影至 MSD 的特征空间后，可以得到

$\bar{Y} = [y_1, y_2, \dots, y_d]^T$ 的矩阵维度，接着使用最大似然法来估算出目前讯号的接收位置坐标。透过贝氏定理 (Bayes' rule)，我们可以得到事后机率 $P(\bar{r}_i | \bar{Y})$ ，如下式(7)

$$P(\bar{r}_i | \bar{Y}) \propto P(\bar{Y} | \bar{r}_i) P(\bar{r}_i) \quad (7)$$

其中 $P(\bar{Y} | \bar{r}_i)$ 为似然函数(likelihood function)，此机率函数表示在已知位置 \bar{r}_i 中，会发生 \bar{Y} 的机率。在本实

	Reference Position # 1				...	Reference Position # L			
	\bar{Y}_1	\bar{Y}_2	...	\bar{Y}_N		\bar{Y}_M
	$t = \Delta t$	$t = 2\Delta t$...	$t = N\Delta t$...	$t = \Delta t$	$t = 2\Delta t$...
Projection #1	feature	feature	...	feature	...	feature	feature	...	feature
Projection #2	feature	feature	...	feature	...	feature	feature	...	feature
...
Projection #d	feature	feature	...	feature	...	feature	feature	...	feature

Figure 2. The structure of the MSD feature space
图 2. MSD 特征空间之构造图

验中，我们所取的参考地点都是在一个有限的范围内，故事前机率 $P(\bar{r}_i)$ 可以假设为均匀分布，从式(7)中可以得知，事后机率 $P(\bar{r}_i | \bar{Y})$ 是取决于似然函数 $P(\bar{Y} | \bar{r}_i)$ 。为了使估算似然函数 $P(\bar{Y} | \bar{r}_i)$ 更加容易，通常都把似然函数 $P(\bar{Y} | \bar{r}_i)$ 假设为高斯分布(Gaussian distribution)。意思是在各个参考地点 \bar{r}_i 中， \bar{Y} 的每个主成份都为高斯随机变数。所以第 d 个投影特征成份 y_d 发生在参考地点 \bar{r}_i 的机率为：

$$P(\bar{Y} | \bar{r}_i) = \frac{1}{\sqrt{2\pi\sigma_{i,d}^2}} \exp\left\{-\frac{(y_d - \mu_{i,d})^2}{2\sigma_{i,d}^2}\right\} \quad (8)$$

因为讯号 \bar{Y} 的每个主成份为统计独立(statistically independent)，所以可以借由乘上 \bar{Y} 的每个不同主成份的高斯机率来计算出似然函数 $P(\bar{Y} | \bar{r}_i)$

$$P(\bar{Y} | \bar{r}_i) = \prod_{d=1}^D \frac{1}{\sqrt{2\pi\sigma_{i,d}^2}} \exp\left\{-\frac{(y_d - \mu_{i,d})^2}{2\sigma_{i,d}^2}\right\}, \quad i = 1, \dots, L \quad (9)$$

其中 $\mu_{i,d}$ 和 $\sigma_{i,d}^2$ 是在第 i 个参考位置 \bar{r}_i 中，预先测量 M_i 个讯号数据样本投影到 MSD 特征空间算出的平均值和标准差，如图 3 所示。可以将投影至特征空间的 \bar{Y} 向量，经由式(8)和式(9)，计算出全部参考位置的似然函数 $P(\bar{Y} | \bar{r}_i)$ 并由最大的似然函数值来决定出定位结果。

为了使定位结果更准确，我们进一步计算向量 \bar{Y} 的接收位置坐标，在全部的参考位置中，实际接收位置坐标可以看作是全部参考位置的加权平均，我们可以对似然函数 $P(\bar{Y} | \bar{r}_i)$ 正规化来算出权重值

$$w_i = \frac{P(\bar{Y} | \bar{r}_i)}{\sum_{i=1}^L P(\bar{Y} | \bar{r}_i)} \quad (10)$$

	Receive Position # 1	Receive Position # 2	...	Receive Position # L
Projection #1	$\mu_{1,1}, \sigma_{1,1}$	$\mu_{2,1}, \sigma_{2,1}$...	$\mu_{L,1}, \sigma_{L,1}$
Projection #2	$\mu_{1,2}, \sigma_{1,2}$	$\mu_{2,2}, \sigma_{2,2}$...	$\mu_{L,2}, \sigma_{L,2}$
...
Projection #d	$\mu_{1,d}, \sigma_{1,d}$	$\mu_{2,d}, \sigma_{2,d}$...	$\mu_{L,d}, \sigma_{L,d}$

Figure 3. The mean and standard deviation of each reference in the MSD feature space
图 3. MSD 特征空间内每个参考地点的平均值与标准差

其中 w_i 表示在第 i 个参考位置，正规化后的似然函数值。最后结果位置的坐标可以由正规化之后的似然函数值估算出来。

$$\bar{r}' = \sum_{i=1}^L (w_i \bar{r}_i) \quad (11)$$

其中 $\bar{r}' = (x', y')$ 即为最后计算出来的位置坐标。

4. 仿真结果与分析

本文中室内无线网络定位的实验位置是位于国立成功大学机械工程学系系馆，我们选定了 10 个 ($L = 10$) 参考地点，位置与位置之间间隔为 10 公尺，然后我们利用笔记型计算机来接收讯号，在每个参考地点接收 10 个 ($D = 10$) 讯号来源(即无线网络接收到的讯号数目)，并且在每个参考地点收集 150 笔讯号强度值，每一笔的时间间隔为 1 秒，我们将它们分成两组数据，前 100 笔讯号强度值当成训练(即离线阶段)时所需的训练笔数，而参考地点有 10 个，故训练向量笔数为 1000 笔，并将这些数据放置于数据库，其数据维度为 10×1000 。而后 50 笔讯号强度值当成测试(即在线阶段)，测试向量笔数为 500 笔，将其当成测试定位结果是否正确所需的测试向量。

在本文实验中，我们将收集到的训练向量所集合而成的数据库，利用最大散度差投影至特征空间后便完成训练步骤，并计算出训练完的平均值及标准差，以便测试定位结果时所需，接着进行测试步骤，将测试向量利用最大似然法算出定位的结果，其中整个室内定位算法的流程图，如图 4 所示。

本研究中，由于原始数据库数据量较大，拥有较高的维度，故将原始数据库利用最大散度差算法投影至特征空间，透过选择不同的主成份个数藉此降低原

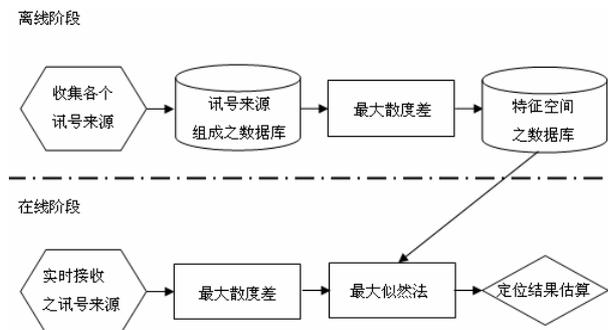


Figure 4. The flow chart of indoor positioning for MSD
 图 4. 最大散度差室内定位算法之流程图

始数据库的维度，当中我们选取前几个较具有特征的主成份，使得在定位计算时可以减少计算量，执行速度加快，同时达到高准确率。由如图 5 所示，主成份个数选择 3 个 ($d = 3$) 和 4 个 ($d = 4$) 主成份时，定位准确率分别为 94.4% 和 96.6%，所以我们只需要选择 3 个 ($d = 3$) 主成份即可达到九成 (94.4%) 以上的定位准确率，相较于原本接收到的原始数据库 ($D = 10$)，减少了七成的计算量，达到在降低数据库维度的同时还拥有高定位准确率。

上述的定位结果，皆以训练向量为 100 笔进行训练，在本研究中，其定位方式是引用指纹特征比对法概念，必须在每个参考地点事先收集训练向量，在离线阶段时需要耗费大量的收集时间，因此若能减少收集训练向量笔数及时间，可提升定位的整体效率，故减少训练向量的笔数对照其定位准确率，如图 6 所示。当中分别取 10 笔、20 笔... 到 100 笔训练笔数做训练，主成份个数我们取 3 个 ($d = 3$) 的情况下，其不同的训练笔数定位准确率，训练笔数取 20 笔、30 笔、40 笔时准确率分别为 89.20%，91.20% 及 92.60%，只需取 30 笔的训练向量，其定位准确率即可达到 90% 以上，大幅节省离线阶段事先收集数据的时间，从结果来看，使用较少的训练向量依然可达到高定位准确率。

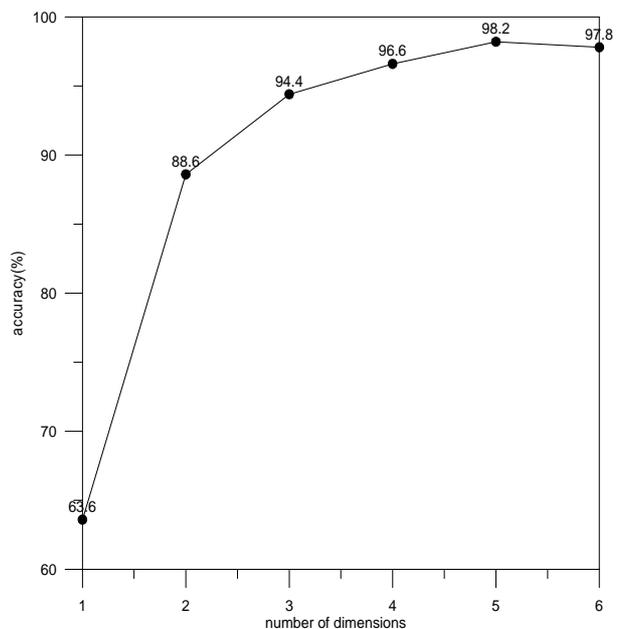


Figure 5. The positioning accuracy versus number of the principal components
 图 5. 选择不同主成份个数的定位准确率

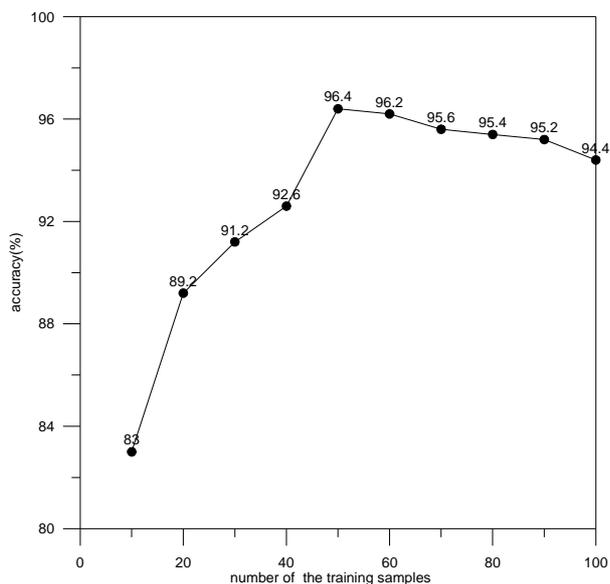


Figure 6. The positioning accuracy versus number of the training samples, as the number of selected components is $d = 3$
图 6. 选择 3 个主成份个数情况下, 不同训练笔数的定位准确率

在本文中所选取的讯号接收位置彼此之间的距离为 10 公尺, 当中我们将实际坐标 (x, y) 和式(10)及式(11)所计算得知的估算坐标 (x', y') 两者之间的距离来当作定位误差, 此距离 Ω 是用欧氏距离来计算。

$$\Omega = \sqrt{(x - x')^2 + (y - y')^2} \quad (12)$$

本文中相邻两个讯号接收位置的间距为 10 公尺, 意即当定位误差小于 5 公尺内即视为成功辨识, 当定位误差越小时, 成功辨识的条件便更不易达到, 在主成份个数依然选取 3 个($d = 3$)的情况下, 定位误差在 5 公尺范围内, 计算出不同定位误差所对应的定位准确率, 如图 7 所示。选择定位误差为 2.5 公尺、3 公尺、3.5 公尺时定位准确率分别为 90.8%, 91.6%, 91.8%, 其中定位误差取的距离越小时, 代表所预估出来的位置与真正的实际位置必须更加接近才可以成功辨识, 从结果可得知, 我们将定位误差仅取 2.5 公尺, 表示每次所预估出来的估算位置与真正的实际位置之间的距离都在 2.5 公尺以内时, 其成功辨识的正确率为 90.8%, 依然可以达到九成以上的高定位准确率。

5. 结论

本研究所应用的 MSD 与 LDA 很相似, 两种算法不同之处在于 MSD 可以解决奇异性矩阵的问题, 它

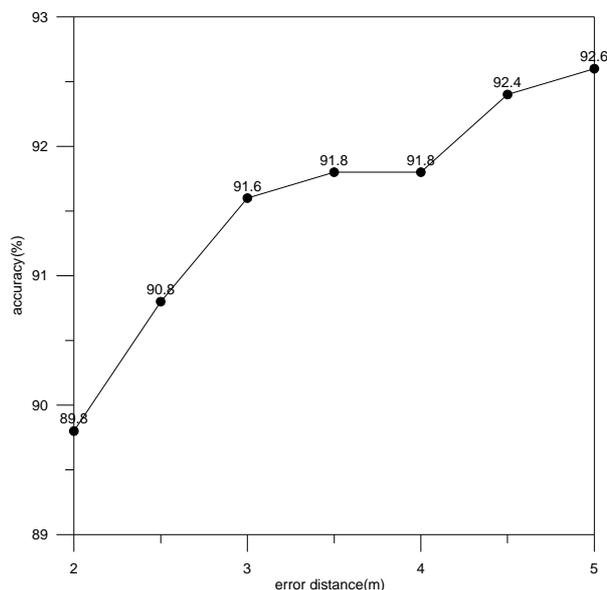


Figure 7. The positioning accuracy versus different error distance
图 7. 选取不同的定位误差的定位准确率

是利用类间散度矩阵与类内散度矩阵之差作为目标函数, 由于不需要构造类内散度矩阵的反矩阵, 故可以避免计算类内散度矩阵遭遇到奇异性矩阵的问题。本研究中应用 MSD 到室内无线网络定位上, 实验结果证明, 只需选取 3 个主成份个数便可以达到 94.4% 的定位准确率, 大量减少定位时的计算量, 并且只需取 30 笔训练向量一样可以达到九成以上的高准确率, 减少事先收集数据库的时间, 并且透过不同的定位误差距离来看出其定位的准确率, 从以上实验结果得知将最大散度差应用于室内无线网络定位是可以实行的。

6. 致谢

感谢行政院国家科学委员会补助专题研究计划 NSC99-2221-E-006-244-MY3 经费补助支持。并感谢我的导师李坤洲教授以及实验室学长, 在本研究过程中, 适时给予我研究上的建议, 使本论文内容能够更加完整与严谨, 在此致上最诚挚的谢意。

参考文献 (References)

- [1] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. Eigenfaces vs. Fisherfaces recognition using class specific linear projection. IEEE Transactions Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720.
- [2] D. L. Swets, J. J. Weng. Using discriminant eigenfeatures for image retrieval. IEEE Transactions Pattern Analysis and Ma-

- chine Intelligence, 1996, 18(8): 831-836.
- [3] K. C. Lee, L. T. Wang, J. S. Ou and C. W. Huang. Radar target recognition by using linear discriminant algorithm on angular-diversity RCS. *Journal of Electromagnetic Waves and Applications*, 2007, 21(14): 2033-2048.
- [4] H. Li, T. Jiang and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks*, 2006, 17(1): 157-165.
- [5] J. G. Wang, Y. U. Lin, W. K. Yang and J. Y. Yang. Kernel maximum scatter difference based feature extraction and its application to face recognition. *Pattern Recognition Letters*, 2008, 29(13): 1832-1835.
- [6] F. X. Song, C. Ke, Y. J. Yu and L. S. Hai. Maximum scatter difference, large margin linear projection and support vector machines. *Acta Automatica Sinica*, 2004, 30(6): 890-896.
- [7] K. Kaemarungsi, P. Krishnamurthy. Properties of indoor received signal strength for WLAN location fingerprinting. *Proceedings of the First Annual International Conference on Mobile and Ubiquitous System: Networking and Services*, Boston, August 2004.
- [8] K. Kaemarungsi, P. Krishnamurthy. Modeling of indoor positioning systems based on location fingerprinting. *23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, 2004, 2: 1012-1022.
- [9] T. K. Moon, W. C. Stirling. *Mathematical methods and algorithms for signal processing*. Upper Saddle River: Prentice Hall, 2000.
- [10] R. E. Ziemer, W. H. Tranter. *Principles of communications: Systems, modulation, and noise*. Wiley, 2010.