Performance Comparison of Similarity Measurements for Database Correlation Localization Method

Juraj Machaj and Peter Brida

University of Zilina, Faculty of Electrical Engineering,
Department of Telecommunications and Multimedia,
Univerzitna 8215/1, 010 26 Zilina, Slovakia
{Juraj.Machaj,Peter.Brida}@fel.uniza.sk

Abstract. User positioning is a very important feature in user adaptive system. The user position can be estimated by various positioning methods. This paper investigates an impact of similarity measurements on localization error in deterministic database correlation method. It is also called fingerprinting. Main idea is to compare widely used Euclidean distance with other similarity measurements. Seven different similarity measurements are implemented to simulation model created in Matlab software tool. Computation complexity of each similarity measurement is investigated and impact of similarity measurements on localization error in normal and extreme conditions is shown.

Keywords: Database correlation method, similarity measurements, fingerprinting localization, indoor positioning.

1 Introduction

The number of LBS (Location Based Services) is rising very fast in last years [1]. Basic requirement for LBS is to know user location. That can be achieved by various ways in dependency on environment, e.g. GPS (Global Positioning System) in outdoor. On the other hand, there can be problem with high signal attenuation in indoor environment and dense urban areas. Alternative positioning solutions have to be used. Generally, they utilized various wireless communication platforms, e.g. cellular networks or IEEE 802.11x etc.

Positioning based on cellular networks is often used as alternative solution in urban environment. In urban environment, A-GPS can be also used to estimate location. Problem with localization is even bigger in indoor environment. Signal fluctuations are large because of multipath signal propagation. Many indoor localization algorithms and systems [2] based on Bluetooth [3], Zig-Bee [4], UWB (Ultra Wide Band) [5, 6], RFID [7] and IEEE 802.11x [8] were developed.

The most popular algorithms used in indoor environment are based on IEEE 802.11 [9 - 12]. Most of they uses signal strength information and are based on fingerprinting algorithm. The biggest advantage is that the algorithm does not need a new infrastructure. Another advantage seems to be multipath propagation resistance.

Fingerprinting algorithm can be implemented in various ways from mathematical point of view. They can be divided into deterministic and probabilistic algorithms. In

N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): ACIIDS 2011, LNAI 6592, pp. 452–461, 2011.

our work we deal with fingerprinting based on deterministic algorithms based on nearest neighbor algorithm (NN), as well as more complicated algorithms k-nearest neighbours (KNN) and weighted k-nearest neighbours (WKNN) [13].

Most of researchers dealing with deterministic fingerprinting algorithms use Euclidean distance as similarity measurements between vector of RSSI (Received Signal Strength Information) collected in on-line phase and vectors stored in radio map. We try to compare Euclidean distance with another six similarity measurements and find which one is the best solution for deterministic fingerprinting localization.

Rest of paper is organized as follows. Section 2 describes related work on fingerprinting localization algorithms. In Section 3 different metrics used in our simulations are described. Simulation model created in Matlab software tool and simulation scenarios are presented in Section 4. In Section 5 simulation results are shown. Finally, Section 5 concludes the paper and provides directions for future work.

2 Related Work

The fingerprinting localization algorithms can be divided in two phases – off-line phase and on-line phase. In off-line phase radio map is created in area, where localization will be performed. In on-line phase position of mobile nodes is estimated.

Radio map construction starts by dividing area of interest into cells [14]. Each cell is represented by one reference point. In this point RSSI value from all transmitters in range – fingerprint is measured for certain period of time and stored in database.

Most of researchers in field of fingerprinting localization use deterministic approach of localization. In deterministic approach position of mobile node is computed as combination of radio map points, using:

$$\bar{x} = \sum_{i=1}^{M} \left(\omega_i \cdot p_i \Big/ \sum_{j=1}^{M} \omega_j \right), \qquad (1)$$

where p_i are coordinates of i-th reference point in radio map, ω_i and ω_j are weights and M is number of reference points stored in radio map.

The mathematical algorithm, which keeps the K biggest weights and sets the others to zero is called the WKNN (Weighted K-Nearest Neighbor) [8]. WKNN with all weights $\omega_i = 1$ is called the KNN (K-Nearest Neighbor) algorithm [14]. The simplest algorithm, where $K = 1$, is called the NN (Nearest Neighbor) [15].

One of possible weight computation is the inverse of distance between two RSSI vectors. Most of authors use Euclidean distance [8, 13-16], Junyang Zhou et al use Mahalanobis distance [17] and Binghao Li et al introduces generalized Minkowski distance [14].

3 Similarity Measurements

In this section similarity measurements that will be used in simulations are introduced. First three distances are from Minkowski distance family, next two distances belongs to L1 family, also called the absolute difference [18]. All of these distances measure difference of two vectors. Last two types of measurements are based on correlation, which means that they measure similarity between two vectors.

3.1 Manhattan Distance

Manhattan distance is also known as city block distance, boxcar distance or absolute value distance [19]. It represents distance between points in a city road grid. It examines the absolute differences between coordinates of a pair of objects, or simply vectors. City Block distance is given by:

$$d_{Mij} = \sum_{k=1}^{n} \left| a_{ik} - b_{jk} \right| , \tag{2}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector $\boldsymbol{A}$ and b_{jk} represents k-th element of vector $\boldsymbol{B}$.

3.2 Euclidean Distance

Euclidean Distance is the most common use of distance. In most cases when people said about distance, they will refer to Euclidean distance. Euclidean distance or simply 'distance' examines the root of square differences between coordinates of a pair of objects. Euclidean distance is given by (3) and represents shortest distance between two vectors in Cartesian coordinate system.

$$d_{Eij} = \sqrt{\sum_{k=1}^{n} \left(a_{ik} - b_{jk} \right)^2} . \tag{3}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector $\boldsymbol{A}$ and b_{jk} represents k-th element of vector $\boldsymbol{B}$.

3.3 Minkowski Distance

Minkowski distance is the generalized metric distance, it is given by (4). When $m = 1$ it becomes city block distance and when $m = 2$, it becomes Euclidean distance. This distance can be used for both ordinal and quantitative variables.

$$d_{Wij} = \sqrt[m]{\sum_{k=1}^{n} \left(a_{ik} - b_{jk} \right)^m} . \tag{4}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector $\boldsymbol{A}$ and b_{jk} represents k-th element of vector $\boldsymbol{B}$ and m is root level.

3.4 Canberra Distance

Canberra distance examines the sum of series of a fraction differences between two vectors. Each term of fraction difference has value between 0 and 1. If one of coordinate is zero, the term become unity regardless the other value, thus the distance will not be affected.

$$d_{Cij} = \sum_{k=1}^{n} \frac{\left| a_{ik} - b_{jk} \right|}{\left| a_{ik} \right| + \left| b_{jk} \right|} . \tag{5}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector ***A*** and b_{jk} represents k-th element of vector ***B***.

Note that if both elements are zeros, we need to be defined as 0/0=0. This distance is very sensitive to a small change when both elements are near to zero.

3.5 Sorensen Distance

Sorensen distance is sometimes also called Bray Curtis distance. It is in fact a normalization method that is commonly used in many science fields. It views the space as grid similar to the city block distance. Sorensen distance is given by (6) and has a nice property that if all elements are positive, its value is between zero and one. Zero Sorensen distance represents exact similar vectors.

$$d_{Sij} = \sum_{k=1}^{n} \left| a_{ik} - b_{jk} \right| \Big/ \sum_{k=1}^{n} \left(a_{ik} + b_{jk} \right) . \tag{6}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector ***A*** and b_{jk} represents k-th element of vector ***B***.

If both vectors have zero elements, the Sorensen distance is undefined. The normalization is done using absolute difference divided by the summation.

3.6 Angular Separation

Angular separation represents cosine angle between two vectors. It measures similarity rather than distance or dissimilarity. Thus, higher value of Angular separation indicates the two objects are similar. Angular separation is given by:

$$s_{Aij} = \sum_{k=1}^{n} a_{ik} \cdot b_{jk} \Bigg/ \left(\sum_{k=1}^{n} a_{ik}^2 \cdot \sum_{r=1}^{n} b_{jr}^2 \right)^{\frac{1}{2}} . \tag{7}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector ***A*** and b_{jr} represents r-th element of vector ***B***.

The value of angular separation is [-1, 1] similar to cosine. It is often called as Coefficient of Correlation.

3.7 Correlation Coefficient

Correlation coefficient is standardized angular separation by centering the vectors to its mean value. The value is between -1 and +1. Same as angular separation it measures similarity rather than distance or dissimilarity. Correlation coefficient is given by:

$$s_{Cij} = \sum_{k=1}^{n} (a_{ik} - \bar{a}_i) \cdot (b_{jk} - \bar{b}_j) \Bigg/ \left(\sum_{k=1}^{n} (a_{ik} - \bar{a}_i)^2 \cdot \sum_{r=1}^{n} (b_{jr} - \bar{b}_j)^2 \right)^{\frac{1}{2}}, \tag{8}$$

Where n is number of elements in vector, a_{ik} represents k-th element of vector $\boldsymbol{A}$ and b_{jr} represents k-th element of vector $\boldsymbol{B}$ and $\bar{a}_i$ and $\bar{b}_j$ are mean values of vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively.

Correlation coefficient measures the strength and the direction of a linear relationship between two vectors.

4 Simulation Model

Simulation model created in Matlab software tool was used for investigation of impact of similarity measurement on localization error. Fingerprinting is based on signal strength measurements, therefore simulation model can be divided into two parts: radio channel and fingerprinting method. Three mathematical algorithms introduced in section 2 – NN, KNN and WKNN were implemented in the simulation model.

Received signal strength is modelled by two independent parts: path-loss and immediate variations of signal strength. Path-loss is based on multi-wall-and-floor model (MWF). The MWF model considers the nonlinear relationship between the cumulative penetration loss and the number of penetrated floors and walls. Total loss L_{WMF} in distance d can be computed from equation:

$$L_{MWF} = L_0 + 10n\log(d) + \sum_{i=1}^{I}\sum_{k=1}^{K_{wi}} L_{wik} + \sum_{j=1}^{J}\sum_{k=1}^{K_{fj}} L_{fjk}, \tag{9}$$

Where L_0 is path loss in distance of 1m in dB, n is power decay index, d is distance between transceiver and receiver in meters, I is number of walls types, K_{wi} is number of traversed walls of category i, L_{wik} is attenuation due to wall type i and k-th traversed wall in dB, J stands for number of floor types, K_{fj} is number of traversed walls of category j and L_{fjk} represents attenuation due to wall type i and k-th traversed wall in dB.

Immediate variations of signal strength could be caused by objects motion at observed area. These variations influence RSSI measurements and add measurement error. Behavior of the variations was derived from experimental measurements. Experimental measurements were performed on notebook Asus N series, with use of WirelessMon software. RSSI values from access point (AP) in range were measured 400 times. Achieved results are depicted in Fig. 1 (up chart).

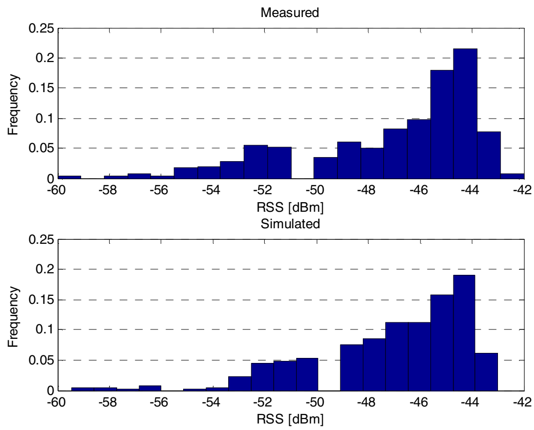

Fig. 1. Histogram of simulated and measured RSSI

According achieved measured data, immediate variations of RSSI were simulated as random variable E computed as product of two random variables with lognormal and uniform distribution respectively. Histogram of 400 simulated RSSI values is shown in Fig. 1.

Simulations of similarity measurements were done in two different scenarios. In Scenario 1 we simulated fingerprinting localization in normal conditions. This means that in on-line and off-line phase same propagation conditions were assumed. We used 6 access points to cover area of 512 square meters. Reference points were chosen in grid with 2 m distance between them. Position of mobile node is randomly chosen from all points in area. In this simulation measured RSSI values in both on-line and off-line phase were simulated with use of MWF affected with random variable E.

The problem of fading was partially eliminated by estimation of local average power in both scenarios. It is calculated as

$$\overline{RSSI} = \frac{1}{N_s} \sum_{i=1}^{N_S} RSSI_i \, , \tag{10}$$

where N_s is number of samples, in this case $N_s = 20$ was used.

In on-line phase all mathematical algorithms were used with combination of all similarity measurements. In situations when KNN and WKNN algorithms were used number of used reference point was set to 4.

In Scenario 2 we assume different environment conditions in on-line and off-line phase. Thus this scenario can be marked as extreme conditions scenario. In off-line phase, when radio map is created, fading error is simulated as random variable with uniform distribution with values from -4 dBm to 20 dBm, and in on-line phase the same distribution as in previous scenario was used. In this case, $N_s = 5$ (local average power), so fading problem is not eliminated well. All the other simulation properties were same as in Scenario 1. Simulations in both scenarios were performed with 10000 independent repetitions.

5 Simulation Results

Simulations results are introduced in this section. The results provide detailed analysis of positioning accuracy in terms of root mean square error (RMSE). The RMSE is calculated as follows:

$$RMSE = \sqrt{(x_r - x_L)^2 + (y_r - y_L)^2}\ , \tag{11}$$

where $[x_r, y_r]$ are coordinates of real (accurate) MS position and $[x_L, y_L]$ are estimated coordinates of MS computed by given mathematical algorithm.

In Fig. 2, simulation results for Scenario 1 can be seen.

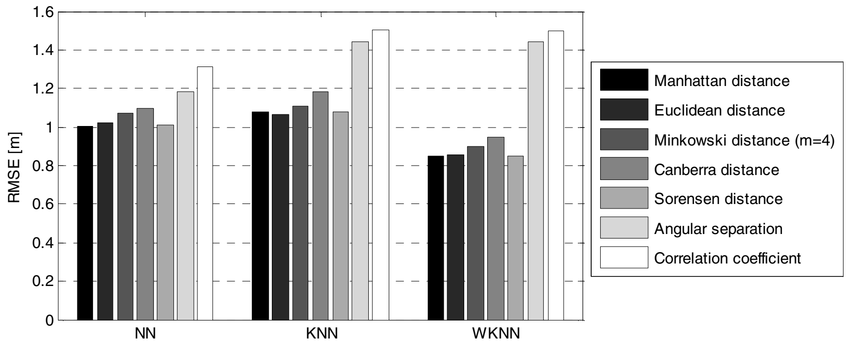

Fig. 2. Mean RMSE values for different algorithms and similarity measurements

On the basis of shown results it can be noted that particular similarity measurements have impact on RMSE regardless of mathematical algorithm. It is evident that Angular separation and Correlation coefficient metrics achieved the worst results compare with other similarity measurements. Results of remaining five metrics are almost same for individual mathematic algorithms. On the other hand, following similarity measurements Manhattan distance, Euclidean distance and Sorensen distance obtained a little bit higher accuracy. Euclidean distance performs slightly worse in combination with NN algorithm.

Difference between observed mathematical algorithms is not big from global point of view. WKNN algorithm achieved the best results (the smallest positioning error), RMSE is approximately 15 % lower. Positioning results for NN and KNN is almost same.

It is known that Manhattan and Euclidean distances are special cases of Minkowski distance. Hence, next simulation was designed to find out how Minkowski distance is affected by root coefficient *m*. Simulation results are shown in Fig. 3. It is clear that the best results are achieved in case of $m = 1.5$ for all algorithms. The greater the root coefficient *m,* the higher is the positioning error. This simulation confirms fact from previous one that WKNN is the most accurate mathematical algorithm.

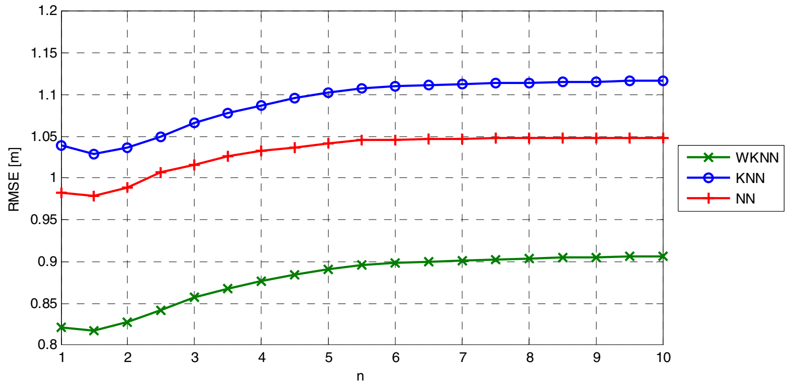

Fig. 3. RMSE values for different algorithms and root coefficient of Minkowski distance

Results of simulations in the extreme case (Scenario 2) are shown in Fig. 4. From results can be seen that best results can be achieved by same distances as in ideal case. Euclidean distance performs slightly better in combination with NN algorithm, but in more sophisticated KNN and WKNN algorithms performance of Manhattan, Euclidean, Minkowski and Sorensen distances is almost the same. Angular separation and correlation coefficient shows the worst results.

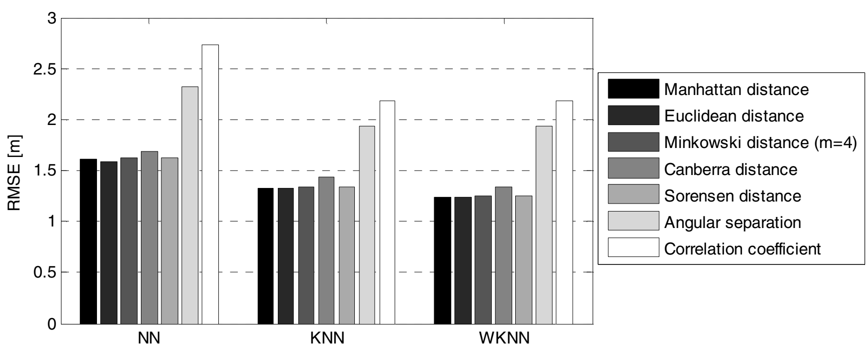

Fig. 4. Mean RMSE values for different algorithms and similarity measurements under extreme conditions

According the simulations results can be assumed that Manhattan and Sorensen distances performs well in both normal and extreme conditions, Euclidean distance performs almost same as those, difference is only with use of NN algorithm.

Last simulation results are aimed to reveal complexity of similarity measurement methods. In Fig. 5 mean computing time of each similarity measurement can be seen. From the figure it is clear that lowest complexity has Manhattan, Canberra and Sorensen distances.

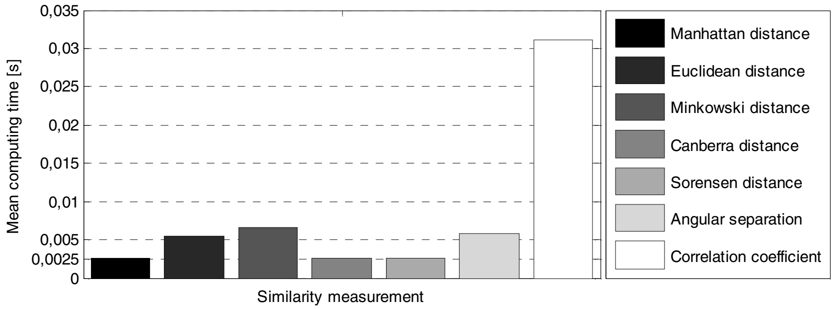

Fig. 5. Mean computing time of similarity measurements

Low complexity is very important in situations, when localization is offered for wide area with high number of users. If system must localize all of them in real time it needs to use computation methods with the lowest possible complexity.

6 Conclusion and Future Work

From results shown in this paper it is clear that Euclidean distance is not best similarity measurement for fingerprinting localization in WLAN networks. On the basis of achieved simulation results can be assumed that Manhattan and Sorensen distance performs better or same as Euclidean distance. Another advantage of Manhattan and Sorensen distances is lower complexity of computation, so these similarity measurements can be used in localization systems covering wide areas with high number of users, with better results than commonly used Euclidean distance.

For future Sorensen and Manhattan distance will be implemented into real localization system WiFiLOC, to verify results of simulation in real environment. There is also space for modification of localization algorithms and development of new mathematical algorithms to improve localization accuracy.

Acknowledgement. This work was partially supported by the Slovak Research and Development Agency under the contract No. LPP-0126-09 and by the Slovak VEGA grant agency, Project No. 1/0392/10 “The research of mobile nodes in wireless sensor networks”.

References

1. Mohapatra, D., Suma, S.B.: Survey of Location Based Wireless Services. In: IEEE International Conference Personal Wireless Communications 2005, pp. 358–362 (2005)
2. Hui, L., Darabi, H., Banarjee, P., Jing, L.: Survey of Wireless Indoor Positioning Techniques and Systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(6), 1067–1080 (2007), ISSN: 1094-6977

3. Chawathe, S.S.: Low-latency Indoor Localization Using Bluetooth Beacons. In: 12th International IEEE Conference Intelligent Transportation Systems, ITSC 2009, pp. 1–7 (2009), ISBN: 978-1-4244-5519-5
4. Yao, Z., Liang, D., Jiang, W., Hu, B., Fu, Y.: Implementing Indoor Positioning System via ZigBee Devices. In: Signals, Systems and Computers 2008, pp. 1867–1871 (2008)
5. Zheng, L., Dehaene, W., Gielen, G.: A 3-Tier UWB-based Indoor Localization Scheme for Ultra-low-powersensor Nodes. In: IEEE International Conference Signal Processing and Communications, ICSPC 2007, pp. 995–998 (2007), ISBN: 978-1-4244-1235-8
6. Bai, Y., Lu, X.: Research on UWB Indoor Positioning Based on TDOA Technique. In: Electronic Measurement & Instruments ICEMI 2009, pp. 167–170 (2009)
7. Ni, L.M., Liu, Y., Lau, Y.C., Patil, A.P.: LANDMARC: Indoor Location Sensing Using Active RFID. Wireless Netw. 10(6), 701–710 (2004)
8. Bahl, P., Padmanabhan, V.N.: RADAR: An in-building RF-based User Location and Tracking System. In: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 775–784 (2000) ISBN: 0-7803-5880-5
9. Krejcar, O.: Problem Solving of Low Data Throughput on Mobile Devices by Artefacts Prebuffering. EURASIP Journal on Wireless Communications and Networking 2009, Article ID 802523, 8 (2009), doi:10.1155/2009/802523, ISSN: 1687-1499
10. IEEE Standard for Information Technology-telecommunications and Information Exchange Between Systems-local and Metropolitan Area Networks-specific Requirements - part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) Specifications, IEEE Standard 802.11-2007 (2007), ISBN: 978-0-7381-5656-9
11. Machaj, J., Brida, P., Tatarova, B.: Impact of the Number of Access Points in Indoor Fingerprinting Localization. In: 20th International Conference Radioelektronika, Radioelektronika 2010, pp. 83–86 (2010) ISBN 978-1-4244-6320-6
12. Krejcar, O., Frischer, R.: Detection of the Internal Defects of Material on the Basis of the Performance Spectral Density Analysis. Journal of Vibroengineering 12(4), 541–551 (2010) ISSN: 1392-8716
13. Tsung-Nan, L., Po-Chiang, L.: Performance Comparison of Indoor Positioning Techniques based on Location Fingerprinting in Wireless Networks. In: International Conference Wireless Networks, Communications and Mobile Computing 2005, vol. 2, pp. 1569–1574 (2005)
14. Li, B., Salter, J., Dempster, A. G., Rizos, C.: Indoor Positioning Techniques Based on Wireless LAN. Technical Report, School of Surveying and Spatial Information Systems, UNSW, Sydney, Australia (2006)
15. Saha, S., Chauhuri, K., Sanghi, D., Bhagwat, P.: Location Determination of a Mobile Device Using IEEE 802.11b access point signals. In: Wireless Communications and Networking, WCNC 2003, vol. 3, pp. 1987–1992 (2003), ISBN: 0-7803-7700-1
16. Yeung, W.M., Ng, J.K.: An Enhanced Wireless LAN Positioning Algorithm Based on the Fingerprint Approach. In: IEEE Region 10 Conference TENCON 2006, pp. 1–4 (2006)
17. Junyang, Z., Yeung, W.M.-C., Ng, J.K.-Y.: Enhancing Indoor Positioning Accuracy by Utilizing Signals from Both the Mobile Phone Network and the Wireless Local Area Network. In: 22nd International Conference on Advanced Information Networking and Applications, AINA 2008, pp. 138–145 (2008)
18. Cha, S.-H.: Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. International Journal of Mathematical Models and Methods in Applied Science 1(4) (2007)
19. Krause, E.F.: Taxicab Geometry: An Adventure in Non-Euclidean Geometry. Dover Publications, Inc., New York (1986) ISBN: 0-486-25202-7