Sensors & Transducers, Vol. 160, Issue 12, December 2013, pp. 256-261

www.sensorsportal.com

Sensors & Transducers

© 2013 by IFSA

http://www.sensorsportal.com

Based on the Wireless Sensor Network Localization Algorithm of Locality Preserving Canonical Correlation Analysis

Hao YANCHUN

Jilin Railway Vocational and Technology College,
No. 43, Zhongxing Street, Changyi District, Jilin City, 132101, Jilin Province
E-mail: hao_spring@163.com

Received: 7 November 2013 /Accepted: 22 November 2013 /Published: 30 December 2013

Abstract: Wireless sensor network localization involves the dataset of physical space and signal space, and its layout accord with the characteristic of manifold. This paper uses the own characteristic of topological structure and signal space ,striving to search a right modeling approach to realize the mapping between these two spaces, thus the usage of locality preserving canonical correlation analysis (LPCCA)is put forward to realize the manifold structure and localization. To further improve the accuracy of localization algorithm, through combining specific positioning problem as regression rather than dimensionality reduction to adapt new application, developed an original wireless sensor network localization algorithm LE-LPCCA (location estimation-LPCCA) which can reflect network topological structure information. *Copyright © 2013 IFSA.*

Keyword: Wireless sensor network, Localization, Algorithm.

1. Introduction

The self-positioning of sensor node is the foundation of wireless sensor network application, the positioning process of nodes are based on statistic model, extracted characteristic quantities which can represent the coordinate position of sensor nodes to establish models and on this account to confirm positioning algorithm. And then according to the algorithm to estimate the position of unknown sensor nodes to calculate its space (location) coordinate. But now the canonical localization algorithm has not taken the full use of the correlation between network canonical topological structure and its signal space, the positional accuracy is remained to be improved.

This paper tries to use the topological structure of wireless sensor network and its own characteristic of signal space to further improve the accuracy of localization algorithm. At present the research work in this respect is relative small, especially in the domestic. Reference [7] puts forward algorithm LE-KCCA to get the required mapping by the realization of kernel canonical correlation analysis (KCCA).But the calculation of KCCA adapts a uniform nonlinearity mapping and without scrupling the local distribution characteristic of network, that is manifold structure. In addition, the wireless sensor network with complicated and fickle topological structure also makes KCCA difficult to fully reflect the network structure information.

In response to these two shortcomings, this paper proposed a locality preserving canonical correlation analysis [LPCCA] which can efficient handing two set of manifold structure data respectively at the same time as the modeling tool of localization algorithm, tried to use local information of signal

space and physical space, put forward an original algorithm LE-LPCCA (location-estimation-LPCCA) of wireless sensor network localization based on manifold. Algorithm LE-LEPPA with a view to the canonical correlation of each neighborhood node, through decomposing global nonlinear problems into multiple local linear problems to get overall correlation. This algorithm by means of introducing the global and local information of network distribution, achieved a double effects with significantly improve node localization accuracy and stability.

2. Based on Manifold Learning Algorithm LE-LPCCA

As the sensor nodes distributed in earth surface or physical spaces consists with several local Euclidean space, so the distribution of wireless sensor network accord with the characteristic of manifold. The wireless sensor node arrangement can use for the sampling of manifold information, the more intensive of arrangement, the better of sampling results. With regard to its corresponding signal space, as the scope of signal-radiation in each sensor node is a sphere theoretically, the node only located in sphere can receive the signal launched by this node, which is the same identical with the distribution of Euclidean space.

This paper in view of the limitation on the generalization performance and flexibility of algorithm KCCA, through using the own characteristic of signal space to improve the accuracy of localization algorithm, put forward using LPCCA to realize manifold structure and localization and by combing specific positioning problem as regression rather than dimensionality reduction to adapt new application, developed an original wireless sensor network localization algorithm LE-LPCCA (location estimation-LPCCA) which can reflect network topological structure information.

Fig. 1 shows the neighbor relationship figure of sensor network nodes

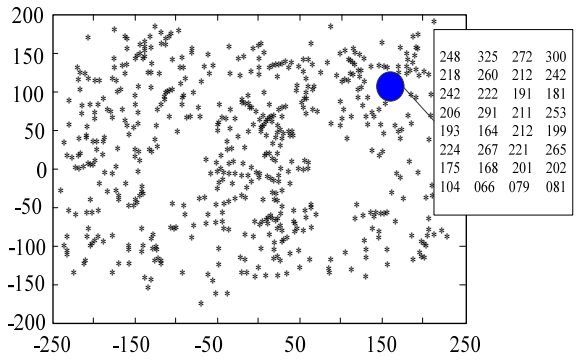

Fig. 1. Neighbor relationship figure of sensor nodes.

Local neighbor adapts k -neighbor definition method, he valuation of K is 4, the 4 nodes within the scope of circle can receive the signal launched by 8 beacon nodes, the critical space distribution gathered together, the figure shows the corresponding received signal strength are close. For example, in the first line of the matrix in this dataset, means 4 nodes received the signal launched by No. 1 signal indication, the signal strength is about 300. As the received signal strength actually is negative number, unit: db, in order to facility research and experiment, we handle all the data to positive number, the larger of the numerical value, the stronger of the signal.

3. The Formation and Evolution of Algorithm LE-LPCCA

3.1. CCA

Given there are a batch of pairs sample observation X and Y, suppose

$$X=[x_1,\cdots,x_n]\in R^{p\times n}$$

and $X=[y_1,\cdots,y_n]\in R^{q\times n}$,

and the samples have been centralization. The object of algorithm CCA is seeking two sets of base vector $w_x\in R^p$ and $w_y\in R^q$ for X and Y, respectively, so that the correlation of random variable between $w_x^T x$ and $w_y^T x$ maximize. In other word, after a certain conversion of original data X and Y, to achieve the maximum correlation. Specially can indicate to the question of maximizing correlation coefficient of algorithm (1) [5]:

$$\rho=\frac{E[w_x^T xy^T w_y]}{\sqrt{E[w_x^T xx^T w_y]}}=\frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}, \quad (1)$$

There $E[]$ is mathematical expectation, $C_{xx}=XX^T\in R^{p\times p}$ and $C_{yy}=YY^T\in R^{q\times q}$ is with-set covariance matrix, $C_{xy}=XY^T\in R^{p\times q}$ is between-set covariance matrix, and $C_{yx}=C_{xy}^T\in R^{p\times q}$.

Pay attention to the correlation function ρ defined by algorithm (1) is not relevant to the size of w_x and w_y, so CCA can express as the solution of formula (2) optimization problem:

$$\left.\begin{aligned}&\max_{w_x,w_y} w_x C_{xy} w_y\\ &s.t. w_x C_{xx} w_x=1, w_y C_{yy} w_y=1\end{aligned}\right\} \quad (2)$$

Reference [5] gives a method to transform the CCA equation (2) solving to characteristic equation (3) solving:

$$C_{xy}C_{yy}^{-1}C_{yx}w_x = \lambda^2 C_{xx}w_x, \quad (3)$$

Find the solution of base vector w_x and eigenvalue λ, then $w_y = \frac{1}{\lambda}C_{yy}^{-1}C_{yx}w_x$. Usually, we can obtain a group of characteristic vector (base vector) to $(w_{x_i}, w_{y_i})(i = 1, \cdots, d)$, and can combine them to two mapping matrix: $W_x = [w_{x_1}, \cdots, w_{x_d}]$ and $W_y = [w_{y_1}, \cdots, w_{y_d}]$. We can use any set of w_{x_i} and w_{y_i} and process a characteristic extract such as $w_{x_i}^T x$ and $w_{y_i}^T x$, there $w_{x_i}^T x$ and $w_{y_i}^T x$ can be named as canonical variate.

3.2. Algorithm LPCCA

Algorithm LPCCA introduces the local structure information of samples to CCA, transforms the original global nonlinear problems to several local linear problems, calculated canonical correlation problems of each neighbor area; and the find the solution of these subproblems, gained a group of sample projection base vector through optimal computation. Therefore, through local linear method to achieve the aim of solving nonlinear problems. This method maintains the local structure information of various samples, at the same time gained the maximize correlation information between these two groups of samples.

Algorithm LPCCA defined similarity matrix $C_X = \{C_{ij}^X\}_{i,j=1}^n$ and $C_Y = \{C_{ij}^Y\}_{i,j=1}^n$, which including the local information of data distribution, that is when the larger C_{ij}^X （or C_{ij}^Y） is, it means the closer the x_i and x_j （or y_i and y_j） is. So the optimization problem of LPCCA can be expressed as algorithm (4) and shows as [9]:

$$\left.\begin{array}{l} \max\limits_{w_x, w_y} w_x^T \mathbf{X} C_{XY} \mathbf{Y}^T w_y \\ s.t. w_x^T \mathbf{X} C_{XX} \mathbf{X}^T w_x = 1 \\ w_y^T \mathbf{Y} C_{YY} \mathbf{Y}^T w_y = 1 \end{array}\right\}, \quad (4)$$

There, $\mathbf{X} = \{x_1, \cdots, x_n\}$, $\mathbf{Y} = \{y_1, \cdots, y_n\}$, $C_{XY} = D_{XY} - C_{X\circ}C_Y$, $C_{XX} = D_{XX} - C_{X\circ}C_X$, $C_{YY} = D_{YY} - C_{Y\circ}C_Y$, symbol ∘ suppose as operator; $D_{XX}(D_{XY}, D_{YY})$ is diagonal matrix with the size of $n \times n$, the i diagonal element is equal to the sum of the i line element (because its symmetry, or the i line) of matrix $C_X \circ C_X (C_X \circ C_Y, C_Y \circ C_Y)$.

On the usage of Lagrange multiplier method, find the solution of merit function (4), we can easily get generalized eigenvalue equation as follows:

$$\begin{pmatrix} & \mathbf{X}C_{XY}\mathbf{Y}^{\mathbf{T}} \\ \mathbf{Y}C_{YX}\mathbf{X}^{\mathbf{T}} & \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{X}C_{XX}\mathbf{X}^{\mathbf{T}} & \\ & \mathbf{Y}C_{YY}\mathbf{Y}^{\mathbf{T}} \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} \quad (5)$$

After getting base vector (w_x, w_y), we can process linear transformation of original data such as $w_x^T x$ and $w_y^T y$. Attention, LPCCA have no need for centralization of data, which is different from CCA and KCCA. Throughout the calculation process, LPCCA keep localization divisor C_{ij}^X, C_{ij}^Y unchanged, and tried to preserve neighborhood data point of the original space neighborhood after transformation [4].

4. Algorithm LE-LPCCA

The self-positioning of sensor nodes is the basic of wireless sensor network application, the common positioning technology is RSSI (received signal strength indicator) technology [1]. On the mechanism of RSSI-based wireless sensor network positioning, the localization of nodes usually divided into offline training phase and online localization phase [2]. In the training phase, through the training of known node data to learn positioning algorithm; in the localization phase, using the learned algorithm to calculate the position of unknown nodes. Suppose $S = [s_1, s_2, \cdots, s_n] \in R^{m \times n}$ is the signal strength received by the N wireless sensor known nodes, where the dimension of each signal vector $s_i (i = 1, \cdots, n)$ is M, which means have M beacon node. If the $j(j = 1, \cdots, m)$ beacon node located beyond the receive scope of i point, then $s_{ij} = 0$. $L = [l_1, l_2, \cdots, l_n] \in R^{2 \times n}$ is the physical coordinate of corresponding nodes. The object of algorithm design is: through the signal strength $s_g = (s_{g1}, s_{g2}, \cdots, s_{gm})'$ received by unknown node g to calculate the position coordinate y_g of g point.

Definition 1. In signal space, $ne(i)$ means node label set similar with the signal strength received by node i, that is the local neighborhood sample set under X; in physical space, $ne(i)$ means the node set which neighbor to node i. Where the local neighbor is divided by K-neighborhood definition method: if $x_j (y_j)$ is the K-neighborhood sample of $x_i (y_i)$,

then means $x_j(y_j)$ is the local neighbor of $x_i(y_i)$. According to the above description, the similarity matrix of wireless sensor network topological structure on signal space and physical space is $C_s = \{C_{ij}^S\}_{i,j=1}^n$ and $C_L = \{C_{ij}^L\}_{i,j=1}^n$, where, matrix elements[4]:

$$C_{ij}^S = \begin{cases} \exp(-\|s_i - s_j\|^2 / t_s), j \in ne(i) / i \in ne(j) \\ 0, other \end{cases}$$
$$C_{ij}^L = \begin{cases} \exp(-\|l_i - l_j\|^2 / t_l), j \in ne(i) / i \in ne(j) \\ 0, other \end{cases} \quad (6)$$

Parameter t_s is the average distance $\sum_{i=1}^n \sum_{j=1}^n \|s_i - s_j\|^2 / n(n-1)$ of signal space, parameter t_l also do similar processing. Therefore, similarity matrix including the local information of network topological structure, namely C_{ij}^S (or C_{ij}^L) rely on the layout of sensor nodes. Therefore, C_{ij}^S (or C_{ij}^L) can varies from network topological structure so that reflect its high flexibility.

According to the above description, we can obtain optimization problem as follows:

$$\left.\begin{array}{c} \max\limits_{w_S, w_L} w_S^T S C_{SL} L^T w_L \\ s.t. w_S^T S C_{SS} S^T w_S = 1 \\ w_L^T L C_{LL} L^T w_L = 1 \end{array}\right\}, \quad (7)$$

where $C_{SL} = D_{SL} - C_S \circ C_L$, $C_{SS} = D_{SS} - C_S \circ C_S$, $C_{LL} = D_{LL} - C_L \circ C_L$ use the Lagrange multiplier method and elimination method to find the solution of the above merit function set, we can get mapping matrix $\mathbf{W_S}$, $\mathbf{W_l}$.

The following are the concrete steps of algorithm LE-LPCCA, where Step1~Step3 is completed in training phase, Step4~Step6 is completed in localization phase.

Input: The intensity matrix received of known nodes are $S = (s_1, s_2, \cdots, s_n)$, $L = (l_1, l_2, \cdots l_n)$; the signal vector received of unknown node g is $s_g = (s_{g1}, s_{g2}, \cdots, s_{gm})'$.

Output: The position coordinate of point g.

Step1.By algorithm(6),calculates similarity matrix C_S and C_L.

Step2. By calculating algorithm (7),processing a locality preserving canonical correlation analysis to collected data of two spaces, learning the relationship between physical space and signal space, finding the solution of mapping matrix W_S, W_l,and generalized eigenvalue λ.

Step 3. Make a linear transformation of known data $\{S, L\}$: $\{P^S, P^L\} = \{W_S^T \times S, W_L^T \times L\}$, namely make a mapping to the data of original space, so that maximize the correlation between P^S and P^L, as well as transform the data of original dimension to two-dimension.

Step 4. About the signal vector $s_g = (s_{g1}, s_{g2}, \cdots, s_{gm})'$ received of unknown node g, uses W_S to make linear transformation and obtain $P^g = W_S^T s_g$.

Step 5. After calculating point g and making linear transformation of all known node, obtained Euclidean distance with authority[3]:

$$d_i = \sum_{j=1}^{2} \lambda_j (P_{ji}^S - P_j^g)^2, i = 1, \cdots, n, \quad (8)$$

where d_i means the distance between the i node to point g; λ_{ij} is the j component of the i eigenvalue. which means the correlation of the j largest, P_{ji}^S means the j component of the i node after linear transformation, P_j means the j component of point g after linear transformation.

Step6. Calculating the minimum value of $d_i(i, \cdots, n)$ of K-neighborhood nodes in point g and its corresponding K node (namely the value of i). From the characteristic of neighborhood preserving of the data transformation of LPCCA, the closer of the distance between signal space, the shorter of the distance between corresponding physical space. Therefore, find the physical coordinate $\{y_1, y_2, \cdots, y_K\}$ of node K, and make its centroid be the physical coordinate of node g:

$$y_g = \frac{y_1 + y_2 + \cdots + y_K}{K}, \quad (9)$$

At this moment y_g is the position and calculation result of node g.

5. Experiments

5.1. Description of Experiments.

Experimental data come from the computer science department of Hong Kong university of science and technology. Experimental setup is in the building of this department and structured by Cross-bow MICA2. Sensor node deploys in the indoor with four corridors, and distributed eight signal

nodes(including non-coplanar). Therefore, the signal strength vector received of each sensor node is eight dimensions, namely simultaneously receive the signal transmitted of eight beacon node. Experimental facility is an IEEE 802.11b wireless network with 2.4 GHz frequency bandwidth. All the data are collected through IBM laptop linking to an external wireless USB network adapter [3].

This experiment divided the entire network plane into several grids with the size of 1.5 m × 1.5 m. Each grid collects 13~14 sample point, altogether have 2999 samples. A part of which are used fro training samples, to learn positioning algorithm; and the rest are used fro testing samples, to calculate the accuracy rate of positioning.

The localization errors of this experiment are represented by δ ,and experimental accuracy are represented by the following three indicators respectively: $\delta \leq 1.5$ m (namely the nodes distributed in the same grid), $\delta \leq 3.0$ m (namely the node can distribute in the neighborhood grid) and average error $\overline{\delta}$ of localization calculation. For convenience, we also use LE-KCCA, CCA and the location algorithm LeMan [7] based on manifold normalization method. The selection of LE-KCCA and LeMan both are Gaussian kernel, and the kernel parameter is 0.05 [7]. Experimental environment is processing in MATLAB7.0. Data sample has been made normalization processing before experiment. In order to guarantee the accuracy of result, each group of experiment all operated more than 20 times.

5.2. Analysis of Experimental Results

Fig. 2 shows the localization accuracy of four kinds of algorithm when the training sample (namely known nodes) selects 10 % ~ 70 %. The selection of these experimental training samples all from random selection, but each time the data with same proportion applying to 4 algorithms is the same.

From the Fig. 2 we can see that the position result of algorithm LE-LPCCA shows a bigger advantage. As for $\delta \leq 1.5$ m (as Fig. 2 (a) shows), the accuracy rate of positioning maintain at about 85 %; for $\delta \leq 3.0$ m (as Fig. 2 (b) shows), the rate maintain at about 100%; the average error $\overline{\delta}$ of location computed is minimum, maintain under about 1 m (as Fig. 2(c) shows). From Fig. 2 we also see that with the increase of training sample, the position result is more stability.

When algorithm LeMan ate the moment of $\delta \leq 1.5$ m (as Fig. 2(a) shows), the best accuracy rate of location is about 78.92 %, the average error (as Fig. 2(c) shows) all more than 1 m. Although this algorithm based on manifold method, it only considers the manifold structure of signal space, and the correlation strength of two dataset is inferior to LPCCA. In addition, this algorithm may be more applicable to the spaces with relatively intensive data collection, the node density of reference [3] is about 2.34 times than this paper.

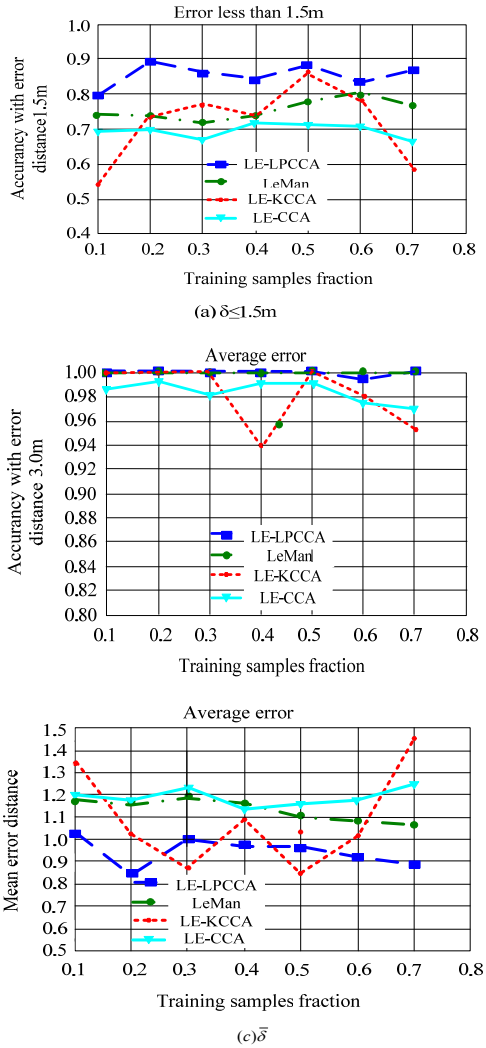

Fig. 2. When the proportion of training sample gradually increases, the comparison of the computational result of these four algorithms.

For algorithm LE-KCCA, the accuracy rate of position result is not very stability. Only when the training samples are about 50 %, the location results will be better. The accuracy rate is 86.02 % when δ≤1.5 m; the accuracy rate is close to 100 % when δ≤3.0 m; and the average error is 0.873 m. But because of the distribution of training samples is random selection, it can not guarantee of completely uniform distribution (It stays the same with practical application), and the selection of kernel parameter and kernel function may not appropriate. Therefore, the result is rather instability when uses algorithm LE-KCCA of whole nonlinear to calculate. After observing the original training data obtained randomly, we discover when the distribution of training sample is uniformity and the number is relatively large, the location result of algorithm LE-KCCA is relatively better, the reverse is different.

The accuracy of algorithm LE-CCA is relatively low, the accuracy rate only maintain at 70 % within the error of 1.5 m; and average error only at 1.2 m. This is because itself is only the global linear mapping of two group of spaces, LE-CCA is no longer applicable to complex nonlinear space; on the other hand, its simplification also make it relative stable.

5.3. The Impact of Parameter on Positioning

In the algorithm LE-LPCCA, the main parameter selection is the assignment of K value in K-neighborhood method used to $ne(i)$ calculation. Suppose the distribution of sensor nodes are uniformity, Fig. 3 shows under the situation of training samples and test samples each about 50 %, the change of positional accuracy with the different value of K.

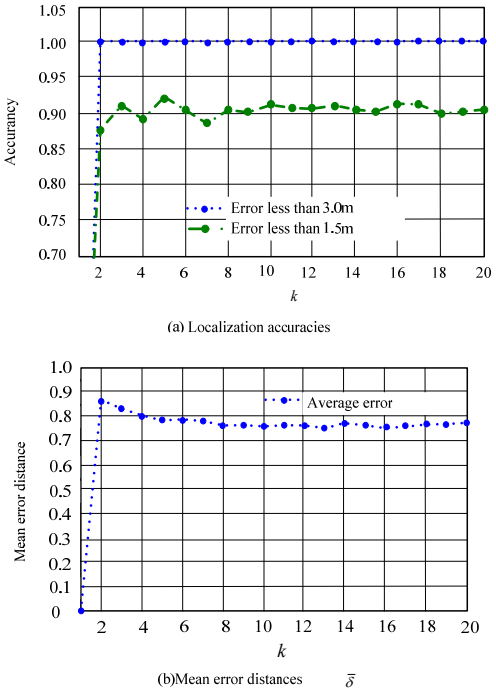

Fig. 3. (on the test set). The change of positional accuracy and average error with K.

When K values 0 or 1, network topological structure can not express well, so we value form k=2. When K valued 2 and gradually increases. as shows in Fig. 3 (b) the average error of position is gradually decreased. The more neighborhood nodes are, the more accurate the position is. But when k valued 8 or more than 8, the change is relative small, the average error is about 0.75 m, and its result tend to be stable. The accuracy rate of $\delta \leq 3.0$ m always is 100 %, while the accuracy rate of $\delta \leq 1.5$ also maintains more than 85 %. Through different experiments of parameter selection, with the local information of wireless sensor network as the focused investigation of algorithm LE-LPCCA, and with well stability to different K, which shows it's not very sensitive to K.

6. Conclusion

The location algorithm LE-LPCCA proposed in this paper through the usage of the local information of wireless sensor network structure, utilizes manifold method to establish the mapping from signal space to physical space, which can efficient handling the complex problem of topological structure in wireless sensor network. Compared with similar algorithm, algorithm LE-LPCCA has the advantages of high accuracy rate and stable performance. But because it finds the solution to local at the beginning and then sum up, so the time consumption is rather high. The further work is to improve and increase efficiency on the basis.

References

[1]. X. L. Nauyen, M. L. Jordan, B. Sinopoli, A kernel-based learning approach to ad hoc sensor network localization, *ACM Transactions on Sensor Networks,* Vol. 1, Issue 1, 2005, pp. 134-152.

[2]. W. Shitong, X. Xiaojiang, L. Weihua, Network business system architecture, standardization and key technical issues, *Communication Technology and Standards,* Vol. 1, 2010, pp. 44-48.

[3]. D. Madigan, P. Krishnan, A. Krishnakum, A. S. Zorych I., Location estimation in wireless networks: A Bayesian approach, *Statistica Sinica*, Vol. 16, Issue 2, 2006, pp. 495-522.

[4]. P. Biswas, T. C. Lian, T. C. Wang, Y. Ye, Semi definite programming based algorithms for sensor network localization. *ACM Transactions on Sensor Networks*, Vol. 2, Issue 2, 2006, pp. 188-220.

[5]. A. Harter, A. Hopper, P. Steggles, A. Wark, P. Webste, The anatomy of a context-aware application, *Wireless Networks*, Vol. 8, Issue 2, 2002, pp. 187-197.

[6]. L. Girod, D. Estrin, Robust range estimation using acoustic and multimodal sensing, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'01),* Vol. 3, Maui: IEEE Robotics and Automation Society, 2001, pp. 1312-1320.

[7]. D. Niculescu, B. Nath, Ad hoc positioning system (APS) using AoA, in *Proceedings of the IEEE INFOCOM*, 2003, pp. 1734-1743.

[8]. W. Fubao, S. Long, R. Fengyuan, The self-positioning system and algorithm in wireless sensor network, *Journal of Software*, Vol. 16, Issue 5, 2005, pp. 857-868.

[9]. S. Limin, L. Jianzhong, C. Yu, Z. Hongsong. Wireless sensor network, *Tsinghua University Press,* Beijing, 2005, pp. 141-146.

2013 Copyright ©, International Frequency Sensor Association (IFSA). All rights reserved.
(http://www.sensorsportal.com)