

# The Hybrid HMM for RSS-based Localization in Wireless Sensor Networks

Zang yanhong, Wang Jinsong, Ling Lin, Lu Peizhong  
 Southwest Inst. of Electron & Telecom. Techn  
 Shanghai, China  
 shisi\_z@163.com, pzlu@fudan.edu.cn

**Abstract**—We propose a method of RSS-base localization in WSN (Wireless Sensor Network), called Hybrid HMM, to improve the stability of node localization based on RSS (Received Signal Strength). This model utilizes HMM (Hidden Markov Model) to take into account the time factor when receiving the RSS sequence, and converts the action of ranging into an operation of classification. For the received RSS used for localization, our Hybrid HMM will compare it with the preset RSS threshold value, and put the result into one of two categories for subsequent processing: If the received value is higher than the threshold value, the distance value will be drawn from the *signal propagation model*. If lower, the information will be obtained from a trained HMM. Experimental results show that the Hybrid HMM method can greatly improve the localization accuracy.

**Keywords**- Wireless Sensor Network; Localization; Received signal strength; Hidden Markov Model

## I INTRODUCTION

WSN is composed of a large number of sensor nodes that autonomously self-organize into a cooperative network. Considering the cost constrains and energy consumption, only few nodes' positions are predetermined. So the position of node is one of the important subjects in WSN, which is useful for coverage, routing, location service, target tracking and rescue. Those nodes whose position are already known called anchor, and the WSN localization is the process to determine the position of the unknown nodes through the information interaction (such as angle, time difference or distance) between the anchor nodes and unknown nodes. Localization algorithms for WSN are plentiful, and we can find a comprehensive review of WSN localization technologies and techniques in [1].

Among numerous localization methods, approaches based on RSS (Received Signal Strength) have drawn much attention with advantages of simplicity and convenience for practical applications. For this reason, we focus on RSS-based localization in this paper.

The localization methods based on RSS fall into two categories: propagation-model based and machine-learning based. The propagation-model based methods will first calculate the distance between the nodes based on the *signal propagation model*, and then use the three-point positioning or maximum likelihood estimation method to figure out the position of the unknown nodes. However, as the RSS is very vulnerable to environmental impact, and the signal attenuation varies in NLOS (Non Line of Sight), these will face the RSS-based localization with great challenges, and many researchers have made a lot of efforts and

improvements on this matter [2][3]. The machine-learning based methods will input the observed data into a trained dynamic model to learn the approximate position of the node. For example, the commonly used fingerprint positioning method collects the RSSI values of the reference locations in the monitored area to construct a database. Then, they will compare the observed RSSI with the database to find the nearest reference position, and take it as the approximate position of the unknown node. Although this mechanism can adapt to different environments and obtain a relatively good localization effect, it requires a lot of offline statistical work in the early stage, and overly massive data can be a challenge for the real-time matching algorithm [4][5].

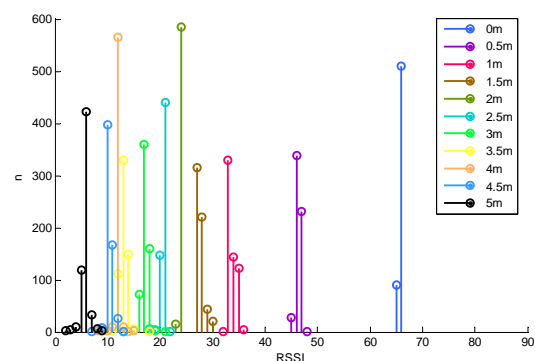


Figure 1 the numbers of RSSI with different distance

Due to the RSS is varying with time, many researchers using HMM to model the localization problem as a classification problem. In [6], John et al., designed a localization system called LOCADIO. It used two HMMs which taking into account the node motion, building's floor plan, expected pedestrian etc. to estimating the position of node. In [7], Carlo proposed an algorithm based on an HMM Bayesian approach that models the node moving capabilities. Simulations show that performances achieved by keeping into consideration mixed LOS/NLOS conditions for all radio links are similar to those obtained in an ideal LOS propagation environment. Vincent et al. [8] carefully model the signal variation with time and employ a semi-supervised HMM framework, which appropriately combine both labeled data and unlabeled data together. It can successfully transfer out-of-date model to fit a current model through learning, even though the training data have very different distributions. Those papers all apply HMM into WSN localization, and achieved high accuracy localization results. Although the above mentioned algorithms are

machine-learning based, the propagation-model based methods are more practical. We propose an algorithm based on propagation-model which uses an HMM to preprocess the RSSI. Our methods only need to filter the environmental noise from RSS in preprocessing process. In order to demonstrate it more intuitively, we use the IRIS node as a transmitting node, and collect 600 RSSI values respectively at several fixed distances. Distance value:  $D = \{d_1, d_2 \dots d_{11}\}$ ,  $d_i = (0.5 * i - 0.5)m$ ,  $d_i$  is held as the *reference distance*. As shown in Figure 1, different colors represent different distances,  $x$  stands for the collected RSSI values, and  $y$  stands for the numbers of RSSI. From the graph, we not only can find that RSSI value decreases as the distance increases, but also can find that when the transmitting node is close to the receiving node, the signal is stable; and when they are far from each other, e.g. at 3m or above as shown in the graph, the RSSI values collected from different distances are partially overlapped. As a result, the distance estimation performance can be grossly inaccurate when the distance between nodes is relatively far.

Therefore, in this paper, we mainly focus on the pretreatment of RSSI values in response to this phenomenon, and on training a dynamic model, called the Hybrid HMM. The main procedures of our method are briefly introduced as following:

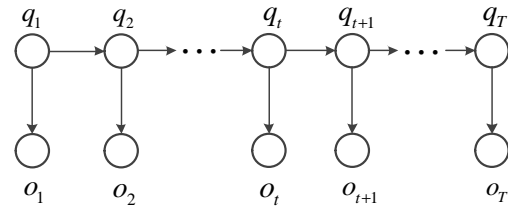
- 1) Find the appropriate RSSI threshold value  $\gamma$  based on the collected data, and use it to divide the collected data into two parts according to distance;
- 2) If  $RSSI > \gamma$ , combine the traditional attenuation equation and the measured RSSI values to get an attenuation curve adapted to the environment, and then calculate the distance value corresponding to the RSSI values;
- 3) If  $RSSI < \gamma$ , use the trained HMM to calculate the distance.

Our goal in this paper is to improve the stability of RSS-based localization. To this end, we introduce an HMM which can preprocess the received RSS. Our main contributions are as follows:

- 1) We propose a novel localization method called Hybrid HMM, which introduce HMM to convert the ranging into classification. Our method improves the stability of node localization when the distances between nodes are relatively far.
- 2) We compare our method with the method based of *signal propagation model*. Experiment shows the presented algorithm performs well and can gain better performance when add random noise.

## II HIDDEN MARKOV MODEL

HMM is a statistical Markov model in which the system being modeled is modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMMs are especially known for their application in temporal pattern recognition such as speech, handwriting, partial discharges and bioinformatics.



$$q_i \in D = \{d_1, d_2 \dots d_N\}$$

$$o_i \in R = \{r_1, r_2, r_3 \dots r_M\}$$

Figure 2 HMM Overview

A basic HMM consists of the following elements:

- 1)  $N$ : the set of state values is  $D = \{d_1, d_2, \dots d_N\}$ , where  $d_i$  stands for the  $i$ th state.  $N$  is the number of states.
- 2)  $M$ : the set of observed values is  $R = \{r_1, r_2, \dots r_M\}$ , where  $r_j$  stands for the  $j$ th observed RSSI.  $M$  is the number of observations.
- 3) The state transition matrix is

$$A = \begin{bmatrix} a_{d_1, d_1} & a_{d_1, d_2} \dots & a_{d_1, d_N} \\ a_{d_2, d_1} & a_{d_2, d_2} \dots & a_{d_2, d_N} \\ \vdots & \vdots & \ddots \\ a_{d_N, d_1} & a_{d_N, d_2} \dots & a_{d_N, d_N} \end{bmatrix}$$

where  $a_{d_i, d_j}$  stands for the probability from state  $d_i$  to state  $d_j$ .

- 4) The observation probability matrix is

$$B = \begin{bmatrix} b_{d_1, r_1} & b_{d_1, r_2} \dots & b_{d_1, r_M} \\ b_{d_2, r_1} & b_{d_2, r_2} \dots & b_{d_2, r_M} \\ \vdots & \vdots & \ddots \\ b_{d_N, r_1} & b_{d_N, r_2} \dots & b_{d_N, r_M} \end{bmatrix}$$

where  $b_{d_i, r_j}$  stands for the probability of getting the observed value  $r_j$  at the state  $d_i$ .

- 5) The initial state probability vector is  $\mathbf{p} = (p_1, p_2 \dots p_N)$ , where  $p_i$  stands for the probability of the initial state being  $d_i$ .
- 6) The end state probability vector is  $\mathbf{q} = (q_1, q_2 \dots q_N)$ , where  $q_i$  stands for the probability of the random process ending at the state  $d_i$ .

Once a system can be described as an HMM, three problems can be solve.

- a) *Evaluation*

Consider the problem where we have a number of HMMs describing different systems, and a sequence of observations. We want know which HMM most probably generated the given sequence. We use the forward algorithm to calculate the probability of an observation sequence given a particular HMM, and hence choose the most probable HMM.

- b) *Decoding*

Finding the most probable sequence of hidden states given some observations. We use the Viterbi algorithm to determine the most probable sequence of hidden states given a sequence of observations and an HMM.

c) *Learning*

The third, and much the hardest, problem associated with HMMs is to take a sequence of observations, known to represent a set of hidden states, and fit the most probably describes what is seen. The forward-backward algorithm is of use when the matrices A and B are not directly measurable.

### III ALGORITHM DESCRIPTION

In this paper, what the HMM does is actually to pretreat the received RSSI value sequence in order to reduce the environmental impact on the RSSI value. RSS-based ranging usually infers the distance value according to the attenuation of a group of received RSSI sequence. Due to the environmental influences on RSSI, even the RSSI value collected at the same position has some undulation over time, as shown in Figure 1. Within a relatively stable period, there are also certain probabilities that different RSSI values would appear at the same position. In contrast, the method based of *signal propagation mode* simply takes average of the collected RSSI sequence, and substitutes the mean value into the *signal propagation mode* to obtain the corresponding distance  $d$ , but ignores the relationship between RSSI and time or probability. Thus, it is now considered as a Hidden Markov process.

- Data collection and pretreatment:

Let  $= \{d_1, d_2, \dots, d_N\}$  be the set of *reference distances*. Let  $R_i$  be the RSSI sequence of  $d_i$ , and therefor  $R_i = \{r_i^1, r_i^2, \dots, r_i^n\}$ , where  $r_i^j$  stands for the  $j$ th value among the  $n$  RSSI values with distance of  $d_i$ .

Through the pretreatment of  $R_i$ , find out the suitable RSSI threshold value  $\gamma$ , pursuant to the selection criteria as: when  $RSSI > \gamma$ , the overlapping of RSSI should be as little as possible between different distances; when  $RSSI < \gamma$ , there will be partially overlapping RSSI values measured at different distances.

- Path loss exponent estimation

We use least square method over real detected data to estimate the path loss exponent. Then we can obtain the attenuation formula, and use it to calculate the corresponding distance of a detected RSS value.

- Training the HMM Model

Use the threshold value  $\gamma$  to determine the state vector and the observation vector, construct the HMM model, and train it with the collected data.

#### A *Path loss exponent estimation*

It is known that the transmission power is attenuated correspondingly the distance between the transmitter and receiver. The traditional *signal propagation mode* is

$$RSSI(d) = RSSI(d_0) - 10\lambda \lg\left(\frac{d}{d_0}\right) \quad (1)$$

where  $RSSI(d)$  is the  $RSSI$  value ( $dBm$ ) received at the distance  $d$  from the transmitting source,  $RSSI(d_0)$  denotes the received  $RSSI$  ( $dBm$ ) at a reference distance  $d_0$  (typically  $d_0$  is 1m) and  $\lambda$  is the path loss exponent related to the environment (typically between 2 and 4). In this paper, we according the real-world data to determine the path loss exponent  $\lambda$ .

Take average of the  $RSSI$  sequence at different reference distances, i.e.  $\{\bar{r}_1, \bar{r}_2, \dots, \bar{r}_n\}$ , where  $\bar{r}_i$  stands for the  $RSSI$  mean value when the reference distance is  $d_i$ . Use these values to do the least square fitting of the attenuation formula, and get an attenuation function conforming to the specific environment, i.e.  $\min \sum_{i=0}^n (RSSI(d_i) - r_{d_i})^2$ , and get the corresponding path loss exponent  $\lambda$ . According to the obtained data in this experiment, we have got the attenuation formula as:

$$RSSI(d) = 34 - 35 \lg(d) \quad (2)$$

In Figure 3, the red line is the  $RSSI$  values with different distances, and the blue line is the fitting curve from attenuation formula, taken as a prior knowledge.

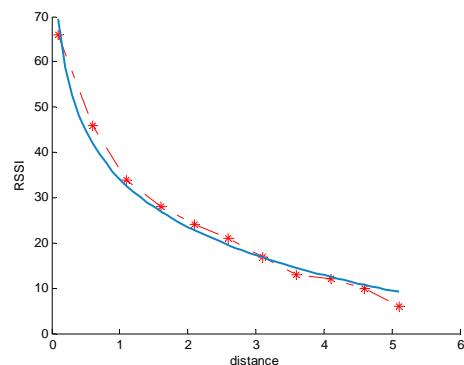


Figure 3  $RSSI$  Attenuates with Distance

#### B *Training the HMM*

- Model parameter estimation

Model parameter estimation is to, under certain known conditions, use different algorithms to select and optimize the parameters of the model  $A, B, p, q$ . To identify the optimal parameters of HMM model is an NP complete problem in mathematics. The usual way is to first give a set of inaccurate parameter, then adjust the model parameters through stepwise iteration refinement, and finally keep the parameters stable within an acceptable precision range.

In this paper, according to the  $RSSI$  values collected on-site at different distances, and based on the aforementioned  $RSSI$  threshold value setting standards, we set the threshold value as  $\gamma = 21$ , and thus the *reference distances* for  $RSSI < \gamma$  will be as following:

$$\{d_7 = 3m, d_8 = 3.5m, d_9 = 4m, d_{10} = 4.5m, d_{11} = 5m\}.$$

Since the distance values are independent of each other, in order to simplify the calculation, we can reference the commonly used method in speech recognition, and reduce a five-state HMM into five simple HMM models. Each simple HMM model is as:  $hmm_i = (D_i, R_i, A_i, B_i, p_i, q_i)$ ,  $i \in (1,5)$ , Where  $i$  stands for the  $i$ th state, the number of states being  $D_i = 1$ , the number of observed values  $R_i = n$  as the number of the RSSI values collected from the trained data, the state transition matrix as  $A_i = 1$ , the symbol transition matrix as  $B_i = (b_{r1}, b_{r2}, \dots, b_{rn})$ , which can be inferred from the appearance probability of the RSSI values in the trained data, and the initial state probability  $p_i$  can be set as the uniform value.

- Estimated state

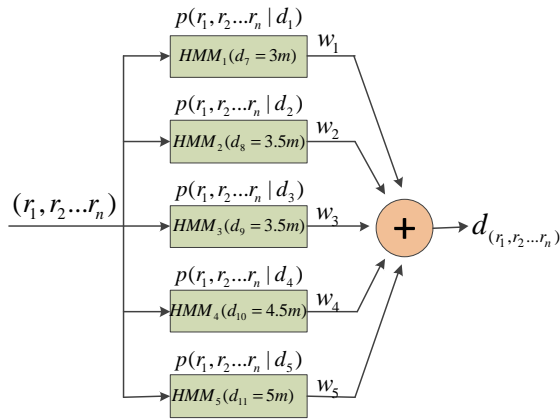


Figure 4 HMM Overview

As shown in Figure 3, we input an RSSI sequence in line with HMM processing, and use the forward algorithm in HMM to calculate the probability of the sequence appearing under each state. In order to obtain the intermediate values, we multiply the probability value of each state  $p(r_1, r_2, \dots, r_n | d_i)$  by the weight  $w_i$ . The estimated distance can be obtained from the weighted summation of the states.

$$d = \sum_{i=1}^5 w_i * p(r_1, r_2, \dots, r_n | d_i) * d_i \quad (3)$$

#### IV EXPERIMENTAL VERIFICATION

Our Hybrid HMM can be used for all RSS-based localization applications. In order to verify the strengths and weaknesses of the model, we have adopted the localization algorithm based on propagation model.

##### A Environment Setup

We have chosen the IRIS node of Crossbow for the hardware experiment, and set the experiment site at a 5m\*5m area in an open playground with no obstacle on site. As shown in Figure 4, the anchor nodes are placed in the red node positions in the area, and the RSSI values are collected from the four anchor nodes at the blue dot positions, with the coordinates of the anchor nodes already known.

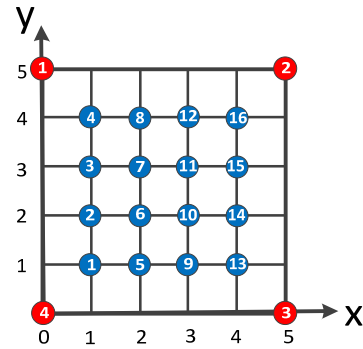


Figure 5 Experimental Nodes Deployment

##### B Node Localization

Here let us take position  $i$  as an example.

- Ranging stage:

The RSSI received from the four anchor nodes for position  $i$  ( $1 < i < 16$ ) is:  $R_i = \{R_i^1, R_i^2, R_i^3, R_i^4\}$ , where  $R_i^1$  is the RSSI value received from the first anchor node for position  $i$ . We input it into the Hybrid HMM ranging model so as to get the corresponding distance value  $d_i = \{d_i^1, d_i^2, d_i^3, d_i^4\}$ , where  $d_i^1$  is the value of the distance between position  $i$  and the anchor node inferred from the RSSI values, as shown in Figure 5.

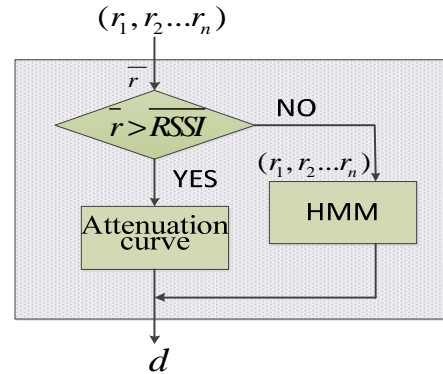


Figure 6 Hybrid HMM system

- Localization stage:

Use the distance values acquired from the ranging stage, and apply the method of maximum likelihood estimation to identify the location of position  $i$ . Based on the distance interconnection, we can get the equation:

$$AX_i = b \quad (4)$$

$$\text{Where, } A = \begin{bmatrix} 2(x_1 - x_4) & 2(y_1 - y_4) \\ 2(x_2 - x_4) & 2(y_2 - y_4) \\ 2(x_3 - x_4) & 2(y_3 - y_4) \end{bmatrix},$$

$$b = \begin{bmatrix} x_1^2 - x_4^2 + y_1^2 - y_4^2 + d_4^2 - d_1^2 \\ x_2^2 - x_4^2 + y_2^2 - y_4^2 + d_4^2 - d_2^2 \\ x_3^2 - x_4^2 + y_3^2 - y_4^2 + d_4^2 - d_3^2 \end{bmatrix}, \quad X_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}.$$

The estimated coordinates of position  $i$  obtained by utilizing the method of standard least mean square criterion is:

$$\hat{X}_i = (A^T A)^{-1} A^T b \quad (5)$$

### C Experimental Results

As the IRIS node has an effective outdoor ranging of 5m, we do not consider the positions beyond the range, i.e. the position of ID 1, 4, 13, 16; Moreover, taking symmetry into consideration, we only need to measure the RSSI sequences for the positions with an ID of 9, 10, 11, 12, 14, 15. We obtained 350 groups of RSSI data at each position, and put 5 groups into localization test at a time, repeating 70 times. We compared Hybrid HMM method with the method based of *signal propagation model*, and took the Euclidian distance between the estimated point and the actual position as the deviation.

As shown in Table 1, which exhibits the mean value of the localization deviation of the 70 sets of positions inferred from the measured data, where  $\bar{e}(L1)$  stands for the localization deviation of the *signal propagation mode* method in measurement, and  $\bar{e}(H1)$  stands for the deviation of the Hybrid HMM method. We can see that Hybrid HMM improves the localization accuracy in noise outdoor environments.

Table 1 localization accuracy

Position ID	9	10	11	12	14	15
$\bar{e}(L1)$	0.447	0.694	0.481	0.284	0.767	0.433
$\bar{e}(H1)$	0.379	0.630	0.417	0.347	0.777	0.335

As the outdoor environment is relatively stable, and the advantage of this method cannot be fully showcased, we simulated the NLOS scenario with the actual data, and randomly added 10% random noise, repeated the simulation 100 times, in order to get the average localization deviation values as shown in Table 2, where  $\bar{e}(L2)$  stands for the localization deviation of the *signal propagation mode* method after adding the random noise, and  $\bar{e}(H2)$  stands for the deviation of the Hybrid HMM method after adding the random noise. It is indicated that in the presence of a random noise, our method has a 14.3% higher localization accuracy than the *signal propagation mode* method on average.

Table 2 Localization accuracy

Position ID	9	10	11	12	14	15
$\bar{e}(L2)$	0.509	0.687	0.525	0.362	0.877	0.463
$\bar{e}(H2)$	0.404	0.630	0.415	0.349	0.777	0.341

As shown in Figure 7, when random noise is added, the localization deviation of a linear interpolation method increases, while the deviation of the Hybrid HMM method basically has no change, demonstrating a favorable

robustness. The deviation values of the 70 localization actions for each location is detailed in Appendix 1.

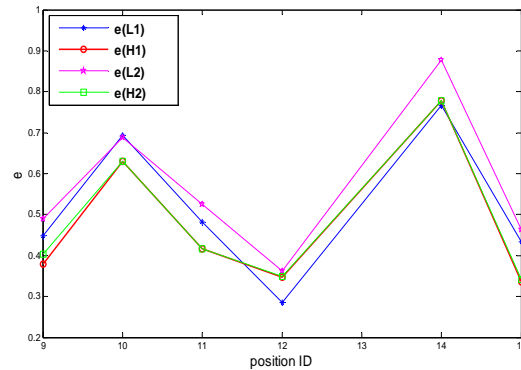


Figure 3 Robustness to random noise

### V CONCLUSION

As the RSSI value is not only influenced by the surrounding environment, but also dependent on the hardware conditions of the node itself, such as voltage and etc. In this paper we fully explored the relationship between RSSI and distance, and figured out the RSSI is not stable while the distance is relatively far. We overcome this problem by introducing an HMM preprocessing process on the ranging stage. Experimental results show that our algorithm is practical and has a relatively high fault tolerance. Furthermore, our model can also be easily combined with other RSS-based applications. In our future work, we will consider to extend this method to mobile wireless networks. In these applications, the mobile node localization and tracking can be modeled as an HMM process.

### REFERENCES

- [1] K.Stone, T.Camp, "A survey of distance-based wireless sensor network localization techniques", International Journal of Pervasive Computing and Communications, vol.8, pp.158-183, 2012
- [2] D.Macii, F.Trenti, P.Pivato, "A robust wireless proximity detection technique based on RSS and TOF measurements", Measurements and Networking Proceedings, 2011 IEEE International Workshop, pp.31-36, 2011.
- [3] M.Soleimanifar, M.Lu, I.Nikolaidis,S.Lee, "A robust positioning architecture for construction resources localization using wireless sensor networks", the 2011 Winter Simulation Conference, pp.3357-3567, 2011.
- [4] Y.Zhao, H.Zhou, M.L, "A novel overlap area matching algorithm based on location fingerprinting in wireless networks", Wireless communications, Networking and Mobile Computing, 2009, WiCOM'09.5<sup>th</sup> International Conference, pp.1-5, 2009.
- [5] A.Coluccia, F.Ricciato, "Maximum Likelihood trajectory estimation of a mobile node from RSS measurements", Wireless on-demand Network Systems and Services(WONS), 2012 9<sup>th</sup> Annual Conference, pp.151-158, 2012.
- [6] J.Krumm, E.Horvitz, "LOCADIO:Inferring Motion and Location from Wi-Fi Signal Strengths", First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA, USA, pp.4-13, 2004.
- [7] C.Morelli, M.Nicoli, "Hidden Markov Models for Radio Localization in Mixed LOS/NLOS Conditions", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 55, NO. 4, pp1525-1542, APRIL 2007

[8] V.Zheng, E.Xiang, Q.Yang, D.Shen, "Transferring Localization Models Over Time", Twenty-Third AAAI Conference on Artificial Intelligence, pp1421-1426, 2008.

[9] R. Arthi, K.Murugan, "Localization in Wireless Sensor Networks by Hidden Markov Model", Advanced Computing (ICoAC), 2010 second International Conference, pp14-18, 2010.

[10] [http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html\\_dev/main.html](http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/main.html).

### Appendix 1

