

Multi-Floor Semantically Meaningful Localization Using IEEE 802.11 Network Beacons

Uzair AHMAD^{†a)}, Brian J. D'AURIOL[†], Young-Koo LEE[†], *Nonmembers*, and Sungyoung LEE[†], *Member*

SUMMARY This paper presents a new methodology, *Beacognition*, for real-time discovery of the associations between a signal space and arbitrarily defined regions, termed as *Semantically Meaningful Areas (SMAs)*, in the corresponding physical space. It lets the end users develop semantically meaningful location systems using standard 802.11 network beacons as they roam through their environment. The key idea is to discover the unique associations using a beacon popularity model. The popularity measurements are then used to localize the mobile devices. The beacon popularity is computed using an 'election' algorithm and a new recognition model is presented to perform the localization task. We have implemented such a location system in a five story campus building. The comparative results show significant improvement in localization by achieving on average 83% SMA and 88% Floor recognition rate in less than one minute per SMA training time.

key words: *positioning and ranging, pattern discovery and recognition, ubiquitous and mobile computing*

1. Introduction

Location awareness is a key enabling technology for ubiquitous computing spaces. Although satellite based localization, e.g. GPS, is a defacto positioning method; many researches have pointed out its shortcomings such as low availability in the indoors and the requirement of special hardware [6], [8], [10], [11]. The potential ubiquitous computing applications can be realized by embedding the sense of location into commodity devices connected through the modern communication networks. There have been several branches of this research which utilize different radio communication networks for localization such as FM radio [10], GSM channels [5], and 802.11 access points [4], [6].

Most of these systems strive to estimate the actual location of the mobile device in terms of geometrical distance. However, as reported in [7], the geometrical definition of location is not suitable or required for many emerging indoor location enhanced applications. Instead such applications perceive a location as an 'area of interest' identified by everyday names e.g. Gents Garments Section, Food Court or Customer Service Center in a super market. The 'area of interest' which can have arbitrary boundaries serves richer semantic value and provides a defining block of a location system. We refer to such an area as *Semantically Meaningful Area (SMA)* in this paper. Two main categories of several applications of such localization scheme are: i)

multimedia content adaptation systems, e.g. tour guides, advertisement and collaborative games, and ii) high level activity recognition for context-aware ubiquitous computing spaces. Some studies have demonstrated such localization via beacon identification [3], [6], [7] or using probabilistic modeling [9], [15]. However, these methods require the prior information such as beacon positions, target locations or propagation models to train a localization algorithm. Such a localization scheme results in longer development time. Moreover, multi-floor environments pose a *resembling signal space* phenomena [14] which renders previous approaches ineffective to distinguish between adjacent floors.

Our main contribution is a new semantically meaningful localization methodology, referred to as radio **Beacon** signal measurement and **recognition** (*Beacognition*), which includes two components. First, an interactive, election algorithm which discovers the best representative beacons for an SMA. This discovery is performed online by presenting the beacon signatures to the algorithm. Second, a new location recognition method which computes the most probable location of the mobile device based upon an intuitive model. *Beacognition* is employed to build a semantically meaningful location system using IEEE 802.11 standard network beacons in a real multi-floor environment. In our view, this methodology is equally beneficial for localization in other wireless networks.

The rest of the paper is organized as follows. Related work is given in Sect. 2. A beacon localization primer is presented in Sect. 3. The election algorithm is discussed in Sect. 4. Section 5 explains the recognition model for location queries. Experimental results and comparative analysis are provided in Sect. 6. The conclusions are presented in Sect. 7.

2. Related Work

Recently, 802.11 wireless network based location systems have gained a significant attention from the research community [9], [11], [13], [15], [16], as well as industry [2], [6]. This is mainly due to the pervasive availability of 802.11 in indoor environments and proliferation of wireless network enabled commodity hand held devices. The WiFi location systems can be broadly categorized based upon: i) the resolution of location estimate and ii) location inference method.

Coarse resolution location systems, such as Intel's Place Lab [6], provide 20 to 30 meter accuracy in the outdoor scenarios, whereas fine resolution location systems

Manuscript received February 27, 2008.

Manuscript revised June 1, 2008.

[†]The authors are with the Department of Computer Engineering, Kyung Hee University, Republic of Korea.

a) E-mail: uzair@oslab.khu.ac.kr

DOI: 10.1093/ietcom/e91-b.11.3450

claim up to three meter accuracy in the indoor environments. The fine resolution systems model the physical space into geometrical grid [16] or topological cells [9]. Such systems require high density sampling, referred to as radio map, of target environment which provides the basis for developing a mapping function between physical space and signal space.

The division based on location inference methods is present in both coarse grained and fine grained systems. The low resolution systems often incorporate simple inference methods such as k -nearest neighbors [3]. However the fine resolution systems employ more complex classifiers such as statistical modeling [9], [11], [13], vector quantization [12] or neural networks [1].

The BeaconPrint [7] employs GSM and 802.11 beacon response-rate signatures to learn and recognize places in an outdoor environment. The Nibble [15] system builds a bayesian network to capture the dependencies between signal to noise ratio (SNR) and corresponding locations. It supports incremental development of a system, however prior knowledge of beacons and training data is required for off-line calculation of the conditional probabilities. The NearMe [8] system provides rapid development without radio map scheme. It is a proximity server which provides a list of neighboring devices based upon similarity of their beacon signatures. The SkyLoc [3] system specifically addresses floor recognition problem in multi story buildings. It improvises GSM (Global System for Mobile communications) beacons instead of 802.11 access points.

The Beacognition targets a coarse resolution but semantically meaningful location system in multi-floor buildings. The salient features of our methodology are as follows: i) it provides interactive development scheme to facilitate location system development, ii) no requirement of any prior knowledge lowers the entry barrier for location system developers, iii) the location to signal space mappings are discovered while a device is roaming through the environment.

The intuitive reason behind better localization performance of Beacognition has two aspects: i) unlike other approaches of using all detectable beacons as references, the Beacognition discovers only the best-representative beacons and their features via the Election algorithm. Apparently, using all detected beacons as references may be sufficient to distinguish different locations. However, due to the noise in indoor environments and resembling signal space problem, this approach faces difficulty to accurately recognize different locations, and ii) the Beacognition SMA recognition model measures the signal strength as well as ranking similarity between the detected beacon set and the reference beacons. Moreover, the ranking similarity measure gives weight not only to the rank of a detected beacon but also to the missing beacons. This approach better resolves the confusion between similar points in signal space representing different points in physical space.

3. Beacon Based Localization Primer

The core issue of inferring location information from radio beacons is to discover the association between a signal space and respective physical space. The area where the signal of an identifiable beacon can be detected is referred to as *Signal Coverage Area* (SCA). A device can infer its location whenever it receives signal from a particular beacon by searching in a list of SCAs. The boundary of an SCA depends on several factors, e.g. propagation environment, transmitter power and frequency band. The meaningfulness of an SCA boundary may often not coincide with the semantic requirements of an indoor location based service.

Multiple neighboring radio beacons, typical of urban and indoor environments, create overlapping signal space which can leverage important clues for localization. The overlapping signal space exhibits two important complementary facets for inferring location information.

1. The single beacon Distinguishable Signal Coverage Area (sDSCA) is the area where signal of multiple beacons can be detected but the signal intensity of a beacon remains stronger than others. A device can be straightforwardly localized with the location of a beacon with the strongest strength.

2. The multiple beacon Distinguishable Signal Coverage Area (mDSCA) is created by overlapping signal space of more than two beacons. It represents more complex boundaries in the physical space given that the appropriate associations between signal and physical space are discovered. The localization is performed by comparing the membership similarity between set of detected beacons v with the all mDSCAs in ($mList$). Then the maximally similar mDSCA to v represents the location of the device as following.

$$l = \underset{i \in N}{\operatorname{argmax}}(\|v \cap mList_i\|) \quad (1)$$

It is understandable that all locations identifiable by beacons may not be the SMA for a location enhanced system. Ideally, the semantic needs of a location based application system should define the boundaries of an SMA. The formation of best representative beacons and their related properties is referred to as Semantically Meaningful Area Recognition Template (*smart*) in the rest of the paper.

In multi-floor environments, beacon based localization face a resembling signal space at vertically similar areas on adjacent floors. This phenomenon is also observed in a cross floor signal propagation study for three radio frequencies in a multi-floor indoor environment [14]. The authors identify that, in certain situations, signal levels tend to remain quasi-constant in adjacent floors. Furthermore, the signal attenuation for a single floor separation is often lower than same-floor signal attenuation. These cross floor propagation characteristics implicate the localization task and require special care for discovering *unique* associations as well as adequate similarity measures for accurate recognition.

In [4], we presented a preliminary version of our bea-

con based localization scheme, referred to as 2-D *smart* in this paper. Here we present an enhanced version of the election algorithm which: i) allows developers to explicitly create an SMA, ii) computes the statistical properties of each access point, iii) detects and prohibits redundant *smart* formations. Moreover, a new SMA recognition model is presented which significantly improves the performance in multi-floor scenarios.

4. The Election Algorithm

The algorithm views all detectable beacons in an SMA as contestants of an election, hence termed the election algorithm. It ranks all contestants based upon the votes they ‘win’ from different points in an SMA. Once a valid *smart* is discovered, the less representative beacons are pruned out so that only the winner survives. Two important operating conditions of this algorithm distinguish it from the conventional localization algorithms: i) the signal to location mapping is created in real-time while the device is roaming in the target area and ii) no prior information about beacons, e.g. beacon ID, position or radio map, is available.

4.1 A Motivating Analogy

The intuition behind this algorithm derives from an analogy with the political election process in the real life where multiple candidates contest for winning the representative offices (*smart*) of a constituency (SMA). The beacons are analogous to the candidates and the individual points in physical space are the voters. The polling results produce a ranking which reflects the degree of representiveness, or popularity, of each candidate such that highest office is awarded to the candidate who wins maximum votes from a sample population in that constituency, then the next office is given to the next highest winner.

The polling process of Election algorithm differs from conventional ones in the time of voting and results computation and duplication of votes. In ordinary elections, polling takes place simultaneously at different locations of a constituency at a given time and the results are compiled afterwards. This polling method is applicable because all the contestants know their constituencies. However, consider a situation in which candidates do not know their constituencies but an election must be held. In the same sense, our aim is to ‘discover’ appropriate beacon representatives for SMAs while ID or position of any beacon is not known. Owing to the special nature of the task, the algorithm conducts polling sequentially and results are compiled on the fly as the voting continues. Allowing duplicate votes is another deviation from the normal elections. However, since the main objective is to facilitate end users to develop location systems, it is not possible for a developer to visit certain pathway only once. Therefore the algorithm detects the duplicate voting internally and only unique *smart* formations are created.

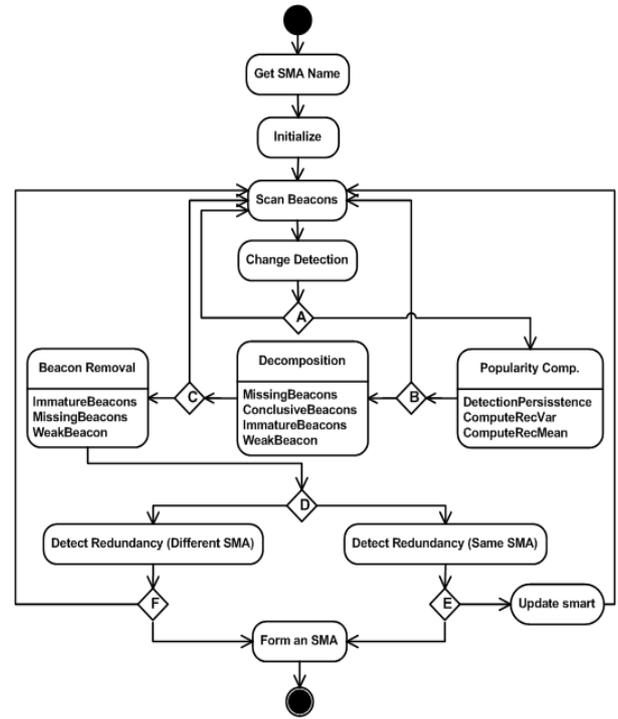


Fig. 1 SMA creation dynamics of the Election Algorithm. Description of Condition Labels: **A.** Change detected ? **B.** Missing beacons detected ? **C.** Smart size constraint is met OR Conclusive beacon missing OR Intrusion ? **D.** Same SMA or different SMA ? **E,F.** Enough different ?

4.2 The SMA Creation Dynamics

The state chart presented in Fig. 1 shows the dynamics of the Election algorithm, and the detail of each state is given in the following subsections.

The SMA creation process starts by asking the location system developer to name the SMA to be created, referred to as ‘*this* SMA’ in the following discussion. The initialization state puts the device into the polling mode by activating the scanning/voting procedure of the device. Incidentally, each scanning operation detects n candidate beacons who want to represent this SMA. As long as a device is roaming through an SMA, a set $T_b = \{b_1, b_2, b_3 \dots b_n\}$ is maintained whose elements are all the beacons which get detected during this time. The structure of the i th individual beacon b_i is composed of three properties: i) MAC address, ii) detection persistence and iii) mean and variance of its signal strengths. The time and method of computing the second and third property is explained in Sect. 4.2.2.

As the device keeps scanning, two primary beacon sets of detected beacons are observed: i) the N_b contains all beacons that get detected at time t , and ii) the L_b contains all beacons that get detected at time $t - 1$. Upon each scanning operation, an important condition is checked for further proceeding based on these two primary beacon sets as follows.

4.2.1 Change Detection

Detecting a change in the location or signal space is important from system development point of view. Considering the election analogy, for the sake of fairness, it is necessary that all locations cast equal number of votes which requires that equal number of scanning operations should be performed at all locations. However, it is very unlikely that the human carrier of the device will keep a consistent speed and visit each location an equal number of times. Therefore, the consequent popularity computations can get biased towards the detected beacons at a location where device stays longer. However, if scanning/voting detects the same set of beacons ($N_b = L_b$) then it would not affect their popularity. Hence, we repose the issue of fairness by changing the equal vote counting condition to that of dissimilar vote counting. This means that valid polling occurs only when there is a change detected in signal space. This change is important even if the device is either stationary or moved at scanning time. It eliminates the consistent speed constraint placed upon the system developer as well as subsequent redundant operations. In case the location is changed but there is no change in the signal space, a new location is covered by the same signal space. Therefore, no computation is required. The change detection state observes the change in signal space by checking the memberships of N_b and L_b . The device perceives a changed signal space in either or both of two cases: i) a new beacon appears in N_b , and ii) a beacon is missing in N_b which was detected in L_b .

4.2.2 Popularity Computations

In case of change detected in the signal space, the process continues to the next state ‘PopularityComp.’ In this state the election algorithm computes the popularity of individual beacons in T_b . The popularity of a beacon, referred to as *detection persistence* (dP), is computed as how persistently it gets detected in an SMA. The dP is measured as the ratio of detection count and total number of scans (or votes) for each beacon as follows:

$$b_i^{dP} = \frac{b_i^{D_c}}{S_c} \quad \forall b_i \in t_B \quad (2)$$

where D_c denotes the detection count for b_i and S_c is the total scan count. After computing dP for each beacon in T_b , the D_c for all N_b beacons is incremented and the scan count S_c is incremented for next round of voting.

In the same state, the mean and variance of signal strength values for each beacon in N_b is computed. These computations have an important role in resolving the overlapping signal space problem that was discussed in the previous section. This information contributes in both discovering the SMAs, at the detection time of redundant *smarts*, and in recognition to achieve inter-floor localization. Since our aim is to avoid the requirement of any prior information about signal space, we choose to compute mean and

variance recursively. The recursive mean exhibits attractive properties of reaching directly to the core of the data and being ‘lazy’ to move away from there due to long term memory. The algorithm computes μ^{ss} , where ss is the signal strength, for each beacon in N_b as $\mu^{ss} = \frac{D_c-1}{D_c} \mu_{D_c-1}^{ss} + ss_{D_c}$. At the same time, the signal strength variance of same beacons is updated using standard recursive variance formula as $\sigma^2 = \frac{D_c-1}{D_c} \sigma_{D_c-1}^2 + \frac{1}{D_c-1} (ss_{D_c} - \mu^{ss})^2$.

4.2.3 Decomposition of Beacon Trace

After computing reputations, if any beacon is found missing in N_b , the control reaches this state. One or more missing beacons indicate that the device might have moved out of the SMA under creation and it ensues a possible *smart* formation. At this point the system segregates legitimately representative beacons from non-representative ones based upon their popularity. However, before determining a beacon as missing, the election algorithm observes a grace period strategy because a beacon may be temporarily missing due to noise. The temporal absence of beacons is a commonplace phenomenon especially in indoor environments. This occurrence can cause abrupt removal of a beacon from candidates set and if it reappears after short absence all its previous popularity is lost. A cushion is provided to overcome this potentially perturbing situation. The $m\tau$ is an externally specifiable parameter which allows the system to tolerate temporarily missing beacons. Due to this mechanism an absent beacon is not considered to be missing until it is consistently not detected more than $m\tau$ times. When a beacon is not detected even after the grace period elapsed, it is marked as missing from the beacon trace.

After ensuring that one or more beacons are missing, the election algorithm divides the beacon trace into subsets, referred to as beacon containers, with respect to their detection persistence. The missing beacons are put into m_B ‘container’ which is a set of beacons which had appeared in T_b but are missing at this moment. The beacons trace is decomposed into three further subsets: i) the ‘winners’ are the possible *smart* beacons which got votes more than certain threshold, e.g. ($dP \geq .70$), ii) the ‘weak’ beacons w_B are the ones which won lesser votes, ($.70 > dP \geq .30$), and iii) the ‘immature’ beacons i_B are the beacons who got even lesser votes, ($dP < .30$). The choice of dP threshold values can be externally determined by developer.

The notion of immature beacons capture the possibility of those beacons who just started to appear in T_b . This situation can occur when a device is moving in the boundary of the currently formatting SMA. Such beacons present might have better contribution to represent some other SMA but for current SMA they are considered as less-representative beacons. However all or some of the missing beacons can be the ones who are yet missing but they won significant votes until now. These beacons are potential boundary markers of the SMA and should not be ignored. Therefore system divides missing beacons into two further subset: i) the ‘to be removed’ r_B , and ii) the ‘conclusive’ con_B . Each of these

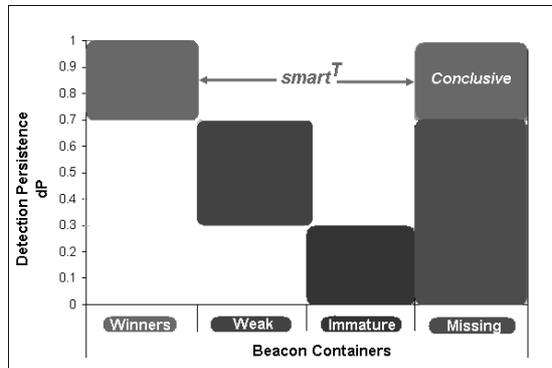


Fig. 2 Beacon containers: decomposed beacon trace.

have an opposite role as explained next. The distribution of decomposed T_b with respect to dP is shown in Fig. 2.

4.2.4 Beacon Removal and Transitive *smart*

The decomposition of the T_b into beacon containers ensures the removal of missing and less-representative beacons from the system. At this point, the system observes another externally specified constraint of maximum smart size denoted as τ :

$$\|T_b\| - (\|m_B\| + \|w_B\| + \|i_B\|) \leq \tau \quad (3)$$

In [4] we detailed the impact of different values of τ on customizing the boundaries of an SMA. It is observed that, in many cases, fewer beacons represent large gDSCAs and more beacons represent small gDSCAs. This constraint further delays the final creation of a *smart* giving another chance to the temporarily missing beacons to make up. There are two other situations when the system creates a transitive smart: i) when a conclusive beacon is detected to be missing, and ii) the developer intrudes the polling process to explicitly create an SMA. Once the second case happens, system tries to find the best representative beacons and create a corresponding *smart* immediately. However, this intrusion is treated as just a suggestion. If there is no corresponding distinguishable signal space, then, the system ignores this suggestion. Ultimately, the remains of t_B are the ones who have won the election in *this* SMA. At this point, these beacons are referred to as transitive *smart*^T.

4.2.5 Detection of a Redundant *smart*

In order to avoid the redundancy, before finalizing the *smart*^T to be a *smart*, the system ensures that a similar set is not already existing. A similar *smart* could have been formed previously in two cases.

Case 1. The developer is roaming in the same SMA even after detection of a valid *smart*. This case can be detected by measuring the similarity between *smart* S_a and *smart*^T S_b using Dice's coefficient [17] denoted as d^C . It measures asymmetric information in two sets which may contain dissimilar elements. The d^C reflects the weight of

Table 1 Structure of a *smart*.

smart			
SMA	Recognition Template		
SMA name	MAC	SS Mean	SS Variance
<i>this</i> SMA	4259	-47	1.3
	5659	-53	2.4
	9235	-69	1.1

common elements in two *smarts*, S_a and S_b , as follows:

$$d^C = \frac{2 \|S_a \cap S_b\|}{\|S_a\| + \|S_b\|} \quad (4)$$

Case 2. The developer has reentered in an SMA which is represented by an already valid *smart*. This case requires more careful treatment for the redundancy detection because only d^C can cause *smart*^T to be regarded as redundant due to the similar signal space on adjacent floor. The election algorithm measures this similarity as an average of signal strength confidence and d^C . Measuring the signal strength confidence is explained in Sect. 5.1.

In either case, if the similarity between S_a and S_b is more than certain threshold, 0.70 in our experiments, then *smart*^T is considered to be redundant thus ignored. However, since the device is roaming in the same SMA, the statistical information of common beacons is updated.

4.2.6 Creation of a *smart*

After passing the redundancy check, the system creates a *smart* by associating the final set of beacons and their statistical information with *this* SMA. The control shifts to scanning state if developer wants to continue otherwise it stops at this point. The structure and example members of a *smart* are shown in Table 1.

5. The Location Recognition Model

The second component of the Beacognition methodology is to recognize the location of a query beacon signature. The influence of the election analogy does not end at the discovery of *smarts* and formation of SMAs. The final localization decision is again made based upon a voting strategy. At that time, a device puts forward the detected beacons as location query v to all representatives, *smarts*, for finding its appropriate location. The member beacons of respective *smarts* vote their confidence to represent the query beacon set v . The location recognition model evaluates the vote of confidence of each s in all *smarts* S . Consequently, the location of querying device is inferred as the SMA of the winner *smart* w who is most confident to represent v as follows:

$$w = \operatorname{argmax}_{s \in S} f(s, v) \quad (5)$$

where $f(s, v)$ is computed as an average of two components of the recognition model as discussed next.

5.1 Signal Strength Confidence

Signal strength of a beacon carries important clue for resolving confusion about the correct location among neighboring SMAs. A simple signal strength similarity measurement formula is devised to compute the combined impact of all beacons in a *smart* on the localization within [0, 1] range.

$$sv^S = \sum_{i=1}^n \frac{1}{n} \text{ where } n = |s|$$

$$\Rightarrow \left[i|\forall (s_i = v_i) \in (s \cap v) \text{ and } \left(\frac{|v_i^{ss} - s_i^{\mu}|}{s_i^{\sigma}} \right) \leq 2 \right] \quad (6)$$

It assumes that within an SMA, the spread of signal strengths follows a normal distribution. By definition, a standard normal distribution can explain 95.4% of the variation within 2 standard deviations from the mean. Therefore, if the normalized signal strength falls in the range [-2, 2] then it is considered to be in favor of the *smart* *s*.

5.2 Ranking Confidence

The ranking confidence sv^R is measured on signal strength based ranking of *smart* s' and query beacon signature v' . The sv^R measures it in the range of 0, for no confidence, to 1, for high confidence. The ranking similarity measurement needs special treatment because of two common situations which are ignored in the previous works: i) the s' and v' may contain different beacons, and ii) the number of beacons in both signatures are different. First the difference between two sets is measured as a fraction of ‘excitation’ and (Ex) and ‘inhibition’ (In) of the differences in ranking. Subtracting the total difference from 1 gives the ranking confidence as given in Eq. (7).

$$sv^R = \left(1 - \frac{Ex}{In} \right) \quad (7)$$

Let s_i and v_j denote the individual beacons in s and v where i and j are their respective ranks. The Ex is calculated as:

$$Ex = \begin{cases} Ex + |i - j| & \text{if } s_i = v_j \in c \\ Ex + 1 & \text{if } v_j \in d \end{cases}$$

where $c = s \cap v$ and $d = s - v$. The In is calculated as,

$$In = \begin{cases} In + (i + j + (k - i)) & \text{if } s_i = v_j \in c \\ & \text{and } i = j \\ In + (i + j) & \text{if } s_i = v_j \in c \\ In + 1 & \text{if } v_j \in d \end{cases}$$

where $k = ||s||$. The Ex is the summation of all differences in the rank of common beacons in s and v as well as uncommon beacons. The In sums up the similarities for common beacons to inhibit the excitement in differences. The occurrence of a beacon at the same rank in both sets points to high similarity between them for that beacon. This event inhibits the difference with more weight such as

$In = In + (i + j + (k - i))$. However, the weight of a missing element adds a constant 1 to both Ex and In , in order to avoid detection of ‘0’ difference for completely different sets.

6. Experiments

In order to ensure that the experimental environment reflects the settings available in common multi-floor buildings, we have taken three important steps: i) we used WiFi access points of a public network ‘Nespot’ deployed by Korea Telecom and the deployment and position of the beacons is not manipulated to better suit required SMA boundaries, and ii) the boundaries of SMA were arbitrarily defined independent of the access point deployments, and iii) we collected multiple data sets on different day in order to evaluate the affect of the temporal variability in signal space which can cause biased location estimation performance. We have conducted extensive experiments in a five floor campus building which contains multiple departments as well as associated labs and class rooms. Table 2 lists the SMAs defined in five floors of engineering building and 802.11 access points deployed in respective floors.

We used hp iPAQ PDA model h4150 device running Pocket PC Embedded 2003 version of Microsoft Windows. The PDA has in built 802.11 network interface card. The WiFi signal collection system was implemented using C# in. Net to acquire data for training and test evaluations of our system. Since signal space can exhibit different properties at different times due to the indoor environmental factors. In order to evaluate the effect of time on localization capability, our training and test data collection spanned over one month at different days and times of the day. Besides time difference, the effect of the size of the training data was also evaluated by employing different sizes of the training set. Table 3 shows a listing of training data sets which were

Table 2 Experimental physical space and 802.11 beacons.

Floor	SMAs	Beacon MAC (Last four Digits)
1	Comp Physics Labs	8135
1	Photo-Electronics Labs	5139
1	Institute of Natural Sci	5035
1	Natural Sci Lec Rooms	5883
2	Robotics Labs	9235
2	Biomed Lecture Rooms	7199, 9207
2	App Biomedical Engg	
2	Admin Offices	8203
3	Comp Engg Labs	7195
3	Radio Engg Labs	9239
3	Imp/Img Res Labs	2243
3	Faculty Offices	5823
4	Std Unions Offices	
4	Radio Engg Rooms	5551
4	Bio-Medical Labs	
4	Lecture Rooms	5535, 5543
5	Laser Engineering Labs	5659
5	Communication Labs	6079
5	Astrophysics Labs	
5	Micro/Ultrasonic Labs	5559

Table 3 Training beacon data specification.

Data Set	Scans/sec per SMA	File Size(kB)
Trg ₁	40	51
Trg ₂	50	62
Trg ₃	50	67
Trg ₄	100	201

used for training our system as well as other systems given in [3], [8] and [9]. For the sake of demonstrating short development time, the training beacons were collected while a volunteer was walking on a normal speed in the SMAs. For the largest training set, Trg₄, the average time spent in each SMA was less than two minutes while the device was scanning the beacons every second. For the test data collection, we spent 15 hours in all SMAs (on average 45 minutes per SMA) and collected 20,000 beacon signatures. All methods were first trained on each training set and tested on the extended test data set.

The localization performance is evaluated in two respects. Firstly, the SMA recognition is measured as: i) ‘Correct SMA’ is the ratio of correct estimates in total test signatures. ii) ‘1 SMA off’ is the ratio of estimates which deviate from actual SMA not more than 1 neighboring SMA in total test signatures. Secondly, floor recognition-ability of each method is computed as ratio of ‘Correct Floor,’ ‘1 Floor Off,’ ‘2 Floors Off’ deviations of estimates from the actual floor.

6.1 NearMe

The NearMe system provides a list of neighboring devices by comparing the beacon signature of querying device with the signatures received from all other devices connected to the system. It proposes a general heuristic for computing the physical distance of two devices from their signatures. An elaborate description of this method can be found in [8].

In order to evaluate the SMA and Floor recognition ability of NearMe system following test setting was created. Suppose that *j*th test signature $T_j^{S_i}$, collected in *i*th SMA S_i , represents the device which wants to enquire for neighboring devices. Similarly, the signatures in the *k*th training beacon set Trg_{*k*} are collected from devices which are located all over the building including the S_i nearby $T_j^{S_i}$. Let *n* denote the number of training signatures which were collected from S_i . For finding the nearby devices, guided by the NearMe heuristic, the distance of each test signature $T_j^{S_i}$ with all the beacon signatures in Trg_{*k*} is computed. Since we know the exact number of neighboring devices is *n*. Therefore nearest *n* distances should be the devices which are within the proximity of query device. If the NearMe system qualifies *m* remote devices to the neighborhood of $T_j^{S_i}$ then the localization error e^{S_i} is computed as $e^{S_i} = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{n}$ where *N* is the total number of test signatures. The SMA recognition and Floor recognition results of NearMe system are shown in Figs. A·1 and A·2 respectively. The SMA recognition-ability is similar across different experiments. The Floor recognition results show that it can successfully localize

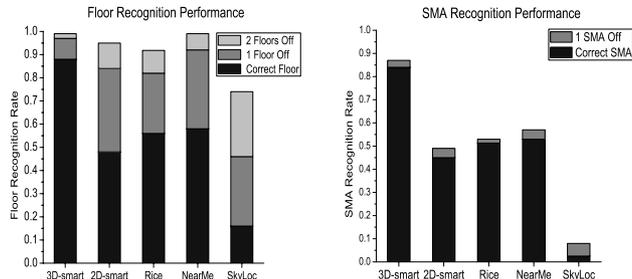


Fig. 3 Overall recognition results averaged over all training data sets.

within 2 floors. However the exact floor recognition rate remains lower than 60% on average.

6.2 SkyLoc

SkyLoc system employs GSM beacons to identify the floor of a mobile device in a multi-floor building [3]. It employs the *k*-nearest neighbors algorithm for floor identification. We can not obtain the GSM signatures due to the unavailability of specific hardware and GSM infrastructure. Therefore, all experiments were conducted using only the 802.11 network beacons and we implemented the SkyLoc system to evaluate its performance in such environments. Detailed results for each floor are given in Fig. A·3 and overall localization performance is shown in Fig. 3. The main reason behind performance degradation of SkyLoc is an essential incapability of Euclidean distance in signatures to represent the physical distance. Incidentally, if majority of the beacons in *t* and *v* are different but a minority, or just one beacon, is similar then the distance can be minimal. The possibility that a non-representative beacons may exhibit similar signal strength at different floors makes this distance almost zero. Thus majority of physically distant locations fall in *k* nearest neighbors of signal space. We refer to this effect as overshadowing of non-representative beacon over representative beacons. Due to the same reasons the SMA recognition performance is degraded. Individual SMA recognition results are not presented here due to space constraints.

6.3 Rice

A topological location model based localization scheme is presented in [9], here referred to as ‘Rice.’ It divides the target environment into cells akin to the SMAs. Like typical radio map based location systems, it requires a priori information about the signal space to capture the signal strength variations at different cells of the target environment. A succinct description of their method can be found in [9]. We implemented this system for our target environment with same data preparation and modeling specifications. The only difference is the number of beacons we use for training the system whereas ‘Rice’ requires higher density network for claimed localization. However, they have shown that lowering the density of beacons shall reduce the localization accuracy. Our implementation results confirm this as shown

in Figs. A·4 and A·5 respectively.

6.4 Beacognition

We evaluated the performance of Beacognition using the same data sets as used for the other systems. Even though it is a real-time discovery method, that is, it does not require a priori availability of training signatures, for the sake of comparisons we feed the target device same training and test data sets. In [4] we presented the first version of election algorithm, here referred to as 2-D *smart*, for single floor scenarios. It was demonstrated that the 2-D *smart* could achieve 87% accuracy in a two dimensional physical space. However in a multi-floor scenario its performance suffers from the resembling signal space problem as with the other systems. The enhanced version of Beacognition for multi-floor localization is referred to as 3-D *smart*. We conducted two types of experiments to evaluate the performance of the 3-D *smart* with the 2-D *smart* as well. For both types we chose the values of τ and $m\tau$ parameters as 5 and 3 respectively. Figures A·6 and A·7 show performance of the 3-D *smart*.

Due to the enhanced election algorithm and similarity measurement model, the 3-D *smart* significantly outperforms other methods. The average correct floor recognition performance over all experiments is 89%. Whereas the SMA recognition is 83% for correct SMA and 86% for 1 SMA Off. Overall performance of all methods is summarized in Fig. 3. The requirement of minimum SMA and Floor recognition rates largely depends on target applications. However, for coarse resolution based applications the expected base line 83% SMA recognition rate is very practical while the location information also contains the semantic meanings. In order to sustain sporadic inaccurate estimates, a general heuristic can be applied which utilizes: i) averaging the response of multiple location queries, and ii) the last known location and build a graph model to estimate next most probable location given the location estimate.

In order to evaluate the effect of training data size of localization performance we conducted a series of experiments with varying data amounts. For each experiment we used a portion of our training data sets Trg_i one by one. Figure 4 shows the SMA recognition performance of 3-D *smart* with respect to different training sets and samples sizes. The ‘Correct SMA Training’ is the ratio of correct estimates against total training signatures. It reflects that how successfully Election algorithm can create SMAs from training data. The Floor recognition results are shown in Fig. 5. It is noted that both SMA and Floor recognition performance improves upon increasing the training samples. However, as can be observed from Trg_4 experiments, increasing the training signatures beyond 50 does not have significant effect on the performance. It suggests that a device needs to roam in an SMA for less than one minute while scanning beacons every second.

Another straightforward merit of our methodology is that it requires minimal storage for storing its knowledge about location to signal space mapping. Once the winner

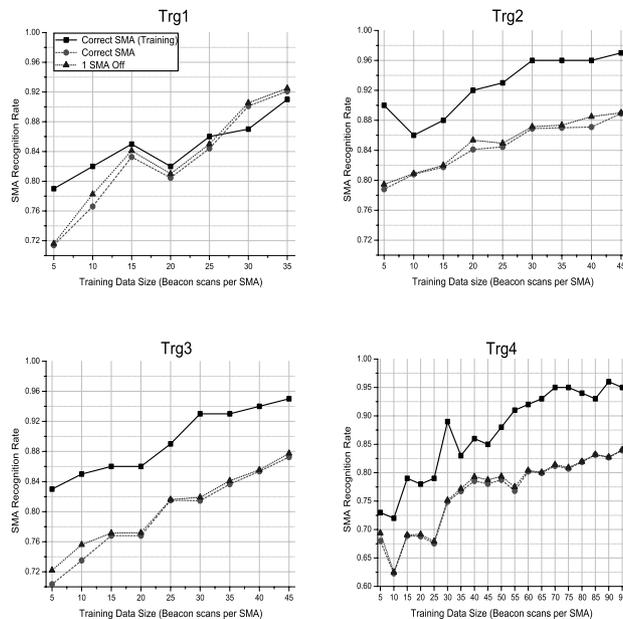


Fig. 4 3-D smart SMA recognition results for different training data size. The legend of Trg1 applies to all graphs.

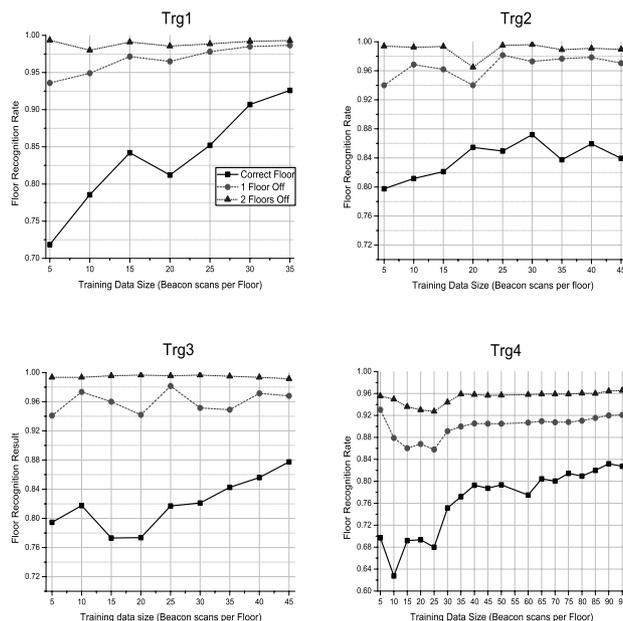


Fig. 5 3-D smart Floor recognition results for different training data size. The legend of Trg1 applies to all graphs.

beacons are discovered, all it needs to store for an SMA S is those best representative beacons along with their popularity information. In our experiments the size of the file containing all *smarts* remains less than 9 kB. Contrarily, the other systems require all training signatures collected from S to be stored. Later at the recognition time the computational requirements of our method are also significantly lower than other systems. Each test signature, location query, is compared only with the *smarts* whereas other methods compare it with all training signatures.

Even though our experiments were conducted using only IEEE 802.11 WLAN access points, the underlying radio infrastructure does not matter as long as the joint coverage area of the beacons is sufficient enough for the target applications. Beaconing can be applied to other commonly available radio signals (e.g RFID, Bluetooth, GSM, CDMA) for indoor localization very easily. However, the short range beacons such as RFID shall require denser deployments on large scale.

7. Conclusions

This paper offers a new methodology for indoor semantically meaningful localization using IEEE 802.11 network beacons. It aims at two objectives: i) short development time by providing a real-time discovery algorithm which requires no prior knowledge about signal and physical space, and ii) it involves end users in the system development by providing an interactive development scheme. Both objectives serve the purpose of rapid development of indoor location systems in a multi-floor environment. The comparative results show that Beaconing achieves significantly better performance, both in terms of SMA recognition as well as Floor recognition. The *smarts* require minimal storage which makes it suitable for stand alone, privacy observant location systems. In this case a device can localize itself by passively scanning the environment without even being connected to the network. However, a centralized location system is feasible as well by placing all *smarts* on a location server while mobile devices place their location queries in terms of detected beacon signatures. Despite significantly better performance of Beaconing, it inherits a limitation of the beacon-based localization systems which is the requirement of fairly dense deployment of radio beacons. In case of a sparse deployment the boundaries of the target SMAs may not exactly correspond to the needs of target applications. As a future work, we will extend our work to incorporate CDMA signals as well and implement specific applications on top of the Beaconing.

Acknowledgments

This research was supported by the MKE (Ministry of Knowledge Economy), under the ITRC (Information Technology Research Center) support program 'IITA-2008-C1090-0801-0002' and by the MIC (Ministry of Information and Communication), Korea, Under the ITFSIP (IT Foreign Specialist Inviting Program) 'C1012-0801-0003,' both supervised by the IITA (Institute of Information Technology Advancement), Korea. This work is also supported by the Ministry of Education(MOE), the Ministry of Commerce, Industry and Energy (MOCIE) and the Ministry of Labor(MOLAB) through the fostering project of the Lab of Excellency.

References

- [1] U. Ahmad, A. Gavrilov, Y.-K. Lee, and S. Lee, "Context-aware, self-scaling Fuzzy ArtMap for received signal strength based location systems," *J. Soft Computing—A Fusion of Foundations, Methodologies and Applications* (Springer), vol.12, no.7, pp.699–713, 2008.
- [2] Ekahau Positioning Engine, <http://www.ekahau.com>, Ekahau Inc., 2008.
- [3] A. Varshavskya, E. de Lara, J. Hightower, A. LaMarca, and V. Otsason, "GSM indoor localization," *Proc. Pervasive and Mobile Computing*, pp.698–720, 2007.
- [4] U. Ahmad, A. Gavrilov, B.J. d'Auriol, S. Lee, and Y.K. Lee, "The election algorithm for semantically meaningful location-awareness," *Proc. ACM/IEEE Mobile and Ubiquitous Multimedia*, pp.55–63, Oulu, Finland. Dec. 2007.
- [5] M. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca, F. Potter, I. Smith, and A. Varshavsky, "Practical metropolitan-scale positioning for GSM phones," *Proc. Ubicomp*, Orange County, California, USA, Sept. 2006.
- [6] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powlledge, P. Borriello, and B. Schilit, "Place lab: Device positioning using radio beacons in the wild," *Proc. Pervasive*, Munich, Germany, 2005.
- [7] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes, "Learning and recognizing the places we go," *Proc. UbiComp*, pp.159–176, 2005.
- [8] J. Krumm and K. Hinckley, "The NearMe wireless proximity server," *Proc. Sixth International Conference on Ubiquitous Computing*, pp.283–300, Sept. 2004.
- [9] A. Haeblerl, E. Flannery, A.M. Ladd, A. Rudys, D.S. Wallach, and L.E. Kavvaki, "Practical robust localization over large-scale 802.11 wireless networks," *Proc. 10th Annual International Conference on Mobile Computing and Networking*, pp.70–84, 2004.
- [10] J. Krumm, G. Cermak, and E. Horvitz, "RightSPOT: A novel sense of location for a smart personal object," *Proc. Ubicomp*, pp.36–43, 2003.
- [11] M.A. Youssef, A. Agrawala, and A.U. Shankar, "WLAN location determination via clustering and probability distributions," *Proc. IEEE PerCom*, March 2003.
- [12] T. Ogawa, S. Yoshino, and M. Shimizu, "Location determination method for wireless systems based on learning vector quantization," *NTT Network Innovation Laboratories*, Japan, vol.1, no.9, pp.27–36, 2003.
- [13] T. Roos, P. Myllymaki, and H. Tirri, "A statistical modeling approach to location estimation," *IEEE Trans. Mobile Computing*, vol.1, no.1, pp.59–70, Jan.–March 2002.
- [14] P. Vlach, B. Segal, J. LeBel, and T. Pavlasek, "Cross-floor signal propagation inside a contemporary ferro-concrete building at 434, 862, and 1705 MHz," *IEEE Trans. Antennas Propag.*, vol.47, no.7, pp.1230–1232, July 1999.
- [15] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A probabilistic room location service for wireless networked environments," *Proc. Ubicomp 2001*, pp.18–34, 2001.
- [16] P. Bahl and V.N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," *Proc. IEEE Infocom 2000*, pp.775–784, March 2000.
- [17] L.R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol.26, pp.297–302, 1945.

Appendix: Detailed Experimental Results

[1] U. Ahmad, A. Gavrilov, Y.-K. Lee, and S. Lee, "Context-aware,

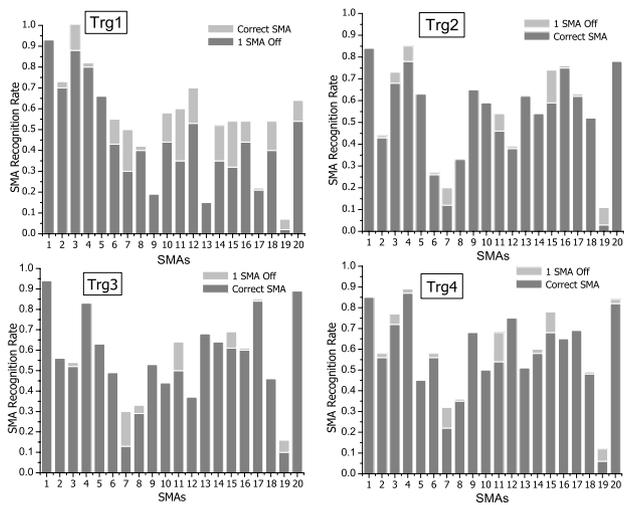


Fig. A-1 NearMe SMA recognition results.

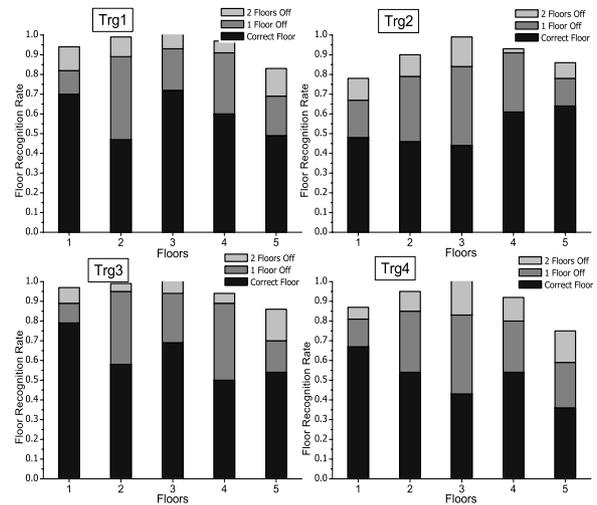


Fig. A-4 Floor recognition results of rice.

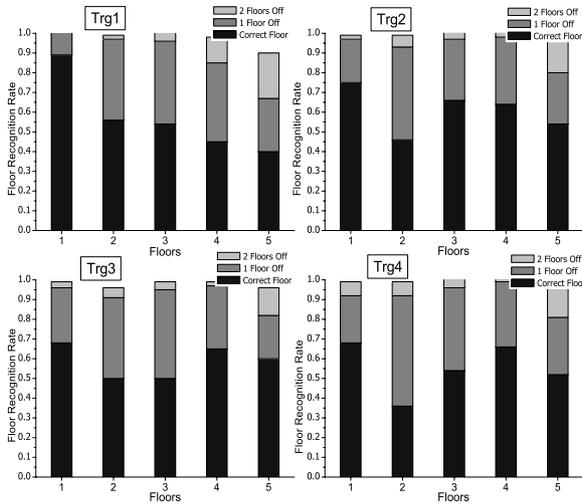


Fig. A-2 NearMe floor recognition results.

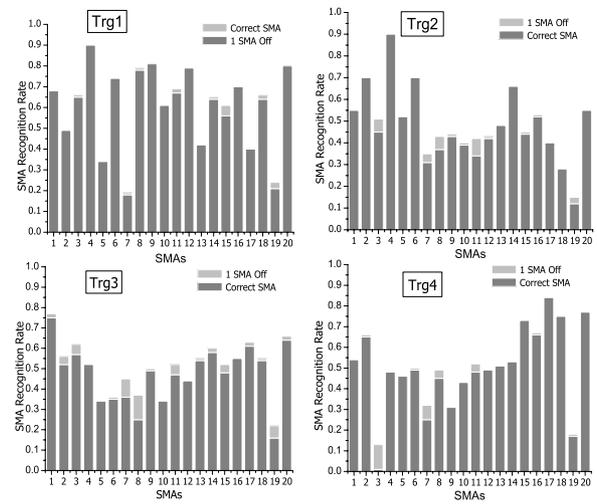


Fig. A-5 SMA recognition results of rice.

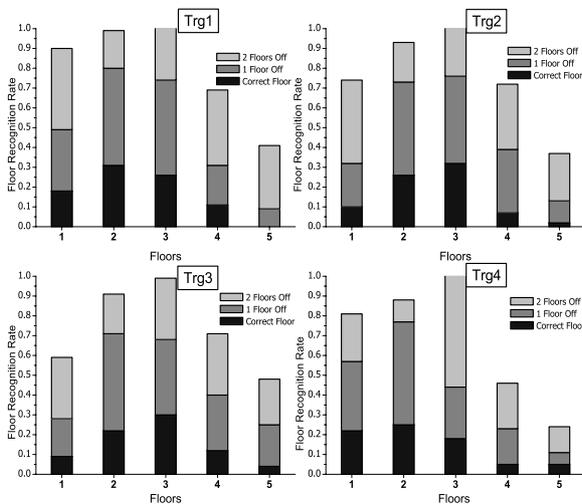


Fig. A-3 SkyLoc floor recognition results.

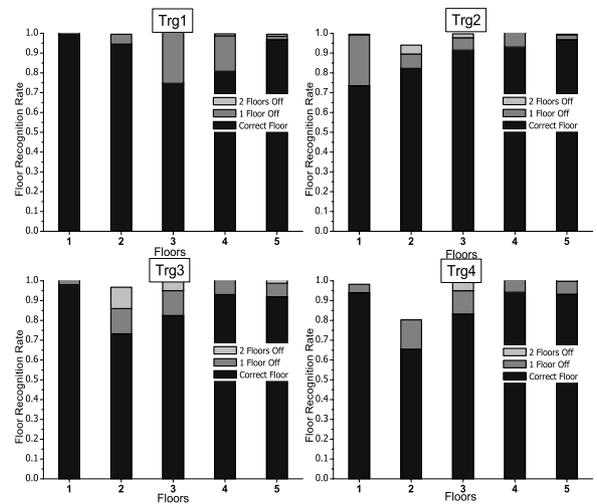


Fig. A-6 3-D smart floor recognition results.

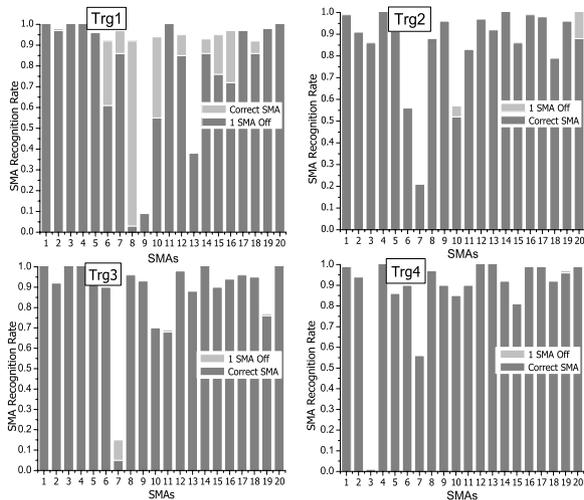


Fig. A-7 3-D smart SMA recognition results.



Sungyoung Lee received his B.S. from Korea University, Seoul, Korea. He got his M.S. and Ph.D. degrees in Computer Science from Illinois Institute of Technology (IIT), Chicago, Illinois, USA in 1987 and 1991 respectively. He has been a professor in the Department of Computer Engineering, Kyung Hee University, Korea since 1993. He is a founding director of the Ubiquitous Computing Laboratory, and has been affiliated with a director of Neo Medical ubiquitous-Life Care Information Technology Research Center, Kyung Hee University since 2006. He is a member of the ACM and IEEE.



Uzair Ahmad received the Masters (IT) degree from the National University of Sciences and Technology (NUST), Pakistan and Ph.D. degree from Kyung Hee University, Republic of Korea, respectively in 2002 and 2008. His research includes Distributed Systems, Reflective Middleware, Location-Aware Computing, and Pattern Recognition algorithms. He is a member of IEEE, and INNS (International Neural Network Society).



Brian J. d'Auriol received the B.Sc. (CS) and Ph.D. degrees from the University of New Brunswick in 1988 and 1995, respectively. He joined the faculty of computer engineering at Kyung Hee University, Republic of Korea, in 2006 as research faculty. His research includes information and data visualization with specialization in program visualization of high performance computing codes as well as in genome sequence visualization in bioinformatics, optical-communication parallel computing models with specialization in optical bus models, communication and computation modeling in parallel and distributed systems, and recently, ubiquitous sensor networks. He is a member of the ACM and IEEE.



Young-Koo Lee got his B.S., M.S. and Ph.D. in Computer Science from Korea Advanced Institute of Science and Technology, Korea. He is a professor in the Department of Computer Engineering at Kyung Hee University, Korea. His research interests include ubiquitous data management, data mining, and databases. He is a member of the IEEE, the IEEE Computer Society, and the ACM.