RF Localization in Indoor Environment

Maja STELLA, Mladen RUSSO, Dinko BEGUŠIĆ

Dept. of Electronics, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, R. Boškovića 32, 21000 Split, Croatia

mstella@fesb.hr, mrusso@fesb.hr, begusic@fesb.hr

Abstract. *In this paper indoor localization system based on the RF power measurements of the Received Signal Strength (RSS) in WLAN environment is presented. Today, the most viable solution for localization is the RSS fingerprinting based approach, where in order to establish a relationship between RSS values and location, different machine learning approaches are used. The advantage of this approach based on WLAN technology is that it does not need new infrastructure (it reuses already and widely deployed equipment), and the RSS measurement is part of the normal operating mode of wireless equipment. We derive the Cramér-Rao Lower Bound (CRLB) of localization accuracy for RSS measurements. In analysis of the bound we give insight in localization performance and deployment issues of a localization system, which could help designing an efficient localization system. To compare different machine learning approaches we developed a localization system based on an artificial neural network, k-nearest neighbors, probabilistic method based on the Gaussian kernel and the histogram method. We tested the developed system in real world WLAN indoor environment, where realistic RSS measurements were collected. Experimental comparison of the results has been investigated and average location estimation error of around 2 meters was obtained.*

Keywords

Indoor localization, Received Signal Strength (RSS), Cramér-Rao Lower Bound (CRLB), location fingerprints.

1. Introduction

Localization using radio signals was first introduced in the World War II to locate soldiers in emergency situations. During the war in Vietnam the Global Positioning System (GPS) was introduced and became available for commercial applications in the 90s of the last century. Although it is the most popular positioning system for open outdoor environments, there is an unmet need for a reliable positioning system that can work indoors, where the microwave radio signals used by the GPS are greatly attenuated [1], [2], [3].

Accurate indoor localization is an important and novel emerging technology [1]. There are numerous important applications in industrial, commercial, public safety, everyday life and military settings [4]. In commercial applications for residential and nursing homes there is an increasing need for indoor localization systems to track people with special needs, the elderly, and children who are away from visual supervision, to navigate the blind, to locate in-demand portable equipment in hospitals, and to find specific items in warehouses. In public safety and military applications, indoor localization systems are needed to track inmates in prisons and navigate policemen, fire fighters, and soldiers to complete their missions inside buildings [1].

The ability of an accurate location determination leads to substantial context aware computing [5] and a great number of useful Location Based Services (LBS). Examples of such applications include asset tracking, context aware computing, pervasive computing, wireless access security, mobile advertising [6], and various personal robotics applications [7].

As new mobile technology comprising highly sophisticated devices as smartphones or notebooks experiences a massive growth these days, context defined by location of the mobile devices grows in importance. Total LBS service revenue in the EU is estimated to exceed $8.3 billion by 2013 (with dominant part of advertising) [2]. Today, many practical applications exploiting the location of people are available. For example, Google offers a service called Latitude which shows the position of the user on Google Map [8], another example is the feature Places of Facebook, where the features of the social networks are consolidated with the actual position of the users [9].

To determine the location of the users within the network it is preferable to employ the existing wireless communications infrastructure. In indoor areas, the wireless communications infrastructure is primarily based on the wireless local area networks (WLANs), in particular the IEEE 802.11 standard. Our focus here thus lies on the experimental results with an IEEE 802.11g WLAN.

In this article we propose a user location determination system in indoor environment based on the RF power measurements and the RSS fingerprinting. In Section 2, we discuss the various wireless positioning techniques used in indoor environments. Section 3 describes the location determination technique based on fingerprinting. In section 4, the Cramér-Rao Lower Bound (CRLB) of localization accuracy for RSS measurements is derived. Section 5 describes the measurement setup of the developed positioning systems and the localization results obtained by an artificial neural network, k-nearest neighbors, probabilistic method based on the Gaussian kernel and the histogram method. We close this paper with a conclusion in Section 6.

2. Indoor Wireless Positioning Techniques

The main elements of the wireless positioning system are a number of location sensing devices that measure metrics related to the relative position of a mobile terminal (MT) with respect to a known reference point (RP), and a positioning algorithm that processes metrics reported by location sensing elements to estimate the location coordinates of MT. The location metrics may indicate the approximate arrival direction of the signal or the approximate distance between the MT and RP. As the measurements of metrics become less reliable, the complexity of the position algorithm increases [3].

The angle of arrival (AOA) is the common metric used in direction-based systems, where additional hardware is needed in order to measure the angle of incidence of the received signal [10]. The received signal strength (RSS) methods use the signal propagation models in estimation of distance of transmitter and receiver [11], [12]. Time of arrival (TOA) [13], and time difference of arrival (TDOA) of the received signal are the metrics used for estimation of distance between transmitter and receiver, both require precise clock synchronization and expensive infrastructure [14], [15]. In indoor environments where conditions of signal propagation are severe (multipath, non-line of sight (NLOS) signal propagation path between the transmitter and receiver), the traditional parametric positioning techniques (RSS, AOA, TOA, TDOA) or their combinations (TDOA with AOA or RSS) fail to provide adequate location accuracy. For these techniques, all the paths used for triangulation must have a line of sight (LOS) to ensure an acceptable accuracy, a condition that is not always met in an indoor environment. Positioning based on the received signals' fingerprint performs better in such an environment when appropriate signatures and pattern-matching algorithms are used [16], [17], [18]. The basic operation of a pattern recognition positioning algorithm is simple. Each indoor environment has unique signal propagation characteristics; each spot in a building would have a unique signature in terms of RSS, TOA, and/or AOA, observed from different sensors in the building. A pattern recognition system determines the unique pattern features (i.e., the location signature) of the area of interest in a training process, and then this knowledge is used to develop rules for recognition.

Broadly speaking, positioning techniques can be classified into two groups: *device-oriented* – where the location is principally determined by measurements of signals emitted from the infrastructure and received by the mobile device, *network-oriented* – where the location is principally determined by measurement of signals emitted from the mobile device and received by sensors within the network infrastructure [19].

Device-oriented approach imposes greater hardware and processing requirements on the mobile device than does the network oriented approach. The biggest advantage is that it allows a mobile device to control the exposure of its location information since the location is determined by the mobile device and not by the infrastructure. The GPS is the most popular and well-known device-oriented location determination system but is limited only to outdoor usage.

Cricket Location Support System [15] uses a combination of RF and ultrasound technologies to track the location of mobile objects in an indoor environment. In the Cricket infrastructure, it is the individual mobile device, rather than a centralized server in the network that is responsible for computing its own location using information gleaned from beacons emitted by wall or ceiling transmitters. As an additional feature, Cricket is also designed to provide orientation information. Significant advantage of this system is its accuracy, it can accurately delineate 4x4 square feet regions within a room. However, the use of ultrasound requires a great deal of infrastructure in order to be highly effective and accurate, and the costs make it inaccessible to most users.

RADAR [18] is a device-oriented RF-based system for locating and tracking users inside buildings. The system requires each mobile wireless device to physically measure the received signal strength of beacons emitted by multiple 802.11 access points at receiver location to estimate the user's coordinates. The system has three access points or fixed stations and covers the entire zone of interest. A pattern-matching algorithm, which consists of the nearest neighbor(s) in signal space, is used to estimate the user's location. Experiments with RADAR indicated ~3 m resolution accuracy at the 50th percentile.

Nibble [20] provides a probabilistic estimation of the location coordinates by incorporating a Bayesian model for predicting the likely origin of 802.11 signal based on the signal quality observed at multiple access points. It relies on a fusion service to infer the location of an object from measured signal strengths. Data are characterized probabilistically and input into the fusion service. The output of the fusion service is a probability distribution over a random variable that represents some context.

In contrast to the device-centric technologies, network-oriented approaches require relatively less sophisticated equipment on the mobile device, since the device merely needs to generate appropriate beacons or pulses. In many cases (e.g., the technologies based on Active (RFID) tags), the mobile device does not directly generate the beacons, but simply relays back appropriately modulated "echoes" of beacons transmitted by the infrastructure nodes.

Active Badge [21] was one of the first indoor location systems. Each active badge is associated with a unique identity (ID), which is broadcast as part of the IR beacons emitted by the badge worn by a user. The Active Badge infrastructure consists of a group of Badge readers (sensors) that are deployed in various locations in the smart environment. The location of a particular active badge (mobile user) is associated with the sensor that currently reports reading from the badge. Active Badge technology does not use any correlation technique (such as triangulation) to further refine the location of a specific badge. The system also requires significant installation and maintenance costs and performs poorly in the presence of direct sunlight, which is likely to be a problem in rooms with windows.

The Active Bat [22] technology was developed as a follow-on to the Active Badge system to obtain higher-resolution location information, since the properties of infrared signals imply that an active badge can be tracked at the granularity of individual rooms by identifying the unique sensor that is currently receiving beacons from the badge. In contrast to active badges, active bats employ ultrasonic (sound waves) technology. Each active bat essentially emits short ultrasonic beacons (pulses), which are then captured at multiple (a minimum of three) receivers or sensors mounted at well-defined reference locations on the ceiling. By accurately synchronizing the clocks between the sensors and an active bat and measuring the time of flight of the ultrasonic signal, each receiving sensor can compute the distance between itself and the bat being tracked. Experiments [23] with the Active Bat system indicate that the trilateration techniques, which may also exploit additional statistical techniques and reflection elimination algorithms to filter out spurious measurements, can provide readings with an accuracy of ~5 – 10 cm in 95 % of cases. However, the performance is influenced by the reflection and obstacles between tags and receivers. Also, deploying a large number of sensors on the ceiling in each room is a time-consuming and costly task.

The LANDMARC (Location Identification Based on Dynamic Active RFID Calibration) prototype [24] is a localization system that employs Active RFID tags. It deploys a group of Active RFID readers over the smart environment with partial overlap between the coverage area of different readers, governed by the power levels associated with each reader. LANDMARC system uses a set of RT reference tags (called landmarks) in a manner similar to RADAR's use of reference locations. The LANDMARC readers receive updates not only from the tag being tracked, but also from this set of static reference tags, whose location is well known. The actual estimated location is computed as a weighted sum of the location of the k-nearest landmark tags. Experimental results demonstrate that the maximum location estimation error using this technique is around 1 – 2 m. The accuracy of LANDMARC, however, depends on the appropriate a priori positioning of the reference tags; determining a good set of locations for a group of RT reference tags is still an open problem.

We base our work on the localization of users in the widely available IEEE 802.11 WLAN network, e.g. RF networks offer a significant advantage over IR networks in terms of range, scalability, deployment and maintenance. Radio waves can travel through walls and human bodies easier, thus the positioning system has a larger coverage area and needs less hardware comparing to other systems. Unlike RFID technology, WLAN technology has been implemented in public areas such as hospitals, train stations, universities, etc. WLAN-based positioning systems reuse the existing WLAN infrastructures in indoor environments, which lowers the cost of indoor positioning. Also, a person already carries possible positioning devices around with them in their daily life such as smart phones, laptops and tablets with WLAN interface. The RSS indicator can be easily read in every 802.11 interface which makes the solution cost effective since only software deployment is required.

Another important aspect of any positioning technology is the support for privacy. Since we implement a device-oriented approach with WLAN passive scanning, privacy can be guaranteed. Advantages of passive scanning are its low power consumption, because no communication is required, and the fact that the user's privacy is completely preserved since his existence is not even revealed. Therefore the wireless user can determine its position but remain private if desired.

Generally, the accuracy of localization systems is affected by various elements in indoor environments such as movement and orientation of human body, walls, doors, etc. Although radio waves can travel through walls and human bodies (unlike the other technologies), these issues still significantly affect the performance. In the WLAN environment, there is also an issue of possible interference with other APs or other RF sources in the same band. One may think that the RSSs from two APs operating in the same channel might interfere with each other. However, results in [25] as calculated by the correlation indicate that both RSSs are independent and do not interfere with the reception of each other. This is due to the way in which the 802.11 MAC operates where a transmission is either not heard or is deferred if a competing transmission exists. Disadvantage of an RF based approach is that the RSS values are highly susceptible to multipath effects, and the signal strength cannot be easily expressed with some propagation model. To deal with this issue, fingerprinting approach can be applied to construct the radio map of entire area. However, significant changes in the environment, such as moving furniture or large equipment, could require a reconstruction of the fingerprint database.

Having the major advantage of exploiting already existing 802.11 network infrastructures, currently the most viable solution for RSS-based indoor positioning is the fingerprinting architecture [26], [27], [28] [29], [30].

3. Location Determination Based on Fingerprinting

A location fingerprint based on RF characteristics such as RSS is the basis for representing a unique position or location. It is created under the assumption that each position or location inside a building has a unique RF signature. The process is composed of two phases: a phase of data collection called off-line phase and a phase of locating a user in real-time (Fig. 1). The first phase consists of recording a set of RSS fingerprints in a database as a function of the user's location covering the entire zone of interest and using this data as input and as the target of pattern matching algorithm. During the second phase, a RSS fingerprint is measured by a receiver and applied on pattern-matching algorithm to obtain location.

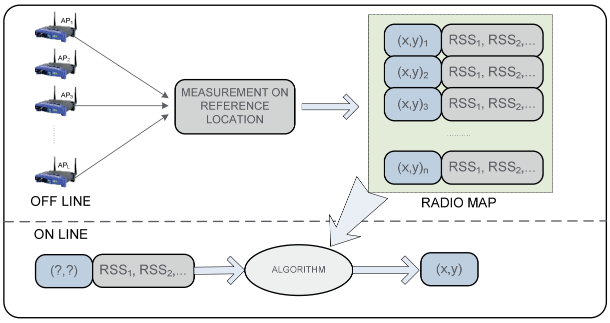

Fig. 1. Location determination based on RSS fingerprints.

Pattern matching algorithms can be classified into deterministic and probabilistic types based on the approaches that model the relationship between location fingerprints and location. The deterministic types of algorithms are those that are based on the nearest neighbor classifiers and the neural network classifiers. The probabilistic types of algorithms are those that are based on the statistical learning theory. Several localization systems using the fingerprinting technique have been recently deployed in outdoor and indoor environments. The main differences between these systems are the types of fingerprint information and pattern matching algorithms [18], [29], [31].

3.1 Artificial Neural Network

A trained artificial neural network can perform complex tasks such as classification, optimization, control and function approximation [32], [33]. Artificial neural network (ANN) can be used to establish a relationship between pattern of RSS samples and location. The pattern-matching algorithm of the system can be viewed as a function approximation problem consisting of a nonlinear mapping from a set of input variables (RSS from N access points) into two output variables representing the two dimensional location *(x, y)* of the mobile station.

An ANN is consisting of processing units which communicate by sending signals to each other over a large number of weighted connections. The total input to unit k is simply the weighted sum of the separate outputs from each of the connected units plus a *bias* or offset term θ_k:

$$s_k(t) = \sum_j w_{jk}(t) y_j(t) + \theta_k(t) \,. \tag{1}$$

Generally, for activation function y_k some sort of threshold function is used: a hard limiting threshold function (a sgn function), or a linear or semi-linear function, or hyperbolic tangent function. One of the most popular ANNs is the MultiLayer Percepton (MLP). It is a feed-forward layered structure. Each layer consists of units which receive inputs from units from layer directly below and send their output to units in a layer directly above. There are no connections within a layer. Back-propagation learning rule is used for finding the optimal weights.

3.2 Distance based Approach

Deterministic distance based approach requires a set of constant location fingerprints from N access points which include mean vector and standard deviation vector. Let fingerprint F_i (labeled with a location information L_i) be an RSS vector on ith referent location measured during the off-line phase and F denotes RSS vector on unknown location measured during the on-line phase. In order to determine the location, a form of discriminant function is commonly used to classify a sample of RSS fingerprints into a location [34]. Location L can be estimated based on the distance measurements in signal space. Some of the distance metrics are given in Tab. 1.

Name	L
p-distance	$L_p = \frac{1}{N}(\sum_{i=1}^{N} \lvert F - F_i \rvert^p)^{\frac{1}{p}}$
Modified p-distance	$L_{mp} = \frac{1}{N}(\sum_{i=1}^{N} \frac{1}{w_i} \lvert F - F_i \rvert^p)^{\frac{1}{p}}$
Mahalanobis distance	$L_M = \sqrt{(F - F_i)^T \sum^{-1} (F - F_i)}$

Tab. 1. Distance in signal space.

Location estimation is obtained from the training examples whose observation vector has a minimal distance when compared with test observation.

The simplest distance is for $p = 1$, it is a Manhattan distance in which the sum of the absolute differences of signal is computed.

If $p = 2$ it is an Euclidian distance, a well-known distance metric used for classifying the positions [18], [35]. Modified p distance is a weighted distance, where w_i are weighting factors (w_i<*1*) that can promote or demote some RSS component in a fingerprint. If all weights are equal, it is a k-nearest neighbor method, and for *k=1* is simply a nearest neighbor method. K-nearest neighbor is based on the assumption that the averaging may yield to an estimate that is closer to the user's true location than any individual neighbor.

Mahalanobis distance is based on correlations between location fingerprints, where Σ is the covariance matrix for the location fingerprint. If RSS values from different APs are assumed to be mutually independent then the covariance matrix becomes a diagonal matrix [36].

3.3 Probabilistic Approach

Probabilistic approach is based on the empirical model that describes the distribution of received signal power at various locations. Let L denote location and F denote an observation variable or vector in the same area A. We assume that the observation variable is a vector of received signal strength values for a set of access points. The training data D consists of n examples, denoted by (L_i, F_i). Let $p(\ )$ denote all probability distributions for either discrete or continuous variables. For the location estimation problem different models that estimate the probability distribution of the observation variable can be used. Simply applying the Bayes rule, we can then obtain the so-called posterior distribution of the location:

$$p(L \mid F) = \frac{p(F \mid L)p(L)}{p(F)} = \frac{p(F \mid L)p(L)}{\sum_{i \in L} p(F \mid L_i)p(L_i)} \tag{2}$$

Where $p(F|L)$ is the likelihood function, the prior distribution $p(L)$ is the prior probability of being at location L before knowing the value of the observation variable, $p(F)$ doesn't depend on location variable L, and can be seen as a normalizing constant, and the summation goes over the set of possible locations. The prior distribution $p(L)$ gives an ability to incorporate background information such as personal user profiles.

There are several approaches for computing likelihood $p(F|L)$ [34]. Some of them are given in Tab. 2.

Name of method	$P(F \mid L)$
Histogram	$\sum_i H(F_i)$
Kernel	$\frac{1}{nh}\sum_i K(\frac{F-F_i}{h})$
Gaussian	$\frac{1}{n}\sum_i \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{(F_i-\mu)^2}{2\sigma^2}\right)}$
Log-normal	$\frac{1}{n}\sum_i \frac{1}{F_i\sqrt{2\pi}} e^{\left(-\frac{(\ln F_i-\mu)^2}{2\sigma^2}\right)}$

Tab. 2. Likelihood for location L.

Estimation of the likelihood function is obtained from the measurements of received signal strengths at each location. In case of the missing RSS values from some access points at some location, special heuristic for handling the missing data is required, but the simplest way is to replace the missing values with a constant smaller than any of the measured RSS values [37].

4. Theoretical Analysis on Lower Bound of Location Estimation Error

Since nowadays WLAN infrastructure, besides providing data communication, is currently the most promising approach for localization systems, during network deployment both coverage and localization requirements should be taken into account. Impact of some parameters (such number of APs, geometry of deployed APs, etc.) can be investigated through the properties of the lower bound on location estimation error. Performance of the RSS based localization system can be theoretically expressed by the Cramér-Rao Lower Bound [38]. For the theoretical propagation model we can derive lower bound on the variance of location estimation error for any unbiased estimator defined by CRLB. Basic estimation problem for n observations $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is to define the "best" estimator for unknown parameter θ on which these observations are dependent. Precisely, basic estimation assumption is that the joint probability density function of $X_1, X_2,\ldots,X_n$ given by $f_X(x_1, x_2, \ldots, x_n, \theta)$ depends on the unknown parameter θ. Since all information is contained in the observations and the corresponding probability density function (PDF), estimation accuracy will be directly dependent of that PDF. Thus it is unlikely to expect that the parameter θ can be accurately estimated if the PDF is weakly, or in extreme case, not dependent on that parameter. Generally, the more dependent the PDF is on the unknown parameter, the better it can be estimated.

Let $\hat{\theta}(X)$ denote the estimator for θ. Obviously $\hat{\theta}(X)$ is a function of only the observations. In ideal case the estimate $\hat{\theta}(X)$ would coincide with the unknown parameter θ. This may not be possible, and the estimation will almost always result in an error given by $e = \hat{\theta}(X) - \theta$. Generally, the estimator has to minimize that error. The Cramér-Rao Lower Bound theorem defines the lower bound on the variance estimation error, where the variance of any unbiased estimator $\hat{\theta}$ must satisfy the lower bound

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta^2}\right]}, \tag{3}$$

It is assumed that the probability density function (PDF) satisfies regularity condition

$$E\left[\frac{\partial \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta}\right] = 0 \text{ for all } \theta .$$

Expression in the denominator of the inequality denotes Fisher information $J(\theta)$ for data X

$$J(\theta) = -E\left[\frac{\partial^2 \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta^2}\right]. \tag{4}$$

In order to investigate the theoretical lower bound on variance estimation error of the RSS based localization system CRLB has been derived as follows.

Received signal strength is generally a random variable and a shadowing model is often used for describing the propagation of signal strength [39]

$$P(d) = P(d_0) - 10n\log(\frac{d}{d_0}) + X_{sh} \tag{5}$$

where $P(d)$ in dBm denotes path loss at distances d, d_0 is a reference distance, n is the path-loss exponent which indicates the rate at which path loss increases with distance and X_{sh} is a log-Gaussian distributed random variable with standard deviation σ in dBm which describes losses due to random shadowing effects.

Received signal strength can be described as a probability density function

$$f(P;\theta) = \frac{1}{\sqrt{2\pi}\sigma P(d)} e^{-\frac{\left(\ln P(d) - P(d_0) + 10n\log\frac{d}{d_0}\right)^2}{2\sigma^2}} \tag{6}$$

where $d = \sqrt{(x-x_i)^2 + (y-y_i)^2}$, (x_i,y_i) are the coordinates of i-th access point and (x,y) are the coordinates of a mobile terminal. For N access points we have a joint probability density function

$$f(P;\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma P(d_i)} e^{-\frac{\left(\ln P(d_i) - P(d_0) + 10n\log\frac{d_i}{d_0}\right)^2}{2\sigma^2}} . \tag{7}$$

If $\hat{\theta} = (\hat{x}, \hat{y})^T$ is a location estimate of $\theta = (x, y)^T$ then the variance of location estimation is

$$Var(\hat{\theta}) \geq \{J(\theta)\}^{-1}, \tag{8}$$

where

$$J(\theta) = E\left\{\left[\frac{\partial \ln f(P)}{\partial \theta}\right]^2\right\}. \tag{9}$$

It can be derived that

$$J = -\frac{20n}{\sigma^2 \ln 10}\begin{bmatrix} \sum_i \frac{(x-x_i)^2}{d^4} & \sum_i \frac{(x-x_i)(y-y_i)}{d^4} \\ \sum_i \frac{(x-x_i)(y-y_i)}{d^4} & \sum_i \frac{(y-y_i)^2}{d^4} \end{bmatrix}, \tag{10}$$

from which follows the lower bound on the variance of location estimation

$$Var(\hat{\theta}) \geq \frac{\sigma^2 \ln 10}{20n} \cdot \frac{\sum_i \frac{(x-x_i)^2 + (y-y_i)^2}{d^4}}{\sum_i \frac{(x-x_i)^2}{d^4} \sum_i \frac{(y-y_i)^2}{d^4} - \left(\sum_i \frac{(x-x_i)(y-y_i)}{d^4}\right)^2} \tag{11}$$

From (11) it is evident that the CRLB depends on the

- number of access points N,
- geometry of access points (x_i,y_i),
- geometry of mobile terminal (x,y),
- propagation model parameters σ and n.

In order to investigate the influence of the number and geometry of access points on performance of a localization system, simulation has been made. In Fig. 2, the CRLB for a case of four APs placed on the edges of the square area is shown. 3D and 2D plots of the lower bound on the variance are given in the left and right panels, respectively. It can be seen that the variance is smaller in the center of the area while rapidly grows towards the edges.

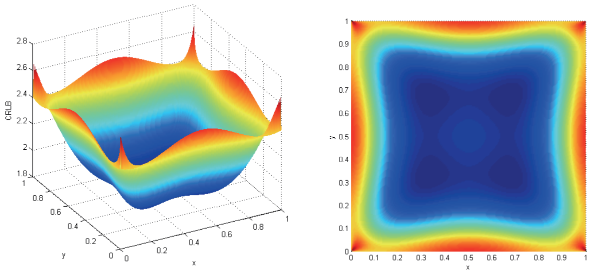

Fig. 2. Lower bound on variance of location estimation.

It can be shown that the optimal geometry occurs when APs are placed on the vertices of a regular polygon [40]. First, we analyzed the arrangement of four APs on vertices of squares of different sizes. Test has been made for 10 different arrangements, from A1 which denotes the arrangement of APs on the edges of the biggest square to A10 for the smallest square. Fig. 3 shows the 2D plots of the CRLB for the arrangements A1, A4, A7 and A10.

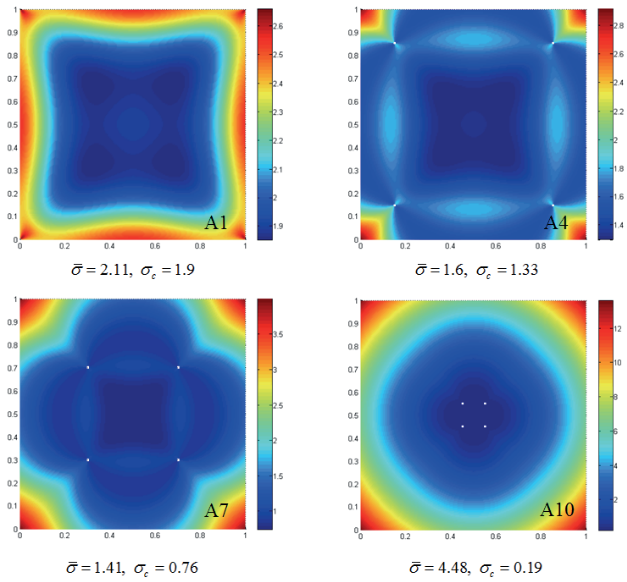

Fig. 3. Lower bound on the variance of estimation for different arrangement of four Aps.

For each of these arrangements of four APs, the variance in the center σ_c and the mean variance $\bar{\sigma}$ have been calculated and shown below each figure. It can be seen that mean value of the variance is decreasing and after the arrangement A6 begins to increase. It can be explained by the fact that the area inside the polygon has much lower variance than the area outside. With the approaching of APs to the center, variance in the center gets smaller with each arrangement, but the space inside the polygon also gets smaller, and on average the variance increases. It is illustrated in Fig. 4, where the mean value of the variance for the whole area and the variance value in the center are shown.

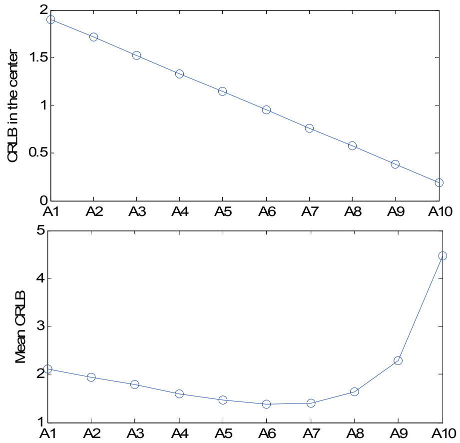

Fig. 4. Influence of the arrangement of four APs on the variance in the center and the mean variance.

We have also analyzed the influence of a different number of APs placed on the vertices of a regular polygon starting from triangle, square, pentagon, etc. In Fig. 5, 2D representations of the lower bound on the variance for 3, 4, 5 and 6 APs arrangement on vertices of regular polygons are given. Generally, all APs are arranged on the same circle for fair comparison (circle is denoted with dashed green line). For each of these arrangements, the variance in the center σ_c and the mean variance $\bar{\sigma}$ have been calculated and shown below each plot. In Fig. 6, the mean value of the variance for the whole area and the variance value in the center are shown.

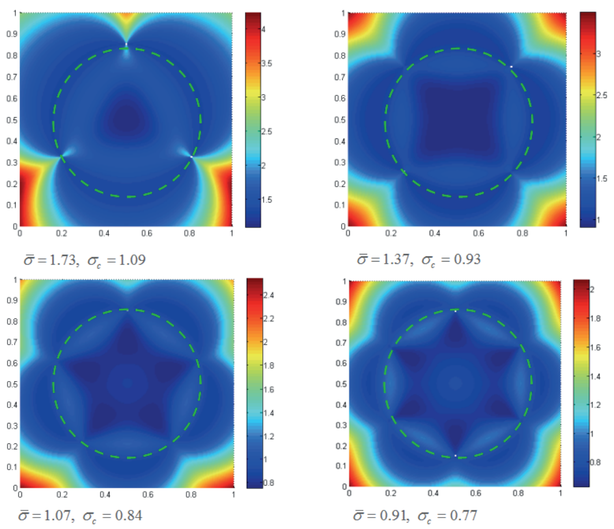

Fig. 5. Lower bound on the variance of estimation for different regular polygon arrangements of APs.

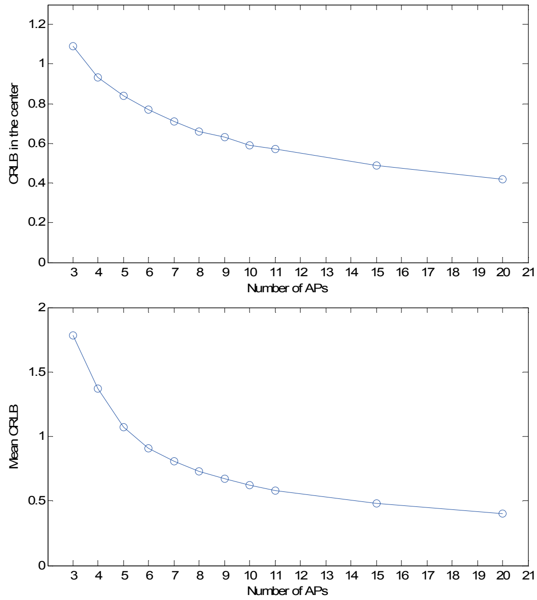

Fig. 6. Influence of different arrangements of APs on the variance in the center and the mean variance.

From the calculated values of the variances in the center and the mean variance values it can be seen that both of them become lower with the increasing number of APs. Obviously, it is better to have a greater number of APs, but after some point improvement becomes less significant.

5. Empirical Results

Our developed positioning system is based on the WLAN using IEEE 802.11 standard in six storey building. Measurements were made in the part of the fourth floor, dimensions of approximately 28m×15m, total area 420m^2. Area includes 4 offices, 3 laboratories, a classroom and a hallway. The layout of the floor and locations of the APs are shown in Fig. 7.

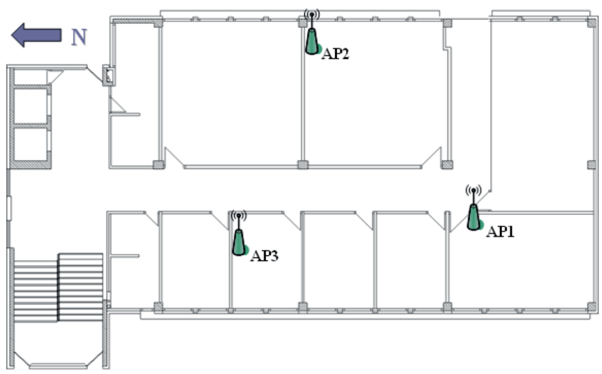

Fig. 7. The test location layout with positions of the access points.

We used three Access Points (AP) WRT54GS from Linksys which are IEEE 802.11b/g compatible. For collection of the RSS samples from APs we used a Fujitsu-Siemens laptop with the Network Stumbler software [41]. The WLAN Proxim Orinoco card was plugged into the

PCMCIA slot on the right side of the laptop. To collect the RSS samples, the laptop was placed on the box approximately one-meter high.

The radio frequency channels of IEEE 802.11b/g are in the 2.4 GHz band which is shared by other equipment in the industrial, scientific, and medical (ISM) band such as the Bluetooth. The number of non-overlapping channels for 802.11b/g is three [42]. The radio channels we used for each AP are channel 1, 6, and 11, respectively, as can be seen in Fig. 8.

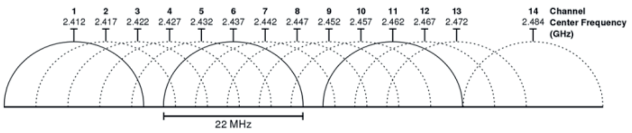

Fig 8. Graphical representation of 802.11 channels in 2.4 GHz band [43.]

We observed that the RSS values reported by the WLAN card were an average value over a sampling period and in integral steps of 1 dBm. The received signal sensitivity of the WLAN card also limits the range of the RSS to be between -94 dBm and 0 dBm. Nevertheless, the highest typical value of the measured RSS was approximately -15 dBm (WLAN card near the AP antenna). The measurement was done by sampling the RSS data every one second. The vector of RSS data at each location forms the location fingerprint with RSS elements in the vector. The locations of APs on this floor are labeled as AP1, AP2 and AP3.

Locations in terms of coordinates for the measurement of RSS have been chosen and stored together with three measurements of RSS values for given location. Total number of measurements was 125, 110 for training and 15 for testing. Collecting enough statistics for creating location fingerprints is the key to achieving good performance with any indoor positioning system. The duration of data collection in the literature is different due to the sampling period. For instance, RADAR [18] used a 0.25-second sampling period. The sampling period is limited by either the software or the hardware. The software limits depend on how often a device driver can be accessed and how often the BSSID scan list is updated. Some wireless cards have the capability to scan for APs' signal in the background [44]. The hardware limits depend on how the vendor implements the scanning cycle and the amount of the channel dwelling time.

The RSS sampling period in our measurement was one second, with 400 samples per location. Measurement locations were not forming the regular grid due to office and laboratory equipment, inaccessible areas, etc. In Fig. 9, RSS values from three APs are shown at one measurement location. It can be seen that the measured signal strength at a fixed position varies over time and the variations can be up to 10 dBm.

In Fig. 10, 2D propagation of the signal strength of AP1 is plotted. Colors denote signal strength, blue presents the weakest signal and red the strongest signal. For AP1 signal strength is from -86.4 dBm to -45.8 dBm.

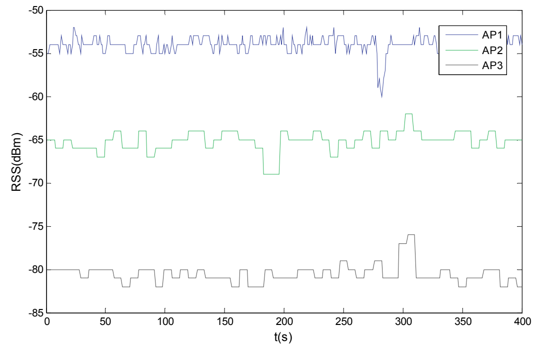

Fig. 9. RRS values from three AP.

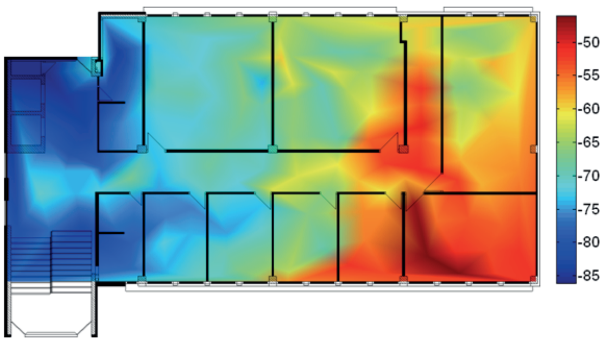

Fig. 10. 2D propagation of the signal strength of AP1.

Location fingerprint presents a vector of RSS values at each location from all APs. Since we used three APs, location fingerprints can be visualized on 3D graph where on coordinate axes the RSS values of each AP are labeled, as can be seen in Fig. 11. Fingerprints of four locations are shown with red, green, blue and magenta circles (green and red belong to two close locations – 1.93m). It can be seen that patterns of the same location are grouped together as a cluster and concentrate around average value. Fig. 11 indicates that the RSS patterns can be separated by some pattern matching algorithm (artificial neural network, nearest neighbor, probabilistic approach).

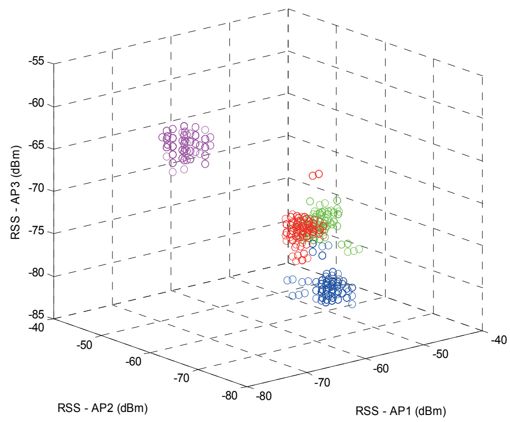

Fig. 11. Location fingerprinting.

The first method we applied was the Multi-Layer-Perceptron (MLP) feed-forward artificial neural network (Fig. 12) consisting of three inputs (received RSS from three APs), outputs with two neurons (corresponding to location of a user (x,y) and one hidden layer with different number of neurons.

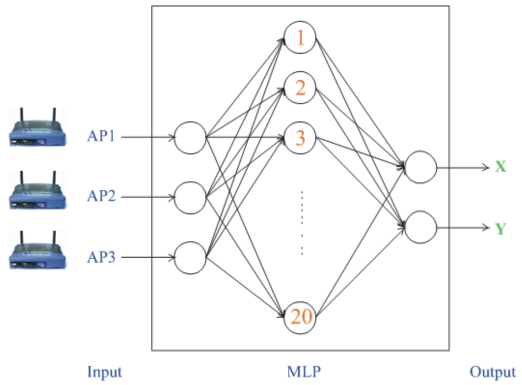

Fig. 12. MLP neural network.

From the 125 measured data, 100 patterns have been employed to train the network, 10 for the validation purpose and the remaining 15 non-training patterns have been applied to the network for testing the proposed location system. In order to train the network, these patterns have been applied to the pattern-matching neural network together with location coordinates. Criterion for stopping of the network training was chosen as a moment after which the performance of validation set terminated to enhance. Fig. 13 shows train, test and validation performance. Iteration at which the validation performance reached a minimum is denoted with green circle.

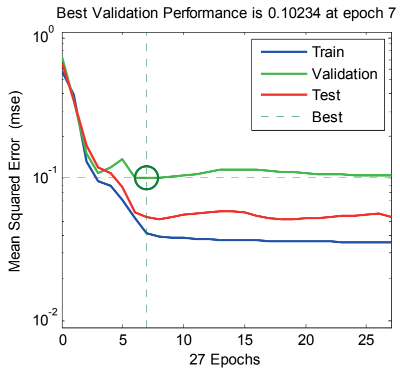

Fig. 13. Train, test and validation performance.

During the training, we experimented with several topologies with a different number of hidden layers, but the results were quite similar. Experimenting with a number of hidden layer neurons, we found that 20 neurons are adequate to achieve minimal mean distance test error.

Second method we used was the k-nearest neighbors. It has been shown that the maximum improvement can be obtained for k = 2-4, for larger k accuracy degrades rapidly because points far removed from the true location also are included in the averaging procedure, thereby corrupting the estimate [18]. Thus we used three nearest neighbors in our positioning system.

In this work we also used the statistical approach. In the statistical approach, to determine the position it is necessary to know the probability distribution for each training location. Probability distribution can be expressed with the histogram or can be approximated by the Gaussian kernel with the appropriate mean value and the variance. In this work we used both approaches, the histogram based and the Gaussian kernel based approach. As a performance metric of developed systems, we used the Euclidian distance between the estimated and true location.

The results of positioning accuracy for all four methods are given in Tab. 3 (mean error, 50 percentile error and 95 percentile error) in meters for location determination based on the Multi-Layer-Perceptron (MLP), three nearest neighbors (3NN), statistical approach based on the histogram (HIST) and statistical approach based on the Gaussian kernel (GAUSS).

Location Estimation Errors			
Method	mean± variance	50 %	95 %
MLP	2.35 ± 1.62	2.23	6.38
3NN	2.79 ± 1.19	3.29	5.41
HIST	4.14 ± 3.56	2.82	9.76
GAUSS	2.05 ± 0.98	2.20	7.83

Tab. 3. Location Estimation Errors (mean, 50 and 95 percentiles in meters).

Positioning accuracy indicated by the cumulative percentage of localization error is plotted in Fig. 15.

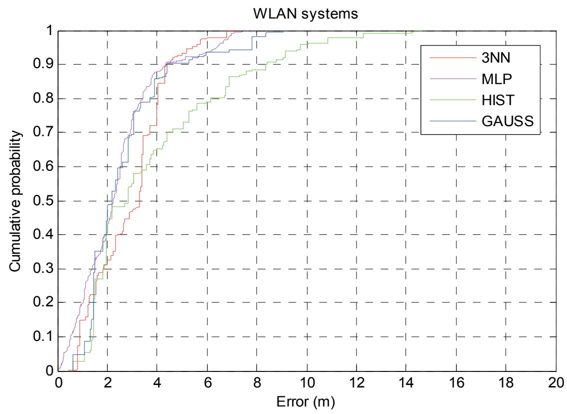

Fig. 14. Accuracy comparison.

It can be seen that all four methods performed well in our experiments, with the probabilistic method based on the Gaussian kernel leading with slightly lower localization error of around 2 m on average. Finally, our experimental results demonstrate that practical applications based on these location estimation methods are quite feasible.

In comparison with other positioning systems (some of them briefly described in Section 2), one should consider the accuracy of the systems as well as their costs/complexities. Furthermore, the accuracy cannot be directly compared since the measurement areas are completely different in terms of size and complexity (some systems are tested only in hallways, others in highly complex areas). Our system is deployed in 420 m^2 area with 4 offices, 3 laboratories (with various electronic equipment and telephone exchange along the wall in one lab), classroom and hallway. Also, a different number of APs are used which also affects the performance. E.g. RADAR experiments [18] (measured only in hallways) indicate ~3 m resolution accuracy at the 50th percentile, LANDMARC experiments [24] (in the small 40 m^2 area) demonstrate that the location error is around 1-2 m. On the other hand, ultrasonic ActiveBat system [22] achieves less than 10 cm location errors, but as discussed in Section 2, this type of positioning system requires costly infrastructure. The advantage of our approach based on WLAN technology is that it does not need new infrastructure (it reuses already and widely deployed equipment), and the RSS measurement is part of the normal operating mode of wireless equipment (RSS indicator can easily be read in every 802.11 interface) which makes the solution cost effective since only software deployment is required.

6. Conclusion

In this paper, we studied the WLAN user location estimation problem in the machine learning framework. We derived the CRLB of localization accuracy for RSS measurements. We investigated the influence of geometry and quantity of APs on this bound of localization error. In the experimental part of this work, four different machine learning approaches were considered: two non-probabilistic methods – k-nearest neighbor method and artificial neural network and two probabilistic approaches – one based on the Gaussian kernel and one based on the histogram. We compared the performance of the four developed positioning systems and obtained quite similar results, with the probabilistic method based on the Gaussian kernel leading with slightly lower localization error of around 2 m on average. The results showed that practical applications for location based services based on this type of machine learning approaches are quite feasible, especially considering the fact that they can be deployed on existing wireless communications infrastructure.

Acknowledgements

Authors are grateful to the anonymous reviewers for their helpful comments. They also wish to note that the research work has been performed within the project "Advanced Heterogeneous Network Technologies" (023-0231924-1660) supported by the Ministry of Science and Technology of the Republic of Croatia.

References

[1] SAYED, A. H., TARIGHAT, A., KHAJEHNOURI, N. Network based wireless location: challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Processing Magazine*, 2005, vol. 22, no. 4, p. 24 - 40.

[2] FANG, S., LIN, T. N. Indoor location system based on discriminant – adaptive neural network in 802.11 environments. *IEEE Transactions on Neural Networks*, 2008, vol. 19, no. 11, p. 1973 - 1978.

[3] PAHLAVAN, K., XINRONG, L., MAKELA, J. P. Indoor geolocation science and technology. *IEEE Communications Magazine*, 2002, vol. 40, no. 2, p. 112 - 118.

[4] GUSTAFSSON, F., GUNNARSSON, F. Mobile positioning using wireless networks. *IEEE Signal Processing Magazine*, 2005, vol. 22, no. 4, p. 41 - 53.

[5] HIGHTOWER, J., BORRIELLO, G. Location systems for ubiquitous computing. *IEEE Computer*, 2001, vol. 34, no. 8, p. 57 - 66.

[6] YUNOS, M., GAO, J. Z., SHIM, S. Wireless advertising's challenges and opportunities. *IEEE Computer*, 2003, vol. 36, no. 5, p. 30 - 37.

[7] JENSFELT, P. *Approaches to Mobile Robot Localization in Indoor Environments*. PhD Thesis. Stockholm (Sweden): Royal Institute of Technology, 2001.

[8] *Techcrunch*. [Online]. 2010. Available at: http://techcrunch.com/ 2010/05/06/ google -latitude-users-check-in/

[9] *Facebook*. [Online]. Available at: http://www.facebook.com/ places/.

[10] NICULESCU, D., BADRI, N. Ad hoc positioning system (APS) using AOA. In *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2003.* San Francisco (USA), 2003, p. 1734 - 1743.

[11] MAZUELAS, S., et al. Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks. *IEEE Journal of Selected Topics in Signal Processing*, 2009, vol. 3, no. 5, p. 821 - 831.

[12] MORAVEK, P., KOMOSNY, D., SIMEK, M., GIRBAU, D., LAZARO, A. Energy analysis of received signal strength localization in wireless sensor networks. *Radioengineering*, 2011, vol. 20, no. 4, p. 937 - 945.

[13] VEGNI, A. M., DI NEPI, A., NERI, A., VEGNI, C. Local positioning services on IEEE 802.11 networks. *Radioengineering*, 2008, vol. 17, no. 2, p. 42 - 47.

[14] XINRONG, L., PAHLAVAN, K. Super-resolution TOA estimation with diversity for indoor geolocation. *IEEE Transactions on Wireless Communications*, 2004, vol. 3, no. 1, p. 224 - 234.

[15] PRIYANTHA, N. B., CHAKRABORTY, A., BALAKRISHNAN, H. The cricket location-support system. In *Proceedings of ACM International Conference on Mobile Computing and Networking (MOBICOM'00).* Boston (USA), 2000, p. 32 - 43.

[16] TAHERI, A., SINGH, A., AGU, E. Location fingerprinting on infrastructure 802.11 wireless local area networks. In *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, LCN '04*. Washington (USA), 2004, p. 676 - 683.

[17] NERGUIZIAN, C., DESPINS, C., AFFES, S. Geolocation in mines with an impulse response fingerprinting technique and neural networks. *IEEE Transactions on Wireless Communications*, 2006, vol. 5, no. 3, p. 603 - 611.

[18] BAHL, P., PADMANABHAN, V. N. RADAR: an in-building RF-based user location and tracking system. In *Proceedings of IEEE Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00)*. Tel Aviv (Israel), 2000, p. 775 - 784.

[19] COOK, D., DAS, S. *Smart Environments: Technology, Protocols and Applications.* Hoboken (NJ, USA): Wiley-Interscience, 2004.

[20] CASTRO, P., CHIU, P., KREMENEK, T. MUNTZ, R. A Probabilistic room location service for wireless networked environments. In *Proceedings of Ubicomp 2001: Ubiquitous Computing*. Atlanta (USA), 2001, p. 18 - 34.

[21] WANT, R., HOPPER A., FALCAO, V., GIBBONS, J. The active badge location system. *ACM Transactions on Information Systems*, 1992, vol. 10, no. 1, p. 91 - 102.

[22] HARTER, A., HOPPER, A., STEGGLES, P., WARD, A., WEBSTER, P. The anatomy of a context-aware application. In *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, MobiCom '99*. Seattle (USA), 1999, p. 59 - 68.

[23] ADDLESSE, M., CURWEN, R., HODGES, S., NEWMAN, J., STEGGLES, P., WARD, A., HOPPER, A. Implementing a sentient computing system. *IEEE Computer Magazine*, 2001, vol. 34, no. 8, p. 50 - 56.

[24] NI, L., LIU, Y., LAU, Y. C., PATIL, A. LANDMARC: Indoor location sensing using active RFID. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PERCOM'03)*. Dallas (USA), 2003, p. 407 - 415.

[25] KAEMARUNGSI, K., KRISHNAMURTHY, P. Properties of indoor received signal strength for WLAN location fingerprinting. In *Proceedings of First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04)*. Boston (USA), 2004, p. 14 - 23.

[26] FANG, S. H., LIN, T. N. Location fingerprinting in a decorrelated space. *IEEE Transactions on Knowledge and Data Engineering*, 2008, vol. 20, no. 5, p. 685 - 691.

[27] SUN, G., CHEN, J., GUO, W., LIU, K.J. Signal Processing Techniques in Network-Aided Positioning. *IEEE Signal Processing Magazine*, 2005, vol. 22, no. 4, p. 12-23.

[28] KJĘRGAARD, M. B., TREU, G., LINNHOFF-POPIEN, C. Zone-based RSS reporting for location fingerprinting. In *Proceedings of Pervasive 2007, the 5th International Conference on Pervasive Computing*. Toronto (Canada), 2007, p. 316 - 333.

[29] BATTITI, R., BRUNATO, M., VILLANI, A. *Statistical Learning Theory for Location Fingerprinting in Wireless LANs*. Technical Report, 2002.

[30] YOUSSEF, M. A., AGRAWALA, A., SHANKAR, A. U. WLAN location determination via clustering and probability distributions. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, PerCom '03*. Dallas (USA), 2003, p. 143 - 150.

[31] FANG, S. H., LIN, T. N. A dynamic system approach for radio location fingerprinting in wireless local area networks. *IEEE Transaction on Communications*, 2010, vol. 58, no. 4, p. 1020 - 1025.

[32] KROSE, B., SMAGT, P. *An Introduction to Neural Networks*. Amsterdam (The Netherlands): University of Amsterdam, 1996.

[33] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. New York (USA): Maxwell Macmillan, 1994.

[34] DUDA, R. O., HART, P. E., STORK, D. G. *Pattern Classification*. 2nd ed. New York (USA): Wiley, 2000.

[35] PRASITHSANGAREE, P., KRISHNAMURTHY, P., CHRYSANTHIS, P. On indoor position location with wireless LANs. In *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. Lisbon (Portugal), 2002, p. 720 - 724.

[36] KAEMARUNGSI, K. Efficient design of indoor positioning systems based on location fingerprinting. In *Proceeding of International Conference on Wireless Networks, Communications and Mobile Computing*. Maui (USA), 2005, p. 181 - 186.

[37] ROOS, T., MYLLYMAKI P., TIRRI H., MISIKANGAS P., SIEVANEN, J. A probabilistic approach to wlan user location estimation. *International Journal of Wireless Information Networks*, 2002, vol. 9, no. 3, p. 155-164.

[38] KAY, S. M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[39] RAPPAPORT, T. S. *Wireless Communications: Prinples and Practice*. Prentice-Hall Inc, 1996.

[40] MAHTAB HOSSAIN, A. K. M., SOH, W.-S. Cramér-Rao bound analysis of localization using signal strength difference as location fingerprint. In *Proceedings IEEE INFOCOM*. San Diego (USA), 2010.

[41] *NetStumbler.com*. [Online] Available at: http://www.netstumbler .com/ .

[42] *IEEE std 802.11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*. IEEE, 1997.

[43] Wikipedia. *2.4 GHz Wi-Fi channels.* [Online]. Available at: http://en.wikipedia.org/wiki/File:2.4_GHz_Wi-Fi_channels_(802. 11b,g_WLAN).svg .

[44] Microsoft Corporation. *IEEE 802.11 Network Adapter Design Guideline for Windows XP.* (White paper). 2003.

About Authors ...

Maja STELLA was born in 1976 in Split, Croatia. She graduated in 1999 from the University of Split, Croatia. She has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, Croatia since 2001, where she received the M.S. and Ph.D degrees in 2007 and 2011, respectively. Her research interests include wireless positioning, signal processing and pattern recognition. She is a member of IEEE and CCIS.

Mladen RUSSO was born in 1977 in Split, Croatia. He graduated in 2001 from the University of Split, Croatia. He has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, Croatia since 2001, where he received the M.S. and Ph.D degrees in 2006 and 2010, respectively. His research interests include speech recognition, signal processing and wireless positioning. He served as a general secretary of the International Conference on Software, Telecommunications and Computer Networks SoftCOM from 2005 to 2007. He is a member of IEEE and CCIS.

Dinko BEGUŠIĆ received the B.S. degree in electrical engineering from the University of Split, Croatia in 1983, and the M.S. and Ph.D degrees from the University of Zagreb, Croatia, in 1988 and 1992, respectively. Since 1985, he has been with the University of Split, Croatia. He is a professor and the chair of the communication technologies and signal processing at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB). He was a visiting researcher with the Universite Libre de Bruxelles, Bruxelles, Belgium (1990/1991) and with the King's College London, London, UK (1992). In 1997/1998 he was with the University of Texas at Dallas, Richardson, TX, as a Visiting Assistant Professor. His research interests include information and communication systems and networks, software engineering in telecommunications, digital signal processing for communications and adaptive algorithms. Dr.Begusic is a Co-Chair of the Conference on Software, Telecommunications and Computer Networks SoftCOM. He is a member of IEEE and CCIS.