Identification and Position Estimation Method with K-Nearest Neighbour and Home Occupants Activity Pattern

Alfatta Rezqa Winnersyah
Department of Informatics, FST
UIN Syarif Hidayatullah
Jakarta, Indonesia
arwinnersyah@gmail.com

Feri Fahrianto, M. Sc
Department of Informatics, FST
UIN Syarif Hidayatullah
Jakarta, Indonesia
feri.fahrianto@uinjkt.ac.id

Nenny Anggraini, M.T
Department of Informatics, FST
UIN Syarif Hidayatullah
Jakarta, Indonesia
nenny.anggraini@uinjkt.ac.id

Abstract- As the number of home-based Internet of Things (IoT) applications such as home automation and monitoring of occupants behavior, indoor location information becomes a necessity. Previous research has been able to estimate indoor position using the fingerprinting method. In this paper, proposed a method that is able to estimate the indoor position of the occupants of the house by combining fingerprinting techniques with the home occupants activity pattern using K-Nearest Neighbour algorithm with Euclidean Distance. From the experimental results, this paper get results that the method that the proposed able to estimate with accuracy of up to 87.8% for accuracy below 2 meters with an average error of 0.82 meters.

Keyword—Indoor Positioning System, K-Nearest Neighbour, Activity Pattern

I. INTRODUCTION

Location is one of the important information in getting a variety of context-aware services. With many home based applications of Internet of Things (IoT) such as home automation and behavioral monitoring of occupants, indoor location information becomes a necessity [4]. The current popular location determination method is the Global Positioning System (GPS), but GPS is not able to work well in an indoor environment [23]. GPS signals are blocked by building construction and do not work well to get locations in space [22].

Indoor position estimation methods can be done in various ways, such as using WiFi, Bluetooth, Ultra-Wide Band (UWB), and RFID. WiFi is one pretty good approach. Can be quite good because WiFi technology has been implemented in many buildings even in smartphones have embedded WiFi technology. In addition WiFi has a wider signal coverage compared to Bluetooth and RFID [4].

WLANs can be used to determine a person's position in room by utilizing the value of Received Signal Strength Indication (RSSI) of multiple access points and with certain algorithms such as Support Vector Machines, Neural Networks, Weighted K Nearest Neighbors, or Bayesian approach. Conducted a learning of the system at a predetermined location point to obtain learning data to be used in the next process [11].

Machine learning deals with how to build computer programs that automatically get better with experience [9]. One approach in machine learning is classification. Classification is a job of valuing data objects to include them in a particular class of available classes. In the classification there are two main work done, namely (1) the development of the model as a prototype to be stored as memory and (2) the use of the model to perform the introduction / classification / prediction on another data object to be known in the class where the data object is in the model which he has saved [12].

Based on the way of training, the classification algorithms can be divided into two types: eager learner and lazy learner. The algorithms included in the lazy learner category do little training (or none at all), only saving part or all of the training data, then using them in the prediction process. This resulted in a long prediction process because the model had to re-read all the training data in order to give the correct class label output on the test data provided. The advantages of such algorithms are the fast-paced training process. The classification algorithms that fall into this category, among them, are K-Nearest Neighbor (K-NN), Fuzzy K-Nearest Neighbors (FK-NN), Linear Regression, and so on [12].

A common method used by calculating short-range or matching of data to be performed with existing learning data [4] [22] [23]. But from the closest distance is still often found the data made the results will be a wrong draw. In determining the position of the occupants of the house, one of the

indicators associated with the position is the activity of the occupants of the house. Activities of occupants can easily be known because home activities belong to a certain pattern. For example, every morning everyone likes to get up and brush their teeth or breakfast [24]. The way home occupants can be used as one of the parameters in determining the position of the occupants of a house at a certain time. This pattern of activity will be used as a comparison so that the results of the WiFi form produce more accurate data.

II. Related Works

Previous research [4] has proposed an AP-Centered Indoor Positioning System compared to Mobile HH based. Fingerprinting technique used to combine MH-Based with AP-Centered, that is during the training process using MH but when the online phase only use sniffy program in AP. This becomes more efficient because no application is installed in MH during the online phase. This study also produced a maximum error distance of 4.5 meters. But because it is centered in the AP then have to configure the AP in order to install the sniffy program created. And the estimation data is not given to the detected person.

Next research [11] analyzed the measured RSS value on two university room divided into 21 point. Measurement using laptop with Netsurveyor software. In this study compared the measurement values of 4 APs with people inside, 4 APs without people inside, 2 APs with people inside, and 2 APs with no people inside. The results of this study indicate that RSS values will change as environmental conditions change. The challenge of the system is the construction of buildings, furniture, and the human body.

Other research [22] took a case study on a room at the university. Using fingerprinting techniques, researchers perform techniques by sending the RSS value of the user's android and processed on the server. Researchers used 6 APs that had been installed at a height of 1.2 meters from the floor. The K-Nearest Neighbors Algorithm with Euclidean Distance was used in this study. The area is divided into 20 points . The results of this study show that implementation is easy and more cost-effective. From experimental data it is known that displacement in the room area affects RSS values and decreases accuracy. Accuracy is obtained 42.5% -64.5% within 2 meters if there are people in the room. While the estimation results are returned to the user and displayed.

Next research [23] measured RSS with smartphone for 3 data sets. The first data set as training data while the second data set is the test data with the situation is placed and the third data set is test data with handheld state. The area is one floor of the university building which is divided into 6 dots. This research uses Density-Based Clustering Combined Localization Algorithm technique in making fingerprint. High enough accuracy of 97.28% to 99.85% for the second data set and 97.28% to 99.85% for the third data set. This shows the position of the smartphone when measurement greatly affects the accuracy of the results.

From some of these research results, it is known that positioning estimation technology is currently highly developed. The most effective method used in previous research is to use the fingerprinting technique, which takes some data for training data and then collects the data and is matched with the training data to get the calculation result. Research [23] is capable of obtaining high accuracy but with considerable spacing between dots because one floor of the university building is only divided into six dots. While research [22] produces applications that can determine the position of its users with a system that works on the server and use android as a sender of RSS value. The result, the accuracy has reached 42.75% to 64.5%. This is influenced by the number of people in the room. While in the study [4], a proposed non-application method for users using OpenWRT access point was developed by developing a sniffy program to monitor Probe Request from nearby devices. The result, able to predict the position with a maximum error distance of 4.5 meters, but the calculation results are not given to the user directly.

III. Identification And Position Estimation

A. Fingerprinting Technique

The fingerprinting technique in position estimation is by recording the RSS value of multiple access points at a point. There are two main phases in fingerprinting technique, namely offline phase (training) and online positioning phase. The position of the reference point is determined by a distance of one or two meters. In the offline phase the RSS value is measured and stored into the database along with reference point information. Then in the positioning online phase, the RSS value is measured and compared to the RSS values available in the database to find the most suitable data.

B. K-Nearest Neighbour

The K-NN algorithm is an algorithm that classifies the proximity of the location (distance) of a data with other data. The value of K on K- NN means the number of nearest neighbors used as test data. One suitable distance calculation method used in the K-NN algorithm is the euclidean distance. Euclidean distance is a straight distance between two points contained in the Euclidean space. For distance between points p and q having n dimensions, the distance is calculated by the formula:

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad (1)$$

C. Home Occupants Activity Pattern

Activities of occupants can easily be known because home activities belong to a certain pattern. The fact is someone tends to do the same activity in the same environment. This show relation between environment and activities that a person undertakes in that environment. In estimating position of the occupants in the house, the activities of the occupants can be used as one indicator. From

a known pattern of activity, the pattern will be used as a comparison for the results of a WiFi-based method to produce more accurate data.

IV. Experimental Study

A. *Determine Reference Point*

In this paper, Inkscape software is used for designing home layout. The house that used has an area of 104.53 m2 with a length of 20.12 m and width of 6.42 m. Each reference point is located 1 m away. Access points are placed in 3 corners of the house which will be reachable by any reference point.

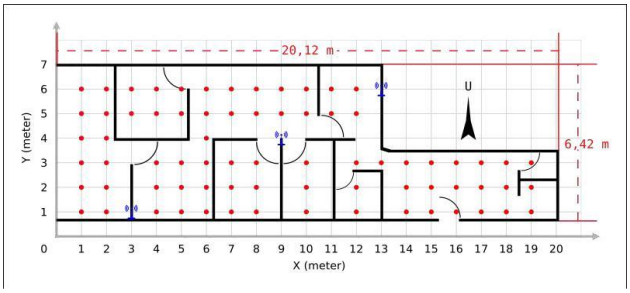
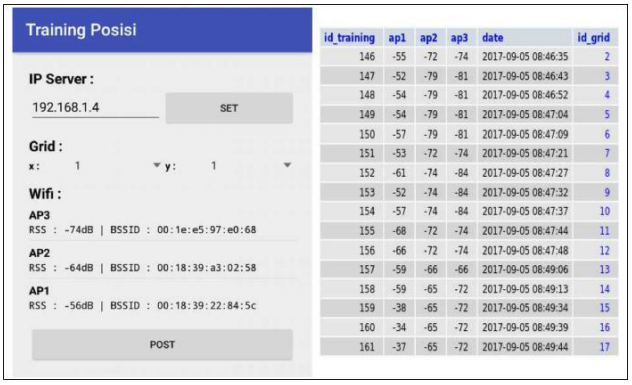


Fig. 1 Reference Points

B. *Training Phase*

Training data conducted in 5 days using the application with two times the data collection per day. Data is captured by capturing data every 5 seconds. Then the data is taken the average value. After that the data is sent to the web server to be stored into the database. From this training process obtained 720 data from 72 reference points that have been predetermined. Data is in the form of RSS value of 3 access points and reference point number.



id_training	ap1	ap2	ap3	date	id_grid
146	-55	-72	-74	2017-09-05 08:46:35	2
147	-52	-79	-81	2017-09-05 08:46:43	3
148	-54	-79	-81	2017-09-05 08:46:52	4
149	-54	-79	-81	2017-09-05 08:47:04	5
150	-57	-79	-81	2017-09-05 08:47:09	6
151	-53	-72	-74	2017-09-05 08:47:21	7
152	-61	-74	-84	2017-09-05 08:47:27	8
153	-52	-74	-84	2017-09-05 08:47:32	9
154	-57	-74	-84	2017-09-05 08:47:37	10
155	-68	-72	-74	2017-09-05 08:47:44	11
156	-66	-72	-74	2017-09-05 08:47:48	12
157	-59	-66	-66	2017-09-05 08:49:06	13
158	-59	-65	-72	2017-09-05 08:49:13	14
159	-38	-65	-72	2017-09-05 08:49:34	15
160	-34	-65	-72	2017-09-05 08:49:39	16
161	-37	-65	-72	2017-09-05 08:49:44	17

Fig. 2 Training Phase

C. *Home Occupants Activity Pattern Observation*

Observation is done by recording all the activities of occupants while in the house. The data recorded is when and where (based on the reference point) the activity is performed. Observation of occupant habits conducted in 7 days. From the results of this observation, occupants activities grouped into 4 time groups, morning (04:01-09:00), noon (09:01-15:00), evening (15:01-19:00), and night (19:01-04:00) to get the pattern in each time group. Once grouped, then calculated probability value in the existing time group and then store them in the database for use in the estimation process.

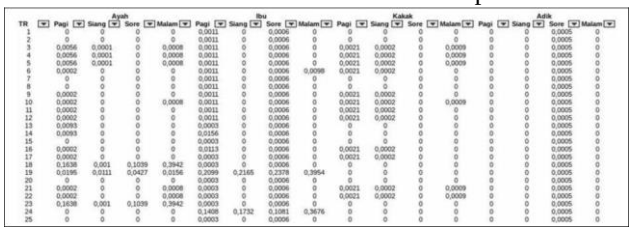


Fig. Activity Pattern Observation

D. *Estimation Phase*

In this paper, estimation process is centralized in server. Input data will be compared with training data an then compared with subject's activity pattern. Estimation phase conducted in 5 days. There are 18874 data that successfully estimated.

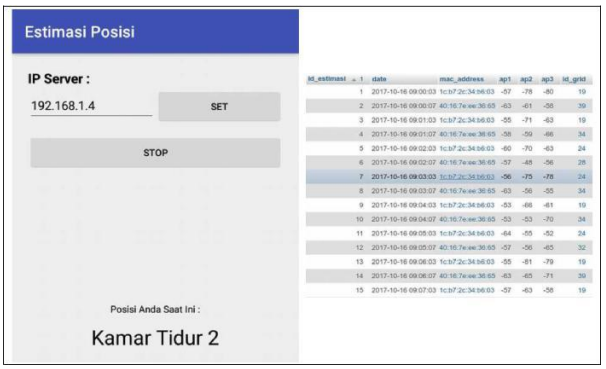
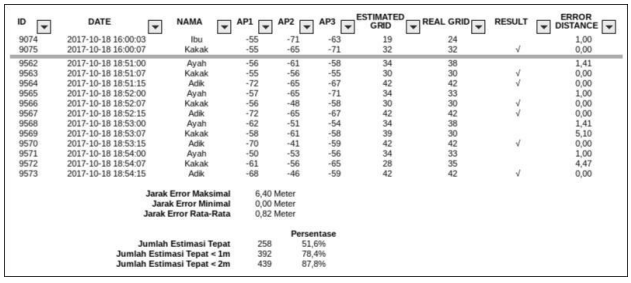


Fig. Estimation Phase

E. *Testing Estimation Result*

Estimation result tested with manual observation to 500 data samples that randomly chosen. Data samples compared with actual subject position.



ID	DATE	NAMA	AP1	AP2	AP3	ESTIMATED GRID	REAL GRID	RESULT	ERROR DISTANCE
9074	2017-10-18 16:00:03	Ibu	-55	-71	-63	19	24		1,00
9075	2017-10-18 16:00:07	Kakak	-55	-65	-71	32	32	√	0,00
9562	2017-10-18 18:51:00	Ayah	-56	-61	-58	34	38		1,41
9563	2017-10-18 18:51:07	Kakak	-55	-56	-55	30	30	√	0,00
9564	2017-10-18 18:51:15	Adik	-72	-65	-67	42	42	√	0,00
9565	2017-10-18 18:52:00	Ayah	-57	-65	-71	34	33		1,00
9566	2017-10-18 18:52:07	Kakak	-56	-48	-58	30	30	√	0,00
9567	2017-10-18 18:52:15	Adik	-72	-65	-67	42	42	√	0,00
9568	2017-10-18 18:53:00	Ayah	-62	-51	-54	34	38		1,41
9569	2017-10-18 18:53:07	Kakak	-58	-61	-58	39	30		5,10
9570	2017-10-18 18:53:15	Adik	-70	-41	-59	42	42	√	0,00
9571	2017-10-18 18:54:00	Ayah	-50	-53	-56	34	33		1,00
9572	2017-10-18 18:54:07	Kakak	-61	-56	-65	28	35		4,47
9573	2017-10-18 18:54:15	Adik	-68	-46	-59	42	42	√	0,00

Jarak Error Maksimal	6,40 Meter
Jarak Error Minimal	0,00 Meter
Jarak Error Rata-Rata	0,82 Meter

		Persentase
Jumlah Estimasi Tepat	258	51,6%
Jumlah Estimasi Tepat < 1m	392	78,4%
Jumlah Estimasi Tepat < 2m	439	87,8%

Fig. 5 Testing Estimation Result

V. Discussion

From the results already described earlier, it is known that this method is able to prove that the pattern of home occupancy activities can be one important indicator in the estimation of positions in the house which can not be reached by Global Positioning System (GPS) technology. The result of estimation accuracy of the method that the proposed is also quite good because it is able to estimate with an average error distance only 0.82 meters.

Many things affect the estimation result using this method, as this method uses fingerprinting technique which use wireless signal from access point, where wireless signal is known to be very fluctuating to the condition of the surrounding environment so that if there is a significant change of environmental condition then the estimation result will not be good. In addition, the pattern of activities of the occupants of the house sometimes not always the same, there are times when occupants suddenly perform activities that have not been done before.

VI. Conclusion

Based on the discussion that has been described previously, can be concluded that the method that the proposed has been able to estimate the position of the occupants of the house with an average error below 1 meter. This method was built starting with training data collection, activity pattern observation, and final estimation. Combining fingerprinting techniques with home occupant activity patterns can improve the estimation accuracy. The fingerprinting technique using K-Nearest Neighbour algorithm with Euclidean Distance as a classification algorithm.

Acknowledgment

This work was supported in part by the Institute for Research and Community Services (LP2M) UIN Syarif Hidayatullah Jakarta.

References

[1] Akkaya, K., Guvenc, I., Aygun, R., Pala, N., & Kadri, A. (2015, March). IoT-based occupancy monitoring techniques for energy-efficient smart buildings. In Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE (pp. 58-63). IEEE.

[2] Alpaydin, Ethem. Introduction to Machine Learning – 2nd ed. 2010. MIT Press.

[3] Bhartiya, Shekhar. 2016. 3 step roadmap for building machine learning systems, https://www.linkedin.com/pulse/3-step-road map-building-machine-learning-systems-shekhar-bhartiya, diakses pada 30 Desember 2016

[4] Du, X., Wu, J., Yang, K., & Wang, L. (2016, December). An AP-Centred Indoor Positioning System Combining Fingerprint Technique. In Global Communications Conference (GLOBECOM), 2016 IEEE (pp. 1-6). IEEE.

[5] Hasil Survei 2016 Asosiasi Penyelenggara Jaringan Internet Indonesia (APJII) https://www.apjii.or.id/survei2016/download/FBY9pDwIce3TO4HALX71bvqxfnZzRG diunduh 04 Maret 2017 08:29 WIB

[6] Kendall, K.E & Kendall, J.E. 2011. System Analysis and Design Eighth Edition. USA: Pearson Education, Inc.

[7] Laman Resmi PHP. PHP: What is PHP? http://php.net/manual/en/intro-whatis.php diakses 14 Juni 2017 11:24 WIB

[8] Laman Resmi PHP. PHP: What can PHP do? http://php.net/manual/en/intro-whatcando.php 14 Juni 2017 11:26 WIB

[9] Mitchell, Tom M. 1997. Machine Learning. USA: McGraw-Hill Companies.

[10] Nugroho, Adi. 2011. Perancangan dan Implementasi Sistem Basis Data. Yogyakarta : ANDI

[11] Pirzada, N., Nayan, M. Y., Hassan, M. F., Subhan, F., & Sakidin, H. (2016, August). WLAN location fingerprinting technique for device-free indoor localization system. In Computer and Information Sciences (ICCOINS), 2016 3rd International Conference on (pp. 650-655). IEEE.

[12] Prasetyo, Eko. 2012. Data Mining : Konsep dan Aplikasi menggunakan MATLAB. Yogyakarta : ANDI

[13] Press Release StatCounter 2017. Android overtakes Windows for first time. http://gs.statcounter.com/ press/android-overtakes-windows-for-first-time diakses 22 Mei 2017 22:31 WIB

[14] Raharjo, Budi. 2011. Belajar Otodidak Membuat Database Menggunakan MySQL. Bandung : Informatika

[15] Raharjo, Budi dkk. 2012. Modul Pemrograman WEB (HTML, PHP, & MySQL). Bandung : Modula

[16] Rizky, Soetam. 2011. Konsep Dasar Rekayasa Perangkat Lunak. Jakarta : PT. Prestasi Pustaka

[17] Rosa A.S., M. Shalahuddin. 2014. Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek. Informatika. Bandung

[18] Safaat, Nazruddin H. 2012. Android : Pemrograman Aplikasi Mobile Smartphone dan Tablet PC Berbasis Android. Bandung : Informatika

[19] Sangaji, Etta Mamang & Sopiah. 2010. Metodologi Penelitian – Pendekatan Praktis dalam Penelitian

[20] Suprianto, Dodit & Agustina, Rini. 2012. Pemrograman Aplikasi Android. Yogyakarta : MediaKom

[21] Sutanta, Edhy. 2011. Basis Data dalam Tinjauan Konseptual. Yogyakarta : ANDI

[22] Thuong, N. T., Phong, H. T., Do, D. T., Van Hieu, P., & Loc, D. T. (2016, January). Android application for WiFi based indoor position: System design and performance analysis. In Information Networking (ICOIN), 2016 International Conference on (pp. 416-419). IEEE.

[23] Toh, C., & Lau, S. L. (2016, October). Indoor localisation using existing WiFi infrastructure—A case study at a university building. In Virtual System & Multimedia (VSMM), 2016 22nd International Conference on (pp. 1-5). IEEE.

[24] Yan, Y., Ricci, E., Rostamzadeh, N., & Sebe, N. (2014, October). It's all about habits: Exploiting multi-task clustering for activities of daily living analysis. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 1071-1075). IEEE.