Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 37 (2014) 32 – 39

Procedia Computer Science

The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)

Fingerprint Time Length Reduction for Developing an Indoor Location Model for Smartphones

Carlos E. Galván-Tejada[a,*], Juan P. García-Vázquez[a], Jorge I. Galván-Tejada[b], Juan R. Delgado-Contreras[a], Ramon F. Brena[a]

[a]*Tecnológico de Monterrey, Monterrey, Nuevo León, México, Autonomous Agents in Ambient Intelligence.*
[b]*Tecnológico de Monterrey, Monterrey, Nuevo León, México Bioinformatics.Escuela de Medicina.*

Abstract

In our previous work, we proposed an approach that consists in using feature extraction to reduce the magnetic-field signal data to generate a *signature*, which can be used to estimate the user location in an indoor environment. Each signature contains 10 seconds of magnetic-field data. In this paper, we investigate whether by reducing the time signature length decrease or not the accuracy of the indoor location system (ILS). In order to find out, we perform experiments in two indoor environments: an office building and a residential home. In both environments, we collect information of the magnetic-field and we variate the fingerprint length in 0.1, 1, 2, 5 and 10 seconds, to verify whether its length affects the accuracy of the ILS Model. The results indicate that signature length is an important issue to be considered in the development of the ILS, since it does affect the accuracy of the system.

Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.
Keywords: Fingerprint, Magnetic-Field, Location, Indoor Positioning, Location Estimation, Feature Extraction,

1. Introduction

A common method used by the indoor location systems (ILS) is the *Fingerprinting* scheme. This commonly is composed of two phases[1]: *training* and *location determination*. First, a radio map of observed signal strength values from different locations is recorded during a training phase. Then, in the position determination phase, the signal strength values observed at a user device are compared to the radio map values using proximity matching algorithms, such as k-NN[2] and other classifiers[3,4].

In several research works, a reduction of the fingerprint size has been attempted to improve the performance of the system, and increase its capacity to storage signal information. For instance, *Kamaladas* et al.[5] uses a wavelet transform to extract the main features of audio files in order to increase the capacity of their song recognition system, while Manjunath et al.[6] use a Fast Fourier Transform (FFT) to extract features from an image to improve the response time of their system.

* Corresponding author. Tel.: +52-81-8358-2000
E-mail address: ericgalvan@uaz.edu.mx

1877-0509
Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.
doi:10.1016/j.procs.2014.08.009

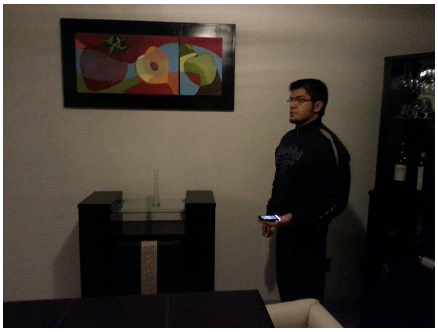


Fig. 1. Smartphone at the user's waist

In our previous work[7], we used a feature extraction process to reduce the magnetic-field signal data required to estimate the user location in an indoor environment, since we are considering that our ILS will be implemented in a mobile device with reduced computational capabilities.

In this paper, we discuss the hypothesis that reducing the time signature length of the magnetic-field fingerprint does not decrease the accuracy of the ILS, because we know that a huge amount of data can lead to misinterpretation of the information[8]. Additionally, the uniqueness of the magnetic field[9] could ensure that with small amounts of data we would be able to identify indoor locations. In order to demonstrate it, we performed experiments in two different indoor environments: an office building floor and a ground floor of a residential home. In those environments, we collected data from the natural Earth's magnetic-field using the magnetometer of a smartphone. We create a data set of the local magnetic field signal for each environment. The data was segmented to create different sets of length fingerprints (0.1, 1, 2, 5 and 10 seconds). Each set of fingerprints was used in our location estimation methodology in order to test the aforementioned hypothesis. The results indicate that fingerprint length is indeed an important issue that indoor location system designers must consider when using the fingerprinting method.

This paper is organized as follows: after this introduction, the methodology used for ILS is presented in section 2. The experimentation setup is presented in section 3. the experimental results are presented in section 4. Finally, our conclusions and future work are presented in section 5.

2. Location Estimation Methodology

Using a summary of our ILS methodology presented in Galván-Tejada et al.[7]. This consists of two phases described in the following sections.

2.1. Data Collection

To collect magnetic-field information from an indoor environment [1], a set composed of 1,000 data points from a magnetometer sensor of a smartphone is used to generate a *signature*. To get data points, the user must walk around the indoor environment with an approximate speed of 1 m/s during 10 seconds with the smartphone at the user's waist as is shown in figure 1. The number of seconds (10) to collect the signature was chosen because in average 10 seconds is enought time to cover an common indoor room location of 6 m^2 walking at the proposed speed of 1 m/s.[2]

In order to estimate the number of signatures needed to create a model, the equation 1 proposed by *Eberhardt*[10] was used to determine the minimal number of experiments in multivariable process with aim to have statistical validation. In the equation 1, x is the minimum number of experiments, and N is the number of variables.

$$x = log_2(N) + 1 \quad (1)$$

[1] Data sets are available in: http://aaami.mty.itesm.mx/?page_id=24

[2] In previous works we have shown that the exact speed is not critical with respect to precision.

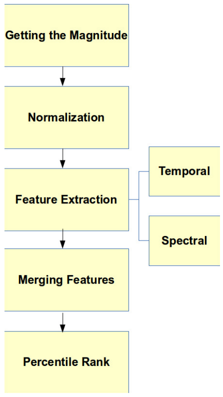


Fig. 2. Activities of Data Analysis

2.2. Data Analysis

This phase consists of five activities, as is shown in figure 2, which are described as follows.

1. *Getting the Magnitude*. The magnetic-field measures were modeled as a vector of three components Bx, By, and Bz[9]; we can compute the total magnitude of the field as described in Eq. (2), where Mx, My, and Mz are the three physical axes along x, y, and z respectively.

$$|M| = \sqrt{Mx^2 + My^2 + Mz^2} \tag{2}$$

2. *Signature Normalization*. After the magnetic field magnitude is obtained, we eliminate spatial scaling and shifting by normalizing each signature using Eq. (3), where $z_{i,d}$ is the normalized reading, $r_{i,d}$ refers to the i^{th} observation of the signature in dimension d; μ_d is the mean value of the signature for dimension d and σ_d is the standard deviation of the signature for dimension d.

$$\forall i \in m : z_{i,d} = \frac{r_{i,d} - \mu_d}{\sigma_d} \tag{3}$$

   Eq. 3 is applied for all dimensions in R^d
3. *Feature Extraction*: This process consists of the magnetic-field data reduction, in order to extract the main signal features sufficient to characterize the signal behavior. We extract features from two domains : *time* and *frequency*.

   (a) *Temporal Shape Features*: These features are computed from the signal waveform. From the temporal shape were extracted 16 features, as is shown in table 1 .
   (b) *Spectral Shape Features*: In order to extract spectral features, the spectral signal is acquired by performing a P-point Fast Fourier Transform to each signature[11], as shown in Eq. (4), where ES_i is the i^{th} energy signature of the normalized signal, and NS_i is the i^{th} normalized signature.

$$\forall i \in n : ES_i = FFT(NS_i) \tag{4}$$

4. *Merging Signal Features*. Once all the features are computed, all of them are merged into a data set of features that summarize the behavior of the signal reducing the amount of data from 1,000 data points to 46 per signature.
5. *Percentile Rank*. Once all the features were extracted and merged, a percentile rank was done to each feature to keep them in a range 0 to 1.

Table 1. Features Extracted

Features	Temporal Domain	Frequency Domain
Kurtosis	*	*
Mean	*	*
Median	*	*
Standard Deviation	*	*
Variance	*	*
Coefficient of Variation (CV)	*	*
Inverse CV	*	*
1,2,3 Quartile	*	*
1,5,95,99 Percentile	*	*
Trimmed Mean	*	*
Shannon Entropy		*
Slope		*
Spectral Flatness		*
Spectral Centroid		*
Skewness		*
1-10 Spectrum Components		*

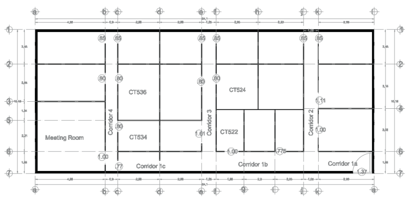


Fig. 3. Office Environment Layout

3. Experimental Setup

3.1. Testing Environment

Our experiments were carried out in two different indoor environments to demonstrate the appropriateness of the methodology for indoor location and the accuracy of the fingerprint lenght to estimate the user location.

- *Building Environment*. It was on a common office building, as shown in figure 3, it consists of 11 rooms or offices, which were considered in this experimentation given the spatial characteristics. We define the names of the 11 rooms as follow: CT542, CT536, CT534, CT522, CT524, Corridor 1a, 1b, 1c, Corridor 2, Corridor 3, and Corridor 4.
  The data collected in the office building, was collected with the integrated sensors of a smartphone (Samsung S4 i905), which includes a magnetometer (model: YAS532).
  The number of seconds required for this environment was calculated with the equation 1 presented in section 2.1, where x is the minimal number of experiment, and N is the number of variables. In this case N is equal to 506

Fig. 4. First floor house plans with furniture

Table 2. Fingerprint sets with different length

Length (in seconds)	Number of Fingerprints per room
0.1	1000
1	100
2	50
5	20
10	10

considering 46 features multiplied by 11 rooms; from the equation we get 9.98 rounded to 10 signatures. This mean 100 seconds of measurements from the magnetic field for each room considering 10 seconds as the longest fingerprint to be tested. This 10 seconds as longest fingerprint was chosen because all the rooms can be fully covered walking at 1m/s, as is proposed in the methodology.

- *Residential Home Environment*. It was a ground floor of a residential house, which consists of 4 rooms and stairs as it is shown in figure 4. The different rooms considered for experimentation are highlighted with a color: living room (blue), dining room (red), kitchen (green), and bathroom (pink).
  Data used in this work was collected from the sensors of a Smartphone device (Samsung S3 i9300) because it has the sensors that we need (magnetic sensor).
  In the residential home the number of signatures obtained with equation1 was 8.59 rounded to 9 signatures, but in order to keep the experiments under the same conditions, we collected 100 seconds of magnetic field measurements per room.

3.2. Fingerprint sets

These number of seconds of measurements from both indoor environments were divided in 5 different sets of fingerprints per room, with different lengths as is shown in the table 2.

Table 3. Confusion Matrix Using the 5 Seconds Set

	CT522	CT524	CT534	CT536	CT542	Corridor1a	Corridor1b	Corridor1c	Corridor2	Corridor3	Corridor4	Error
CT522	15	4	1	0	0	0	0	0	2	3	1	0.4230769
CT524	2	12	2	1	2	0	0	0	2	4	1	0.5384615
CT534	0	2	16	2	0	4	1	0	1	0	0	0.3846154
CT536	1	0	3	4	6	5	0	4	2	0	1	0.8461538
CT542	0	0	0	2	15	4	0	3	2	0	0	0.4230769
Corridor1a	0	0	1	7	2	12	2	2	0	0	0	0.5384615
Corridor1b	0	0	1	0	0	4	19	1	0	0	1	0.2692308
Corridor1c	0	0	0	2	5	0	1	18	0	0	0	0.3076923
Corridor2	1	1	3	3	3	1	1	0	13	0	0	0.5
Corridor3	1	3	1	0	0	0	1	0	0	20	0	0.2307692
Corridor4	1	0	0	1	0	2	1	0	2	0	19	0.2692308

Table 4. Confusion Matrix Using the 2 Seconds Set

	CT522	CT524	CT534	CT536	CT542	Corridor1a	Corridor1b	Corridor1c	Corridor2	Corridor3	Corridor4	Error
CT522	58	3	3	1	0	0	0	0	0	0	0	0.1076923
CT524	4	50	5	4	1	0	0	0	0	1	0	0.2307692
CT534	0	4	55	5	0	1	0	0	0	0	0	0.1538462
CT536	0	1	3	50	6	0	1	2	0	0	2	0.2307692
CT542	0	0	1	2	53	6	1	1	0	0	1	0.1846154
Corridor1a	0	0	0	0	7	50	4	4	0	0	0	0.2307692
Corridor1b	0	0	1	0	0	6	56	2	0	0	0	0.1384615
Corridor1c	0	1	0	0	4	1	2	51	4	1	1	0.2153846
Corridor2	0	2	0	0	0	3	2	3	52	3	0	0.2
Corridor3	1	0	1	0	0	0	0	0	5	56	2	0.1384615
Corridor4	0	1	0	0	0	0	0	0	1	6	57	0.1230769

3.3. Extraction process

To extract the features a script was programmed in *R Project for Statistical Computing* software. This script, which implements the process in figure 2, extracts all the features. Once all the features were extracted, a percentile rank was applied to each feature to keep them in a range 0 to 1. Once all the features are extracted and ranked, then a random forest composed of 5000 trees was trained in order to obtain a prediction model. The Random Forest (RF) algorithm was chosen because it is an ensemble supervised machine learning technique and is based on bagging and random feature selection[12]. Further, RF was chosen because it allows us to calculate the error during the model generation, instead of requiring splitting the data set into training and testing sets to estimate the error with a blind test, using the out-of-bag (OOB) error estimation which has been demonstrated to be unbiased and that avoids the overfitting problem.

4. Experimental Results

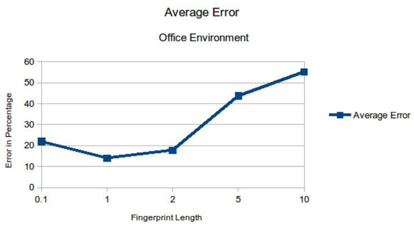
After experimentation, The obtained results in the both study cases are presented.

4.1. Building Environment Results

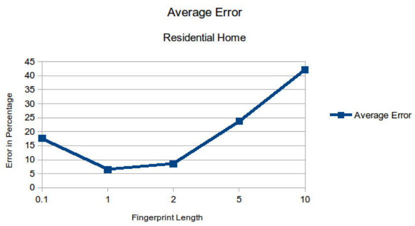
In the tables 3 and 4 the confusion matrix of the prediction model obtained with the random forest for the set of 5 and 2 seconds fingerprint length in the office environment are shown. We can see how the classification error rate goes down in the 2 seconds fingerprint in comparison with the 5 seconds fingerprint. This process was done for the 5 sets of fingerprints; table 5 shows the average classification error for each set. In figure 5a we can observe the variation of the error, and how the valley of the curve is located over the 1 second fingerprint length. From this observation we can see how 1 second contain enough information to estimate the location and less and more seconds lead to a misclassification given the lack or overwhelming amount of information.

Table 5. Average Error in the Office Environment

Fingerprint Length (in seconds)	Average Error (In percentage)
0.1	22.04
1	14.06
2	17.75
5	43.75
10	55.1



a)



b)

Fig. 5. Average error with the different fingerprint sets: a)Office Environment, b)Residential Environment

Table 6. Confusion Matrix Using the 5 Seconds Set in a Residential Home

	Bathroom	Kitchen	Dinning Room	Living Room	Error
Bathroom	16	3	1	0	0.2
Kitchen	6	13	1	0	0.35
Dinning Room	1	2	16	1	0.2
Living Room	1	2	1	16	0.2

Table 7. Confusion Matrix Using the 2 Seconds Set in a Residential Home

	Bathroom	Kitchen	Dinning Room	Living Room	Error
Bathroom	48	1	0	1	0.04
Kitchen	2	45	3	0	0.1
Dinning Room	0	2	45	3	0.1
Living Room	0	2	3	45	0.1

4.2. Residential Home Environment results

In the Residential home environment case, the same evaluation was carried on. The tables 6 and 7 show the confusion matrix of the prediction model obtained with the random forest for the set of 5 and 2 seconds fingerprint length for the 4 rooms in the residential home. In residential home environment we get the same behavior than in the office building environment, that it, a decreasing error when the fingerprint is shorter. The 5 sets of fingerprints were evaluated; table 8 shows the average classification error in the residential home environment presented in figure 5b, which follow the same behavior observed in the office environment with the lowest error in the 1 second fingerprint set.

Table 8. Average Error in the Residential Home Environment

Fingerprint Length (in seconds)	Average Error (In percentage)
0.1	17.6
1	6.5
2	8.5
5	23.75
10	42.1

5. Conclusions and Future Work

In this paper, we present a practical analysis of the fingerprint length needed to develop an indoor location system based on the magnetic-field signal.

Our results sugest that the uniqueness of the magnetic field indeed allows us to know the location of the user in real time using a fingerprint approach. Furthermore, that an ideal fingerprint length of magnetic field should be in the interval from one to two seconds, since in our experiments we identify that those fingerprints contain enough information to create a model to estimate the user location with high accuracy. However, out of these length range the accuracy of the ILS drops.

We conclude that fingerprints greater to 2 seconds, although they have more information, could lead to a misclassification problem given the redundant information, while a few data points increased the error rate too, given the lack of information needed to estimate the location.

As future work, we are considering to collect fingerprints from other sources, such as, wi-fi, environmental audio and indoor light; in order to test our location estimation methodology and verify whether it has the same behavior in relation to the length of the signature.

References

1. A. Taheri, A. Singh, A. Emmanuel, Location fingerprinting on infrastructure 802.11 wireless local area networks (wlans) using locus, in: Local Computer Networks, 2004. 29th Annual IEEE International Conference on, IEEE, 2004, pp. 676–683.
2. X. Liang, X. Gou, Y. Liu, Fingerprint-based location positoning using improved knn, in: Network Infrastructure and Digital Content (IC-NIDC), 2012 3rd IEEE International Conference on, IEEE, 2012, pp. 57–61.
3. M. S. Bargh, R. de Groote, Indoor localization based on response rate of bluetooth inquiries, in: Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments, MELT '08, ACM, New York, NY, USA, 2008, pp. 49–54.
4. J. Machaj, P. Brida, B. Tatarova, Impact of the number of access points in indoor fingerprinting localization, in: Radioelektronika (RADIOELEKTRONIKA), 2010 20th International Conference, 2010, pp. 1–4. doi:10.1109/RADIOELEK.2010.5478585.
5. M. D. Kamaladas, M. M. Dialin, Fingerprint extraction of audio signal using wavelet transform, in: Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on, IEEE, 2013, pp. 308–312.
6. B. S. Manjunath, W.-Y. Ma, Texture features for browsing and retrieval of image data, Pattern Analysis and Machine Intelligence, IEEE Transactions on 18 (8) (1996) 837–842.
7. C. E. Galván-Tejada, J. P. García-Vázquez, R. Brena, Magnetic-field feature extraction for indoor location estimation, in: Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction, Springer, 2013, pp. 9–16.
8. P. Indyk, R. Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, in: Proceedings of the thirtieth annual ACM symposium on Theory of computing, ACM, 1998, pp. 604–613.
9. W. Storms, J. Shockley, J. Raquet, Magnetic field navigation in an indoor environment, in: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010, 2010, pp. 1–10. doi:10.1109/UPINLBS.2010.5653681.
10. F. Eberhardt, C. Glymour, R. Scheines, N-1 experiments suffice to determine the causal relations among n variables, in: Innovations in machine learning, Springer, 2006, pp. 97–112.
11. W.-H. Tsai, Y.-M. Tu, C.-H. Ma, An FFT-based fast melody comparison method for query-by-singing/humming systems, Pattern Recognition Letters 33 (16) (2012) 2285–2291. doi:10.1016/j.patrec.2012.08.020.
URL `http://www.sciencedirect.com/science/article/pii/S016786551200284X`
12. V. Y. Kulkarni, P. K. Sinha, Pruning of random forest classifiers: A survey and future directions, in: Data Science & Engineering (ICDSE), 2012 International Conference on, IEEE, 2012, pp. 64–68.