

Towards an Indoor Location System using Indoor-light as Information Source

Carlos E. Galván-Tejada¹, Jorge I. Galván-Tejada¹, Selene Ramírez-Rosales¹,
F. J. Martínez-Ruiz¹, S. Paulizeth Maldonado-Luján²

¹ Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica,
Zacatecas, Zac., México

² Universidad Autónoma de Durango, Campus Zacatecas, Maestría en Diseño
Arquitectónico, Zacatecas, Zac., México

`ericgalvan@uaz.edu.mx`

Abstract. The location of individuals is a key element for commercial and assistive location-based applications. Since the Global Positioning System, the most effective technology for positioning a mobile object in the outdoors does not work in indoor environments. Several technological approaches have been proposed to tackle this problem. In this paper, we present an alternative approach based on the use of indoor lighting variations, feature extraction of the signal and a deterministic selection to get a model in order to estimate localization of an individual in indoor environments.

Keywords: indoor location, mobile systems, indoor luminosity, location estimation.

1 Introduction

Indoor location systems using natural signals (environmental signals) are becoming relevant in new approaches. Because it can be assumed that a natural signal must be present everywhere, avoiding the constraint of the coverage of non natural signals (i.e. WiFi, RFID) and the need for extra infrastructure deployment. However, natural signals generate a lot of data that must be processed in order to get information of them. This treatment implies more computing time and storage problems. To tackle these problems, Galván-Tejada et al. proposes a work developing ILS [10], exploring feature extraction and feature selection, in order to reduce the amount of data required to develop an ILS using natural signals, in our case, magnetic-field signal.

Nevertheless, from our experience, we know that using Genetic Algorithms during the selection phase, despite performance, can carry a high computational cost, given the stochastic nature of the approach. In this paper, we use the methodology proposed in [9] to extract features from the signal, and a feature selection thought a *Chi-Squared Filter* [15] to avoid stochasticity and finally be used to classify using well-known Random Forest Classifier [3]. Finally, we

propose using luminosity. Since this signal in controlled environments is more robust than other natural signals in terms of changes.

This paper is organized as follows. In section 2, we present a briefly review of Indoor Location Systems development approaches (ILS). The methodology to develop an ILS is explained in section 3. In section 4 we described our proposal. The results of experiments are presented in section 5, and finally, our conclusions and future work are presented in section 6.

2 Related work

Given the importance of the location context variable to develop ILS, several technological approaches have been proposed to develop ILS using different approaches. For instance, the use of technology that must be deployed as RFID, Wi-Fi, Bluetooth, Ultra Wide-band (UWB), Ultrasonic Sensors, among others [16,4]. Depending on the technique and the technology, these approaches allows us to estimate user location. Nevertheless, the use of signals that must be generated, as Wi-Fi or Bluetooth, present one main disadvantage. They require a dedicated infrastructure (e.g. access points, array of sensors or cameras) that must be deployed to develop an ILS. On the other hand, if infrastructure is already deployed. We have to consider other issues (limited Coverage, scalability, etc).

Therefore, new technological approaches have been proposed to tackle these issues. Approaches using the signals that are already available in the indoor environment [20,2,14,19,6]. For instance, the use of barometers to detect location is one of the options that emerges to know a "high-precision indoors location system". Barometers can be used it on smart phones in different kinds of buildings like hospitals, airports, etc. However, has several disadvantages. These ILS are very sensitive to "floor selection". they are just focused on plane location. The challenge is knowing, exactly in what floor of the building users are located. Besides that, we cannot know accurately location in the horizontal axe . To tackle this disadvantage, Xia et al [22] proposes the use of multi-reference barometer floor positioning (MBFP) as the answers under some conditions, also using a mobile device. This method proposes "comparing the altitude with the height of the floors in a building" to estimate in which floor is the user. Also, this method "is robust and less sensitive to factors such as temperature and humidity". The method MBFP is instable, the results that it have in the experimental face are different, it uses different sensors in the same altitude and environment. MBFP provides facilities in the new lifestyle using a sensor in the mobile device, but it needs a combination with another method to get better results.

Other available environmental signal is the light, which can be generated artificially with bulbs or naturally by the sun. It can be sensed in lumens using a sensor, which nowadays is included in almost all the smartphones in the market. For instance, Rahman et al [17] propose a Visible Light Communication (VCL) system using lighting LEDs and image sensors, several LEDs transmit the 3D coordinate information and it is demodulated by the image sensors. The 3D

array of LED helps to send location by two image sensors through two optical sensors. Also the positioning on the VCL has a limited area of response. This limitation introduces a positioning error with a few centimeters of difference.

These works use irregularities of available natural signals to estimate the location. However, huge amount of data must be analyzed to estimate the location, and it must be combined with other techniques to improve the accuracy. Galvín-Tejada et al [10] we propose a methodology to reduce the amount of data and use a single signal to develop an ILS using a natural signal (i.e. geomagnetic-field signal), but from our experience we know that feature selection, carried on through a Genetic Algorithm (GA) is computationally expensive.

The present work enhances Galvín-Tejada et al methodology through a deterministic method in the feature selection step, which is used to reduce the amount of information required. Finally, GAs are avoided since they are the most time consuming process of the methodology.

3 Methodology

In this section, we present the new version of the methodology presented in *Galvín-Tejada et al.* [9] and [10]. This methodology is modified in order to use light as information source. Beside that GAs are not used. This methodology consists of three steps described below.

3.1 First step: data collection

In data collection step, lumen data of an indoor environment is collected. To record this data, user must walk around the indoor environment carrying on a device running our mobile application, which is able to record lumen data through a lumen sensor. This walk must be done in an average walk speed during 10 seconds. The recording device must be positioned with the sensor up side, regardless any other position in the spatial axes as is shown in Figure 1. The data lectures are done at a 100 lectures per second frequency. We call *signatures* this basic set of data that represent a room.

The lumen/light sensor returns a float value indicating lumens in the environment at the moment when the lecture is done.

3.2 Second step: signature collection

The minimal number of lumen signatures is calculated using the equation 1 presented by Eberhardt [8], where x is the minimal number of needed signatures and N is the number of independent features that will compose the final lumen fingerprints.

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

To avoid the spatial scaling an Z norm normalization is applied to each light lecture signature using the equation 2, $z_{i,d}$ where $z_{i,d}$ is the normalized reading,



Fig. 1. Disposition of sensor in order to record light lectures.

$r_{i,d}$ 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 .

$$\forall i \in m : z_{i,d} = \frac{r_{i,d} - \mu_d}{\sigma_d} \quad (2)$$

Equation 2 is applied for all dimensions in R^d

Once light signatures are collected in the rooms and stored in a data set created during the Signature Collection. Feature extraction and selection are carried on to get the classification model that allows us to estimate the location of users.

3.3 Third step: feature extraction

The feature extraction is a process that consists in carrying out an efficient data reduction while preserving the appropriate amount of signal information. In this sense, the feature extraction step allows us to reduce the amount of data to develop the indoor location estimation model. To apply this step we must select the features that will be extracted from the signal, these activities are described below.

Signal features Following Galvn-Tejada et al methodology and given the fact that behavior of lumen signal(under the assmption of a stochastic behavior). We select 46 features collected from literature review in digital signal processing

(DSP). Besides first and second order statistic parameters [13,1,5,11] shown in Table 1, where 16 features are from temporal domain and 30 ones are from spectral domain. *Temporal features* are computed from the waveform of lumen signal generated from the signature previously recorded, and *Spectral features* extracted from the signal after performing a P-point Fast Fourier Transform to each signature [21]; to achieve it, we apply Equation (3), where ES_i is the i^{th} energy signature of the normalized lumen's signal, and NS_i is the i^{th} normalized signature.

We keep the first 10 components from the spectral evolution, because from observation (we presume) that the light phenomena tends to be chaotic too.

$$\forall i \in n : ES_i = FFT(NS_i) \quad (3)$$

Table 1. Extracted features

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		*

Once all the features are computed, all of them are merged into a data set of features that summarizes the behavior of the signal, reducing the amount of data from 1000 data points to 46 per signature.

To keep all the features in a range between 0 to 1 and make them have the same impact in the model development, a percentile rank is applied using Equation (4), where x is the x^{th} feature.

$$PR = \frac{trunc(rank(x))}{length(x)} \quad (4)$$

Model development: feature selection For getting a model with less features that allows us to estimate user location. A feature selection process is applied. Feature selection use multiply techniques. For instance, filter, wrapper and embedded methods [18]. From literature [7] we propose apply a *filter method*, because it returns features that can be used with different classifiers and is less computationally expensive.

Filter methods that allow us to get a feature selection at low computational cost includes: *Chi-Squared Filter* [15], because is capable of evaluate the data in function of weights; features with a low weight are removed based on the chi squared statistics [12]. Once the features are selected, these can be used with several classifiers to get a model to estimate the location, which is the core of an ILS.

4 Experimentation

The experimentation was carried on in a residential home shown in Figure2, using available data from AAAMI research group website ³. We select four rooms because are contiguous. This characteristic involved a extra complication. The spaces share light data in some parts of rooms. The light is provided by 7 bulbs composed by 107 LED's warm light.

In order to determine the minimal number of signatures required, the equation 1 was applied. We consider 46 features per each room, which means 184 variables, getting a total of 8.5 lumen signatures, finally rounded to 9.

Once we got the minimal number of signatures. We choose a Samsung Galaxy S3 i900 to record lumen data, using an Android application, previously developed in order to read lumen sensor.

To perform feature extraction step, we develop a script in R Project for Statistical Computing software ⁴, a free multi-platform software (GNU project) environment for statistical computing, Getting a matrix composed of 36 rows (one for each signature) and 46 columns (one for each feature). Percentile rank is applied to all the values to keep them between 0 to 1.

After feature matrix normalized. The filter method (Chi-Square Filter) proposed is used to select features. This step select 6 features from the original 46 extracted from the signatures.

The Chi-Square Filter is applied using the CARET package [12] in R, which does the work in five phases:

1. Data equalization and normalization.
2. Use the filter to obtain significative features.
3. Obtain classification rate using the filtered model with Random Forest.
4. Obtain the classification rate using 46 selected features with Random Forest.
5. Calculate the Confusion Matrix for comparison.

³ <http://aaami.mty.itesm.mx/>

⁴ <http://www-r-project.org/>

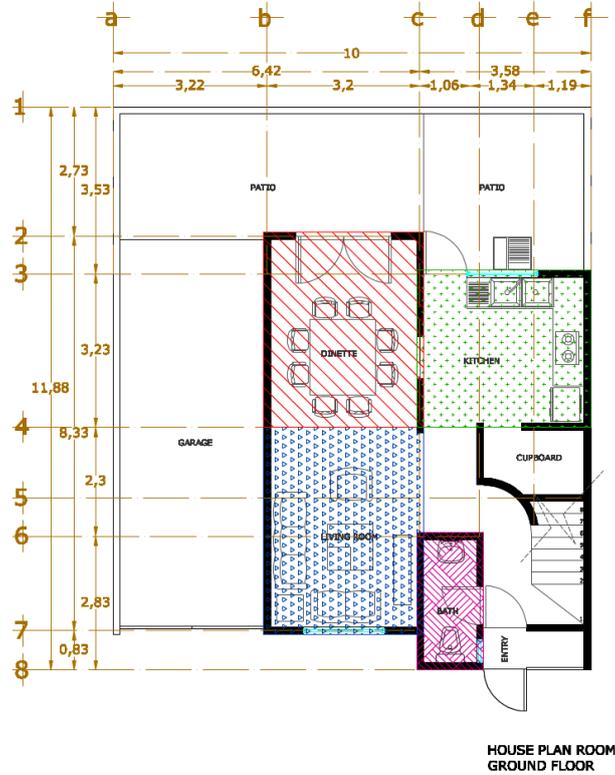


Fig. 2. Plane of residence used in our experiment. bathroom in pink, living room in blue, dining room in red and kitchen in green.

With selected features, a Random Forest classifier composed of 5,000 trees is used to evaluate indoor location model conformed by features selected by filter method. This classifier was proposed by Breiman et al. [3]; and selected because it provides tree ensembles that depends on the values of a random feature vector and provides the same distribution to all the trees included in the forest. The decision of this classifier depends on the decision of several trees.

All classification process was performed using Random Forest R package.

5 Results

After experimental test, the final filtered model was a 17 feature model of classification. Table 3 shows the confusion matrix acquired from the classification model of the random forest using 17 features model from data set. The obtained model can classify 95.83 percent of the samples, correctly. In table 2 filtered features model is presented.

Table 2. Filtered Features

Features	Temporal Domain	Frequency Domain
Mean		*
Median	*	
Standard Deviation	*	
Variance		*
Coefficient of Variation (CV)	*	*
Inverse CV	*	*
1,2,3 Quartile	*	
1,5,95,99 Percentile	*	
Slope		*
6 Spectrum Component		*

Table 3. Signatures Classification For Different Locations Using filtered features

	Bathroom	Dining Room	Kitchen	Living Room
Bathroom	6	0	0	0
Dining Room	0	6	0	0
Kitchen	1	0	5	0
Living Room	0	0	0	6

The forty six features were tested in order to produce a classification model. Table 4 shows the confusion matrix acquired from classification model of the random forest using 46 features model. The obtained model can classify correctly 91.66 percent of the samples.

Table 4. Signatures Classification For Different Locations

	Bathroom	Dining Room	Kitchen	Living Room
Bathroom	4	0	2	0
Dining Room	0	6	0	0
Kitchen	0	0	6	0
Living Room	0	0	0	6

The obtained features are used to generate a random forest classification model which is compared against the 46 features random forest model. As shown in Table 5.

Table 5. Comparison of Different Approaches

Approach	Classification Rate (in percentage)
Filtered Model	95.83
Unfiltered features	91.66

6 Conclusions and future work

This paper presents a novel approach where light-based models for geo spatial positioning are used for the first time. Instead of using stochastic techniques. Like in the first version of the methodology. A classification model based on spatial characteristics of light emission and the behavior of the same in different environments within a residence space was introduced.

The following important aspects regarding the classifiers were identified :

- *Sensitivity and Specificity*: Since the model found by using Chi Square filter is relatively large, their sensitivity and specificity are high, as can be seen in its confusion matrix , resulting in a very high percentage of correct classification of spaces that were used for this study. It is clear that reducing variables by filtering not only improves speed of classifier , it helps increase the percentage of prediction. Because some statistically rivalry between variables. Therefore, causing a lower model performance.
- *Computational cost* compared to the use of genetic algorithms, the calculation time is much smaller. In this version, it was reached an acceptable performance through fewer variables and more concise models. In our tests the GA version has required 144 minutes (with 2 features) while this version required 2 minutes and 5 seconds (with 17 features) .The improvement is considerable.
- *Time is the most important issue to take into account*. Along the evaluation performed in this work, we also show that time is the variable with higher variability. Time became the most important characteristic to take into account when developing an ILS that requires recalibration.

For future work the possibility of sampling a larger space is proposed. Taking into account different types of lighting in each scene and purpose of their interiors. Also, using other tools to improve the performance of mobile phones and tablets .

References

1. Agostini, G., Longari, M., Pollastri, E.: Musical instrument timbres classification with spectral features. EURASIP Journal on Applied Signal Processing 2003, 5–14 (2003)

2. Bilke, A., Sieck, J.: Using the magnetic field for indoor localisation on a mobile phone. In: Krisp, J.M. (ed.) *Progress in Location-Based Services*, pp. 195–208. *Lecture Notes in Geoinformation and Cartography*, Springer Berlin Heidelberg (2013)
3. Breiman, L.: Random forests. *Machine learning* 45(1), 5–32 (2001)
4. Chen, L., Pei, L., Kuusniemi, H., Chen, Y., Krger, T., Chen, R.: Bayesian fusion for indoor positioning using bluetooth fingerprints. *Wireless Personal Communications* 70(4), 1735–1745 (2013), <http://dx.doi.org/10.1007/s11277-012-0777-1>
5. Chen, P.C., Pavlidis, T.: Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. *Computer graphics and image processing* 10(2), 172–182 (1979)
6. Chung, J., Donahoe, M., Schmandt, C., Kim, I.J., Razavai, P., Wiseman, M.: Indoor location sensing using geo-magnetism. In: *Proceedings of the 9th international conference on Mobile systems, applications, and services*. pp. 141–154. *MobiSys '11*, ACM, New York, NY, USA (2011)
7. Delgado-Contreras, J.R., García-Vázquez, J.P., Brena, R.F., Galván-Tejada, C.E., Galván-Tejada, J.I.: Feature selection for place classification through environmental sounds. *Procedia Computer Science* 37, 40–47 (2014)
8. Eberhardt, F.: A sufficient condition for pooling data. *Synthese* 163(3), 433–442 (2008), <http://dx.doi.org/10.1007/s11229-007-9293-3>
9. Galván-Tejada, C.E., García-Vázquez, J.P., Brena, R.: Magnetic-Field Feature Extraction for Indoor Location Estimation. In: *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*, pp. 9–16. Springer (2013)
10. Galván-Tejada, C.E., García-Vázquez, J.P., Brena, R.F.: Magnetic-field feature reduction for indoor location estimation applying multivariate models. In: *MICAI (Special Sessions)*. pp. 128–132. IEEE (2013)
11. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on* (6), 610–621 (1973)
12. from Jed Wing, M.K.C., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., the R Core Team: caret: Classification and Regression Training (2014), <http://CRAN.R-project.org/package=caret>, r package version 6.0-24
13. Lambrou, T., Kudumakis, P., Speller, R., Sandler, M., Linney, A.: Classification of audio signals using statistical features on time and wavelet transform domains. In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. vol. 6, pp. 3621–3624 (1998)
14. Li, B., Gallagher, T., Rizo, C., Dempster, A.G.: Using geomagnetic field for indoor positioning. *Journal of Applied Geodesy* 7(4), 299–308 (2013)
15. Liu, H., Setiono, R.: Chi2: Feature selection and discretization of numeric attributes. In: *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*. pp. 388–388. IEEE Computer Society (1995)
16. Ni, L., Liu, Y., Lau, Y., Patil, A.: Landmarc: Indoor location sensing using active rfid. *Wireless Networks* 10(6), 701–710 (2004)
17. Rahman, M.S., Haque, M.M., Kim, K.D.: Indoor positioning by led visible light communication and image sensors. *International Journal of Electrical and Computer Engineering (IJECE)* 1(2), 161–170 (2011)
18. Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. *bioinformatics* 23(19), 2507–2517 (2007)

19. Storms, W., Shockley, J., Raquet, J.: Magnetic field navigation in an indoor environment. In: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010. pp. 1–10 (oct 2010)
20. Subbu, K.P., Gozick, B., Dantu, R.: Locateme: Magnetic-fields-based indoor localization using smartphones. *ACM Trans. Intell. Syst. Technol.* 4(4), 73:1–73:27 (Oct 2013), <http://doi.acm.org/10.1145/2508037.2508054>
21. Tsai, W.H., Tu, Y.M., Ma, C.H.: An FFT-based fast melody comparison method for query-by-singing/humming systems. *Pattern Recognition Letters* 33(16), 2285–2291 (2012), <http://www.sciencedirect.com/science/article/pii/S016786551200284X>
22. Xia, H., Wang, X., Qiao, Y., Jian, J., Chang, Y.: Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning. *Sensors* 15(4), 7857–7877 (2015)