Wifigrams: Design of Hierarchical Wi-Fi Indoor Localization Systems Guided by Social Network Analysis

Jose M. Alonso[1], Noelia Hernández[2], and Manuel Ocaña[2]

[1] European Centre for Soft Computing, Mieres, Asturias (Spain)
jose.alonso@softcomputing.es
[2] Department of Electronics, University of Alcalá, Madrid (Spain)
{noelia.hernandez,mocana}@depeca.uah.es

Abstract. This work introduces the notion of *Wifigrams*, i.e., visual Wi-Fi maps which comprise hundreds of topological positions (nodes) linked in accordance with their co-visibility degree of Wi-Fi Access Points (APs). Both design and analysis of *Wifigrams* are guided by the most popular Social Network Analysis (SNA) techniques. Thus, the structure of a hierarchical topology-based Wi-Fi indoor localization system emerges naturally when looking carefully to the related *Wifigram*. It shows graphically the closeness among topological positions in terms of their pattern of AP co-visiblity, yielding several zones of interest which are automatically discovered through community mining techniques. This novel proposal has been successfully tested in a large real-world environment, namely the first floor of the west wing in the Polytechnic School at the University of Alcalá (UAH), yielding encouraging results.

Keywords: Wi-Fi Technology, Topology-based Hierarchical Localization, Social Network Analysis.

1 Introduction

Effective coordination and collaboration between humans and robots is becoming essential for more and more applications (for instance deployment of rescue teams in emergency situations, proactive care for the elders, context-aware services, navigation assistance, and so on) [12, 15]. Moreover, wireless communications and localization systems are deemed as core elements for such applications.

This work is made in the context of the ABSYNTHE[1] project which is aimed at handling "Abstraction, Synthesis, and Integration of Information for Human-Robot Teams" [1]. We have recently developed a hierarchical topology-based Wi-Fi indoor fingerprint localization system supported by the ability of an expert to manually dividing the environment into sub-regions [7]. Unfortunately, dividing the environment by hand is not always straightforward. Moreover, it becomes a really hard task when dealing with real-world large and complex environments.

[1] Official website: http://absynthe.softcomputing.es

R. Moreno-Díaz et al. (Eds.): EUROCAST 2013, Part II, LNCS 8112, pp. 9–16, 2013.

This work proposes a new localization system which carries out topology-based indoor localization in a high abstraction level in order to be valid and useful for both humans and robots. The localization task is addressed as a high dimensional hierarchical classification task supported by social network analysis (SNA) techniques. To do so, we first introduce the novel *Wifigrams* by inspiration on the scientograms previously developed by Vargas-Quesada and Moya-Anegon [14] as novel tools for visualizing the structure of science. In a *Wifigram*, pairs of topological locations (to be identified by the localization system) are linked in a social network regarding their co-visibility of Access Points (APs). Then, the experimental environment is automatically split into sub-regions of interest that are made up of those locations more strongly connected in the network. The analysis of *Wifigrams* through SNA techniques, mainly community mining algorithms, allows us identifying groups of locations strongly related.

The rest of the paper is structured as follows. Section 2 explains how to generate and analyze *Wifigrams*. Moreover, it describes how to use the valuable information provided by *Wifigrams* with the aim of guiding the design of a hierarchical topology-based Wi-Fi indoor localization system. Section 3 discusses the reported experimental results. Finally, some conclusions and future works are sketched in Section 4.

2 Proposal

First of all, we build a Wi-Fi fingerprint database (M) which is made up of Received Signal Strength (RSS) data samples from N locations:

$$M = \begin{pmatrix} RSS_{AP_1}(P_1,1) & \dots & RSS_{AP_Z}(P_1,1) \\ RSS_{AP_1}(P_1,2) & \dots & RSS_{AP_Z}(P_1,2) \\ \dots & \dots & \dots \\ RSS_{AP_1}(P_1,T) & \dots & RSS_{AP_Z}(P_1,T) \\ RSS_{AP_1}(P_N,1) & \dots & RSS_{AP_Z}(P_N,1) \\ \dots & \dots & \dots \\ RSS_{AP_1}(P_N,T) & \dots & RSS_{AP_Z}(P_N,T) \end{pmatrix} \tag{1}$$

where Z is the number of APs and T samples were collected in each reference location. $RSS_{AP_i}(P_k,t)$ is the RSS data sample t taken at location P_k from AP_i.

Second, we compute the degree of visibility ($VIS_{AP_i}(P_k)$) of each single AP_i from each location P_k, according to M:

$$VIS_{AP_i}(P_k) = \frac{1}{T}\sum_{t=1}^{T} d_{ik}(t), \quad d_{ik}(t) = \begin{cases} 1 \text{ , } RSS_{AP_i}(P_k,t) > RSS_{th} \\ 0 \text{ , } otherwise \end{cases} \tag{2}$$

It is computed as the percentage of samples that were collected with $RSS_{AP_i}(P_k,t)$ greater than a predefined threshold RSS_{th}[2].

[2] This threshold was set to -200 dBm in the experiments.

Third, we generate the visibility matrix, M_V $(N \times Z)$, by counting the number of APs that are visible from each single location:

$$M_V = \begin{pmatrix} VIS_{AP_1}(P_1) & \dots & VIS_{AP_Z}(P_1) \\ \dots & \dots & \dots \\ VIS_{AP_1}(P_N) & \dots & VIS_{AP_Z}(P_N) \end{pmatrix} \quad (3)$$

$$VIS(P_k) = \sum_{z=1}^{Z} d_z(P_k), \quad d_z(P_k) = \begin{cases} 1 \text{ , } VIS_{AP_z}(P_k) > VIS_{th} \\ 0 \text{ , } otherwise \end{cases} \quad (4)$$

Notice that a given AP is deemed as visible if and only if its visibility degree is greater than a predefined threshold VIS_{th}[3].

Then, we generate the co-visibility matrix M_{co-VIS} $(N \times N)$:

$$M_{co-VIS} = \begin{pmatrix} 0 & m_{12} & \dots & m_{1N} \\ m_{21} & 0 & \dots & m_{2N} \\ \dots & \dots & \dots & \dots \\ m_{N1} & m_{N2} & \dots & 0 \end{pmatrix}, \quad m_{ij} = \begin{cases} \frac{co-VIS(P_i,P_j)}{\sqrt{VIS(P_i) \cdot VIS(P_j)}} & , i \neq j \\ 0 & , i = j \end{cases} \quad (5)$$

$$co - VIS(P_i, P_j) = \sum_{z=1}^{Z} d_z(P_i, P_j) \quad (6)$$

by counting the number of APs that are visible at the same time from each pair of given locations P_i and P_j:

$$d_z(P_i, P_j) = \begin{cases} 1 \text{ , } d_z(P_i) = 1 \;\; \& \;\; d_z(P_j) = 1 \\ 0 \text{ , } otherwise \end{cases} \quad (7)$$

Notice that matrix M_{co-VIS} contains the edge weights of a *Wifigram* which comprises N topological locations (nodes) linked in accordance with their co-visibility degree of Wi-Fi APs. Both nodes and edges can be displayed in an aesthetically pleasing graph thanks to the so-called force-based drawing algorithms which are commonly used in the context of SNA [11]. We have selected the Kamada-Kawai [9] algorithm. It starts from a circular position of nodes and searches for an optimal layout through maximizing the use of available space, minimizing the number of crossed links, and so on. As result, it assigns 2D coordinates to the nodes yielding a graph with the most important elements placed toward the center of the resultant image.

In addition, it is worthy to note that the initial *Wifigram* is usually quite dense and difficult to analyze even for a reduced number of locations. Fortunately, we can resort to very efficient and effective network scaling methods that were developed in the context of SNA. Among the existent methods, we opt for Pathfinder [6] which prunes the graph while keeping the most important links. Pathfinder removes every path that connect two nodes that violate the triangle inequality, having an associated distance greater than any other path between the same two nodes composed of up to Q links.

[3] This threshold was set to zero in the experiments.

Once the *Wifigram* is scaled then it can be analyzed by community mining algorithms [17]. They are aimed at finding out groups of nodes that are strongly connected thus constituting a community, i.e., inner connections among group members are stronger than outer connections with members of other groups. We have chosen the Fast Modularity (FM) algorithm [16]. It is able to point out communities exhibiting more edge density than other parts of the graph. Each community represents a zone of interest which is made up of a set of strongly related topological locations to be identified by the localization system.

Localization is actually carried out in two hierarchical classification levels. In the first level, the root classifier is in charge of distinguishing among the zones of interest. Then, in the second level, another classifier is in charge of pointing out the closest reference location inside the related zone that was previously identified. Notice that, this way we perform a topology-based indoor localization. It is a human friendly symbolic localization, which is made in a high abstraction level, in order to be valid and useful for both humans and robots.

3 Experiments and Results

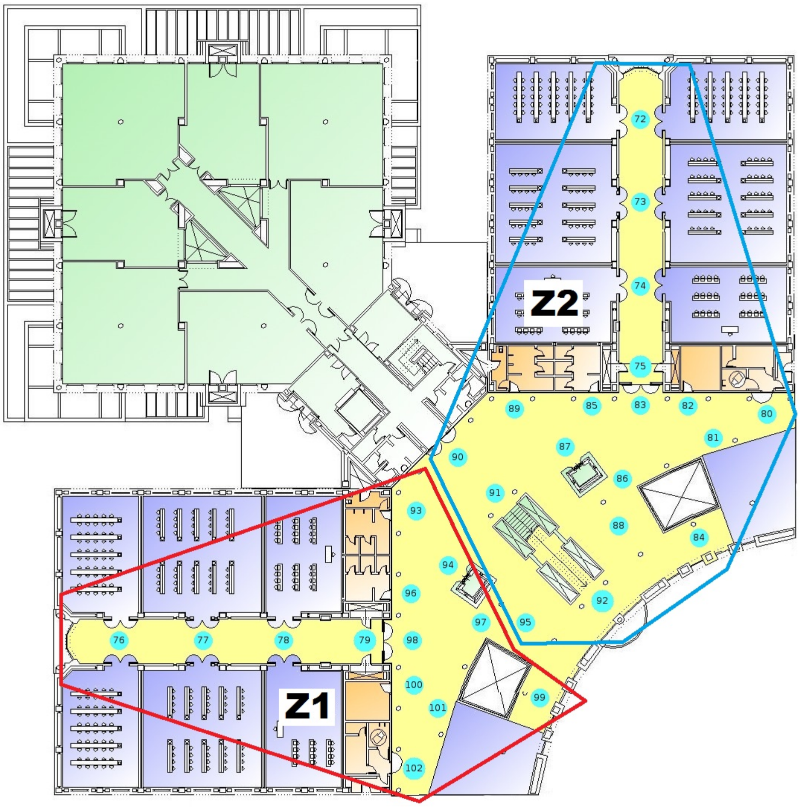
We have tested our proposal with experiments that were carried out in the first floor (west sector) of the Polytechnic School at the University of Alcala (UAH). The test-bed environment is depicted in Fig. 1. It comprises 31 topological locations which are represented by circled numbers. Each reference location is placed several meters (between 2.5m and 9.5m) apart from the nearest neighbor location to be recognized.

Data acquisition for building a fingerprint database was made in two independent weeks. In fact, collecting data for all the locations was made in discontinuous periods of time, at different hours, all along each week considering both mornings and afternoons. At each reference location 60 RSS data samples from each visible AP were measured and saved in a file. Thus, we built two datasets (one per week). In a first experiment, the dataset corresponding to the first week was taken as training set, while the dataset related to the second week was considered as test set. Then, we tried the other way around, i.e., the dataset related to the second week was taken as training set while the dataset corresponding to the first week was considered as test set. Of course, averaged results were reported.

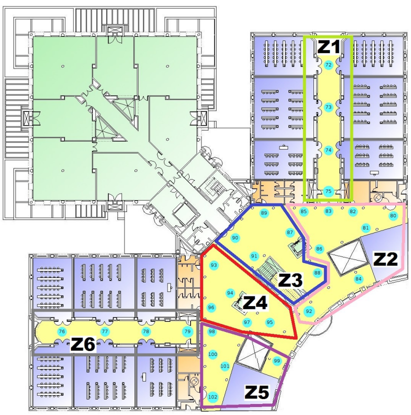
Then, just for comparison purpose, we applied the well-known K-means clustering technique [13]. It provided a simple division of the environment in two main zones (Fig. 1(a)).

Then, we applied the *Wifigram* analysis, previously presented in Section 2, with the aim of discovering the most salient zones. To do so, we took the implementation of the Fast Modularity (FM) community mining algorithm [16] provided by Meerkat[4] [3], a free software tool for community mining. It offers two variants of FM, namely “Weighted FM” and “Un-weighted FM”.

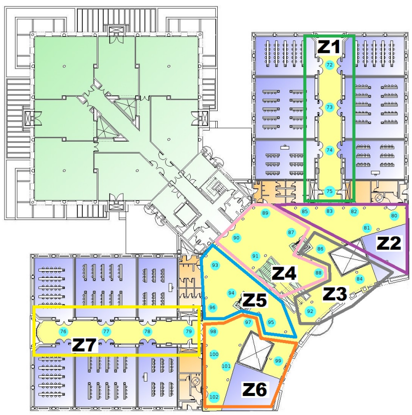
[4] Official website: `http://www.aicml.ca/node/41`



(a) K-means clustering.

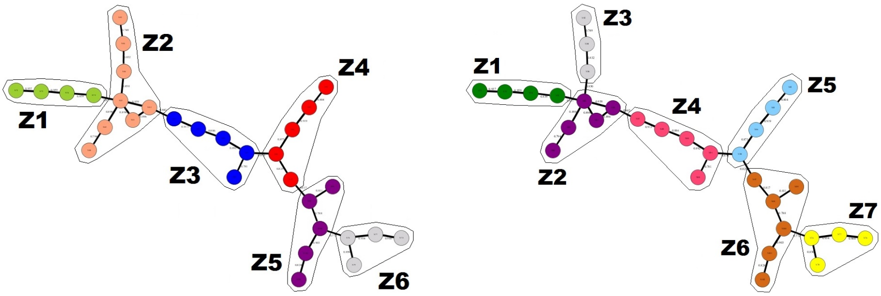


(b) Weighted FM community mining.



(c) Un-weighted FM community mining.

Fig. 1. UAH test-bed environment. Automatically discovered zones of interest (Zi).



(a) Weighted FM community mining. (b) Un-weighted FM community mining.

Fig. 2. Wifigram analysis

Figs. 2(a) and 2(b) show the generated *Wifigrams* along with the discovered zones remarked with different colors. Such zones are also highlighted in Figs. 1(b) and 1(c). Notice that, they have different sizes (since they comprise different number of reference locations) and geometric shapes. Even though the number of zones is different (six zones in case of Weighted FM versus seven zones when considering Un-weighted FM), the discovered zones of interest become consistent and quite similar. In fact, the zones related to corridors are exactly the same. Of course, there are slight differences when looking at the open area where it is not so easy to set an arbitrary optimal division. Such task is extremely hard for a human expert. Therefore, automatic division is required.

Once the environment was divided into a consistent set of zones of interest, it was time to face the generation of the related classifiers in charge of the hierarchical localization task. We considered three options:

– K-Nearest Neighbor (K-NN) [10]. It is an instance-based classifier which is defined as a variation of the well-known Nearest Neighbor algorithm for dealing with noise, where the most popular class of the K nearest examples is used for prediction. Notice that K-NN is usually considered as base line for comparisons when evaluating topology-based Wi-Fi fingerprint localization systems since the Bahl's pioneer paper [2] where a Wi-Fi localization system based on the use of the Nearest Neighbor algorithm with fingerprinting was first proposed.
– Fuzzy Unordered Rule Induction Algorithm (FURIA) [8]. It is a fuzzy rule induction classifier which extends the well-known RIPPER algorithm [4]. FURIA learns fuzzy rules instead of conventional rules like the ones produced by RIPPER, and unordered rule sets instead of rule lists.
– Support Vector Machines (SVM) [5]. It is a kernel-based classifier which constructs a hyperplane or set of hyperplanes in a high-dimensional space which separates input classes.

Table 1. Summary of Localization Results

Group	Classifier	Accuracy
None	FURIA	44.99%
	SVM	64.87%
	K-NN	58.04%
Un-weighted FM	FURIA	49.02%
	SVM	72.71%
	K-NN	63.65%
Weighted FM	FURIA	52.25%
	SVM	73.62%
	K-NN	63.51%
K-means	FURIA	49.7%
	SVM	70.29%
	K-NN	61.82%

Table 1 summarizes the reported localization results. The first column gives the method applied for dividing the experimental environment into zones of interest. "None" means that division was not considered, i.e., only one classifier was in charge of identifying the closest reference location. The second column presents the classification technique that was considered. The last column shows the percentage of locations properly identified. It is worthy to note the high accuracy (above 70%) reported by the proposed hierarchical localization approach. The highest accuracy (73.62%) was reported by Weighted FM + SVM.

4 Conclusions

On the light of the reported results, *Wifigrams* emerge as a powerful visual tool for supporting the design of topology-based Wi-Fi fingerprint indoor localization systems. Even more, we achieve results so promising that they encourage us to explore the applicability of other advance community mining techniques but also additional SNA techniques. Thus, in the near future we plan checking, for instance, the use of centrality measures with the aim of identifying the most salient nodes in *Wifigrams*. In addition, we will further explore the quest for optimal values for the considered parameters *RSSth* and *VISth*.

Acknowledgment. This work has been funded by the "Spanish Ministry of Economy and Competitiveness" through TIN2011-29824-C02-01 and TIN2011-29824-C02-02 (ABSYNTHE project). It has also been partly supported by the European Centre for Soft Computing.

References

1. Alonso, J.M., LeBlanc, K., Ocaña, M., Ruspini, E.: ABSYNTHE: Abstraction, synthesis, and integration of information for human-robot teams. In: International Workshop on Perception in Robotics, IEEE Intelligent Vehicles Symposium, pp. P14.1–P14.6 (2012)
2. Bahl, P., Padmanabhan, V.: RADAR: An in-building RF-based user location and tracking system. In: Proceedings of the IEEE Computer and Communications Societies, pp. 775–784 (2000)
3. Chen, J., Fagnan, J., Goebel, R., Rabany, R., Sangi, F., Takaffoli, M., Verbeek, E., Zaiane, O.: Meerkat: community mining with dynamic social networks. In: IEEE International Conference on Data Mining Workshops, pp. 1377–1380 (2010)
4. Cohen, W.W.: Fast effective rule induction. In: Proceedings of the International Conference on Machine Learning, pp. 115–123. Morgan Kaufmann (1995)
5. Cortes, C., Vapnik, V.: Support-vector networks. In: Machine Learning, pp. 273–297 (1995)
6. Dearholt, D., Schvaneveldt, R.: Properties of pathfinder networks. In: Schvaneveldt, R. (ed.) Pathfinder Associative Networks: Studies in Knowledge Organization, pp. 1–30. Ablex Publishing Corporation (1990)
7. Hernández, N., Alonso, J.M., Magro, M., Ocaña, M.: Hierarchical WiFi localization system. In: International Workshop on Perception in Robotics, IEEE Intelligent Vehicles Symposium, pp. P21.1–P21.6 (2012)
8. Hühn, C., Hüllermeier, E.: FURIA: an algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery 19, 293–319 (2009)
9. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. Inf. Processing Letters 31, 7–15 (1989)
10. Kibler, D., Aha, D.: Learning representative exemplars of concepts: an initial case study. In: Proceedings of the International Workshop on Machine Learning, pp. 24–30 (1987)
11. Kobourov, S.G.: Force-directed drawing algorithms. In: Tamassia, R. (ed.) Handbook of Graph Drawing and Visualization, ch. 12. CRC Press (2012)
12. López, E., Barea, R., Bergasa, L.M., Escudero, M.S.: A human-robot cooperative learning system for easy installation of assistant robots in new working environments. Journal of Intelligent and Robotic System 40(3), 233–265 (2004)
13. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297 (1967)
14. Moya-Anegón, F., et al.: A new technique for building maps of large scientific domains based on the cocitation of classes and categories. Scientometrics 61, 129–145 (2004)
15. Mousannif, H., Khalil, I.: Cooperative data management in wireless sensor networks. International Journal of Computational Intelligence Systems 5, 403–412 (2012)
16. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Physical Review E 74 (2006)
17. Porter, M.A., Onnela, J., Mucha, P.J.: Communities in networks. Notices of the AMS 56(9), 1082–1166 (2009)