

# Environment Classification Using RSS for Indoor Localization

Maximilien Dufau<sup>1</sup>, Elizabeth Colin<sup>1</sup> and Vighnesh Gharat<sup>2</sup>

<sup>1</sup> Efrei Paris, 30-32 avenue de la République, Villejuif, 94800, France

<sup>2</sup> Ela Innovation, 297, rue Maurice Bédart, Montpellier, 34080, France

## Abstract

Indoor localization using radio frequency is challenging because of multipath propagation and complex phenomenon that make standard positioning techniques, such as multilateration, inaccurate most of the time. However, knowing information about the environment can help overcome these issues. In this paper, we use Machine Learning to classify the environment between a Hall and a Corridor in order to use a more precise propagation model for Received Signal Strength (RSS) based localization. The classified samples are series of RSS values taken from RFID tags. Our method achieves 98.57% accuracy. The novelty of this approach is that it uses only a small series of RSS and thus is lighter than the existing techniques. To achieve this result, we use quadruplets of tags, we compare the best quadruplets for classification that we found with the best quadruplets for localization.

## Keywords

Environment classification, RSS, UHF-RFID, hall localization, corridor localization, tag positioning

## 1. Introduction

Nowadays, Indoor positioning is achieved through many techniques that all face different specific issues [1]. Techniques based on Machine Learning, especially Fingerprinting, needs a tedious pre-deployment phase and are not resilient to even slight modifications of the environment [2]. The complex behavior of electromagnetic wave in indoor environment makes Received Signal Strength (RSS) based techniques less accurate than competing ones based on time or angles. However, as this behavior depends on the type of environment, knowing information about the current environment can improve the accuracy of the positioning. Moreover, this information can be used to support context-aware applications.

Environment classification in an indoor localization scenario is traditionally achieved using Radio Frequency (RF) signature [3-5]. Channel Transfer Function (CTF) is used to extract the different multipaths that compose the signal, other features like the Frequency Coherence Function, the complex autocorrelation of CTF, are also often used as features for Machine Learning. Different algorithms such as k-nearest neighbor or convolutional neural networks are then used to classify the environment. These systems achieve classification accuracies of over 99%. However, they are more complex and require more data than using only RSS measures. In this form, they are thus not applicable in many localizations' scenarios.

In this paper, we present a new classification method based on only a small series of RSS measures from RFID tags that can classify the environment between a hall, and a corridor. Previous papers have shown that using only RSS gives poor classification results [5]. However, we introduce the use of four tags, also called quadruplets, to significantly improve the classification accuracy. Three different classification algorithms are considered, k-nearest neighbors, classification tree and bagged tree, so that we can compare their performance for different model parameters and analyze the tags' quadruplets behavior for each of them.

---

IPIN 2021 WiP Proceedings, November 29 -- December 2, 2021, Lloret de Mar, Spain

EMAIL: maximilien.dufau@efrei.net (A. 1); elizabeth.colin@efrei.fr (A. 2); vighnesh.gharat@elainnovation.com (A. 3)

ORCID: 0000-0001-5796-7974 (A. 1); 0000-0002-5992-8124 (A. 2)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

In the end, our solution is easier to apply than the previous ones in a practical scenario. It can be added to existing RSS based location systems with little effort to improve its accuracy as it allows using a specific propagation model for each environment.

We also discuss the best positioning of tags for classification purposes and compare it with the best positioning for localization that was found in [6-8].

This paper is organized as follows, section 2 presents the classified environments and the measures. The classification methods with a single tag and quadruplets are presented in section 3. In section 4 we discuss the best tags' quadruplets characteristics, and we compare them with the best quadruplets for localization. We conclude in section 5 with a discussion of the presented results and future work.

## 2. Measures & context

### 2.1. Environments

Experiments are performed in a hall of approximately 205 square meters, and in a corridor of dimensions 30 m x 2 m, which can be seen in Fig. 1. Tags are alternatively placed at a 1.3 meter and a 2.1 meters height. Both environments are meshed with sixty active RFID tags.

We needed to verify that the measurements taken in the narrowest part of the hall were sufficiently different from the ones in the corridor to be labeled as hall measurement. Indeed the dimensions of these parts were similar. We used k-means and hierarchical clustering to do so, with the conclusion that the natural labelling of the data was indeed the best one.

### 2.2. RFID equipment

For this work we used active UHF-RFID tags at 433 MHz band. The tags can be detected from as far as 20 meters in an indoor environment. A bidirectional traffic detection RFID reader, having two  $\frac{1}{4}$  wave antennas separated by 10.0 cm, is considered for the experiment. RS232 connector is used to connect the reader to the computer. The tag model "Coin ID" and reader model "UTP Diff 2", used for this work are from RFID manufacturer Ela Innovation [9]. The reader provides an 8-bits RSSI (RSS Information) corresponding to the received power as shown in (1). The higher the RSSI, the lower power.

$$P_{dBm} = 30.84 - 0.632 * RSSI \quad (1)$$

Where  $P_{dBm}$  is the received power in dBm.  $RSSI$  is the received signal strength in decimal.

### 2.3. Measurements

The measurements that we will be considering here are similar to the ones presented in [6-8]. The RFID reader is used to read RSS coming from sixty active tags. At a given position, 2x60 RSS acquisitions are realized, one for each tag for the two antennas. Multiple series of measurements are

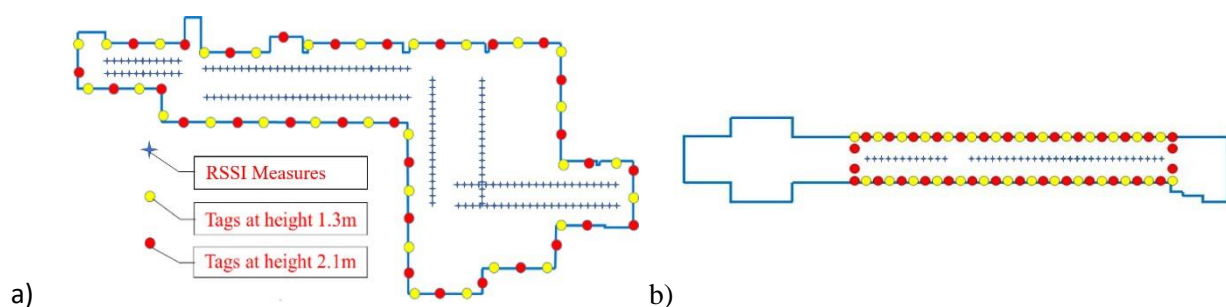


Figure 1: a) Tags' placement in the Hall, b) Tags' placement in the Corridor

taken, separated by 30 cm each and there has been no passing in both environments during the measures' collection. The measures were done at reference positions as can be seen on Fig.1.

## 2.4. Scenario and data sample

Our baseline scenario is the one of a moving agent only able to record Signal Strength at regular intervals, e.g. a robot or a pedestrian carrying a smartphone and walking in an indoor environment. The relative distance between successive RSS measures needs to be the same for each sample, in order for them to always describe the same type of information.

In a practical use scenario this can be achieved using dead reckoning or with the help of an accelerometer for example. However, these kinds of techniques are only suitable for a few steps and quickly become inaccurate due to sensor drift [10]. Thus, we must consider that we cannot guarantee an accurate relative position for measurements taken over large distances.

Moreover, reducing the amount of data for each sample implies less calculation and energy consumption. Considering that our solution aim to be used for real time localization on embedded systems these parameters are important to consider.

Therefore, for optimization and accuracy reasons, our classification method needs to use the less possible data. Moreover, each samples' measures are related to a single tag.

The samples we use for Machine Learning are therefore series of 2x6 RSS measures taken along a line, as shown in Fig. 2. Distance between antennas is 10 cm, six measures separated by 30.0 cm approximately correspond to at least two measures per wavelength at 433 MHz, which gives a good compromise between a reduced data set and the amount of information. We used a total of 1200 samples distributed as follow, 70% for the training set and 30% for the test set. We used oversampling to balance the quantity of samples between the hall and the corridor.

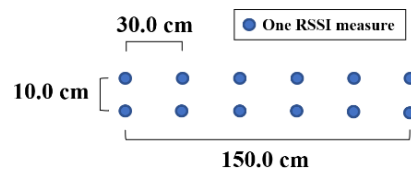


Figure 2: A typical sample data used for Machine Learning

## 3. Environment Classification

### 3.1. Features and results of single tag classification

To classify the two considered environments (hall and corridor), we established a set of features that describe our samples and discriminate them efficiently. We extracted mean, max, min, root mean square and other classical statistics functions of our measures set. We also extracted two subsets of data: the difference between the two antennas at a given position and the difference between two pairs of adjacent measures. The same statistics functions are applied for these two subsets. This allows the algorithms to take into account the sudden RSS variations that are characteristic of the multipath and of the fading of this kind of environments. A total of 14 features were extracted for each data sample.

Three algorithms' results are compared in Table 1. The best classification accuracy is achieved with the Classification Tree algorithm (82% of prediction accuracy). However, this system only considers the data coming from one tag, we improve it in subsection 3.2 by considering multiple tags for classification.

One important information to notice is that we do not use tags' ID information to perform the classification, even though it could contain useful clues about the environment. As said in section 2.2, our tags have a 20 meters range, which means a single tag's signal can be received in many different rooms. Moreover, using ID information would require an additional pre-deployment step. Thus, we decided not to take this information into account in this first iteration of our solution. However, when deployed as a practical solution, ID information could indeed be useful to perform sanity check on the model output.

**Table 1**

Prediction accuracy for three methods

Methods	k-NN	Classification Tree	Bagged Tree
Accuracy	75.32 %	82.28 %	79.75 %

### 3.2. Using quadruplets of tags to improve performance

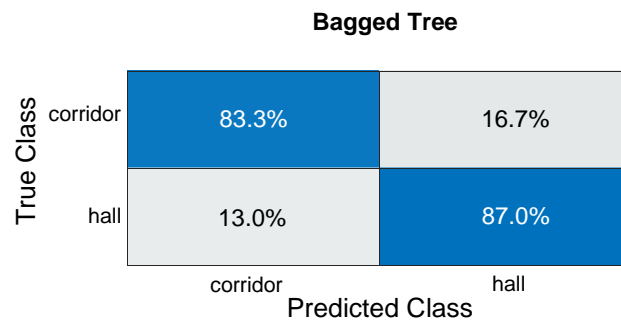
As our classification method will be used in a localization scenario using multilateration with RSS, we can assume that at least four tags will always be available for this purpose. Thus, we built a classification method that considers the output of our previous algorithms for quadruplets. We choose the new classification output by majority voting. With 60 tags there are 487635 possible quadruplets, the accuracy of our method mainly depends on the choice of goods quadruplets. Therefore, different selection criteria were considered to assess the quadruplet efficiency.

As shown in Table 2, choosing randomly from available tags does not significantly improve the classification accuracy, in some cases it even degrades it compared to single tag classification (82% for the best single tag classification). On the one hand, choosing the four tags that have the lowest mean RSS gives good results (higher than 85%) and on the contrary selecting tags with the highest mean considerably weakens the model (less than 70%). Indeed, high RSS value from a tag tends to indicate that the tag is close and therefore that the dominant path is generally the line-of-sight path, which is the same in all environments, contrary to the multipaths that depend on the environment geometry, building materials, etc. Therefore, as a high RSS value generally implies that the line-of-sight path is dominant in the signal, it means that the considered sample does not contain a lot of environment-related information, which weakens the classification accuracy. On the other hand, signal from tags that are too far are strongly attenuated and the signal strength is around the noise floor level, making it not suitable for classification. Selecting the tags that are closest to the 35<sup>th</sup>-percentile allows to find a good trade-off between strength of the signal and a sufficient distance between the sensor and the quadruplets tags. It gives very satisfying results, from 91.43% with Bagged Trees algorithm to 98.57% with k-NN algorithm.

**Table 2**

Prediction accuracy for different quadruplets selection methods

	k-NN	Classification Tree	Bagged Tree
Random	73.25 %	77.33 %	83.22 %
Maximum mean	35.71 %	45.71 %	68.57 %
Minimum mean	92.86 %	88.57 %	85.71 %
35 <sup>th</sup> -percentile mean	98.57 %	94.29 %	91.43 %

**Figure 3:** Confusion matrix with minimum mean selected quadruplets

## 4. Describing best classification quadruplets

In this paper we meshed both environments with sixty tags. However, this high density is not desirable in a practical usage scenario due to the cost of the tags as well as interference issues that reduce the tags' signal range. Therefore, we need to identify the optimal placement of tags, i.e. the tags' quadruplets that give the best classification results, to reduce their number. This is the subject of this section.

### 4.1. Best quadruplets for classification

We take two criteria to rank quadruplets. Firstly their accuracy ( $acc$ ), i.e. the ratio of correct classifications. Secondly their visibility ( $vis$ ), i.e. the number of samples they are able to classify within the test set. The visibility notion is important because a high accuracy on very few samples is not statistically significant. For example, a quadruplet that has an accuracy of 100% on only two of the test set samples will be ranked lower than a quadruplet that has an accuracy of 95% on thirty test set samples. We thus define the following score to rank quadruplets :

$$rankScore = (acc - C) * \log(vis) * e^{acc-1} \quad (2)$$

Where  $acc$  is the accuracy,  $vis$  the visibility and  $C$  a threshold used to set below 0 the rankScore of quadruplets that have a very low accuracy. It was set to 0.4 for this article's analysis so that the scores of quadruplets that achieve less than 40% classification accuracy are negative.

#### 4.1.1. Best quadruplets in the corridor

The three best quadruplets for classification in the corridor environment for each mentioned algorithm appear on Fig. 4. They are very similar. They all share:

- A triangle shape, moreover, the angles that are adjacent to the side with three tags are always both acute.
- Two very close tags.

One noticeable point is that the corridor is symmetric whereas the best quadruplet layouts are not. This could be explained by an electromagnetic disturbance that is not caused by the corridor geometry, e.g. a particularly important presence of ferro-magnetic materials in a wall.

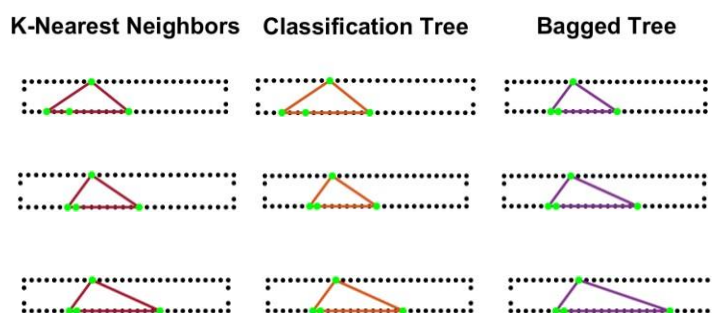


Figure 4: Best quadruplets in the corridor environment

#### 4.1.2. Best quadruplets in the hall

Best Quadruplets in the hall environment appear on Fig. 5. The best quadruplets are not strictly identical for each algorithm; however, they share common characteristics which are:

- Omniscience, i.e., at least one tag has a direct line-of-sight for the whole hall.
- Alignment, i.e. three tags always approximately share a common straight line. Just like in the corridor, hall quadruplets have a triangular shape and two tags at least are close to each other.

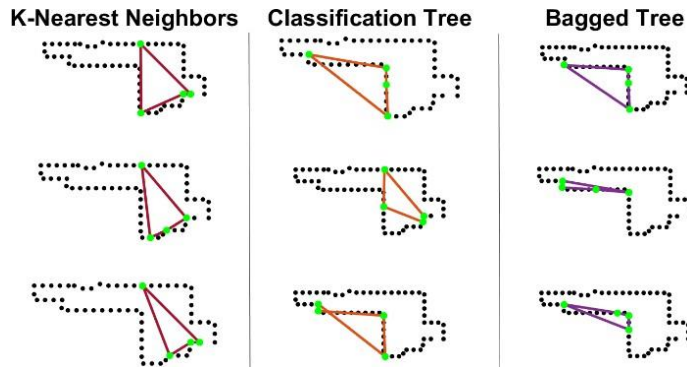


Figure 5: Best quadruplets in the hall environment

Moreover, there is no significant correlation between the length of the quadruplets' maximal diagonal and their rankScore, however best tags, i.e. the top 20% of rankScore values, tend to be clustered closely with a span of around 14.56 m long, as shown in Fig. 6.

#### 4.2. Best quadruplets for classification vs best quadruplets for localization

In [7-9] best tags' quadruplets for localization in both environments were analyzed. We compared their characteristics with the ones we identified above. A priori, good positioning quadruplets must differ from good classification quadruplets because they need opposite signal characteristics. The first ones require the less possible multipaths whereas the second ones need these multipaths to identify the environment.

According to [8], quadruplets positioned on the extremity of the corridor achieve better positioning accuracy. We observe an opposite behavior with quadruplets for classification in the corridor. Indeed, they are centered with respect to the corridor. On the other hand, [8] points out the importance of having two close tags out of the four quadruplet's tags, which is similar to our results. According to [7], the closer the quadruplets' tags are from the sample, the better the positioning accuracy is. However, we found that for classification the tags need to be not too close from the sample position.

Best quadruplets for positioning in the hall described in [6] share the alignment characteristic we identified for classification's quadruplets. However, the average span of the positioning quadruplets is much greater than the classification ones.

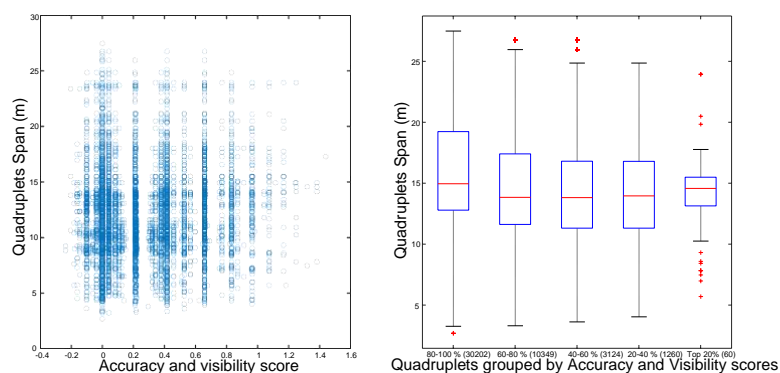


Figure 6: Quadruplets Span against rankScore



## 5. Conclusion and future work

We built an indoor environment classification method that can recognize a hall and a corridor with 98.57 % accuracy. As a comparison the method presented in [2] classifies four indoor environments with an accuracy of 97.21 % and a similar size training set. Our method thus achieves an equivalent classification accuracy to that of the previous methods, but it requires less measures and is thus simpler to use in a practical real-time localization scenario. However, additional data from other similar environments are needed to test and improve the scalability of our solution to other halls and corridors. Other environments will also need to be added to make the system complete.

We identified best classification tags' quadruplets characteristics for both environments. Good tags' quadruplets for classification correspond to an extend to the good tags for localization which confirms our solution can be use in pair with a traditional localization method, such as received RSS, with little additional deployment effort.

Our future work will be to expand the classification method to other environments and to test its performances with other type of signals such as WiFi as it is more widespread than RFID and thus requires less deployment effort. It will also be necessary to test our method in a localization scenario to confirm its value.

## 6. References

- [1] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, thirdquarter 2019, doi: 10.1109/COMST.2019.2911558.
- [2] Antoni Pérez-Navarro, Joaquín Torres-Sospedra, Raul Montoliu, Jordi Conesa, Rafael Berkvens, Giuseppe Caso, Constantinos Costa, Nicola Dorigatti, Noelia Hernández, Stefan Knauth, Elena Simona Lohan, Juraj Machaj, Adriano Moreira, Pawel Wilkv, "Challenges of Fingerprinting in Indoor Positioning and Navigation, 2019, Intelligent Data-Centric Systems, Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation, Academic Press, ISBN 9780128131893, doi : 10.1016/B978-0-12-813189-3.00001-0.
- [3] G. Zhu, F. Dong and N. Pang, "Classification of Indoor Environments Based on Mixed Graph Similarity using UWB Signals," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 197-201, doi: 10.1109/SSCI47803.2020.9308294.
- [4] Z. Chen, M. I. AlHajri, M. Wu, N. T. Ali and R. M. Shubair, "A Novel Real-Time Deep Learning Approach for Indoor Localization Based on RF Environment Identification," in *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1-4, June 2020, Art no. 7002504, doi: 10.1109/LESENS.2020.2991145.
- [5] M. I. AlHajri, N. T. Ali and R. M. Shubair, "Classification of Indoor Environments for IoT Applications: A Machine Learning Approach," in *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 12, pp. 2164-2168, Dec. 2018, doi: 10.1109/LAWP.2018.2869548.
- [6] V. Gharat, E. Colin, "Active RFID Tags Placement Configuration for Accurate Positioning in a Hall Environment" in *Proceedings of IPIN 2015*, 2015.
- [7] E. Colin, A. Moretto and M. Hayoz, "Improving indoor localization within corridors by UHF active tags placement analysis," 2014 IEEE RFID Technology and Applications Conference (RFID-TA), 2014, pp. 181-186, doi: 10.1109/RFID-TA.2014.6934224.
- [8] V. Oruganti, V. Gharat, E. Colin and A. Moretto, "Location performance law according to the dimensions of the corridor using trilateration," 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2014, pp. 511-517, doi: 10.1109/IPIN.2014.7275523.
- [9] Ela Innovation active RFID tag and reader manufacturer. <https://www.elainnovation.com>. Accessed May 2021
- [10] L. Ojeda and J. Borenstein, "Personal dead-reckoning system for gpsdenied environments," in *Safety, Security and Rescue Robotics, 2007. SSRR 2007. IEEE International Workshop on*. IEEE, 2007, pp. 1-6.