

RFID indoor localization based on support vector regression and k-means

Everton Luís Berz
PUCRS University
Email: everton.berz@acad.pucrs.br

Deivid Antunes Tesch
PUCRS University
Email: deivid.tesch@acad.pucrs.br

Fabiano Passuelo Hessel
PUCRS University
Email: fabiano.hessel@pucrs.br

Abstract—Systems need to know the physical locations of objects and people to optimize user experience and solve logistical and security issues. Also, there is a growing demand for applications that need to locate individual assets for industrial automation. This work proposes an indoor positioning system (IPS) able to estimate the item-level location of stationary objects using off-the-shelf equipment. By using RFID technology, a machine learning model based on support vector regression (SVR) is proposed. A multi-frequency technique is developed in order to overcome off-the-shelf equipment constraints. A k-means approach is also applied to improve accuracy. We have implemented our system and evaluated it using real experiments. The localization error is between 17 and 31 cm in 2.25 m² area coverage.

I. INTRODUCTION

Localization of objects and people in indoor environments has been widely studied due to security issues and because of the benefits that an indoor positioning system (IPS) can provide. To determine location in outdoor environments, GPS technology (global positioning system) has grown increasingly popular, becoming the de facto standard in the location and navigation of vehicles, people and other objects. However, this technology is not suitable for indoor environments because it requires a direct line-of-sight communication to satellites. Indoor environments are more complex than outdoor environments because of the high density of obstacles and interference phenomena in a reduced space. Considering these problems, localization systems focused on indoor environments bring new challenges for the future of communication systems [1], [2].

In the last years, there is a growing demand for industrial applications that need IPSs. In most cases an IPS locates individual assets for industrial automation and find specific items in distribution centers. In public security and military use, such systems are needed to aid in the navigation of police officers, firefighters and soldiers in their missions inside buildings [3].

Localization of objects in mobile scenarios tends to be easier than in static scenarios. Scene analysis in a mobile environment provides fingerprints that change for a given target, which can help in object tracking.

There is a lack of research on low-cost IPSs with item-level accuracy applied to stationary objects. This work proposes a new IPS to meet these requirements. To achieve better accuracy, machine learning models based on support vector

regression (SVR) and k-means are employed. This work presents an IPS able to perform localization of stationary objects using off-the-shelf equipment.

The remainder of this paper is organized as follows: Section II presents a summary of related work. An overview of the proposed system is provided in Section III. Sections IV and V discuss the offline and online phases of the proposed system, respectively. Experiments and results are presented in Section VI and finally, Section VII contains the conclusion.

II. RELATED WORK

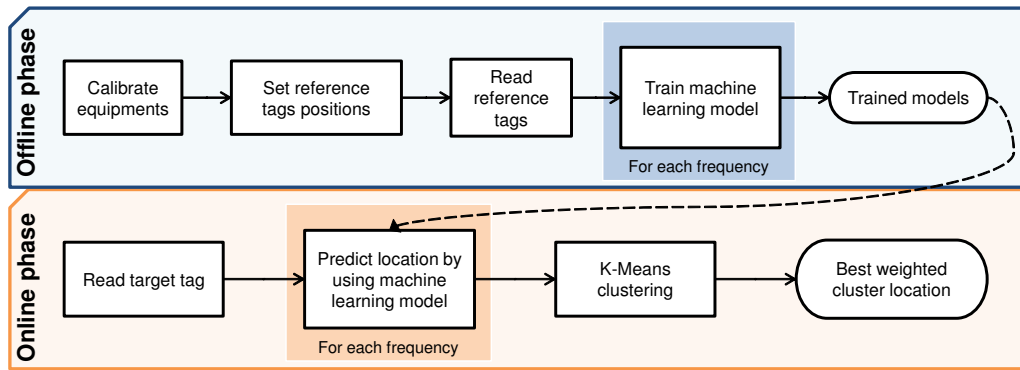
RFID (Radio-Frequency Identification) technology provides identification and localization of goods equipped with RFID tags at low cost. One characteristic that makes it attractive is the small size of its components and low power consumption, especially in passive tags, which are battery-free. When a reader reads a tag, it can also obtain the received signal strength indicator (RSSI) from the tag. RSSI information is a measure of power, usually represented in dBm [4].

In LANDMARC [5], RFID reference tags are placed on the environment, and the RSSIs are sensed by RFID readers. Tags in unknown positions are sensed, and their RSSIs are used in the nearest neighbor algorithm to find the closest reference tags and predict the unknown tag position. In [6], the LANDMARC technique is compared to a localization model based on artificial neural network (ANN). During the training phase, for each reader antenna, RSSI from the reference tags feed the network input. In the output layer, the coordinates (x, y) and orientation angle of the tags are given. The results show that the localization accuracy is 7 cm better than in the Landmarc system.

In [7], [8], a path-loss shadowing model is used to generate the RSSI fingerprint of the indoor environment. RSSI values from each reader antenna and the coordinates of the tags are provided as ANN inputs and outputs, respectively. Both works are based only on simulations and isotropic antennas.

Wille et al. [9] presents a support vector regression (SVR) localization approach for a medical navigation system. RFID phase difference is used as a nondeterministic indicator to train and run the SVR model. Experiments were performed inside a small plastic basin designed to emulate the human body and head. Phase data were collected by applying grids with 5 and 10 mm step sizes. The results showed an accuracy between 0.6 and 6.6 mm.

Fig. 1: Block diagram of system design. The inputs are the RFID readings and the output is the target object position.



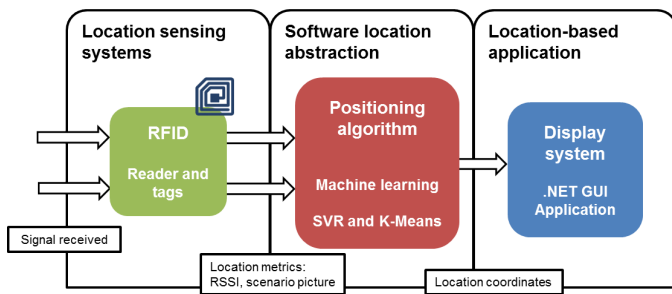
In [10], the LANDMARC approach is fused with a back-propagation network (BPN) model. First, LANDMARC uses measured RSSI values to calculate target tag coordinates. Due to the relationship between RSSI and distance is dynamic, the BPN adjusts the calculated coordinates to increase location accuracy. Results shown a 56 cm error rate when reference tags are 30 cm apart from each other. Contrary to our approach, reference tags must be present during online phase, which can difficult deployment and maintenance of the system.

III. SYSTEM OVERVIEW

The main elements of an IPS are (i) location sensing devices that measure metrics related to the relative position of a target object; (ii) an algorithm or localization technique that processes metrics reported by location sensing devices; and (iii) a display system that graphically illustrates the location of the target object to users [3]. As shown in Figure 2, these elements can be associated with a three-tier architecture [1]. In this work, the following technologies were used and associated to each layer: (i) location sensing: RFID; (ii) technique: machine learning models; (iii) graphical interface: .NET GUI application.

This IPS proposal is applied to a scenario where RFID tags are attached to objects we need to locate. The RFID tag uniquely identify each object in the scenario. Thus, using support vector regression (SVR) and k-means models, the IPS must be able to estimate the position of each tag present in the scenario.

Fig. 2: Layered architecture associated with a block diagram containing fundamental components of proposed IPS.



The RFID reader is connected to a computer running the system. A Matlab library [11] has been integrated into .NET C# to train and run the SVR model. A .NET GUI application was developed to automate tests and graphically show results to users.

The proposed localization technique works in two phases, online and offline. The offline phase is performed only once for the chosen scenario. The online phase is performed as often as necessary for each object we want to locate. Figure 1 shows the two phases, processes and the flows between them. The phases are detailed in the next sections.

IV. OFFLINE PHASE

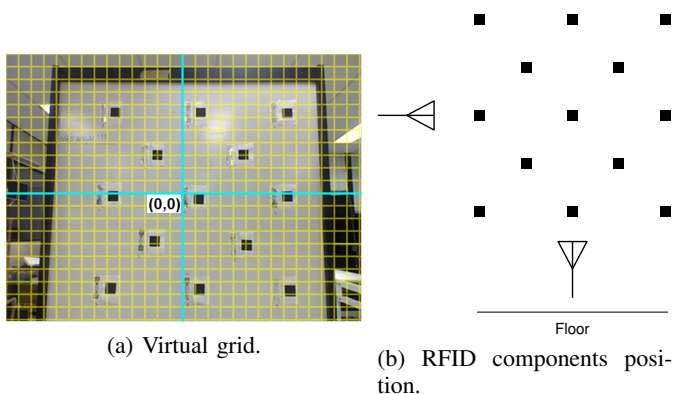
Initially, the parameters and positions of the RFID equipment are set to values that must remain fixed throughout system operation. Some calibrations performed at this phase are of the reader power, region of operation (frequency range) and antennas positions.

The proposed RFID system is based on the RSSI value of each tag to estimate its location in the scenario. As a statistical model is proposed, data collection from reference tags is required. This step requires that reference tags be uniformly distributed in the environment. During the experiment test bed, tag positions in simple grid and diagonal mesh scenarios were evaluated. Diagonal mesh design (Figure 3) obtained better performance, and it was chosen for the rest of this work.

After initial adjustment, reference tag positions must be stored in the system configuration. Spatial coordinates (x, y) are translated to coordinates of a virtual grid created over an image. A scenario picture is taken, and the grid is defined over the image (Figure 3a). Thus, the system may output the position of any cell within the limits of the captured image.

Reader antenna position plays a key role in the accuracy of the IPS. In initial tests, two antennas were placed in front of the tags, but the RSSI values of tags in different positions were the same, making it impossible to have a reasonable RSSI interpolation during prediction. Thus, for each axis of the virtual grid, it was decided to place antennas in positions such that RSSI values decrease as the distance increases. Therefore, in a 2D scenario, at least two antennas must be present in the

Fig. 3: Virtual grid over a picture captured from the training scenario (a) and positioning of reference tags and antennas over the diagonal mesh design (b). RFID tags are represented by black squares.



system (x axis and y axis). This arrangement can be seen in Figure 3b.

After these configuration steps, the reference tags are read and the data are collected. The RFID reader is activated for a fixed time period, and the system collects the following data: the antenna ID that senses the tag, the frequency in MHz, the RSSI and the position (x, y) of the reference tags present in the scenario.

A. Multi-frequency

RSSI values have been widely used in IPSs. However, even in static environments, this value can vary according to the operating frequency. This difference could be due to physical characteristics of the equipment, interference (like other RF equipment in the same frequency), obstacles, and other factors [12].

In attendance to federal regulations like FCC (USA) and ANATEL (Brazil), UHF RFID equipment cannot stay on the same frequency for more than 0.4 seconds in a 10 second interval [13], [14]. Addressing this constraint, RFID readers hop on to each available 250 KHz channel, limiting the possibility of running on a fixed frequency. This feature can be considered a constraint if we wish to use off-the-shelf equipment in the IPSs. For example, RSSI values trained at the frequency 915.25 MHz may vary considerably from values measured in the online phase at the frequency 923.25 MHz.

To overcome these limitations, we propose to partition the data collected in both phases using the operation frequency. Thus, in the offline phase, machine learning models for each sensed frequency are created. Consequently, in the online phase, each model uses data from the respective frequencies to estimate the location of the target objects. This method aims to separate the RSSI values of distinct frequencies, avoiding the mentioned constraints. Thus, it is possible to run statistical localization systems by using equipment that complies with federal regulations.

B. SVR model

Support vector machine (SVM) is a supervised learning algorithm for classification. To apply it in non-linear regression models, it has been modified and referred to as support vector regression (SVR) [9], [15]. The localization problem of this work is a regression problem instead of a classification problem. As stated in Section IV, the target tag position is given by spatial coordinates rather than a region or proximity.

Given a training dataset $\{(x_1, y_1) \cdots (x_n, y_n)\} \subset X \times \mathbb{R}$, where X denotes the space of the input patterns, x_i and corresponding target values y_i are a combined training set. The SVR goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data. Thus, the linear approximation function is described as:

$$f(x) = \langle w, x \rangle + b \text{ with } w \in X, b \in \mathbb{R} \quad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in X . However, the problem is not always feasible, because there are points that violate the restrictions. To avoid overfitting, one should add a capacity control term, which in the SVM case is $\|w\|^2$. Formally, we can write this problem as an optimization given by

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{subject to} \quad & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (2)$$

where C is a regularization parameter that controls the trade-off between penalizing violations of the accepted interval ε (denoted by ξ and ξ^*) and the complexity of the decision function $f(x)$. A solution of the convex optimization problem is usually found by means of an equivalent dual formulation.

The dual formulation of the SVR problem provides an alternative to working in a high dimensional space. Thus, it is possible to map the data into higher dimensional spaces in the hope that the data could become more easily separated or better structured. To accomplish this, kernel functions approaches are used.

In our proposal, we use a Matlab implementation [11] of SVR with a wavelet kernel [16]:

$$K(x, z) = \prod_{i=1}^n \left[\cos \left(1.75 \frac{x_i - z_i}{a} \right) \exp \left(-\frac{(x_i - z_i)^2}{2a^2} \right) \right] \quad (3)$$

where x , z and a are the wavelet dilation and translation coefficients. More details and concepts about SVR can be found in Cristianini [17] and Smola [15].

Data collected in the offline phase feed the SVR training process. All collected data are used, and any data removal or aggregation are performed at this stage. As stated in Section IV-A, the data are separated by operation frequency, and a SVR model is created for each frequency. RSSI values for

each antenna are presented as inputs, and the virtual grid coordinates (x, y) of each reference tag are the target output data.

Most RFID readers collect individual tag readings per antenna. To submit samples to the model, these records should be merged. Thus, for each reference tag position, the total number of samples will be equal to the record count of the antenna that had fewer readings. Figure 4 shows an example of samples collected by the RFID reader and the merge operation.

Fig. 4: Collected samples (left) and data merge (right). Frequency: 923.25 MHz.

Antenna	RSSI (dBm)	Position (x, y)
2	-61.5	(-7, 11)
1	-44.5	(-7, 11)
1	-44.5	(-7, 11)
2	-61	(-7, 11)
2	-60.5	(-7, 11)
1	-44.5	(-7, 11)
1	-44.5	(-7, 11)
1	-44.5	(-7, 11)
1	-44	(-7, 11)
...
1	-49.5	(8, 3)
1	-50	(8, 3)
1	-50	(8, 3)
2	-55.5	(8, 3)
1	-49.5	(8, 3)
2	-56	(8, 3)
2	-55.5	(8, 3)
1	-49.5	(8, 3)
1	-50	(8, 3)
1	-50	(8, 3)
...

Input: RSSI (dBm)		Output: position	
Antenna 1	Antenna 2	x	y
-44,5	-61,5	-7	11
-44,5	-61	-7	11
-44,5	-60,5	-7	11
...
-49,5	-55,5	8	3
-50	-56	8	3
-50	-55,5	8	3
...

Discarded

In SVR, only one target value is possible for each calculus, so we create one SVR model for each target coordinate x and y . We cross-validated values for SVR coefficients, and based on the results, they were set as $\varepsilon = 0.00025$, $c = 40000$ and $a = 4$ (wavelet).

V. ONLINE PHASE

This phase determines the final location of the target object. The system estimates target object positions using the trained model and the k-means method. During the online phase, no reference tags need to be present in the scenario, and an unknown RFID tag is read during a fixed period of time. The data collected by the RFID reader are antenna ID, frequency and RSSI. The data merge procedure (Fig. 4) and multi-frequency technique (Section IV-A) performed in setup phase are also applied on this phase.

Once the SVR model has been optimally trained, data from an unknown tag are presented to predict its location. For each frequency, RSSI values from an unknown tag are presented to the trained SVR model, and tag location is predicted. Each output coordinate has its own SVR model. Thus, each respective model is evaluated in order to estimate coordinates x and y .

A. K-means

The RFID reader collects dozens of readings for each tag. Currently, equipment and protocols allow a large number of

readings in a short period of time. For example, in 3 seconds sensing a tag, 46 readings are collected. RSSI values for the same tag suffer significant variations that can affect estimated positions, i.e., different positions for the same tag are predicted. This may occur if a given frequency model has low performance in the training process or due to interferences, obstacles and multipath effects.

Thus, some technique is required to provide the final location of the target object. Initial tests show that a simple mean of the location predictions would not bring about the desired results. To merge these predictions and provide a reasonable location, the k-means technique is used.

K-means clustering is a partitioning method that partitions data into k mutually exclusive clusters. K-means operates on actual observations and creates a single level of clusters. K-means treats each observation as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible and as far from objects in other clusters as possible [18].

Given a set of observations (x_1, x_2, \dots, x_n) , where each observation is a d -dimensional real vector, k-means clustering aims to partition the n observations into $k (\leq n)$ sets $J = \{J_1, J_2, \dots, J_k\}$.

Each cluster in the partition is defined by its members (i.e., observations) and by its centroid, or center. The centroid for each cluster is the point at which the sum of distances from all members in that cluster is minimized. K-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The algorithm aims at minimizing an objective function

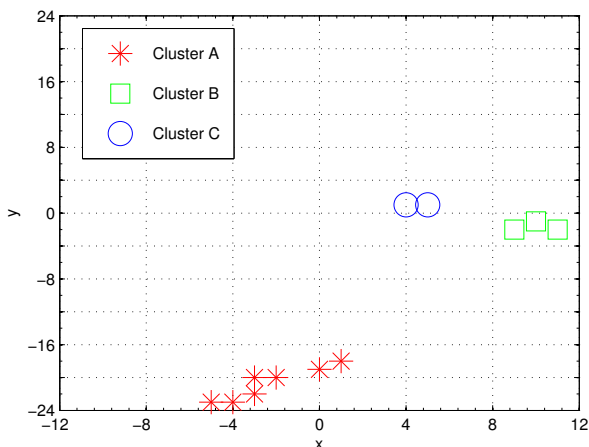
$$S = \sum_{j=1}^k \sum_{i=1}^x \left\| x_i^{(j)} - c_j \right\|^2 \quad (4)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center c_j , and an indicator of the distance of the n data points from their respective cluster centers.

In our model, tag locations predicted by machine learning techniques are the k-means observations, and squared Euclidean is the distance measure. As estimated positions from certain frequency models may differ from other frequencies, k is defined as $k = d - 1$, where d is the number of sensed frequencies. Thus, it is more likely that predictions from noisy frequencies are grouped in their own clusters.

We also define a weight for each cluster from member count information. In a good weighted cluster there are more members than in a bad weighted cluster. That is, a good weighted cluster aggregates more closest predicted locations than a bad cluster. In most cases, a cluster with few members presents spurious positions or locations where machine learning models did not perform well. The best weighted cluster is the cluster that has more similar locations.

Fig. 5: K-means clustering applied to predicted locations over virtual-grid coordinates (x, y) .



Finally, the centroid location of the best weighted cluster is defined as the final target location.

Figure 5 shows clusters extracted from a set of locations predicted for a given tag. Samples from four operation frequencies between 923 and 924 MHz were used. In this case, cluster A is the best weighted as it has more similar locations.

VI. EXPERIMENTS AND RESULTS

In the experiments, the localization system was run in a laboratory where tags were attached on a whiteboard, which is 1.5 m in width and height (2.25 m² area). In the offline phase, reference tags were positioned in diagonal mesh over the board and antennas placed on each side, as discussed in Section IV. Diagonal distance between each reference tag was 28 cm.

A Speedway Revolution R420 RFID reader and a Threshold RFID antenna, both from Impinj, were used in the experiments. Threshold is a far-field antenna, which operates in a frequency range of 902 – 928 MHz. The RFID tag used is a RafSec DogBone Wet Inlay, which operates in the frequency range of 860 – 960 MHz. Reader operation frequency was defined to use the follow values: 923.25, 923.75, 924.25 and 924.75 MHz. Reader power was set to a maximum value of 32.5 dBm.

The system was evaluated in four different places of the laboratory (P1-P4). RFID reader antennas were placed under and on the right side of the whiteboard. During the offline phase, 13 reference tags were used, and the RFID reader was activated for 10 s. The number of samples collected to feed the SVR model was 500 on average. Figure 6 shows the test bed environment and all system components.

In the online phase, the goal of the experiment was to locate six target tags distributed in the environment. Three tags were in positions already used in the offline phase, and three tags were in unknown positions. For each tag, RFID reader was activated by 3 s.

Fig. 6: Test bed environment and all system components.

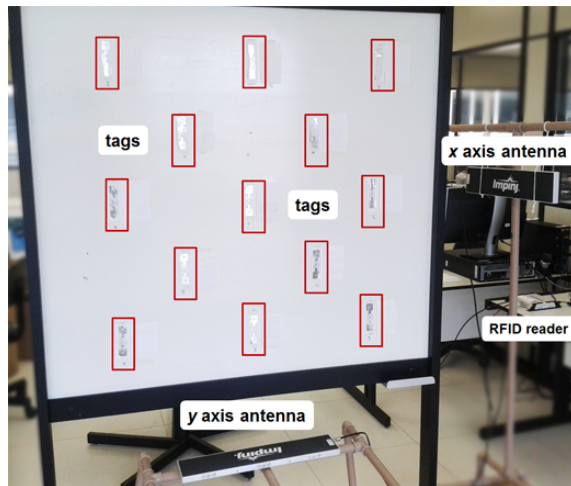


Fig. 7: Screenshot of the system running. Target tag ID: 16; localization error: 11.1 cm.

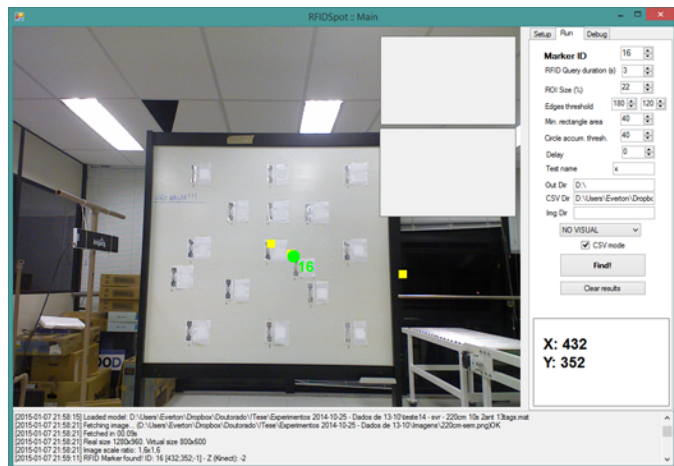


Figure 7 presents a screenshot of the system showing SVR predictions (yellow squares) and the k-means final location (green circle). The graphical user-interface helps users to easily identify the tag location. Time latency to localize each tag is less than one second, excluding RFID reader time.

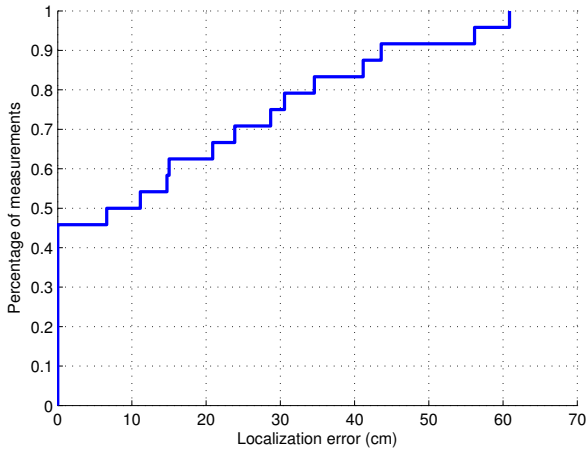
Typically, localization error is given by the Euclidean distance between estimated and actual locations. The cumulative distribution function (CDF) of the error distance for the SVR model is presented in Figure 8.

CDF results show the localization error is 0 cm for 48 % of the experiments. In worst case, the error reaches 61 cm. To summarize the localization accuracy for each scenario, the root mean square error (RMSE) of the location estimates is calculated as the difference between the predicted and actual location as

$$RMSE = \sqrt{\frac{\sum_{t=1}^k (\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2}{k}} \quad (5)$$

where \hat{x}_t, \hat{y}_t describe the estimated locations, x_t, y_t are the

Fig. 8: Cumulative error distance for SVR model localization.



actual positions and k is the number of predictions. Table I shows the RMSE performance of the system.

The results show a localization error between 17 and 31 cm in the range of 1.5 m. Places P3 and P4 have the worse performance, mainly due to multipath effects and interferences present in online phase. The overall RMSE is 25 cm, which brings some limitation in scenarios where many items are close to each other.

TABLE I: Localization performance (RMSE in cm) for each scenario.

Place	P1	P2	P3	P4	RMSE
Error (cm)	17.6	18.2	30.2	30.9	25.0

In comparison to related works, the proposed system performs 31 cm better than a neural network RFID-based approach [10], where the distance between reference tags is similar to our work.

VII. CONCLUSION

We have presented a system for localization of stationary objects using off-the-shelf equipments. Given objects with RFID tags attached, the system localize them. In order to achieve this, we proposed machine learning model able to learn RSSI fingerprints and to predict tags locations. A multi-frequency technique is proposed to overcome constraints from off-the-shelf RFID readers. A k-means technique were applied to enhance the localization.

We conducted real world experiments to evaluate the localization performance. Results show a 17.6 cm accuracy in best case scenario and demonstrate a 55% improvement over other techniques.

This system can be easily extended to 3D scenarios, which would provide the depth distance of objects. In future works, systems scalability will be tested through experiments in large environments using multiple readers.

REFERENCES

- [1] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 13–32, 2009. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4796924>
- [2] T. Sanpechuda and L. Kovavisaruch, "A review of RFID localization: Applications and techniques," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*. IEEE, May 2008, pp. 769–772. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4600544>
- [3] K. Pahlavan and J. Makela, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, 2002. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=983917>
- [4] L. Ni, D. Zhang, and M. Souryal, "RFID-based localization and tracking technologies," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 45–51, Apr. 2011. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5751295>
- [5] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID," *Wireless networks*, vol. 10, no. 6, pp. 701–710, 2004.
- [6] A. Chattopadhyay and A. R. Harish, "Analysis of low range Indoor Location Tracking techniques using Passive UHF RFID tags," in *2008 IEEE Radio and Wireless Symposium*. IEEE, Jan. 2008, pp. 351–354. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4463501>
- [7] A. S. M. Sala, R. G. Quirós, and E. E. López, "Using neural networks and Active RFID for indoor location services," in *RFID Systech 2010 - European Workshop on Smart Objects: Systems, Technologies and Applications*, 2010, p. 9. [Online]. Available: <http://www.vde-verlag.de/proceedings-en/453282020.html>
- [8] —, "Active RFID Reader clustering and neural networks for indoor positioning," in *The Third International EURASIP Workshop on RFID Technology*, 2010, pp. 62–65.
- [9] A. Wille, M. Broll, and S. Winter, "Phase difference based RFID navigation for medical applications," in *2011 IEEE International Conference on RFID*. IEEE, Apr. 2011, pp. 98–105. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5764608>
- [10] H.-Y. Kung, S. Chaisit, and N. T. M. Phuong, "Optimization of an RFID location identification scheme based on the neural network," *International Journal of Communication Systems*, p. 20, Nov. 2013. [Online]. Available: <http://doi.wiley.com/10.1002/dac.2692>
- [11] R. Clark, "A MATLAB implementation of Support Vector Regression (SVR)," 2014. [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/43429-support-vector-regression>
- [12] A. Whitney, J. Fessler, J. Parker, and N. Jacobs, "Received signal strength indication signature for passive UHF tags," in *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, vol. 0503. IEEE, Jul. 2014, pp. 1183–1187. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6878242>
- [13] Federal Communications Commission (FCC), "Electronic Code of Federal Regulations - Section 15.247," 2014. [Online]. Available: <http://www.ecfr.gov/>
- [14] Brazilian Agency of Telecommunications (ANATEL), "Resolution nº 506 - Article 40," 2008.
- [15] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," Tech. Rep., 1998.
- [16] L. Zhang, W. Zhou, and L. Jiao, "Wavelet Support Vector Machine," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 34–39, Feb. 2004. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1262479>
- [17] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press, 2000.
- [18] The MathWorks Inc., "k-Means Clustering," 2014. [Online]. Available: http://www.mathworks.com/help/stats/k-means-clustering.html#bq_679x-19