A Hybrid RFID and CV System for Item-Level Localization of Stationary Objects

Everton Luís Berz[1], Deivid Antunes Tesch[2], Fabiano Passuelo Hessel[3]
PUCRS University
Porto Alegre, Brazil
[1]everton.berz@acad.pucrs.br, [2]deivid.tesch@acad.pucrs.br, [3]fabiano.hessel@pucrs.br

Abstract

In order to optimize the user experience and solve logistical and security issues, many systems require the physical location information from objects and people. Indoor positioning systems (IPSs) based on more than one technology can improve localization performance by leveraging the advantages of different technologies. This work proposes a hybrid IPS able to estimate the item-level location of stationary objects using off-the-shelf equipment. By using RFID technology, machine learning approaches based on artificial neural networks (ANNs) and support vector regression (SVR) are proposed. A k-means technique is also applied to improve accuracy. A computer vision (CV) subsystem detects visual markers in the scenario to enhance RFID localization. To combine the RFID and CV subsystems, a fusion method based on region of interest (ROI) is proposed. We have implemented our system and evaluated it using real experiments. The localization error is between 9 and 29cm in the range of 1 and 2.2m scenarios. In a machine learning approach comparison, ANN performed 31% better than SVR approach.

Keywords

Indoor positioning systems, RFID, sensor fusion, visual analysis

1. Introduction

Localization of objects and people in indoor environments has been widely studied mainly because of security and logistical issues. Indoor environments have a high density of obstacles and interference phenomena in a reduced space, which makes them more complex than outdoor environments. Due to these reasons, localization systems focused on indoor environments bring new challenges for the future of communication systems [1, 2].

Many applications require more precise information on the location of smaller objects (item-level). For example, item retrieval is a natural extension of traditional inventory systems. In this application, a human or robot needs to find the location of a given item, whereas the location must be accurate enough for the item to be collected properly even if there are nearby objects or any other type of interference.

The motivation of this work is to propose a low-cost and high-accuracy IPS that can be used in a large amount of items (as we have in logistics/distribution centers). The current IPSs do not meet these requirements and, nowadays, companies choose to manufacture again goods that are lost in distribution centers, increasing their production costs. There are few researches on low-cost IPSs with item-level accuracy applied to stationary objects. This work proposes a new IPS to meet these requirements. To achieve a better accuracy, the proposal also aims to define a new hybrid mechanism based on RFID and computer vision (CV). This work presents a hybrid IPS able to perform item-level localization of stationary objects using off-the-shelf equipment.

Our contribution to the state of indoor positioning systems are novel machine learning models and a sensor fusion method able to perform item-level localization of stationary objects using off-the-shelf equipment. A novel multi-frequency technique is proposed to allow the use of off-the-shelf RFID equipment. Besides that, passive RFID tags and a cheap camera are used, which represents reduction in the cost of infrastructure. Additionally, while many IPSs work only under dynamic scenarios, our system is able to localize objects on stationary environments.

The remainder of this paper is organized as follows: Section 2 presents a summary of related work. An overview of the proposed system is provided in Section 3. Sections 4 and 5 discuss the offline and online phases of the proposed system, respectively. Experiments and results are presented in Section 6 and finally, Section 7 contains the conclusion.

2. Related Work

Technologies used in IPSs have different costs and performance. For instance, usually the cost of a RFID passive system is lower than systems based on technologies like active RFID, ultrasound, infrared and UWB. Another systems, like WiFi and ZigBee, are not so expensive but they have low accuracy (2 m to 3 m) or require a large infrastructure (i.e. a large number of antennas) to achieve an item-level performance [3, 4]. In literature, there are surveys that compare the state-of-art of IPSs from distinct technologies [5].

In our previous work [6], a machine learning model based on support vector regression (SVR) is proposed for localization of stationary objects using off-the-shelf equipments. This model learns RSSI fingerprints during an offline phase and then predict tags locations in an online phase, where no reference tags are needed. Experiments were performed in four different places inside a laboratory where tags were attached on a whiteboard, which is 1.5 m in width and height. This technique presented a location error between 17 and 31 cm in 2.25 m^2 area coverage.

Wille et al. [7] presents a support vector regression (SVR) localization model for a medical navigation system. In order to train and run the SVR model, the RFID phase difference is used as a nondeterministic indicator. Phase data were collected by applying grids with 5 and 10 mm step sizes. The results show an accuracy between 0.6 and 6.6 mm.

In [8], a back-propagation network (BPN) model is fused with the LANDMARC system. First, LANDMARC uses measured RSSI values to calculate target tag coordinates. Due to the dynamic relationship between RSSI and distance, the BPN adjusts the calculated coordinates to improve location accuracy. Results show 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 the online phase, which can hamper the system's deployment and maintenance.

An RFID localization system combined with other technologies such as optical, inertial and ultrasonic systems is a growing trend in the field. Nick et al. [9] presents a tracking system for trolleys carrying boxes leaving or coming into a mail distribution center. Using a RFID reader and four antennas attached to the ceiling, the relations between the RSSI and different measured distances are stored and later estimated. In the CV system, sample images from the target object are captured, and thresholding and morphological operations are then applied to recognize the object in the image. Sensor data fusion is performed by a constrained unscented Kalman filter (CUKF) technique. Localization errors were 26 cm and 36 cm for stationary and moving scenarios, respectively.

3. System Overview

This IPS proposal is applied to a scenario where RFID tags and visual markers are attached to objects we need to locate. The RFID tag and its visual marker uniquely identify each object in the scenario. In this work, each pair of RFID tag and visual marker is simply referred as a marker. Thus, the IPS must be able to estimate the position of each marker present in the scenario.

The proposed system is divided into two subsystems. The first employs machine learning models to predict the location using RFID technology, while the second uses a camera and CV algorithms to enhance the predictions estimated by the first subsystem. The IPS works in two phases, online and offline. The offline phase is performed only once for the chosen scenario. The online phase is run as often as necessary for each target object we need to locate. Figure 1 illustrates the two subsystems, processes, phases and the flows between them.

4. Offline phase

The proposed RFID subsystem predicts the position of each tag through RSSI values. As a probabilistic model is proposed, data collection from reference tags is needed. In this step, reference tags are uniformly distributed in the environment. On the experiment test bed, tag positions were evaluated in diagonal mesh and simple grid scenarios. For the rest of this work, diagonal mesh design (Figure 2) was chosen due to its better performance.

After initial configuration, reference tag positions are stored in the system. Spatial coordinates (x, y) are translated to coordinates of a virtual grid over a scenario picture (Figure 2a).

Reader antenna position plays a critical role in IPS accuracy. For each axis of the virtual grid, antennas were arranged such the RSSI values decrease as the distance increases. Thus, at least two antennas are needed in a 2D scenario (x axis and y axis).

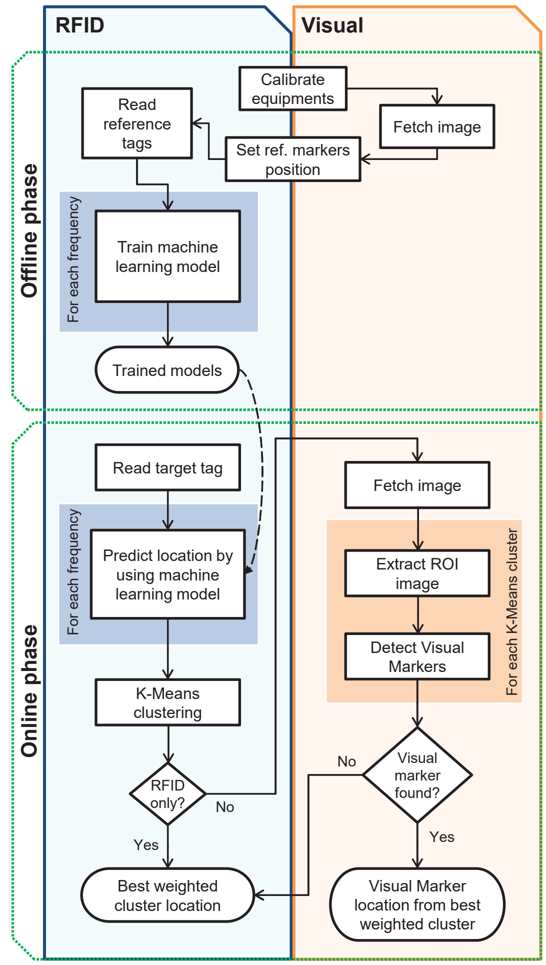


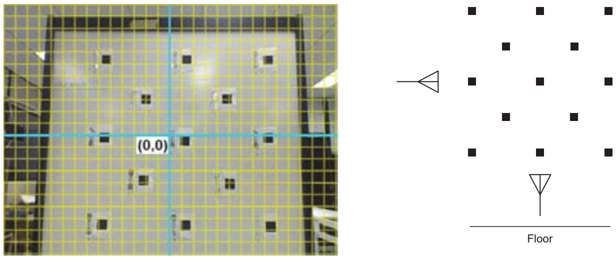
Figure 1: Block diagram of system design. The inputs are the RFID readings and the camera picture. The output is the target object position.

After the configuration steps, the RFID reader is turned on for a fixed time range, and the system gathers the following information: the frequency in MHz, the antenna ID that senses the tag, the RSSI, and the position (x, y) of the reference tags present in the scenario.

4.1 Multi-frequency

In meeting federal regulations like ANATEL (Brazil) and FCC (USA), UHF RFID equipment cannot stay on the same frequency for more than 0.4 seconds in a 10 second range [10, 11]. Addressing this requirement, RFID readers hop on to each available 250 KHz channel, limiting the ability of running on a fixed frequency. If off-the-shelf equipment are used in IPSs, this feature can be a constraint. For instance, RSSI values trained at frequency 915.25 MHz may have significant variation from values measured in the online phase at frequency 923.25 MHz.

To overcome these constraints, we propose to partition the data collected in both phases by operation frequency. Thus, in the offline phase, machine learning models for each detected frequency are designed and trained. Accordingly, in the online phase, each model uses data from the respective frequencies to predict objects' location. Figure 3 illustrates an example of



(a) Virtual grid. **(b)** RFID components position.

Figure 2: Virtual grid over a picture captured from the training scenario (a) and positioning of reference tags and antennas over the diagonal mesh design (b).

this technique. This method intends to partition RSSI values of distinct frequencies, avoiding the aforementioned constraints. In this way, statistical localization systems can use equipment that complies with federal regulations.

4.2 ANN model

Data gathered in the offline phase feed the ANN (Artificial Neural Network) training process. All data are used, and any aggregation or data removal is performed at this step. As stated in Section 4.1, the data are partitioned by operation frequency, and a neural network model is created for each frequency. RSSI values for each antenna are the network inputs, and the virtual grid coordinates (x, y) of each reference tag are the target output data.

Reference tag data are divided into three subsets: training, validation and testing, randomly divided into the ratio of 0.7, 0.15 and 0.15. The training set is used for computing the gradient and updating the network weights and biases. The validation set ensures that there is no overfitting in the final result. Testing set error is useful to indicate a poor division of the data set.

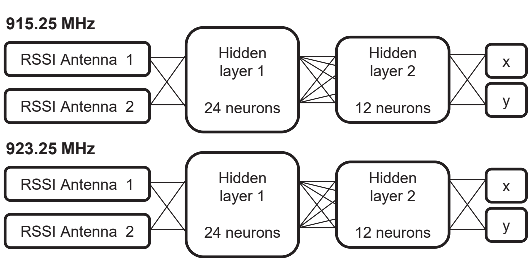


Figure 3: ANN models for each frequency.

A feedforward backpropagation network with four layers is modeled, which consists of n neurons in the input layer, 24 neurons and 12 neurons in each hidden layer, and 2 neurons in the output layer. The number of neurons in the input layer must be equal to the number of antennas in the scenario. Figure 3 shows the network configuration. The hidden layer numbers have been chosen because beyond or below these numbers, the performance does not improve and computational time increases considerably. The Levenberg-Marquardt backpropagation algorithm was used to update weight and bias values. The mean square error (MSE) (also known as the validation error) was used to measure the performance of the network.

4.3 SVR model

Support vector machine (SVM) is a supervised learning algorithm for classification. It has been modified to be applied in non-linear regression models, becoming known as support vector regression (SVR) [7, 12]. The localization problem of this work is a regression problem instead of a classification problem. As aforementioned in Section 4, the target marker location is given by spatial coordinates rather than by region or proximity.

In our proposal, we use a Matlab implementation [13] of SVR with a wavelet kernel [14]:

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

where x, z and a are the wavelet dilation and translation coefficients. More details and concepts about SVR can be found in Cristianini [15] and Smola [12].

SVR is modeled similar to ANN, where the RSSI values sensed by each antenna is presented as training datasets and the virtual grid coordinates of each reference tag are the output data. In SVR, only one target value is possible for each calculus, so we create one SVR model for each target coordinate x and y. The multi-frequency technique (Section 4.1) is also applied. 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).

5. Online phase

In this phase, a sensor fusion approach is proposed. Sensor fusion seeks to improve accuracy and precision by integrating many location or positioning systems to form hierarchical and overlapping levels of resolution [16]. This phase runs the RFID and visual subsystems to determine the final location of the target object. RFID subsystem estimates target object positions using the trained models and the k-means method. In this phase, the RFID subsystem aims to detect regions of interest (ROIs), which will later be used by the visual subsystem to estimate a more accurate location of the target object. A ROI can be defined as an area of limited size, smaller than the size of the complete scenario. Because other localization techniques can be applied to this small area, a better performance in the global localization can generally be achieved.

In the online phase, no reference tags are needed in the scenario, and an unknown RFID tag is read during a fixed period of time. The RFID reader collects the following data: antenna ID, frequency and RSSI. The multi-frequency technique (Section 4.1) is also applied in this phase.

Once the network has been optimally trained, data from an unknown tag are presented to estimate its location. For each frequency, a respective ANN model is automatically chosen, and RSSI data from each antenna are presented as input to the network. Tag location (virtual grid coordinates) is predicted and presented as the network output.

In the SVR approach, the online procedure is very similar to that of the ANN model. The trained SVR model receives RSSI values from an unknown tag, and tag location is estimated. In SVR, each output coordinate has its own SVR model. In this way, each respective model is evaluated in order to predict coordinates x and y.

5.1 K-means

As the RFID reader gathers dozens of readings for each tag, some technique is needed to provide the final position of the target object. An initial evaluation shows that a simple mean of the predictions would not bring about the desired results. Thus, the k-means technique is used to merge these predictions and provide a reasonable location.

In our approach, the estimated tag locations, obtained by the machine learning technique, are observations of the k-means model, and the squared Euclidean is the measured distance. Since predicted locations from models of a given frequency may be different from models at other frequencies, the variable k is defined as $k = d - 1$, where d is the number of detected frequencies. In this way, the predicted values from noisy frequencies are more likely to be grouped in their own clusters.

Figure 4 shows clusters with samples from four operation frequencies between 923 and 924 MHz. These samples were extracted from a set of estimated positions for a given tag. Cluster A has more similar locations and it is the best weighted cluster.

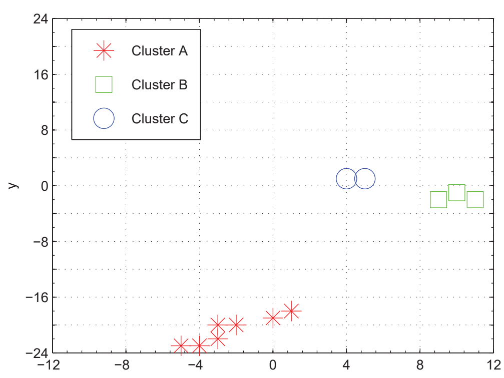


Figure 4: K-means clustering applied to predicted locations over virtual-grid coordinates (x, y).

If the system is running in RFID-only mode, the centroid location of the best weighted cluster is defined as the final target location. Otherwise, centroids and weights from all clusters are given as ROIs for localization enhancement using CV, presented in the next section.

5.2 Computer Vision for fine localization

To refine the results obtained by the RFID subsystem, we propose a sensor fusion based on RFID estimates and CV recognition. This work proposes a multiple ROI approach, where ROI is a small image extracted from the scenario picture that will be analyzed using CV. As the RFID subsystem estimates more than one location, CV method explores multiple regions in order to find a visual marker.

Figure 5: GetVisualMarkerLocation(img, C)

input: img, photo captured from scenario
input: set C, k-means clusters positions (C^p) and weights (C^w)
output: Final marker location

1: $bestWeight \leftarrow 0, M \leftarrow \emptyset, i \leftarrow 0$
2: **foreach** $cluster$ in C **do**
3: $subImg \leftarrow$ CropImage$(img, cluster^p)$
4: $S \leftarrow$ DetectShapesPos$(subImg)$
5: **foreach** $shapePosition$ in S **do**
6: $M_i^p \leftarrow$ GetImagePos$(shapePosition)$
7: $M_i^w \leftarrow cluster^w$
8: $i \leftarrow i + 1$
9: **if** $cluster^w > bestWeight$ **then**
10: $bestWeight \leftarrow cluster^w$
11: **end if**
12: **end for**
13: **end for**
14: **if** $M \neq \emptyset$ **then** ▷ Visual markers found
15: $B \leftarrow \{$indexes$(M^w) \mid M^w = bestWeight\}$
16: $F \leftarrow \{M_i^p \mid i \in B\}$
17: **return** $(\frac{1}{n}\sum_n^{i=1} F_i^x, \frac{1}{n}\sum_n^{i=1} F_i^y)$
18: **else** ▷ No markers found (RFID-only)
19: **return** C_i^p, where $i = \max(C^w)$)
20: **end if**

We use a simple visual marker represented by a black square on a white background. Each side of the visual marker is 4.5 cm long. Some advantages of this type of marker are its easy creation, low cost, fast detection and good accuracy. Because the marker is black and white, ambient lightning issues also tend to be reduced. This kind of marker also facilitates the adoption of another visual marker and CV algorithm as they do not need to be uniquely identified.

The Algorithm in Figure 5 shows the sequence of operations in the visual subsystem. For each k-means cluster, a small image of the scene is cropped, creating a sub-image. The center of the sub-image is based on the k-means cluster centroid location and its size is between 15% and 30% of the original scenario photo.

To detect the square shape in the sub-image (Alg. 5, Line 4), a canny edge detector is employed. The threshold value is set to 180, and the edge linking value is 120. From the canny edges image, polyline contours are detected and shapes whose angles are between $80 - 100$ degrees are selected. If square shape area is bigger than a configurable minimum size, it is recognized as a visual marker. We used EmguCV (.NET wrapper to OpenCV) [17] to implement this subsystem.

Finally, the target location is given by the position of the visual marker detected in the better weighted cluster position (Alg. 5, Lines 15-16). If more than one visual marker is detected in this cluster, the simple centroid of all finite points is calculated. If no visual marker is detected, the RFID-only location is given.

6. Experiments and Results

The experiments were performed in a laboratory (10 m x 7 m). Markers (tag and visual) were attached to a whiteboard, which is 1.5 m in height and width (2.25 m^2 area). In the offline phase, reference tags were placed in diagonal mesh over the board and antennas positioned on each side, as stated in Section 4. Each reference tag was placed at 28 cm equidistant from each other, diagonally.

The RFID reader operation frequency was defined to use the follow values: 923.25, 923.75, 924.25 and 924.75 MHz.

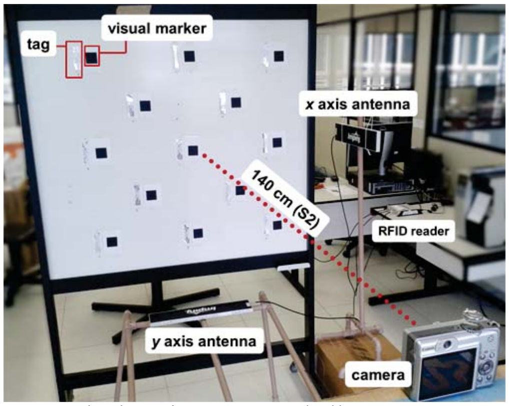
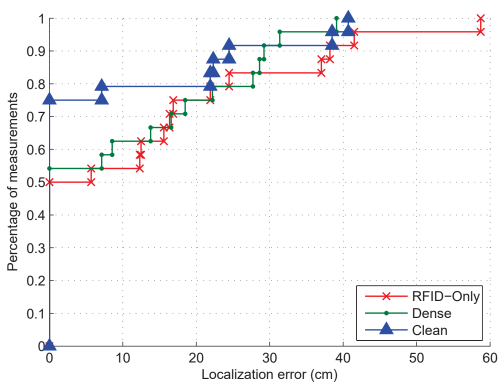
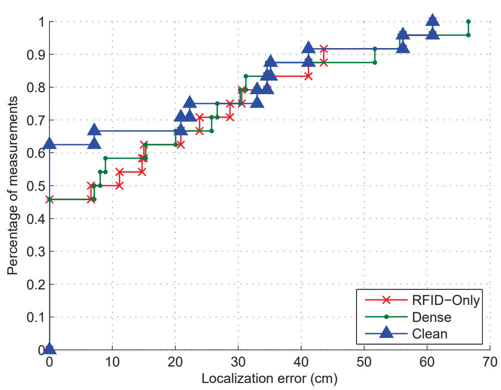


Figure 6: Test bed environment and all system components.

Reader power was set to a maximum value of 32.5 dBm. For visual localization, an inexpensive off-the-shelf camera with 1.3 megapixels (1280x960) and a 1/4" sensor was used. This type of camera was used to demonstrate that the system is able to operate with low-cost equipment and low-resolution images that allow fast CV analysis.



(a) CDF for ANN model localization.



(b) CDF for SVR model localization.

Figure 7: Cumulative error distance for both machine learning approaches.

The system was evaluated in four scenarios: S1, S2, S3 e S4. In each scenario, the distance between camera and markers was 100, 140, 180 and 220 cm, respectively. RFID reader antennas were positioned under and to the right side of the whiteboard. During the offline phase, 13 reference tags were used, and the RFID reader remained on for 10 s. The number of samples gathered to feed the models was 500 on average.

The training set MSE of neural network was 1.14 cm. Figure 6 shows the test bed environment and all system components.

In the online phase, the experiment's aim was to locate six target markers distributed in the environment, of which three were in locations already used in the offline phase, and three were in unknown locations. For each tag, RFID reader remained active for 3 s. Minimum area of the visual marker was 40 px. ROI size was defined to 30, 22, 17 and 15% for scenarios S1, S2, S3 and S4, respectively.

Three validation tests were performed. The first was RFID-only, i.e., visual markers were not present. In the dense test, 16 markers were attached on the whiteboard. In this test, the camera captured an image with all visual markers simultaneously. In the last test, referred to as the "clean" test, just one visual marker was present in the environment at each run.

Typically, localization error is given by the Euclidean distance between estimated and actual locations. The cumulative distribution functions (CDFs) of the error distance for the ANN and SVR models are presented in Figures 7a and 7b, respectively.

CDF results show the localization error is 0 cm for most experiments. For ANN model, 75% of the clean tests show an optimal accuracy (0 cm) and both hybrid tests do not exceed 40 cm error. The hybrid system did not detect the visual marker in 20% of the clean tests. In this case, RFID-only location is used as output. In dense test, the system detected a wrong visual marker in 25% of the experiments, most of them on long distances scenarios. To summarize the localization accuracy for each scenario, the root mean square error (RMSE) is calculated as the difference between the predicted and actual location (Table 1).

Table 1: Localization performance (RMSE in cm) for each scenario, validation test and machine learning approach.

Scenario	ANN			SVR		
	RFID-Only	Hybrid Dense	Hybrid Clean	RFID-Only	Hybrid Dense	Hybrid Clean
S1	12.1	12.2	9.4	17.6	28.4	16.8
S2	17.3	13.6	9.1	18.2	14.1	19.7
S3	33.3	23.7	22.9	30.2	26.8	28.4
S4	12.1	13.0	10.0	30.9	29.3	26.5
RMSE	20.6	16.3	14.1	25.0	25.4	23.3

In comparison between ANN and SVR approaches, the ANN model has better performance than SVR in most scenarios and tests. Overall results show that the ANN model performs 31% better than the SVR approach on average.

The results for the ANN approach show a localization error between 9 and 29 cm in the range of 1 and 2.2 m scenarios. Scenario S3 has the worst performance, mainly because RFID subsystem did not have a good accuracy due to multipath effects and interferences present in online phase.

Focusing in the ANN model, the hybrid system has better results than the RFID-only approach. Localization is improved by 21% and 32% for dense and clean tests, respectively. This demonstrates the effectiveness of the improvement brought about by the integration of the visual subsystem, even using simple visual markers and low-cost equipments.

The overall RMSE in dense and clean tests are 16.3 and 14.1 cm, respectively. Scenarios where the distance between camera and markers are shorter have the best results. These results demonstrate the system can be applied to item-level localization. However, the approach still has some limitations in scenarios where many items are close to each other.

In comparison to related works, in [7] the target scenario is very small (millimeter scale) and a direct comparison is not reasonable. However, when considering the same ratio between accuracy and reference tags distance, our system is 33% more accurate. The proposed hybrid system performs 40 cm better than a neural network RFID-based approach [8], where the distance between reference tags is similar to our work. Also, our IPS decreased the localization error in 45% than other stationary hybrid system [9].

7. Conclusion

We present a hybrid system for item-level localization focusing on stationary objects using off-the-shelf equipment. From objects with RFID tags attached, the system locates them precisely with the aid of visual markers. To achieve this goal, we propose machine learning models that learn the RSSI fingerprints and estimate the markers location. A multi-frequency approach is proposed to overcome the off-the-shelf RFID equipment limitations. To enhance the localization a k-means technique and computer vision algorithms were used.

Real-world experiments were performed to evaluate the localization performance and to compare the machine learning models. Results demonstrate a 32% improvement over RFID-only localization and a precision of 9.1 cm in the best-case scenario.

This system could provide in-depth distance to objects, if extended to 3D scenarios. In future work, experiments in large environments using multiple readers and cameras will be used to test the system scalability.

8. 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.

[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.

[3] M. Aravecchia and S. Messelodi, "Gaussian process for rss-based localisation," in *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2014, pp. 654–659.

[4] D. Taniuchi and T. Maekawa, "Robust wi-fi based indoor positioning with ensemble learning," in *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2014, pp. 592–597.

[5] Z. Farid, R. Nordin, and M. Ismail, "Recent Advances in Wireless Indoor Localization Techniques and System," *Journal of Computer Networks and Communications*, vol. 2013, pp. 1–12, 2013.

[6] E. L. Berz, D. A. Tesch, and F. P. Hessel, "RFID indoor localization based on support vector regression and k-means," in *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*. IEEE, jun 2015, pp. 1418–1423.

[7] 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.

[8] 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.

[9] T. Nick, S. Cordes, J. Gotze, and W. John, "Camera-assisted localization of passive RFID labels," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Nov. 2012, pp. 1–8.

[10] Federal Communications Commission (FCC), "Electronic Code of Federal Regulations - Section 15.247," 2014.

[11] Brazilian Agency of Telecommunications (ANATEL), "Resolution nº 506 - Article 40," 2008.

[12] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," Tech. Rep., 1998.

[13] R. Clark, "A MATLAB implementation of Support Vector Regression (SVR)," 2014.

[14] 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.

[15] 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.

[16] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," *Computer*, vol. 34, no. August, pp. 57–66, 2001.

[17] EmguCV, "EmguCV," 2014.