IET Cyber-Physical Systems: Theory & Applications

Special Issue: Reliability and Quality Control for Cyber-Physical Systems

Machine-learning-based system for multi-sensor 3D localisation of stationary objects

ISSN 2398-3396
Received on 31st May 2017
Revised 31st October 2017
Accepted on 6th December 2017
E-First on 22nd March 2018
doi: 10.1049/iet-cps.2017.0067
www.ietdl.org

Everton L. Berz[1] ✉, *Deivid A. Tesch*[1], *Fabiano P. Hessel*[1]

[1]*Department of Informatics, Faculty of Computer Science, PUCRS University, Porto Alegre RS, 90619-900, Brazil*

✉ *E-mail: everton.berz@acad.pucrs.br*

Abstract: Localisation of objects and people in indoor environments has been widely studied due to security issues and because of the benefits that a localisation system can provide. Indoor positioning systems (IPSs) based on more than one technology can improve localisation performance by leveraging the advantages of distinct technologies. This study proposes a multi-sensor IPS able to estimate the three-dimensional (3D) location of stationary objects using off-the-shelf equipment. By using radio-frequency identification (RFID) technology, machine-learning models based on support vector regression (SVR) and artificial neural networks (ANNs) 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 localisation. To combine the RFID and CV subsystems, a fusion method based on the region of interest is proposed. We have implemented the authors' system and evaluated it using real experiments. On bi-dimensional scenarios, localisation error is between 9 and 29 cm in the range of 1 and 2.2 m. In a machine-learning approach comparison, ANN performed 31% better than SVR approach. Regarding 3D scenarios, localisation errors in dense environments are 80.7 and 73.7 cm for ANN and SVR models, respectively.

1 Introduction

Indoor positioning systems (IPSs) have been required for different sorts of commercial applications. Such technology helps to find specific items in hospitals or distribution centres. There are also military and public security uses for these systems, as police officers, firefighters and soldiers use it for a better navigation during missions inside buildings [1]. Communication networks and telecommunication applications require several types of information about the environment, people and devices in order to offer flexible and adaptive services. In the future of communication systems, location information can bring many benefits such as the autonomous organisation of sensors or devices in *ad hoc* networks [2].

For smaller objects, on an item level, applications tend to depend on more specific information. In item retrieval, for instance, a robot or a person needs to find the exact location of a product. Accuracy must be high in order to complete the task properly, without interferences of any kind such as nearby objects. This application is a natural extension of regular inventory systems. Another one, called location assurance, is able to verify which items are in pre-specified locations [3].

Radio-frequency identification (RFID) is a low-cost technology which allows this identification and location process to happen. What makes it attractive is that it basically requires items to be equipped with RFID tags, which are small-sized, low-power-consumption equipment. Passive tags are even battery free. These components can also transmit the received signal strength indicator (RSSI), a measure of power commonly represented in dBm [4].

The motivation of this work is to propose a low-cost and high-accuracy IPS that can be used in a large number of items (as we have in logistics/distribution centres). The current IPSs do not meet these requirements and, nowadays, companies choose to manufacture again goods that are lost in distribution centres, increasing their production costs. 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 a better accuracy, the proposal also aims to define a new hybrid mechanism based on RFID and computer vision (CV). This work presents a multi-sensor IPS able to perform localisation of stationary objects over three-dimensional (3D) space using off-the-shelf equipment.

Our contribution to the state of IPSs is novel machine-learning models and a sensor fusion method able to perform localisation of static items on 3D space. The models and the fusion method provide a low error distance (item level) with a good success probability of position estimations, improving the localisation performance. Besides that, passive RFID tags and a cheap camera are used, which represent a reduction in the cost of infrastructure.

Scene analysis in mobile scenarios provides fingerprints that change, which helps to track a given target. Therefore, many IPSs work only under mobile environments, considering it is an easier task. Our system, on the other hand, is also able to localise goods on static environments.

The remainder of this paper is organised 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

RFID passive systems might be cheaper than IPSs based on other technologies such as active RFID, ultra-wideband, infrared or ultrasound. Yet, costs and performance can vary even more. Wireless fidelity and ZigBee, though not so expensive, have a very limited accuracy (3 m at most) or may require a large number of antennas in order to guarantee item-level performance. The state-of-the-art of these technologies has been already compared in the literature [5].

To perform the indoor localisation of objects and people, one of three main techniques can be chosen: triangulation (distance estimation), proximity and fingerprint analysis. Belhadi and Fergani [6] present a comparison between distance estimation (lateration) and the fingerprint techniques *k*-nearest neighbour (NN) and artificial neural network (ANN) for RFID indoor localisation. Simulation results show that the ANN algorithm outperforms the other techniques, though it requires a large deployment of reference tags.

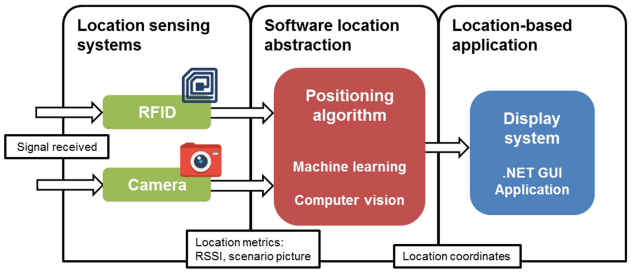

Fig. 1 *Layered architecture associated with a block diagram containing fundamental components of the proposed IPS*

RFID reference tags used in LANDMARC [7] are placed on the environment and RFID readers sense the RSSIs. Then, tags in unknown positions are sensed and the NN algorithm uses RSSIs to find nearby reference tags, thus predicting the position of an unknown tag. In [8], this technique is compared with an ANN-based localisation model. For each reader antenna, in the course of the training phase, RSSI from the reference tags feed the network input. The orientation angle and the coordinates (x, y) of the tags are given in the output layer. Results show localisation accuracy is 7 cm better than in the LANDMARC system.

In [9], a backpropagation network (BPN) model is merged with the LANDMARC approach. LANDMARC first uses RSSI values and calculates coordinates for target tags. Second, BPN adjusts such coordinates to a more accurate location. This is possible because the relationship between distance and RSSI is dynamic. In experiments where reference tags are 30 cm apart from each other, results show a 56 cm error rate. Conflicting with our approach, reference tags must be present during online phase. This can complicate system's development and maintenance.

An RFID localisation system combined with other technologies such as optical, inertial and ultrasonic systems is a growing trend in the field. Building a system that combines these technologies is an existing challenge. This combination is referred to as a hybrid system or sensor data fusion. Hybrid systems can improve localisation performance by leveraging the advantages of the different technologies [4, 10].

In our previous work [11], a machine-learning model based on support vector regression (SVR) is proposed for localisation of stationary objects using off-the-shelf equipments. This model learns RSSI fingerprints during an offline phase and then predicts 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. In a most recent work [12], a hybrid approach is presented to locate objects on bi-dimensional scenarios. The localisation error was between 9 and 29 cm in the range of 1 and 2.2 m scenarios.

In [13], active reference tags are attached to the floor, while an RFID reader is carried by each person present in the environment. In the setup phase, an RSSI fingerprint table is created. In the online phase, the RSSI sensed by the reader from each reference tag is related to the fingerprint table to obtain the estimated region of the person. In the visual location system, the background subtraction method is applied. Finally, the fusion system matches RFID and visual locations according to their sequence in the coordinate axis, defining the visual position as a final result. Average accuracy was between 95 and 100% for 1–2 m range. Despite a good accuracy, the system was limited only to persons' localisation.

Nick *et al.* [14] present a tracking system for trolleys carrying boxes leaving or coming into a mail distribution centre. Using an 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 recognise the object in the image. Sensor data fusion is performed by a constrained unscented Kalman filter technique. Localisation errors were 26 and 36 cm for stationary and moving scenarios, respectively.

In [3], a new RFID-based hardware called wireless identification and sensing platform (WISP) is attached to the target object. The WISP activates a light-emitting diode (LED) each time a passive RFID tag is written. When the LED illuminates, it is recognised by an optical system. The optical system collects images in pairs (one LED on and one off), performs the difference computation on board, and locates the maximum brightness change. The system was able to locate objects between 1.5 and 2 m under 3D scenarios. Results showed high accuracy, locating objects with errors between 1 and 2 cm.

Deyle *et al.* [15] present a multi-sensor IPS based on RFID, optical and laser technologies. The system considers the output of three approximately coincident sensors with overlapping fields of view: the RSSI image, a low-resolution camera image from a rectified camera and a 3D point cloud from a tilting laser range finder. From the camera image, colour histograms are employed as the visual feature. The probability that a tag is at a given location uses a Bayesian inference model. After all sensor images are fused, the maximum likelihood pixel is selected, and the corresponding 3D location from the laser is chosen. The system correctly located the target object in 17 out of 18 trials (94.4%).

3 System overview

An IPS has three main elements: (i) there are location sensing devices which provide metrics regarding the relative position of an object; (ii) there is a localisation technique, or an algorithm, responsible for processing those metrics; and (iii) there is a display system for graphical illustrations of the target object's location [1]. Fig. 1 shows such elements might be associated with a three-tier architecture [2]. In this work, for the location sensing layer, RFID and digital camera were used; technologies regarding the location technique were machine-learning models and CV algorithms; for the graphical interface, .NET graphical user-interface (GUI) application was used.

In this work, every pair of RFID tag and the visual marker is referred to just as a marker. Markers are attached to objects and they identify each item. Then, the system is able to estimate specific locations in the scenario.

The proposed IPS is divided into two subsystems. One uses machine-learning models and RFID technology to predict location. The other enhances these estimated predictions by employing CV algorithms and a camera. 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. This process is further described in Fig. 2.

In the offline phase, the RFID equipment and the camera are calibrated, reference tags and visual markers are positioned, and supervised machine-learning models are trained. Two machine-learning models are proposed and compared. One model is based on ANNs and the other on SVR. The role of the RFID subsystem is to collect RSSI fingerprints from the environment to train the machine to learn models. In this phase, visual subsystem processes only capture scenario images in order to set reference marker positions.

In the online phase, the RFID subsystem collects fingerprints to use as the input of the models created in the offline phase. The k-means approach is proposed to group preliminary locations in weighted clusters. Finally, the visual subsystem refines the results through visual markers detection. Both the subsystems and phases are detailed in the next sections.

The RFID reader and camera are connected to a computer running the system. MATLAB libraries [16, 17] have been integrated into .NET C# to train and run machine-learning models. A .NET GUI application was developed to automate experiments and graphically show results.

4 Offline phase

The proposed RFID subsystem is based on the RSSI value of each tag to estimate its location in the scenario. Data collection from reference tags is needed in order to provide a statistical model –

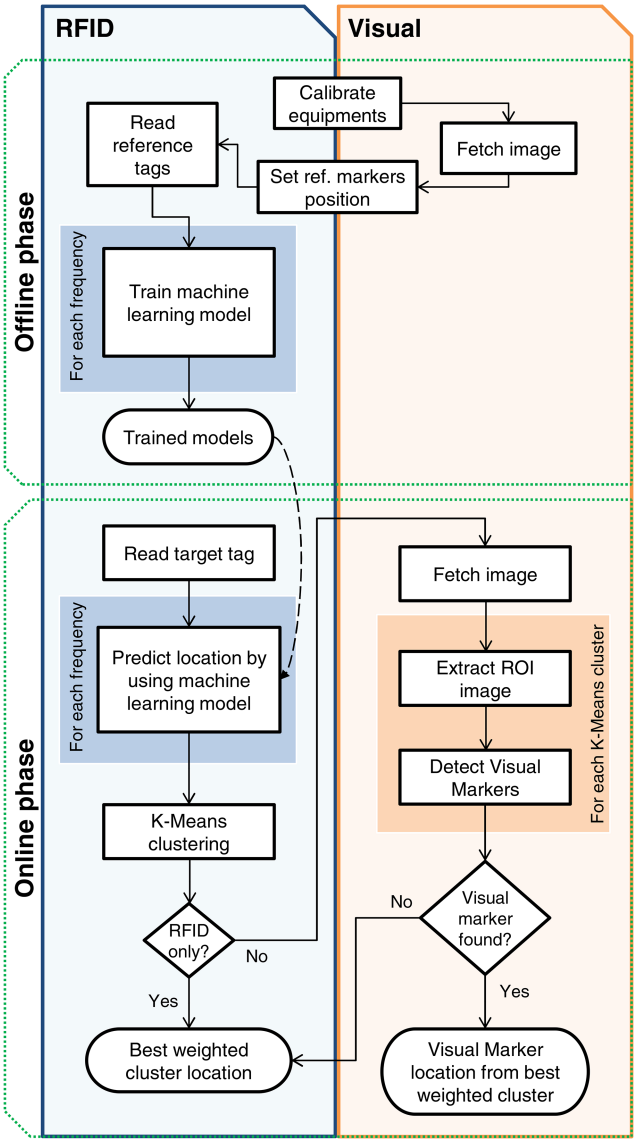

Fig. 2 *Block diagram of system design. The inputs are the RFID readings and the camera picture. The output is the target object position*

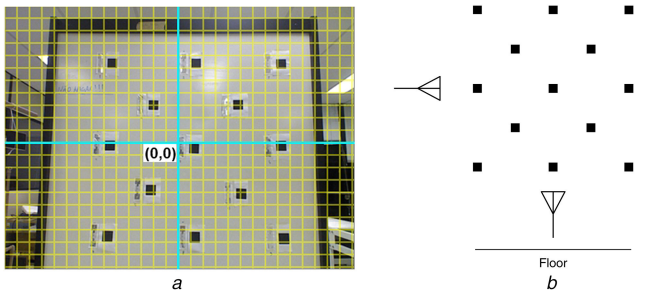

Fig. 3 *Virtual grid over a picture captured from the training scenario (a) and positioning of reference tags and antennas over the diagonal mesh design (b)*
(a) Virtual grid, ***(b)*** RFID components position

therefore, such tags must be evenly distributed. During the experiment test bed, diagonal mesh design (Fig. 3) performed better, as opposed to simple grid format, being consequently chosen for the rest of the project.

Some initial adjustment is required. After reference tag positions are stored in the system, spatial coordinates (x, y) are translated to a virtual grid, which is created and defined over an image of the scenario, as seen in (Fig. 3*a*). Thus, the system is able to show any cell within the borders of the captured picture.

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 system (x-axis and y-axis). This arrangement can be seen in Fig. 3*b*. In a 3D scenario, the infrastructure needs an additional antenna. The third antenna is placed in front of the target objects, providing signal data to predict the z-axis distance.

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 megahertz, the RSSI and the position (x, y, z) of the reference tags present in the scenario.

4.1 ANN model

ANNs learn the non-linear mapping between inputs and outputs through non-linear activation functions and hidden neurons. ANNs are effective for localisation problems because they act as universal interpolators. One of their main characteristics is that no prior knowledge about environment geometry (position of rooms, walls and obstacles) is needed. Interference factors such as multipath propagation of RF signals and the presence of other electronic devices are all embedded in the training samples collected onsite. Knowledge about the propagation channel and reader antenna positions are also not necessary [18, 19].

Data collected in the offline phase feed the ANN training process. All collected data are used, and any data removal or aggregation are performed at this stage. Data are separated by operation frequency, and a neural network is created for each frequency. RSSI values for each antenna are presented as network inputs, and the virtual-grid coordinates (x, y) of each reference tag are the target output data. In a 3D scenario, the depth distance (z) is also presented as output.

Reference tag data are divided into three subsets: training, validation and testing, randomly divided into the ratios 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 dataset.

A feedforward BPN is modelled. It consists of four layers and has n neurons in the input layer, 24 neurons and 12 neurons in each hidden layer and 2 neurons in the output layer. Neurons in the input layer and antennas in the scenario must appear in the same amount. As for the hidden layers, the numbers are such because performance does not improve above or below this point – and computational time would increase substantially, as well. Weight and bias values were upgraded with the Levenberg–Marquardt backpropagation algorithm. Performance of the network was measured by the mean square error (MSE).

4.2 SVR model

Support vector machines (SVMs) have originally been proposed as a supervised learning algorithm for binary classification. In a modified formulation they can also be applied to regression tasks, which are then referred to as SVR [20, 21]. The localisation problem of this work is a regression problem instead of a classification problem. As stated in Section 4, the target marker 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

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

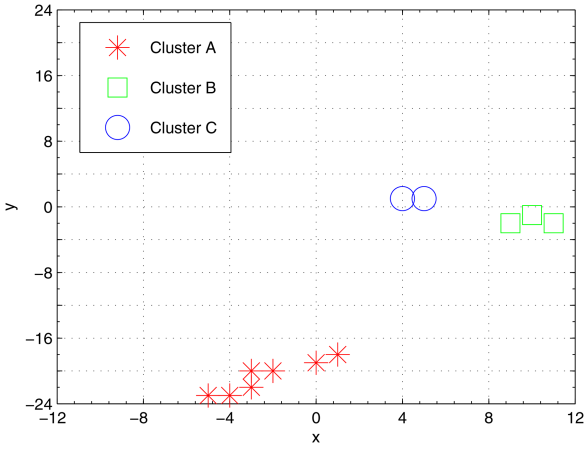

Fig. 4 *k-Means clustering applied to predicted locations over virtual-grid coordinates* (x, y)

where $\langle .,. \rangle$ 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 optimisation given by

$$\text{minimise} \quad \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \varepsilon + \xi \\ \langle w, x_i \rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{cases}$$

where C is a regularisation parameter that controls the trade-off between penalising 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 optimisation problem is usually found by means of an equivalent dual formulation.

The dual formulation of the SVR problem provides an alternative to work 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 [16] of SVR with a wavelet kernel [22]

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

where x, z and a are the wavelet dilation and translation coefficients. More details and concepts about SVR can be found in Cristianini and Shawe-Taylor [23] and Smola and Schölkopf [21].

SVR is modelled 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, y and z (3D environment). We cross-validated values for SVR coefficients, and based on the results, they were set as $\varepsilon = 0.00025$, $c = 40{,}000$ and $a = 4$ (wavelet).

5 Online phase

In this phase, a sensor fusion approach is proposed. By integrating several positioning systems, accuracy tends to improve, once hierarchical and overlapping levels of resolution are formed [24]. In this phase, the RFID subsystem seeks to detect regions of interest (ROIs). These are limited areas, smaller than the size of the whole environment. The trained models and the k-means method are used by the RFID subsystem to estimate target object positions. Later, the visual subsystem uses ROIs to predict more precise locations. Other techniques can be applied to this limited area, so localisation performance is usually higher.

During the online phase, no reference tags are necessary for the environment, but an unknown RFID tag is read during a fixed period of time. When the network is well trained, data from an unknown tag are used to estimate its location. An ANN model is automatically chosen for each frequency. RSSI data from the antennas work as input to the network, and virtual-grid coordinates are estimated and shown as the network output.

In both SVR approach and ANN model, online procedures are quite similar. The trained SVR model receives RSSI values from an unknown tag, thus estimating tag location. Output coordinates have, each, their own SVR model. Every model is evaluated to predict coordinates x, y and z (3D scenario).

5.1 k-Means

The RFID reader, during the two phases of the system, gathers a large number of readings for every tag, and it does so in a short time. For instance, for 3 s sensing a tag, there are 46 readings available. If obstacles or interferences disturb the training process, the performance of a given frequency model might be low. Hence, RSSI values can lead to different estimated positions for the same tag.

Thus, some technique is required to provide the final location of the target object. Initial tests show a simple mean to predict the location of target objects might not provide a precise result. Therefore, the k-means technique is required to achieve our goals.

In our model, 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. 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.

On the basis of member count information, we can also define a good weighted cluster has more members, and, thus, assembles more and closest predicted positions than a bad weighted one. In fact, the best-weighted cluster presents more similar locations. On the other hand, false positions are common in clusters with few members, in which machine-learning models show a poor performance.

Fig. 4 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.

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

5.2 CV for fine localisation

Our proposed sensor fusion is based on RFID estimates and CV recognition in order to filter and improve the results the RFID subsystem obtains. With a multiple ROI approach, images from the scenario are analysed by CV; as the RFID subsystem estimated more than one location, CV method explores multiple regions in order to find a visual marker.

In this work, visual markers do not need to be uniquely identified. In other words, the visual markers do not need to be machine-readable and store data. This feature can facilitate the adoption of new visual markers and CV algorithms. The visual marker should always be placed as close as possible to RFID tag. Each side of the visual marker is 4.5 cm long.

Algorithm 1 (see Fig. 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 centre 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 photograph.

To detect the square shape in the sub-image [Algorithm 1 (Fig. 5), line 4], a canny edge detector is employed. The threshold

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 $subImg \leftarrow \texttt{CropImage}(img, cluster^p)$;
4 $S \leftarrow \texttt{DetectShapesPos}(subImg)$;
5 **foreach** *shapePosition in S* **do**
6 $M_i^p \leftarrow \texttt{GetImagePos}(shapePosition)$;
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**
12 **end**
13 **end**
14 **if** $M \neq \emptyset$ **then** // Visual markers found
15 $B \leftarrow \{\texttt{indexes}(M^w) \mid M^w = bestWeight\}$;
16 $F \leftarrow \{M_i^p \mid i \in B\}$;
17 **return** $\left(\frac{1}{n}\sum_n^{i=1} F_i^x, \frac{1}{n}\sum_n^{i=1} F_i^y\right)$;
18 **else** // No markers found (RFID-only)
19 **return** C_i^p, *where* $i = \texttt{index}(\max(C^w))$;
20 **end**

Fig. 5 *Algorithm 1: GetVisualMarkerLocation(img, C)*

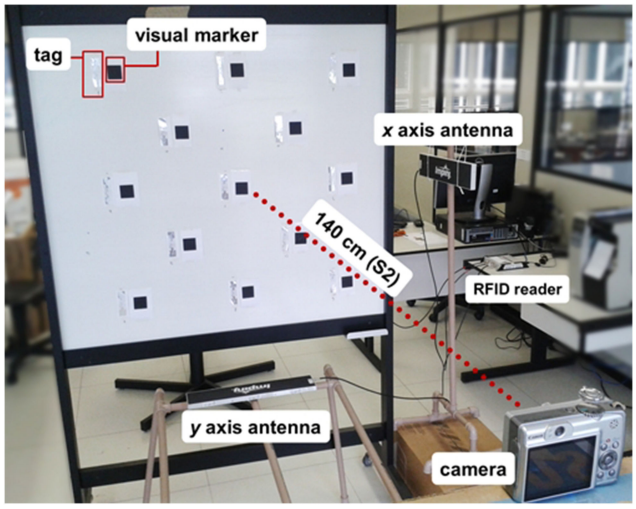

Fig. 6 *Test bed environment and all system components*

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° and 100° are selected. If square-shaped area is bigger than a configurable minimum size, it is recognised as a visual marker. We used EmguCV (.NET wrapper to OpenCV) [25] to implement this subsystem.

As the visual marker is detected in the sub-image and we want to know the marker location in the scenario picture [Algorithm 1 (Fig. 5), line 6], the following equation must be used:

$$(x, y) = \left(x_m + x_{\text{img}} - \frac{w}{2}, y_m + y_{\text{img}} - \frac{h}{2}\right) \quad (3)$$

where $(x_{\text{img}}, y_{\text{img}})$ are the centre coordinates of the sub-image in the original picture, (x_m, y_m) are the visual marker positions in sub-image and w and h are the sub-image width and height, respectively.

The best-weighted cluster position provides a visual marker, and its position gives the final target location [Algorithm 1 (Fig. 5), lines 15–16]. It might happen this cluster has more than one visual marker, so the system calculates the simple centroid of all finite points. If there are no visual markers, location is provided only by RFID.

6 Experiments and results

In the experiments, the localisation system was run in a laboratory (10 m × 7 m), where markers (tag and visual) were attached on a whiteboard, which is 1.5 m in width and height (2.25 m^2 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 4. 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. The threshold is a far-field antenna, which operates in a frequency range of 902–928 MHz and its coverage range is 3 m. The antenna maximum power is 30 dBm and it provides a maximum gain of 5 dBi. 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 following values: 923.25, 923.75, 924.25 and 924.75 MHz. Reader power was set to a maximum value of 32.5 dBm. For visual localisation, an inexpensive off-the-shelf camera with 1.3 Mp (1280 × 960) 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.

6.1 2D scenario

In the 2D experiments, the system was evaluated in four scenarios: S1, S2, S3 and S4. In each scenario, the distances between camera and markers were 100, 140, 180 and 220 cm, respectively. RFID reader antennas were placed under and on the right-hand 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 machine-learning models was 500 on average. The training set MSE of the neural network was 1.14 cm. Fig. 6 shows the test bed environment and all system components.

During the online phase of the experiment, our aim was to find six target markers on the scenario. Three positions were already used in the previous, offline phase, while the other three markers were placed in unknown locations. RFID reader was activated by 3 s for each tag. ROI size was set to 30, 22, 17 and 15% for scenarios S1, S2, S3 and S4, respectively, and the visual marker had a minimum area of 40 px.

The first validation test had no visual markers, named 'RFID-only'. In the second, 'dense' test, the camera captured an image which showed, simultaneously, all the 16 markers that were attached to the whiteboard. For the last one, called the 'clean' test, each run had only one marker present in the scenario.

Fig. 7 presents a screenshot of the system showing RFID locations (yellow squares) and final target location (green circle). The GUI helps users to easily identify the marker location. Time latency to localise each marker is <1 s, excluding RFID reader time.

The Euclidean distance between estimated and actual points gives the location error. Figs. 8*a* and *b*, respectively, show the cumulative distribution functions (CDFs) of the error distance for the ANN and for the SVR models.

CDF results show the localisation 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 the dense test, the system detected a wrong visual marker in 25% of the experiments, most of them on long distances scenarios. SVR results are also better in clean tests, while dense and RFID-only are similar. Contrary to ANN approach, the dense test has the worst performance, which means the SVR model predicted more locations close to wrong visual markers.

To summarise the localisation accuracy for each scenario, the root MSE (RMSE) of the location estimates is calculated as the difference between the predicted and actual location as

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{k}(\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2}{k}} \quad (4)$$

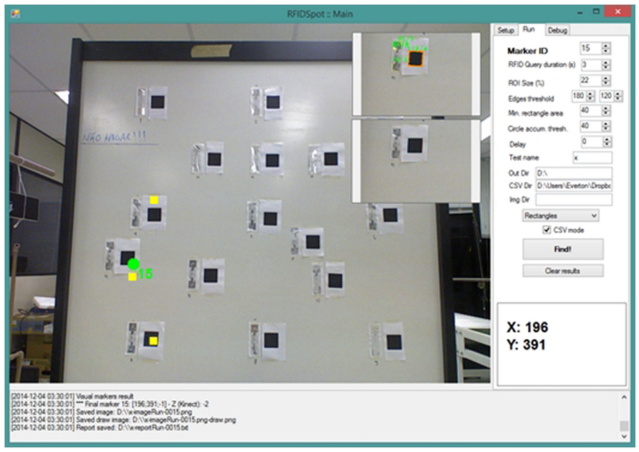

Fig. 7 *Screenshot of the system running. Scenario: S2; test: dense; machine-learning model: ANN; target marker ID: 15; and localisation error: 0 cm*

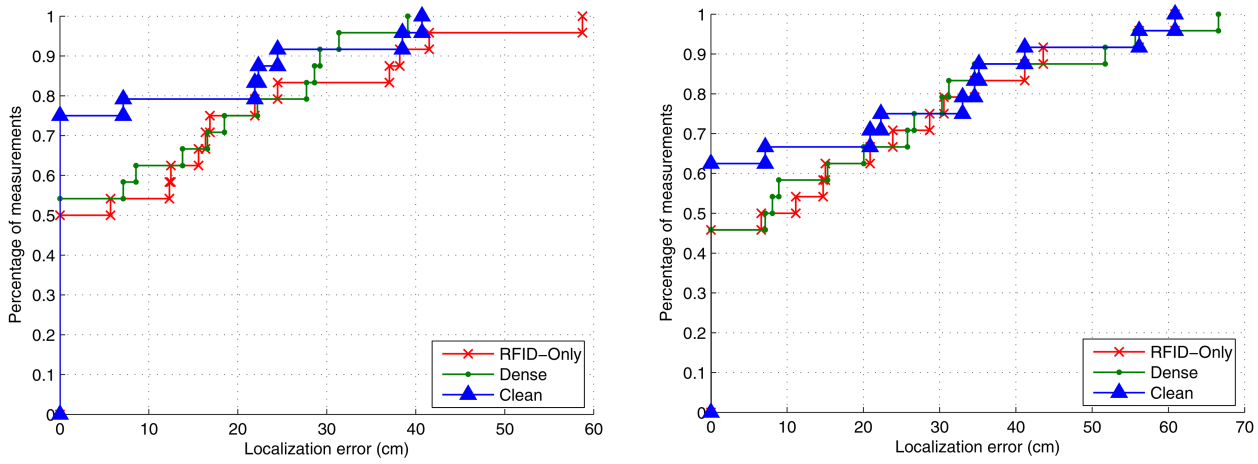

Fig. 8 *Cumulative error distance for both machine-learning approaches*
(a) CDF for ANN model localisation, ***(b)*** CDF for SVR model localisation

Table 1 Localisation performance (RMSE in centimetres) for each scenario, validation test and machine-learning approach

Scenario	ANN			SVR		
	RFID-only	Hybrid		RFID-only	Hybrid	
		Dense	Clean		Dense	Clean
*S*1	12.1	12.2	9.4	17.6	28.4	16.8
*S*2	17.3	13.6	9.1	18.2	14.1	19.7
*S*3	33.3	23.7	22.9	30.2	26.8	28.4
*S*4	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

where $\hat{x}_t$, $\hat{y}_t$ describe the estimated locations, x_t, y_t are the actual positions and k is the number of predictions. Table 1 shows the RMSE performance of the system.

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 of the ANN approach show a localisation 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 on the ANN model, the hybrid system has better results than the RFID-only approach. Localisation 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 RMSEs 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 localisation. However, the approach still has some limitations in scenarios where many items are close to each other.

In comparison to related works, the proposed hybrid system performs 40 cm better than a neural network RFID-based approach [9], where the distance between reference tags is similar to our work. Also, our IPS decreased the localisation error by 45% than other stationary hybrid system [14].

6.2 3D scenario

In the 3D scenario, an additional antenna was placed in front of the target objects, along with the digital camera. To feed the training dataset, reference tags were read at distances (d_z) 100, 140, 180 and 220 cm. During the offline phase, 4400 samples were collected to feed the machine-learning models. The neural network had a 23.6

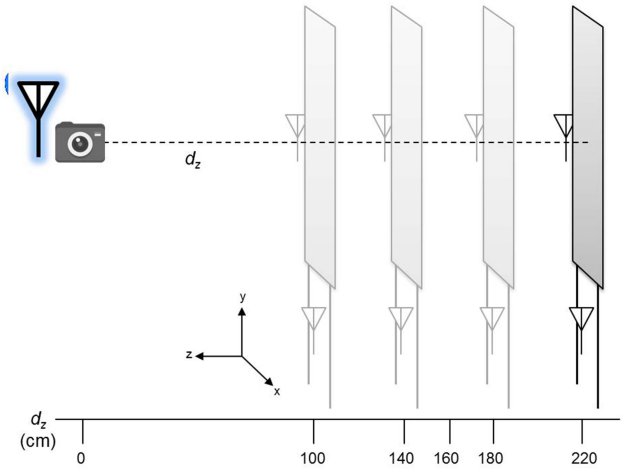

Fig. 9 *Infrastructure and antennas arrangement of the 3D scenario*

Table 2 Localisation performance (RMSE in centimetres) under 3D scalability

Distance	ANN			SVR		
	RFID-only	Hybrid		RFID-only	Hybrid	
		Dense	Clean		Dense	Clean
160 cm	79.0	77.6	83.3	83.5	81.3	86.5
220 cm	80.7	83.7	78.2	62.7	65.3	74.6
RMSE	79.8	80.7	80.8	73.9	73.7	80.8

cm MSE on the training set. Fig. 9 shows the infrastructure needed and the arrangement scheme for a 3D experiment execution.

In the online phase, the goal of the experiment was to provide 3D coordinates (x, y and z) of six target markers. Evaluated depth distances (d_z) were 160 and 220 cm, and validation tests are the same of 2D scenario experiments, i.e. RFID-only, dense and clean.

Table 2 summarises systems performances in the 3D scenario. In contrast to 2D results, SVR model is 5, 3% superior to ANN model. Results from validation tests of the hybrid system and RFID-only system are very similar. Regarding ANN model, error in the hybrid system increased 1 cm when compared with RFID-only tests.

Regarding to SVR model, the hybrid system ('clean' test) has 12 cm higher accuracy than an RFID-only test. In these cases, among the regions analysed by the visual subsystem, the one with a visual marker resulted in high-accurate values for x and y coordinates, but low-accurate predictions for the depth distance (z).

In comparison to two related systems that work with 3D scenarios [3, 15], the proposed system performance is lower in both cases, as the error increased about 80 cm. However, both related works need a new hardware and could not be deployed under an existing infrastructure. Besides that, Deyle *et al.* [15] predict the depth distance by using additional laser sensors.

7 Conclusion

Our multi-sensor system uses affordable equipment to locate stationary items with great accuracy, due to the combination of visual markers and RFID tags attached to objects. The proposed machine-learning models in this work can learn RSSI fingerprints and, thus, predict markers' position. Also, localisation is enhanced by the use of CV algorithms and a k-means technique.

Real-world experiments helped us evaluate accomplishments and compare models. In bi-dimensional scenarios, our best case demonstrated a 9.1 cm accuracy, as well as a 32% improvement over localisation systems based only on RFID. Regarding 3D scenarios, localisation errors in dense environments are 80.7 and 73.7 cm for ANN and SVR models, respectively.

In future works, experiments in larger scenarios, with multiple readers and cameras, will help us test scalability.

8 References

[1] Pahlavan, K., Makela, J.: 'Indoor geolocation science and technology', *IEEE Commun. Mag.*, 2002, **40**, (2), pp. 112–118

[2] Gu, Y., Lo, A., Niemegeers, I.: 'A survey of indoor positioning systems for wireless personal networks', *IEEE Commun. Surv. Tutor.*, 2009, **11**, (1), pp. 13–32

[3] Sample, A.P., Macomber, C., Jiang, L.T., *et al.*: 'Optical localization of passive UHF RFID tags with integrated LEDs'. 2012 IEEE Int. Conf. RFID (RFID), 2012, pp. 116–123

[4] Ni, L., Zhang, D., Souryal, M.: 'RFID-based localization and tracking technologies', *IEEE Wirel. Commun.*, 2011, **18**, (2), pp. 45–51

[5] Farid, Z., Nordin, R., Ismail, M.: 'Recent advances in wireless indoor localization techniques and system', *J. Comput. Netw. Commun.*, 2013, **2013**, pp. 1–12

[6] Belhadi, Z., Fergani, L.: 'Fingerprinting methods for RFID tag indoor localization'. 2014 Int. Conf. Multimedia Computing and Systems (ICMCS), 2014, pp. 717–722

[7] Ni, L.M., Liu, Y., Lau, Y.C., *et al.*: 'LANDMARC: indoor location sensing using active RFID', *Wirel. Netw.*, 2004, **10**, (6), pp. 701–710

[8] Chattopadhyay, A., Harish, A.R.: 'Analysis of low range indoor location tracking techniques using passive UHF RFID tags'. 2008 IEEE Radio and Wireless Symp., 2008, pp. 351–354

[9] Kung, H., Chaisit, S., Phuong, N.T.M.: 'Optimization of an RFID location identification scheme based on the neural network', *Int. J. Commun. Syst.*, 2015, **28**, (4), p. 20, doi: 10.1002/dac.2692

[10] Goller, M., Feichtenhofer, C., Pinz, A.: 'Fusing RFID and computer vision for probabilistic tag localization'. 2014 IEEE Int. Conf. RFID (IEEE RFID), 2014, pp. 89–96

[11] Berz, E.L., Tesch, D.A., Hessel, F.P.: 'RFID indoor localization based on support vector regression and k-means'. 2015 IEEE 24th Int. Symp. Industrial Electronics (ISIE), 2015, pp. 1418–1423

[12] Berz, E.L., Tesch, D.A., Hessel, F.P.: 'A hybrid RFID and CV system for item-level localization of stationary objects'. 2017 18th Int. Symp. Quality Electronic Design (ISQED), 2017, pp. 331–336

[13] Wang, C., Cheng, L.: 'RFID & vision based indoor positioning and identification system'. 2011 IEEE Third Int. Conf. Communication Software and Networks, 2011, pp. 506–510

[14] Nick, T., Cordes, S., Gotze, J., *et al.*: 'Camera-assisted localization of passive RFID labels'. 2012 Int. Conf. Indoor Positioning and Indoor Navigation (IPIN), 2012, pp. 1–8

[15] Deyle, T., Nguyen, H., Reynolds, M., *et al.*: 'RF vision: RFID receive signal strength indicator (RSSI) images for sensor fusion and mobile manipulation'. 2009 IEEE/RSJ Int. Conf. Intelligent Robots and Systems, 2009, pp. 5553–5560

[16] Clark, R.: 'A MATLAB implementation of support vector regression (SVR)', http://www.mathworks.com/matlabcentral/fileexchange/43429-supportvector-regression, accessed November 2014

[17] The MathWorks Inc.: 'MATLAB neural network toolbox', https://www.mathworks.com, accessed December 2014

[18] Ng, W.W.Y., Qiao, Y., Lin, L., *et al.*: 'Intelligent book positioning for library using RFID and book spine matching'. 2011 Int. Conf. Machine Learning and Cybernetics, 2011, pp. 465–470

[19] Martínez-Sala, A., Guzmán-Quirós, R., Egea-López, E.: 'Active RFID reader clustering and neural networks for indoor positioning'. The third Int. EURASIP Workshop on RFID Technology, 2010
[20] Wille, A., Broll, M., Winter, S.: 'Phase difference based RFID navigation for medical applications'. 2011 IEEE Int. Conf. RFID, 2011, pp. 98–105
[21] Smola, A.J., Schölkopf, B.: 'A tutorial on support vector regression', 1998
[22] Zhang, L., Zhou, W., Jiao, L.: 'Wavelet support vector machine', *IEEE Trans. Syst. Man Cybern. B, Cybern.*, 2004, **34**, (1), pp. 34–39
[23] Cristianini, N., Shawe-Taylor, J.: '*An introduction to support vector machines and other kernel-based learning methods*' (Cambridge University Press, New York, NY, USA, 2000)
[24] Hightower, J., Borriello, G.: 'Location systems for ubiquitous computing', *Computer*, 2001, **34**, (8), pp. 57–66
[25] http://www.emgu.com, accessed December 2014