Fusing RFID and Computer Vision for Probabilistic Tag Localization

Michael Goller
Enso Detego GmbH
Graz, Austria
Email: michael.goller@enso-detego.com

Christoph Feichtenhofer
Institute of Electrical Measurement
and Measurement Signal Processing
Graz University of Technology, Austria
Email: feichtenhofer@tugraz.at

Axel Pinz
Institute of Electrical Measurement
and Measurement Signal Processing
Graz University of Technology, Austria
Email: axel.pinz@tugraz.at

***Abstract*—The combination of RFID and computer vision systems is an effective approach to mitigate the limited tag localization capabilities of current RFID deployments. In this paper, we present a hybrid RFID and computer vision system for localization and tracking of RFID tags. The proposed system combines the information from the two complementary sensor modalities in a probabilistic manner and provides a high degree of flexibility. In addition, we introduce a robust data association method which is crucial for the application in practical scenarios. To demonstrate the performance of the proposed system, we conduct a series of experiments in an article surveillance setup. This is a frequent application for RFID systems in retail where previous approaches solely based on RFID localization have difficulties due to false alarms triggered by stationary tags. Our evaluation shows that the fusion of RFID and computer vision provides robustness to false positive observations and allows for a reliable system operation.**

I. Introduction

Radio Frequency Identification is considered as one key enabling technology for fully transparent supply chains from the manufacturing stage to the end customer. Under optimal conditions, RFID systems provide a reliable identification of tagged products in various scenarios, such as distribution centers or retail stores. The wireless transmission channel is characterized by multipath propagation which leads to undefined, environment dependent antenna radiation patterns. This fact is limiting the applicability of RFID systems in certain applications, such as EAS (Electronic Article Surveillance). Commercially available EAS systems suffer from false alarms caused by tags which are unintentionally identified. The typical approach to overcome this problem involves specialized antenna designs with an extremely narrow radiation pattern to minimize the percentage of unwanted reads. Alternatively, tag localization schemes are employed to determine the tag position with respect to a given reference frame [1]–[3]. However, tag localization in passive RFID systems is a challenging task due to the limited bandwidth and the inherent multipath channel characteristic [4].

To overcome the limitations in practical scenarios, this paper presents a localization approach based on the fusion of RFID and computer vision (CV) systems. For this purpose, we combine the information from a monocular camera with a scalable RFID system to estimate the location and movement of tags in a scene. We provide an in-depth description of the proposed localization scheme and conduct a series of experiments to evaluate the performance in a practically relevant EAS scenario.

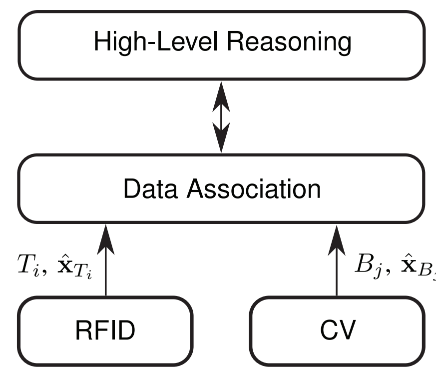


Fig. 1. Block diagram for a combined RFID and CV localization system. The RFID and CV subsystems detect and localize tags T_i and motion blobs B_j, respectively. The data association layer subsequently finds the most probable match between identified tags and detected blobs in the scene. Based on the combined information, a high-level reasoning scheme can be employed to provide abstract information rather than low-level trajectory data.

II. Related Work

CV systems have progressed to a technological state where they allow for a reliable tracking of individual objects or people from image sequences at low costs. Several authors have investigated on the combination of RFID and CV systems for localization and tracking in different applications. For example, Germa *et al.* [5] have developed a fusion system consisting of an RFID reader and a camera on a mobile robot platform. In particular, the authors combine the sensor information from the CV and the RFID system using a Particle Filter framework to track and follow individual people in a scene. A similar system [6] uses a fixed camera and a moving RFID reader. Nick *et al.* [7] have investigated on the localization of passive UHF tags by a combined RFID and CV system. For this purpose, the authors employ a deterministically found RSSI (Received Signal Strength Indicator) model and a template based visual object detection to identify and track tagged items in a warehouse portal scenario. The three discussed studies show a reasonable accuracy for a very

978-1-4799-3587-1/14/$31.00 ©2014 IEEE

specific setup but do not include the case of multiple objects or people and the related data association problem [8], which limits the applicability to practical scenarios. The apprach presented by Dibitonto *et al.* [9] uses an ultra-wideband (UWB) RFID system and a CV system to monitor people by matching pairs of RFID and CV trajectories for anomaly detection. Although this system employs UWB localization which potentially allows for a higher accuracy, the presented results indicate a suboptimal performance. Wu *et al.* [10] use a dynamic Bayes network to fuse RFID, CV and common-sense knowledge for activity recognition. While RFID readings are helpful to learn the object models without the use of manual annotation, it is shown that RFID does not provide an enhancement in terms of object and activity recognition. Furthermore, the localization is limited due to the binary response of the used RFID system. The localization approach presented by Sample *et al.* [11] provides high accuracy due to a precise optical localization, however it requires a line-of-sight to the individual tags which cannot be guaranteed in our targeted application. In contrast, our approach includes a general vision-based tracker and a robust data association which allows for an operation in realistic, potentially non line-of-sight environments with EPCglobal compliant off-the-shelf RFID hardware.

III. Approach

To tackle the problem of a multi-target tracking scenario, a combined CV and RFID system requires four major building blocks as shown in Fig. 1. First, the RFID system needs to provide at least a rough location estimate of individual tags T_i in a region of interest. Second, the CV system needs to provide an estimate for the current location and trajectory of moving objects in the scene. For this purpose, a mechanism to detect and track moving objects (referred to as motion blobs B_j) is required. Third, a way to combine the individual location estimates needs to be found, which can be formulated in terms of a data association problem. Finally, the combined information can be interpreted in terms of a high-level reasoning scheme such as a trajectory classification block to provide abstract information rather than low-level trajectory data.

The discussed system architecture provides a flexible way to combine the two complementary sensor modalities in a probabilistic framework. This forms the basis for an elegant fusion approach and furthermore allows for the integration of prior information, for example in terms of a floor plan.

A. RFID Subsystem

The task of the RFID system in this context is not only to identify RFID tags, but also to provide a location estimate $\hat{\mathbf{x}}_{T_i}$ in order to track individual tags over time. Typically, elaborate localization schemes require a high number of read events which limits their applicability to practical multi-object scenarios. For this reason, we choose a model-based location by proximity approach to provide location estimates for a given number of tags. To allow for an efficient data fusion, an appropriate formulation for the location estimate together with the associated uncertainty is required. This enables us to interpret the RFID related information in a Bayesian sense and forms the basis for a recursive update scheme when new observation data becomes available.

The idea behind the localization approach is to use several antennas $A_1 \ldots A_K$ that cover a specific region of interest. The individual antenna interrogation zones in the horizontal $x-y$ plane are modeled by means of a two dimensional Gaussian kernel g, specified in terms of the antenna position $\boldsymbol{\mu}_i$ and the covariance matrix $\boldsymbol{\Sigma}_i$. Consequently, the entire region of interest can be modeled using a Mixture of Gaussians (MoG), where each mixture component represents one particular antenna. For every RFID observation $\mathbf{z}$, the estimated tag location $\hat{\mathbf{x}}$ can be expressed using the mixture model

$$P(\hat{\mathbf{x}} \,|\, \mathbf{z}) = \sum_{i=1}^{K} w_i g(\hat{\mathbf{x}} \,|\, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{1}$$

where $w_i \propto \bar{r}_i$ are the weights of the mixture components proportional to measured RSSI values on each antenna. For a simplified scenario with $K = 3$ antennas, a qualitative RFID sensor model is shown in Fig. 2 with the individual mixture components. Since the considered tag is closest to antenna A_1, the RSSI value (and consequently, the weight of the mixture component) for this antenna is dominating over antennas A_2 and A_3. The chosen modeling approach has several implications in terms of complexity, accuracy and practical aspects. First, the model requires a parameter initialization to estimate the characteristics of the Gaussian mixture components. This parameter estimation needs to be carried out during the deployment phase in order to accurately reflect the characteristics in a given scenario. Second, the abstract representation of the interrogation zone removes the need for an accurate channel model and hence forms a compromise between accuracy and model complexity. Finally, the chosen approach is relatively robust to changes in tag orientation, since we rely on the ration of individual RSSI measurements rather than using the absolute values.

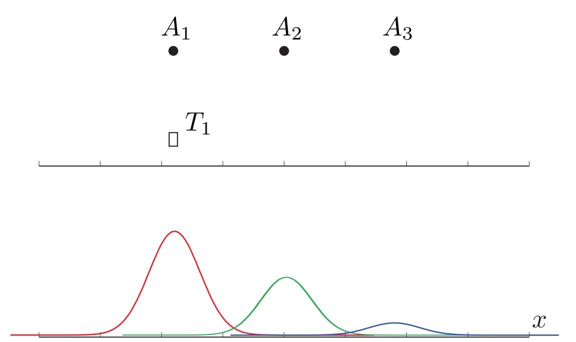


Fig. 2. RFID sensor model for a simplified 1D case with $K = 3$ antennas. The sensor model incorporates a mixture of Gaussians (MoG) to provide a likelihood for the location of tag T_1 based on the mean RSSI values. The parameters of the mixture model depend on the antenna characteristics (position and radiation pattern) and need to be learned in a calibration step.

B. Motion Blob Detection and Tracking

As complementary sensor modality, the CV system is used to monitor the region of interest in order to detect and track moving objects. For this purpose, we use a monocular camera system. For indoor scenarios, it is advantageous to mount the camera on the ceiling to provide a bird's-eye view of the scene. In addition, this allows us to align the camera field of view with the RFID interrogation zone such that we can easily establish a common reference frame. See Fig. 3 for an illustration of our experimental environment.

Fig. 3. Setup of the proposed sensor fusion system. The RFID antennas and the camera are mounted at the ceiling to cover a specific region of interest. By means of that, the two sensor modalities can be aligned to operate in a common reference frame.

The blob detection and tracking mechanism includes several processing steps that are applied to each recorded frame as shown in Fig. 4. The first block implements a foreground/background segmentation which is based on an MoG model representing the RGB (red, green, blue) values of each pixel $\mathbf{x}$. We employ the background subtraction technique first presented by Zivkovic [12], which is an improved version of the model introduced by Stauffer *et al.* [13]. The background segmentation scheme learns a background model BG by representing each pixel as an MoG. Note that this provides robustness to changes in the lighting conditions and allows us to represent multi-modal dynamic texture patterns such as blinking lights. Based on the learned model, each pixel $\mathbf{x}$ in a new frame is classified by evaluating the likelihood $p(\mathbf{x} \mid \mathrm{BG})$. The segmentation scheme implements an online update mechanism and can therefore adapt to gradual illumination changes in the scene. This is of particular importance for practical applications without strictly defined lighting conditions. In addition, a shadow detection mechanism [14] is used to suppress false positive blobs caused by moving shadows. This mechanism is based on a non-parametric approach which introduces the two additional classes *highlighted* and *shadowed* background for the classification of each pixel.

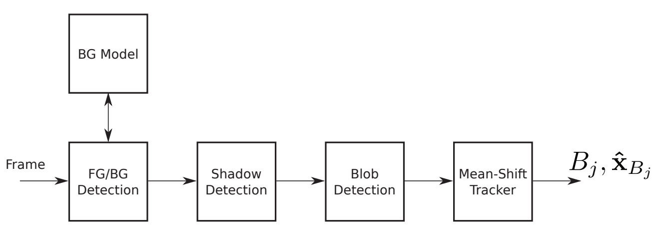


Fig. 4. Blob detection and tracking: Each frame is first segmented into a foreground (FG) and background (BG) region using an MoG model which represents the RGB values of each pixel $\mathbf{x}$. Subsequently, a shadow detection mechanism is applied to suppress false positive blobs from moving shadows in the scene. The actual blob detection incorporates several constraints regarding blob size and velocity. Each detected blob is tracked over consecutive frames using the mean-shift algorithm. Consequently, the blob detector provides an annotated trajectory for every detected blob in an image sequence.

A new blob instance B_j is detected based on the local area of a segmented foreground region. To track the individual blobs over a given image sequence, a Mean-Shift tracking algorithm is employed [15]. Each blob is represented as a histogram of the image intensity distribution which allows for a robust tracking of non-rigid objects in the scene (e.g. people). For every blob B_j, we obtain an estimate for the current position $\hat{\mathbf{x}}_{B_j}$ in image coordinates and can construct the trajectory $\hat{\mathbf{x}}_{B_j}(t)$ over time.

For the exemplary lab environment in Fig. 3, the camera view with a person detected as blob is shown in Fig. 5. The visualization shows the blob center (cross), the estimated blob dimensions (ellipse) as well as the moving direction and velocity (arrow). The proposed CV system can be easily integrated in an existing RFID deployment, does not require knowledge about the intrinsic camera parameters and allows for a tracking in real-time.

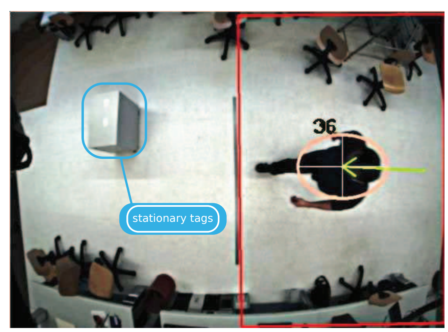


Fig. 5. Motion blob detection and tracking: Camera view of the lab environment shown in Fig. 3. In this scenario, a moving person is identified as blob with ID 36. On the left hand side, there are two stationary tags in the scene.

C. Data Association

For every processed frame, the two described subsystems provide a set of tags $\mathbf{T}$ and a set of blobs $\mathbf{B}$, incorporating the identifier and location estimates for every tag and blob, respectively. From these two sets, we want to establish an assignment between individual tags and blobs such that a subset $(\mathbf{T}_i \subset \mathbf{T})$ is assigned to a particular blob B_j. Furthermore, we need to consider the possibility of stationary tags, *i.e.*, tags that belong to the scene background rather than a particular blob. The described problem can be formulated in a data association context which considers the spatial distance $d_{i,j} = \sqrt{x_{i,j}^2 + y_{i,j}^2}$ between tag T_i and blob B_j, as shown for one tag and two blobs in Fig. 6. The spatial distance can

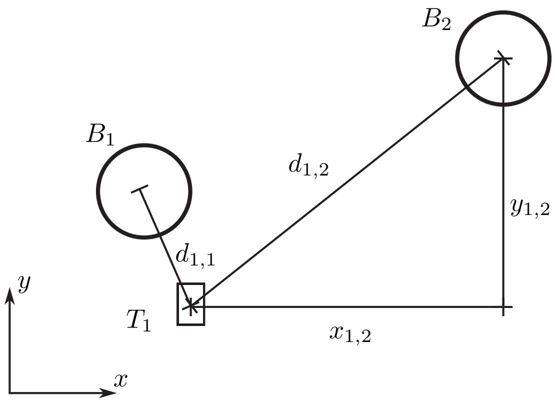


Fig. 6. Data association problem for a tag T_1 and two motion blobs B_1, B_2. The goal is to find the most likely assignment between the tag and the motion blobs B_j. For this purpose, the spatial distance $d_{i,j}$ is estimated and transformed to a probability measure $p_{i,j}$ by means of a Gaussian kernel. In addition, we account for the dynamic of tag-blob assignments by incorporating the association history. Stationary tags are considered using a velocity based background model.

be transformed into a probability measure

$$p_{i,j} = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(\frac{-x_{i,j}^2}{2\sigma_x^2} + \frac{-y_{i,j}^2}{2\sigma_y^2}\right) \tag{2}$$

using a zero-mean Gaussian kernel with specific covariance

$$\mathbf{\Sigma}_d = \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix} \tag{3}$$

according to the localization uncertainty of the RFID system. In addition, we explicitly consider the possibility of stationary tags in terms of a velocity based background model. In particular, we estimate the tag velocity

$$\hat{\mathbf{v}}_i = \frac{\mathrm{d}\hat{\mathbf{x}}_{T_i}(t)}{\mathrm{d}t} \tag{4}$$

as derivative of the trajectory with respect to time. In analogy to the spatial distance, the velocity is transformed to a probability measure by means of a Gaussian kernel with zero mean and a covariance matrix $\mathbf{\Sigma}_v$. For each observed tag, we can hence build an assignment matrix

$$\mathbf{M} = \begin{array}{c} \\ T_1 \\ T_2 \\ \vdots \\ T_N \end{array} \begin{array}{c} \begin{array}{ccccc} \text{BG} & B_1 & B_2 & \cdots & B_M \end{array} \\ \begin{pmatrix} p_{1,\text{BG}} & p_{1,1} & p_{1,2} & \cdots & p_{1,M} \\ p_{2,\text{BG}} & p_{2,1} & p_{2,2} & \cdots & p_{2,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{N,\text{BG}} & p_{N,1} & p_{N,2} & \cdots & p_{N,M} \end{pmatrix} \end{array} \tag{5}$$

holding the individual probability measures for each tag $\leftrightarrow$ blob pair and the background BG. A tag T_i can then be assigned to the most likely class (motion blobs B_j or background) by finding the maximum value in each row. The data association needs to be tolerant to noisy and incomplete data such as missing RFID observations for a particular observation period. To provide the required robustness, we integrate the history of previous assignments by considering the individual blobs and the background in a discrete state-space setting

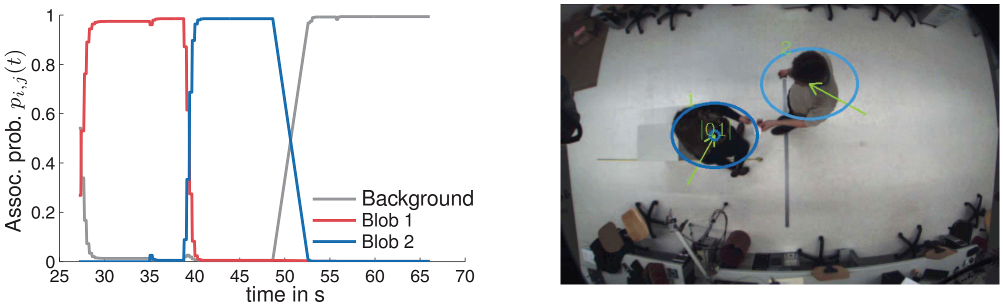
$$\mathcal{B} = \{\text{BG},\ B_1,\ B_2,\ \ldots,\ B_M\} \tag{6}$$

The idea is that the tag $\leftrightarrow$ blob assignment is rather stationary, *i.e.*, a particular tag T_i is typically connected to a blob B_j for a given observation period. Based on the defined state-space, we can hence employ a Hidden Markov Model (HMM) $\lambda_A = (\boldsymbol{\pi},\ \mathbf{A},\ \mathbf{B})$ consisting of a prior state distribution $\boldsymbol{\pi}$, a transition matrix $\mathbf{A}$, and an observation matrix $\mathbf{B}$ to filter the estimated assignments. Since there is usually no prior information available, the prior state probability vector $\boldsymbol{\pi}$ represents a uniform distribution over all blobs and the scene background. The requirements for the transition model $\mathbf{A}$ are twofold: First, it needs to consider the discussed stationary assignment characteristic by means of appropriate self-transition probabilities. Second, the transition model also needs to allow for a transition from one blob to another, for example when a tag is handed over. For this reason, a compromise between a robust assignment and the capability to follow the scene dynamics needs to be found. The observation model $\mathbf{B}$ accounts for the uncertainty regarding missing RFID observations and the limited localization accuracy.

Based on the described HMM, we can filter the estimated assignments by means of a recursive Bayes update which is implemented in the Forward algorithm [16]. For an exemplary scenario with one tag and two blobs, the filtered assignment probabilities $p_{i,j}$ over time are shown in Fig. 7(a). In this scenario, the tag is first carried by blob B_1 and then handed over to blob B_2. Finally, the tag is left stationary in the region of interest. The camera view of this hand over procedure is shown in Fig. 7(b). The association probabilities show a smooth behavior since the HMM filter introduces a low-pass characteristic due to the history of previous associations. This provides a considerable robustness for the data association process and reduces the negative impact of noisy observations.

D. Calibration

The combination of the two different sensor modalities requires a calibration procedure consisting of two main steps. First, a common reference frame for the RFID and the CV system needs to be established such that the RFID interrogation zone and the camera field of view are properly aligned. For this purpose, the antenna and camera positions $\boldsymbol{\mu}_i$ as well as the camera's field of view need to be determined. Second, the characteristics of the Gaussian mixture components, represented by the covariance matrices $\mathbf{\Sigma}_i$ need to be estimated. These characteristics directly reflect the spatial extend of the antenna interrogation zone and are the crucial parameter for the location by proximity approach.



(a) Association Probabilities

(b) Camera view

Fig. 7. Data association problem: (a) shows the filtered association probabilities over time for the scene shown in (b). The considered tag is first carried by the person detected as motion blob B_1 and then handed over to the motion blob B_2. Finally, the tag is left stationary and both blobs leave the scene. The HMM based filtering approach adds a low-pass characteristic to the association probabilities which considerably increases the robustness to noisy and incomplete observations.

To estimate the characteristics of the interrogation zone, a 2D tag grid in the region of interest is required for which RFID read events are recorded over a specific observation period. Fig. 8 shows a schematic representation of the calibration setup with the equidistant tag grid. From the read events, we obtain

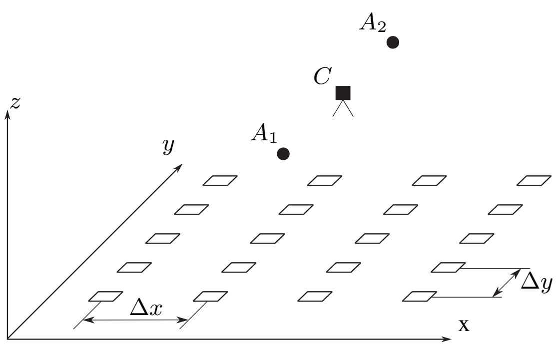


Fig. 8. Schematic representation of the calibration. A tag grid with equidistant spacing Δx and Δy is used to obtain an observation statistic for each tag-antenna pair. The observation statistic comprises the mean RSSI value and the corresponding standard deviation. The statistic is used to estimate the parameters of the Gaussian mixture model that represents the spatial extend of the interrogation zone in the $x-y$ plane.

an observation statistic for every tag-antenna pair. In particular, the statistic comprises the mean RSSI and the corresponding standard deviation for each tag and antenna. Since the K antenna positions $\boldsymbol{\mu}_i$ are known with respect to the reference frame, we need to fit a set of K Gaussian kernels to the recorded calibration data. This directly yields an estimate for the covariance matrices $\boldsymbol{\Sigma}_i, i = 1 \ldots K$. The measured interrogation zone and the corresponding Gaussian approximation for a single antenna (Kathrein Widerange with $70°$ / $30°$ half power beamwidth) are shown in Fig. 9. The measured interrogation zone exhibits local maxima and minima (dead zones) as a direct consequence of the antenna radiation pattern and the multipath channel environment. Furthermore, the interrogation zone shows a certain asymmetry due to reflections caused by a concrete wall located at $x = -1.5\,\mathrm{m}$. Fig. 10 shows the measured and modeled interrogation zone for the principal antenna axes in the x- and y- direction. The antenna radiation pattern introduces local RSSI maxima due to the prominent main-lobe and the two side-lobes in the x- direction.

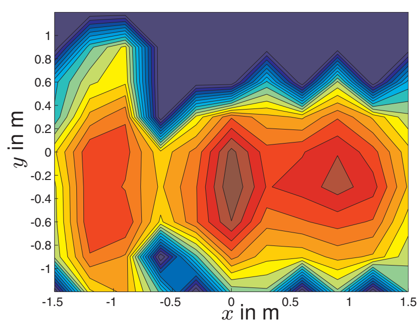
Fig. 10 indicates that the Gaussian kernel, which can be efficiently estimated from the calibration data, represents a reasonable approximation for the interrogation zone. In a setup with K antennas, the combined sensor model can then be obtained as a superposition of the individual Gaussian components. This allows for a concise formulation of the likelihood function for the tag location.

The presented information fusion concept provides a concise, yet powerful way to incorporate the two different sensor modalities. The next section presents an evaluation in an RFID driven article surveillance scenario.

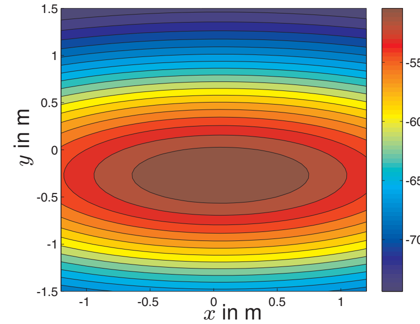
IV. Experiments

To evaluate the presented information fusion concept, we investigate on an EAS scenario which is a typical application for RFID systems in apparel retail. The functional requirement for an EAS system is to trigger an alarm as soon as an article leaves the shop without a preceding transaction at the checkout desk. The main challenge for RFID driven EAS systems is the occurrence of observations that trigger a false alarm. With the presented localization and tracking capability, the idea is to track the tag in the region of interest and trigger the alarm as soon as it crosses an imaginary line marking the shop exit. More generally, we can define an *exit region* which a moving tag is not allowed to enter. In addition to the localization and tracking task, this adds a high-level reasoning step which can be elegantly integrated in the proposed framework. In contrast to current RFID driven EAS solutions, this provides a considerable degree of robustness in practical applications and a mechanism to deal with observations from stationary tags.

From the physical point of view, not only the localization, but also the tag detection is challenging in this scenario due to the limiting forward link budget [17]. The power transferred from the reader to the tag suffers from absorption due to the water content in the human body and the angle dependent antenna gain. The evaluation is hence focused on the two key performance metrics for an RFID system, the detection and false positive probability, rather than the geometric localization uncertainty. For the detection probability, we define two

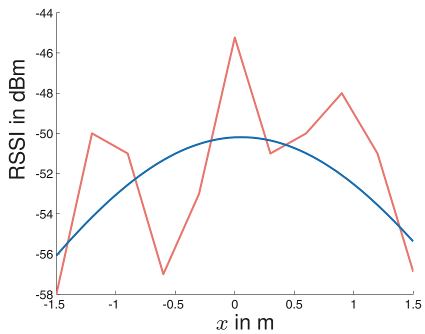


(a) Measurement

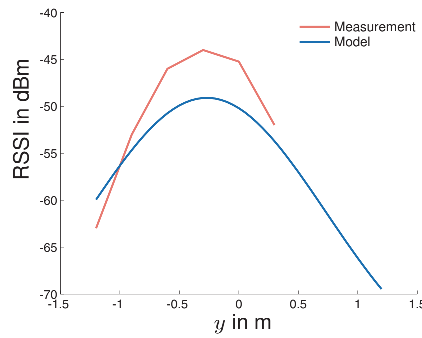


(b) Model

Fig. 9. RFID sensor model: Contour plot. (a) shows the measured interrogation zone (RSSI values) for a $\Delta x = \Delta y = 30\,\mathrm{cm}$ tag grid and (b) shows the approximation by means of a 2D Gaussian. The considered system comprises a Kathrein Widerange antenna with 70° / 30° half power beamwidth, mounted at a height of $h = 2.5\,\mathrm{m}$.



(a) x- axis



(b) y- axis

Fig. 10. RFID sensor model for a single antenna: RSSI values along the principal antenna axis for the x-direction in (a) and the y- direction in (b), together with the Gaussian approximation. Along the x- axis, the interrogation zone shows a characteristic peak and two local maxima from the side lobes of the antenna radiation pattern.

measures: On the tag-level, the detection probability P_{D} is defined as the ratio of correctly detected tags over the total number of tags that move through the scene and enter the exit region. On the test-run level, $P_{\mathrm{D,\,Run}}$ denotes the percentage of runs where at least one tag has been successfully identified as stolen. The false positive probability P_{FA} denotes the ratio of false positive observations over the number of stationary tags in the region of interest.

The evaluation is based on an RFID system comprising an Impinj Revolution R420 reader and $K = 3$ Kathrein Widerange antennas together with a monocular camera (see Fig. 3). The camera features a $1/3$" Aptina CMOS sensor, coupled with a $1.8\,\mathrm{mm}$ wide-angle lens to provide an appropriate field of view. The chosen image resolution is $752 \times 480\,\mathrm{px}$ at a framerate of $50\,\mathrm{fps}$.

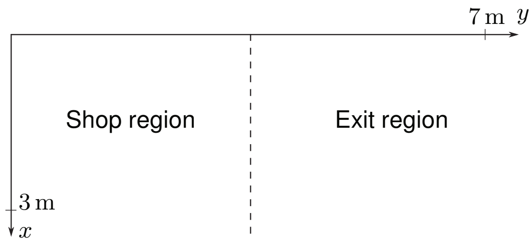
Typically, practical scenarios face certain restrictions regarding the placement of the system components due to shop design and spacing considerations. For this reason, we follow the common approach to mount the RFID antennas and the camera system on the ceiling. The floor plan of the geometric setup and the camera view of the evaluation environment are shown in Figure 11(a) and Figure 11(b), respectively. Note that the radial distortion effects at the image borders could be compensated by applying an image rectification. However, the EAS scenario only requires accurate localization at the image center and therefore we do not apply an undistortion to allow for an operation in real time.

Approach	P_{D}	P_{FA}
RFID only	0.86	0.40
RFID + CV	0.96	0

TABLE I
COMPARISON FOR RFID BASED AND HYBRID RFID + CV APPROACH.

In a preliminary experiment, we compare the performance of a purely RFID based EAS system to the proposed hybrid RFID + CV system in a simple scenario with two stationary tags and one moving tag which transitions from the shop to the exit region. For the purely RFID based approach, we employ a simple threshold based detection which identifies tags moving into the exit region by means of their RSSI response.

Table I compares the resulting performance for the purely RFID based approach with the hybrid RFID and CV system over 50 test runs. For the threshold based RFID system, we have chosen the optimal threshold value in terms of the Receiver Operating Characteristic (ROC). The results indicate that the purely RFID based approach does not provide satisfactory results in practical applications due to the noisy RFID readings which make it difficult to identify the transition



(a) EAS scenario

(b) Camera view with highlighted exit region

Fig. 11. EAS scenario: (a) floor plan. The considered scenario features a $3 \times 7\,\mathrm{m}$ region of interest, split into a *shop region* and an *exit region*. The task of the EAS system is to trigger an alarm as soon as a tag enters the exit region without a preceding transaction at the checkout desk. In addition to the localization and tracking of individual RFID tags, this requires a high-level reasoning step to decide whether a tag has crossed the imaginary line to the exit region. (b) shows the camera view with highlighted *exit region*.

Scenario	test runs	mov. tags	stat. tags	$P_{\mathrm{D,\,Run}}$	P_{D}	P_{FA}
SP-ST	150	1	0	0.9733	0.9733	-
SP-MT	150	2	0	1.000	0.9750	-
SP-MT	150	4	0	1.000	0.9125	-
SP-MT	80	2	4	1.000	0.9312	0.0188
MP-MT	80	2	4	1.000	0.8438	0.0219

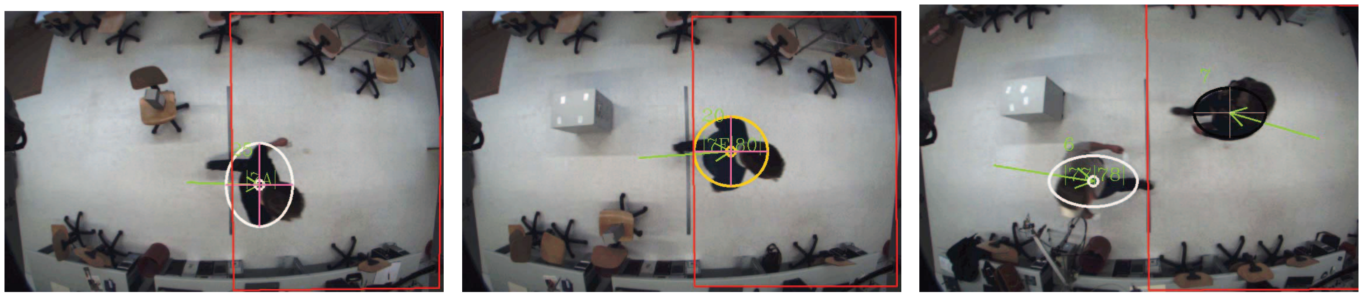
TABLE II
RESULTS FOR VARIOUS EVALUATION SCENARIOS. THE NUMBER OF TRIALS, MOVING TAGS, STATIONARY TAGS AND THE CORRESPONDING AVERAGE DETECTION RATES ARE REPORTED FOR EACH SCENARIO.

between the shop and the exit region. This agrees with the results found by other researchers in similar applications [18] showing that simple, threshold based methods are not able to perform a precise localization. Similarly, a system solely based on CV is not appropriate for the described scenario since it cannot provide information about potentially occluded objects. In contrast, the hybrid RFID + CV system reliably suppresses false positives due to the accurate localization and our robust data association. This demonstrates the advantage of a hybrid system over a purely RFID or CV based approach.

To further investigate on the performance of the combined RFID and CV system, we define different scenarios (see Fig. 12) with an increasing level of complexity. First, we investigate on the detection performance in a single person, single tag scenario which gives insight to the combined detection performance of the blob detector and the RFID system. The second and third part of the evaluation cover a single person, multiple tag scenario with and without stationary tags in the region of interest. Finally, we evaluate a multi person, multi tag scenario with stationary tags. In each scenario, the tags are carried through the region of interest along different, randomly displaced trajectories. The stationary tags are placed in the region of interest such that they are continuously visible to the RFID system. The results of the conducted evaluation are summarized in Table II.

For the single person, single tag (SP-ST) scenario, we perform 150 trials to estimate the detection probability. In this scenario, all tags have been detected at least once by the RFID system, and only four tags could not be identified as stolen. This is caused by the CV system failing to identify a moving blob or the background model which erroneously considers a particular tag as stationary. For the single person, multiple tags (SP-MT) scenario, the evaluation is conducted with two and four moving tags per run, respectively. In this case, the results agree to the intuitive assumption that the detection probability depends on the scenario complexity, which in this case is determined by the overall number of tags. This is due to the fact that the number of read events per tag and unit time decreases as the number of tags increases. This issue has direct impact on the localization accuracy and hence on the overall system performance. For the considered scenario with four moving tags, the achieved detection probability of over 90% allows for an efficient system operation since the theft event is detected for all performed test runs. To further increase the scenario complexity, we add four stationary tags to the region of interest. Consequently, we can evaluate the detection and false positive probability. The resulting performance for a total of 80 test runs shown in Table II indicates that 149 tags could be identified as stolen with six false alarms caused by the stationary tags. Although the detection probability on the tag level decreases, the actual theft event is robustly detected by the proposed system.

Finally, the multi person, multi tag (MP-MT) scenario features two people in the region of interest: Whereas one person is walking around at random, the other carries two (moving) tags per run. For this scenario, the detection performance on the tag level further decreases due to the increased scenario complexity and the required data association between the RFID tags and motion blobs in the region of interest. This is especially the case when people in the scene are closely spaced. An erroneous data association introduced by the limited localization accuracy of the RFID system is hence the main cause of false positives. For this reason, we can conclude that the system performance is directly affected by the scenario complexity and limited by the RFID system rather than the CV system. Even in this complex scenario, the proposed system provides an ideal detection performance on the trial level and considerable detection performance on the tag level by simultaneously suppressing false positives.



(a) SP-ST (b) SP-MT (c) MP-MT

Fig. 12. Different scenarios for evaluation. The blobs are indicated by a number around the ellipsoidal blob region and the RFID tags are indicated by $|\cdot|$. Arrows correspond to the direction and velocity of the moving blobs. In (a), a single person, holding a single tag, enters the exit region and therefore the corresponding blob is marked by a pink cross. In (b), a single person enters the exit region with two tags, while the four tags placed on the grey box are not detected due to stationarity. Finally, in the MP-MT scenario (c), two people walk towards each other with one person holding two tags, again four stationary tags are placed on the grey box.

Since the increased number of tags also adds a certain kind of diversity, the actual theft event can be reliably detected in the defined setup. The presented information fusion approach can be directly applied to the discussed EAS scenario and shows a robust performance under realistic environmental conditions.

V. Conclusion

RFID tag localization can be considered as the key enabling technology for the successful deployment of certain applications. In state-of-the-art systems, the limited system bandwidth together with the multipath channel characteristics impose serious limitations to the achievable accuracy. To overcome these limitations, we have presented an information fusion concept as a powerful approach to combine the localization capabilities of CV with the strengths of an RFID system.

With the use of a flexible, yet compact RFID sensor model and a location by proximity approach, we can estimate the location and velocity of individual RFID tags in a region of interest. The information fusion approach forms the basis for high-level reasoning schemes which allow for a direct application to practical scenarios. The experimental evaluation in a practical EAS scenario showed that the presented approach provides a robust system performance with respect to the detection of stolen articles and the suppression of false positives. The performance depends on the scenario complexity and typically decreases with the total number of tags and visible people in the scene. For the considered EAS scenario, the information fusion approach is superior to currently established systems which only try to minimize the number of false positive observations by means of specialized antenna designs and therefore cannot deal with stationary tags in the interrogation zone.

In order to further improve the performance of the proposed system, future research will target the integration of additional prior information. In particular, we envision a unified CV and RFID background model which can be learned autonomously in an online manner. In addition, future work will address a general calibration scheme to allow for a more flexible camera and antenna positioning which is an important aspect for practical deployments.

References

[1] P. V. Nikitin, R. Martinez, S. Ramamurthy, H. Leland, G. Spiess, and K. Rao, "Phase based spatial identification of UHF RFID tags," in *Proc. IEEE RFID*, 2010, pp. 102–109.

[2] R. Miesen, R. Ebelt, F. Kirsch, T. Schafer, G. Li, H. Wang, and M. Vossiek, "Where is the tag?" *Microwave Magazine*, vol. 12, no. 7, pp. S49–S63, 2011.

[3] S. Shao and R. J. Burkholder, "Passive UHF RFID tag localization using reader antenna spatial diversity," in *Proc. IEEE ICWITS*, 2012.

[4] G. Li, D. Arnitz, R. Ebelt, U. Mühlmann, K. Witrisal, and M. Vossiek, "Bandwidth dependence of CW ranging to UHF RFID tags in severe multipath environments," in *Proc. IEEE RFID*, 2011, pp. 19–25.

[5] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat, "Vision and RFID data fusion for tracking people in crowds by a mobile robot," *Computer Vision and Image Understanding*, vol. 114, no. 6, pp. 641–651, 2010.

[6] Z. Babic, M. Ljubojevic, and V. Risojevic, "Indoor RFID localization improved by motion segmentation," in *Proc. IEEE ISPA*, 2011, pp. 271–276.

[7] T. Nick, S. Cordes, J. Götze, and W. John, "Camera-assisted localization of passive RFID labels," in *Proc. IEEE IPIN*, 2012, pp. 1–8.

[8] C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects," *Tran. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 560–576, 2001.

[9] M. Dibitonto, A. Buonaiuto, G. L. Marcialis, D. Muntoni, C. M. Medaglia, and F. Roli, "Fusion of radio and video localization for people tracking," in *Ambient Intelligence*. Springer, 2011, pp. 258–263.

[10] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, "A scalable approach to activity recognition based on object use," in *Proc. IEEE ICCV*, 2007, pp. 1–8.

[11] A. P. Sample, C. Macomber, L.-T. Jiang, and J. R. Smith, "Optical localization of passive UHF RFID tags with integrated leds," in *Proc. IEEE RFID*. IEEE, 2012, pp. 116–123.

[12] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. IEEE ICPR*, vol. 2, 2004, pp. 28–31.

[13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE CVPR*, vol. 2, 1999.

[14] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," *Tran. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.

[15] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE CVPR*, vol. 2, 2000, pp. 142–149.

[16] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[17] A. Lázaro, D. Girbau, and D. Salinas, "Radio link budgets for UHF RFID on multipath environments," *Trans. Antennas and Propagation*, vol. 57, no. 4, pp. 1241–1251, 2009.

[18] T. Keller, F. Thiesse, J. Kungl, and E. Fleisch, "Using low-level reader data to detect false-positive RFID tag reads," in *Proc. IEEE IOT*, 2010, pp. 1–8.