2016 Eighth International Conference on

Knowledge and Systems Engineering (KSE)

IEEE Catalog Number
CFP1603I-ART
ISBN 978-1-4673-8929-7

2016 Eighth International Conference on
Knowledge and Systems Engineering (KSE)

IEEE

2016 Eighth International Conference on
Knowledge and Systems Engineering (KSE)

Hanoi, Vietnam, October 6-8, 2016

Editors
Minh Le Nguyen
Le Sy Vinh
Lam Thu Bui
Van-Giang Nguyen
Yew-Soon Ong
Keiji Hirata

Celebrating 50th Anniversary of Le Quy Don Technical University

Accurate object localization using RFID and Microsoft Kinect Sensor

Thi-Son Nguyen, Thanh-Hai Tran, Hai Vu
Computer Vision Department
International Research Institute MICA
HUST - CNRS/UMI - 2954 - INP Grenoble
Email: {thi-son.nguyen, thanh-hai.tran, hai.vu}@mica.edu.vn

***Abstract*—Object localization is the first requirement for many applications such as navigation, obstacle avoidance, object grasping. In this paper, we present a new method that combines two techniques of localization: RFID (Radio Frequency Identification) based and RGB-D camera based. In our method, each RFID tag with an unique ID will be assigned to one object. Based on the RSSI (Received Signal Strength Indication) received from RFID readers, we make a coarse localization of the object. This localization result is then projected on the image captured by a Kinect sensor to limit the region of search (RoS). If the Kinect sensor provides depth in this RoS, depth distribution of the RoS will be computed and served to narrow again the RoS. Finally, object position is refined by applying a HOG-SVM detector [1] on the RoS of the RGB image. The combination of RFID and RGB-D is twofold. It avoids both false positives and negatives when using only RGB-D information. It reduces the computational time. We have evaluated our method in a real-scene with different positions of object. The combination of RFID and RGB-D helps to reduce the localization error from 1.02m to 0.16m in average compared to using solely RFID. The HOG-SVM detector applied on the RoS obtained higher precision (100%) than applied on the whole RGB image (72.86%) while keeping the same recall (98.96%). It also reduced the computational time from 1.038s per image to 0.39s.**

I. Introduction

Detecting and localization objects in environment is one of the most important requirements for many applications such as navigation, obstacle avoidance or object grasping [2], [3], [4]. Many researches have been conducted in the literature. Depending on specific applications, various types of sensors could be utilized, among them camera are the most suitable for indoor environment.

Vision based object detection have been studied more than 30 years ago. Although a lot of successful results have been obtained, methods using conventional camera usually suffer from shadow, light condition changing and occlusion. Shadow could make bigger object boundary than itself. Light condition changing and occlusion usually produce false positives or negatives. Using depth sensor helps to avoid light sensitivity. But depth sensor is limited by a measurable range.

In this paper, we propose to use RFID for coarse localization of object. The main benefit of RFID is it provides a general framework for object identification, which addresses the main shortcoming of visual sensor based systems. RFID based always determmines the existing area where the object resides. We then do the search only in this area by projecting it on RGB-D map to refine object localization. When depth is available, we compute depth distribution to narrow still the region of search. Finally, we deploy the HOG-SVM technique for object detection on the RoS of RGB image. The combination of RFID and RGB-D is twofold. It avoids both false positives and negatives when using only RGB-D information. It reduces the computational time.

The remaining of this paper is organized as follows. In section II, we present related works. In section III, we describe our proposed method. Section VI presents experimental results. Finally, we conclude and give some ideas for future works.

II. Related works

Vision based object detection and localization is an attractive research topic in computer vision and robot navigation. Many methods were proposed using conventional RGB, stereo camera or RGB-D sensors. In this section, we present briefly some object detection methods that combine RGB sensors with RFID that are nearly related to our work.

[5] presented a method that uses two RFID readers placed somewhere in the room to roughly locate objects on the floor. Each object (bed, chair, table, can) could have more than one tags depending on their size. These object locations are refined by ceil cameras. Firstly, HSV histogram of the existing area is computed and the backproject is created by replacing the value of the observed pixel of the input image by the value of the probability of that pixel given the distribution (histogram). Then, the image is thresholded and the largest connected component is selected. Experiments have been conducted with 10 various objects in a space 6mx5m. The average recognition accuracy with and without using the RFID system is 93.1% and 68.6%. The computational time reduces from 0.49s to 0.21s.

[6] presented an approach that constructs a RSSI image from panned and titled RFID antennae. This image is then smoothed by a Gaussian filter and normalized to [0,1] range. Then a procedure of registering camera sensor with RSSI image is applied. Finally a probabilistic model of fused multi modal data is built to predict the position of objects. The method has been evaluated on three objects at 18 positions in the scene. Using RSSI image increased the system performance of 11%.

978-1-4673-8929-7/16/$31.00 ©2016 IEEE

In [7], a method that combines RF and visual sensors was presented for human activity analysis or object detection. First, coarse location estimation of object is estimated with RFID tags. Then object location is refined using visual sensors. For vision based detection, an image difference technique was utilized to separate the object from background. Finally, the object is recognized using a supervised classifier (kernel PCA and EM algorithm). This method has been evaluated in kitchen environment with 16 activities and 33 kitchen objects. Using RFID additional to Visual Sensors do not improve the detection rate. Object localization error was not reported.

In [8], the authors proposed a method for multiple cattle detection and localization using RF and camera network. First, cattle are located by RFID. Then the narrow region will be verified by the nearest camera. Each cattle has one active tag made by several LEDs which provide illumination invariant color information. Then a simple red filtering is applied for detecting active tags. The indoor test-bed size is 4mx6m. The performance of object detection is 79.16% and localization performance is 71.52%. This approach will fail when active tags are confused with spot lights of the sun.

It is noticed that all presented methods show the good benefit of RFID for improving localization error and detection rate. However, conventional vision based methods always suffer from illumination. Recently, low cost RGB-D sensors providing both RGB and depth information help to overcome this limitation. To the best of our knowledge, there are no work that combines RFID with RGB-D sensors. This motivates our work on exploiting the benefits of three modalities for a more accurate object localization.

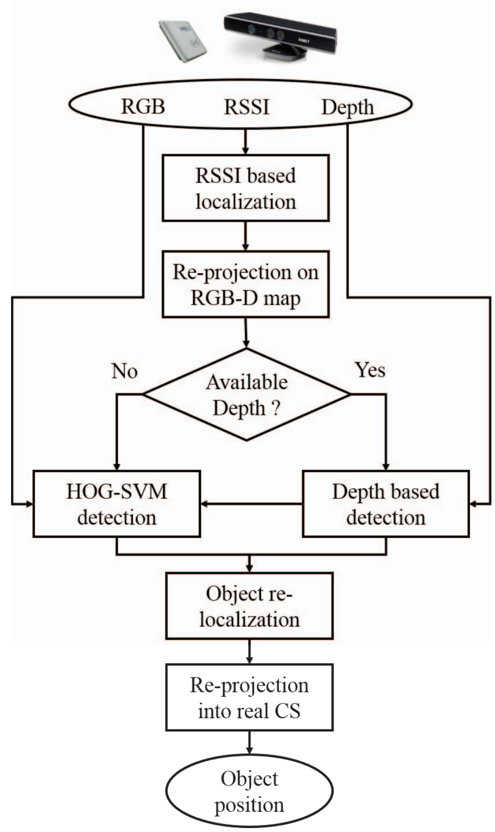

Fig. 1. Main components of the proposed localization system

III. Methodology

A. General framework

We propose a framework for object localization as illustrated in Fig.3. It consists of following main steps:

- RFID based localization: it analyzes the RSSI received from RFID readers to localize the object. The localization results is then re-projected on the RGB-D map acquired by a Kinect sensor to limit the region of search (RoS).
- Depth based localization: if depth is available in the RoS, the distribution of depth in the RoS is analyzed to narrow still the RoS.
- RGB based localization: Finally, we learn a HOG-SVM model and verify if one of slided windows on the RGB-RoS satisfies the model. The position of object will be back-projected to the world coordinate system for further applications such as assisting visually impaired people to find objects of interest.

B. System calibration and coordinate transformation

We consider three coordinate systems: the world coordinate system (WCS); the coordinate system of the Kinect sensor (KCS); and the image coordinate system (ICS) (see Fig. 2). We define the world coordinate system (WCS) with the origin O at a certain point on the floor. The plane O_{xy} fits the floor plane. The axe z upwards. The KCS has origin (0,0,0) in the sensor; its x,y and z-axis go in right, up and outward direction respectively. The ICS has (0,0) at the top-left corner, x and y go in right and down direction.

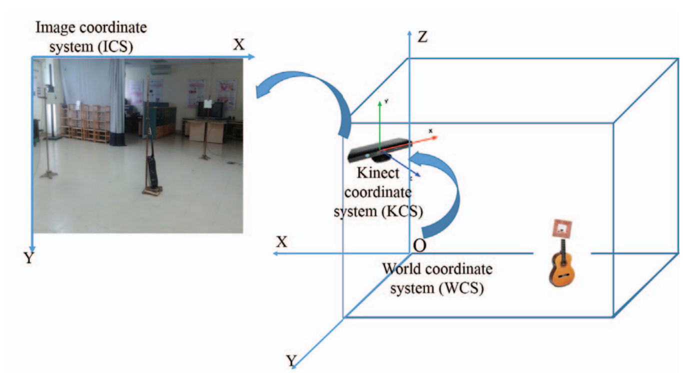

Fig. 2. Illustration of coordinate systems

RFID based localization provides the location of object in WCS: $P^W = (X^W, Y^W, Z^W)$. In order to project this point in ICS $p^I = (x^I, y^I)$ or back-project a point from ICS to WCS, intrinsic and extrinsic parameters of the Kinect sensor must be computed. We use a Matlab tool [9] for camera calibration on RGB data.

The Kinect sensor provides RGB and depth data. But these raw data are not aligned because the CMOS sensor for capturing RGB data and the IR sensor providing depth data are put at a distance. To find the depth d of the corresponding (R, G, B) at a pixel (x^I, y^I) in the RGB map, we have to calibrate depth and RGB data. In our work, we use Microsoft Kinect SDK to convert depth coordinate to color coordinate then use parameters from [10] to convert to 3D coordinates.

C. RFID based object localization

The RFID based object localization system consists of RFID antenna-readers, passive RFID tags and a PC. Antennae are mounted at fixed positions in the environment. Tags are attached on objects of interest. The antennae transmit and the readers receive feedback RSSI from tags. By analyzing the value RSSI and comparing with the RSSI values at reference points, we can determine approximately the position of the object.

1) Building the database of reference points: We make a grid composed of cells of 0.4m x 0.4m on the floor and define reference points at corners of every two cells. We put the tagged object at every reference point, compute the average of RSSI values returned during one second and store the average value in the database. If one reader cannot receive the RSSI from a tag, we set a default value -90dB. This phase is processed one time off-line. When the environment changes, it is recommended to re-measure RSSI values at reference points. At the i^{th} reference point, we store $(X_i^W, Y_i^W, Z_i^W, R_{i_1}, R_{i_2}, ..., R_{i_M})$, where (X_i^W, Y_i^W, Z_i^W) are the coordinates in the WCS of the i^{th} reference point. $Z_i^W = 0$ because we consider only points on the floor plane. R_{i_k} is the average RSSI value received from the k^{th} reader with $k \in [1, M]$, M is the number of antenna-readers. In our experiment, $M = 2$.

2) Computing the position of tag based on reference points: At runtime, we measure RSSI from every tag. For each measured $RSSI$ value, we determine approximately the position of tag as presented in Algorithm 1.

Data: RSSI received from M readers $(R_1, R_2, ..., R_M)$
RSSI at N reference points:
$S = (X_i^W, Y_i^W, Z_i^W, R_{ri_1}, R_{ri_2}, ..., R_{ri_M})$, $i \in [1, N]$.
Result: Position of the tag (X_c^W, Y_c^W)
for *i=1:N* **do**
  $diff[i] =$
  $L_1\{(R_1, R_2, ..., R_M), R_{ri_1}, R_{ri_2}, ..., R_{ri_M}\}$;
end
Sort $\{diff[i], i = [1, N]\}$ and take K smallest values.
for *j=1:K* **do**
  $w[j] = \frac{diff[j]}{\sum_{k=1}^{K}(diff[k])}$
end
$X_c^W = \frac{\sum_{k=1}^{K} w[k] X_k^W}{\sum_{k=1}^{K} X_k^W}$; $Y_c^W = \frac{\sum_{k=1}^{K} w[k] Y_k^W}{\sum_{k=1}^{K} Y_k^W}$;

Algorithm 1: Algorithm of RFID based localization

3) Projecting the estimated position on image plane: The object is approximatively localized at the position (X_c^W, Y_c^W) in the world coordinate system. Let δ be the localization error. We determine four corners of the square surrounding (X_c^W, Y_c^W) as follows $\{(X_1^W = X_c^W - \delta, Y_1^W = Y_c^W - \delta), (X_2^W = X_c^W + \delta, Y_2^W = Y_c^W - \delta), (X_3^W = X_1^W, Y_3^W = Y_c^W + \delta), (X_4^W = X_2^W, Y_4^W = Y_3^W)\}$. Then we project these points on the image using calibrated parameters of the Kinect. Suffered from perspective projection, the square becomes a quadrilateral on the image $\{(x_1^I, y_1^I), (x_2^I, y_2^I), (x_3^I, y_3^I), (x_4^I, y_4^I)\}$. We determine $x_{min} = min\{x_1^I, x_2^I, x_3^I, x_4^I\}$, $x_{max} = max\{x_1^I, x_2^I, x_3^I, x_4^I\}$, $y_{max} = max\{y_1^I, y_2^I, y_3^I, y_4^I\}$. Finally, we determine a rectangle by two top-left and bottom-right corners: $(x_{min}, y_{max} - 2 * (x_{max} - x_{min})), (x_{max}, y_{max})$. This rectangle (RoS_1) is smaller than the original image.

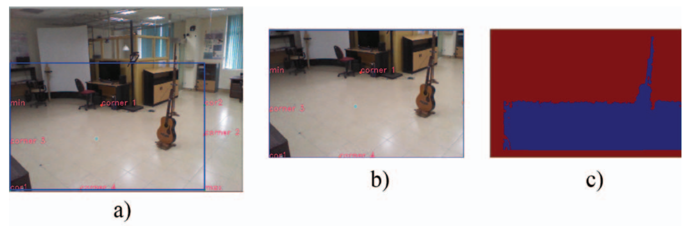

Fig. 3. a) Project the point localized by RFID and four corners on the image (cyan dot); b) cut out the minimum surrounding rectangle (RoS); c) the corresponding depth of b).

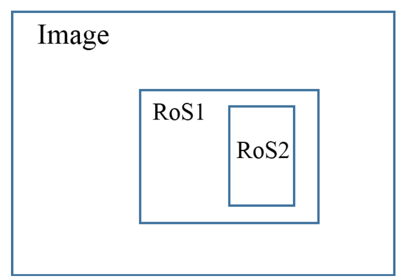

Fig. 4. Illustration of region of search (RoS)

D. Depth based object localization

We process following steps for object detection on the (RoS_1) of the depth map.

1) Removing pixels of floor plane: A number of floor plane detection methods have been proposed. In our work, we choose a ground plane detection technique provided by Microsoft Kinect SDK because of its good performance. The floor plane is described by the equation: $aX + bY + cZ + d = 0$. Each pixel of the depth map is described by a triplet (X, Y, Z). We remove all points satisfying the following condition:

$$\delta = \frac{aX + bY + cZ + d}{\sqrt{a^2 + b^2 + c^2}}; \delta < \theta_d \quad (1)$$

where θ_d is a pre-defined threshold.

2) Morphological operation and connected component analysis: We apply opening operation to remove single pixels and fill holes due to the noise in depth. Then, we apply a connected component analysis technique to group nearby pixels into a region. Due to the feature of Kinect sensor, some black regions of the object can not be measured (See in Fig.3) leading to the defragment of object region.

Based on the assumption that the object stands on the floor, we compute the distribution of the depth map and combine regions belonging to the same range in $X \in [X_1^W, X_2^W]$ and in $Y \in [Y_1^W, Y_2^W]$ in the WCS. Finally, we remove too small regions and keep only regions with the size suitable to the size of the object. We call the region of search now RoS_2.

Then we apply a HOG-SVM detector on the candidate region RoS_2. If we do not find any region due to the important missing of depth data, we apply the HOG-SVM detector on RoS_1. The HOG-SVM detector on RGB data is presented in the next section.

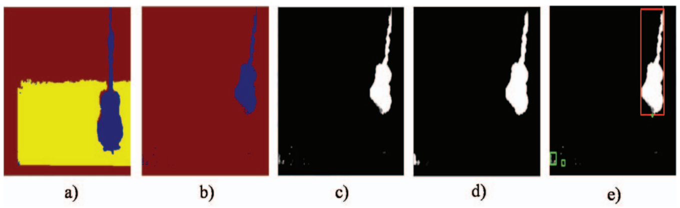

Fig. 5. Illustration of steps for depth based object detection:a) depth map; b) after removing floor points; c) thresholding; d) morphological operation; e) detected object

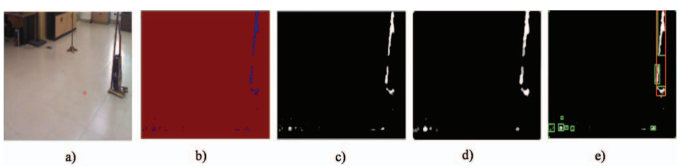

Fig. 6. Illustration of steps for depth based object detection in case of noisy depth:a) RGB image; b) depth map after removing floor pixels; c) thresholding; d) morphological operation; e) fusion of regions (red rectangle is RoS_2)

E. RGB based object localization

HOG-SVM is well known to be an efficient object detector [1]. We have to prepare positive and negative examples to train the detector and choose suitable parameters of HOG-SVM to our context. Specifically, we train two detectors: frontal detector detects the object from frontal view and side detector detects the object from side view. We have built 168 frontal positive samples, 160 side positive samples and 600 negative samples. Linear SVM is used. First, each sample is resized to 32x80. The number of bins is set to 9, each cell has a size of 8x8 pixels. Each feature vector consists of 972 elements. Fig.7 shows examples of positive and negative samples.

IV. Experimental Results

A. Material and environment setup

We conduct our experiment in a space of size 4.8mx2.4m in the Showroom of MICA Institute. We mount a Kinect sensor on a frame at height of 1.5m, the angle view of the Kinect towards to the space so that it could observe the most space.

Fig. 7. First row: Negative samples. Second row: frontal positive samples. Third row: side positive samples

The Kinect is connected to a nearby PC through USB port. There are two antennae (Thing Magic). Each is mounted on a frame of height 1.0m. One or two tags are attached on the object. In this work, we test with one object that is the guitar standing on the floor plane. Fig.8 is a snapshot of the environment and materials.

Fig. 8. Illustration of environmental setup

B. Evaluation of RFID based localization

We fix the tag on the object. We put the object at 18 positions on the floor. At some positions the RSSI is unmeasurable from antennae, we set a default value RSSI to -90dB. The tag could toward to the antenna in any direction. In this experiment, we study four cases with different positions/ directions of tags and antennae. The objective of this experiment is to estimate the range of RFID based localization error RMSE (Root Mean Square Error).

- Case study 1: One tag is attached to the object. The tag towards the antenna 1.
- Case study 2: One tag is attached to the object. The tag towards the angle of 45^o.

- Case study 3: Two tags are attached to the object, one tag towards the antenna 1, one tag towards the antenna 2. Suppose that (X_1^W, Y_1^W) and (X_2^W, Y_2^W) are two positions determined using tag 1 and tag 2 respectively. The position of the object is determined by the average in X and Y. $X^W = \frac{X_1^W + X_2^W}{2}$; $Y^W = \frac{Y_1^W + Y_2^W}{2}$
- Case study 4: Two antennae opposite to each other. Two tags are attached to the object, each tag towards to one antenna.

Figure 9 shows the results obtained for each case study. We found that the first case has the best result. In the first case study, the $RMSE \in [0, 1.5m]$ where $RMSE \in [0, 3.12m]$ in the second case. The use of two tags do not give better results (case 3 and 4 have $RMSE \in [0, 3.57m]$ and $[0, 1.73m]$ respectively). The RMSE is biggest at corners of other sides of space without antenna. This could be improved when we put each antenna at each side of the rectangle. From this result, we configure antenna-readers and attach only one tag to the object as the first case study.

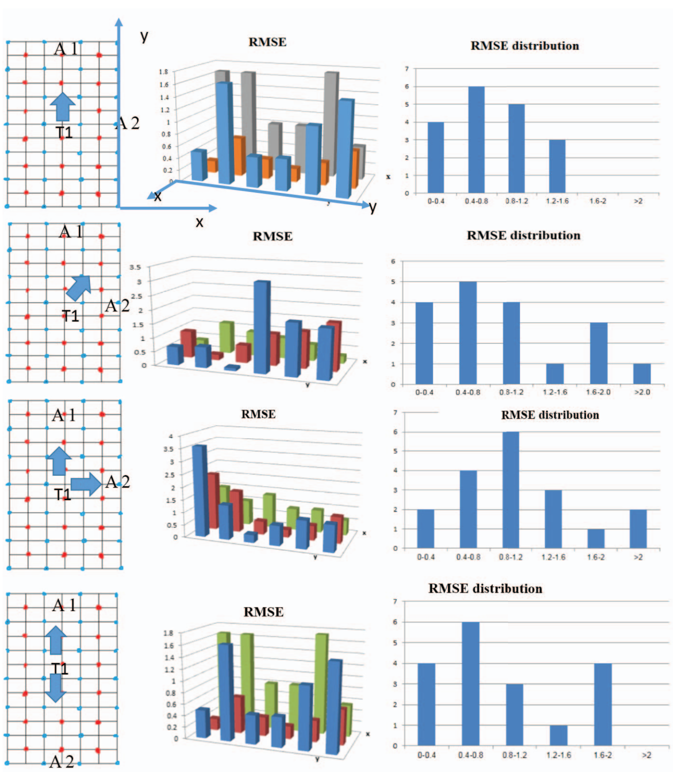

Fig. 9. Four rows show the RMSE corresponding to four case studies: first column: tags and antennae setting, the reference points are blue, the test points are red; second column: the RMSE at each test point; third column: the RMSE distribution

With the configuration of the first case study, we build four datasets of RSSI for four different orientations of the tag in the space. We found that when referring to four dataset, the average localization error $\delta = 1.04m$. This error is reasonable for further re-localization.

TABLE I
COMPARISON OF OBJECT LOCALIZATION PERFORMANCE

Case study	Error range	Average error
1	**[0-1.50]**	**0.77**
2	[0-3.12]	1.02
3	[0-3.57]	1.15
4	[0-1.73]	0.85

C. Evaluation of RGB-D based localization

Notice that the tag is fixed on the object. We put the object at different positions on the floor so that the depth could or could not be measured from the depth sensor of the Kinect. When the orientation of object changes, the tag changes and the RSSI received from this tag will be different. In this experiment, we put the object in four directions to evaluate the influence of tag direction on localization error.

Totally, we have 80 images among which one image has not observation of the object. We obtained Precision = 100% and Recall = 98.68%. Fig. 10 illustrates two examples of object detection using depth information. The red dots are the positions localized by RFID. We could see these positions are far from the object. Using depth, it is re-localized more precisely. In the figure, lots of green rectangles with small size have been not considered as object.

Fig. 10. Illustration of object detection based on depth

When using RFID, the RMSE of localization is 1.02m. When using depth based re-localization, the RMSE reduced to 0.16m. The RMSE varies in function of object orientation. When the object is frontal to the Kinect sensor, the RMSE is smallest (0.08m). When the object seen from the back, the black color is absorbed then the depth data is unmeasured or noisy. In that case, the RMSE is biggest (0.30m) (see Fig.11).

TABLE II
COMPARISON OF LOCALIZATION ERROR

Method	RMSE(m)	Precision (%)	Recall (%)	Time (s)
Only RFID	1.02	100	100	-
Only RGB	-	72.86	98.68	1.038
RFID+RGB-D	**0.16**	**100**	**98.68**	**0.39**

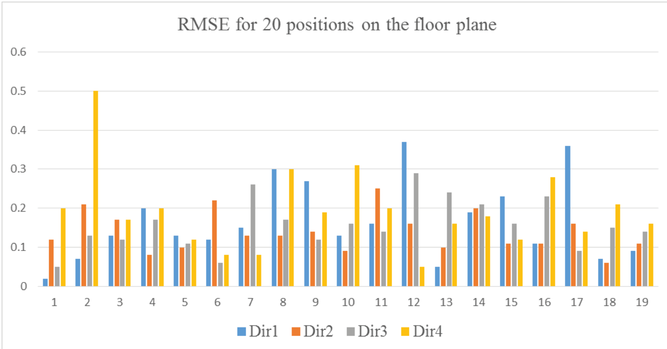

Fig. 11. RMSE for 4 directions of the object at 20 positions on the floor plane. In this figure, at one position, the camera can not observe the object so we do not show the RMSE.

Fig. 12 shows two true positives and one false positive of RGB based detection. In this figure, the red dot is the position localized by RFID which is still far from the true position of the object. It is obvious to see that HOG-SVM detector improves the localization of the object. The false negative is due to the fact that the object is too near from other objects in the scene (see Fig.12). Thanks to the coarse localization of RFID, the search region is reduced then the computational time of RGB-D object detection reduces from 1.038s to 0.39s per frame.

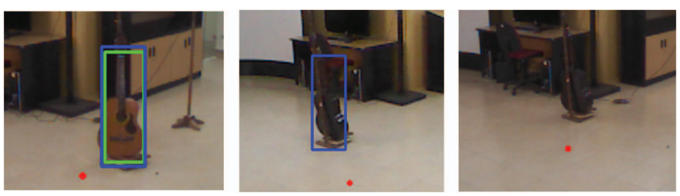

Fig. 12. Results of object detection using HOG-SVM. The red dot is the position localized by RFID, projected in image space. The object detected is bouned by the rectangle (blue, green) from both frontal and side HOG-SVM detectors.

V. Conclusions

In this paper, we presented a new method for object localization using RFID and RGB-D sensors following a serial processing steps. RFID gives coarse localization. Then object position was refined using RGB-D information. We have conducted the experiment in indoor environment. The experimental results showed that the combination improved significantly the localization error as computational time. However, the current methods have some limitations. First, we have evaluated with only one object. Multiple object detection and localization should be considered. The localization based of RFID is limited to the configuration of RFID readers as well as tags. If we have more readers and more tags, the coarse localization could be still improved. For example, a triangulation method could be deployed for RFID based localization. Besides, currently we have used only one Kinect sensor. Multiple Kinect sensors will give a multi-view observation of the object, then could improve RGB-D object localization.

Acknowledgment

This work was supported by the Visually impaired people assistance using multimodal technologies project funded by the Vlaamse Interuniversitaire Raad (VLIR) in the framework of the VLIRs Own Initiatives Program 2012 under the grant number ZEIN2012RIP19.

References

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, 2005, pp. 886–893.
[2] Q.-H. Nguyen, H. Vu, T.-H. Tran, D. Van Hamme, P. Veelaert, W. Philips, and Q.-H. Nguyen, “A visual slam system on mobile robot supporting localization services to visually impaired people,” in *Computer Vision-ECCV 2014 Workshops*. Springer, 2014, pp. 716–729.
[3] Q.-H. Nguyen, H. Vu, T.-H. Tran, and Q.-H. Nguyen, “Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment,” *Multimedia Tools and Applications*, pp. 1–25, 2016.
[4] V.-H. Le, H. Vu, T. T. Nguyen, T.-L. Le, T.-T.-H. Tran, M. Vlaminck, W. Philips, and P. Veelaert, “3d object finding using geometrical constraints on depth images,” in *Seventh International Conference on Knowledge and Systems Engineering (KSE), 2015*. IEEE, 2015, pp. 389–394.
[5] P. Kamol, S. Nikolaidis, R. Ueda, and T. Arai, “RFID Based Object Localization System Using Ceiling Cameras with Particle Filter,” in *Future Generation Communication and Networking (FGCN 2007)*, vol. 2, 2007, pp. 37–42. [Online]. Available: http://stefanosnikolaidis.net/papers/Kamol_FGCN_2007.pdf
[6] T. Deyle, H. Nguyen, M. Reynolds, and C. C. Kemp, “RF Vision: RFID Receive Signal Strength Indicator (RSSI) Images for Sensor Fusion and Mobile Manipulation,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 5553 – 5560.
[7] G. C. Karmakar, L. S. Dooley, N. C. Karmakar, and J. Kamruzzaman, “Object analysis with visual sensors and RFID,” in *Chipless and Conventional Radio Frequency Identification: Systems for Ubiquitous Tagging*, 2012, pp. 234–250.
[8] J. Kim and W. Y. Jung, “Vision-based Cattle Detection and Localization System in an RGB Color Space,” in *The Second International Conference on Intelligent Systems and Applications*, 2013, pp. 173–175.
[9] *http://www.vision.caltech.edu/bouguetj/calib_doc/.*
[10] *http://nicolas.burrus.name.*