Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 32 (2014) 1095 – 1100
Procedia Computer Science

3rd International Workshop on Cooperative Robots and Sensor Networks (Robosense-2014)

# Distributed sensor network for multi-robot surveillance

A. Pennisi[a,*], F. Previtali[a], F. Ficarola[a], D. D. Bloisi[a], L. Iocchi[a], A. Vitaletti[a]

[a]*Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Italy*

**Abstract**

Monitoring of populated indoor environments is crucial for the surveillance of public spaces like airports or embassies, where the behavior of people may be relevant in order to determine abnormal situations. In this paper, a surveillance system based on an integration of *interactive* and *non-interactive* heterogeneous sensors is described. As a difference with respect to traditional, pure vision-based systems, the proposed approach relies on Radio Frequency Identification (RFID) tags carried by people, multiple mobile robots (each one equipped with a laser range finder and an RFID reader), and fixed RGBD cameras. The main task of the system is to assess the presence and the position of people in the environment. This is obtained by suitably integrating data coming from heterogeneous sensors, including those mounted on board of mobile robots that are in charge of patrolling the environment. The robots also adapt their behavior according to the current situation, on the basis of a Prey-Predator scheme. Experimental results carried out both on real and on simulated data show the effectiveness of the approach.

Selection and Peer-review under responsibility of the Program Chairs.

*Keywords:* Mobile Robots; Wireless Sensor Networks; Multi-Robot Systems; Multi-Robot Surveillance

## 1. Introduction

A critical infrastructure (CI) is a system which is essential for the maintenance of vital societal functions. The damage to a CI, due to terrorist attacks, criminal activities or malicious behaviors may have a significant negative impact for the entire society. Usually, CIs are monitored by passive cameras and appropriate computer vision techniques are used for tracking people and understanding their behaviors. However, in addition to the well-known problems affecting vision-based surveillance (e.g., changes in illumination conditions, occlusions, and re-identification), passive vision-based systems can result ineffective when dealing with realistic scenarios, since relaying only on passive fixed sensors it is hard to identify and tracking a person in a large environment and to obtain relevant information about him/her. Moreover, vision systems can be subject to malicious physical attacks[1].

In this paper, the problem of monitoring a populated indoor environment is faced by combining data coming from multiple heterogeneous sensors. We consider a system in which authorized personnel wear Radio Frequency Identification (RFID) tags, fixed RGBD cameras with RFID receivers are placed in the scene, and multiple mobile robots, equipped with laser range finders and RFID receivers, patrol the environment. Laser scans, RFID tag data, and RGBD images are merged in order to acquire information about the position and the identity of people in the environment. The system works in a distributed fashion in order to verify normal behavior of people and to automatically raise alarms when abnormal conditions are detected.

* Corresponding author. Tel.: +39-06-772-74157 ; fax: +39-06-772-74106.
*E-mail address:* pennisi@dis.uniroma1.it


Selection and Peer-review under responsibility of the Program Chairs.
doi:10.1016/j.procs.2014.05.538

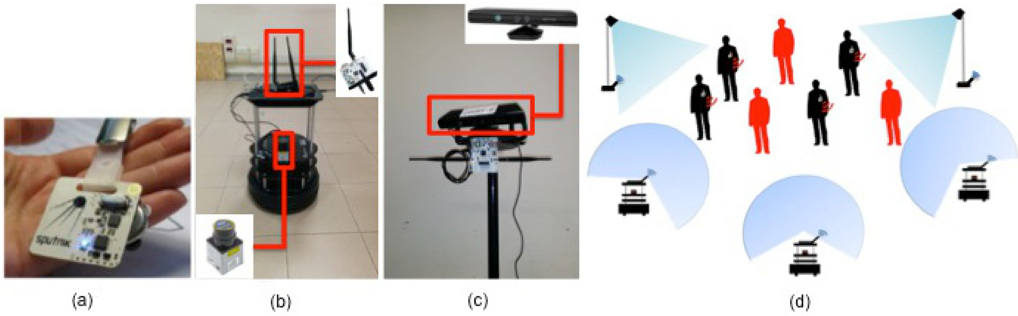


Fig. 1: (a) An RFID tag. (b) A Turtlebot robot equipped with a laser range finder and an RFID receiver. (c) A fixed RGBD camera and an RFID receiver. (d) The proposed system combining robots, RFID tags, and fixed RGBD cameras to monitor a populated environment.

The remainder of the paper is organized as follows. Related work is analyzed in Section 2, while Section 3 provides the definition of the addressed problem. The proposed system is described in Section 4 and the results coming from a first set of simulated and real experiments are discussed in Section 5. Conclusions are drawn in Section 6.

## 2. Related Work

There exists a large literature about the problem of people detection in indoor environments by using cameras. However, since a variety of factors, including illumination conditions, occlusions, and blind spots, limit the capacity of pure vision-based systems, it is possible to consider a combination of multiple heterogeneous sensors to achieve better results. The systems dealing with multiple sensors can be divided into two main categories, namely *interactive methods*, where each person has an active role during the detection process (e.g., by dressing an RFID tag as shown in Fig. 1a), and *non-interactive methods*, where the role of the person is passive and the analysis is computed only by the detection system (e.g., a robot equipped with a laser or an RGBD camera). In order to improve the accuracy of the information of the monitored environment, a combination of interactive and non-interactive sensors can be chosen (e.g., by using a robot equipped with a range finder and an RFID receiver as shown in Fig. 1b).

*Interactive Methods.* One of the first experiments about collecting information from a group of people in a physical real context was carried out by Hui *et al.*[2], where 54 individuals attending to a conference were dressed with an *Intel iMote* device, a Bluetooth radio and a flash memory. However, the choice of using Bluetooth did not allow a fine-grained recording of social interactions.

Following projects that focused on the collection of huge data sets from social interactions were developed by the *SocioPatterns* collaboration. Partners participating in this collaboration were the first to record fine-grained contacts by using RFID sensors. SocioPatterns realized several installations in different social contexts (e.g., conferences[3], hospitals[4], primary schools[5], a science gallery[6]) and made some data sets publicly available on their website. Experiments similar to the SocioPatterns' ones were deployed by Chin *et al.*[7], in which each person wore an active RFID badge during a conference. A remarkable result of the experiment was that, for social selection, more proximity interactions lead to an increased probability for a person to add another as a social connection. Recently, Becchetti *et al.*[8] collected data coming from wireless active RFID tags worn by 120 volunteers moving and interacting in an indoor area to assess the performance of Population Protocols[9], a fully decentralized computational model, on real dynamic social networks.

While the above approaches target the analysis of social human behaviors, in this paper we investigate the use of data acquired from interactive tags for surveillance applications. Indeed, we aim at integrating the SocioPatterns sensing platform together with other sensing technologies, including laser range finder and RGBD cameras, to overcome the problems related to traditional automatic surveillance. It is worth noticing that a scenario in which 1) authorized personnel wear RFID tags and 2) other people (e.g., visitors, travelers, spectators) have an RFID transmitter (e.g., included in a ticket or a passport or a boarding pass) is a quite plausible one (e.g., airports, embassies, theaters).

*Non-Interactive Methods.* Non-interactive methods are based on passive sensors. Since the literature on vision-based systems is huge, we limit our description to the approaches that use technologies other than vision for overcoming problems such as blind spots and occlusions. In the field of laser-based systems, Cui *et al.*[10] introduced a feature extraction method based on accumulated distribution of successive laser frames. A pattern of rhythmic swing legs was used to extract each leg of a person and a region coherency property was introduced to construct an efficient

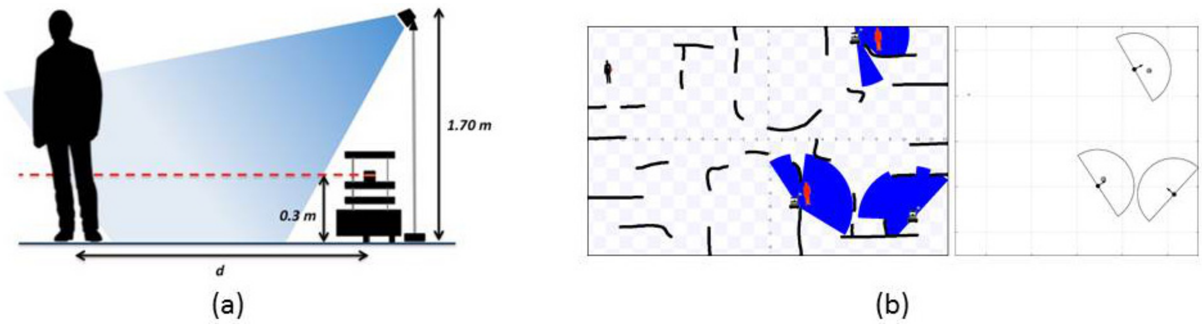


Fig. 2: (a) Real experiment evaluation: a person located at a known distance is detected by both the laser range finder and the fixed RGBD camera. (b) The simulated environment in Stage.

measurement likelihood model. A combination of independent Kalman filter and Rao-Blackwellized Monte Carlo data association filter (RBMC-DAF) was used to track people. However, this approach is not effective for people moving quickly or partially occluded.

Xavier *et al.*[11] developed a feature detection system for real-time identification of lines, circles, and legs from laser data. Lines was detected by using a recursive line fitting method, while leg detection was carried out taking into account geometrical constrains. This approach cannot handle scan data of a dynamic scene including moving people and not well separated structures.

Other methods integrate laser sensors with vision. Shao *et al.*[12] presented a solution for human-robot interaction based on a combination of visual and laser range information. Legs were extracted from laser scans and, at the same time, faces were detected analyzing the images of a camera. The information was integrated in a detection procedure returning the direction and the distance of the surrounding people, and it was used by a mobile robot to approach and to start interacting with humans. However, the swinging frequency results low for people tracking and detection.

While in the above cited papers the analysis of sensor data is limited to detect a single person and it cannot be easily extended when multiple people are grouped together, in this paper we propose an approach that can deal with groups of people.

## 3. Problem Definition

The problem of monitoring a populated environment can be modeled as a *Prey-Predator* game. Given the predator and the prey species, it is possible to formalize the monitoring task as follows. *A predator tries to catch preys and a prey runs away from predators*.

The game consists of preys and predators living in the same environment. It is usually defined as a game where both predators and preys has a score and any individual can gain or lost points over time. A metric distance is assigned to each prey and to each predator as the game score. The goal for each prey is to maximize its distance from the predators, while the aim of each predator is to minimize its distance from the preys.

In our setting, the preys are the people moving in the monitored environment, while the predators are sensor nodes that can detect the presence and the position of a person. A sensor node is made of an RFID reader and other sensors (e.g., an RGBD camera or a laser range finder). Moreover, some sensor nodes are mounted on mobile robots and thus they move around the environment.

The goal of catching a prey is achieved whenever a sensor node is close enough to a person, since it can read the RFID tag and possibly determine that such a person does not wear an RFID tag. The same performance metrics defined for the *Prey-Predator* scheme can be used for evaluating the approach. Experimental results are reported in Section 5.

## 4. System Description

The proposed approach is composed of three modules, namely 1) Perception, 2) Data Fusion, and 3) Dynamic Task Assignment. Each module will be detailed in the following.

*Perception.* The perception task is performed by using a combination of interactive and non-interactive methods. In particular, mobile robots equipped with laser and RFID receiver sensors, RFID tags worn by people, and fixed RGBD cameras are considered (Fig. 1).

*RFID tags and receivers.* The two main entities of our sensing platform, designed and developed by the SocioPatterns research collaboration, are the OpenBeacong active tags (Fig. 1a) and the OpenBeacon Ethernet EasyReader PoE II device (Fig. 1b). The tags are electronic wireless badges equipped with a micro-controller and a transceiver. They are powered by batteries ensuring a lifetime of about two weeks. The tags are programmed to periodically broadcast beacons of 32 bytes at different levels of signal strength. Every beacon contains the tag identifier, the information about the current signal strength, and other fields useful for debugging. The RFID receivers collect the data sent by the tags via a wireless channel. In our experimental scenarios, a receiver is mounted on each robot and it is used to read the signal strength and the ID of a tag, in order to detect if a person in the environment is actually wearing a tag.

*Laser person detection and tracking.* A mobile sensor composed by a Turtlebot equipped with a range finder and an RFID receiver is used to monitor a limited indoor environment (Fig. 1b). The robot has the map of the environment and it is well localized on it. The person detection is carried out by means of a *distance map*, indicating the probability that a given point belongs to the map, that is used to detect the foreground objects, i.e., sets of points that are far enough from the map points. From each object a set of features is extracted (i.e., the number of points of the object, the standard deviation, the bounding box, and the radius). Then, the features are sent as input to an Ada-Boost based person classifier, trained with about 1800 scans. People tracking relies on a multi-hypothesis approach based on a set of Kalman Filters [13]. Data association is used to determine the relationship between observations and tracks, and multiple hypotheses are maintained when observations may be associated to more than one track. Finally, each track is combined with the signal detected by the RFID receiver mounted on each robot, in order to verify if a person is wearing the RFID tag.

*RGBD fixed cameras.* To compute an accurate foreground detection both color and depth information are used. An RGB image and a 16 bit depth map are stored for each captured frame. A statistical approach, called IMBS [14], is used to create the background model that is updated every 30 seconds for dealing with illumination changes. The positions of the foreground blobs are computed by combining the foreground mask and the depth map. A surface normal approach [15] is used to recognize the floor. Given the set of 3D points of each blob, the problem of determining the normal to a point on the surface is approximated by estimating the normal of a plane tangent to the surface, thus resulting in a least-square plane fitting estimation problem. Therefore, surface normal estimation is reduced to an analysis of the eigenvectors and eigenvalues of a covariance matrix created from the nearest neighbors of the query point, where the sign of the normal is assigned on the basis of the view point of the scene. Knowing the position of the floor, it is possible to align the data coming from the RGBD camera and the laser (Fig. 2a).

*Data Fusion.* The previous modules produce data in a time interval $\mathscr{T} = [t, ..., T]$, during which each robot executes the patrol task and each RGBD camera monitors a portion of the environment. Thus, the information collected during $\mathscr{T}$ needs to be merged. Data coming from the range finders and the RGBD cameras are fused by using a Kalman filter-based approach [16] in order to obtain a global occupancy map $M$.

To detect and identify a person entering the scene, a predefined entrance area is defined to create an event detection area. The system maintains a representation of the scene $S$, consisting of a list of IDs, where the positions in the list correspond to the order in which the people enter the scene. A set $U$ of currently detected IDs is generated by analyzing the set $S$ to check if a new ID has been detected by the RFID receiver. In case of a new ID detection, $S$ is updated to $S'$ for including the new ID ($S' \leftarrow push(S, newID)$). After every patrol task, the following data are available: the current status of the scene $S$, the current global occupancy map $M$, and the current set $U$ of the detected IDs, producing a new status of the scene $S'$ and possibly alarms if the scene rules are violated. In the following, the notation $set(S)$ denotes the set of IDs included in $S$ (without considering their position), thus if $set(S) = U$, then the same IDs in the scene $S$ are present in $U$.

Both the detection and the identification of a person leaving the scene are carried out in a way similar to the one used for the previous task. People are detected in a predefined exit area by using the laser range finders and the RGBD cameras, while the difference between $set(S)$ and $U$ is used to detect the leaving ID. If it belongs to $S$, then $S' \leftarrow delete(S, leavingID)$, otherwise the system launches an alarm representing that a person without tag is exiting the scene.

When no person is entering or leaving the scene, the following checks are executed. 1) if $set(S) = U$, than the current set of IDs corresponds to the IDs in the scene; 2) if $count(M) = |S|$ than the estimated number of people in the scene given by the range finder and the RGBD analysis corresponds to the size of the set of the IDs. If both the conditions are true, then $S$ is updated to $S'$, otherwise particular alarms can be sent out. As an example, if $set(S) \neq U$ and $count(M) = |S|$, then an alarm is generated for a person who is still in the scene, but not having any more the tag (e.g., voluntary switch off).

Table 1: Results in the real scenario.

| Sensor Type | Real Distance | Detected Distance (avg ± std. dev.) | Error (avg ± std. dev.) | Error Robot/Kinect Localization |
|---|---|---|---|---|
| Kinect | 1 *m* | 1.441 ± 0.002 *m* | 0.441 ± 0.002 *m* | ± 0.12 *m* |
| Laser | 1 *m* | 1.029 ± 0.013 *m* | 0.029 ± 0.013 *m* | ± 0.11 *m* |
| Kinect | 2 *m* | 2.404 ± 0.013 *m* | 0.404 ± 0.013 *m* | ± 0.12 *m* |
| Laser | 2 *m* | 2.040 ± 0.011 *m* | 0.040 ± 0.011 *m* | ± 0.11 *m* |
| Kinect | 3 *m* | 3.464 ± 0.015 *m* | 0.464 ± 0.015 *m* | ± 0.12 *m* |
| Laser | 3 *m* | 3.068 ± 0.006 *m* | 0.068 ± 0.006 *m* | ± 0.11 *m* |
| Kinect | 4 *m* | 4.533 ± 0.021 *m* | 0.533 ± 0.021 *m* | ± 0.12 *m* |
| Laser | 4 *m* | 4.066 ± 0.038 *m* | 0.066 ± 0.038 *m* | ± 0.11 *m* |

*Dynamic Task Assignment.* The *Dynamic Task Assignment* (*DTA*) is performed by using a greedy algorithm[17,18] that assigns a prey to a predator. A predator creates a new *bid* each time it has seen a prey. A bid is a list of costs and information gains for catching the prey. Bids are asynchronously sent to all the predators and the DTA algorithm makes the assignment on the basis of the current bids. During the chasing, a predator could change the prey to chase, therefore, in order to handle this situation, the DTA algorithm assigns the prey no longer chased to another predator.

## 5. Experimental Evaluation

Preliminary results, performed both in a real scenario and by using a simulator, are reported in the following. The experiments carried out in the real scenario have been useful to compute the error model of the sensors (both for the RGBD cameras and the laser range finders), that has been considered in the simulated environment to quantitatively evaluate the effectiveness of our approach. More extensive experiments will be carried out in order to confirm the significance of the proposed method.

*Experiments in a real scenario.* In order to demonstrate the effectiveness of the system and to compute typical error models for the used sensors, an experimental setting made of a Turtlebot robot (equipped with an Hokuyo laser) and a fixed Kinect camera has been considered. The set up is shown in Fig. 2a and it has been used to measure the error in the detection of a person in the environment. Landmarks at known distances have been placed on the floor in order to register ground-truth data. Several measurements for each landmark have been performed by using the approach discussed in Section 4. The obtained results are reported in Table 1.

As expected, the accuracy of the laser-based method is higher than the one of the RGBD-based technique. However, when mounted on the robot also the accuracy of the self-localization of the robot must be taken into account. The above results, although preliminary and incomplete, are useful to determine a suitable error model for the sensors involved in the system. A sensor model taking into account such an error has been used in the simulated experiments described below to obtain more realistic observations during the simulations.

*Experiments in a simulated environment.* The goal of the experimental evaluation on simulated data is to quantitatively evaluate the performance of our method. We run all the experiments by using the simulator Stage. In Stage, both the sensor nodes and the people are represented as robotic agents. The estimation of the position of the simulated people (i.e., the implementation of the virtual sensors) is obtained by generating observations with the addition of an error calculated accordingly to the error model of the real sensors calculated in the experiments discussed above. Fig. 2b shows a screen-shot from an experiment in which three robots (i.e., predators) are chasing the preys (i.e., people without an RFID tag). The results obtained during the simulations are reported in Table 2. The low value of the standard deviation demonstrates a remarkable reliability of the proposed approach.

The reported results show that the integration of data coming from heterogeneous sensor nodes composed of active RFID tags, RGBD cameras, and mobile laser range finders can be used to deal with the problem of monitoring a populated environment. A more accurate experimental analysis for measuring false positive/false negative rates in different situations and integration with other techniques (e.g., vision) would further improve the assessment of the quality of the system.

## 6. Conclusions

Integrating multiple technologies for surveillance applications is an important step in order to develop and deploy effective systems. In this paper we propose a method for integrating heterogeneous fixed and mobile sensor nodes in order to determine the presence and the position of people in an indoor environment. Different technologies (RFID tags, laser range finders, and RGBD cameras) are combined through a distributed data fusion method that is robust to perception noise and is scalable to multiple heterogeneous sensors. The reported preliminary results show the

Table 2: Results in the simulated environment.

| **Experiment** | **Prey-Predator Distance (avg ± std. dev.)** | **Experiment** | **Prey-Predator Distance (avg ± std. dev.)** |
|---|---|---|---|
| 1 | 0.81 $m \pm$ 0.13 $m$ | 6 | 0.64 $m \pm$ 0.17 $m$ |
| 2 | 1.22 $m \pm$ 0.21 $m$ | 7 | 0.79 $m \pm$ 0.31 $m$ |
| 3 | 0.83 $m \pm$ 0.15 $m$ | 8 | 1.18 $m \pm$ 0.35 $m$ |
| 4 | 1.43 $m \pm$ 0.08 $m$ | 9 | 1.03 $m \pm$ 0.22 $m$ |
| 5 | 1.38 $m \pm$ 0.13 $m$ | 10 | 1.39 $m \pm$ 0.28 $m$ |

feasibility of the approach and the overall capabilities of the system. Automatic monitoring and detection of abnormal activities are possible and performance in this task can be good enough for an actual deployment. However, additional work must be done in order to make the techniques more precise and more robust.

## Acknowledgements

This work has been partially supported by the TENACE PRIN Project (n. 20103P34XC) funded by the Italian Ministry of Education, University and Research and has been possible thanks to the cooperation with the SocioPatterns collaboration (`www.sociopatterns.org`).

## References

1. Liu, C., James, T.Y.. An analysis of dos attacks on wireless lan. In: *Wireless and Optical Communications*. 2006, .
2. Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., Diot, C.. Pocket switched networks and the consequences of human mobility in conference environments. In: *Proc. of ACM SIGCOMM First Workshop on Delay Tolerant Networking and Related Topics*. 2005, p. 244–251.
3. Barrat, A., Cattuto, C., Szomszor, M., den Broeck, W.V., Alani, H.. Social dynamics in conferences: analysis of data from the live social semantics application. In: *Proceedings of the 9th International Semantic Web Conference*. 2010, p. 17–33.
4. Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. In: *PLoS ONE*. 2011, p. e17144.
5. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.F., et al. High-resolution measurements of face-to-face contact patterns in a primary school. In: *PLoS ONE*. 2011, p. e23176.
6. Van den Broeck, W., Quaggiotto, M., Isella, L., Barrat, A., Cattuto, C.. The making of sixty-nine days of close encounters at the science gallery. *Leonardo* 2012;:285–285.
7. Chin, A., Xu, B., Wang, H., Wang, X.. Linking people through physical proximity in a conference. In: *Proc. of the 3rd International Workshop on Modeling Social Media*. 2012, p. 13–20.
8. Becchetti, L., et al. Population protocols on real social networks. In: *Proceedings of the Fifth Workshop on Social Network Systems*. 2012, p. 15:1–15:2.
9. Angluin, D., Aspnes, J., Eisenstat, D., Ruppert, E.. The computational power of population protocols. *Distributed Computing* 2007; :279–304.
10. Cui, J., Zha, H., Zhao, H., Shibasaki, R.. Laser-based detection and tracking of multiple people in crowds. *Computer Vision and Image Understanding* 2007;:300–312.
11. Xavier, J., Pacheco, M., Castro, D., Ruano, A.. Fast line, arc/circle and leg detection from laser scan data in a player driver. In: *Proc. of the IEEE International Conference on Robotics and Automation*. 2005, p. 3930–3935.
12. Shao, X., Zhao, H., Shibasaki, R., Shi, Y., Sakamoto, K.. 3d crowd surveillance and analysis using laser range scanners. In: *IROS*. 2011, p. 2036–2043.
13. Bloisi, D.D., Iocchi, L., Marchetti, L., Monekosso, D.N., Remagnino, P.. An adaptive tracker for assisted living. In: *IEEE Conf. on Advanced Video and Signal Based Surveillance*. 2009, p. 164–169.
14. Bloisi, D.D., Iocchi, L.. Independent multimodal background subtraction. In: *Proceedings of the Third International Conference on Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications*. Rome, Italy; 2012, p. 39–44.
15. Rusu, R.B.. Semantic 3d object maps for everyday manipulation in human living environments. *Articial Intelligence (KI - Kuenstliche Intelligenz)* 2010;.
16. Sasiadek, J., Hartana, P.. Sensor data fusion using kalman filter. In: *Third International Conference Information Fusion*. 2000, p. WED5–19.
17. Simmons, R., Apfelbaum, D., Burgard, W., Fox, D., Moors, M., Thrun, S., et al. Coordination for multi-robot exploration and mapping. In: *AAAI/IAAI*. 2000, p. 852–858.
18. Palmer, D., Kirschenbaum, M., Murton, J., Zajac, K., Kovacina, M., Vaidyanathan, R.. Decentralized cooperative auction for multiple agent task allocation using synchronized random number generators. In: *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*; vol. 2. 2003, p. 1963–1968.