

# A System for Multimodal Context-Awareness

Georgios Galatas  
Computer Sci. and Eng. Dept.  
University of Texas at Arlington  
Arlington, TX, USA IIT, NCSR Demokritos  
Athens, Greece

Fillia Makedon  
Computer Sci. and Eng. Dept.  
University of Texas at Arlington  
Arlington, TX,  
USA

**Abstract**—in this paper we present the improvement of our novel localization system, by introducing radio-frequency identification (RFID) which adds person identification capabilities and increases multi-person localization robustness. Our system aims at achieving multi-modal context-awareness in an assistive, ambient intelligence environment. The unintrusive devices used are RFID and 3-D audio-visual information from 2 Kinect sensors deployed at various locations of a simulated apartment to continuously track and identify its occupants, thus enabling activity monitoring. More specifically, we use skeletal tracking conducted on the depth images and sound source localization conducted on the audio signals captured by the Kinect sensors to accurately localize and track multiple people. RFID information is used mainly for identification purposes but also for rough location estimation, enabling mapping of the location information from the Kinect sensors to the identification events of the RFID. Our system was evaluated in a real world scenario and attained promising results exhibiting high accuracy, therefore showing the great prospect of using the RFID and Kinect sensors jointly to solve the simultaneous identification and localization problem.

**Keywords**—Multimodal; Context-awareness; Microsoft Kinect; RFID; Localization; Identification

## I. INTRODUCTION

An assistive ambient intelligence environment is a smart space that aids the inhabitants with its embedded technology. The proliferation of ambient intelligent environments has triggered research related to applications, such as monitoring Assistive Daily Living (ADL), fall detection, risk prevention and surveillance [1, 2]. For achieving these goals, activity recognition performed in a natural and unintrusive way is of utmost importance. The most fundamental step towards activity monitoring and ultimately context-awareness is successful multi-person identification and localization. By utilizing the location of the person in a domestic setting, related activities can be derived. Accurate person localization plays an essential role in all the aforementioned applications and has been dealt with using many different approaches. Nevertheless, when used domestically, most current implementations can be considered as invasive. Our novel system uses information from multiple sensors in order to ensure reliable and unintrusive localization of the inhabitants.

### A. Localization:

Applications that rely on localization such as surveillance and monitoring of ADL commonly use video cameras as an affordable and abundant source of information. Many

approaches based on either a single camera or multiple cameras have been proposed in the literature.

In single camera setups, discriminative appearance affinity models [3] and level-set segmentation [4] have been used for tracking, while other approaches based on tracking-by-detection exist [5, 6]. In multi-camera setups, stereo-vision is employed in order to introduce depth perception. In [7], color histograms of the person-shaped blobs are used to disambiguate between people, when they are very close to each other. The system tracks multiple people standing, walking, sitting, entering and leaving in real-time. In [8] two techniques were used to determine the location of a person in 3-D space. These were 1) best-hypothesis heuristic tracking and 2) probabilistic multi-hypothesis tracking to derive the 3-D location of people. The results show similar tracking performance for both approaches. However, the simplistic probabilistic approach produces more false alarms, which may be improved by using a sophisticated probabilistic model.

Solving the problem using only cameras is very challenging for a large space with many people. The reason is that localization requires wide coverage to capture and map the respective locations of many people simultaneously, but identification requires zooming into a person's face. In surveillance applications, cameras are typically mounted on tall poles and configured such that they could provide maximum coverage. Nevertheless, video feeds from such settings may not be sufficient to provide accurate information about a person's face or other biometric features. In addition, the segmentation and tracking problems can be very challenging, thus hindering the system's reliability in a camera-only setup. Furthermore, despite the fact that the use of cameras and computer vision techniques are very promising, extensive use of video cameras in a domestic setting can be considered a violation of privacy [9]. Therefore, our main focus is to achieve the same goal of identifying and localizing multiple people in an assistive environment in a less intrusive manner.

### B. Identification:

RFID (Radio-frequency Identification) systems are frequently being used to track medicine and patients in large hospitals in order to verify the correct medicine reaches the correct patients [10]. RFID sensors have become very popular, as they are cheap, easy to use and provide accurate identification information wirelessly [11]. Although RFID is very effective in identifying objects, it may not be as effective in surveillance applications, since people are required to wear an RFID tag so that the events related to the tag are detected. As a result, such systems may not be able to detect intruders or

This material is based upon work supported by the National Science Foundation under Grants No. NSF-CNS 1035913, NSF-CNS 0923494.

anyone not wearing a tag. However, it constitutes a viable solution for recognizing activities in a smart environment, since its inhabitants can very easily carry a passive RFID tag with them.

Multimodal person identification has become a significant area of research in recent pervasive assistive applications. Some of these applications use existing biometric identification methods, such as face recognition and speaker identification [9, 12, 13] to identify multiple people in smart environments [14]. Nevertheless, these approaches do not convey the location information of the person.

### C. Simultaneous Identification and Localization:

Locating multiple users simultaneously while identifying each one is considered to be the first step to create a context-aware application, such as activity and human behavior recognition. RFID technology has also been used to solve the problem of simultaneous identification and localization. Although radio signal propagation suffers from various problems, such as multipath, line of sight path, diffraction or reflection etc. even in an indoor environment [15], several indoor-based localization algorithms have been proposed in the literature, which, according to [16] can be classified into three categories: 1) distance estimation, 2) scene analysis and 3) proximity. Among them, distance estimation algorithms use different range measurement techniques, such as Received Signal Strength, Time of Arrival, Time Difference of Arrival, Received Signal Phase etc. and apply triangulation to estimate a location. On the other hand, the scene analysis approaches first measure fingerprints of an environment and then, try to match the target's range measurements with the appropriate set of fingerprints for estimating the location. Finally, the proximity-based algorithms determine a target's location by mapping it to the location of an antenna that receives the strongest signal.

Overall, RFID technology possesses a promising solution to identify and localize multiple objects with attached RFID tags. Existing well-known systems, such as LANDMARC [17] use active RFID tags and exploit the signal strength property to correctly localize an object. Passive RFID tags have also been used in the past to identify and locate multiple objects. In [18], the authors have utilized the percentage of tag counts at different power attenuation levels in order to approximate the distance between a reader and a tagged object. Another, indirect way of deriving the location information of an object is to record the location of the reader as the location of an object. But, the location accuracy and precision of such a system heavily depends on the level of deployment of readers and antennas in the space [19, 20].

However, RFID still lacks sufficient localization accuracy especially for the minimal number of deployed antennas and tags in a domestic environment. Simply using RFID to obtain the location of an object can lead to many false readings, e.g., an RFID antenna may miss a tag depending on the tag's position and the antenna's orientation.

In an attempt to improve accuracy, multi-modal person localization has become a significant research area in recent applications. Thus, for a very dynamic environment,

information collected from multiple sources, such as video cameras, microphone arrays, sensors etc. are all combined together such that the system can achieve better identification and localization accuracy. Techniques, such as Hidden Markov Models, K-nearest neighbors etc. can be applied to captured audio-visual signals to extract higher-level semantic information, such as identification and location in real time. A system that combines face and audio based identification along with motion detection, person tracking and audio based localization has been proposed in the literature [21]. Such a system applies state-of-the-art methods to process results from each individual modality and uses particle filtering to fuse both modalities for providing robust identification and localization.

Methods that combine localization using cameras with identification using wearable sensors or accelerometers are also proposed in the literature. Since most of the recent mobile phones contain accelerometers and magnetometers attached to them, mobile phones are considered to be very convenient and fulfill all of the above requirements. In [1] the authors combined an existing CCTV based system with sensors (accelerometers and magnetometers) embedded to a person's mobile phone as a solution. According to this method, the camera captures the location of each person, which is transmitted wirelessly to the mobile phone carried by the respective person. After receiving the location information, the mobile phone resolves the most probable location by matching them with the measurements from its own sensors. The identification process is very easy in this case, as each person is labeled with his/her mobile phone's unique ID.

The deployment of wireless sensor network (WSN) is another common approach nowadays to monitor and localize persons in assistive environments [22, 23]. RFID systems and WSNs can be combined together not only for identifying and localizing objects, but also for real-time monitoring [24]. To identify and localize in open areas, researchers of [2] derived a calibration method for a joint RFID-camera system based on the area of overlap between the field of view (FoV) of a camera and the field of sense (FoS) of RFID sensors.

In our approach we build in prior work [29] utilizing the identification capabilities of RFID and combining them with precise 3D tracking from the Kinect to create an accurate identification and localization solution. The latter is an active sensor, able to accurately measure the position of the person in the 3-D space. Skeletal tracking is carried out using the Kinect sensor's 3D depth images and sound source localization is conducted utilizing microphone arrays of 2 such sensors, to deduce accurate location information. At the same time, the video information is not captured, making this approach less intrusive than using video cameras. RFID is used mainly for discerning between users and also for providing a rough estimate of their location utilizing the RSSI. Our goal is to map the location of multiple people in an ambient intelligence environment at a detailed level that will allow inference of conducted activities (figure 1).

In the following sections we will present the architecture and operation of our system for person identification and localization, the experimental setup and finally our concluding remarks.

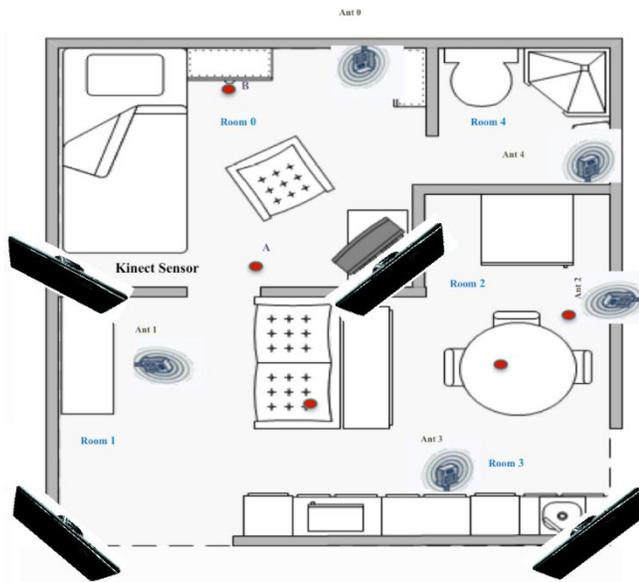


Fig. 1. Example apartment layout with RFID antennas and Kinect sensors deployment.

## II. SYSTEM OVERVIEW

### A. Hardware

The Microsoft Kinect (figure 2) is a novel device mainly used for gesture recognition. It is based on the PrimeSensor design [25] and it incorporates a color camera, a depth sensor and a microphone array. Depth images are acquired using the structured light technique. According to this method, a laser beam passes through a grating, and is split into different beams. The beams are then reflected from an object in the device's field of view (FOV) and captured by an infra-red sensor, making it possible to calculate the distance of the object using triangulation [26]. The range of the depth sensor is 2.3-20 ft. (restricted to 13 ft. by the SDK) The microphone array is comprised of 4 microphones, enabling sound source localization. For our application, we implemented the least intrusive setup possible by capturing data only from the depth sensor and the microphone array, without capturing the actual color video data.

The RFID system we have used is the commercially available Alien 9900+ developer kit, which includes a reader with two circularly polarized antennas. The tags used in our experiment are EPC Class 1 Generation 2 supported by the 9900 readers. Figure 3 shows an example tag and antenna design from Alien. As the antennas are circularly polarized, the tag orientation is not an issue for our experiment. However, for an indoor environment, the antenna read range for the passive RFID tags varies from 20 to 30 ft. Such a read range is sufficient to detect the presence of a person carrying a tag in the simulated rooms of our Heracleia Assistive Apartment, given the tags are within the FOS of the antennas.

### B. System Architecture

The architecture of our system is modular, comprising of 3 main components as shown in figure 4.

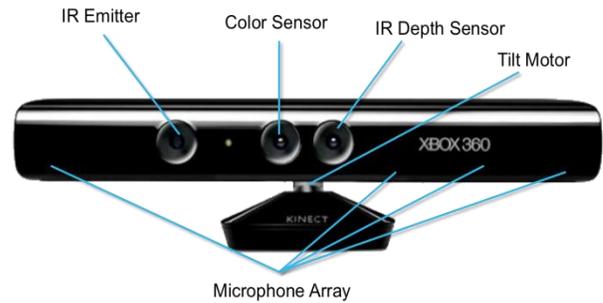


Fig. 2. The MS Kinect and the location of its various sensors and components.

Communication between the modules is based around the Joint Architecture for Unmanned Systems (JAUS) [27], originally developed by the U.S. Department of Defense, to govern the way that unmanned systems are designed. The user datagram protocol (UDP) is used for inter-module communications, which increases the level of interoperability, allowing new software modules to be easily integrated in the system or existing modules to be installed on different systems. Input is provided by the RFID reader and the 2 Kinect devices. One of them is considered as primary, capturing both a stream of depth images and audio, while the secondary captures only audio for performing sound localization. Interfacing with the Kinect is carried out using the MS software development kit (SDK) v1.0 [28]. The 3 modules 1) skeletal tracking based localization, 2) audio localization and 3)RFID tracking are described in detail in the following paragraphs.



Fig. 3. RFID transceiver equipment used and example Alien RFID Tag and Antenna Designs.

### 1) Skeletal Tracking Based Localization Module

Skeletal tracking is used in our system in order to detect and track a person in the FOV of the sensor, as s/he moves in the smart space and it was implemented using the MS Kinect SDK. This module has been explained in our previous work [29] and therefore only briefly described.

When a person is detected to be moving, her/his center of mass is determined and a skeletal model is fitted. The detected skeleton has a unique identifier for a specific session and is defined by the 3-D coordinates of its 20 joints  $\langle X_{di}, Y_{di}, Z_{di} \rangle$ , expressed in meters.

Each joint can be at any of the three associated states: 1) tracked, 2) not-tracked and 3) inferred. Furthermore, two kinds of filters are applied to the joint coordinates due to the nature of the captured data, 1) high frequency jitter and 2) temporary spikes rejection. Localization using such skeletal tracking is very accurate and unintrusive since we only utilize the coordinates calculated from the depth sensor feed.

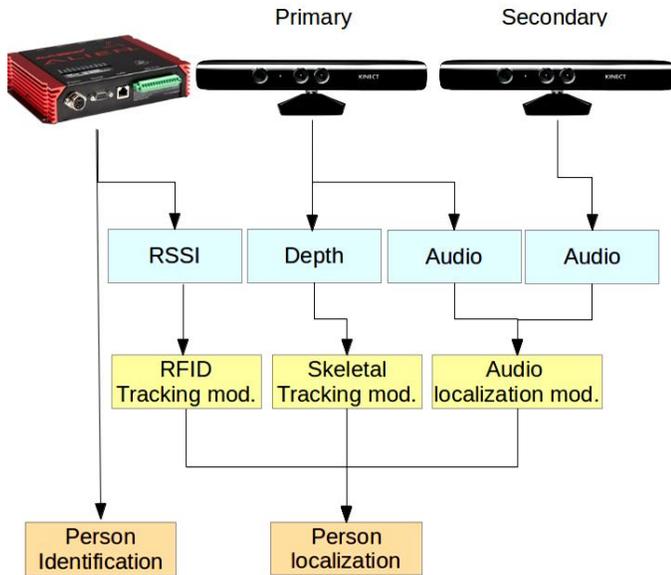


Fig. 4. System architecture showing the devices used (RFID and Kinect sensors), as well as the different types of information captured and processing pipelines for the identification and localization tasks.

### 2) Audio Localization Module

The audio localization module, acts as an auxiliary form of information input. The operation of the module has been described in detail in [29] and it is presented here epigrammatically. The Kinect incorporates a microphone array, comprised of 24-bit ADCs driven by 4 microphones. The frequency response of the microphones is tuned appropriately for speech and their directionality is isotropic for these frequencies. Sound source localization is applied to the audio signal in order to determine the angle of the sound source in relation to the device and acquire the audio signal from that particular direction. The returned values are the sound source angle (in degrees) in relation to the axis that is perpendicular to the device, and a confidence level of the reported angle.

Although the angle of the sound source can be acquired, this information is inadequate in estimating the source's distance. Thus, a second Kinect is introduced, used only for sound source localization (figure 5). The additional information provided by this unit can be used for accurate 2-D localization through triangulation. We denote  $\theta_A, \theta_B$  the angles between the wall and the axes perpendicular to devices A and B respectively and assuming there is a sound source S detected by the two devices, let the corresponding detected angles be  $\phi_A, \phi_B \in (-50, 50)$ . We consider the triangle that is created, with A, S and B as its vertices such that the altitude of the triangle passing from vertex S, divides L into a and b so that  $a+b=L$ . Let the length of the altitude (in our case the distance of

the audio source/person from the wall) be  $X_s$ . Since  $L=a+b$ , the final solution to the system of equations is given by:

$$X_s = \frac{\tan(\theta_A - \phi_A) \cdot \tan(\theta_B + \phi_B) \cdot L}{\tan(\theta_A - \phi_A) + \tan(\theta_B + \phi_B)}$$

$$a = \frac{X_s}{\tan(\theta_A - \phi_A)}$$

$$b = \frac{X_s}{\tan(\theta_B + \phi_B)}$$

This method allows for the calculation of the precise 2-D position of the audio source in the room.

Some additional restrictions concerning this setup were that the sound source angles are taken into account only when the sound level exceeds 50dB, the confidence for both estimated sound source angles is more than 50% and that there is a solution for the equation system and that this solution falls within the boundaries of the room.

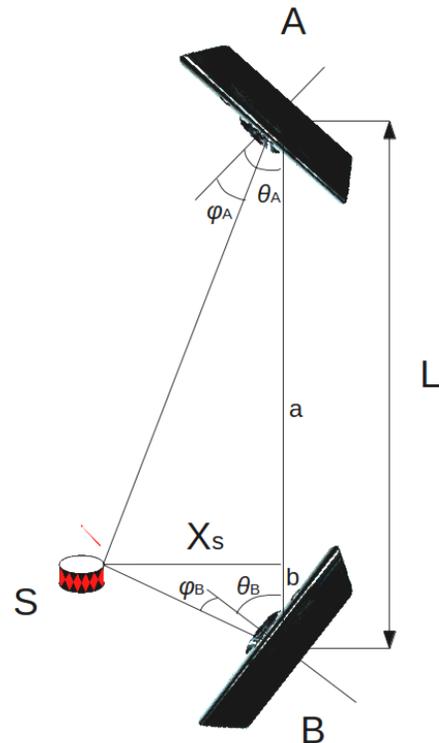


Fig. 5. Audio localization configuration utilizing the microphone arrays of 2 Kinect sensors.

### 3) RFID Based Localization Module

The RFID system that we used was comprised of two antennas and a tag reader. Its main role was to identify the person in its field of sense (FOS), but also to provide a rough estimate of her/his location using the received signal strength indicator (RSSI) from each antenna. The mapping between the RSSI values and the actual position of the tag is accomplished through a calibration process that accounts for both the

directionality of the antennas and the specific layout of the room. Multiple people are identified using their unique RFID tag and tracked as long as they remain in the FOS of the system. Skeletal tracking alone may not be able to discern between different people since a new tracking id is issued each time a person is lost from the FOV of the Kinect and then re-enters. Therefore, we improved our system's accuracy by matching the new RFID tag with the new tracking id as soon as an individual enters the room. This technique allows identification of each individual detected by the skeletal tracker. In the case where an unmatched tag id or skeletal id appears e.g. if a person was not detected upon entrance by either sensor, they are matched when they both appear in the same sector. Finally, when no skeleton is detected in the FOV, but a tag is still being detected, audio localization is utilized in order to increase accuracy (e.g. when only one antenna reads the tag).

Localization is based on a training phase during which statistical regression is applied on pre-specified position signatures (RSSI in our case) in order to build a classifier. More specifically, we divide the entire room into multiple sectors, as shown in figure 6. Next, we collect the RSSI signatures of the detected tags in these different sectors using the antennas. Given the measurement from the Kinect sensor for any particular person, if the measured location falls within that specific sector, then we map that particular person to the location described by the Kinect sensor. As afore-mentioned, in both approaches we use the sound from the microphone array as another modality besides skeletal tracking to resolve ambiguities in mapping.

### III. SYSTEM OPERATION

The main function of our system is person localization utilizing information from all three modules. The main source of location information is the skeletal tracking module. More specifically, this module detects a person as soon as s/he enters the FOV of the sensor and tracks her/him while moving in the room. The accuracy and robustness of the tracker is exceptional due to the nature of the depth sensor, so the person is tracked while standing, walking or even sitting.

We consider the location of the person as the average of the 3-D coordinates of all the tracked joints, expressed as  $\langle \bar{X}_d, \bar{Y}_d, \bar{Z}_d \rangle$ , where:

$$\bar{X}_d = \frac{1}{20} \sum_{i=1}^{20} X_{di} \quad \text{the mean distance from the sensor's plane.}$$

$$\bar{Y}_d = \frac{1}{20} \sum_{i=1}^{20} Y_{di} \quad \text{the mean deviation from the sensor's axis.}$$

$$\bar{Z}_d = \frac{1}{20} \sum_{i=1}^{20} Z_{di} \quad \text{the mean distance from the floor.}$$

Another source of location information is the audio localization module. It should be noted that the audio localization module is capable of estimating the location of the

person in 2 dimensions expressed by  $\langle X_s, a \rangle$ , not accounting for height, as described in the previous section.

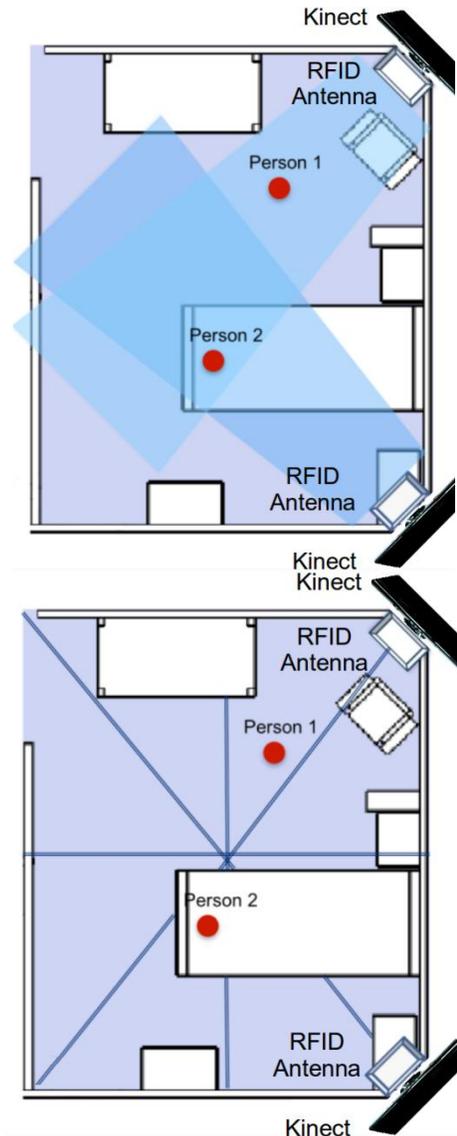


Fig. 6. Deployment setup combining the RFID RSSI and recognized sectors and the Kinect distance information for person localization.

In order to determine the final estimated location of the person we consider the available localization information from all three modules hierarchically, according to our experimental results presented in the next section. So, in the case where one of the modules does not return any coordinates, then the other module's coordinates are considered. The order in which we determine the location of each person is: 1) Skeletal tracker, 2) RFID, 3) Sound source localization. If skeletal tracking information becomes unavailable (e.g. if the person is outside the FOV of the depth sensor), then we rely on RFID. Similarly, if both skeletal tracking and RFID information are unavailable (e.g. tag undetected by 1 antenna), then sound source localization is used. In addition, we experimented by calculating the average location for each person.

More specifically, when a location estimate is available from both the RFID and the skeletal tracker, the average of each of the 2-D coordinates is calculated after proper transformation to match the 2 coordinate systems, while the third coordinate equals that of the skeletal tracking module. For our application, the detected activity is bound to the estimated location of the person. Therefore, if a person is standing by an appliance such as the oven or refrigerator we infer that s/he is using this particular appliance.

#### IV. EXPERIMENTAL SETUP

An extensive set of evaluation experiments were conducted in order to fine-tune the parameters of the setup at our simulated apartment (figure 7). As mentioned earlier, two Kinect devices and two RFID antennas were used, mounted at the opposite sides of one of the walls, facing the entrance. The distance between the two devices was 175.5 inches. The axis perpendicular to the sensors' axes pointed at 45 degrees towards the interior of the apartment, maximizing both the FOS, FOV and microphone coverage.



Fig. 7. An aspect of the Heracleia assistive apartment.

All modules were installed on the same computer, although our system's implementation permits the use of separate computers for each one of the modules. For our experiments we partitioned the space in 8 different sectors, intersecting at the center of the room. The estimated location of the person was considered accurate when the coordinates fell within the boundaries of the corresponding sector. For our application, the detected activity is bound to the estimated sector.

In our experimental setup, we have deployed two antennas at the two corners of the bedroom. We have simulated an experiment for identifying and localizing up to 4 people, limited only to part of the apartment, although the system can be extended to more rooms by adding more Kinect sensors in the apartment. During the experiment, each person wears an RFID tag around her/his neck.

We conducted extensive experiments in our realistic domestic setup. Four individuals participated in our experiments, with one, two or four occupying the apartment simultaneously. Subjects were asked to move in the apartment in 10 sessions and perform 4 activities, namely walk and sit in a chair, at a desk or on a bed. In table 1 we report results for

both the identification and localization tasks after 10-fold cross validation. For both tasks, accuracy degraded for more occupants, due to the people interacting and the resulting occlusions. Identification accuracy using RFID was at very high levels, considering single antenna misdetections. Localization accuracy denotes the percentage of correctly estimated locations for all individuals present in the room and also accounts for misidentifications and mismatches between the detected tag sector and skeletal id location. The accuracy attained using the Kinect was over 90%, and constituted the most accurate source for person location information. The accuracy achieved using RFID was over 80% and 75.9% using sound (only 1 speaker).

#### V. CONCLUSIONS

In this paper we presented the introduction of RFID technology to our existing novel person localization system improving its location estimation robustness and adding identification capabilities. Our system combines the tracking capabilities of the Kinect sensor with identification information from existing RFID technology. 3 types of data were used to solve the localization problem, namely the RSSI, 3D depth and audio information. Accurate position estimation for each person was carried out using the depth sensor and microphone arrays of the Kinect devices as inputs, by means of skeletal tracking and sound source localization respectively. The system was deployed in a simulated apartment and during the experiments conducted, it achieved high localization and identification accuracy for the 4-person localization scenario. More specifically, identification and localization accuracy always remained over 90% even in the 4 person scenario, when using information from RFID and the Kinect respectively. After confirming the effectiveness of our design we plan to extend it by utilizing depth and audio information from additional Kinect devices for increased robustness and coverage.

TABLE I. EXPERIMENTAL RESULTS FOR THE IDENTIFICATION AND LOCALIZATION TASKS FOR ALL THREE MODULES.

Task/Source	4 people	2 people	1 person
Identification/RFID	92.5%	97.1%	100%
Localization/Kinect	90.3%	93.8%	98%
Localization/RFID	82.1%	87.2%	85.4%
Localization/Sound	75.9% (1 speaker)		

#### ACKNOWLEDGMENTS

We would like to thank Shahina Ferdous for her invaluable help.

#### REFERENCES

- [1] T. Teixeira, D. Jung and A. Savvides, "Tasking networked cctv cameras and mobile phones to identify and localize multiple people", In Proc. Ubicomp, pp. 213-222, 2010.
- [2] R. Cucchiara, M. Fornaciari, A. Prati and P. Santinelli, "Mutual calibration of camera motes and RFIDs for people localization and identification", In Proc. ICDCS, 2010.

- [3] C. H. Kuo and R. Nevatia, "How does Person Identity Recognition Help Multi-Person Tracking?", In Proc. CVPR, pp. 1217-1224, 2011.
- [4] D. Mitzel, E. Horbert, A. Ess and B. Leibe, "Multi-person tracking with sparse detection and continuous segmentation", In Proc. ECCV, pp. 397-410, 2010.
- [5] M. Andriluka, S. Roth and B. Schiele, "People Tracking-by-Detection and People Detection by Tracking", In Proc. CVPR, pp. 1-8, 2008.
- [6] B. Wu, R. Nevatia, "Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Part Detectors", In Proc. IJCV, pp. 247-266, 2007.
- [7] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. H. and S. Shafer, "Multi-camera multi-person tracking for easy living", In Proc. IEEE IWVS, pp. 3-10, 2000.
- [8] D. Focken and R. Stiefelbogen, "Towards Vision-Based 3-D People Tracking in a Smart Room", In Proc. IEEE ICMI, pp. 400-405, 2002.
- [9] Y. Chang, R. Yan, D. Chen and J. Yang, "People identification with limited labels in privacy-protected video", In Proc. ICME, pp. 1005-1008, 2006.
- [10] B.S. Ashar and A. Ferriter, "Radiofrequency identification technology in health care: benefits and potential risks", JAMA: The Journal of the American Medical Association, 298:2305-2307, Nov. 2007.
- [11] K. Finkenzerler, "RFID handbook: fundamentals and applications in contactless smart cards and identification", Wiley, New York, NY, USA, 2003.
- [12] T. J. Hazen, E. Weinstein, B. Heisele, A. Park, and J. Ming, "Multimodal Face and Speaker Identification for Mobile Devices," Book Chapter in Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems, R. I. Hammoud, B. R. Abidi, and M. A. Abidi (Editors), pp. 123-138, Springer, 2007.
- [13] N. Fox, R. Gross, P. Chazal, J. Cohn, and R. Reilly, "Person Identification Using Automatic Integration of Speech, Lip and Face Experts," In ACM SIGMM 2003 Multimedia Biometrics Methods and Applications Workshop, Berkeley, CA, 2003, pp. 25-32.
- [14] H. K. Ekenel, M. Fischer, Q. Jin, and R. Stiefelbogen, "Multi-modal person identification in a smart environment," CVPR Biometrics Workshop, Minneapolis, USA, June 2007.
- [15] D. Hahnel, W. Burgard, D. Fox, K. Fishkin and M. Philipose, "Mapping and localization with RFID technology", In Proc. IEEE ICRA, pp.1015-1020, 2004.
- [16] M. Bouet and A. L. D. Santos, "RFID tags: Position principles and localization techniques", Wireless Day, pp. 1-5, 2008.
- [17] L. Ni, Y. Liu, Y. Cho Lau and A. Patil, "LANDMARC: Indoor location sensing using active RFID", Wireless Networks, pages 701-710, 2004.
- [18] P. Wilson, D. Prashanth and H. Aghajan, "Utilizing RFID signaling scheme for localization of stationary objects and speed estimation of mobile objects", In Proc. IEEE International Conference on RFID, pp. 94-99, 2007.
- [19] D. Hahnel, W. Burgard, D. Fox, K. Fishkin and M. Philipose, "Mapping and localization with RFID Technology", TR IRS-TR-03-014, Intel Research, 2003.
- [20] D. Joho, C. Plagemann and W. Burgard, "Modeling RFID signal strength and tag detection for localization and mapping", In Proc. IEEE ICRA, pp. 3160-3165, 2009.
- [21] M. M. Trivedi, K. S. Huang and I. Mikic, "Dynamic context capture and distributed video arrays for intelligent spaces", In IEEE Transactions on Systems: Man and Cybernetics - Part A: Systems and Humans, volume 35, January 2005.
- [22] L. Klingbeil and T. Wark, "A wireless sensor network for real-time indoor localization and motion monitoring", In Proc. International Conference on Information Processing in Sensor Networks, pp. 39-50, 2008.
- [23] J. A. Stankovic, et al, "Wireless sensor networks for in-home healthcare: potential and challenges", In Proc. HCMDSS, 2005.
- [24] L. Zhang and Z. Wang, "Integration of RFID into wireless sensor networks: Architectures, opportunities", In Proc. GCCW, pp. 463-469, 2006.
- [25] "The PrimeSensor Reference Design", [online] available at: <http://www.primesense.com/?p=514>
- [26] C. Liebe, C. Padgett, J. Chapsky, D. Wilson, K. Brown, S. Jerebets, H. Goldberg and J. Schroeder, "Spacecraft hazard avoidance utilizing structured light", In Proc. IEEE Aerospace Conference, pp.10, 2006.
- [27] S. Rowe and C. Wagner, "An Introduction to the Joint Architecture for Unmanned Systems (JAUS)", Technical Report, Cybernet Systems Corporation.
- [28] "The Microsoft Kinect SDK", [online] available at: <http://msdn.microsoft.com/en-us/library/hh855347.aspx>
- [29] G. Galatas, S. Ferdous and F. Makedon, "Multi-modal person localization and emergency detection using the Kinect", International Journal of Advanced Research in Artificial Intelligence, vol. 2 issue 1 pages 41-46.

#### AUTHORS PROFILE



Georgios Galatas is pursuing his PhD in Computer Engineering at the Heracleia Human-Centered Laboratory of the Computer Science and Engineering department of the University of Texas at Arlington. He is also a research fellow of the Institute of Informatics and Telecommunications of the National Center for Scientific Research "Demokritos". He received his combined bachelor's and master's degree from the Electrical and Computer Engineering department of the University of Patras in 2008. His research interests include Computer Vision, Digital Image Processing, Automatic Speech Recognition and Person Localization. He has co-authored several peer reviewed papers published in technical conferences and journals.



Filia Makedon is Distinguished Professor and Department Head of Computer Science and Engineering at the University of Texas at Arlington. She received her PhD in Computer Science from Northwestern University in 1982. Between 1991-2006, she was professor of computer science at Dartmouth College where she founded and directed the Dartmouth Experimental Visualization Laboratory (DEVLAB). Prof. Makedon has received many NSF research awards in the areas of trust management, data mining, parallel computing, visualization and knowledge management. She is author of over 300 peer-reviewed research publications. She directs the Heracleia Human-Centered Laboratory that develops assistive technologies for human monitoring and smart health. She is member of several journal editorial boards and chair of the annual PETRA conference ([www.petrae.org](http://www.petrae.org)).