IMPROVING INDOOR SECURITY SURVEILLANCE BY FUSING DATA FROM BIM, UWB AND VIDEO

M. Rafiee, H. Siddiqui, *A. Hammad,

Concordia Institute for Information Systems Engineering, Concordia University
1515 Saint Catherine Street West,
Montreal, QC, Canada H3G 2W1
*(*Corresponding author: hammad@ciise.concordia.ca)*

IMPROVING INDOOR SECURITY SURVEILLANCE BY FUSING DATA FROM BIM, UWB AND VIDEO

ABSTRACT

Indoor physical security, as a perpetual and multi-layered phenomenon, is a time-intensive and labor-consuming task. Various technologies have been leveraged to develop automatic access control, intrusion detection, or video monitoring systems. Video surveillance has been significantly enhanced by the advent of Pan-Tilt-Zoom (PTZ) cameras and advanced video processing, which together enable effective monitoring and recording. The development of ubiquitous object identification and tracking technologies provide the opportunity to accomplish automatic access control and tracking. Intrusion detection has also become possible through deploying networks of motion sensors for alerting about abnormal behaviors. However, each of the above-mentioned technologies has its own limitations. This paper presents a fully automated indoor security solution that leverages an Ultra-wideband (UWB) Real-Time Locating System (RTLS), PTZ surveillance cameras and a Building Information Model (BIM) as three sources of environmental data. Providing authorized persons with UWB tags, unauthorized intruders are distinguished as the mismatch observed between the detected tag owners and the persons detected in the video, and intrusion alert is generated. PTZ cameras allow for wide-area monitoring and motion-based recording. Furthermore, the BIM is used for space modeling and mapping the locations of intruders in the building. Fusing UWB tracking, video and spatial data can automate the entire security procedure from access control to intrusion alerting and behavior monitoring. Other benefits of the proposed method include higher accuracy and robustness, more complex query processing, and interoperability with other BIM-based solutions. A prototype system is implemented that demonstrates the feasibility of the proposed method.

KEYWORDS

Indoors Security, Data Fusion, BIM, UWB, PTZ Security Camera

INTRODUCTION

Physical security should be enforced in-depth, at least in four layers: (1) environmental design to deter threats (e.g. fences), (2) access control to restrict admission only to authorized people (historically by using mechanical keys and locks), (3) intrusion detection systems (e.g. motion sensors) to alarm suspicious behaviors for appropriate defensive response, and (4) identification and incident verification to prosecute the criminals (e.g. video surveillance). Despite the deployment of layer-specific automated security systems, human workforce is still intertwined in different tasks whether as patrols at checkpoints, administers of access control systems, alarm responders or video observers and analyzers (Wikipedia, 2012).

In an effort to increase automation level and accuracy degree of indoors security, this research targets the implementation of intruder tracking through the following steps: (1) intrusion detection, (2) intruder identification and locating, and (3) visual tracking in a hybrid system that leverages real-time complementary technologies to effectuate the involved tasks. The technologies are adopted based on the capability to detect, identify and locate the persons in a restricted area and finally to capture evidence of intruders in their movement path. Besides the automation level of the solution, which is determined by the degree of technology integration, the effectiveness and applicability of the proposed solution relies on its accurate and real-time response which is improved through data fusion among multiple sources.

Building Information Modeling (BIM) has gained considerable attention over other building modeling approaches by offering many advantages for digital design of facilities and a reliable basis for decision making. BIM, as a cumulative digital representation of physical and functional characteristics of building elements, is shared among different stakeholders at different phases of building life cycle (BIM, 2012).

Ultra-Wide Band (UWB) Real-Time Locating System (RTLS) is a technology for capturing and transmitting large amounts of data over a wide frequency spectrum using short-pulse, low powered radio signals (Kshetrimayum & IIT Guwahati, 2009). UWB has gained widespread use in indoors applications where typical RF-based technologies are unable to provide accurate result. Figure 1 shows the relative position of UWB and its distance from the ideal locating technology in terms of accuracy and coverage tradeoff. UWB being immune to the multipath problem and able to achieve high accuracy in localization is promising for wide area tracking applications (Muñoz, Bouchereau, Vargas, & Enriquez-Caldera, 2009). In contrast, the localization of the other RFID technologies which suffer from the multipath problem is not promising for wide area applications and can be suitable for chokepoint access control (tag identification and access granting to authenticated person).

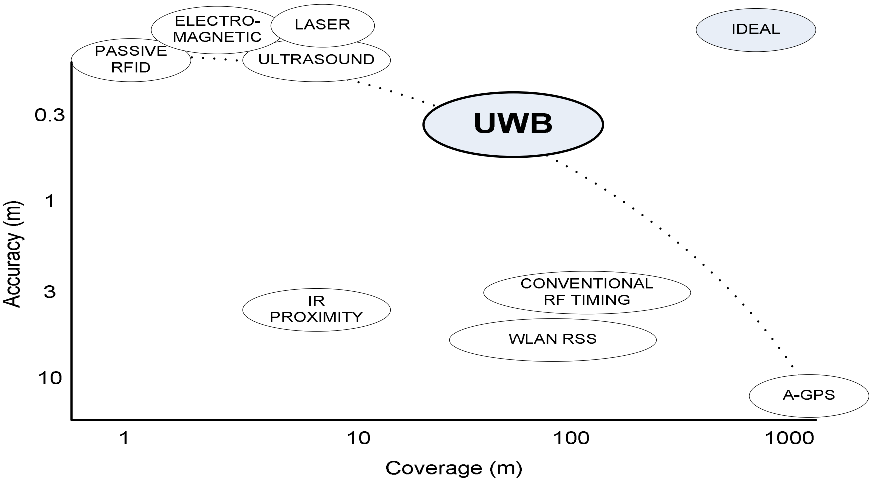

Figure 1 – Comparison of locating technologies **(Ward, 2007)**

Nonetheless, despite superiority of UWB to other RTLS technologies for indoor applications, still its accuracy is influenced by indoor conditions e.g. noise or other error sources such as instrument, measurement or inaccurate reference points (Sato, 2011).

On the other hand, visual target tracking in wide areas is doable either by several fixed cameras or a few number of PTZ cameras that should be managed by a sophisticated handover algorithm. Fixed cameras do not have the physical capability of rotating to follow a target. In order to detect abnormal behaviors, people must be tracked through multiple, mostly disjoint camera views. Tracking people by multiple cameras is not a trivial task due to the considerable changes in orientation and pose of people and lighting conditions between camera views that impose many challenges in target re-identification. However, continuous monitoring over the fixed Field of View (FoV) makes fixed cameras more suitable for small critical regions (e.g. gates or entrances) to not lose any incidences. PTZ cameras are extensively used in wide-area surveillance as they offer acceptable degree of continuity for tracking, thanks to pan and tilt rotations, whilst being able to capture highly detailed video from any area of interest within their FoV. PTZ cameras thereby contribute in simplifying the camera network operation and management by decreasing the number of required cameras for wide-area surveillance applications. Activity monitoring, tracking and behavior analysis in such environments require continuous coverage of the entire area (Davis, 2011). Furthermore, PTZ cameras are programmable for auto-tracking thanks to on-board intelligence (Sony SNC-ER580, 2012).

The video captured from entrance can be analyzed for counting people and be used for intrusion detection and intruder identification. However, accuracy of video analytics is affected by time-varying or bad lighting conditions addressed in (Prosser, Gong, & Xiang, 2008), (Chao-Ho, Tsong-Yi, Je-Ching, & Da-Jinn, 2011), (Loy, Xiang, & Gong, 2009), occlusion (two persons or a person and an obstacle occluding each other, partially or totally), bad resolution or tempering with camera (e.g. spray-painting the lens). Infra Red (IR) night vision helps PTZ cameras to overcome low light conditions.

The coordinates of identified intruders can be computed as inputs for PTZ auto-tracking applications. The powerful optical zoom of PTZ cameras provides high degree of details from relatively distant points from mobile targets. Therefore, the recording of PTZ cameras can be analyzed for detecting abnormal and suspicious behavior over space. However, PTZ operations that achieve a wider dynamic FoV result in losing some parts of the picture. Thus, PTZ cameras are not suitable for monitoring entrances to count people and detect intrusions, as they may lose some occurrences while being rotated or zoomed.

With respect to the diversities of RTLS technologies and camera types, different sensor combinations are possible. Considering aforementioned comparisons among UWB and other RF-based technologies, six possible combinations of camera and RTLS are summarized in Table 1along with key functionalities of each combination.

Although numerous works have been done on fusing radio data from RTLS technologies with video data from surveillance camera networks, none of them leverage a geospatial data source which contributes in the deployment, monitoring and tracking enhancement. To the best of our knowledge the closest work to our research has been done by (Massimiliano, et al., 2011) to fuse UWB and video data for reliable and robust tracking services and more accurate context understanding. Their experimental results demonstrated coarse matching level of their prototype system. Another proof-of concept application named RVid fuses UWB and computer vision location data for tracking people in multiple scenes (Tronci, et al., 2012).

With respect to the abovementioned arguments, this research focuses on the integration of UWB RTLS and a single PTZ camera for wide area intruder tracking. However, future work can extend our proposed architecture to a network of fixed and PTZ cameras that handle wider areas and more complicated scenarios. The proposed framework in this research leverages three environmental data sources, one for space modeling and location data validation, two for real-time complementary data sensing that enable distinguishing and locating intruders for visual tracking. The framework consists of: (1) a BIM, (2) UWB sensor network, and (3) PTZ camera.

The rest of this paper is organized as follows. Next section introduces the concept and system design followed by the operational requirements. Then the proposed methodology is described through some flowcharts. A case study that validates the feasibility and effectiveness of the proposed system is followed by the system's advantages, limitations and future works. A brief conclusion section summarizes our discussions in this paper.

Table 1 – Technology Combinations, Capabilities and Shortcoming

			Camera Types		
			Fixed Camera	**PTZ Camera**	**Fixed + PTZ Cameras**
			Continuous observance over a limited FoV (Suitable for lossless-data applications)	Wider FoV achievable in exchange of loss of continuity during camera rotations (Suitable for target tracking & behavior analysis)	Continuous observance over critical & confined areas (e.g. entrance) with fixed cameras + wide-area on-demand surveillance with PTZ cameras
RF-based Technologies For RTLS	**Traditional RFIDs**	Tag detection & identification + chokepoint locating	Step 1, 2 and 3 are infeasible across wide areas	Although PTZ camera is available for performing step 3, steps 1 and 2 are infeasible across wide areas	Although PTZ camera is available for performing step 3, steps 1 and 2 are infeasible across wide areas
	UWB	RFID capabilities + precise wider range locating	Step 3 is doable only within a very limited area	Step 1, 2 and 3 are feasible	Fixed camera is assigned to collaborate with UWB in steps 1 and 2. PTZ camera is assigned to step 3

CONCEPT AND SYSTEM DESIGN

Figure 2 demonstrates the architecture of the proposed system composed of the following modules:

(1) BIM Module as a visual decision support tool for optimum deployment of sensor networks as well as operation control to filter out noisy location data and assure accuracy of the inferences.
(2) UWB Module for the detection, identification and location update of the authorized persons who have been registered in the system with a unique tag ID and have been requested to wear the pre-assigned UWB tag before entering the restricted area.
(3) Video Module for expanding detection ability of the system from tagged persons to untagged unauthorized persons through running body detection and locating algorithms, as well as PTZ auto-tracking of the intruders.
(4) Data Fusion Module that infers presence of intruders through comparing UWB and video person counts, distinguishes intruders who are untagged among other tagged persons, calculates location of intruders from video data and maps it to real-world coordinates and

other processing such as processing more complex queries across data sources (XQueries). XQuery processing refers to the function of searching each application database for matching data and generating the combined results into a formatted report. A quick and efficient way is necessary for locating matched data from all applications and interfaces utilizing a single query (Cross-query, 2013).

(5) GUI as a medium to represent fused environmental data to users and to input settings, commands and queries.

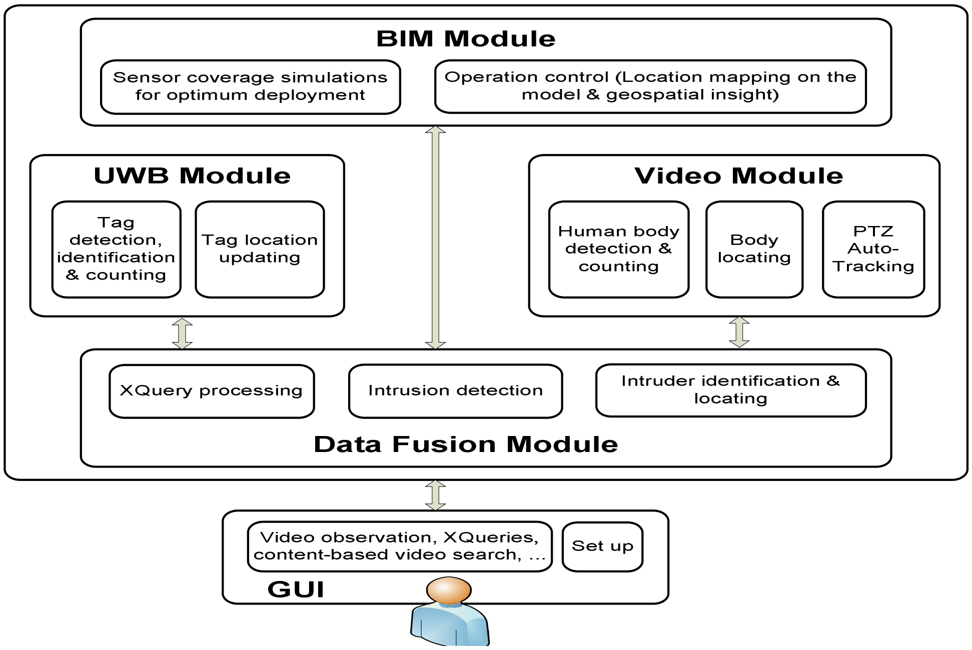

Figure 2 – Architecture of the proposed system

Fused data can be stored for further analysis or as forensics evidence, i.e. linked data storage for inferring behavior. Video-recognized events can be linked with related UWB data while being archived to enable processing cross-data inquiries (Kolias, et al., 2010). Indexing video frames by UWB tagIDs present at each correlated timestamp enables processing XQueries. An appropriate data fusion method can assist processing of XQueries over radio, video and geospatial data such as:

- Which authorized people were present, while event “x” was occurring?
- When exactly event “x” has happened?
- Retrieve video footage from time t_1 to t_2 that tagID = “e” was detected in restricted area.

The followings are the operational requirements of the proposed system:

(1) Authorized persons must be registered in the RTLS by providing them with UWB tags which they are supposed to attach to themselves when entering the restricted area.

(2) Intrusion detection is only achievable in the areas covered simultaneously with the RTLS and video cameras. In partially overlapped areas, depending on the available modules some functionalities of the proposed system cannot be performed.

(3) The update rate of the UWB tags (Hz), must be (optimally) equal or as close as possible to the video update rate (fps) for correct data correlation during data fusion. Maximum frame recording rate varies depending on the camera and its different codec types, for example Sony ER-580

provides maximum of 30 fps for H.264 codec, 20 fps for MPEG-4 and 16 fps for JPEG (Sony SNC-ER580, 2012). RTLS platforms usually have controllable update rates for UWB tags. For example Ubisense provides as fast as 40 Hz and as slow as one update every 14 minutes (Ubisense RTLS, 2012).

METHODOLOGY

In this section, the proposed methodology for performing the three aforementioned tasks are described and illustrated in some flowcharts. The real-time coordinates of the detected individuals (including x and y coordinates; the z coordinate is not necessary for security applications) are collected in two 2D vectors which are defined in the Euclidean space; named: r_t for RTLS reported data at time t and v_t for video extracted data at time t. p_{it}^{r} corresponds to the position coordinates of i^{th} element of vector r_t, where $i=1,\ldots,n$, and p_{jt}^{v} corresponds to the coordinates of j^{th} element of vector v_t, where $j=1,\ldots,m$.

Figure 3 illustrates how the system detects intrusions in a near real-time manner. Sensor modules actively percept events within their overlapping coverage and provide the fusion module with the number of the detected persons together with their coordinates. The fusion module determines whether an intrusion has been occurred or not by comparing the elements of r_t and v_t and generates an alert if an intrusion occurred. We propose adding a BIM to the framework as the third source of environmental data. BIM can provide an accurate basis for mapping sensor-measured locations into real-world coordinates and rejecting noisy data that result in false intrusion alerts. For example if a tag owner who is inside a room and is detected in video, is positioned outside the boundaries of the room by the RTLS, he would not be counted. This will lead to having an incorrect smaller number of people from RTLS, whereas the camera sees this person and gives a number that is bigger than RTLS counter and results in false deduction about intrusion although there is no intruder in the restricted room. Moreover, the BIM including privacy levels of rooms together with a list of authorized persons for each room helps in disambiguating authorization-room relationships; for example a person who is only authorized for the room R_1 will be detected upon entrance to the room R_2.

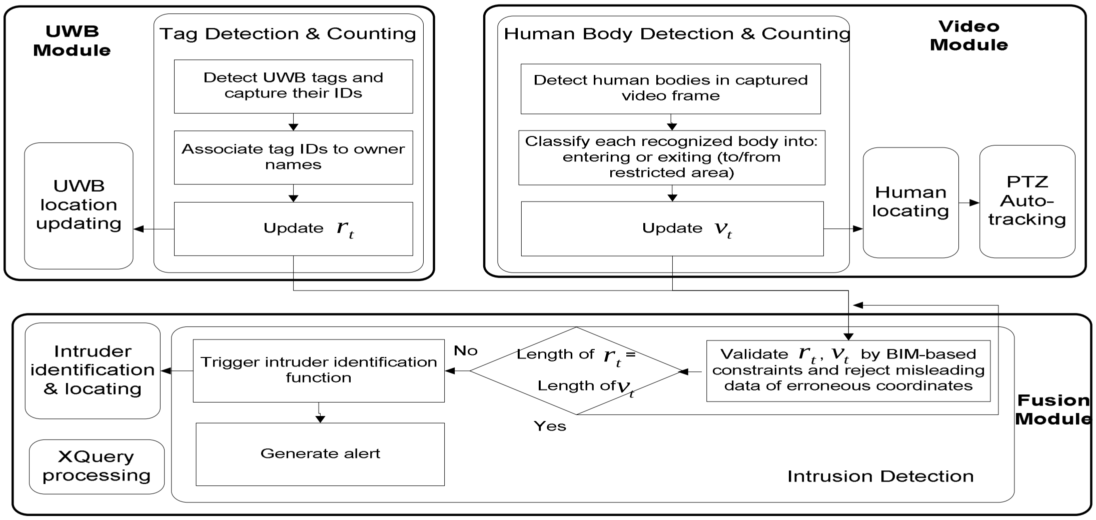

Figure 3 – Intrusion detection and alert generation flowchart

Figure 4 illustrates intruder identification and locating function which is triggered in case of intrusions to resume processing for achieving the intruder tracking goal.

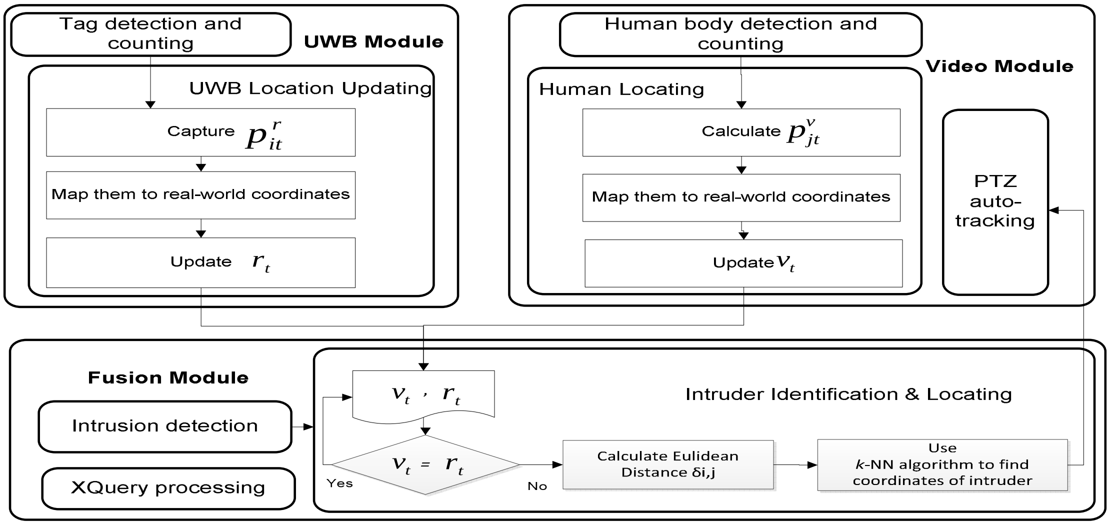

Figure 4 – Intruder identification and locating via fusion module

Euclidean distances ($\delta_{i,j}$) between p_{it}^r and p_{jt}^v are calculated (Equation 1) and saved in a metric space *Eud*. Then nearest elements are searched, using *k-nearest neighbor algorithm* (*k*-NN) (Li & Cheng, 2009; Chávez et al., 2001), to find corresponding p_{it}^r for each p_{jt}^v. The unmatched video vector element p_{jt}^v contains the coordinates of the intruder, as shown in Figure 5, and is passed to the video module in order to track the intruder.

$$\delta_{i,j} = \sqrt{\left(p_{it}^r(x) - p_{jt}^v(x)\right)^2 + \left(p_{it}^r(y) - p_{jt}^v(y)\right)^2} \tag{1}$$

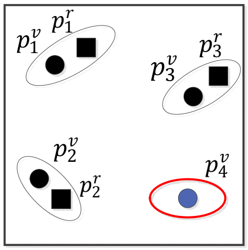

Figure 5 – Intruder identification through *k*-NN

The auto-tracking function converts input (real-world) coordinates to camera's internal rotation parameters, i.e. pan and tilt angles, as shown in Figure 6. Powerful zooming capability of PTZ camera enables capturing high resolution video of the intruder while he is being tracked along his moving path, and this video is recorded as authentic forensics evidence providing enough details of the appearance of the intruder.

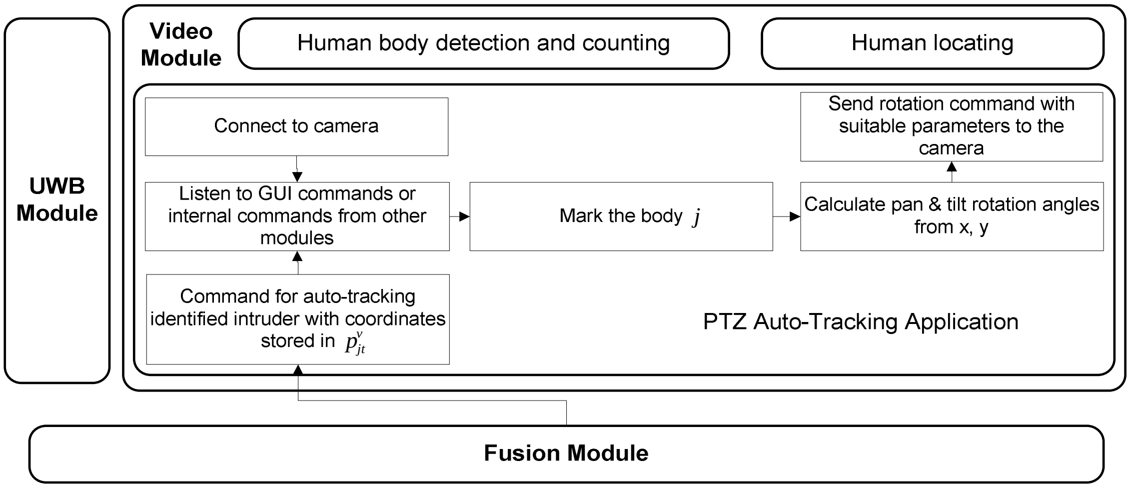

Figure 6 – Intruder auto-tracking via PTZ camera

CASE STUDY

An initial case study has been designed to evaluate the feasibility of the proposed architecture and the reliability of the fused data. Tests have been conducted in a lab at Concordia Institute for Information Systems Engineering, with the dimensions of 7.32 m × 4.16 m × 3 m. Four Ubisense UWB Sensors (Ubisense RTLS, 2012) have been installed at the four corners of the lab and one Sony PTZ Camera (Sony SNC-ER580, 2012) is mounted at the ceiling of the room as shown in Figure 7. The mounted camera covers only a part of the UWB sensors cell in its range of rotation; e.g. when it is posed in a vertical position, the coverage field is 1.57 m × 2.87 m. We tested a scenario with two tagged and one untagged persons in the room. An image was captured when the three persons were in the camera's FoV. EmguCV Open Library (EmguCV, 2013) was used for detecting the persons from the image. Location of their periphery centers are extracted to be compared and fused with the UWB data. The origin of the image is at point $O(2.17,6.36)$ in the room (xy) Coordinate System. Having video image size of 640 by 480 pixels and the above-mentioned FoV, the pixel dimensions can be calculated; w_p for pixel width and h_p for pixel height. Pixel size (w_p, h_p) is used in translating coordinates of detected bodies from image (XY) coordinate system into the room (xy) coordinates (Equations 2a and 2b).

$$x = y_{\mathrm{O}} - \mathrm{Y} \times w_p \tag{2a}$$

$$\mathrm{y} = x_{\mathrm{O}} - \mathrm{X} \times h_p \tag{2b}$$

The Video locations of the three persons were calculated as $p_{1t}^{v} = (1.52, 4.29)$, $p_{2t}^{v} = (1.90, 4.21)$ and $p_{3t}^{v} = (1.17, 5.40)$, and the RTLS locations of the two tagged persons were captured as $p_{1t}^{r} = (1.40, 4.74)$ and $p_{2t}^{r} = (1.97, 4.31)$. These locations were then provided as input to the fusion module implemented in MATLAB using *k*-NN algorithm, which then provided the coordinates of intruder, which is p_3 in our case.

CONCLUSIONS AND FUTURE WORK

This paper presents a framework for tracking an intruder across wide indoor areas by applying data fusion over detection and localization results of UWB RTLS and PTZ video cameras. The introduced framework takes advantage of BIM for removing noisy location data and better control over access authority lists and space restrictions. The proposed approach requires authorized people to be registered in the system with pre-assigned UWB tags in order to wear them before entering the restricted area. It also requires simultaneous UWB and video captures from a completely overlapped coverage field to enable comparing data to deduce intrusion and intruders. Data fusion also brings added-values for the proposed multi-sensory system; e.g. XQuery processing, content-based video retrieval, thanks to video indexing and linked data storage. Nevertheless, the proposed design is still influenced by environmental conditions and data fusion efficiency that demands further research work and enhancement. A single PTZ camera is not sufficient for lossless monitoring and background modeling of the area, this imposes some limitations for 3D locating, multi-target tracking, and fast moving objects. These shortcomings can be alleviated by leveraging more cameras. The proposed architecture is also enabler of other applications than intruder tracking, including: Safety applications such as healthcare or elderly care, workers' safety in risky environments, inventorying (Kolias, et al., 2010) Ambient Intelligence scenarios (Massimiliano, et al., 2011), etc.

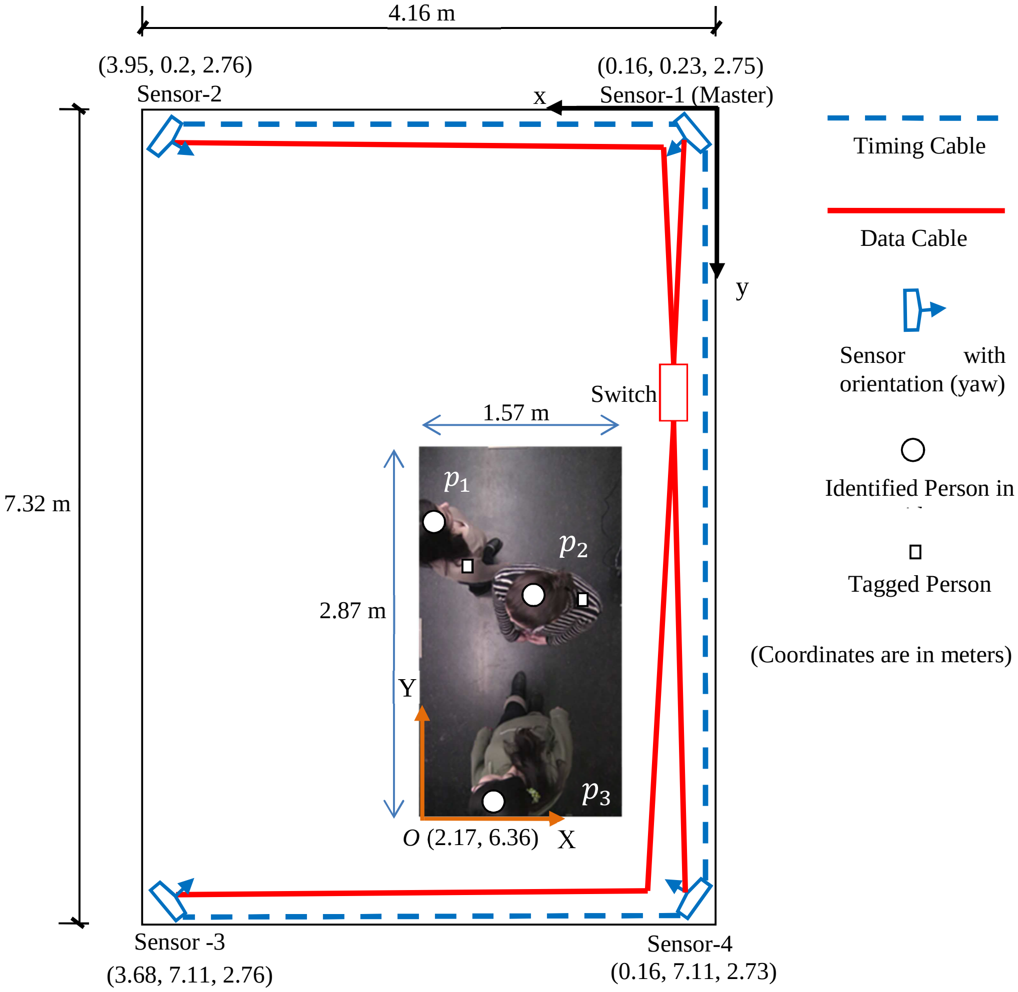

Figure 7 – Settings and coverage of UWB sensors and camera

REFERENCES

BIM. (2012). Retrieved November 20, 2012, from Open Geospatial: http://www.opengeospatial.org/ogc/markets-technologies/bim

Chao-Ho, C., Tsong-Yi, C., Je-Ching, L., & Da-Jinn, W. (2011). People Tracking in the Multi-camera Surveillance System. *Second International Conference on Innovations in Bio-inspired Computing and Applications* (pp. 1 - 4). Kaohsiung, Taiwan: IEEE.

Chávez, E., Navarro, G., Baeza-Yates, R., & Marroquín, J. (2001). Searching in Metric Spaces. *ACM Computing Surveys, Vol. 33, No. 3*, (pp. 273–321).

Cross-query. (2013). Retrieved January 9, 2013, from noris: http://www.noris.org/multi-agency/

Davis, K. S. (2011). PTZ Camera Modeling and Panoramic View Generation via Focal Plane Mapping. In *Computer Vision – ACCV 2010.* Springer Berlin / Heidelberg.

EmguCV. (2013). Retrieved January 1, 2013, from Emgu: www.emgu.com

Kolias, V., Giannoukos, I., Anagnostopoulos, C., Anagnostopoulos, I., Loumos, V., & Kayafas, E. (2010). Integrating RFID on Event-based Hemispheric Imaging for Internet of Things Assistive Applications. *PErvasive Technologies Related to Assistive Environments.* New York, NY, USA: ACM.

Kshetrimayum, R., & IIT Guwahati, G. (2009). An introduction to UWB communication systems. *IEEE, 28*(2), 9 - 13.

Li, Y., & Cheng, B. (2009). An Improved k-Nearest Neighbor Algorithm and Its Application to High Resolution Remote Sensing Image Classification. *17th International Conference on Geoinformatics.* Fairfax, VA, USA: IEEE.

Loy, C., Xiang, T., & Gong, S. (2009). Modelling activity global temporal dependencies using Time Delayed Probabilistic Graphical Model. *Computer Vision, 2009 IEEE 12th International Conference* (pp. 120 - 127). London, UK: IEEE.

Massimiliano, D., Antonio, B., Gian, M., Daniele, M., Carlo, M., & Fabio, R. (2011). Fusion of Radio and Video Localization for People Tracking. *AmI'11 Proceedings of the Second international conference on Ambient Intelligence.* Springer.

Muñoz, D., Bouchereau, F., Vargas, C., & Enriquez-Caldera, R. (2009). *Position Location Techniques and Applications.* Burligton, MA: Academic Press.

Prosser, B., Gong, S., & Xiang, T. (2008). Multi-camera Matching under Illumination Change Over Time. London: Queen Mary University of London.

Sato, Y. (2011). Location-based Responsive Space for Multimedia Environment Using Ultra-wideband Technology. *Masters Thesis*. Montreal, QC, Canada: Concordia University.

Sony SNC-ER580. (2012). Retrieved January 9, 2013, from Sony: http://pro.sony-asia.com/product/resources/en_BA/images/Brochure/Security/SNC-ER580.pdf

Tronci, R., Murgia, G., Pili, M., Ristori, M., Sirena, N., & Roli, F. (2012). *The Ambient Intelligence Laboratory Of The Sardegna District.* Pula - Cagliari (CA), Italy. Retrieved January 9, 2013, from http://prag.diee.unica.it/pra/system/files/amilab-final.pdf

Ubisense RTLS. (2012). Retrieved January 9, 2013, from Ubisense: http://www.ubisense.net/en/resources/factsheets/ubisense-real-time-location.html

Ward, A. (2007). In-building location systems . *Location Technologies, 2007. The Institution of Engineering and Technology Seminar* (pp. 1 - 18). IEEE.

Wikipedia. (2012). Retrieved December 10, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Physical_security