Smart Monitoring of Complex Public Scenes

Collaboration between human guards, security network and robotic platforms

L. Iocchi[1], D. N. Monekosso[2], D. Nardi[1], M. Nicolescu[3], P. Remagnino[4], M. Valera Espina[4]

1 University of Rome "La Sapienza", Italy

2 University of Ulster, UK

3 University of Nevada, USA

4 Kingston University, UK

Abstract

Security operators are increasingly interested in solutions that can provide an automatic understanding of potentially crowded public environments. In this paper, an on-going research is presented, on building a complex system consists of three main components: human security operators carrying sensors, mobile robotic platforms carrying sensors and network of fixed sensors (i.e. cameras) installed in the environment. The main objectives of this research are: 1) to develop models and solutions for an intelligent integration of sensorial information coming from different sources, 2) to develop effective human-robot interaction methods in the paradigm multi-human vs. multi-robot, 3) to integrate all these components in a system that allows for robust and efficient coordination among robots, vision sensors and human guards, in order to enhance surveillance in crowded public environments.

Introduction

The aim of the research presented in this paper is to study a surveillance system composed of human security operators managing different devices such as mobile phones, PDAs and desktop PCs. It is also composed of a network of fixed sensors such as cameras and mobile robotic platforms carrying sensors. The guards, mobile robots and operators at a control room should communicate by means of a wireless network infrastructure, to which different devices such as mobile phones, PDAs, desktop PCs and computational units onboard the robots are also connected. The proposed solution is intended to be modular and considers the integration of the following approaches in a hierarchical architecture (see Figure1):

- **Sensor networks**. Sensor nodes are embedded throughout the environment, and are interconnected by a wireless network. Each node comprises a number of sensors and sufficient local processing capability to carry out onboard the required processing.
- **Computer vision for scene understanding**. The proposed system relies on implementing techniques for robust dynamic scene interpretation in complex, crowded environments to trigger signals to the reactive components of the system (i.e. robots and guards).
- **Multi robot systems**. Mobile platforms are very useful in a surveillance task, since they provide mobility to intelligent sensors.
- **Localization of security personnel.** The position of mobile nodes (including human guards) must be known to the command centre. Position estimation of the personnel will be achieved with a Radio Frequency Identification (RFID) system**.**

The paper is structured as follows: next section describes related work on the components listed above to be integrated to build such autonomous surveillance systems. After this section, the architecture of the proposed system is discussed. Then, the fusion of information coming from different sources is described, following a brief discussion section on human-robot interaction methods. In the next section, a case study where an integration of a fixed node (e.g. camera) and a PDA device is described. In the last section the conclusions and future work are presented.

Related work

The literature and survey in video-surveillance systems is large (Raty 2010)(Valera, Velastin 2005).Traditionally, these systems were based on static sensor devices such as CCTV cameras and later on, smart cameras were used, as computer vision algorithms were able to be embedded on these sensors. The image processing in video-surveillance systems mainly consists in video analysis of the monitoring area. The analysis may be used to interpret stationary objects and people of a scene (Harveille, Dalong 2004)(Zhao et al. 2005) or to interpret dynamic scenes (Fusier et al. 2006). The analysis of dynamic scenes are based on the motion estimation or tracking people (Munoz, Salinas 2009) to perform behavioral analysis or activity recognition. (Buxton,Gong 2002) presents a survey of the state-of-the-art on learning and understanding scene activity.

Lately, cooperative multi-robot solutions (MRS) for surveillance have been proposed (Burgard et al. 2000)(Zhang,Chen, and Xi 2005). MRS provide the advantage of having many robots being distributed within the available space and carrying out different tasks at the same time. However, they cannot be simply regarded as a generalization of the single robot case, therefore the proposed approaches need to be more precisely characterized in terms of assumptions about the environment and in terms of the internal system organization (Iocchi, Nardi, and Salerno 2001). The main challenges that MRS face are the fact that the monitoring area is unstructured and it is shared by humans, rising new challenges such as safety issues due to human interaction (users or security personnel of the system). The effectiveness of a cooperative mixed team formed by many humans and many robots will strongly depend on the effectiveness of the human-machine interface. The data have to be presented to the operators in easy-to-understand and intuitive way. The operators have to be able to quickly “grasp” the situation, to immediately assess the situation, without be overwhelmed by unnecessary information.

Besides the human-robot interaction, to also enhance the monitoring activity on these systems, the security personnel, carrying one or more computing devices containing devices, should be considered as a proactive node on the system. To localize accurately each of those proactive nodes in the system, a Global position Systems (GPS) solution is used in outdoors. However, for indoors applications, localization is less readily achievable. A number of RFID-based localization schemes have been proposed using active or passive tags; these include (Stelzer, Pourvoyeur, and Fischer 2004) using lateration techniques for distance estimation, (Wang, Huang and Hung 2010) that use a method referred to as scene analysis (or received Signal Strength, RSS-based) and proximity (or constraint-based) methods (Tesoriero et al. 2010). Recent trends are towards combining methods to improve accuracy (Hatthasin et al. 2009). In the system described in this paper RSS-based position estimation is proposed for localization of the proactive nodes. This method exploits the fact that an RF signal attenuates as it travels.

System Architecture

The research presented in this paper aims to create a heterogeneous intelligent multi-sensor system to monitor, understand and interpret complex public environments. Therefore, the goal is the use of multi-sensor gathering technology to extract automatically and on demand, information of interest; top-down through a central directive (e.g. behaviour analysis of a suspicious person or the analysis of video related to unattended luggage) or bottom-up prompted by information extracted and analysed from the leaf nodes of the system. Figure 1 illustrates the composition of these leaf nodes.

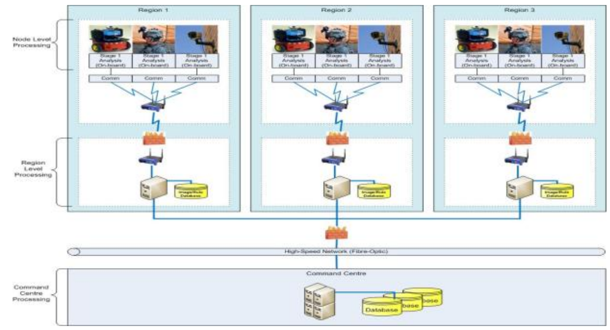

Figure 1. Graphical architecture of the system.

To build a heterogeneous multi-sensor surveillance system there are three main concerns that should be analysed: the design of architecture of the system, the network infrastructure to be able to communicate all the resources available in the system, and the integration of all the data information to enhance the monitoring activity of the system. Figure 2 and Figure 3 shows the system's design architecture, using the solution presented in (Quigley et al. 2009)(Valera, Velastin 2003). The architecture approach is based on a middleware architecture concept; any active node (i.e. sensor) in the system may join a service that suits its purpose (see Figure 2). All services are published on the entire system and in terms of software computing, these services can process the signals provided by the sensors. The communication between active nodes and the services is peer-to-peer and it is completely transparent to them.

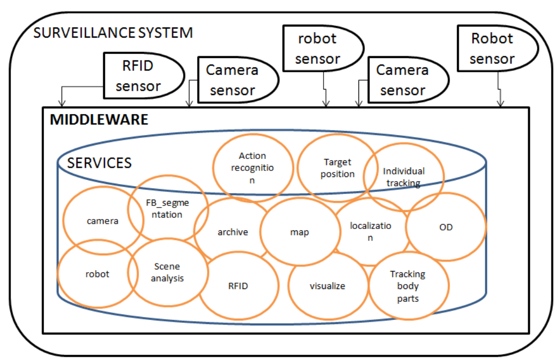

Figure 2. Architecture design of the system.

Notice, that the solution presented in Figure 2, is intended to be distributed as each sensor should be completely autonomous and there is a completely decoupling between active nodes and services. However, the system should create a pervasive layer where information is integrated to automatically interpret a complex and cluttered scene, potentially frequented by a large number of people (see scene analysis in Figure 3).

As mentioned, wireless network technology will be used to manage the communication of such different technology sensors. Therefore, the following issues: routing protocols, security and QoS have to be analyzed. However, the thorough discussion of these three main topics is out of the scope of this paper, as the authors focussed on the description of the components of the system.

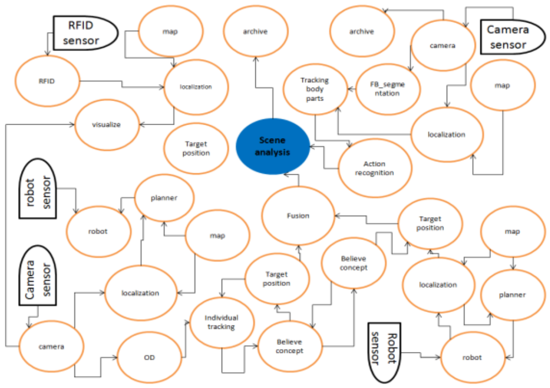

Figure 3. Different services provided by the middleware and the communication with the sensors in the system.

Autonomous multi-robot systems and human-robot interaction

The use of mobile platforms carrying sensors is an important component of the described system. This introduces two issues that must be considered: 1) the robots in the system must act as an autonomous coordinated team: no direct supervision from human should be required; 2) humans (i.e., security personnel non expert in robotics) must be able to interact with robots in a natural way. In the remaining of this section we briefly present the techniques used for autonomous coordinated multi-robot systems and human-robot interaction for security robots.

The main tasks for a team of surveillance robots are: *patrolling*, i.e., the task of continuously visiting relevant locations of the environment where information have to be gathered; *threat response*, that is any specific set of actions needed as a consequence of the detection of a threat. Patrolling can be either passive or active. Passive patrolling is executed without any information from other components of the system, while active patrolling is driven by some specific request for information gathering. In both cases, robots must be able to move in the environment safely and effectively (so standard robotics modules, like mapping, localization, navigation, obstacle avoidance, etc. are provided to the robots), and to act in a coordinated way, by taking into account dynamic task assignment (for example, which location has to be visited by each robot), as well as action synchronization (for example, when two or more robots are required to get combined information from the same source).Coordination techniques for multi-robot patrolling are described in (Iocchi, Nardi, and Salerno 2011), that include also an extensive experimental analysis showing that on-line coordinated behavior, in contrast with predefined off-line strategies, are fundamental for actual deployment of surveillance robots.

As mentioned, the second issue of surveillance robots is their way of interacting with human operators. Such operators need to interact in order to teach the robot about the environment (for example, which locations are important for the security mission), drive the robot to acquire new or additional information about a location or an event, receive the feedback from the user. These operators are expected to interact with robots in a natural way, without being required to have knowledge about the implementation of the system.

In (Randelli, Iocchi, Nardi 2011) we propose an architecture based on multi-modal user interfaces that allows a simple but effective form of human-robot interaction. Multi-modal user interaction include the combined use of a tangible device (e.g., a WiiMote controller), a pointing device (such as a laser pointer), laser and stereo-vision sensors on the robots to acquire metric information from the environment, and speech recognition (using a small portable device: Speaky[1]). With this tool effective knowledge about the environment (for example a semantic map) can be easily acquired by the robots and used for the surveillance task.

Localization of security personnel

Personnel carry active RIFD tags/readers as well as other Wi-Fi enabled communication equipment. The attenuation of the signals emitted from these devices is available. A model of the propagation characteristics can be used to estimate the range at between transmitter and receiver using the difference in signal strength. The nature of the model can be theoretical or empirical. The challenge for such a system is that RF propagation suffers from a number of problems including diffraction, reflection, and scattering of the signal that introduce significant errors; thus probabilistic methods are usually employed. The proposed approach is based on scene analysis with additional reference tags. Scene analysis uses fingerprinting and operates in two phases. In the first offline phase a fingerprint of the environment or a map of the signal strength at all locations is produced. In the second or online phase, the fingerprint is employed to estimate position given signal strength. Kalman filters (Bekkali, Sanson, Matsumoto 2007), particle filters and Bayesian techniques have been used to improve the localization estimate.

Video analysis for scene understanding

In fusion data process of multi-camera sensors, the first step is the object detection and tracking of the target in the scene being monitored.

Object detection and Tracking algorithms

The approach is based on (Harveille, Dalong 2004) work. It consists in tracking using 3D information and Kalman Filters (Bekkali, Sanson, Matsumoto 2007), fixed templates which are implemented as the combination of the height and statistical data of occupancy of the object detected. Figure 4 illustrates the whole process. Background subtraction technique is applied to obtain the object from the scene. Initially, a background model is computed and adapted, applying a well-known statistical method for clus-

[1] Speaky - www.mediavoice.it

tering called Gaussian Mixture Model (GMM) per-pixel, using four channels of information (color and depth). Following equation illustrates the computation of the GMM:

$$P(X_{i,t} \mid X_{i,1}...X_{i,t-1}) = \sum_{j=1}^{K} \delta_{i,t-1,j}\varphi(X_{i,t};\theta_j(\mu_{i,t-1},\Sigma_{i,t-1})) \qquad (1)$$

Where δ, it is the mixing weights of pass observations and $\theta_j(\mu_{i,t-1},\Sigma_{i,t-1})$ it is a Gaussian density function component. Once the object is detected, the two data templates are created by projecting the object detected data to the ground plane, from a top-camera view. The height template is computed using the object detected height information (i.e. the head pixels would provide the highest value and the feet the lowest). The occupancy template is compute as follows; all pixels from the object are projected to the ground plane in an accumulative manner, therefore the region of the pixels belonging to the head and main body of the object provide higher occupancy values as they accumulate more number of pixels (the projection of all pixels is made from top to bottom).See top images in Figure 4.

The 2D position of the center of mass of the computed templates are thus tracked using Kalman Filters. See bottom images in Figure 4. The state vector is composed by the 2D position of the templates, their velocity, and the highest value of the height template and the highest value of the occupancy template. The area to search for correspondence on the tracks, is centered to the estimated 2D position of the ground plane object's projection. The correspondence is resolved via match score, which is computed at all localizations within the search zone. A lower match score implies better match. Following equation illustrates the computation of the match score:

$$\varphi(i,X) = \rho SAD(\tau_H, H_{masked}(X)) + \omega SAD(\tau_O, \theta_{sm}(X)) + \beta\sqrt{(x-x_{pred})^2 + (y-y_{pred})^2} + \alpha\sum_{j<i}\theta_j(X,40) \qquad (2)$$

SAD refers to the sum of absolute differences between height and occupancy templates created from the current frame. The (ρ,ω,β,α) weights are deduced in (Harveille, Dalong 2004).

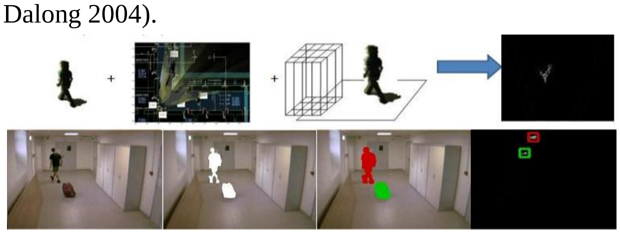

Figure 4. Results of height and occupancy maps and tracking.

Furthermore, to enable in future the recognition of more detailed human actions, in addition to tracking persons, the system is able to detect their body parts (torso and arms). The approach is based on background-foreground segmentation, followed by identifying and tracking the torso and labeling silhouette segments with respect to the torso (using extremal convex points and SIFT tracking for increased robustness (Lowe 2004). Figure 5 shows an example of detected body parts, to be used for recognizing a grabbing action.

Figure 5. Detection of body parts.

As illustrated in Figure 3, the object detection (OD), tracking individuals and tracking human body parts services have defined outputs. For example, the tracking individuals service outputs the certainty or likelihood of the target position being tracked. Therefore, once a target starts to be tracked a certainty map (Munoz, Salinas 09) is created with the target's data, as shown in the following equation:

$$\Psi(x,y) = \begin{cases} 0 & I(x,y)=0 \\ \dfrac{\mathrm{E}[(\frac{Zcam}{f})^3]}{N\cdot\sigma^{\frac{3}{2}}} & I(x,y)\neq 0 \end{cases} \qquad (3)$$

The certainty maps presented in this paper are computed using a third moment (skew) of the Gaussian density function that represents the reliability of the 3D data coming from the camera sensor; i.e. the working range of camera sensor.

Information Fusion

The solution we research consists in the implementation of a system composed of a network of video cameras, radio-frequency devices (RFID) sensors, robotic platforms and portable devices worn by guards. One of the implications of using heterogeneous sensors to monitor the same location is the large amount of redundant data, which can be exploited to improve performance on the system’s monitoring activity (Snidaro, Visentini, Foresti 2009). However, to manage successfully this amount of data, the fusion should be driven by the quality of data to be integrated.

Different approaches have been researched in computer vision to fusion multi-cameras. The common related work is based on finding the topological relation between cameras (Lobaton et al. 2010) using appearance models (Ellis, Markis, Black 2003). Other work is based on fusion using statistical data association (Kettnaker,Zabih 1999), others use temporal matching cues (Zhao et al. 05) or color cues matching (Gilbert et al. 2009) or integrating both information cues using unsupervised learning methods (Chen,Hung 2010)(Vivek, Pradeep 2008). Other related work is based on finding the relative position of the cameras on a common reference system (Mavrinac, Chen, Tepe 2010)(Zhang et al. 2008).

Some of these work present good results. However, all these approaches are hardly based on the assumption of fusion the same type of data from the same type of sensors (i.e. cameras). The approach presented in this paper is focused not only on fusion data but the quality of data to be fused, and also the fact that the representation of this data should be standard as the sensors may be different technology (i.e. RFID, robot and cameras). Thus, the data that they output would be different. The certainty maps used in this approach, which are described in previous section, provide at the same time a degree of data quality (i.e. the sensor output how certain is from the output data). The fusion approach used, is based on distributed independent likelihood pool mechanism (Snidaro et al. 2009) and the decentralized data fusion is obtained by applying the local beliefs concepts (Capitan et al. 2009).

Fusion through distributed independent likelihood pool

As mentioned, the approach to fusion all the data (initially the probability of the position of a target) coming from different sensors is based on distributed independent likelihood pool mechanism (Snidaro et al. 2009) and local beliefs concepts (Capitan et al. 2009). The probability of the target's localization is approximated using a third moment of a Gaussian. Assuming that all the data coming from K sensors are conditionally independent given certain state X with a set of measurements Z, then the posterior distribution of the state X is composed by the independent likelihood pool of the likelihood functions of each sensor:

$$P(X \mid Z^K) = C \cdot P(Xo) \prod_{i=1}^{K} \frac{\psi_i(X)}{P(Xo)} \qquad (4)$$

Therefore, X represents the fusion of the position of the target once all the positions provided by the K sensors are combined through their local beliefs ($\psi_{i(X)}$) and the common information shared (i.e. the prior over the trajectory P(Xo) is removed. C is a constant to normalize the posterior distribution.

Case study

The scenario for the system to be built could be the baggage area of an airport as it is a complex environment where crowded situations may occur at random times. People in this type of scenario with their luggage provide all sorts of shapes and forms interesting to analyse. This section presents an initial prototype of small scale of the final system to illustrate what is presented in Figure 3. The prototype consists of two robots patrolling an area and communicating with a fix camera and also with a control room operator, sees Figure 6. The prototype has also a fix camera operating in a remote localization and providing the results to a mobile operator through a PDA, which it is illustrated in Figure 7.

Figure 6. Robot goes to the position indicated by the camera.

Figure 6 shows the results of implementing the camera, robot, map, localization, OD and tracking individual services illustrated in Figure 3. First, the robots create the map and localize themselves. Then, the personnel from the control room, can operate the robots in two ways: *manually*, the operators send the robots to a position that they want to monitor or *automatically*, once the fix camera sensor detects a predefined event (i.e. someone in a forbidden area) , sends the target position to the robot and the robot goes there.

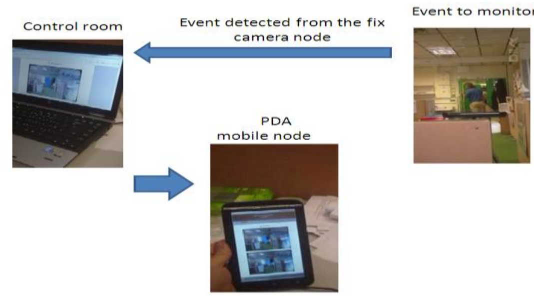

Figure 7. The system transmits the event monitored by the camera sensor to the control room and PDA device.

Figure 7 shows the results of the implementation of camera, archive, visualize and OD services. Once the fix camera sensor detects a target, it archives the output, which then allows a mobile node (i.e. security personnel) to visualize the output through a PDA device.

Conclusions and future work

This paper presented an earlier stage of the research focus on building a heterogeneous multi-sensor system to monitor public environments. The paper discussed the proposed solution and presented the fusion data approach that is going to use between the sensors that conform the system. Although the fusion data approach described in this paper is based on multi-camera sensors, the framework is developed to fusion in future, data coming from different sensors (i.e. RFID, robot and cameras). The approach is based on fusion likelihood functions based on the target's localizations provided by each sensor.

However, research is still required in several topics, which have been explained along the paper as e.g. in multi-human/multi-robot paradigm (MHMR). Also, further work in interaction and feedback mechanism within sensors should be carried out e.g. applying approaches that allow to cooperate sensors on a merit decided by competition (Vivek et al. 2008) optimizing the cooperation between the sensors, increasing the performance of the system.

References

Raty D. T. 2010. Survey on Contemporary Remote Surveillance Systems for public Safety. Part C:Applications and Reviews. *IEEE Transactions on Systems, Man and Cybernetics*. 40(5): 492-315.

Valera M. and Velastin, S.A .2005. Intelligent distributed surveillance systems: a review. IEE Proceedings of Visual Image Signal Process, 152(2):192-204. Calif.: IEE conference of Visual Image Signal Process.

Harville M. and Dalong Li. 2004.Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2: 398-405.

Zhao T; Aggarwal M.; Kumar R.; and Harpreet Sawhney. 2005. Real-time wide Multi-Robot and Multi-Camera. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 976-983. Washington, DC, USA. Calif: IEEE Computer Society.

Fusier F.; Valentin V.; Brémond F.; Thonnat M.; Borg M.; Thirde D. and Ferryman J.. 2006. Video Understanding for Complex Activity Recognition. *In Machine Vision and Applications Journal* , 18(3-4):167-188.

Munoz-Salinas R.; Medina-carnicer R.; Madrid-Cuevas F.J.; Cartmona-Poyato A. 2009. People detection and tracking with multiple stereo camera using particle filters. *Journal Visual Communication Image Representation*. (20):339-350.

Chen. K-W. and Hung Yi-P. 2010. Multi-Cue Integration for Multi-Camera Tracking. In Proceedings of International Conference on Pattern Recognition. 145-148. Calif.: International Conference on Pattern Recognition.

Buxton H., Gong S. 2002. Generative Models for Learning and Understanding Scene Activity. Technical Report DIKU-TR. The University of Copenhagen.

Burgard, W.; Moors M.; Fox D.; Simmons R.; and Thrun S. 2000. Collaborative Multi-Robot Exploration. In Proc. IEEE International Conference on Robotics and Automation. 476-481 .San Francisco. USA. Calif.: IEEE International Conference on Robotics and Automation.

Zhang, F.; Chen W.D.; and Xi Y.G. 2005. Improving Collaboration through Fusion of Bid Information for Market-based Multi-robot Exploration. In Proc. IEEE International Conference on Robotics and Automation. 1157-1162.Barcelona, Spain. Calif.: IEEE International Conference on Robotics and Automation.

Iocchi L.; Nardi D.; and Salerno M. 2001. Reactivity and Deliberation: A Survey on Multi-Robot Systems. Balancing Reactivity and Social Deliberation in Multi-Agent Systems, From RoboCup to Real-World, 3-540-42327-3, 9–34. Mass.: Springer-Verlag, London, UK.

Stelzer A.; Pourvoyeur K.; and Fischer A.2004. Concept and application of LPM - a novel 3-D local position measurement system. *IEEE Trans. on Microwave Theory and Techniques* 52(12):2664–2669.

A. Bekkali, H. Sanson, and M. Matsumoto. 2007. RFID indoor positioning based on probabilistic RFID map and kalman filtering. In Proc. of WiMOB. Calif.: Conference WiMOB.

Wang Ching-Sheng; Huang Xin-Mao; And Hung Ming-Yu. 2010. Adaptive RFID Positioning System Using Signal Level Matrix. *World Academy of Science, Engineering and Technology* 70

Hatthasin, U.; Thainimit, S.; Vibhatavanij, K.; Premasathian, N.; Worasawate, D. 2009. An improvement of an RFID indoor positioning system using one base station. In Proc. 6th International Conference on ECTI-CON. Calif.: Conference on ECTI-CON.

Tesoriero R.; Tebar R.; Gallud J.A.; Lozano M.D. ; and Penichet V.M.R.2010. Improving location awareness in indoor spaces using RFID technology. *Expert Systems with Applications* 37:894-898.

Valera, M. and Velastin, S A. 2003. An approach for designing a real-time intelligent distributed surveillance system. In Proceedings on IEE Workshop on Intelligent Distributed Surveillance Systems. London. 42-48.Calif.: IEE Workshop on IDSS.

Quigley. M.; Gerkey B.;Conley. K.; Faust. J.; Foote. T. 2009. ROS: an open-source Robot Operating System. In Proceedings of IEEE International Conference on Robotics and Automation. Calif.: International Conference on Robotics and Automation workshop.

Iocchi L.; L. Marchetti, Nardi. D. 2011. Multi-Robot Patrolling with Coordinated Behaviours in Realistic Environments. In Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA. Calif.: International Conference on Intelligent Robots and Systems.

G. Randelli. G; Iocchi L.; Nardi. D. 2011. User-friendly security robots. In Proc. of IEEE International Symposium on Safety, Security, and Rescue Robotics. Kyoto, Japan. Calif.: International Symposium on Safety, Security, and Rescue Robotics.

Lowe D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints . *International Journal of Computer Vision* 60(2):91–110.

Snidaro L.; Visentini I.; and Foresti G.L. 2009. Multi-sensor multi-cue fusion for object detection in video surveillance. In Proc. Of AVSS. 364-369. Calif.: Conference on AVSS.

Lobaton E.; Vasudevan R.; Bajcsy ;R. Sastry S. 2010.A distributed Topological Camera Network Representation for Tracking Applications. *IEEE Trans. on image processing* 19(10):2516-2528.

Ellis T. J.; Makris D. ; and Black J. 2003. Learning a multi-camera topology. In Proc. Of Joint IEEE Int. Workshop VS-PETS. 26-33. Calif.: IEEE International Workshop VS-PETS.

Kettnaker V.;Zabih R. 1999. Bayesian Multi-camera Surveillance. In Proc. On IEEE International Computer Vision on Pattern Recognition.2253-2259. Calif.: International conference on Computer Vision. On Pattern Recognition.

Vivek K.S.;Pradeep K.A.2008. Coopetitive multi-camera surveillance using model predictive control. *Machine Vision and applications* 19:375-393.

Gilbert A.; Illingworth J.; Bowden R.; Capitan J.; and Merino L. 2009. Accurate fusion of Robot, Camera and Wireless sensors for surveillance. The ninth IEEE International workshop on Visual Surveillance. Calif.: International conference on Computer Vision.

Mavrinac A.; Chen X.; Tepe K. 2010. An Automatic calibration method for stereo-based 3D distributed smart camera networks. Computer Vision and Image Understanding 114:952-962.

Zhang Z.; Scalon A.; Yin W. ;Yu L. Venetianer P.L. 2008. Video Surveillance Using a Multi-camera tracking and Fusion System. In Proc. Of workshop on Multi-Camera and Multi-modal Sensor Fusion Algorithms and Applications.

Capitan J.;Merino L.; Caballero F.; and Ollero A. 2009. Delayed-state Information Filter for Cooperative Decentralized Tracking. In Proc. on IEEE International Conference on Robotics and Automation. Japan.3865-3870. Calif.: International conference on Robotics and automation.

Acknowledgments

This publication was developed under Department of Homeland Security (DHS) Science and Technology Assistance Agreement No. 2011-ST-108-000021 awarded by the U.S. Department of Homeland Security. It has not been formally reviewed by DHS. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. The Department of Homeland Security does not endorse any products or commercial services mentioned in this publication.