Sensor network reconfiguration and big multimedia data fusion for situational awareness in smart environments

Z. Akhtar, C. Drioli, M. Farinosi, G. Ferrin, G.L. Foresti, N. Martinel, C. Micheloni, C. Piciarelli, D. Salvati, L. Snidaro and M. Vernier

AVIRES Lab - Department of Mathematics and Computer Science, Università degli Studi di Udine
Via delle Scienze, 206, 33100 Udine - Italy

INTRODUCTION

In the last years, an increasing number of environments have been enhanced with smart sensors and have become more and more smart and self-organizing [1]. Situational awareness (SA) in these wide areas covers a huge range of topics and challenges [2]. As matter of fact, understanding activities for situation assessment cannot be achieved locally but it requires to widen as much as possible the monitored area. Several different and new problems must be investigated from the use of single sensors able to adapt internal or external parameters to the cooperation of multiple sensors in a dynamic and smart network [3] that continuously adapt to the scenario evolution by changing topology, controlling sensors parameters, switching-on/off nodes, deciding the best spectrum for specific analysis, etc.

In this contribution, an innovative situational aware system for smart environments is presented. It is based on an original architecture that allows two-way communications and combines content produced by citizens living in the environment with data generated from distributed sensor networks. The main functionalities of the proposed architecture are: (a) to collect geo-referenced multimedia data from multiple sensors, e.g., infrastructure sensors (video cameras, microphones, etc.) or from citizens (e.g., mobile smartphone, social applications such as Twitter or Facebook, etc.); (b) to process and fuse locally rough data acquired by infrastructure sensors; (c) to reconfigure the sensor network according with the locally process data, (d) to communicate data about detected events via a high-speed wireless communication network to the unified operative center; (e) to integrate different multimedia data to obtain a robust situational awareness of the observed environment; (f) to collect on demand additional data (e.g. short video and audio sequences) using unmanned autonomous vehicles (UAVs); (g) to share targeted information amongst different executive monitoring units; (h) to guarantee the system security and at the same time to ensure timely and reliable access to useful information.

SYSTEM ARCHITECTURE

Figure 1 shows the logical architecture of the presented system for situational awareness based on big multimedia data. The proposed architecture is able to capture and aggregate different kind of data: (a) user generated content (UGC) produced by citizens (e.g., during or immediately after a disaster) and shared online through social platforms and (b) data acquired by smart sensors (i.e., intelligent cameras, microphones, acoustic arrays, etc.) distributed on the environment. Data are selected, analyzed, processed and integrated in order to obtain some relevant information for the situational awareness such as emergency response, search and rescue operations in natural disasters, etc.. Video and acoustic data are locally pre-processed and sent to the sensor network reconfiguration module and to the Data Fusion and Situational Awareness module which has the task to detect anomalous events. The final goal is to provide “full digital” solutions to the functioning of situational awareness systems, starting from the sensor level, up to the presentation of integrated big multimedia information to the operators at the control center. Recent progress in low cost, high performance computing networks and in the availability of digital communications on heterogeneous, mobile and fixed broadband networks [4] have allowed both an easy digital interaction between citizens and network infrastructures and the availability of large amount of multimedia data coming from multiple heterogeneous sensors deployed on the environment or managed by non-professional users.

The architecture of the proposed system is organized into three different layers: the sensor layer, the processing layer, the situational awareness layer.

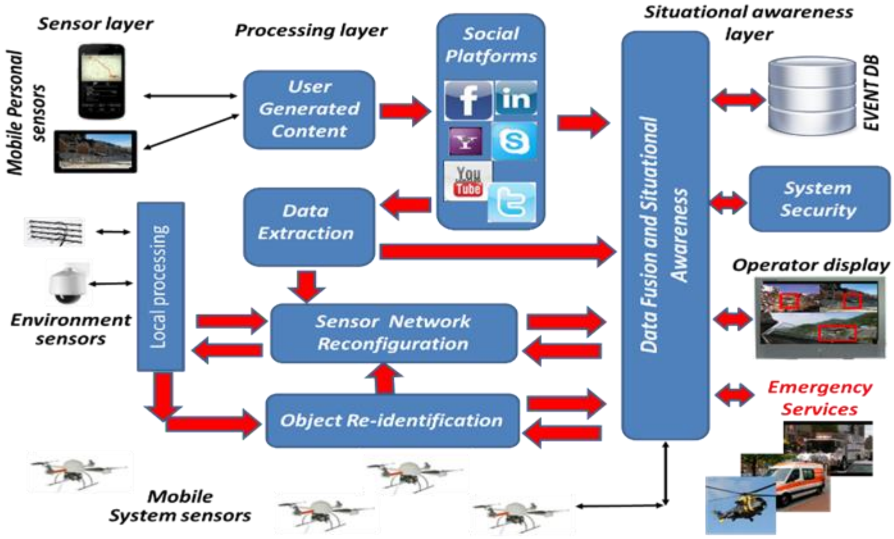


Figure 1: *Logical architecture of the proposed situational awareness system for smart environments*

The sensor layer is composed of different kinds of sensors: (a) environment sensors, distributed permanently on the environment, (b) mobile personal sensors, directly used by citizen and (c) mobile system sensors, placed on-board to unmanned aerial vehicles (UAVs) useful to inspect specific areas with augmented acquisition capability. At the sensor layer, some local processing and fusion mechanisms [5] are applied to multimedia data acquired by distributed heterogeneous sources. Local and network transmissions are managed directly through network operators. The local processing module applies digital compression methods to save bandwidth resources [6]. The communication medium is normally represented by wireless LANs (e.g., IEEE 802.11g, IEEE 802.11n, etc.) or mobile digital devices (e.g. HSDPA for mobile phones) as well as broadband media such as optical fibers, coax cables or IEEE 802.16 WiMAX, which extends over 30 miles and allows a bandwidth of 100 Mbps. Information and data coming from the sensor layer are processed at the intermediate layer. In particular, user generated content uploaded by citizen on social network platforms are filtered and extracted according with specific search parameters (e.g., words such as “earthquake” or “flood” in case of natural disaster or “bomb” or “attach” in the case of terrorism activities, etc.) and sent to the Data Fusion and Situational Awareness (DFSA) module for automatic interpretation of occurring events. Video and acoustic data coming from environment sensors are locally processed and partially integrated to focalize the attention of the system on possible interesting events. Output data are analyzed by the sensor network reconfiguration module to modify the main sensor control parameters such as pan or tilt positions, zoom-in or zoom-out commands of a video camera, microphones sensitivity, etc. in order to increase the quality of acquired data. At the same time, at the situational awareness layer, meta data are finally processed by the DFSA module, where advanced machine learning algorithms can handle both emergencies and prevention to improve citizens’ safety. Moreover, an active user-friendly interface allows to display to the operator multimedia data in an efficient way.

DATA ACQUISITION and PRE-PROCESSING

Several video and audio sensors can be adopted to acquire information from the environment. Optical and infrared cameras, microphones and/or acoustic arrays are the most commonly sensor used and the large amount of data they produce can be easily processed and useful information can be extracted.

Video sensors are normally placed in specific positions in the urban environment as, for example, in the front of a building or on street corners, and generally they can be classified as static or with pan-tilt-zoom (PTZ) capabilities. The large amount of data they are able to acquire are locally processed using well know state-of-art computer vision algorithms. Change detection algorithms can be adopted to highlight significant changes on the environment by focusing on the moving objects [7].

Video sensors can be also placed on Unmanned Aerial Vehicle (UAV) systems which are able to fly over specific areas and acquire useful information of the environment from an orthogonal viewpoint (Figure 2).

The DFSA module works both of on-line and off-line mode: i) during the off-line phase a mosaic of the interesting environment is built by using SURF features and iterative matching techniques such as RANSAC.

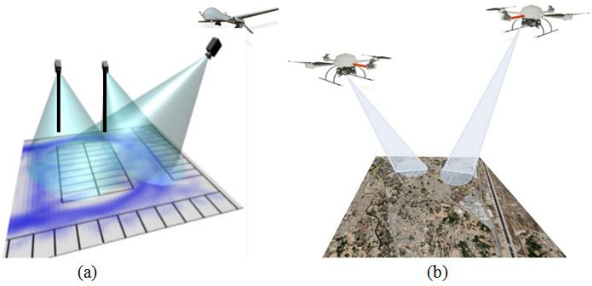


Figure 2: *Unmanned Aerial Vehicles (UAVs) used to monitor the environment and to acquire additional info on system's or operator's request.*

At the end of the iterative process, the homography minimizing the error is chosen; ii) during the on-line phase real-time images acquired by the UAV are matched with the previously stored data and significant changes on the environment are highlighted using specific change detection algorithm. GPS information is used to estimate the location of the current frame within the mosaic.

The system relies also on acoustic scene analysis techniques to incorporate environment acoustical awareness. Single and multiple acoustic source localization, with respect to both the near-field and the far-field, is realized using arrays of microphones and a number of new techniques and algorithms for the accurate estimation of *Direction Of Arrival* (DOA) estimation, fusion of information from distributed sensors, and tracking of moving acoustic sources ([8,9]). Such acoustic front-end provides the system with the ability to tackle a wide number of tasks typical of signal-based surveillance applications and multimodal interactive environments, including vehicles tracking for traffic surveillance (Figure 3), indoor acoustic events detection and localization, events detection and classification for sensors steering and dynamic reconfiguration in intelligent environments.

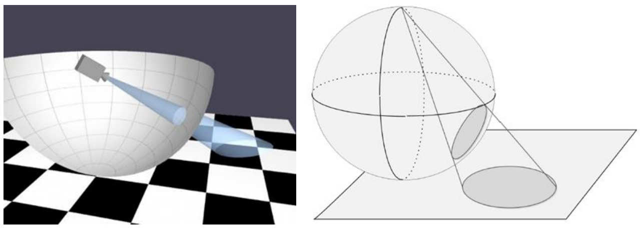
Figure 3: *An example of microphone array configuration for outdoor acoustic source localization and tracking. Left: the sensor array configuration; Right: the acoustic tracking of a vehicle crossing the area under surveillance.*

SENSOR NETWORK RECONFIGURATION

The sensor network reconfiguration system aims at optimizing the external (pan, tilt) and internal (zoom) parameters of a set of PTZ cameras in order to dynamically guarantee an optimal coverage of the monitored environment. The system relies on the notion of relevance maps. A relevance map is a 2D (in [10]) or 3D (in [11]) map expressing the relevance for each visible portion of the environment based on the goals of the system. Its definition is thus highly dependent on the global goal: a typical example could be an activity map expressing the presence of moving objects detected by wide-FOV static cameras. In this case, it is assumed that the zones with higher activity are also the most important ones for surveillance.

Independently of the nature of the relevance maps, we propose a local coverage model that expresses, for each camera, the quality of the current local configuration, defined as the total amount of relevance of the zones observed by the sensor. The goal is to find the network configuration that maximizes the global quality, expressed as the sum of all the local quality measures.

We consider the cone-of-view of each camera and project the relevance of the area covered by the cone on the surface of a unitary sphere centered in the sensor. As shown in Figure 4a, this results in a circle-shaped area on the surface of the sphere, which can be re-projected on a new space (called ***camera space*** - see Figure 4b), by means of the circle-preserving stereographic projection.



(a) (b)

Figure 4*: (a) A circle-shaped area on the surface of the sphere and (b) re-projection on the camera space.*

We approximate the circles in the camera spaces with Gaussian functions, so that the global optimization problem can be solved with a standard Expectation-Maximization approach with isotropic Gaussian functions. The main advantage of this technique is that the optimization problem is unconstrained: any solution – any isotropic Gaussian function in the camera space - corresponds to a valid camera configuration. Applying the same technique directly on the relevance map space would have led to a constrained problem, since the orientation and eccentricity of the ellipses obtained by the intersection of the cone-of-view with the ground plane are strictly related to the camera position and orientation.

OBJECT RE-IDENTIFICATION

In the recent years, the object re-identification problem, formally defined as the task of assigning the same label to an object (typically a person) that moves across camera field-of-views (FoVs), has gained a lot of interest by the community. Such problem is very attractive as, for video surveillance applications, knowing whether a person is present in the monitored area at a precise time instant is of paramount importance.

To address the challenging issues of the person re-identification problem, the community has devoted his effort following three main different classes of approaches: (i) discriminative signature based methods [12-14], (ii) feature transformation learning-based methods [15] and (iii) metric learning-based methods [16].

All of such methods try to capture the discriminative power of each image by extracting local or global color, shape and texture features. While this has been shown to be an effective approach, some of such features are complex and time consuming, thus they can be hardly used in a real-time scenario. Moreover, we also believe that particular image pixels have more discriminative power than such complex features. However, the comparison of pixels at exactly the same location is highly sensitive to the person location and pose. To deal with this issue, we group set of neighboring pixels to have a coarse representation, then, we build upon the idea that there exist a multi-modal transformation of the difference between such localized pixel groups. Still, not all the localized groups may be useful to capture such multi-modal transformation.

As shown in Figure 5 we propose to model the multi-modal transformation of the difference between groups of localized pixels that rely on a linear subspace that best captures their intrinsic dimensionality.

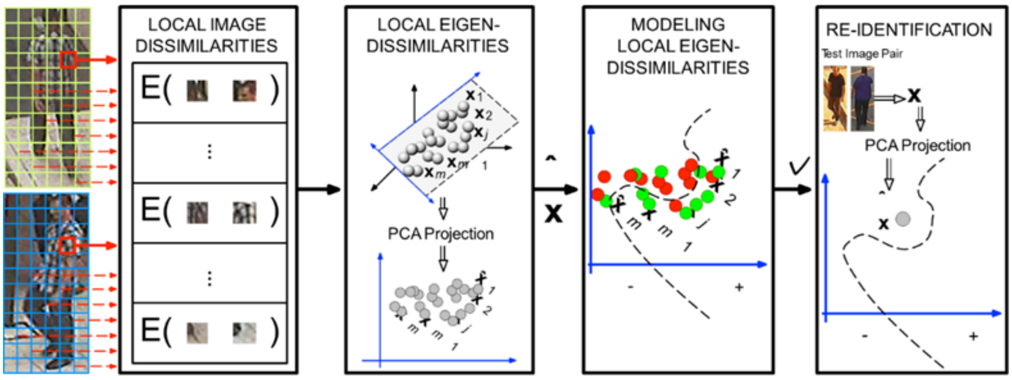


Figure 5: *Person Re-Identification by means of local image eigen-dissimilarities.*

To this aim we compute the local dissimilarity between images of the same person (positive pair) as well as the dissimilarity between images of different persons (negative pair) viewed in two cameras. Then, we learn the linear subspace where set composed of all the local dissimilarities lie. Finally, we use a supervised classification framework to discriminate between the positive and negative pairs in the linear subspace.

SYSTEM SECURITY and AUTHENTICATION

System security is a critical issue since any breach may lead to severe consequences from personal information leakage to a cascade of failures, such as massive blackout and destruction of infrastructures. The system is likely to be exposed to various attacks, involving denial of service (DoS), spoofing, hacking, malicious codes, worms, and viruses as it is also connected to Internet [17]. In the following, we enlist three high-level security objectives in the proposed system.

i) *Availability*: Ensuring timely and reliable access to and use of information is of the most importance in the system. Since, a loss of availability may undermine the emergency response, search and rescue operations in natural disasters.
ii) *Integrity*: Guarding against improper information modification or destruction is to ensure information nonrepudiation and authenticity. A loss of integrity might induce incorrect decision regarding resources management.
iii) *Confidentiality*: Preserving authorized restrictions on resources or information access and disclosure is mainly to protect personal privacy and proprietary information.

Based on the above mentioned security objectives, we consider three types of malicious attacks as below:

i) Attacks targeting availability, also called denial-of-service (DoS) attacks, attempt to delay, block or corrupt the communication in the system (e.g. channel jamming). Cryptography, filtering and network reconfiguration based powerful countermeasures are employed to mitigate effects of DoS attacks.
ii) Attacks targeting integrity aim at deliberately and illegally modifying or disrupting data exchange in the system. The target can be either sensor/s' information or status of decision planning and emergency services. *Authentication*, confirming an identity or origin of a communication partner or a piece of information, is a crucial identification process to egest attacks targeting data integrity. Thus, end-to-end *authentication* schemes (e.g., digital signature and verification) are enforced to prevent such attacks.
iii) Attacks targeting confidentiality intend to acquire unauthorized information or resources in the system. Like spoofing attacks, which can lead to loss of availability, integrity and confidentiality. Such as biometric based access control (e.g., computer-server managing decision, broadcasting the situations and dispatching the emergency services) can be deceived by biometric spoofing attacks [18], as shown in Figure 6. Therefore, we have designed several software-based anti-spoofing countermeasures, however they are challenging due to timing requirements.

To deal with the potential security threats in the proposed system, countermeasures and defense strategies are widely deployed and integrated into system protocols and architectures.

Figure 6: *Examples of biometric spoofing attacks. Left: face photo-attack. Middle: fingerprint silicone-attack. Right: iris photo-attack.*

DATA FUSION and SITUATION AWARENESS

Building and updating a situational picture of the scenario under consideration is the goal of the DFSA module. The scenario generally involves multiple entities and actors where possibly only a few are under direct control of the decision maker. Situation assessment (SA) aims at explaining the observed events (mainly) by establishing the entities and actors involved, inferring their goals, understanding the relations existing (whether permanently or temporarily) between them, the surrounding environment, and past and present events. It is therefore apparent how the SA process inherently hinges on understanding and reasoning about relations. SA is necessary preparatory step to the following phase of Impact Assessment (IA) where the decision maker is interested in estimating the evolution of the situation and the possible outcomes, dangers and threats. SA and IA processes are particularly complex and critical for large-scale scenarios with nearly-chaotic dynamics such as those affected by natural or man-made disasters. Recent developments in information fusion methods for representing and reasoning about relational information and knowledge are exploring methods belonging to the rapidly growing area of Statistical Relational Learning methods that have attracted significant attention lately. In particular, methods that blend the expressiveness of first-order logic and the learning capabilities and uncertainty management of graphical methods are under study [19].

Another important recent line of research is the integration of Contextual Information (CI) in order to improve SA performances [20,22]. Modern Information Fusion (IF) systems must consider the specific characteristics of the application domain in which they have to operate, showing robust and context-sensitive behavior. A system designer must take into consideration different sources of contextual knowledge in addition to immediate sensory data in order to develop an effective, efficient, and useful system in the target domain: such knowledge may include structural models of the scenario, known a priori relationships between entities and the surrounding environment, dynamic relationships necessary to interpret or constrain the system output, user preferences, social norms, etc. Context includes conditions which augment results that enhance meaning.

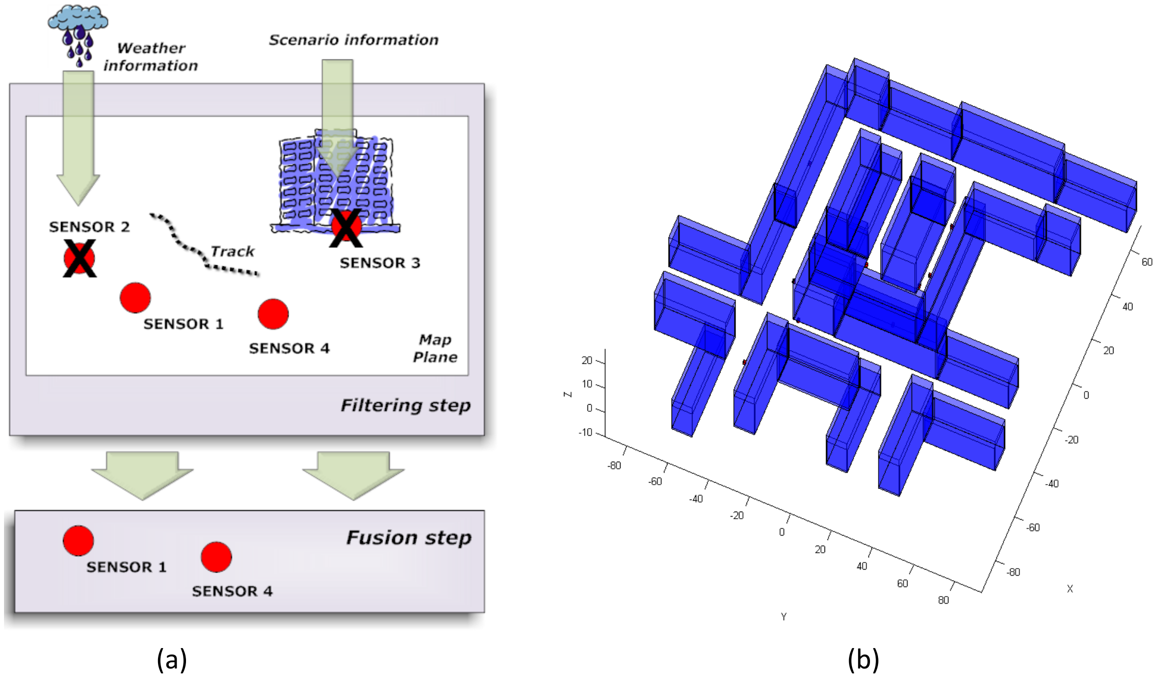


Figure 7 - *Filtering by contextual information: the measurements of several sensors observing a target are fused considering their contribution and their reliability with respect to the map of the area and weather conditions.*

The development of IF systems, to include data-, sensor-, and feature-level fusion, is a necessary engineering process in diverse applications, and new domains are requiring an increasing degree of contextualized solutions and situation-adaptation mechanisms (Figure 7a). The development of IF systems inclusive of contextual factors and information is an opportunity to improve the quality of the fused output, provide solutions adapted to the application requirements, and enhance tailored responses to user queries. Contextual-based strategy challenges include selecting the appropriate representations, exploitations, and instantiations. Context could be represented as knowledge-bases, ontologies, maps, etc. and would form a powerful tool to favor adaptability and system performance. Example applications include context-aided tracking and classification, situational reasoning, ontology building and updating, etc. Contemporary discussions of context are domain and sensor specific for which information fusion enhances performance. For example, context-aided target tracking seeks to determine kinematic movements with domain-constrained sensitivities (e.g. roads and buildings, Figure 7b) [21]. Whereas, context-aided information fusion solutions utilize not only the road information but the social norms of the same geographical information (e.g. traffic rules).

CONCLUSIONS

The integration of the different section gives life to an advanced situational aware system for smart environment. Data generated by environmental sensors (optical and infrared cameras, microphones, etc.) together with data produced by citizens via socio-mobile applications can offer an accurate visualization of the real time situation. Differently to other systems, the proposed architecture allows an active involvement of people, that, thanks to socio-mobile applications, can spread relevant communication about some hyper local situations and cooperate in this way with the public authorities, providing up-to-date information. Collecting geo-referenced multimedia data form multiple sensors, the system allows localizing the critical areas and ensuring timely and reliable access. The combination of grassroots communication practices with a sophisticate intelligent sensor network increases the efficiency of the traditional software for situational awareness.

Further research is needed to integrate additional data from new sensors in order to allow the entire architecture to achieve best performance. This addition will guarantee the possibility of having a greater number of information for the situational awareness. Furthermore, as ordinary citizens acquire largest digital skills, they will be able to provide more detailed and accurate content and information, which in turn will guarantee a more precise intervention on the territory. It will also be possible to develop a specific socio-mobile application to use on the next generation personal devices, in case of various situational awareness.

REFERENCES

[1] J.C. SanMiguel, C. Micheloni, K. Shoop, G.L. Foresti, A. Cavallaro, "Self-reconfigurable smart camera networks", ***IEEE Computer***, (in press), 2014.

[2] G.L. Foresti, M. Farinosi, M. Vernier, "Situational awareness services in smart environments: socio-mobile and sensor data fusion for emergency response to disasters", ***Journal of Ambient Intelligence and Humanized Computing***, 2014.

[3] B. Dieber, C. Micheloni, and B.Rinner, "Resource-Aware Coverage and Task Assignment in Visual Sensor Networks", ***IEEE Transactions on Circuits and Systems for Video Technology***, 21 (10): 1424-1437, 2011.

[4] P. Abad, V. Puente, J.A. Gregorio, Balancing Performance and Cost in CMP Interconnection Networks, ***IEEE Transactions on Parallel and Distributed Systems***, 23 (3): 452-459, 2012.

[5] L. Snidaro, I. Visentini, G.L. Foresti, "Data fusion in modern surveillance," in ***Studies in Computational Intelligence***, vol. 336, P. Remagnino, D.N. Monekosso and L.C. Jain, Eds. Springer-Verlag, 1-21, 2010.

[6] G.L Foresti, C. Regazzoni, R. Visvanathan, "Video Communications, Processing and Understanding for Third Generation Surveillance Systems", ***Proceedings of IEEE***, 89 (10), 2001.

[7] L. Snidaro and G.L. Foresti, "Real-Time Thresholding with Euler Number", ***Pattern Recognition Letters***, 24 (9-10): 1533-1544, 2003.

[8] D. Salvati, C. Drioli and G.L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments", ***IEEE Signal Processing Letters***, 21 (5): 581-585, 2014.

[9] D. Salvati, A. Rodà, S. Canazza and G. L. Foresti, "Multiple acoustic sources localization using incident signal power comparison", ***8th IEEE International Conference on Advanced Video and Signal-Based Surveillance***, Klagenfurt, Austria, August 30 - September 2, pp. 77-82, 2011.

[10] C. Piciarelli, C. Micheloni and G. Foresti, "Occlusion-aware multiple camera reconfiguration," ***4th ACM/IEEE International Conference on Distributed Smart Cameras***, Atlanta, GA, USA, 2010, pp. 88–94.

[11] C. Piciarelli, C. Micheloni, and G. Foresti,, "Automatic reconfiguration of video sensor networks for optimal 3d coverage," ***5th ACM/IEEE International Conference on Distributed Smart Cameras***, Ghent, Belgium, 2011.

[12] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network", **International Conference on Computer Vision and Pattern Recognition Workshops**, Providence, RI, USA, pp. 31–36, 2012.

[13] N. Martinel and G.L. Foresti, "Multi-signature based person re-identification", **Electronics Letters**, 48 (13):765-768, 2012.

[14] N. Martinel, C. Micheloni and C. Piciarelli, "Distributed Signature Fusion for Person Re-Identification", **International conference on Distributed Smart Cameras**, Hong Kong, pp. 1-6, 2012.

[15] N. Martinel, C. Micheloni and C. Piciarelli, "Learning Pairwise Feature Dissimilarities for Person Re-Identification", **International conference on Distributed Smart Cameras**. Palm Springs, CA, 2013.

[16] S. Pedagadi, J. Orwell and S. Velastin, "Local Fisher Discriminant Analysis for Pedestrian Re-identification", **International Conference on Computer Vision and Pattern Recognition,** pp. 3318–3325, 2013.

[17] H. Nakashima, H. Aghajan, J. C. Augusto, " Handbook of Ambient Intelligence and Smart Environments", ***Springer Verlag***, New York, 2009.

[18] Z. Akhtar, "Security of Multimodal Biometric Systems against Spoof Attacks", ***PhD thesis, University of Cagliari,*** Cagliari, Italy, 2012.

[19] L. Snidaro, I. Visentini, K. Bryan and Gian Luca Foresti, "Markov Logic Networks for context integration and situation assessment in maritime domain", **Proceedings of the 15th International Conference on Information Fusion**, July 9-12, Singapore, pp. 1534-1539, 2012.

[20] L. Snidaro, I. Visentini, J. Llinas, G.L. Foresti, "Context as binding element for multi-level fusion", **Proceedings of the 16th International Conference on Information Fusion**, July 9-12, 2013, Istanbul, Turkey.

[21] I. Visentini and L. Snidaro, "Integration of Contextual Information for Tracking Refinement", **Proceedings of the 14th International Conference on Information Fusion**, July 5-8, 2011, Chicago, IL, USA.

[22] G. Ferrin, L. Snidaro and G.L. Foresti, "Contexts, Co-texts and Situations in Fusion Domain", **Proceedings of the 14th International Conference on Information Fusion**, July 5-8, 2011, Chicago, IL, USA.