Ph.D. in Electronic and Computer Engineering
Dept. of Electrical and Electronic Engineering
University of Cagliari

New Challenges in HCI: Ambient Intelligence for Human Performance Improvement

Massimiliano Dibitonto

Advisor: Prof. Fabio ROLI
Curriculum: ING-INF/05

XXIV Cycle
March 2012

Ph.D. in Electronic and Computer Engineering
Dept. of Electrical and Electronic Engineering
University of Cagliari

New Challenges in HCI: Ambient Intelligence for Human Performance Improvement

Massimiliano Dibitonto

Advisor: Prof. Fabio Roli
Curriculum: ING-INF/05

XXIV Cycle March 2012

To my family

Acknowledgements

Even if sometimes, while studying, writing or thinking one can be feel alone doctoral studies are not a lonely journey. They are a path made of people, of guides able to inspire, of obstacles, of successes and defeats, and of knowledge and experience. And in this journey there are many people that I have to thank for their help and for the inspiration that they gave to me. First of all I would like to thank my supervisors, prof. Fabio Roli and prof. Gian Luca Marcialis for their guidance, and advices throughout the Ph.D. A special thank goes to prof. Carlo Maria Medaglia for his encouragement and for the great time working together inside the CATTID laboratories in Rome.

It would have been impossible to arrive until this point without the love and the support of my family and my friends. Finally a thank goes to all the people that in different ways have contributed to this journey.

Abstract

Ambient Intelligence is new multidisciplinary paradigm that is going to change the relation between humans, technology and the environment they live in. This paradigm has its roots in the ideas Ubiquitous and Pervasive computing. In this vision, that nowadays is almost reality, technology becomes pervasive in everyday lives but, despite its increasing importance, it (should) becomes "invisible", so deeply intertwined in our day-to-day activities to *disappear into the fabric of our lives*. The new environment should become "intelligent" and "smart", able to actively and adaptively react to the presence, actions and needs of humans (not only users but complex human being), in order to support daily activities and improve the quality of life. Ambient Intelligence represents a trend able to profoundly affect every aspect of our life. It is not a problem regarding only technology but is about a new way to be "human", to inhabit our environment, and to dialogue with technology. But what makes an environment smart and intelligent is the way it understands and reacts to changing conditions.As a well-designed tool can help us carry out our activities more quickly and easily, a poorly designed one could be an obstacle. Ambient Intelligence paradigm tends to change some human's activities by automating certain task. However is not always simple to decide what automate and when and how much the user needs to have control.

In this thesis we analyse the different levels composing the Ambient Intelligence paradigm, from its theoretical roots, through technology until the issues related the Human Factors and the Human Computer Interaction, to better understand how this paradigm is able to change the performance and the behaviour of the user. After a general analysis, we decided to focus on the problem of smart surveillance analysing how is possible to automate certain tasksthrough a context capture system, based on the fusion of different sources and inspired to the paradigm of Ambient Intelligence. Particularly we decide to investigate, from a Human Factors point of view, how different levels of automation (LOAs) may result in a change of user's behaviour and performances. Moreover this investigation was aimed to find the criteria that may help to design a smart surveillance system.

After the design of a general framework for fusion of different sensor in a real time locating system, an hybrid people tracking system, based on the combined use of RFID UWB and computer vision techniques was developed and tested to explore the possibilities of a smart context capture system.

Taking this system as an example we developed 3 simulators of a smart surveillance system implementing 3 different LOAs: manual,low system assistance, high system assistance.

We performed tests (using quali-quantitative measures) to see changes in performances, Situation Awareness and workload in relation to different LOAs.

Based on the results obtained, is proposed a new interaction paradigm for control rooms based on the HCI concepts related toAmbient Intelligence paradigm and especially related to Ambient Display's concept, highlighting its usability advantages in a control room scenario.

The assessments made through test showed that if from a technological perspective is possible to achieve very high levels of automation, from a Human Factors point of view this doesn't necessarily reflect in an improvement of human performances. The latter is rather related to a particular balance that is not fixed but changes according to specific context. Thus every Ambient Intelligence system may be designed in a human centric perspective considering that, sometimes less can be more and vice-versa.

Contents

List of Figures

Index of Tables

Introduction

"It is four o'clock in the afternoon. Dimitrios, a 32 year-old employee of a major food-multinational, is taking a coffee at his office's cafeteria, together with his boss and some colleagues. He doesn't want to be excessively bothered during this pause. Nevertheless, all the time he is receiving and dealing with incoming calls and mails.

He is proud of 'being in communication with mankind': as are many of his friends and some colleagues. Dimitrios is wearing, embedded in his clothes (or in his own body), a voice activated 'gateway' or digital avatar of himself, familiarly known as 'D-Me' or 'Digital Me'. A D-Me is both a learning device, learning about Dimitrios from his interactions with his environment, and an acting device offering communication, processing and decision-making functionality. Dimitrios has partly 'programmed' it himself, at a very initial stage. At the time, he thought he would 'upgrade' this initial data periodically. But he didn't. He feels quite confident with his D-Me and relies upon its 'intelligent ' reactions. At 4:10 p.m., following many other calls of secondary importance – answered formally but smoothly in corresponding languages by Dimitrios' D-Me with a nice reproduction of Dimitrios' voice and typical accent, a call from his wife is further analysed by his D-Me. In a first attempt, Dimitrios' 'avatar-like' voice runs a brief conversation with his wife, with the intention of negotiating a delay while explaining his current environment. Simultaneously, Dimitrios' D-Me has caught a message from an older person's D-Me, located in the nearby metro station. This senior has left his home without his medicine and would feel at ease knowing where and how to access similar drugs in an easy way. He has addressed his query in natural speech to his D-Me. Dimitrios happens to suffer from similar heart problems and uses the same drugs. Dimitrios' D-Me processes the available data as to offer information to the senior. It 'decides' neither to reveal Dimitrios' identity (privacy level), nor to offer Dimitrios' direct help (lack of availability), but to list the closest drug shops, the alternative drugs, offer a potential contact with the self-help group. This information is shared with the senior's D-Me, not with the senior himself as to avoid useless information overload.

Meanwhile, his wife's call is now interpreted by his D-Me as sufficiently pressing to mobilise Dimitrios. It 'rings' him using a pre-arranged call tone. Dimitrios takes up the call with one of the available Displayphones of the cafeteria. Since the growing penetration of D-Me, few people still bother to run around with mobile terminals: these functions are sufficiently available in most public and private spaces and your D-Me can always point at the closest...functioning one! The 'emergency' is about their child's homework. While doing his homework their 9 year-old son is meant to offer some insights on everyday life in Egypt. In a brief 3-way telephone conference, Dimitrios offers to pass over the query to the D-Me to search for an available direct contact with a child in Egypt. Ten minutes

later, his son is videoconferencing at home with a girl of his own age, and recording this real-time translated conversation as part of his homework. All communicating facilities have been managed by Dimitrios' D-Me, even while it is still registering new data and managing other queries. The Egyptian correspondent is the daughter of a local businessman, well off and quite keen on technologies. Some luck (and income...) had to participate in what might become a longer lasting new relation."[46].

This is one of the scenarios depicted by EU ISTAG in 2001 to offer glimpses about a possible future (the narration took place in 2010) and inspirations for future researches on Ambient Intelligence. Reading this scenario after ten years we can find many elements that nowadays, can be recognized as existing technologies, active research fields but is also highlighted the multidisciplinarity and the link with many aspect of human life and society. But the more interesting point that emerges is the importance of the user in theDNA of Ambient Intelligence. Indeed Ambient Intelligence is new paradigm, rooted in the ideas of Ubiquitous and Pervasive computing, that doesn't focus on a single technology, it rather foster using and ensemble of technologies and techniques to improve the people's quality of life, bringing the into the foreground and making technology disappear. Indeed, with Ambient Intelligence, we don't talk anymore about users or customersbut we talk about people embracing a higher level of complexity related to the needs, goals and intentions of the individuals. Ambient Intelligence aims to bring technology in the background making it "disappear". It happens when using a service (technology enabled) become "natural", perfectly integrated in the flow of user's activities.

Going deeper on this concept, one of the main characteristics of the new paradigm is that technology should understand the user's needs and intentions and the context and answer to them in an intelligent way. While we can judge with a quantitative measure the performance of a certain technology (intended as the ability to perform a task using a certain quantity of resources), is more difficult to define and measure when it becomes "intelligent" for the user.

New technologies and solutions act changing the task the user has to accomplish, whit the need to acquire new skill or changing his/her behaviour. When we use a washing machine it does the hard work for us, however we have to learn how to use it, to reach our goal with success. Even if the example if very simplistic it clearly show how the utility of a technology is in the balance between the complexity and the effort of the new task related with the old one. Ambient Intelligence aims make the environment (the ambient) around us able to support our daily activities, give us new opportunities and improve the quality of our lives. This will happen adding smart technologies in it, automating certain activities thus changing the way people act. However automating a task means both reduce the effort (or the boredom) of the user but also deprive him/her of the control of the situation, putting him/her "out of the loop". There is not a given rule but this balance should be evaluated every time, considering the specific context in a human-centered design of ambient intelligence. For example in a smart home scenario we expect that, in chase of fire, the system automatically turns the fire extinguishers on, while let us take control of the air conditioning system to adjust the temperature of a room according to our desire.Probably, in future years, the success of Ambient Intelligence solutions will be dictated more by their capacity to empower the user, to provide "natural" and intelligent form of interaction.

In this work, after an analysis of the Ambient Intelligence paradigm, of its theoretical and technological enablers and of its general implications related to Human Computer Interaction, focuses on the problem of finding the right level of automation (LOA), in an Ambient Intelligence scenario, that is really able to empower the final user. As the high complexity of analysingthe different contexts

in which the Ambient Intelligence could be applied we decided to focus on a specific scenario: smart surveillance. We carried on our worktrying to touch all the important levels of an Ambient Intelligence system. At first we analysed the problems related to a general surveillance scenario, identifying them mainly as giving the user the right support in order to have a good level of Situation Awareness, intended as the ability to perceive, understand and make future projection of a certain situation. At first we faced the problem from a technological perspective aiming to develop a context capture system able to have a good perception and, in certain case, understanding of a certain situation in order to provide these information to a human user. According to the fundamental characteristic of the Ambient Intelligence paradigm, that doesn't rely on a single technology but rather on the way they are mixed together we used a sensor fusion strategy, designing and implementing an hybrid people tracking system based on the combined use of an RFID UWB real time locating system and a Computer vision system. We proposed an architecture and a fusion model able to harvest the data achieved by the two subsystems using both complementary and redundant strategies. We tested the system to assess its performance and reliability.
In a second step, according to the characteristics of the proposed context capture system, we tested, from an Human Factors perspective, how much different LOA could change human performance in terms of success rate, Situation Awareness and workload. Indeed our hypothesis is that an higher level of automation doesn't necessary match a real user empowerment. Ambient Intelligence implies also different forms of user interfaces and a different organization of the environment. To explore this issue a Natural User Interface was proposed, applying the concept of Ambient Display to a smart control room, highlighting both the new interaction paradigm and also how different devices could be used together to form a smart environment.

Contribution

The main contributions presented in this work are

- a proposal of an hybrid people tracking system[41]based on the combined use of RFID UWB and computer vision system to achieve, through sensor fusion techniques, better performances and reliability. Using cooperative and redundant sensor fusion strategies indeed is possible to achieve a better context understanding, that is recognize to be a fundamental element in many Ambient Intelligence scenarios;
- the investigation on the effects of different Levels of Automationon a human operator performing a video surveillance task in an Ambient Intelligence scenario. Indeed to an higher level of automation, realized through technology, doesn't always match a real advantage for the user. For this reason choosing a correct level of automation could be a success factor in the design of an Ambient Intelligence system.

The techniques and model used in the present work could be generalized and applied to other context offering a framework useful for human-centered ambient intelligence design with a special attention to the effect of the automation on the user.
Therefore, this dissertation aspire to represents a valid and useful contribute to the body of HCI and Human Factors research on Ambient Intelligence paradigms.

Structure

In chapter 1we analyse different definitions of Ambient Intelligence, trying to highlight the most important characteristics of the new paradigm. Moreover we go through the theoretical and technological roots of Ambient Intelligence, giving a rapid look to the concept of Ubiquitous and Pervasive Computing and the Internet of things. In chapter 2the enabling technologies of Ambient Intelligence are briefly revised with special attention to Wireless Sensor Networks. In chapter 3 is highlighted the impact of Ambient Intelligence on the Human Computer Interaction. A quick review and analysis is made about the new metaphors and User Interfaces used in the Ambient Intelligence paradigm. In Chapter 4 we introduce the problem of improving human performance in a smart surveillance scenario, analysing issues related to surveillance tasks. A special attention is given to the concept of Situation Awareness, the different way to measure it and its role in surveillance tasks.
In Chapter 5 a context capture system is proposed and tested to show how, through sensor fusion techniques, is possible to use different systems to have a better understanding of a monitored area. In Chapter 6 is assessed how different levels of automation could affect users performances through test made on simulators based on the system proposed in Chapter 5. In Chapter 7 a Natural User Interface for smart control rooms is proposed and tested.

Chapter 1

Ambient Intelligence: a definition

Ambient Intelligence is a concept developed in the last decades of the last century. It is a wide, multidisciplinary paradigmthat draws a new kind of relationship between humans, their environment and the technology. For this reason a number of definitions can be found in literature, each focusing on a different aspect. A widely accepted definition of the concept of AmI comes from the ISTAG[46], a group in charge of giving advice to the EU Commission on the overall strategy to be followed in carrying out the Information and Communication thematic priority under the European research framework.

"The concept of Ambient Intelligence (AmI) provides a vision of the Information Society where the emphasis is on greater user-friendliness, more efficient services support, user-empowerment, and support for human interactions. People are surrounded by intelligent intuitive interfaces that are embedded in all kinds of objects and an environment that is capable of recognising and responding to the presence of different individuals in a seamless, unobtrusive and often invisible way."

This definition is focused on the users rather than on technology. In an AmI world, massively distributed devices operate collectively while embedded in the environment using information and intelligence that are hidden (and distributed)in a interconnected network.

AmI wasn't meant to increase functional complexity (even if it could be a side effect) but to support peoples' lives in terms of care, wellbeing, education, and creativity, contributing to the development of easy to use and simple to experience products and services, that make sense in the first place.

As observed by Aarts and Marzano [2] in this vision the technology moves in the background, becoming the "Ambient" and the user comes in the foreground. He/she faces new kind of interfaces that allow intelligent and meaningful interactions. Ambient Intelligent environments will show their "intelligence" on one hand by the social nature of the interface, a kind of dialogue with the user, and on the other hand withthe ability of the system to adapt itself to its users and environments. The social character of the user interface will be determined by how well the system's behaviour will meet the social and cultural context of the user and self-adaptability depend on how the system is able to understand the context and react to its changes

Following Aarts and colleagues [3] the notion of ambience in AmI refers to the environment and reflects the need to embed technology in a way that became unobtrusively integrated into everyday

objects while the intelligence is related to the ability of the digital surrounding to exhibit specific form of social interaction with the people that live in the "ambient". Authors identify salient features of AmI as:

- integration through large-scale embedding of electronics into the environment;
- context-awareness through user, location, and situation identification;
- personalization through interface and service adjustement;
- adaptation through learning
- anticipation through reasoning.

Also this definition focuses on the characteristic of AmI to improve the quality of peoples' lives. The technology factor, even if is important and is an enabler, is not enough to reach the full expression of AmI. Indeed an investigation on the human factors, natural interaction, and human behaviour is needed to realize a real user empowerment.
Cook and colleagues [33] have made a survey about different definitions of AmI. In their research they highlighted the features that are expected to be present in an AmI paradigm (and in technologies involved):

- Sensitive: able to perceive information about the context;
- Responsive: able to respond to the presence of persons in the environment;
- Adaptive: able to respond in different way adapting to different situations;
- Transparent: invisible to the user;
- Ubiquitous: present everywhere;
- Intelligent: able to respond and adapt in intelligent way (here the concept of intelligence is related to the Artificial Intelligence paradigms).

In these researches is pointed out the importance of the technology used to perceive information about the environment,and the centrality of the user in the design of the services to be offered in "intelligent way".

Moreover AmI itself is connected to a number of different concept and paradigms. The most important paradigms that merges and gives origin to AmI are:

- Ubiquitous Computing;
- Pervasive Computing;
- Internet of Things.

Approaching the field of Ambient Intelligence it is necessary to refer to Mark Weiser and his "Computer for the 21th century" [155]. In his work Weiser thought that the computer of the 21th century should be "invisible" to its users, indeed it would be embedded in the environment. The "invisibility"is related to the capacity of the technology to help users to reach their goals in the less obtrusive way, as our interaction with ink and printed word. If a page is well presented we rarely dwell the printing technology but we go directly to the information that is contained. Weiser depicted a new kind of relation between humans and computer moving the attention from technology to the user and his goals.Moreover he saw, in the evolution of computing, three main waves, that were preparing the right ecosystem for the "invisible" computer. In the first wave (1960-1980) there were the mainframes, big computers, expensive and difficult to use. Every computer was typically used by

many individuals (usually well trained users). In the second wave (1980-1990) the diffusion of Personal Computers changed the ratio to one user per computer. Moreover in this phase there is the beginning of the concept of "user" with the attention to the Human Factors, as personal computers (and the software to be used) were meant for a broad public, not only for scientist and engineers as in the first wave.

The progresses of technology let the cost drop, the size decrease and the processing power increase. This was the beginning of the third wave (2000 onwards) where a single user could access more than one computer that started to have different shapes and specialized functions as smartphones. PDA, mp3 players etc..

Analysing the three waves of computing, from mainframes to smartphones/smart object/smart ambient, we can see that there are several dimension of changes:

- Physical dimension: during years the scale factor has changed, from mainframes that take an entire room to microprocessors that may fits inside everyday objects.
- Mobility: reduced dimension lead to the possibility of use computer "in the wild", in mobility/nomadic scenario. Of course dimension is not the only enabling factor. Mobility implies also networking capabilities, energy efficiency (for battery duration) and, most important, new kind of user's needs and tack that could be accomplished.
- Processing power: the increase of processing capabilities allowed more complex application but also more attention for the human-computer interfaces. Indeed as the processing power increases a share of it could be dedicate to the human-computer interface in terms of responsiveness and graphical user interface.
- Relation with human and usability: on one hand the change is about the ratio between users and computers. In the mainframe era a single computer had many users with a time/shared organization while in the third wave the proportion is reversed. Moreover the different ratio implies also that computers become "personal" and customizable. The user can shape the device properties in relation to its personal needs and tastes. The user doesn't need to have a deep technical knowledge (as the operators using mainframes) to use a computer but it is designed taking into account the users' needs and his capabilities according to usability criteria.

It is possible to identify also a fourth wave where the ratio will be many users to many computers. When computers will become invisible and disappear in the environment people will use it quite unconsciously, and the computers will be shared among many users in a networked, decentralized service-oriented architecture.

The third wave is also the beginning of the Ubiquitous and Pervasive computing. The two terms are often used as synonyms even if there are ambiguous interpretation. However the concepts they refer to have wide overlapping area so it is not fundamental to have a strong difference Ubiquitous Computing isa term first used by Weiser at Xerox Parc in 1988 andprovides a new scenario in which information is available through a fine-grained distributed network made of a variety of electronic device.

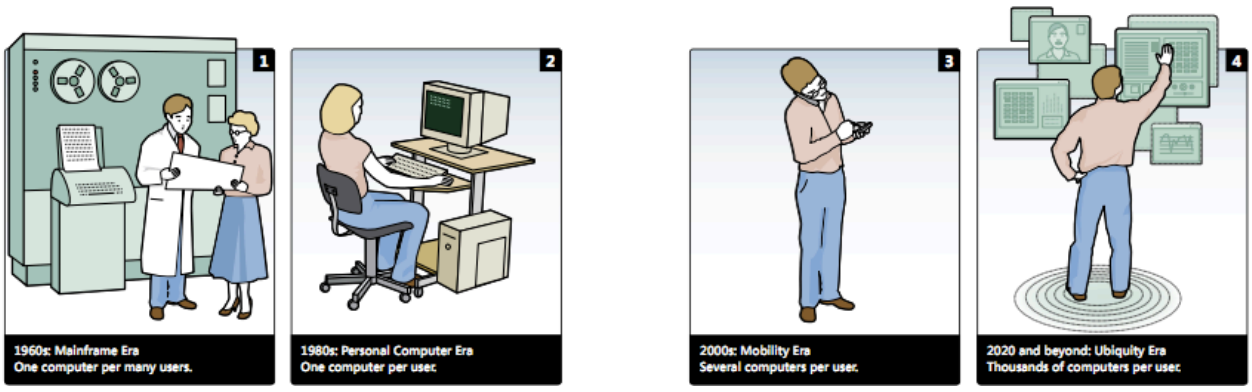

Figure 1 ***The four wave of computing as seen by IBM*** **[75]**

Pervasive Computing is concept defined almost in the same period by IBM. According toMerkand colleagues [105] it is "[…] the convenient access, through a new class of appliances, to relevant information with the ability to easily take action on it when and where you need it". Within pervasive computing environments, computing is spread throughout the environment, users are mobile, information appliances are becoming increasingly available, and communication is made easier – between individuals and things, and between things [7].
According to Bick and Kummer[12]the emphasis of Pervasive Computing is more on the software properties of services than on the device properties as in the case of mobile computing, which result from weight, size, and other physical constraints. The salient properties of Pervasive Computing can be formulated as follows:

- Ubiquitous. Widely present with identical appearance;
- Interactive. Control through multi-modal user interfaces;
- Interoperable. Plug and play with seamless integration and access;
- Distributed. Simultaneous access to resources including databases and processing units.
- Scalable. Adaptation of resources, quality of service and graceful degradation.

Moreover Ubiquitous/Pervasive computing are strictly related with the concept of mobile computing in the sense that mobile devices and their services become ubiquitously available .
The basic properties of mobile computing are well-known and, according to Adelstein and colleagues [4], can be summarized as follows

- Portable: small, battery-operated handheld devices with large footprints and multi-functional properties.
- Wireless: remote wireless connectivity with handover protocols and ad-hoc and TCP properties.
- Networked: remote data and service access with layered protocols.
- Location sensitive: global positioning with information on local position sensing.
- Secure: encryption based with authentication and conditional access securing privacy.

Often Ubiquitous and Pervasive Computing are used as synonyms of Ambient Intelligence but, as observed by Augusto[10], the first two concepts emphasize the physical presence and availability of resources but miss the key element of "Intelligence". According to Augusto "Intelligence" should be

intended as related to the field of Ambient Intelligence. As key features in AmI systems are flexibility, adaptation, anticipation and a proper interface to humans it is necessary a sort of intelligence able to perceive and learn from the environment and make some reasoning to answer to the needs of the users. As the example made by Augusto:*"That is how a trained assistant, e.g. a nurse, typically behaves. It will help when needed but will restrain to intervene unless is necessary. Being sensible demands recognizing the user, learning or knowing her/his preferences and the capability to exhibit empathy with the user's mood and current overall situation."*

In that way AmI concept refers, once again, to behaviour of technology that supports human activities responding with "intelligence" to the users' needs and to the changing variables of the environment. To avoid ambiguity the term "Smart Environment" [110]could be used to focus on the physical infrastructure (sensors, actuator and networks) that supports the system.

Nakashima and colleagues [110] highlight how AmI is a multidisciplinary area that embraces a number of pre-existing field of computer science, engineering but also human-computer interaction and cognitive sciences. While AmI nourishes from all those areas, the sum is bigger than its parts. Ami brings together networks, sensors, human-computer interfaces, pervasive computing, Artificial Intelligence as well as many other areas to provide flexible and intelligent services to the users.

A disrupting innovation in computing paradigms is represented by the exponential growth of telecommunication networks. Nowadays we have cheap and easy way to talk and exchange information with people (and machines) all over the world. Internet and the World Wide Web have been more than a technological innovation, in a decade they have changed (and they are still transforming) the entire society, the way the people communicate and, sometimes, the way they think.

Internet is a structure where can be built services (as email, web etc) to connect people and machines and exchange information. However nowadays a new paradigm is emerging, where objects connects to the network to offer or access information and offer services. This new paradigm is called the Internet of Things.

The concept of Internet of Things was originally coined by Kevin Ashton[9] of the MIT AutoID Center, during a presentation in 1999, to describe the possibility of using RFID tags in supply chains as pointers to Internet databases which contained information about the objects to which the tags were attached. The phrase maintained this meaning, until 2004, when, for the first time a world where "everyday objects [had] the ability to connect to a data network" was conceived [63]representing an higher level of complexity. Innovative concepts such as the extreme device heterogeneity and IP-based, narrow-waist protocol stack were for the first time introduced for what was also called Internet.

In the last years the hype surrounding the IoT grew in proportions. In the last years, quite a few definitions have been given and we will analyse them briefly in order to provide a better definition of the Internet of Things phrase. In the final report of the Coordination and Support Action (CSA) for Global RFID-related Activities and Standardisation (CASAGRAS)[25] project the reader can find a compiled list of definitions which capture different aspects of and meanings given to the concept of Internet of Things:

Initial CASAGRAS definition*: "A global network infrastructure, linking physical and virtual objects through the exploitation of data capture and communication capabilities. This infrastructure includes existing and evolving Internet and network developments. It will offer specific object identification, sensor and connection capability as the basis for the development of independent cooperative services and applications. These will be characterised by a high degree of autonomous data capture, event transfer, network connectivity and interoperability",*

The CASAGRAS definition was given in the first part of year 2009, and was then confirmed in the final report of the project. In this definition the IoT is first and foremost a network infrastructure. This is coherent with the semantic meaning of the phrase which assumes that the IoT builds upon the existing Internet communication infrastructure. The definition is also focused on connection and automatic identification and data collection technologies that will be leveraged for integrating the objects in the IoT.

SAP definition from Stephan Haller [73]: *"A world where physical objects are seamlessly integrated into the information network, and where the physical objects can become active participants in business processes. Services are available to interact with these 'smart objects' over the Internet, query and change their state and any information associated with them, taking into account security and privacy issues".*

We would like to note here the focus on the physical objects which are in the center of the attention as main participants of the IoT. They are described as active participants in the business processes. Besides, the IoT here is more a vision than a global network, as the word "world" would suggest. Also the idea of using services as communication interfaces for IoT

is explicited. Services will soon become one of the most popular tools to broaden the basis of communication interoperability in the IoT vision. Security and privacy, though not related to the definition of IoT, are also highlighted as critical issues.

Future Internet Assembly/Real World Internet definition: The IoT concept was initially based around enabling technologies such as Radio Frequency Identification (RFID) or wireless sensor and actuator networks (WSAN), but nowadays spawns a wide variety of devices with different computing and communication capabilities – generically termed networked embedded devices (NED). […] More recent ideas have driven the IoT towards an all encompassing vision to integrate the real world into the Internet […].

More recent definitions seem to emphasize communication capabilities, and to assign a certain degree of intelligence to the objects [15]:

"*a world-wide network of interconnected objects uniquely addressable, based on standard communication protocols."*

"Things having identities and virtual personalities operating in smart spaces using intelligent interfaces to connect and communicate within social, environmental, and user contexts."

In conclusion, we can thus identify two different meanings (and thus definitions) of the phrase: the IoT network and the IoT paradigm. First and foremost, the Internet of Things is a global network, an extension of the current Internet to new types of devices – mainly constrained devices for WSANs and auto-ID readers –, aiming at providing the communication infrastructure for the implementation of the Internet of Things paradigm. The Internet of Things paradigm, on the other hand, refers to the vision of connecting the digital and the physical world in a new worldwide augmented continuum where users, either humans or physical objects (the things of the Internet of Things), could cooperate to fulfil their respective goals.

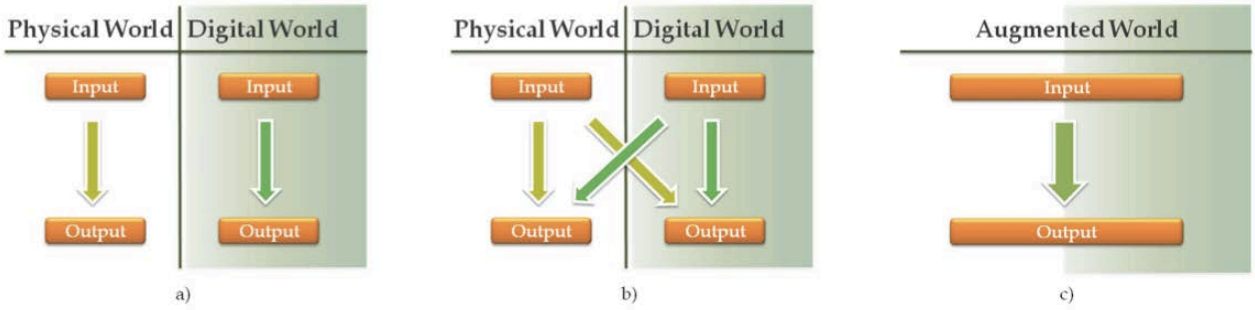

Figure 2: *The paradigm of IoT*: from the current situation where digital and physical environments are uncoupled (a), to one where physical and digital world can interact (b) and finally to one where physical and digital worlds are merged sinergically in an augmented world (c).

In order to realize the IoT paradigm, the following features will be gradually developed and integrated in or on top of the Internet of Things network infrastructure, slowly transforming it into an infrastructure for providing global services for interacting with the physical world:

- object identification and presence detection
- autonomous data capture
- autoID-to-resource association
- interoperability between different communication technologies
- event transfer
- service-based interaction between objects
- semantic based communication between objects
- cooperation between autonomous objects.

The generic IoT scenario can be identified with that of a generic User that needs to interact with a (possibly remote) Physical Entity of the physical world. In this short description we have already introduced the two key actors of the IoT. The User is a human person or a software agent that has a goal, for the completion of which the interaction with the physical environment has to be performed through the mediation of the IoT. The Physical Entity is a discrete, identifiable part of the physical environment that can be of interest to the User for the completion of his goal. Physical Entities can be almost any object or environment, from humans or animals to cars, from store or logistic chain items to computers, from electronic appliances to closed or open environments.
In the digital world Digital Entities are software entities which can be agents that have autonomous goals, can be services or wimple coherent data entries. Some Digital Entities can also interact with other Digital Entities or with Users in order to fulfill their goal. Indeed, Digital Entities can be viewed as Users in the IoT context. A Physical Entity can be represented in the digital world by a Digital Entity which is in fact its Digital Proxy. There are many kinds of digital representations of Physical Entities that we can imagine: 3D models, avatars, objects (or instances of a class in an object-oriented programming language) and even a social network account could be viewed as such. However, in the IoT context, Digital Proxies have two fundamental properties:

- they are Digital Entities that are bi-univocally associated to the Physical Entity they represent. Each Digital Proxy must have one and only one ID that identifies the represented object. The association between the Digital Proxy and the Physical Entity must be established automatically
- they are a synchronized representation of a given set of aspects (or properties) of the Physical Entity. This means that relevant digital parameters representing the characteristics of the Physical Entity can be updated upon any change of the former. In the same way, changes that affect the Digital Proxy could manifest on the Physical Entity in the physical world.

While there are different definitions of smart objects in literature (Kortuem et al., 2009), we define a Smart Object as the extension of a Physical Entity with its associated Digital Proxy.

We have chosen this definition as, in our opinion, what is important in our opinion is thesynergy between the Physical Entity and the Digital Proxy, and not the specific technologies which enable it. Moreover, while the concept of "interest" is relevant in the IoT context (you only interact with what you are interested in) the term "Entity of Interest" [73]focuses too much attention on this concept and doesn't provide any insight on its role in the IoT domain. For these reasons we have preferred the term Smart Object, which, even if not perfect (a person might be a Smart Object), is widely used in literature.

Indeed, what we deem essential in our vision of IoT though, is that any changes in the properties of a Smart Object have to be represented in both the physical and digital world. This is what actually enables everyday objects to become part of the digital processes. This is usually obtained by embedding into, attaching to or simply placing in close vicinity of the Physical Entity one or more ICT devices which provide the technological interface for interacting with or gaining information about the Physical Entity, actually enhancing it and allowing it to be part of the digital world. These devices can be homogeneous as in the case of Body Area Network nodes or heterogeneous as in the case of RFID Tag and Reader. A Device thus mediates the interactions between Physical Entities (that have no projections in the digital world) and Digital Proxies (which have no projections in the physical world) extending both. From a functional point of view, Device has three subtypes:

- Sensors can provide information about the Physical Entity they monitor. Information in this context ranges from the identity to measures of the physical state of the Physical Entity. The identity can be inherently bound to that of the device, as in the case of embedded devices, or it can be derived from observation of the object's features or attached Tags. Embedded Sensors are attached or otherwise embedded in the physical structure of the Physical Entity in order to enhance and provide direct connection to other Smart Objects or to the network. . Thus they also identify the Physical Entity. Sensors can also be external devices with onboard sensors and complex software which usually observe a specific environment in which they can identify and monitor Physical Entities, through the use of complex algorithms and software training techniques. The most common example of this category are face recognition systems which use the optical spectrum. Sensors can also be readers (see Tags below).
- Tags are used by specialized sensor devices, usually called readers in order to support the identification process. This process can be optical as in the case of barcodes and QRcode, or it can be RF-based as in the case of microwave car plate recognition systems and RFID.
- Actuators can modify the physical state of the Physical Entity. Actuators can move (translate, rotate, ...) simple Physical Entities or activate/deactivate functionalities of more complex ones.

It is also interesting to note that, as everyday objects can be logically grouped together to form a composite object and as complex objects can be divided in components, the same is also true for the Digital Entities and Smart Objects which can be logically grouped in a structured , often hierarchical way. As previously said, Smart Objects have projections in both the digital and physical world plane. Users that need to interact with them must do so through the use of Resources. Resources are digital, identifiable components that implement different capabilities, and are associated to Digital Entities, specifically to Digital Proxies in the case of IoT. More than one Resource may be associated to one Digital Proxy and thus to one Smart Object. Five general classes of capabilities can be identified and provided through Resources:

- retrieval of physical properties of the associated Physical Entity captured through Sensors;
- modification of physical properties of associated Physical Entity through the use of Actuators;
- retrieval of digital properties of the associated Digital Proxy; • modification of digital properties of the associated Digital Proxy;
- usage of complex hardware or software services provided by the associated Smart Object.

In order to provide interoperability, as they can be heterogeneous and implementations can be highly dependent on the underlying hardware of the Device, actual access to Resources is provided as Services.

In an AmI scenario Internet of Things is an enabling infrastructure, a fabric to build intelligent systems. Interconnected Smart Objects able to sense, to reason and to respond to human needs becoming also a new kind of interface with humans.

Chapter 2

AmI's Technologies

2.1 Introduction

Since the first vision of Weiser [155] a lot of technical problems have been overcome, many technologies have become cheap and widespread, and a lot of AmI scenarios have become possible. Particularly there are several key technologiesthat could be considered enabler of AmI scenarios. Here we will go through the main enabling technologies trying to point out their contribution and how they influence and are influenced by AmI paradigms.

Before going deeper in single technologies it is necessary to make a brief reflection on the Moore's Law. Gordon Moore, founder of Intel, in 1965 [108], analysing factors affecting profitability of semiconductor manufacturing predicted that economics would have forced to squeeze more and more transistors in single silicon chip (he predicted 65,000 by 1975). The Moore's prediction demonstrates to fit very well the trends of the semiconductor industry and becoming a "law" (considered by someone as a law of nature) that states that the transistors' density doubles each year, with the corollary of an increase of processing power and miniaturization of components. This law and its corollaries are at the base of the ubiquitous/pervasive computing paradigms already analysed.

Kuniavsky [95], looking at Moore's law,cleverly pointed out another corollary. According to him, we are in the "hidden middle" of Moore's law. As the number of transistors and (most important) the processing power increase the cost of the new CPUs remains almost constant, thus meaning a drop of the cost per transistor that translates in a reduction of the cost for older (but useful) technology. As Kuniawsky observes that*"Although old technology gets cheaper, it loses none of its ability to process information. Thus, older information processing technology is still really powerful3 but now it is (almost) dirt cheap"*.

The technologies involved in AmI scenario are not necessarily the most powerful ones, but the ones that better fits a given scenario.

2.2 Context Capture Technologies

A common character in definitions of an AmI system is that it has to be context-aware, having the ability tounderstand and answer to users' needs, without being intrusive. It means that the system needs to acquire all the information that are useful to understand the situation and to properly support the user. Nowadays we are used to deal with services that take advantage of certain knowledge of our context. The most common example is the use of GPS. It can be used inside a car navigator, to calculate the correct route to a destination, but it is used also on smartphones to tag contents uploaded on social networks, to find our car in a big parking to find friends near us, and so on. However location is only one of the many variables that can constitute the "context". Indeed even if context is a "common sense" concept it is hard to define it and enumerate all the elements that may constitute it. In this sense it has more to deal with philosophy and phenomenology that with computer science.
Many authors agreed that is difficult to have a common definition of context. Since the first definition of context-aware computing [131] many definition of context have been written. As in the review made by Zimmermann [162] there are definition that try to be general going by example and synonyms, but they risk to be recursive and not useful, while other definitions enumerates elements that constitute the context, but they risk to be not complete or relative only to some specific scenarios.
A widely accepted general definition of context is the one made by Dey [37]:
"Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between the user and the application, including the user and the applications themselves".
Even if the definition is general it gives some important points. The context is relative to some entity (including the user and the system) and it could be any information. This means that every time there is the need to define which information is context and which is not the criterion is to consider what is relevant for the interaction. Dey saythat:*"If a piece of information can be used to characterise the situation of a participant in an interaction, then that information is context."* So the information should be salient inside an interaction between a user and a system. In relation to this while giving a definition of context-aware computing, Dey introduces the concept of user's task. Indeed the interaction takes place because the user want to achieve some sort of goal and the context will be constituted of the information relevant to the accomplishment of the task.
Another general definition that we can take into account is made by Coutaz [35] :*"context is about evolving, structured, and shared information spaces, and that such spaces are designed to serve a particular purpose. In ubiquitous computing, the purpose is to amplify human activities with new services that can adapt to the circumstances in which they are used"*. Here the important concept is that context evolves and serves a particular purpose.
However we need an operational definition of context to develop effectivecontext aware applications. Zimmermann extends the definition made by Dey adding five categories of information about context: *individuality, activity, location, time*, and *relations.*
"The activity predominantly determines the relevancy of context elements in specific situations, and the location and time primarily drive the creation of relations between entities and enable the exchange of context information among entities."

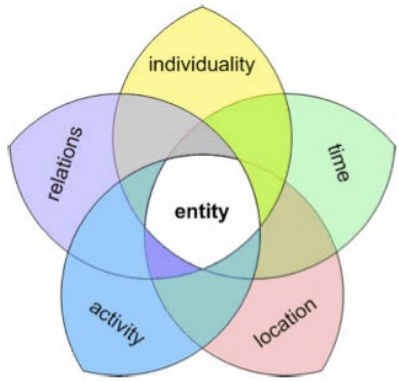

Figure 3:Five fundamental Categories for Context Information [162]

Individuality is about the entity. It could be natural, artificial, human or group of entity, sharing the same attributes. Entity may be real or virtual, may behave differently and have different roles in a context. Basically we can observe the state of the entity through related information.
Activity identifies the needs of the entity. Usually when an entity interacts with a context aware application is because is trying to achieve some goal, performing actions (divided into simple tasks) that will make the entity change his state. Identifying the goal of the entity is one of the most important things to realize an Ambient Intelligence application able to actively and transparently support human activities.
Location and *time* are two main fundamental physical dimension in which an entity and an action exist and take place. They could be real or virtual, absolute or relative.
Then *relations*occur among entities, space, time and activities.
As the entity interacting with the system is usually a human, we have also to consider the social context. The analysis of social context looks beyond the interaction between an individual user and a computer system, but focuses at the context in which that interaction emerges. It focuses on the social, cultural and organizational factors that affect interaction, and on which the user will draw in making decisions about actions to take and in interpreting the system's response. As in the Situated Activity theory as explained by Nardi [111]human activity is not consequence of a clear plan but it is decided in relation to the immediate circumstances and also, as the Situated Cognition theory says, in relation to the previous knowledge of the user.
From a sociological point of view the interaction between people and systems are themselves features of broader social settings, and those settings are critical to any analysis of interaction.
In the next paragraphs we will see how AmI systems are able to capture information about the context and, eventually,use them to take decision and perform actions.

2.2.1 Sensor Networks

To capture information about the real world AmI systems rely on a variety of sensors. In this perspective sensors are the keys that link available computational power with physical applications.
The software perceives the environment through sensors and uses this data to reason about the environment and the actions that eventually can be taken to have effect on it changing its state (i.e

turning on a heater when the temperature sensor say that the room temperature). Sensors have been designed to detect a wide range of phenomena and physical measurements as position measurement, detection of chemicals and humidity sensing, and to determine readings for light, radiation, temperature, sound, strain, pressure, position, velocity, and direction, and physiological sensing to support health monitoring. Indeed they could provide discrete data (on/off state) or analogue values (temperature, brightness). Thanks to miniaturization of electronic components sensors are typically quite small and can be integrated into almost any AmI application. Of course to be useful it is necessary o connect sensors to a processing unit and (eventually) to an actuator. This could be done embedding all the components in a single object (i.e. an emergency lamp turning on when there is a black-out) or over a network.

To monitor large areas (i.e. an house/building) it is necessary to realize network of sensors able to collect data to a central system in charge of performing fusion over different information achieved. As we will see in the present work, are one of the fundamental elements to implement AmI scenarios.

The mainstream research on sensor networks began in the year 90s of the last century in the US. It was due mainly to DARPA and some important Universitieswith the objective to create tiny autonomous computer that could unobtrusively observe their environment. These kinds of sensor platforms were initially called *motes* to indicate devices that are nodes of a network able to report their information to a base. In this phase researches pointed out some of the main characteristics of these sensor networks, namely [14] :

- Primitive processing capability: the sensor should have a primitive but robust processing capability. This is used to perform simple processing on data achieved to reduce the amount of data transmitted, sending only useful information and reducing the computational load of the server;
- Self-organized networking: the sensor platforms should be able to collaborate to create ad-hoc networks even in places where there are no infrastructure. As will be analysed further this characteristic is important especially for wireless sensor networks;
- Low power operation: unobtrusiveness and invisibility of a sensor network means also that the user doesn't have to worry (too often) of batteries. As Motes were meant to be self sufficient, installable in place with no infrastructures, they should be able to operate on battery power for a reasonable amount of time. This also to the detriment of processing and networking hardware (especially when there is a wireless communication);
- Tiny form factor: sensor networks were defined also as "smart dust" [154]also because they are required to be small in order to be placed in the environment even in big quantities remaining unobserved and unobtrusive;
- Target applications: in the initial researches the sensors were designed for specific applications, programming them directly inside the platform's firmware focusing on efficiency rather than flexibility and ease of application development. This is one of the things that have been changes in recent development to fit, as we will see, Internet of Things scenarios.
- Lack of mechanical actuation: in this initial researches passive scenarios were investigated with no need of mechanical actuation. Moreover mechanical actuation usually require more power and the moving parts may affect reliability an long lasting life of motes themselves.

Berkeley Motes are an example of this kind of sensor. They communicate through a wireless network implementing ZigBee mesh networking stacks, and have a microcontroller embedded. One of the

important things about Berkeley Motes is that they use an operative systems called TinyOS. This is an open-source component-based embedded operating system address. Applications are programmed, compiled and statically linked with TinyOS code to ensure efficient application execution but is not meant for easy and rapid application development and for distributed applications.
Even if sensor network could be realized using wires, to be really pervasive sensor must be able to move freely and without s fixed infrastructure (one of the motes requirements). For this reason we will talk about Wireless Sensor Networks (WSN) an Wireless Sensor and Actor Networks (WSAN) to underline the mobility and pervasiveness of sensors in smart environments.
In relation to the spreading of ubiquitous/pervasive computing paradigms research started to investigate new domains (e.g. domotic) and new features were required to sensor networks. If motes were originally meant to be used in primitive environments (e.g. to create networks in battlefields) as they begin to be integrated in spaces equipped with networking and power facilities (and maybe other systems with information processing abilities) they can contribute to the creation of a smart environment, the enabling structure for any ambient intelligence scenario. One of the first example of research about smart spaces is the development of the concept of smart homes, focusing on the enhancement of the quality of life of the inhabitants.
While the basic constrains of the sensor platforms remains (mainly energy efficiency, form factor, networking), new needs raised to enable the integration with other system and also to facilitate the development of new application. According to the taxonomy of Bose [14]:

- Ease of application development: as the sensor platform should be integrated into a smart space, in an Internet of Things paradigm, and they became pervasive not only in the environment but also in human activities, is important to ensure easiness of application development, reducing the requirement of specific knowledge to the developers.
- Programmability: the applications running on the sensor network should be easily deployed and updated. Indeed a smart space could easily evolve to accomplish changing needs of the humans that habit the environment. The mote should be able to accept and execute a new application (with an adequate level of security) an application deployed over the network without a direct intervention (as in the case of Berkeley Motes)
- Mechanical actuation: as the system is required to produce effects on the environment (to manifest its smartness to the inhabitants) motes have to support actuators (e.g. switches , servos and motors) to control different aspects of the smart space.
- Standardized communication protocol: in an Internet of Things perspective a new entity should be able to seamlessly connect to the network (i.e. a person buy a new lamp that has to deal with other appliances for an energy efficiency purpose). For this reason sensor platforms have to implement standard communication protocols (i.e. 802.15.4) ;
- Externally executed application: as the motes have small processing and memory resources the computational load of complex application should be divided between sensors and an external(powerful) processing unit. This means realize a distributed application leaving to the sensors only a small pre-processing on data.

Sun's (Sun Microsystems, Inc.) Small Programmable Object Technology, or Sun SPOT [136] are a good example of easily programmable motes. They are wireless, battery operated and run a Java Virtual Machine called "Squawk". The main features of this VM are [135]:

1. was designed for memory constrained devices,
2. runs on the bare metal on the ARM,

3. represents applications as objects (via the isolate mechanism),
4. runs multiple applications in the one VM,
5. migrates applications from one device to another, and
6. authenticates deployed applications on the device.

The most interesting features are the possibility to write application in Java (a widely used programming language) and (point n.5) the ability to migrate applications among devices.

The extraordinary potential of wireless sensor networks is not so related to high local processing capacity of individual nodes (that is modest instead), but relies mainly in the possibility of the nodes, to coordinate with each other and to self-organize.Another important aspect is the way in which data travels between the base station and the locations where the phenomena are observed (routing). As in a WSN, it is important to lower the bandwidth utilization and the power consumption, medium access control (MAC) and routing algorithms are very important. Moreover routing algorithms should be able to rapidly adapt to change in number and physical distribution of sensors, to the eventual failure of a node or to the change of network topology

For ad hoc networks can be classified according to the way in which information is acquired and maintained and by which this and used to find the paths between nodes.

Generally, each node announces its presence in the network and listens to the communication between other nodes, which become known. Over time, each node acquires the knowledge of all network nodes and one or more ways to communicate with them, and in most cases, the data make many hops transmitting the packet to the nearest neighbour.

There are different technologies and protocols used to build sensor networks infrastructures as ZigBee over IEEE 802.15.4 are WiFi IEEE 802.11 and Ultra Wide Band over IEEE 802.15.4a .

As we can see from table 1, based on a n elaboration of the work of Zhang and colleagues [160] and Lee [99] these technologies differ under many aspects but the more important are energy consumption, data rate, range and other implicit services as localization.

ZigBee is a protocol for communications wireless infrastructure that provides a reliable and robust exchange of information between devices equipped with any kind of sensors. Is out of the scope of this work to deepen the problem of MAC and routing but is important to highlight the characteristic of the ZigBee protocol as it is becoming widely used in implementing WSAN. The ZigBee protocol "meets the unique needs of sensors and control devices. Sensors and controls don't need high bandwidth but they do need low latency and very low energy consumption for long battery lives and for large device arrays" [93]. Indeed ZigBee allows creating self-organized, multi-hop and reliable mesh networks that are also energy efficient. Usually there are two kinds of devices in a ZigBee network (thanks to the IEEE 802.15.4): (1) a full-function device (FFD) that could serve as network coordinator or as a device; (2) reduced-function device that is only able to talk to a FFD. An RFD could be a light switch. It doesn't need big resources, as it only has to signal if the light is on or off and change its state when asked by the coordinator. The coordinator instead may have to deal with more switches i.e. to turn on all the lights in a part of the room. FFD establish star networks with RFD while is able to create mesh networks with other FFDs. Only one FFD can be the global network coordinator.

UWB technologies are drawing great interest in the wireless community thanks to their properties that are suitable for many applications but especially for sensor networks. The principal UWB characteristic is to transmit electromagnetic signals with relatively large proportion of frequency spectrum. The main properties that make it suitable for sensor networks are (1)the resistance to severe

multipath and jamming (that may be frequent in indoor environments), (2) thanks to its noise-like signal creates low interferences to other systems, (3) has a low energy consumption and good time domain resolution allows a precise location and tracking (the use of UWB for localization and tracking will be discussed in paragraph 5.2.2). It has a network topology similar to ZigBee but its higher data rate makes it suitable for transferring higher amounts of data, enabling more complex applications.
WiFi is usually employed to create wireless local area networks. It has an high data rate but also an high energy consumption if compared with ZigBee and UWB. As nowadays WLANs are widely diffused both in private and in public spaces this technology represent an easy way to make thing connect to a network. For this reason we can find a lot of "smart objects", appliances as tv sets, game consoles, media centres that can connect to a WiFi network to share data, offering services or be commanded and configured. The high data rate allows to use these network to deliver big quantity of data, as video, and could be used to collect data achieved by other sensor networks.

	ZigBee	**WiFi**	**UWB**
IEEE spec.	802.15.04	802.11 a/b/g/n	802.11.4a
Frequency	868/915 MHz; 2.4 GHz	2.4 GHz; 5GHz	3.1-10.6 GHz
Data Rate	low, 250 kbps	High, up to 100+ Mbps for 802.11n	Medium, 1 Mbit/s mandatory and up to 27 Mbps for 802.11.4a
Transmission distance	Short, < 30 meters	Long, up to 100 meters	Short, < 30 meters
Location accuracy	Low, several meters	Low, several meters	High, < 50cm
Power consumption	Low, 20mW – 40mW	High, 500mW- 1W	Low, 30mW
Multipath performance	Poor	Poor	Good
Interference resilience Low Medium High with high	Low	Medium	High with high Interference resilience with high complexity receivers, low with simplest receivers
Interference to other systems	High	High	Low
Complexity and cost	Low	High	Low – medium – high are possible

Table 1 ***Characteristics of different technologies for WSN.***

2.2.2 Radio Frequency Identification

Determining the identity of an entity is a very important element for a context capture element. Indeed in many domains decisions are strictly related to identity of objects or people. There are various technologies for the automatic identification based on computer vision or other sensors (i.e. biometry). Among these RFID, for its characteristics, is recognized as one of the most promising technologies for the automatic identification inside Internet of Things and AmI paradigms.

The term RFID stands for Radio Frequency Identification and it indicates a wide range of technologies. It' s possible to generalize defining how RFID technology can uniquely and automatically determine the identity of an object through radio frequency.
RFID wirelessly exchanges information between a tagged object and a reader/writer. An RFID system is comprised of the following components:

- One or more tags (also called transponders), which consist of a semiconductor chip and antenna.
- One or more read/write devices (also called interrogators, or simply, readers)
- Two or more antennas, one or two on the tag and at least one on each read/write device.
- Application software and a host computer system.

Tags are usually applied to items, often as part of an adhesive bar-code label. Tags can also be included in more durable enclosures and in ID cards or wristbands. Readers can be unattended standalone units (such as for monitoring a dock door or conveyor line), integrated with a mobile computer for handheld or forklift use or incorporated into bar-code printers.
The reader sends a radio signal that is received by all tags present in the RF field tuned to that frequency. Tags receive the signal via their antennas and respond by transmitting their stored data. The tag can hold many types of data, including a serial number, configuration instructions, activity history (e.g., date of last maintenance, when the tag passed a specific location, etc.), or even temperature and other data provided by sensors. The reader receives the tag signal via its antenna, decodes it and transfers the data to the computer system through a cable or wireless connection. The data received in this way allow to research, identification, selection, spatial localization and tracking.
RFID technology has certain advantages over optical technologies (bar codes) and magnetic (magnetic stripe). With the radio frequency data transmission, the object to be identified and the system of identification should not be in contact, so it's not required the eye contact as in case of a systems-readable. The amount of data recorded is greater than the traditional magnetic and optical systems data, and also the transmission occurs at a higher rate than the other kind of systems and, depending on the type of chip used, the information can be re-written.
With RFID technology we are able to do multiple identification or to collect information from many code labels, whose distance from the player may range depending on the technology used, and we are also able to transmit this information to the management information system.
Thanks to a high reading reliability, RFID can operate in contaminated and dirty environments and have the ability to resist to the environmental chemical exposure, by using the proper packages. RFID can also operate if immersed in a fluid, inside the object you want to identify, or inside another container, if not completely metallic.
The TAG becomes an identification system that can track the history of a product from the initial stage of processing and that then can be used interactively throughout the production chain, to reach the retail sector and, in some cases, up to the consumer.
Below there's a table comparing the different technologies that allows us to have an overview of the individual features and of the possible fields of application.
To summarize, RFID offers several notable advantages over other forms of data collection:

- RFID enables monitoring and data collection in environment unfit for workers, because tag reading requires no labour.
- More than a thousand reads can be performed each second, providing high speed and great accuracy.

- The data on an RFID tag can be altered repeatedly. RFID does not require direct line of sight between tag and reader, making it suitable for manyapplications where bar codes are not viable.

Thousands of organizations in many industries have exploited RFID's advantages to develop operations that monitor processes, provide real-time data accuracy, track assets and inventory, and reduce labourrequirements.RFID technology can be used in conjunction with bar-code systems and Wi-Fi networks.

2.2.2 Vision

Video cameras are considered high-content sensors, which provide rich sources of information both for human observation and for computer interpretation. Indeed images and videos could be processed to become understandable by a computer in order to automate the extraction of high-level information that, eventually, could be used to trigger some events.
Since it is a very dynamic and broad research area we do not intend to present a complete survey but we will give an overview on the key aspect related to the present works.
Computer vision is an alive and challenging research fields. A lot of methods and techniques have been developed aiming to give to a computer the same (or higher) abilities of humans to understanding images, where understanding means extracting the symbolic information contained in it. Indeed computer vision could be very useful to extract information about the context. Moreover, as there is no direct contact with user or objects that, unlike sensors, doesn't have to wear or touch anything computer vision is unobtrusive for the user, even if requires a proper infrastructure.
We can divide, for the scope of the present work computer vision techniques used just to detect information about the environment and techniques used for a direct interaction with the user.
In paragraph 4.1 several algorithms widely used for automatic video surveillance are presented. They cover the most important aspects of monitoring an area, from people identification and tracking to the behaviour analysis.
In paragraph 3.1 are presented several cases where computer vision is used to build human computer interfaces to interact with computers and complex systems.
A computer vision system may differ from others depending on system's purposes and on the devices that compose it. However we can identify a set of devices that are present in any vision system and that are the core of the basic functionality.
The basic structure of any computer vision system has, therefore, the following elements:

- A processing unit (be it a CPU, an FPGA, a DSP or other).
- An electronic device for image acquisition. This device, known as Frame Grabber, allows the acquisition of the signal from the camera and takes it to the computer's memory, so that it can be analysed.
- A camera. The camera and the instrument able to impress the luminous intensity of the scene on its own sensor (CCD or CMOS), consisting of an array of photovoltaic elements (pixels), and to transform it into a digital signal.
- An optic apparatus. The optic apparatus is responsible for mapping the real world on the camera sensor (CCD). Each opticssufferof problems of image distortion that must be properly contrasted

- An illuminator. Lighting plays a key role in image processing and systems necessary to ensure better stability and may cause huge changes in the behaviour of the system. In addition, the careful choice of a light source can provide greater enhancement of the features that have to be extracted from images, such as edges or corners
- Software for image analysis. The mere presence of all the previous components of a system is not a tool for artificial vision. The presence of a suitable analysis software that applies the techniques and algorithms of computer vision allows extract from the scene displayed a quantity of information.

The quality of a vision system can be drawn from the analysis of these individual components, although for some of these elements, is not easy to obtain objective information of the quality attainable. For example, to assess the quality of a camera sensor it is not sufficient to consider the number of pixels, as this should always be compared to the camera field of view (FOV). For this reason, a 1 Mpixel sensor observing a field of view of 10 cm. defines a pixels/distance of 10 pixels/mm, with an association of 1 pixel = 0.1 mm. In this case, therefore we can assume a precision of one tenth of a millimetre. Nowadays, however, there is a special technique, known as sub-pixelling, which allows to go beyond the Nyquist criteria (for which the minimum appreciable value and twice of the sampling rate) by going to identify with a particular higher accuracy than can be obtained optically. In a few words obtain this accuracy the value of the pixel grey levels and its neighbours are considered. In any case, the accuracy of the system cannot be assessed by the accuracy of individual components, as, for example, the technique of sub-pixel lighting, which requires high stability is not always achievable with ease.

For its flexibility computer vision could be considered a cornerstone of AmI technologies. However many computer vision techniques are based on inference and probability and they could be affected andthe reliability of computer vision techniques may vary depending on the context. In an AmI scenario this problem could be overcome thanks to the integration with other systems. In Chapter 5 is presented an hybrid people tracking system based on the combined use of an UWR RFID system and a computer vision system.

2.2.3 Location and Tracking inside a sensor network

In a context capture the location of an entity is an important information. While in outdoor context the GPS (Global Positioning System) allows a quite precise localization in indoor context locating an entity is still a challenge. Sensor networks could be for location and tracking purpose thank to the ability, though specific techniques, to locate their own nodes. According to Patwari[121] "*in cooperative localization, sensors work together in a peer-to-peer manner tomake measurements and then form a map of the network*".

Before analysing the different strategies that could be used to locate the nodes of a sensor network we can make a different between implicit and explicit localization. The difference is about the main purpose of the network. Indeed as, localization techniques are affected by the network characteristic, the WSN could be build to explicitly support location services or this could be a secondary service and the structure of the network is optimized for other purposes.

First of all localization of the nodes must respect the main characteristic of a WSN, primary is to maintain the energy efficiency and the low cost of the network. For this reason should be avoided the excessive transmission of extra beacons for long distance and the installation of extra hardware (as

GPS chip). Instead we can take advantage of the large number of nodes in a WSN and their ability to exchange information with multi hop strategies. The solution proposed by Patwari is to use a small number of sensors that have a known (an exact) location and are called reference nodes. The rest of the nodes can determine their own coordinates in relation to the reference nodesFigure 4. If sensors are capable of high power and long distance transmission they can exchange information with multiple reference nodes (as in the case of WLANs), while energy efficient sensors could use a multi-hop strategy to reach the reference node (as in the case of ZigBee networks).

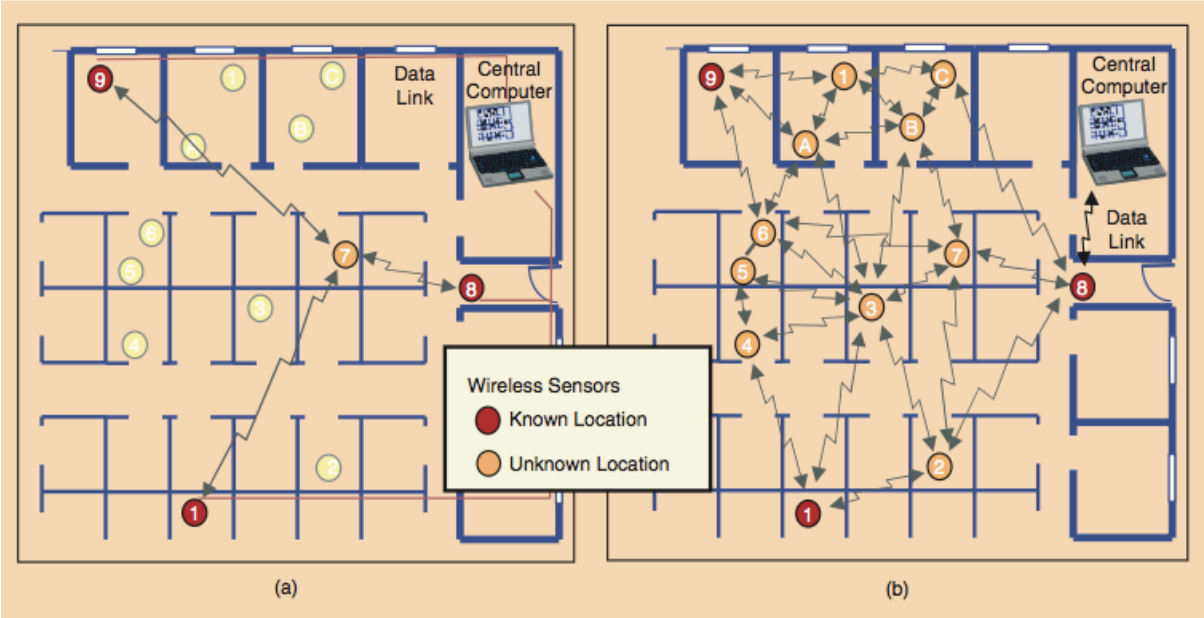

Figure 4 *Localization in a WSN*a) Traditional multilateration or multiangulation is a special case in which measurements are made only between an unknown- location sensor and known-location sensors. In (b) cooperative localization, measurements made between any pairs of sensors can be used to aid in the location estimate.

Different techniques could be used to locate sensor in a cooperative scenario:

- Received Signal Strength (RSS)
- Time of Arrival (TOA)
- Angle of Arrival (AOA)

The RSS technique is based on the strength of the signal received by a receiver. The main advantage is that there is no extra communication needed among sensors. Using the normal message exchanged a sensor could measure the strength of the received signal. Through multilateration algorithms is possible to determine the position of a node. However RSS measurements are unpredictable thus the techniques is prone to error. Multipath and shadowing are the two majors sources of environment dependence

The Time of Arrival technique is based on the measured time at which a signal first arrives at a receiver. The measured TOA is the time of transmission plus a propagation-induced time delay. This time is related to the distance between the transmitter and the receiver. As for RSS the measurements obtained are used in a multilateration algorithm to determine the position. Multipath can be a cause of error. Moreover a correct synchronization is needed among the nodes.

The Angle of Arrival technique is the most complex one, as it has to detect the direction of arrival of the signal coming from neighbour sensors. Moreover this measure cannot be used alone but is considered complimentary to RSS and TOA techniques. To detect the direction of arrival an array of sensors should be used thus there is a major complexity of the hardware that may be make this solution unfeasible in certain context.

Further, in the present work, a fusion strategy is presented to obtain a more precise and reliable localization using an UWB RFID system and a computer vision system.

Chapter 3

AmI a new Challenge in HCI

3.1 Introduction

Since the early history, humans tried to design and build tools to improve their life quality and their capabilities. The impact of these tools was so high to deeply influence all the history of the mankind, as we address certain historical periods by the dominant technology e.g. Iron Age, Bronze Age, Industrial Age. Over the centuries humans have made tools to overcome their physical limits as a force (levers, motors), senses (microscopes, telescopes, sound amplifiers) and even coming to replace parts of their bodies, as in the case of implants. But they have made also tools to improve their cognitive abilities. These could be defined as *cognitive artifacts*[117] that help us performing certain tasks, giving us the impression that our mental abilities are improved. When we take notes on a piece of paper (e.g. a telephone number or the shopping list) we are "expanding" our memory, that doesn't have to keep those information but only the place where the sheet is (not a simple task indeed!).
In this perspective the computer (in all its forms) could be considered as the most evolute form of cognitive artifact, due to the infinite number of human activities that it can support. Moreover it could be considered a meta-cognitive artifactas helps to build new artifacts (i.e. when a developer writes a new application).
But theidea that artifacts enhance or amplify human abilities may be misleading. As Norman [117] clearly points out:
"Artifacts may enhance performance, but as a rule they do not do so by enhancing or amplifying individual abilities. There are artifacts that really do amplify. A megaphone amplifies voice intensity to allow a person's voice to be heard for a greater distance than otherwise possible. This is amplification: The voice is unchanged in form and content but increased in quantity (intensity). But when written language and mathematics enable different performance than possible without their use, they do not do so by amplification: They change the nature of the task being done by the person and, in this way, enhance the overall performance."

In this perspective when the informational and processing structure of the artifact is combined with the task and the informational and processing structure of the human, the result is to expand and enhance cognitive capabilities of the total system of human, task, and artifacts.
Even if for an external viewer may seem that the user is performing the same task with the help of an artifact, for the individual person that has to perform a task, the artifact is something in the middle between him/her and the task itself. Using the artifact means to learn new things, new procedures, and to acquire and use new kind of information. The person's cognitive abilities are unchanged, but the overall performance is enhanced.
In a simplified view, humans are goal oriented[34]. They want to achieve some result (remember something, arrive in a destination) and they plan proper tasks according to context variables and personal experience. There are many possible combinations of tasks that can be used to reach some goal, implying different performance levels.
Norman uses the examples of achecklist as memory aid. From an external point of view, as said at the beginning of this paragraph, writing a list on a piece of paper appears to an external observer as a memory aid, but for the task performer using the list is a taskitself. Without the list we should do a planning and then remember it. With the list the planning task is done in advance so we just need to remember to consult the list and reading and interpreting items. As part of the work is done in advance (and maybe by someone else) the cognitive effort is distributed across time and people. Moreover the "advance planning" could be done in a convenient moment, when there are no other variables affecting the performance (stress, noise, time pressure). For this reason many security procedures include the use of a list made by experts long time in advance. After this example we can agree with Norman about three characteristics of cognitive artifacts as they

- distribute the actions across time (precomputation);
- distribute the actions across people (distributed cognition);
- change the actions required of the individuals doing the activity.

Artifacts are mediators between us and the world:while executing a task, artifacts are between our actions and the resulting effect on the reality; while perceiving, artifacts are in the middle between changes in the world and our detection and interpretation of its state.
Humans act through a feedback mechanism: an individual, which wants to achieve a goal, makes an initial evaluation of the situation and then performs an action that he/she expects to have an effect on reality. Then he perceives and interprets state changes and eventually, performs other actions to achieve his/her goal.
In consequence Norman points out that "*things can make us smart*" and "*things can make us dumb*" [116]in relation to their ability to allow us to perform a task and correctly interpret ambient states.

3.2 Smart Objects, Smart Spaces and Levels of Automation

Nowadays we are facing a deep change of the objects and the artifacts that we use every day. The evolution of technology and the paradigms of Pervasive-Ubiquitous computing are bringing to us new kind of objects, smart objects, that embody information processing, sensing, acting and networking

abilities. Information processing is no more only an objective but it is rather a component of objects that make them more desirable for the users[95].
Kuniawsky[95]calls "*information appliances*" objects designed to process and present information to the users with new human-computer interaction paradigms. Information processing can be viewed as a material that is an inseparable part of the object.

Figure 5 *The Ambient Umbrella*. The handle has light patterns that indicates weather according to forecasts downloaded from accuweater.com.

As an example the Ambient Umbrella (Figure 5) has a handle that illuminates to indicate rain, drizzle, snow or thunderstorms automatically receiving weather data from accuweather (www.accuweather.com). It radically changes the task the user has to accomplish. The user doesn't have to turn on his/her computer, connect to a website and look at the weather forecast, he/she just has to look at the handle and interpret the light signal.
Inside AmI scenarios the environment inhabited by the user starts to be populated by information appliances and smart objects. But these are not standalone elements. They are intelligent and maybe coordinated by a central system. They connect creating smart environments that integrate information, communication, and sensing technologies into everyday objects.
The way the system interacts with the user (and vice-versa) is a key factor to determine its quality, intended as the improvement that it is able to offer to the user. Here the concept of interaction is not related to a single action performed on a I/O device but it rather refers to the integration of technology in the flow of human activities. The system, in relation to its smartness, will manifest behaviour to the user, but, as humans and human activities are complex, it is not easy to have the right behaviour. Technology itself may only provide functionalities but the smartness, and the right impact on humans' quality of life, relies in the overall design of the system and in the human-computer interaction. Streitz[144] makes a distinction between two kind of smart spaces, highlighting their relation with users:

- system oriented, importunate smartness;
- people oriented, empowering smartness

In the first case the system acts even without a human in the loop by basing the decision on data collected. However those decisions may be unwelcomed by the human user. The author draws the example of a smart-fridge auto buying food when it detects to be quite empty. Indeed, sometimes, there are variables that the system may not perceive, or that are the result of an unpredictable event and the food bought by the fridge may not fit the user's needs.

The second type of smart spaces keeps the human in the loop and the user can always decide what to do next, even if in this case, the system risks to bother the user with its requests (when they become too frequent or are asked in inappropriate moments).
These two kinds of systems are two end-points of a line along which we can pose different combination of the two paradigms. As Norman observed [116] it is a problem about the correct balance between letting the system decide autonomously and giving the control to the user. However it is not always easy to decide when the user should take the control. For example, while driving a vehicle in case of a sudden brake the ABS[1] (Auto Blocking System) has to take the control of the brakes and of the throttle, giving the control back to the user when the situation is normalized. In this case the autonomous behaviour is a safety measure as the system could react faster than the driver. But in certain conditions the driver may want to deactivate some safety systems because he/she wants to have a more sportive driving.Moreover systems may not ask continuous confirmations to the user because it would result in a distraction.
The balance between humans and automated systems could be better defined as Level of Automation (LOA). According to Kaber and Endsley[87]:
"Level of automation refers to the level of task planning and performance interaction maintained between a human operator and computer in controlling a complex system".

Even if the context of the research about LOA has been mostly related to the use of "command and control" systemsto execute complex tasks, the findings can be easily applied to AmI scenarios.
Indeed, as seen in previous examples, one of the main problems of automation is to avoid to leave the user out-of-the-loop. It not only can lead to system's decisions that unplease the user, but can also affect the user's performance, situation awareness and mental workload. This could become dangerous when a critical task is executed (e.g. driving, monitoring an environment) as the user may fail to detect and understand problems.
Automation levels are not binary but there could be different levels that refer to different divisionsof the task allocated to humans or to machines. In their seminal work Sheridan and Verplank[134]defined 10 levels of automation:
(1) Human does the whole job up to the point of turning it over to the computer to implement;
(2) Computer helps by determining the options;
(3) Computer helps to determine options and suggests one, which human need not follow;
(4) Computer selects action and human may or may not do it;
(5) Computer selects action and implements it if human approves;
(6) Computer selects action, informs human in plenty of time to stop it;
(7) Computer does whole job and necessarily tells human what it did;
(8) Computer does whole job and tells human what it did only if human explicitly asks;
(9) Computer does whole job and decides what the human should be told;
(10) Computer does the whole job if it decides it should be done and, if so, tells human, if it decides that the human should be told.

[1]An anti-lock braking system (ABS, from German: Antiblockiersystem) is a safety system that allows the wheels on a motor vehicle to continue interacting tractively with the road surface as directed by driver steering inputs while braking, preventing the wheels from locking up (that is, ceasing rotation) and therefore avoiding skidding. (Wikipedia)

This taxonomy, even if general, could be better applied to tasks where a decision needs to be taken and a further implementation should be made. It is possible to distinguish four intrinsic functions related to these different levels[57]:
(1)Monitoring: it is necessary to acquire information about a certain context;
(2) Generating options: in relation with information achieved is necessary to formulate options to achieve goals;
(3) Selecting: deciding for a specific option;
(4) Implementing: carrying out the chosen option.

The different levels differ by the intervention of the user. In certain case the system suggests some options and the users is in charge to decide. In other cases the system selects an options and performs an action, giving to the user the faculty of interrupting the task. In the last level (full automation) the system selects and implements an option without even inform the user.
In this taxonomy there isn't a "manual control" level that, instead, is presented byEndsley[54]:
(1) Manual control—with no assistance from the system;
(2) Decision support—by the operator with input in the form of recommendations provided by the system;
(3) Consensual artificial intelligence (AI)—by the system with the consent of theoperator required to carry out actions;
(4) Monitored AI—bythesystemtobeautomaticallyimplementedunlessvetoedby the operator;
(5) Full automation with no operator interaction.

This taxonomy is related to decision support systems and it clearly highlights some keypoints in the continuum between the full-automation and the manual control, even if there is no perfect symmetry because the user has the right to veto the system's decisions and not vice-versa.
However often the design of automated system could be technology driven, focused in optimizing technical capabilities and performance. Moreover another objective could be to reduce costs through the reduction of human labour and thus human staffing requirements, assigning the task to a system and leaving to the human operator the role of system monitor, resulting in the risk of pushing the user out of the loop. To overcome to the risks of such designed systems a human-centred automation approach can be used.
Sheridan [133] said that human-centred automation points to*"allocate to the human the tasks best suited to the human, allocate to the automation the tasks best suited to it*", through *"achieve the best combination of human and automatic control, where ''best'' is defined by explicit system objectives"*.
The goal of human-centred automation is to create systems that retain the human operator in control loops with meaningful and well-designed tasks that could be well performed optimizing the overall human–machine system functioning. Billings [13]said that human-centred automation should ensure that automation does not leave the human with a fragmented and difficult job. It should define the assignment of tasks between humans and computers in controlling an automated system,considering them as a team[54], [13]. High levels of human–machine system performance may be achieved through human-centred automation by ensuring that the user has the capability to monitor the system, that he/she receives adequate feedback on the state of the system, and that the automation works in predictable ways [13], supporting the achievement of a correct situation awareness.

Kaber and Endsley [87] define two "orthogonal" and complementary approaches to human-centred automation. One approach is the definition of LOA, through a correct assignment of tasks between human and machines. The other approach is the Adaptive Automation that recognizes that the control must pass back and forth between the human and the automation over time.
Adaptive Automation could be defined as varying degrees of computer assistance in complex control systems in relation to the nature of a situation, including task characteristics and the state of the human operator. They proposed structuring human–automation interaction on the basis of 'what' is to be automated, 'when' a task is automated and 'how' it is automated.
The choice between the use of a simple LOA or an AA approach is strictly related to the context and the specific tasks to be performed. Moreover the variables to be considered are relative to the situation as physical and objective factors, but also to subjective state of the user, as mental workload, attention, strain, emotion. Indeed, in AA, many system evaluate the state of the user adapting their behaviour. Indeed, in automotive research, the driver is constantly monitored [42] to react to distraction and drowsiness.
However, if on one hand automated system may help user to stay in-the-loop, on the other hand it may interrupt the user's "flow", disrupting his/her attention. Indeed enriching ordinary objects with technologies that give them new abilities raises the risk to disrupt the human attention during everyday activities. To avoid this, machines have to sense and recognize goals and activities of the humans to invisibly help them. Any unwanted help becomes a distraction.
Mark Weiser [156], in his visionary work, said that:
"*If computers are everywhere they better stay out of the way, and that means designing them so that the people being shared by the computers remain serene and in control"* and *that "when computers are all around, so that we want to compute while doing something else and have more time to be more fully human, we must radically rethink the goals, context and technology of the computer and all the other technology crowding into our lives"* and introduced the concept of "*calm*" computing. Indeed technology is an enemy of calm, interactive systems need our attention, distracting us from the rest of the things happening around us. According to Weiser the difference between an "*enraging*" and an "*encalming*" technology is the way it engage our attention. Calm technology is able to engage both the centre and the periphery of our attention, moving back and forth between the two. Attention has two "places": the centre and the periphery. The centre is the main focus of our attention, we stay in there when we drive or when we talk with someone. Instead we put in the periphery things (and tasks) that are not important or that doesn't require immediate attention. We are almost unaware of the mobile phone in our pocket until it starts to ring and vibrate. In that moment we move it from the periphery to the centre of our attention. This means that while the phone is in stand-by mode it is working for us, being connected to the network ready to receive a call, but we don't need to pay attention to it and we can put other task in the centre. But as the phone rings, we suddenly put it in the centre, taking control of the device. While the centre of our attention is limited, we can focus on few things at a time, the periphery is wide. Using the periphery means to empower the user without challenging his/her attention. AmI systems may enhance user's peripheral reach by bringing more details into the periphery. To be truly useful and unobtrusive the smart environment should work inside our periphery, ready to come in the centre only when needed.

3.3 Natural Interaction

Referring to the way the user is conscious about the technology around him/her and interact with it, we can distinguish between explicit and implicit interaction. As when we want to communicate with someone we can use explicit (i.e. speaking) or implicitways, (i.e. looking repeatedly at the clock communicate to someone else that we are waiting or we want to go), the user can interact explicitly with technology making a direct request (i.e. pushing a button) or, in certain cases, the user can interact implicitly, performing action that are related with the execution of a specific task, and a smart system should be able to recognize this activity and react properly. The simplestexample are the automatic doors at the supermarket: the user's goal is to enter inside, for this reason he/she walk and a sensor recognize his/her action/willand opens the doors. This behaviour can be also defined as "natural interaction".
According to Alessandro Valli [149]:
"People naturally communicate through gestures, expressions, movements. Research work in natural interaction is to invent and create systems that understand these actions and engage people in a dialogue, while allowing them to interact naturally with each other and the environment. People don't need to wear any device or learn any instruction, interaction is intuitive. Natural interfaces follow new paradigms in order to respect human perception. Interaction with such systems is easy and seductive for everyone."
Often the interaction between man and computeris influenced by technological constraints, with the user having to adapt to the interface. Moreover every time a user wants to use a new tool (even a non digital one) he/she has to learn new skills. Thus, even if using the tool will make simpler to achieve a goal, there is an initial overhead of things that have to be learned and a time of practice may be also needed. However if we look at HCI and the history of user interface we can see a path that leads from complexity to simplicity. Computers, especially in AmI paradigms, are now able to adapt to their users. Moreover the user is no more only a brain and a mouse but is considered as a person, with all the complexity related to this concept. The natural interaction paradigm aims atbuilding interfaces that are symbiotic with our everyday actions, sometime anthropomorphic, interpreting our "natural" gestures, expressions and seamlessly interact with us. Persons should be allowed to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do. Unlike traditional WIMP (window, icon, mouse, pointer) interfaces, that has a closed set of commands and behaviours, the concept of what is a natural interface is fuzzy because is not possible to give a definition of what is natural for someone. "Natural" could be intended as what is in the capabilities of a human body (physically speaking) or as what is learned during a whole life and characterized by a specific culture. For this reason methods taken from sociology and ethnography are used to analyse human behaviour, recognize patterns and design natural interfaces.
Using natural interaction paradigms means to reduce the distance between the physical world and computers, "augmenting" our everyday activities and behaviour.
If technology doesn't change our abilities, smart environments are able to change the way we interact with our surrounding environment. And it is happening in the places where we usually spend our lives as the home and the office. Technology becomes (more or less invisibly) a part of our life experience. For this reason the design of the HCI of AmI systems should consider the user not only for his/her primary goals but also as a human, with all his/her needs, characteristics and complexity.

According to Karrayand colleagues [90] the design of a user interaction should take into account three aspects: *physical, cognitive,* and *affective.*
"The physical aspect determines the mechanics of interaction between human and computer while the cognitive aspect deals with ways that users can understand the system and interact with it. The affective aspect is a more recent issue and it tries not only to make the interaction a pleasurable experience for the user but also to affect the user in a way that make user continue to use the machine by changing attitudes and emotions toward the user".
The system should be able to understand the emotions of the user and take them into account during the interaction, to create, itself a good mood. As Norman [114]highlights in his work, emotions are not a surplus but could deeply affect the performance and the quality of the user's experience: *"cognition and affect, from a functional point of view, are deeply intertwined: they are parallel processing systems that require one another for optimal functioning of the organism".*
Moreover, according to the author, when the user is in a good mood his/her mental abilities are more efficient, he/she is more creative and able to evaluate alternatives, and he/she is also more tolerant to system errors. Thereby, on one hand, the system must be able to create a good userexperience, on the other hand it must understand user's emotions and react properly. For example, knowing the user's emotions, the computer can become a more effective tutor in an e-learning application. Synthetic speech with emotions in the voice would sound more pleasing than a monotonous voice. Moreover Normanproposes to program computers to feel emotions (i.e. a computer may be anxious because of hacker attacks) because it could represent a good model to face rapidly changing situations and to communicate with the user.
As in AmI scenarios technology enters in our everyday experience we cannot talk only about Interaction Design but we have to refer to User Experience design. According to Kuniawsky [95]*"The user experience is the totality of end users' perceptions as they interact with a product or service. These perceptions include effectiveness (how good is the result?), efficiency (how fast or cheap is it?), emotional satisfaction (how good does it feel?), and the quality of the relationship with the entity that created the product or service (what expectations does it create for subsequent interactions?)."*
But this definition could be extended to a complex scenario as the interaction with a Smart Environment. In this case the product is the entire environment we live inside and the experience is no more related only to a single goal but is intertwined with our life experience. The more the technology will become invisible the more it will be part of our life, and, as it is becoming more reliable, we will tend to rely more on it and our "natural" gestures, actions and behaviours will be learned from interaction with digital, smart objects.

3.4 New HCI technologies and interfaces in AmI scenarios

Ambient Intelligence paradigms radically change the relation between humans and technology. The user is not interacting with a pc but with an ensemble of devices spread and hidden in the surrounding environment, acting together to create a smart environment.
This changes the classical HCI paradigms used for years with personal computers. According to Butz[19]:

"While in PC-style interfaces, the interaction bandwidth between the human and the computer is relatively low (limited by the use of mouse, keyboard and screen), this bandwidth is much higher for interaction with smart environments. The human user can interact using her or his entire body and multiple senses. Units of information can occupy the same space the user actually lives in."
Unlike the pc paradigm, where there is a fixed set of I/O devices and the used is largely standardized, in the novel field of AmI and smart environments the interaction vocabulary, and the I/O devices to be used, are still in a explorative phase, far from a fixed definition and standardization, still open to an almost infinite range of possibilities.The classical Windows, Icons, Mouse and Pointer (WIMP) paradigm, based only on the use of keyboard, mouse and display is not enough to respond to all the needs that emerge in the interaction with smart environments. Users interact using their whole bodies and senses, in a multimodal ways. Moreover the context acquires a fundamental role in the dialogue between users and technologies.

Indeed the current phase of evolution HCI for ambient intelligence is facing new interesting challenges:

- Multimodal interaction;
- New metaphors that may suite the new scenarios;
- New input/output devices that may catch implicit/explicit user interaction and communicate to all his/her senses;
- New interaction vocabulary and grammar.

Whatever constitutes the smartness or intelligence of AmI andsmart environments has to manifest itself to the human user through the human senses. Interaction with the environment can only take place through phenomena that can be perceived through senses and through physical actions executed by the human. Therefore, the devices which create these phenomena (e.g., light, sound, force, etc...) or sense these actions, are the user's contact point with the underlying smartness or intelligence.
Moreover AmI systems have to be unobtrusive and "natural" as human-to-human communication or as interaction with objects of common use. Indeed, as in a conversation, to have an immersive and unobtrusive interaction, we have to use multiple channels to dialogue with the user.
"A multimodal interface acts as a facilitator of human-computer interaction via two or more modes of input that go beyond the traditional keyboard and mouse. The exact number of supported input modes; their types and the way in which they work together may vary widely from one multimodal system to another. Multimodal interfaces incorporate different combinations of speech, gesture, gaze, facial expressions and other non-conventional modes of input"[119] .
With flexible multimodal interfaces users can take advantage of more than one oftheir natural communication modes during human-computer interaction, selecting thebest mode or combination of modes that suits their situation and task. Moreover these kinds of interfaces augment the bandwidth of communication. In human-to-human dialogue we use words, the main communication channel, but we add other information through non-verbal communication as gestures (especially in certain cultures), body posture, face expressions, voice modulation, and, if needed, we can also start sketching our ideas on a piece of paper. All these secondary information (implicit and explicit) makes us sure to be understood by the other speaker(s). In addition to this we can use these multiple channels as alternatives while adapting to a situation. Referring to the human-to-human communication example, when we are speaking in noisy place we can seamlessly switch from words to gestures (this time used

explicitly) to continue, even if with lower efficiency, the dialogue. An ideal smart environment lets the user interact in different, explicit and implicit, alternative or combined ways, ensuring a more robust and natural interaction.
From a technical point of view the multimodal interaction is enabled by the progresses made in electronics, sensing and processing technologies (already discussed).
The human senses are sight, touch, hearing, smell, and taste. The input modalities of many computer input devices can be considered to correspond to human senses: cameras (sight), haptic sensors (touch), microphones (hearing), olfactory (smell), and even taste. Many other computer input devices activated by humans, however, can be considered to correspond to a combination of human senses, or to none at all: keyboard, mouse, writing tablet, motion input (e.g., the device itself is moved for interaction), galvanic skin response, and other biometric sensors. Here we can insert body movements, expression and gestures that may be symbolic (with the explicit aim to communicate) or non-symbolic.
The wide range of computing devices available (as smartphones), differing for computational power and input/output capabilities, suggest that the future of computing is likely to include novel ways of interaction.
To give a brief overview of the technologies that, nowadays, are characterizing the UIs od AmI system we can start by how humans perceive from them and how they interact.

Displays

One of the most used human senses is sight. Many actions in our everyday lives are connected with sight. Through our sight we can acquire a lot of information at a glance as shapes, colours, depth, position of objects/people in the space, and our brain is able to use them to get more high-level information. Since the first graphical user interface has been created, a lot of interaction with computers have been based on visualization, making displays one of the most used output devices. Moreover in last years there has been a huge increment of interactive touch and multi-touch displays making the input and the output functionsconverge on a single device.
Regarding the interaction paradigm and their employment in AmI scenarios displays can be classified using these factors:

- Dimension;
- Position/orientation;
- Mobility;
- Interactivity-

The dimensions of a display could be related to the way it is used. According to the work of Want and colleagues [153] about the ParcTab we can distinguish between tabs, pads and boards. Tabs are hand sized, portable and personal; pads are larger, as a book , portable, and may be used to share information among persons; boards could be very big, usually not portable, and made for a collective use. Moreover the position and the orientation of a display could affect its use. If a board size display is put in vertical, it could be used as a pure output device and interaction happens from distance. While if the display is in a horizontal position it becomes a table that, is enriched with multitouch or other input technology, becomes more prone to interaction. Of course form factor influences mobility. If a display is small probably it wouldn’t require unpacking (i.e. a smartphone vs. a laptop). Large

displays could be used in AmI scenarios to communicate information in a "calm" and metaphorical way. These kind of displays are called Ambient Displays".
The ambient display goes beyond the traditional notion of typical displays found on PC, mobile phones, kiosks and other interactive devices. They can display information based on context and on user interaction without requiring his/her full attention. The *Hello.Wall* presented in the work of Streitzand colleagues [144] used led clusters to display information to people walking near it. As seen in Figure 6 it has three different communication zones depending on the distance of the user from the display. The authors call it "*distance-dependent semantic*",as the distance from the smart artifact defines the kind of information shown and the interaction offered. The *Hello.Wall* is also capable to stream information to user's mobile device, enabling a public/private level of communication.

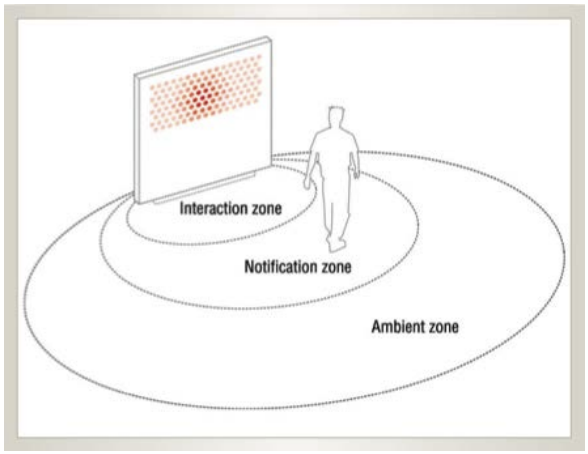

Figure 6 ***Example of Ambient display*** **[144].**

Moreover this kind of interfaces could communicate information using a natural metaphor with a strong aesthetic and emotive impact, influencing also the mood of the social body around it and the atmosphere of the place.

Tangible User Interfaces

As already said, nowadays many displays are made to be interactive. This is due to allow a more natural interaction. In the real world if we want to take a cup of coffee we draw our hand until we reach our objective, with a perfect coincidence between what we are looking at and what we touch while with computer we "take" object on the screen using a mouse. The coincidence of input and output on the screen gives the user the sensation of a more direct (not mediated) interaction. Moreover multi touch screens, that can detect multiple touch points, enable users to use gestures to reach their goal. As an example many multitouch interfaces uses the thumb and the index finger to zoom contents (usually images). Moving apart the two fingers will cause a zoom in, while the opposite movement will cause a zoom out. However someone may argue that gestures on touch screens may not be so natural, and this is true. Gestures may be designed looking at natural behaviours, but they are transposed in a metaphorical way and, sometimes, they could be totally unrelated to reality. As many gestures and behaviours we do everyday, interactive gestures implemented in this kind of systems are

influenced by our culture and must be learned as, in our life, we learn that (in italian culture) shaking hands could means "hello". In an interesting research, made by Wobbrock and colleagues[158]a guessability study was made to determine the most intuitive gestures to be used on a tabletop computer, resulting in a taxonomy useful for the implementation of those systems.
The ability of an interactive table to recognize more than one set of touches at a time enables multiple users to interact or work collaboratively, but in this case the emerging problem is to distinguish among users.
Humans are used to interact with the surrounding environment with their body, through the direct manipulation of objects, with a huge use of hands.For this reason are emerging a lot of Tangible User Interfaces (TUI)[83]but also systems able to recognize human gestures and behaviours. A tangible user interface provides the user a physical object to grasp (Fitzmaurice, 2005). These objects could be used as information placeholders but also as input/output device able to be an interface between humans and complex systems. To better explain the concept we borrow an example from Ishii 63]:
"Among other historical inspirations, we suggested the abacus as a compelling prototypical example. In particular, it is key to note that when viewed from the perspective of human-computer interaction (HCI), the abacus is not an input device. The abacus makes no distinction between "input" and "output." Instead, the abacus beads, rods, and frame serve as manipulable physical representations of numerical values and operations. Simultaneously, these component artifacts also serve as physical controls for directly manipulating their underlying associations"
The main advantage of a TUI is its physicality and the fact that humans have learned to manipulate physical objects during all their life. Manipulating a physical placeholder, handle or tool hence borrows from human life experience. Moreover these objects could provide haptic feedbacks. The term "haptic" comes from ancient Greek word "*happthai*" meaning the sense of touch and, applied to human computer interface, refers to a variety of stimuli that could be perceived. These stimuli are mainly related to the sensation that a contact with a surface could give as geometry, roughness, smoothness, slippage, and temperature, and the force feedback we could have interacting with a physical object as inertia, weight resistance to the force impressed. It is not easy to simulate all these stimuli with electronic actuators. Many devices use haptic feedback to realize multimodal interactions. For example mobile phone uses vibration joint with sound to notify the user of an incoming call. Moreover some touch interfaces (ie. Android gingerbread operating system interface for touch smarthphones) uses vibration when the users touches a sensible area (i.e. a button) to advice him/her that the touch could fire some action. Some game controllers uses the force impressed by the user as an input and could give resistance feedback. Changing the sensation that a surface could give is still a challenge in HCI researches. The MudPad [85] is a multitouch display that could dynamically change the tactile feedback and haptic texture overlay of the touch surfaces thanks to a smart fluid (magnetorheological fluid) that is able to change its viscosity when a magnetic filed is applied.
As in AmI scenarios users will likely interact with objects instead that with computers is important that these objects (smart objects) could perceive and communicate a wide range of information in a "calm" way, without overwhelm user's attention.

Vision techniques: motion, gestures and emotions

As we have said humans interact and express themselves with the whole body. Moreover human gestures or expressions may express emotion, and the same gesture could have different meanings, if

related to different context. To better understand humans and realize a more natural interaction many computer vision techniques could be used.
These have been classified by James and Sebe[84]in:

- Large scale body movements;
- Hand gestures;
- Gaze.

Moreover authors made a distinction between commands (actions that could be used to explicitly execute commands) and non-commandsinterfaces(actions or event used to indirectly tune the system to the user's needs)
Through computer vision techniques is possible to understand the pose and the motion of an individual. This could be used both for commands (i.e. waving a hand in front of the TV means change the channel watched) and for non-commands (i.e. a fall detection inside a smart home could mean that somebody is probably in danger and need assistance). A very visible example is the Microsoft Kinect, the game controller for the console Xbox 360. This device allows users to play games without using any physical controller. The device is able of simultaneously tracking up to six people with a feature extraction of 20 joints per player (maximum 2 players). The device was reverse engineered in few days [65]and is now used, especially after the release of the official SDK, to develop a number of natural user interfaces, applied to different domains.
Computer vision techniques could be used to detect also more fine movements as gaze. As will be discussed further in the present work, gazes could provide a number of information about the user especially in relation with his/her cognitive processes and his/her mental workload (see par 4.3.1).
Detecting and analysing gazes could help to understand where the user is focusing his/her attention. We can imagine a car that automatically detects that the driver is not paying attention to the road, because he/she is looking at the radio, and call his/her attention with a sound or an haptic feedback. Moreover the gaze could be used as a command. We can imagine a large Ambient Display that automatically enlarges some areas when the user looks at them.
People are able to perceive one's emotional state based on their observations about one's face, body, and voice. Research in multimodal systems have been conducted to allow theinferring of one's emotional state, based on various behavioural signals. A bimodal system based on fusing the facial recognition and acoustic information, provided an accurate classification of 89.1 per cent in terms of emotion recognition of 'sadness, anger, happiness, and neutral state' [90].

Spoken Interfaces

Spoken dialogue, which is the most natural mode of communication between humans, is now being applied successfully to many aspects of human– computer interaction[104],[49]. The speech processing, or the study of the speech signal and the methods needed for the analysis and the comprehension of it, makes possible a verbal interaction with the computer (or the Smart Environment). The speech processing can be seen as the intersection of digital signal processing techniques (DSP) and Natural Language Processing techniques(NLP).
The fields of research of this discipline can be divided into:

- Speech synthesis (speech synthesis): is the artificial reproduction of speech through a variety of synthesis algorithms.

- Speech recognition (automatic speech recognition - ASR) analysis of the linguistic content of the signal;
- Speaker recognition (speaker recognition) in which the goal is to recognize the identity of who is talking;
- Speech coding (coding of speech) is responsible for seeking the optimal solutions for the transmission of speech information, such as compression techniques and noise reduction.

ASR together with speech synthesis is the first, basic step toward a natural verbal human computer interaction.
Spoken dialogue systems have traditionally been deployed for automated self-service interactions in which users employ spoken language on the telephone to perform well-defined tasks such as making travel inquiries and reservations. Recently more open-ended applications have been developed, for example, to support senior citizens and people with disabilities in the management of daily activities within smart homes, or to provide drivers and pedestrians with route planning and point-of-interest information.
A number of challenges concern specific aspects of spoken dialogue technology, including methods for dialogue control, representation of dialogue and user states, and the use of learning to enable systems improvement over time.
In traditional dialogue systems dialogue control is often achieved using methods that are appropriate to the application—for example, system-directed for applications involving predetermined services and transactions, and user initiative for more open-ended applications such as voice search and question answering. Mixed- initiative dialogue has been explored mainly in research systems involving collaborative problem solving or human-like conversation.
In AmI environments dialogue control will need to be more open and more diverse. On occasion users will want to query the system for information using unrestricted natural language, but it might be necessary for the system to switch to a more constrained dialogue to elicit particular details. Within a system- directed dialogue users may also wish to query or clarify things with the system, or take over the initiative to introduce a different topic. Or it might be necessary for the system to interrupt the on-going dialogue because there is some important information to be conveyed. More generally, dialogues may be less well structured than the form-filling variety in which the interaction is determined by a sequence of slots to be filled. They may be more opportunistic and distributed, as the user obtains some information from one application and then switches seamlessly to a different application, perhaps to engage in a casual conversation.

3.3 AmI's Metaphors

Every time there is the need to design an interface there is also the need to allow the user to easily understand it and interact with it. For doing it, the designer of the UI should use a language and a structure that could be understandable and familiar to the user. Applying a metaphor, that includes every element in the interface, is the most used method.
Metaphor is a linguistic concept that implies mapping one category of ideas to another. According to the MIT Enciclopedia of Cognitive Sciences [157]metaphor is a "*class inclusion assertion*". Thus, with a metaphor, two objects that could be very different are compared as they belong to the same

class of objects i.e. a lawyer is a shark.A metaphor allows to reason about unfamiliar concepts using more familiar ones, even if details may not match exactly. According to Lackoff and Johnson [97]our reasoning is highly metaphorical. We often use metaphors also to reason about abstract concepts, more difficult to be imagined. When we say that we are "saving time" we are treating the time as a physical thing, that we cold take and stack, and to be used in another moment.

Alfter the first wave of line command interfaces (70s, 80s), thanks to Xerox Park and Apple, a GUI started to be used to access to computer functions and contents. Inside this interface function and contents where represented as common office objects (as paper sheets or folders) in a office desk metaphor.

This metaphor has proven to be very effective if compared with the command line interface and is still used on different operating systems, even if some details doesn't match correctly (i.e. the recycle bin is ON the desk but in the real world nobody wants to keep the garbage bin on his/her desk or keep a stack of glass windows on it).

However, since the office desktop metaphor has been conceived, a lot of new technologies have appeared. Especially with the paradigm of ubiquitous/pervasive computer the user has been surrounded by many unfamiliar devices and technologies and, to understand and interact with them, he/she has had to rely on familiar concepts.

Kuniawsky[95] tried to identify the metaphors that are currently used in ubiquitous computing scenarios. The author operates a useful distinction between metaphors for organization and for interaction. Organizational metaphors are used by designers and users to figure out how different systems, forming an AmI system, relate each others and to people who use them. Interaction metaphors are used to build the interaction between people and technologies.

According to the author, the most used organizational metaphors are:

- Factory: technological systems are seen from their automation point of view. The user is the owner of his/her personal factory that is used for labour-saving purposes (i.e. a smart home able to clean itself).
- Public Service: "*information processing is a utility, like electricity*". In this metaphor information processing is considered to be everywhere. As for us is normal to plug our electrical appliances everywhere it would be normal to find information processing (and connectivity) and plug our smart device to it. This metaphor is used in telecommunication networks. Internet may treat packets as electrons. Some bits may be video or text as some electrons may light a lamp or turn on a motor. For the network is the same.
- Nature: making technology invisible means to come back to nature. Function and relation of smart technologies is inspired to nature, with the aim to free people from technology driven behaviours.
- Vapor (cloud): information and information processing surround us like a cloud. We could access then anywhere and anytime through a number of different devices. As every cloud it doesn't have defined shape and could extends beyond user's reach.
- Parallel Universes: "Technology gives us access to a parallel universe with different laws". Information and information processing are not material for the user; they constitute a parallel universe, a virtual layer that is superimposed to the real one he/she inhabits. This metaphor is used in Augmented Reality application where we superimpose digital images to a live camera feed, having a look at a world where virtual and real coexists.

Interesting interaction metaphors are:

- Terminals Everywhere: everything is an interface. We expect that every object around us could have smartness and be able to receive our inputs and display information. The spreading of multitouch displays almost everywhere is the sign that digital interactivity is now embedded everywhere.
- Invisibility: it is one of the most important concept in ubiquitous computing and, as a consequence, in ambient intelligent. The user expects the computer to be invisible, he/she doesn't have to change habits and behaviours but the computer must be able to understand their need and react properly.
- Animism: one broad definition of animism is the belief that objects have will, intelligence, and memory, and that they interact with and affect our lives in a deliberate, intelligent, and somehow conscious way. Without entering into the problems of the Philosophy of Mind, in a simplistic way, we tend to attribute a mind even to unanimated objects that (seem) to express a behaviour (i.e. when we blame to our computer because it crashed). Smart objects could be designed to express a certain behaviour, sometimes miming the behaviour of humans or animals, as in the case of the smart vacuum cleaner Roomba that simulates the behaviour of an insect [96].
- Prosthetics: smart technologies could be used to extend our bodies and faculties. Digital notebooks could extend our memory while electric motors could extend our strength.
- Enchanted Objects: for the user some technologies are like magic, making things behave like enchanted objects. An automatic door that opens when the user is near it like the magic door in the Aladdin fairy tale.

Using metaphor could be an advantage and a risk. If a metaphor could help users to easily understand how to use a device or a service through analogies with familiar conceptual frameworks. However a metaphor, if used too literally could also bring useless constraints and become inappropriate. Moreover metaphors are always culture and context related thus a single design doesn't fit all the needs.

Of course current metaphors are base on our experience (and the experience of designers) inside and outside the world of computers. In current metaphors we continue to use references to physical objects that are now almost totally replaced by their digital equivalent. We have "phonebooks" on mobile phones even if many young users would have never tried the experience to search for a person (maybe with a very common surname) on a big paper phonebook. Likely new metaphors will have as a reference digital objects and digital "culture" and, even if technology will tend to be invisible, it would be more and more present in our behaviour and culture.

Chapter 4

Improving Human Performance through AmI

4.1 Introduction

Technology cannot improve humans' physical and cognitive abilities but can help them to reach their goals. Indeed technology changes the task that the user has to perform, with simpler and more affordable ones.
In Chapter 3 the concept of Level of Automation has been introduced as the allocation of task to humans and machines in order to achieve a certain goal. As humans have to face complex activities they tried to build automatic system able to do some part of the task, usually to achieve better reliability, efficiency, effectiveness, cost efficiency, allowing the user to concentrate on more high level (and sometimes satisfying) tasks.
For example in many critical scenarios as management of power plants, piloting of aircrafts, and air and naval traffic control, operators are almost always supported by technological systems designed to capture certain information about the environment and detect possible critical situations. These technological progress have meant that today, operating in many domains, one of the main tasks performed by operators is monitoring or supervising these systems.
Indeed as many human activities rely on the knowledge of some information, technology can be used also to provide, through a proper user interface, more detailed information that are useful to have a better perception and understanding of a given situation and (eventually) take decisions.
However the quality of information acquired and the design of the user interface have a strong impact on user performance and on the overall effectiveness of the system.
This is important especially in Decision Support Systems [124] that are "*interactive computer-based system or subsystem intended to help decision makers use communications technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks, and make decisions.*" Moreover it becomes crucial in safety-critical systems, in which a human error may lead to disastrous consequences in terms of safety.

Endsley and colleagues [89] show that there isn't direct correlation between an higher level of automation and a benefit for the user in terms of overall performance and user satisfaction. Indeed, in certain cases, too much automation may bring the user to lose the control of the situation (being out-of-the-loop) .
AmI is going to introduce a lot of automation in many domains and in many fields of our lives and changing not only the way we interact with technology but our relationship with our environment. The successfulness of this relationship is related mostly on the way the system composed by human and "smart technologies" will be designed, involving problems that are related with technology, human factors, usability and user experience.
As every domain interested by AmI and automation is very specific and it is not possible to generalize, we decided to focus on the specific field of Smart Surveillance, applying the paradigms of AmI, aiming to find the right balance between human and automation in an Ambient Intelligence scenario.
Indeed technology must supports human activities and improves the overall quality of live safety is, of course, an important matter. The concept of safety can be declined in many ways, depending on the context. One of the main aims of safety is to preserve human life. It could happen by preventing hazardous conditions (i.e. in an industrial plant an automated security system avoid workers to enter in a room where is any danger) or it could react after some accident has occurred (i.e. a smart home detecting the fall of an old man and some anomaly in his vital parameters could call an ambulance) or raising an alert for a human operator that has to take a proper decision. One of the characteristics of AmI systems is to have an understanding of the environment and the ability to take "smart" decision. This capability could be used, for instance in Smart Surveillance systems, to preserve the safety of people in an automatic way or simplifying the work of a human operator.
In the following chapters we analyse the concept of Smart Surveillance with the related technologies introducing the concept of situation awareness and the issues related to the specific field investigated. Then we will propose a Smart Surveillance system using Ami paradigms and technologies able, through a proper human computer interface, to improve the performances of a human operator.

4.2 Smart surveillance

There are many technologies and strategies involved to solve safety issues, however one of the most used is video surveillance. There are a lot of cameras installed everywhere, in public and private space. However as Haering and colleagues [72] point out, monitoring is expensive and ineffective. Cameras alone produce a constant flow of data, but it is necessary the intervention of a human operator to understand the content of the video, recognize dangerous situations and react properly. For human operators the problem is related to the Situational Awareness (see paragraph 4.2) intended as the ability to have the constant control of a situation and react properly if some event occurs. In their study the authors found that after 20 minutes of monitoring, the human attention lowers to an ineffective level. Moreover Regazzoni and colleagues [127] confirm that, due to the difficulty of a real-time analysis, the majority of video feeds are recorded and used for post-event analysis.
In a work of Wallace and colleagues [151] about the ergonomics of CCTV systems, is pointed out how each operative can only really monitor 1-4 screens at time while, according to Dee and Velastin[36] the actual rate is higher, up to 6 screens. To avoid this unefficiency the monitoring process could be automated using systems that are able understand and react to events. This evolution

of video surveillance is called Smart Surveillance to underline the "smartness" introduced by technology. As the main source of information in a video surveillance system are images, the most used technique are related to computer vision. Tian and colleagues [147] define Smart Video Surveillance as:
"the use of computer vision and pattern recognition technologies to analyse information from situated sensors".
Thanks to these techniques smart surveillance systems are able to work without human intervention, alerting the operative only when something happens (or according to the different degrees of automation seen in 3.2). For example, a system could monitor the drop-off area of an airport parking, detecting if a car stops for an excessive amount of time (maybe also identifying it by the plate number). Computer vision research and development have advanced the state-of-the-art in video related algorithms in conjunction with the increasing processing power available. However these systems are not yet fully reliable and robust. Indeed one central issue here is the problem of robustness. Robustness is defined by the IEEE [147] as *"the degree to which a system or component can function correctly in the presence of invalid inputs or stressful environment conditions"*, and this property is clearly vital for automated visual surveillance. Within computer vision, the term robustness is often used in the related, statistical sense: robustness in statistics is the ability for a test to handle data, which deviate from its assumptions (e.g. the ability for a Gaussian-based model to handle non-Gaussian noise). In real world conditions changes in illumination, occlusions and other phenomena could affect the effectiveness and robustness of the system, so the intervention of a human operator in some part of the loop is necessary. For this reason is important to evaluate the human factor aspect of the system, as will be further discussed in the present chapter.
Video analysis and video surveillance are active area of research both in Academia than in industry. The key areas of interest are

- detection and tracking of single or multiple person or objects;
- person identification;
- object identification;
- large-scale surveillance systems.

According to the work of Hampapur and colleagues [74] there are several technical challenges that need to be faced for the development of smart surveillance systems. The authors highlight three main interesting challenges:

- the multiscale challenge: it refers to the ability to acquire information at multiple scales, to ensure an effective situation awareness for the human operator. It means extracting extra data from the details in the observed scene. For instance, into a hotel's lobby is interesting to observe what people are doing but also the expression on their faces. This would help to understand their behaviour and to predict their actions. The problem can be seen in two different perspectives. On one hand there is a need of technologies able to get information on details (i.e. high resolution images of faces) and, on the other hand, there is the need to relate all the information acquired. Moreover to get other detailed information about the observed scene, the video surveillance system can be integrated with other kind of sensors (inertial, temperature, orientation). This topic will be further discussed in 4.2.1 .
- Contextual event detection challenge: to actively support an human operator, the smart surveillance system must be able to interpret data to detect events of interest and identify

trends. A lot of technologies and techniques are involved in this task that is closely dependent by the information available about the context.

- Large system deployment challenge: this means using the single findings to build an efficient system to monitor large areas. This means minimizing costs, having auto calibrating hardware and a proper management of the huge amount of data gained.

The majority of computer vision systems for surveillance are organised with low-level image processing techniques feeding into tracking algorithms that in turn feed into higher level scene analysis and/or behaviour analysis modules.

The lowest level is constituted usually by motion detection and background subtraction systems.

Dee and Velastin[36]address several computer vision systems and algorithm applicable to the problem of surveillance, mainly:

- Tracking and Occlusion reasoning: are algorithms used to identify foreground pixels over time as belonging to a particular moving or occasionally stationary object using, usually, Kalman filter or particle filter. However one of the well-known problems of computer vision systems is the handling of occlusion. We have an occlusion when an something falls between the camera and the objective to be tracked. This could be a fixed obstacle, hat may be mapped inside a scene modelling, or a moving obstacle (object or person).
- Scene Modelling: a correct modelling of the observed scene could help to improve surveillance applications. It is important to define aspects that may also influence the behaviour of the targets to be tracked. For example is important to know the entrance and the exit of a room to for a correct tracker initialization. Also the knowledge of fixed obstacles and the paths followed by agents in certain scene could help to disambiguate difficult cases.
- Behaviour analysis and detection of unusual behaviour: in the video surveillance domain behaviour is a word used in most general sense as the observable actions of agents. The system tracks the behaviour of agents and models it. The mathematical representation of the behaviour could be classified through statistical methods. Usually Hidden Markov Models and Bayesian Networks are the approach used in the literature for the classification of behaviours.
- Detection of specific alarms: based on the modelling of the scene and of the behaviour of the agents, the occurrence of a pattern of data that statistically does not conform to the norm can give rise to the issuing of an alarm. In operational conditions such an alarm could be presented to an human operator that is in charge to react (i.e. alerting an emergency team) or giving a feedback to the system to correct a wrong detection (false-positive) or, as in the work of Zhong and colleagues [160] could re-label an event that was not perceived by the system as unusual (false-positive).

4.2.1 Multi-Sensor and Multi-Modal video surveillance

Since their first generation video surveillance systems have been based on multiple sensors. With the evolution from analogue to digital IP-camera has become easier to build networks of video sensors to extend surveillance coverage over wide areas. However multiple sensors could be used also to monitor the same area, thus allowing redundant or possibly improved data that can be exploited to improve

detection and robustness, enlarging monitoring coverage, reducing uncertainty. The sensors used could be homogenous (i.e. cameras) or heterogeneous (e.g. temperature sensors, depth s., pressure s.).
In literature and in real world installation there are a number of works using multi-camera systems for smart surveillance. They can use different kind of cameras, usually fixed and PTZ (pan, tilt and zoom). A primary system made of fixed cameras could be used to automatically have a first understanding of the scene, eventually driving a secondary system made of PTZ cameras point on region of interest to get more detailed data to realize multi-scale systems.
In these systems is also possible to use algorithms based on the combined use of the information gained from all the cameras, as cross-camera tracking of moving objects/persons. A centralized engine could fuse the information gained to combine the multiple representations of objects, which are in the overlapping camera field of view, and building object trajectories even over non-overlapping field of views.
Heterogeneous sensors could be used to build multi-sensor and multi-modal surveillance systems. The advantages in technologies, especially regarding sensor networks have made simpler to use multiple sources of information to model a scene. However, according to Snidaro and colleagues[139], this topic still has to be deeply investigated in literature. As highlighted by Pratiand colleagues[123], despite the efforts made by the researchers in developing a robust multi-camera vision system, computer vision algorithms have proven their limits to work in complex and cluttered environments. These limits are mainly due to two classes of problems. The first is that "*non-visible areas cannot be processed by the system*". This trivial statement is of particular importance in cluttered scenes and can be partially lessened by using multiple sensors (not only cameras). The second class of problems, instead, is due to the limited resolution of cameras. Having infinite resolution and zooming capabilities would make the job easier, but in addition to be unfeasible, it would exponentially increase the computational load and it is typically too expensive.
In particular using different sensor with data fusion techniques that may solve classical problems of video surveillance systems like detection, localization and tracking or person identification.
In the work of Prati and colleagues[123] a multi- modal sensor network that integrates a wireless network of PIR-based sensors with a traditional vision system is developed to provide a more robust and accurate tracking of people. The problem of using sensor fusion in smart survelliance system is analyzed in detail in Chapter 5, with a proposal of an hybrid people tracking system.

4.3 Situation Awareness

About forty years ago, some military researchers began to define the concept of "Situational Awareness" in order to understand the behaviour of pilots in critical situations such as those of air battles.However, a formal definition of the concept named "Situation Awareness (SA) was provided only in the 1980s by Endsley [57]. According to the definition most cited in the cognitive ergonomics literature, the term "situation awareness" indicates the ability of the operator to: (1) perceive the elements present in the environment in a given unit of space and time, (2) understand their meaning, and (3) make projections on their status in the near future [56],[57]. Basically, when individuals operating in complex and changing systems, situational awareness is essential to detect changes in the system and respond to unexpected events.However, as the definition states, this concept includes and

defines human cognitive abilities (i.e., perception, comprehension and prediction) that everyone uses constantly in their daily activities.

The most diffused definition of the concept of "situational awareness" is provided by Endsley[56], who, in cognitive ergonomics literature, has assumed the role of the reference model. Endsley[56] has divided the model of SA in two parts.The first (named situation awareness) shows those processes (perception, comprehension and projection) directly responsible for the acquisition of SA.The second part (more complex) includes various factors that may indirectly influence the SA.These were grouped into four classes:the outside world, the task factors and of the system, individual factors and a series of specific domain-dependent factors (e.g., skills, experience and training).

The three levels of SA, which represent the major components of this model have been defined as follows:

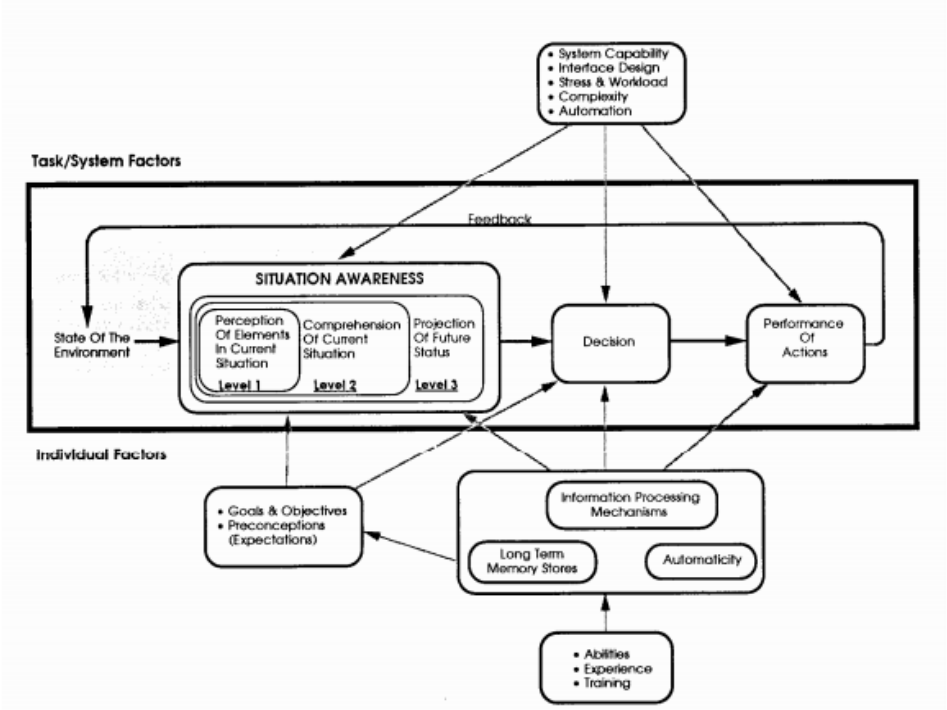

Figure 7:Three levels Situation Awareness model.

LEVEL 1 – *Perception of the elements in the environment.* This represents the first stage in order to reach a certain degree of situational awareness.The individual must acquire the relevant information on the elements in the situation.This information interacts with the knowledge contained in long-term memory and the individual is then able to classify them and categorize them in a comprehensive mental representation of the situation.In other words, requests (or the objectives) of the job guide a process of information selection that allows the individual to gain knowledge about the most relevant elements of the situation.

LEVEL 2 – *Comprehension of the current situation.* This level is a synthesis of all those elements that at the first level remain separate amongst themselves.Previous knowledge in the form of cognitive patterns and mental models, allows the operator to integrate new knowledge (acquired at level 1) in a mental representation of the situation, which is updated and

organised according to the requests of the task that the operator is carrying out at that specific moment.

LEVEL 3 – *Projection of future status.*The third level refers to the ability of projection (or prediction).At this level, the operator must be able to project, in the near future, all elements and events that has he/she has perceived (level 1) and of which he/she has understood the meaning in relation to the current situation (level 2).In other words, the individual is required to possess a mental representation of how the current situation will be in the near future.

Therefore according to this model, when the individual is carrying out a task it is essential that he/she anticipates future consequences of what is happening at that moment.

The main feature of this model is to consider the perception, comprehension and projection as three skills among them strictly linear and hierarchical.As a result, any error at level 1 or at level 2 would lead the individual to inevitably commit errors also at a predictive level.Another peculiarity is that while theoretically Endsley defined SA only as a cognitive state, within his model, conditions and cognitive processes are interconnected.Subsequent works of the same author have not added any further explanation or clarification to this regard.

Basically, one of the major benefits that the concept of Situation Awareness has introduced in research into Ergonomics is to try to accurately define the interaction and the essential inseparability between the individual (awareness) and the operating environment surrounding him/her (situation).However, the more traditional research on this concept appears to have increasingly focused on component "awareness", omitting a systematic description of situations where individuals act.For example, in the context of air traffic control, numerous studies have thoroughly investigated the link between the level of experience of the controller, his/her mental models, memory, the workload, the attention and the acquisition of situation awareness.In contrast, the characteristics of scenarios whereby flight controllers operated and those of the interfaces used were only generically described, without providing relevant information on the interaction between the objective features of the situation (for example, objects, events and communication) and the achievement of a certain degree of situational awareness.

For example, if an operator has the task of monitoring a series of displays for long periods of time, the physical characteristics of the interfaces used should favour the operator to make the best use of all his/her perceptual functions, maintaining attention and information processing so that he/she can perform the necessary actions in order to reach his/her objective (i.e., maintain a high and constant level of performance over time).

A different definition of Situation Awareness was provided by Sarter and Woods[130]. They defined the SA as "the accessibility of a comprehensive and coherent situation representation which is continually updated through the results of evaluations of the situation". This definition focuses primarily on the continuous interaction between the dynamics of the operating environment (taking into account also the assessment of the situation) and the mental representation that the operator has of the situation.In this case, therefore, the operator, based on his/her past experience, already possesses a mental representation of the scenario and integrates it with the information (in other words with the evaluations) regarding what is happening in the present situation.This continuous updating of mental representation must be understandable and consistent in function of the tasks that the operator must

perform.Of course, the operator must also have sufficient cognitive resources to be able to quickly access quickly this representation.

A simple example of this articulated definition of "situational awareness" can be represented by the daily use of computers.All of us, according to our experience as users, have a certain mental representation of an operating system (GUI Windows, OS X or Linux).However, each time we use our computer we must update our representation of the system based on what the situation is really running at that particular time.For example, assuming that we clicked an icon of a text file, the start of the writing program inevitably updates the “static” mental image that we previously had of the operating system.This update should also keep a mental representation of the system that is understandable and completely consistent with what we expected and with what we have to start doing (e.g. writing a report). Inmost ordinary situations, for example in operating systems that do not give errors and in situations where we are not carrying out too many tasks at once (for example, talking on the mobile phone or leafing through a book while the system is starting the program), our representation of the "updated" system is also easily accessible and our awareness of what is happening in the system at that moment is high.

The definitions of SA and the example described above, show how factors such as (1) perception, (2) interpretation and understanding, (3) projection, anticipation and expectations, (4) updating and evaluation are key elements that enable individuals to play all those activities that require the resolution of problems through a good situational awareness.

As previously introduced, despite the concept of SA describes an important "phenomenon" that can affect the performance of operators, this was still not fully developed and a unique and universally accepted definition in literature has not yet been provided.One of the main limitations of this concept lies in its dual interpretation, as a condition or as a cognitive process. For example, the definition of Situation Awareness by Endsley [56],[57] explains the SA as a cognitive condition (i.e., a state of knowledge) that is separate from the cognitive processes underlying its attainment.For more clarity, a "*cognitive state*" is defined as the result of a process at a given instant of time.Differently, a "*cognitive process*" is one continuous progression of different and composite cognitive activities.Within the SA concept the distinction between state and process should be finalized in order to avoid confusion in the understanding of how an individual reaches a certain degree of situation awareness. Endsley [52] justifies the importance of this distinction through the following grounds:"different individuals can use different processes (e.g. different methods for capturing information) reaching the same state of knowledge, or while using the same cognitive processes, can reach different states of knowledge due to different insights, predictions, acquired information or different mental models."

Elaborating on this distinction, Endsley includes the processes tied to the SA in the concept of "*situation assessment*". This concept represents the mental activity of perception, memory, attention and categorization that should underpin the production of a certain degree of awareness.

The models of SA, now present within the scientific literature of Ergonomics, are numerous, however, none of these seem to have thoroughly clarified the role of objective features of the operating environment, as well as cognitive processes, the acquisition and maintenance of situational awareness by the operator.

This issue does not, at present, provide a clear interpretation of the results obtained when measuring the degree of situational awareness that an operator has during the course of a task.To this day, in fact, many measurements of SA could easily lead to ambiguous interpretations, tending towards circularity,

for example:"*How can you know if the operator lost situational awareness?Why did the operator respond inappropriately?Because situational awareness was lost*"[59].
As is easy to understand, these types of interpretations arise from lack of theoretical and methodological foundations that should allow to make inferences for example on a deterioration of some cognitive abilities or worsening of certain operating-environmental conditions that can cause a loss of situational awareness.

4.3.1 Measuring the Situation Awareness

Each problem regarding the definition and creation of models of Situation Awareness is particularly critical for the development of valid and reliable techniques for its measurement.When we speak about the need for a "systematic definition" of the situation we mean that every element, event and agent that constitute a certain operating scenario should be described in detail starting from their most significant features.This description should then be placed in relation to skills and to cognitive and behavioural limitations of the individual when he/she must reach a goal.This type of procedure should be implemented first to measure any increase or decrease of SA [137]. Otherwise, the interpretation of the measurements obtained would be too subjective and, as often is criticized in the literature, any operator error may be attributed to a general loss of situational awareness (see also [59]).
Endsley [52] has attempted to divide existing measurement techniques to estimate the SA, in different categories.The main methodologies for measuring situation awareness can be grouped into the following categories:

- SA requirements analysis;
- Measures based on the "freeze" technique;
- Measures of SA in "real time";
- Self-assessment techniques;
- Observation techniques;
- Performance measures;
- Indexes of SA cognitive processes.

Below will be provided a brief description of each of these categories and major measurement techniques that later will be used in this work.

Analysis of SA requirements

The analysis of the requirements to obtain a degree of situational awareness is the first step necessary to determine what really affects the operator's SA in the task or in the specific environment in which he/she must act. Endsley [55] has described this procedure of analysis which includes (1) an unstructured interview to an expert operator, (2) the analysis of the task that needs to be carried out and (3) the administration of one or more questionnaires to determine how each requirement identified may be relevant to have an adequate situational awareness.In other words, through this procedure, you will find all the knowledge and skills that an operator must already possess (or acquire during the execution of the task) in order to have a suitable situational awareness.

The results of this analysis should then be used during the development of an SA measurement technique, which should take into account relevant elements in that specific situation in order to perform that particular task.

Measures based on the “freeze” technique

The “freeze” technique is, to this day, the most widely used technique to measure the individual's awareness of the situation.This technique provides an interruption of the task and the administration of a screenshot of the scenario.During the *freeze* one or more elements of the scenario are hidden (or obscured).Subsequently, the individual is prompted to answer a set of questions on the elements of the scenario that were blacked out. Empirical evidence [55]demonstrated that the *freeze* technique does not show intrusive effects neither on the execution of the task nor in maintaining situational awareness. The "situation awareness global assessment technique” (SAGAT) is the main SA measurement based on the "freeze" technique.This was developed by Endsley [55] to estimate the degree of awareness of the situation of military pilots involved in air missions.Subsequently, SAGAT was also adapted to air traffic control.During experimental simulation of air traffic control, SAGAT provided the "freeze-frame" of the radars display and communications between controllers and pilots.The image of the aircraft on the radar was replaced by a simple point on the monitor and the controller was asked to:bring back the flight path of the aircraft, its speed, its altitude, the instructions that were given to the pilot and those that must be given in the future.
Hauss and Eyferth [78] developed the "SALSA" measuring technique to specifically measure situational awareness of air traffic controllers.The SALSA implements the typical *freeze* technique to test situational awareness of the operators (see example in Figure 4).The main difference of SALSA compared to SAGAT concerns the allocation of a main value to each element that during the *freeze* is hidden. In other words, while in the SAGAT the elements of the scenario are all considered with the same functional value, in the SALSA a score based on its importance for the proper performance of the task is assigned to each element.Consequently, if the operator commits several errors in reporting information on the most important elements of the task, the calculation of his/her degree of situational awareness is particularly low.

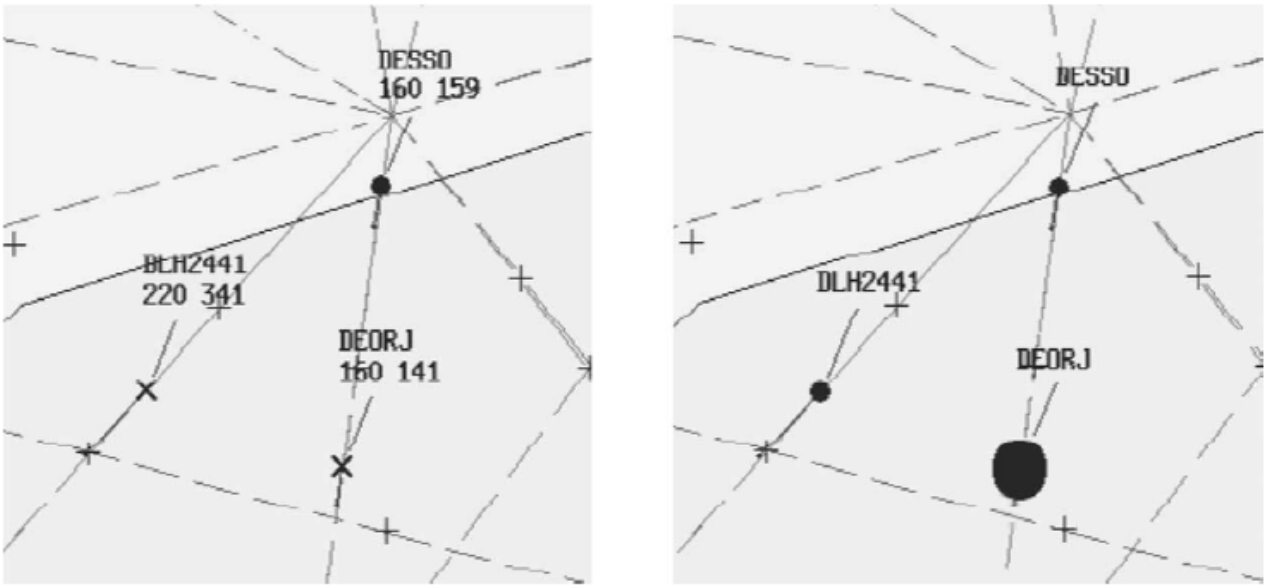

Figure 8 ***Example of the freeze technique in SALSA.*****The left image shows a screenshot of the screen-radar during the interruption of the task.The right image shows the same area of the screen-radar adapted for administration of the SA questionnaire. Only one aircraft (DEORJ) is highlighted to indicate that the subject must meet a series of questions on the characteristics of this flight (for example, direction, speed and altitude).**

Measures of SA in "real time"

When the simulation used does not allow the use of the *freeze* technique, for example because the scenario of the task cannot be interrupted, the SA measures in "real time" represent a valid alternative.Usually, these are based equally on a set of questions that are carefully processed during the first phase of analysis of SA requirements.The elements of the operational situation that are deemed more critical to the execution of the task are selected and questions about their characteristics are developed.These are then usually administered when executing the task without stopping it completely.The "situation present assessment method" (SPAM) was developed by Durso and colleagues[48]to study the awareness of the situation of air traffic controllers.While the controller is carrying out his task, he is required to report certain information "by phone" on the aircraft and on the activities (e.g., aircrafts that are entering or exiting from airspace jurisdiction) that are particularly important.Practically, during the execution of the task, the controller receives a call that he/she can accept as soon as he/she is available.From the moment that the controller accepts the call, he/she responds to the question of SA using the time he/she needs.Unlike the other measures mentioned above, SPAM takes into account only correct answers and the degree of "situational awareness" is taken from the time that the operator used to answer the questions (correctly).In other words, with this method, the measure of "situational awareness" is based on the time that the operator uses to retrieve (or remember) the data needed to respond appropriately to the questions.Finally, the time that the operator employs to accept the call is used as a workload indicator.For example, if the task requires the operator to use a high level of cognitive resources and the execution of numerous actions he/she

must essentially postpone the acceptance of the incoming call and the task will be evaluated as more demanding.

Self-assessment techniques

Self-assessment measures have several advantages for measuring how much an individual is (or rather, was) aware of the situation during the execution of a task.These techniques usually consist in questionnaires that are administered at the end of the task.Such measurement techniques are simple to use (fast and low-cost) and do not interfere with the execution of the task.However, these types of measurements are based on subjective evaluations of the individual who has just completed a task.The subjective nature of these techniques represents their greater criticality.In fact, the individual, after carrying out the task, must remember certain events whereby he/she is asked to give a personal assessment (for example, how much they felt able to control the operational situation).It appears evident that these types of measures can be easily influenced by what is remembered (or forgotten) and the subjectivity of the answers.Consequently, these self-assessment measures are often criticized for their low sensitivity in estimating those variables that can actually be related to high or low degrees of situational awareness.

Performance measures

Performance measures provide an indirect measure of SA.For example, in military battle drills the performance can be measured in the number of enemies hit, the number of ammunitions used and success or failure of the mission.In an assessment of "situational awareness" of the driver, Gugerty [70] measured the identification of hazards and the obstacles on the road and the correct avoidance of possible accidents during simulated driving.Although performance measures are easy to obtain and are not intrusive because the data is collected during the execution of the task, they have many critical issues that affect the general relationship between "situation awareness" and performance.For example, a very experienced driver may be able to maintain an acceptable level of performance driving even in times when his/her awareness of what is happening in the street is not adequate.In contrast, an inexperienced driver can have a high awareness of everything that is happening in the street, but for other aspects tied to little driving experience, may not be able to achieve high levels of performance.

Measuring indexes of SA cognitive processes

The indexes of cognitive processes involved in the acquisition and in maintaining situational awareness are achieved through recordings (for example, of eye movements) carried out during the execution of a task.Therefore, the researcher, while the individual is carrying out the task, must sample those variables (e.g., behavioural and/or psycho-physiological) that reflect in a more sensible and reliable way a few variations of those cognitive processes that are involved in the skills (of perception, comprehension and projection) that define the concept of SA.
The measurement technique most widely used to obtain information about the cognitive processes of SA is the recording of eye movements (eye tracking).An eye-tracking device can be used to register

the location (on a screen or on an interface) of ocular fixations and their duration.These data can be used in order to obtain information on how an operator allocates his/her attention during the course of a given task and how he/she visually explores a specific interface.The eye movements are typically used in measurement of SA to avoid all the problems tied to the interruption of the task that is necessary instead for the administration of questionnaires with the freeze technique.Therefore deriving the measures of SA through the recording of eye movements has the main objective of obtaining measurements with a high ecological validity. In orderto apply this type of measurement SA requirements are needed in the analysis, locating all the elements in the experimental scenario that must be explored and/or monitored to capture the necessary information in order to obtain a good situational awareness during execution of the task. Hauland [77] investigated the different strategies of visual attention to measure perceptual aspects of SA.In particular, it was suggested that a more concentrated visual attention (i.e. eye fixations close together) would indicate that the acquisition of individual information was focused only on a few elements of the scenario. Otherwise, a more distributed visual attention (i.e. eye fixations more distant from one another) should reflect a strategy aimed to acquire information on many items in the scenario.Taking into consideration the times of visual scanning instead, a more focused strategy had to be associated with prolonged visual exploration (of at least one second) on each item noted, while, when the scan was distributed, visual elements were to be focused on for a short time (i.e. for less than a second).A priori, all the more important elements of flight scenarios that should have been monitored by the controller were defined as items of interest in order to acquire a good situational awareness.The results of the study conducted by Haulandshowed that the distributed attention strategies used by the flight planning controllers correlated positively with the performance levels.Similarly, a prolonged observation of the items on the radar display (i.e. a more strategy of visual attention) positively correlated with performance obtained by radar controllers.From these results, Hauland concluded that the frequent use of a distributed exploration strategy, in the control of flight planning, could indicate an appropriate use of perceptual processes of SA.

Despite several attempts to measure the awareness of the situation by analysing visual exploration strategies of individuals, in the more traditional research on the concept of Situation Awareness the eye-tracking is still a technique that is not used very often. This fact is often criticized because of its sensitivity to environmental conditions (e.g. brightness) and to the difficult relationship between attention and eye movements (see also Salmon and coll. [129]).In fact, through eye-tracking it is possible to know which element of a visual scenario does an individual stare at, but not how many and which characteristics he/she acquired and processed.Otherwise, the individual may have noticed an object that is present in the visual scenario on the edges of his/her visual field without having it stared at it.These limits, marginal as they may be (or exceeded) in some very controlled experimental conditions and specially designed, are instead very critical in the use of highly realistic simulators or in experimentations in ecological conditions.

4.3.2 Operator attention in Video Surveillance Task

In a video surveillance task the operator usually have to monitor some screen looking for events or elements that are considered dangerous. Usually the activity consists in tracking multiple independent entities (persons or objects). In the work of Vural[150] is highlighted how the ahuman ability to track multiple object is related to the number of objects, their speeds, and spatial configuration. Moreover it

could be related to subjectives operator differences, attention levels, strategies adopted,and workload. As the human attention is limited and not constant, and can be affected by physical and contextual conditions, operators may have problems and limits in performing these tasks. In many studies in the field of Psychology it is shown that a human can track only a small number of independently moving objects at the same time, particularly some studies found that individuals can track4 to 6 dependent objects by dividing their attention [26]. On the other hand, humans can only track one object with sustained visual attention.

According to the work of Vural[150], multiple moving objects with similar trajectories form a group called virtual object. Grouping may help the operator in track a larger number of objects [125]. Moreover also the speed may affect the ability to track objects. Lower speed allows a better precision, with a lower mental and temporal demand. Tracking is also hard for independently moving objects that collide and occlude each other. Attention levels and periods are subjective and depend on expertise, emotional situation, and workload. However, during surveillance tasks, there could be long period of inactivity, resulting in an underload and boredom for the operator. On the other hand, attention levels of overloaded operators will drop in a short time.

Moreover operator's performance could be affected by a poor design of the human-computer interface. Indeed the main research focus in this area has been on the computer vision side and not on usability, reliability and efficiency of the interface.

From a human factors and Human Computer Interaction (HCI) perspective, it is important to identify whether CCTV systems support human operators effectively, and are fit for the purpose.

Investigation into the design and operation of CCTV control centres has revealed significant problems with how the systems are set-up, managed and used [141]. Whilst the implementation of CCTV can be extremely expensive, this does not guarantee its effectiveness and all CCTV systems rely to some extent on the competence of the human operator. As a consequence, as with all technology, there is always a risk that too much attention is paid to perfecting the technical solution rather than studying how humans will interact with it [140]. For example, cameras poorly positioned and located inappropriately produce low quality images in the control room, which in turn lead to difficulties for operators in trying to identify an object or recognise a person [91]. Similarly, when operators are typically expected to view a high number of monitors simultaneously, vigilance deteriorates as a function of the number of screens being attended to [43].

These issues identify fundamental aspects of HMI where the user is overwhelmed by information or processes to the point where they cannot perform their usual tasks effectively [80]. It also highlights issues of trust and transparency of automated systems and aspects of team-working between users and technology. The CCTV operator and surveillance system interact, forming a working team, and just as a conventional team of humans operate, modern automated systems are characterised by trust in the system, functionality of team members, communication within the team, and where authority should be invested in the team [146]. It is crucial that the operator remains 'in the control loop' and aware of the overall situation at all times [20]. To achieve this, a certain degree of transparency must exist, which relates to the user's ability to understand what the automated processes are doing and 'see through' the system [117].Thus, the lower the transparency, the more removed the user is from the information processing, which might have serious implications for their overall awareness of a situation. Situation awareness can be defined as the user's knowledge of both the internal and external states of the system, as well as the environment in which it is operating [51].

In their work Keval and Sasse [91] reviewed a number of previous studies and performed an on field qualitative test to discover the main tasks and the main HCI issues related to video surveillance.
The main task identified by the authors are:

(1) Proactive surveillance requires operators to “spot” suspicious behaviour and individuals by scanning activity across several cameras using the monitor wall or by inspecting camera by camera on their inspection (spot) monitor(s).
(2) Reactive surveillance requires operators to react to audio or visual cues about incidents. Usually operators must follow some procedure and perform actions like a deeper inspection of the event, focusing on the right camera(s) and then dispatching the alarm through some communication device.
(3) CCTV video review and tape administration tasks. A post-hoc activity made to analyse recorded videos, hand-label images, make logs etc..

The main usability problems revealed (excluding the specific one related to the scenario investigated) were:

- high camera to operator ratio, with a difficulty by the operatives to monitor all the screens maintaining a good level of situation awareness;
- difficulty in searching and locating scenes as many systems observed doesn’t have a correct mapping between cameras and their position in the environment. In this way for the operator is hard to get oriented in the space he is observing.

Chapter 5

The proposed system

5.1 Introduction

As described previously the “intelligence” of an AmI solution is both in the ability of the system to understand the context and support human activity also through a good human-computer (or human ambient) interface. To demonstrate this hypothesis we developed a multisensor context capture system and an interface to be used in a smart video surveillance scenario.
The main purpose of the system is to use the data gained from a context capture system to improve situation awareness of a human operator performing surveillance activity using an human computer interface that can actively support the human operator. The context capture system is based mainly on sensor fusion techniques used to develop an hybrid person tracking system based on the combined use of computer video techniques and RFID UWB localization with the aim of outperform the two subsystems.
To have a better understanding of the requirements of a real system we had informal interview with an operator of a nuclear power plant in France and with some officers of the local police of Rome, in charge to control surveillance systems on freeways and in public transport stations. In the first case the interesting factor emerged is that often, in designing security systems, legacy technologies are preferred as already well known for their reliability. Moreover, when some accident occurs automatic systems are in charge to apply general safety procedures, however they rely mainly on CCTV to identify if there are people to rescue.
From the interviews with police officers instead we found that they have problems following people with multi camera systems and also, in case of emergency, identifying officers belonging to other corps (firemen, police) operating in the observed area.
The proposed system is intended as a proof of concept so it is not complete but is meant to highlight some core functionalities and characteristicsthat address some well known problem of video surveillance systems. Indeed we decide to focus on several aspect that highlights the potential of the context capture system, enable by a smart environment, andof the user interface able to use these data without increasing the complexity but improving the performance of the user.
These functions are:

- Identification of people;
- Tracking people over wide areas/multicamera tracking;
- Dynamic application of rules;
- Support users in case of emergencies;
- Integration with other AmI systems.

We will give a brief description of these problems pointing out the issues regarding technological implementation, usability and the impact on situation awareness.

Identification of People: there are many scenarios where there is the need to identify people. Most of them are related with security, i.e. checking if someone has the rights to access certain areas, but this data could be used also to provide services tailored on user's profile, especially inside smart environments, that are supposed to react in an intelligent way to one's needs. Determine the identity of someone with a reasonable degree of certainty could be a challenging problem both for humans than for technological systems. In fact someone's identity could be based on a number of characteristics. Some of them are embodied others are related to some object (data carrier) and could be automatically detected by sensors. Nowadays many video-surveillance systems rely on the use of images for the remote recognition of people. However in certain context it could be very difficult, as where there are crowds of people or when their faces are covered with dresses or protection garment or when images have a poor quality. Moreover some characteristics change over time or could be artificially altered to cheat both humans and automatic systems. Other systems use magnetic or RFID HF badges, but they present two main issues: (1) as they are external "identity" carriers people may not use the or exchange tem, cheating the system (2) they are read only in certain point, not allowing a constant identification and tracking of people.
As we explain further, with the combined use of a computer vision system and an RFID UWB system, is possible to obtain a more reliable identification of people, that can be used also as an enabler of other services like multi-camera people tracking or to apply rules based on specific roles and privileges assigned to individuals.
By the User Interface side, in a video surveillance system, having a robust identification means a lower workload for the operator and the possibility to connect a wide range of information, i.e. the personnel database of a company, that helps the operative to get a deeper understanding of the context.

Tracking people over wide areas/multicamera tracking: in a real world scenario usually there is the need to monitor large areas like an industrial plant ora warehouse, but also to monitor irregularspaces like a perimeter or some important spots. To achieve this goal often multiple cameras are used. However for a human operator is difficult to follow an individual over large areas among different cameras. In fact, due to the disposition of the cameras around the area under video surveillance, operators may have to recognize people seen from different perspective. For example ceiling mounted cameras will see people from top while wall mounted will see them with an quite frontal perspective. Moreover videos coming from multiple cameras are not always displayed according to a "natural mapping". So operators have to recognize people and figure out the mapping of the cameras. "Blind" areas between the areas monitored may also create a pause between the disappearing and the reappearing of someone adding an extra difficulty to the operator.

Form a technical point of view performing tracking across multiple cameras is a challenging task. It implies the non–trivial problem of automatic re-identification of people across disjoint camera views, a matching task that is made difficult by factors such as lighting, viewpoint and pose changes and for which absolute scoring approaches are not best suited[126].
While is easier to give an identity to a certain "blob" in a given scene and track it, with multiple video streams is harder to be sure hat a person exited from a scene is the same entered in another one. In fact, different camera may present different perspectives on subjects observed but also different lighting conditions. Moreover the automatic identification could be different in context where people wear uniforms, helmets, and protective glasses.
Thanks to the use of RFID UWB (that will be further described) is it possible to obtain a more reliable identification thus enabling a more robust multi camera tracking. The automatic recognition allows adopting several visualization strategies, as the automatic highlighting of people on the UI, to help the user to follow someone across multiple display. Moreover the constant tracking of people allows to apply location based rules and having a constant awareness about the presence of people in a certain area.

Dynamic application of rules/ triggering alarms: controlling a place means applying many kinds of rules that may result in an alarm on in a request for an action of the human operator. Rules could be relative to people, places, specific parameters/conditions or could be also a mix of these elements i.e. in case of fire (specific condition) only the rescue team is allowed to enter in the spaces on fire. In consequence of identification and tracking of people rules could be based on roles and places. As an example only certain workers could enter in a certain area. Moreover also the behaviour could be a parameter subject to rules. In an area where vehicles are suppose to move constantly (i.e. a tunnel) nobody can stop more that a certain time, but if in the meantime there is a fire people standing still may be hurt and need help. Usually a lot of these evaluations have to be made by human operators. Automatic system may have a fixed set of rules and give an alert to the human operator. However, as situation may rapidly change is important (1) to have a correct situation awareness and fast reaction to unwanted events (2) apply dynamic rules based on the knowledge base available by the system and on the commands of the operator.
For the operator setting-up a new rule should be easy and fast. It could be done by a certain level of visual programming or with suggestion of the system.

Support users during emergencies: as a consequence of the rise of an alarm the system should support operators giving a correct representation of the situation and, eventually suggesting action to be performed.
Moreover there is the need to give the correct balance between the decision of the system and of the operator. For example in nuclear power plant, in case of emergency, human operators cannot interrupt the automatic procedures while in other environments operators have to confirm the decision of the system.

Integration with other AmI systems :the surveillance system may be only one of the elements of a more complex command and control system. It can integrate with other systems that are part of a wider Ambient Intelligence environment. As in the case of the iNeres[159]the surveillance system can send messages to mobile devices of people that have to escape from a building. The real time locating

system is able to detect presence and position of people and automatically calculate the faster/safer route to escape the building considering also data acquired from other sensor (fire, smoke etc.). Or it can be used to trigger some behaviour of smart objects inside the environment. We can imagine that if, analysing the behaviour of a worker, the systems detects that he is not paying attention to a machinery, it can be automatically deactivated or the system could operate some interface adaptation according to user's profile.
In further paragraphs we will describe the overall architecture of the system, the hybrid context capture system and the user interface and test performed on the system and on its effect on human performances.

5.2 The hybrid tracking system

People identification and tracking, used traditionally for video-surveillance applications or human behaviour analysis, has become an enabler for Ambient Intelligence applications[139]. Many different approaches have been studied using various technologies and methods. Among them computer vision techniques and wireless localization systems can be successfully used to track and identify people. These two systems can offer a good accuracy but in real world scenario their performances can be affected by several environmental factors. Moreover the two systems strengths and weaknesses appear to be complementary and can be used to build hybrid system able to overcome performances of the single systems.
The proposed system integrates a UWB-RFID-based remote localization system and a computer vision system. Current version of our system has been conceived for indoor applications.
The main motivation of such an integration is due to two elements: (1) the high level of precision given by the UWB-RFID system, and also the possibility to save ancillary information as the highness of the TAG and tracking people with TAG reliably; (2) the possibility to exploit advanced computer vision and pattern recognition tools for tracking people and also extracting important information as biometrics (faces) and other ancillary information.

5.2.1 Sensor fusion for context capture

Sensors are device able to capture certain types of information about a given context and make them available to systems to take decisions and act consistently within the same environment There are many types of sensors that can capture different aspects of the context in which they are in (temperature, pressure, acceleration, humidity, pollution etc.), with different modes (discrete, continuous) and with a certain frequency of detection. Data from different sensors or sensor networks can be fused to have a better understanding about the context or the entity that is to be monitored. In particular there are different ways to integrate such data according to the objective that is to be achieved and the characteristics of the available sensors. As evidenced by [109] the fusion of such data can be used to have more accurate and reliable information but also to describe more fully the environment and make inferences. The sensors are designed to operate in an environment where they can become fallible or even useless at the occurrence of certain conditions such as large changes in temperature, brightness, and pressure. Also, like all measuring instruments, the sensors have a

physiological inaccuracy related to the technology and the method by which the phenomena to be observed is detected. The sensors are also able to cover a certain spatial area not always uniformly (e.g. temperature sensor) and with a certain frequency of sampling and detection thus there is an intrinsic loss of certain information.
To overcome these limits, while designing a sensor network, three properties should be guaranteed:

- Redundancy;
- Cooperation;
- Complementarity.

Redundancy involves the presence of multiple sensors that measure the same parameter of the same phenomenon in the same space/time interval. Due to the overlap of these measures it is possible to reduce the errors and get more accurate values. An example could be several sensors measuring the temperature of a room. Each sensor will measure a temperature that is slightly from the others. Therefore is possible to calculate an average value and, more important, exclude sensors that gives an out of range value due to error.
Cooperation means having sensors capable of detecting different environmental parameters that can be combined in order to allow the formulation of inferences that otherwise would not have been possible only based on the data of individual sensors. Examples of cooperation are meteorological stations that are able to combine (through models) data about pressure, humidity, temperature and wind, to forecast weather conditions.
Complementarity can be described as the ability have a complete view of an environment adding the data from individual sensors that are able to sense the same parameter of different (partial) parts of an environment. This is the case of multi camera surveillance system. Through the use of multiple cameras pointed toward different zones is possible to monitor large areas.
These three different fusion strategies are shown inFigure 9. However these strategies could also be used together, depending on the sensor network used. For example two sensors may overlap only for a little portion of their range. In this case both redundancy and complementarity strategies may be used. Moreover groups of sensors could be used for redundancy and then their output could be used for cooperative o complementary fusion. However these strategies present also critical aspects mainly network congestion and error propagation. Indeed as sensor networks are often (see paragraph 2.1) made to transmit small quantity of data, a wrong dimensioning of the system may led to a network congestion, with delays or packet losses, or problems in elaborating those data. Moreover in certain case, if algorithms aren’t designed properly, sensor fusion may lead to propagation of error.

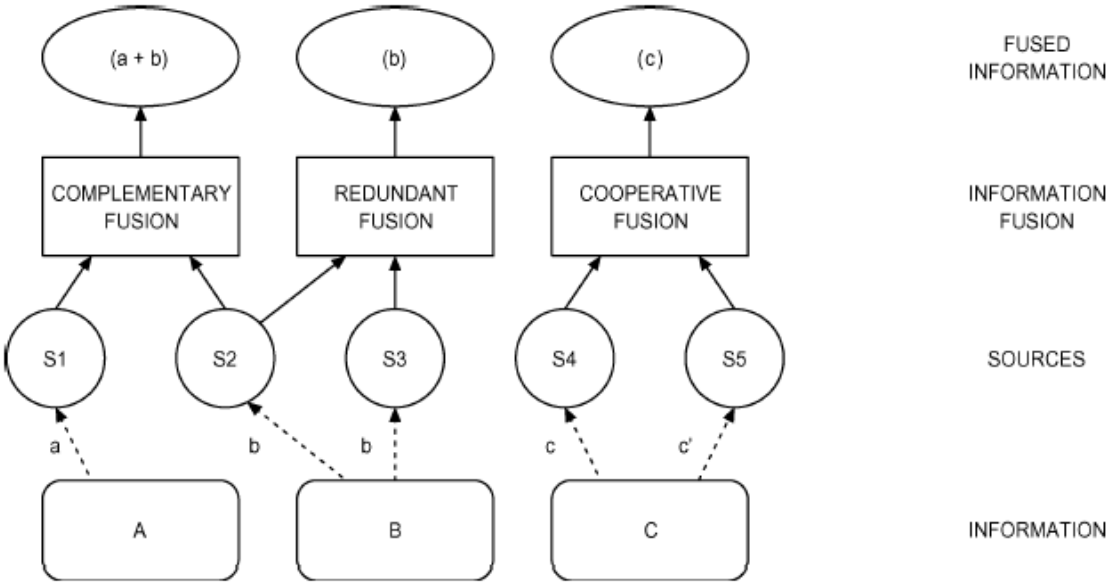

Figure 9: *Fusion strategies* based on the relationship between sources [109].

The integration of different types of sensors is a complex problem that usually generates systems composed of multiple elements. In this case it is useful to use a model of abstraction [79]able to identify the individual modules and functions involved in the information fusion system.
In particular in the model proposed by [107]a distinction was made between physical, informative and cognitive domains. The physical domain includes the sensors modules, each of which represents a sensor that interacts with the physical world. Each module in turn contains a sensor model. A sensor model is an abstraction of the process by which the same sensor captures information. It describes the information that the sensor is able to provide, how these are influenced by the environment, how it can be improved by information from other sensors or by subsequent processing. In this domain can be placed also actuators, if is needed to produce changes in the environment.
The information domain constitutes the centre of the system and contains modules for data fusion, application and control of resources and the interface with the end user. The data fusion module in turn consists of more data fusion sub-modules that are combined in a coherent "vision".
The cognitive domain relates to the presentation to the user of the information obtained and possibly the ability to act on actuators or operate additional modifications on the data.

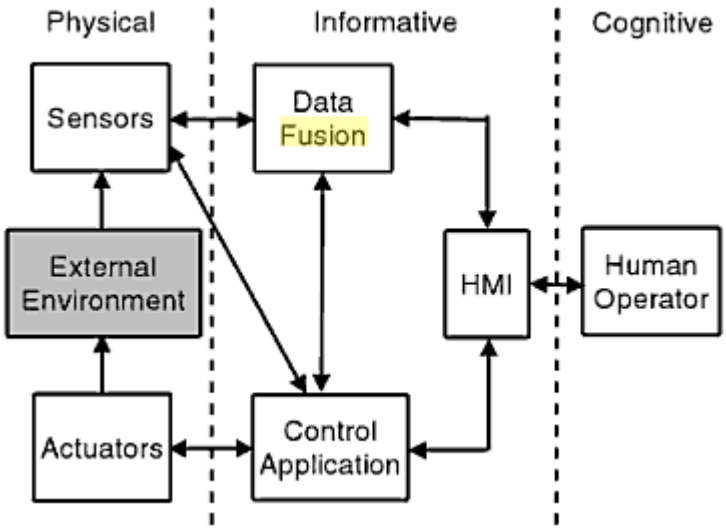

Figure 10 ***The different modules inside a multisensor data fusion system.***

There are countless interpretations and representations of the fusion process of data acquired from sensors. One of the most widely used models is the JDL [71] developed by the Joint Directors of Laboratories (JDL) Data Fusion Working Group. This model, conceived in 1987, was revised several times while maintaining its initial framework. The model is designed to be generic and applicable in many application fields and it identifies processes, functions and techniques applicable to data fusion. The model consists of a two-levels hierarchy and its fundamental structure is visible in Figure 11.

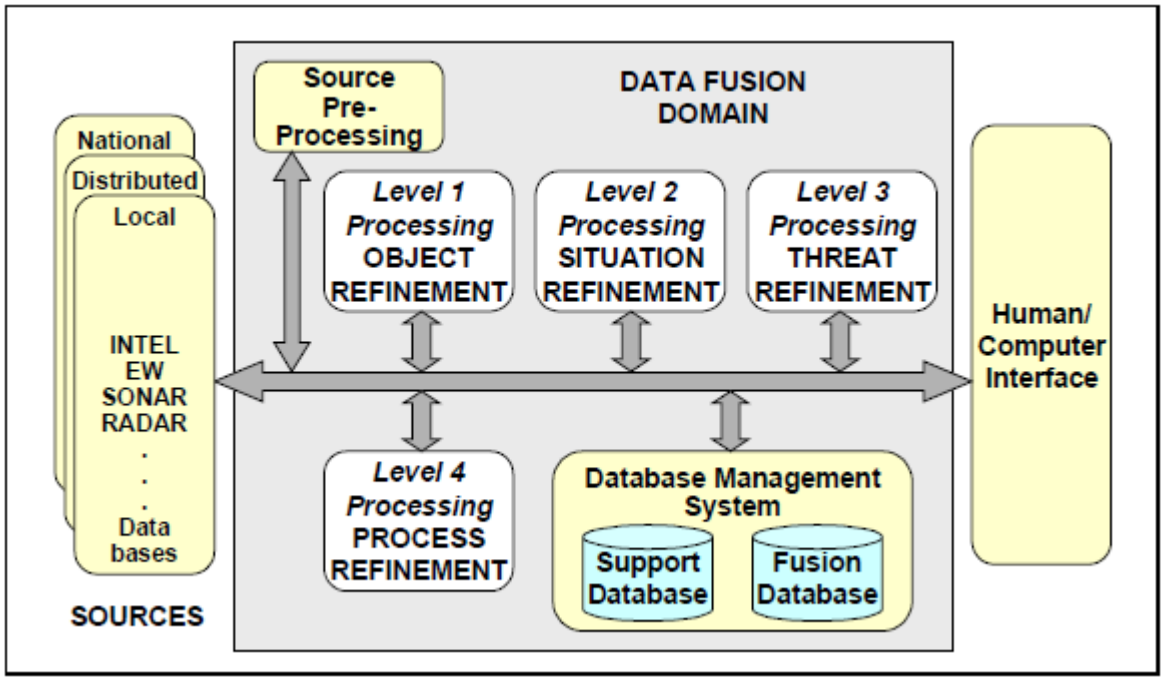

Figure 11 ***JDL data fusion model*** **[71].**

The data fusion process consists of: sensor input, source preprocessing, 4 refinement sub-processes database management, and human-computer-interaction. The levels functions are following explained:
Source Pre-Processing: combines signal levels/pixels to obtain initial information with respect to an objective's characteristic observed. At this stage a preliminary filtering on the data can be operated to delete irrelevant ones and reduce the workload of the system.
Level 1(Object Refinement): uses different types of data to get a more accurate representation of the target (identity, location, parameters...);
Level2 (Situation Refinement): draws a more accurate description of the environment looking for relationships between objects and events within the context observed and grouping objects within groups;
Level3 (Threat Refinement): performs a projection into the future of the current situation through inferences.
Level4 (Process Refinement): is a meta process that controls the overall performance of the fusion process. In particular carries out 3 tasks:

1) Monitoring of short and long term performance of the fusion system;
2) Identification of information required to improve the output of the multilevel fusion process;
3) Determines the conditions in which the sensor is able to acquire relevant information.

An important element of the system is the module that takes care of database management. The database is able to preserve the history of all data that, because of their heterogeneity, can make the management process particularly challenging. On the left in Figure 11we can see the Human Computer Interface module. This module is not only a simple input/output interface with the user, where it receives commands and presents the data obtained from fusion process. It should be considered as a complex system that takes into account the workload of its users, aiming to present the information in the most effective way, in order to maintain a constant level of attention and raise it in case of alarm.
In the specific case of integration of localization systems, specific models should be considered instead, such as the one proposed by [79]. The authors' objective is to create, for the data fusion in the context of localization, a model similar to the one for formalized communication networks in the 7 layers of the Open System Interconnection. In fact, this model allows different entities to talk to each other thanks to the standardization of all elements that are part of the communication network.
Similarly in a scenario that is ever more heterogeneous in terms of devices and protocols it is necessary to establish a common language and a clear framework of reference.
The model described consists of 7 levels as in Figure 12.

These levels can be described as follows:
Sensors:

Contains sensors (hardware and software) that can detect several logical and physical phenomena.

Exports to the next level raw data in various formats

Measurements:

Contains algorithms to transcribe raw data from sensors in the normalized types. It is able also to assess the bias based on the model (sensor model) of the sensor that generated them.

Exports: a stream of measures of distance, angle, proximity, position and non-geometric characteristics.

Fusion:

Contains a general method for merging continuously streams of measures in a probabilistic representation of the position and orientation of the object. During this process the different characteristics of sensors, redundancy and the contradictions are used to reduce uncertainty.If necessary, in this level a unique identifier is assigned to the objects.

Exports an interface that, based on events or requests, can provide the location of objects and the uncertainty of that data. More complex information can be offered performing calculations on obtained data.

Arrangements:

Contains a probabilistic engine to evaluate the relationship between two or more objects (proximity, geometric formations). The engine can also convert the given position to different coordinate systems.

Exports an interface that, based on events or requests, can provide the type of relationship between two or more objects located by the Fusion level.

Contextual:

Contains a system for position data merging with other types of contextual data such as personal data, ambient temperature etc.

Exports a specific interface as a system based on rules and triggers that activate specific applications.

Intentions:Contains the users' intentions.

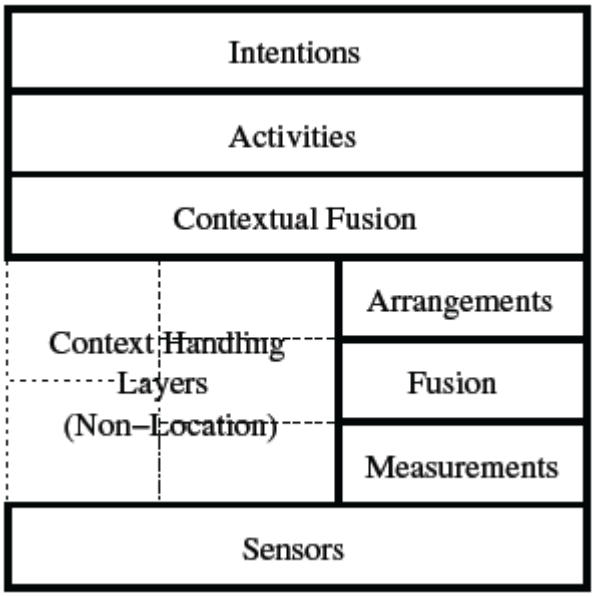

Figure 12 ***The Location Stack*****.**

5.2.2 Fusion of Computer Vision and RFID system

According to a bibliographic research conducted on works that propose the fusion of computer vision and RFID systems, three areas of application were identified:

- recognition of objects and people
- localization of objects and people
- recognition of human actions and behaviours

The three areas have multiple points of contact that can solve common problems. Hereinafter will be presented different works that are deemed significant and organized according to the three above-mentioned areas. A summary table of the characteristics of the different technologies used can be found at the end of each paragraph.

Recognition of objects and people

Various works [50],[17],[18],[28],[103],[86]explore the possibility of improving the performance of systems that recognize objects and people with the integration of RFID technology. In particular, many works use the possibility to store data relative to the objects on the database, relating them to the ID of the tags applied to them, or in the memory of the tag itself. This data is used to facilitate the computer vision system's recognition. In fact, the tags can contain information relative to the colour, shape or other parameters of the specific object by reducing the number of models which the systems must compare to the sample image. The RFID technology has the advantage of being able to distinguish between individual instances of the same objects, which are not possible in a computer vision system. However, since this is an external data carrier, it is not able to detect changes in the object itself, up until the limit case of separation of the tag from the object. In this case, the computer vision is able to take over by detecting a greater number of data and details on the object in question.Among the most interesting works, Cerrada [28] analyses the fusion of 3D vision techniques and RFID for object recognition in complex scenes. The authors show that the computational complexity of algorithms for the three-dimensional recognition/localization increases proportionally to the number of objects present in the database, which the scene will be compared to. Thus, they propose a system that is able to reduce this complexity and, as a consequence, the processing time. The system consists of a stereo camera and a UHF RFID reader. The RFID reader presents a list of tags, corresponding to the objects in the scene, as an output. The algorithm of recognition / localization will compare the scene only with the subset of objects detected by the reader as exemplified in Figure 13.

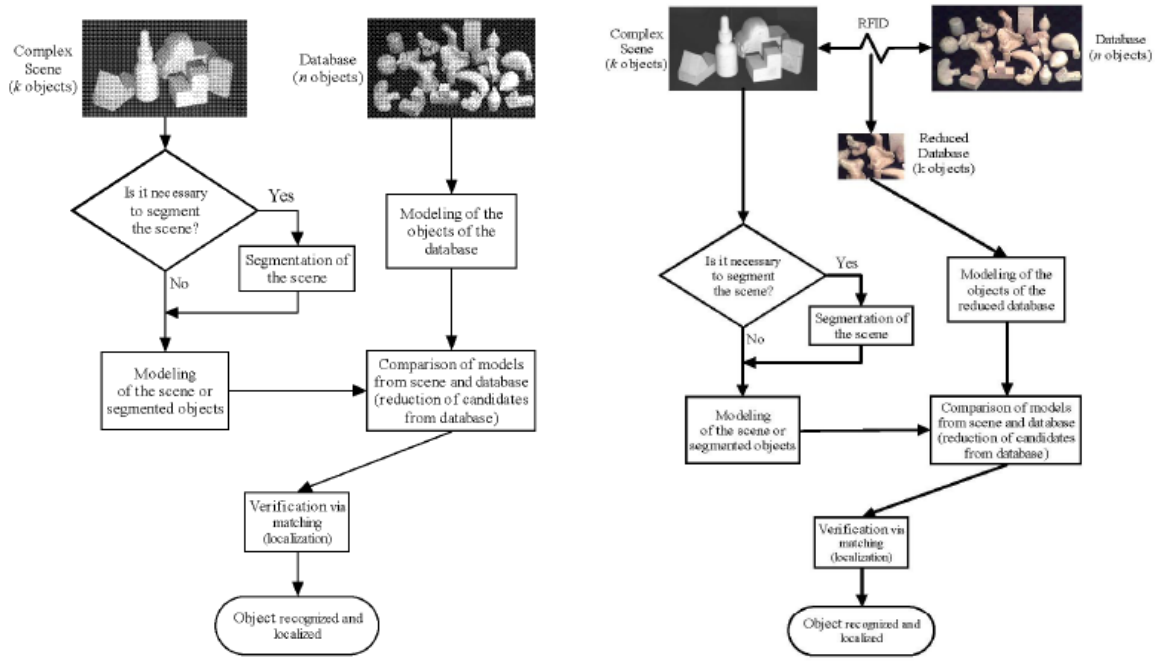

Figure 13 ***Hybrid object recognition.*****Difference between an object recognition schemes based only on computer vision and a system that also uses RFID technology.**

The authors present experimental results that show the true benefit in terms of processing time of the proposed system.

In [86], the author proposes a system to facilitate the recognition of obstacles and people by a robot that is equipped with a camera stereo and an RFID reader. In this system, RFID tags are applied to the objects and people. Information regarding the object are stored within the tag's memory so that the robot is able to recognize the characteristics of the object to be identified beforehand. The system applies a probabilistic model to determine the position of the object by iterating the tag reading cycles during the robot's movements. In this way it is possible to establish a grid of points where the tags and, thus, the obstacle is most likely to be found. Based on this calculation, the region of interest (ROI), which will be processed for the recognition of the person according to the scheme inFigure 14, is extracted.

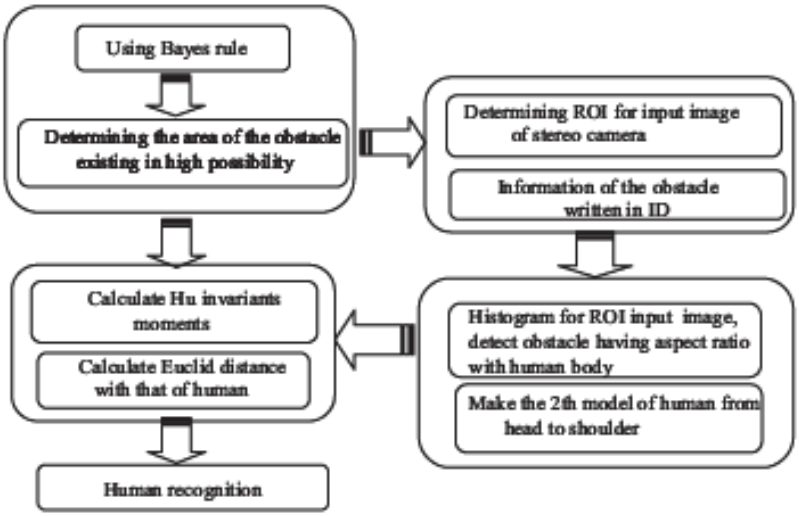

Figure 14 ***People recognition process.***

Recognition	
RFID	**CV**
Recognition of the single instance (unambiguous id)	Object class recognition
External data carrier	Distinctive characteristic of the object/person
Contemporaneous multiple identifications	Contemporaneous multiple identifications
NLOS	LOS
External data carrier	Distinctive characteristic of the object/person
Subject to environmental factors (destructive interference)	Subject to environmental factors (destructive interference, absorption)
Intrusive (need to apply tags/readers on objects/people)	Non-intrusive
Failure To Enroll	False Match - False Non Match

Table 1 ***Comparison of the characteristics of the different recognition technologies examined.***

Localization

Locating an object involves unambiguously establishing the position in a given environment or in a coordinate system. There are several localization technologies, characterized by the degree of accuracy (difference between estimated and actual position), area of use (indoor and outdoor) and reliability.

There are affirmed and mature technologies, such as the GPS, that ensure reliability and accuracy for outdoor localization. For indoor localization, there are numerous solutions that have different degrees of accuracy, reliability and complexity. This last context uses techniques based on wireless technologies and computer vision.

Below an analysis of two works that explore these two different scenarios:

- localization of multiple entities within an environment (such as video surveillance);

- an entity that must establish its own position and the position of objects or people compared to their own (for example, the orientation of a robot).

In [6], the authors carry out an in-depth analysis of the possible synergies between passive RFID, Wi-Fi and computer vision for locating and tracking people in indoor environments. RFID technology is defined "asynchronous and discrete": it allows to detect objects within the reader's range of action (50-100 cm for passive RFID) when the object is close but it doesn't allow an overall "view" of the considered environment. Additionally, the authors noted that according to the standards implemented and the size of the tag and the reader's antenna, the reading range might vary from 5 to 100 cm, introducing the problem of intentionality of the reading and of multiple readings. It is obvious that with a range of a few centimetres we have precise localization (the position of the tag almost coincides with the position of the reader) but decreases the possibility of random readings generating the necessity of forced paths leading the tag to approach the reader. Increasing the reading range, (which can reach several metres in the case of active RFID technology) the precision of the localization decreases (I know that an object is in a certain area) and the number of tags that can be read at the same time increases, generating ambiguity. In the mentioned system, there are various RFID readers placed in fixed positions and tags are assigned to the people present in the environment. The readings collected by various readers allow to track the movement of people within the environment.
Unlike RFID, the vision-based identification system provides synchronous information (equal to the capture rate) and allows monitoring larger areas of the considered environment with spatial continuity, albeit with growing problems of occlusion due to fixed and mobile obstacles. While the RFID system is able to unambiguously identify people with tags, the vision-based system assigns a temporary identity to the things present within the environment, which may be fallible due to displacements and occlusions. The proposed system identifies and tracks people with tags that enter within the RFID readers' range of action. By analysing the images, all people within the environment (with and without tags) are followed. The data collected from both technologies are compared by transforming the plane coordinates into the scene coordinates by an homographic matrix and by placing into correspondence to check the correctness of the identification of a person.The same authors also explore the possibility of combining Wi-Fi localization and computer vision. Wi-Fi localization differs from RFID systems, as it is spatially continuous. Wi-Fi localization offers an accuracy of 3 metres, which may vary depending on the number of "visible" access points and environmental conditions. Wi-Fi localization algorithms are normally based on the intensity of the signal, which are sensitive to the variation of the environmental conditions (e.g. entering of a group of people into an environment). Similarly to the system previously described, the data coming from both technologies are compared. Every system is able to localize a person or an object by defining an area of different dimensions related to the dimension of the system itself. In this case, the small area, defined by the vision-based system, is intersected to the largest area defined by Wi-Fi system by assigning a unique identity to the object located in this intersection.

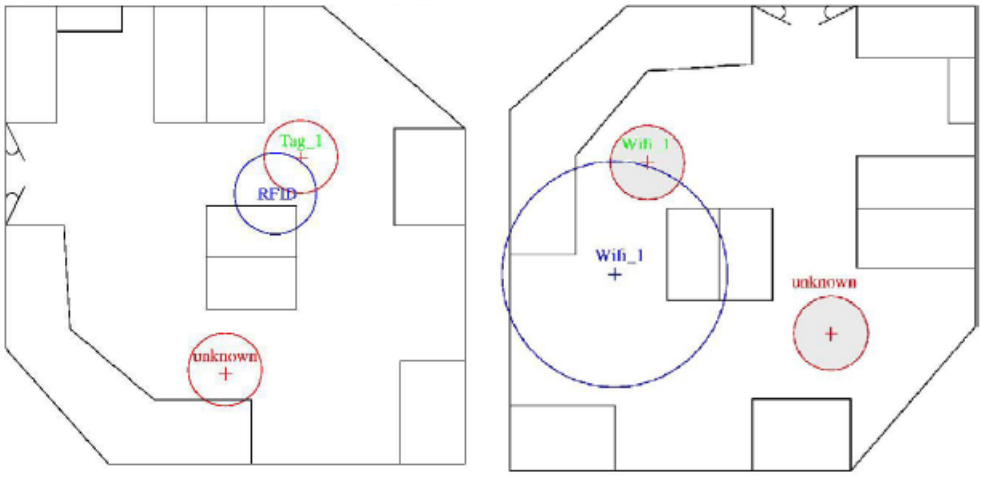

Figure 15: *Hybrid localization*.The image on the left shows the result of the localization via RFID and computer vision. The image on the right shows the result of the localization via Wi-Fi and computer vision

The so proposed fusion of information shows how various technologies can act synergistically with each other. In this view, an open application infrastructure that can collect data from different sources needs to be designed.

In[29] the scenario, in which a robot wants to calculate its own position and orientate itself within an environment is examined. In particular, the authors show a mixed RFID-Vision system for a robot in an indoor environment. The environment is divided into regions where passive RFID tags are installed. Thanks to a RFID reader, the robot is able to detect these tags by determining the region in which it is located. This process occurs by the reading of more tags. To minimize the localization error (attribution of a wrong region), to every tag is assigned a weight proportional to the tags distance from the boarders of the region.

Then a vision system is used to refine the localization of the robot and allow orientation within the environment. The vision system uses the Scale-Invariant Feature Transform that allows identifying a set of features for the matching, which has invariant properties to illumination from a rotation and scaling point of view. The comparison of the *feature descriptors* allows to find the best image in the visual map. However, this procedure only allows to identify the robot's angle of view. Estimating the distance between the current image and two images that are close to each other, of which we know the angle relative to the reference orientation and their distance in the space, it is also possible to calculate the robot's distance and obtain a more accurate localization.

Localization		
RFID	**Wi-Fi**	**CV**
Asynchronous discrete localization	Discrete and synchronous localization	Continuous and synchronous localization
Recognition of the single instance (unambiguous id)	Recognition of the single instance (unambiguous MAC)	Object class recognition
External data carrier	External data carrier	Distinctive characteristic of the object/person

Contemporaneous multiple identifications	Contemporaneous multiple identifications	Contemporaneous multiple identifications
NLOS	NLOS	LOS
Variable localization error depending on the technology used (5 cm ~ 1 m)	Localization error about 3 meters	Depending on system's characteristics.
Failure To Enroll	Failure To Enroll	False Match - False Non Match

Table 2 ***Comparison of the characteristics of the different localization technologies examined.***

Recognition of behaviour and actions

The recognition of human behaviours and actions is a complex task. In fact, it concerns the dynamic activity made up of many elements and variables (movement, interaction with objects and environment) that may have different meanings depending on the context. In this scope of application, the recognition and localization techniques analysed above are combined. In the works examined, the method used is the tracking of the movements of objects and people (with the relative problems of recognition and localization analysed above) and the subsequent application of probabilistic models to establish correlations.

By using unambiguous identifications in each tag, the RFID technology allows an accurate and reliable discrimination between specific instances of objects equal to the eye. Furthermore, it is possible to detect small objects and detect their presence even in poor lighting conditions or in case of obstructions. The vision based systems, albeit subject to errors, allow the detection of more details on the scene examined and are less intrusive to users.

In [81],the authors experiment a system to assess the behaviour of a single individual during the study. The system consists of a table with a RFID reader and a video camera installed in front of the user. The table is able to detect objects with tags placed above it while the camera can detect the presence of the person, the permanence and some basic behaviours (study, distraction, rest). The system combines the data according to predetermined rules that consider the behaviour detected by the camera and by the objects on the table. For example, if the RFID reader finds a book on the table and the vision system detects that the user's glance is directed downwards, we will infer that the person is studying.

In [122], the same technologies are considered in the wearable computing and tangible interfaces outlook. The system requires the user to provide a glove containing an antenna connected to a RFID reader and glasses with a micro camera and a system able to recognise characters and simple shapes. In particular, the system is designed to create new forms of human computer interaction. The information obtained from the RFID system and from the vision system can be used to recognize the behaviour of the user and activate software processes in relation to the natural actions of the user. The authors define a decision-making model based on rules and related only to one specific case (payment by credit card through the recognition of paper and a specific gesture).

An interesting work by Krahnstoeverand colleagues [94], explores the possibility of combining the visual tracking of human motion with the tracking of the objects through RFID technology with the aim to arrive to an accurate estimate of high-level interaction between people and objects. The system is made up of three main elements (Figure 17):

- A motion tracking module that uses a stereo camera able to estimate the movement of a user's

head and hands within the space
- A module that recognizes and tracks objects based on RFID.
- An activity recognition module that uses the information of the two previous modules.

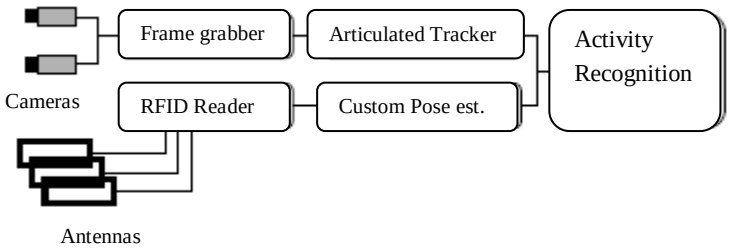

Figure 16 ***Elements of the system***

The motion-tracking module produces as its output a temporal series of the estimated position of the head and hands, and the presence of a user. The output of the RFID system is a list of tags present within the reader's range of action. The system was modified to obtain the value of the voltage emitted by the tag in order to estimate motion, orientation and distance. The activity recognition system compares these two outputs by comparing them with the rules established within the scenarios.
In particular, the system is able to detect simple actions that may be combined in order to identify more complex actions. For example, if the movement of the user's hand is detected and, at the same time, the RFID reader detects the appearance / disappearance or a variation in the intensity of the tag's field that is combined with the object *k*, the system infers that the user is manipulating the object *k*.

Recognition of actions and behaviours	
RFID	**CV**
High temporal resolution of the tracking of objects	Temporal resolution relative to the acquisition rate
Recognition of the single instance (unambiguous id)	Object class recognition
External data carrier	Distinctive characteristic of the object/person
Contemporaneous multiple identifications	Contemporaneous multiple identifications
NLOS	LOS
Subject to environmental factors (destructive interference)	Subject to environmental factors (destructive interference, absorption)
Intrusive (need to apply tags/readers on objects/people)	Non-intrusive
Failure To Enroll	False Match - False Non Match
Discrete and synchronous localization	Continuous and synchronous localization
Contemporaneous multiple identifications	Contemporaneous multiple identifications
NLOS	LOS
Variable localization error depending on the technology used (5 cm ~ 1 m)	Depending on system’s characteristics.

Table 2 ***Comparison of different strategies for behaviour recognition.***

5.2.3 Hypothesis for Location Fusion Models

Computer vision systems and real-time locating systems are sensor systems capable of detecting information from an environment that can be used to determine the position of a given entity (e.g. a person or an object).
Such systems are capable of detecting both information of the same type, enabling redundant and complementary strategies, and information of different types, allowing cooperative strategies. The information of the same type are location and identification. They will be redundant when referring to an overlapping area or complementary when referred to different areas. Cooperative data differs depending on the system and may cover physical parameters about the environment or about entities occupying it.
In particular the RTL system on its tags can have sensors capable of detecting certain physical parameters of the entity on which it is installed or the environment where it is placed in.
A computer vision system will provide instead semantic or geometric information about the environment. Geometric information will be related to the size and position of objects while the semantic ones are referring to the interpretation of images or videos, such as recognition of gestures, expressions, and behaviours.
However the fusion process cannot be restricted to simple comparison or integration of data but the information collected can also be used to give feedback on the system itself or to trigger events in the other system. Moreover the three fusion strategies seen could be combined indifferent ways and is possible to hypnotize two different models:

- Parallel model;
- Sequential model.

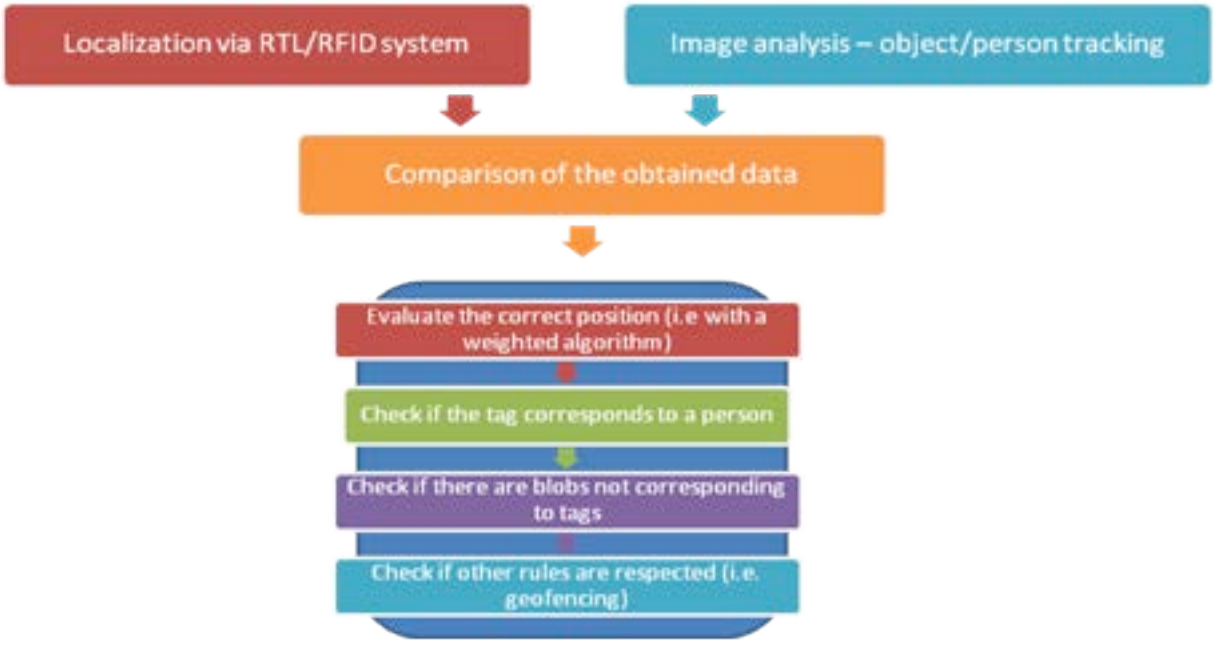

Figure 17 ***Parallel model.***

In the parallel model Figure 18the two systems produce, independently, an assessment of the position of an entity in the monitored context, a fusion level compares data taking into account the characteristics of the systems in order to obtain a more accurate localization. The characteristics of the systems identified in the sensor model may be represented by "weight" within the localization algorithm, giving greater importance to the system considered more reliable. At this level there is a check on some pre-set rules, possibly causing signalling of events. Complementary data are purified of evidently incorrect values (result of errors of the sensor) and fused simply by addition and possibly used for verifying pre-set rules. These rules could be based, for example, on the fulfilment of certain criteria of entry/exit from some areas (geofencing), but could also consider integrity policies on the same system as, for example, verify that each tag is associated with a blob and vice versa. Indeed the latter could generate ambiguity in the fusion process. In the case of blob detection not associated with a tag or vice versa in addition to the generation of an event data can be used in a cooperative manner, considering both trackings valid (ie. On the scene there are two people, one with the tag and one without) and eventually filtering them through other probabilistic networks.
In the sequential strategy the information acquired from one of two systems, the one generally considered more reliable, is used to direct the other system that is in charge to perform refinement functions and data completion. In the model proposed in Figure 19the wireless localisation system is considered more reliable, particularly for its ability to give unique identity to each tag. When the presence of an entity is detected its location is used to define the region (or regions) of interest (moving PTZ cameras) to be submitted to the image analysis system, to reduce the computational load of the system. However, this model does not provide the ability to use data cooperatively and identify cases where there is a blob but not a tag associated with it.

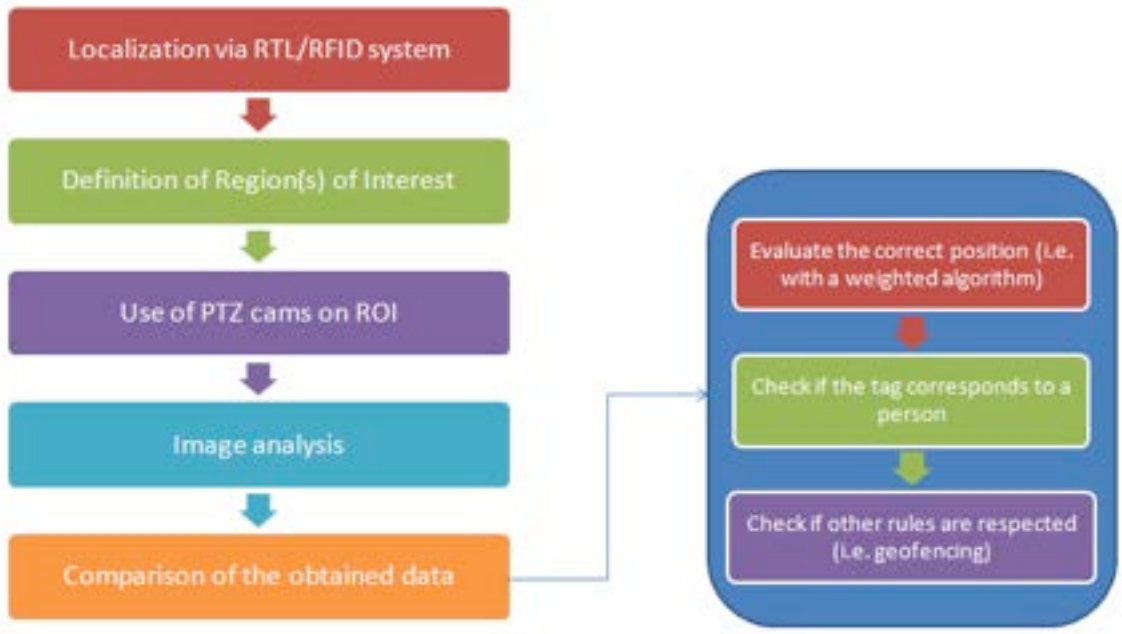

Figure 18 ***Sequential model.***

The proposed models (parallel, sequential) can however be integrated in order to sum the strengths and minimize their weaknesses. It is possible to hypothesize a "hybrid" model Figure 20 in which,

before the fusion level, there is a coordinator who can determine in real-time which is the most reliable system in that moment through indicators collected by the systems themselves but also by external systems like sensor networks (WSN). We can hypothesize, for example, that through a sensor or with the computer vision system, we can detect a massive entry of persons into the environment which is monitored. This phenomenon could cause an increase in uncertainty of the wireless system due to the decay of the intensity of the received signal caused by the presence of numerous subjects. In a condition like that even the computer vision system would encounter possible errors because of possible obstructions caused by movement of a large number of people. In this context, the Coordinator would apply to the system the parallel model by adjusting "weights" of the localization algorithm in order to reduce the uncertainty of localization. Conversely, in a scenario where it finds in an empty room the entry of people from one direction, such as through a door equipped with RFID locks attached to the system, the coordinator may decide to apply a sequential model where the RTL system determines the region of interest.

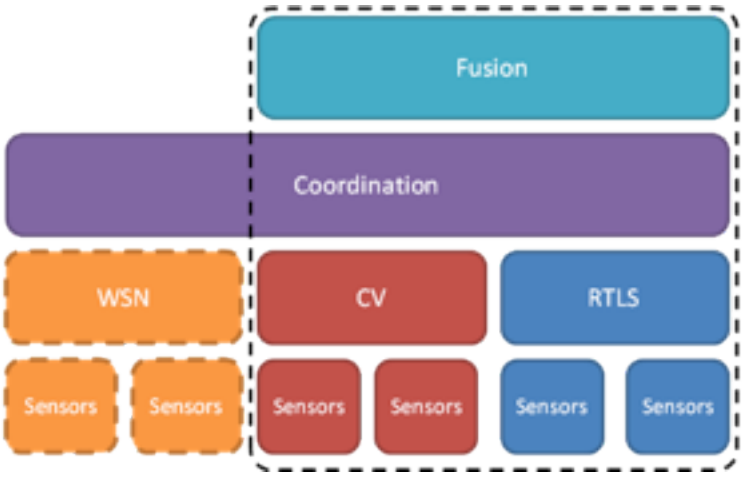

Figure 19 ***Hybrid model.***

5.2.4 Fusion Stack

Starting from the models explained in previous paragraphs we can produce an abstract model that helps to clearly identify the various components of the system, define their capabilities, and their role, including all levels, from physical sensors to the interaction with the final user.

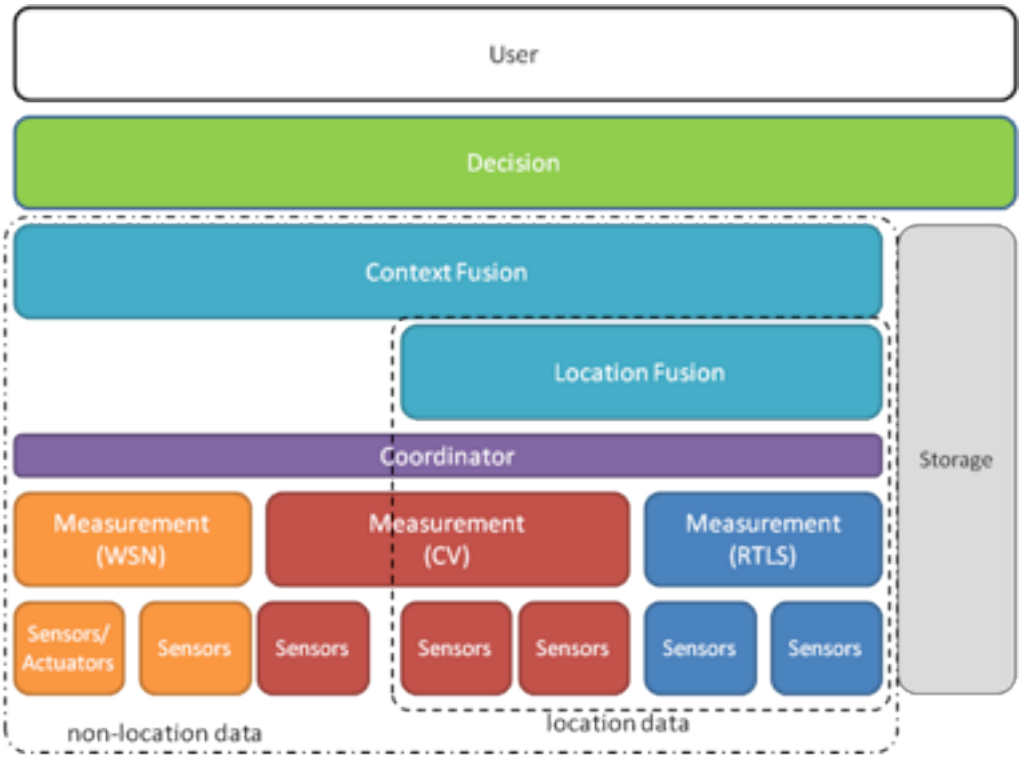

Figure 20 *Fusion Stack.*

The model Figure 21 involves the integration of different systems that produce heterogeneous data that require different processing. It is necessary, therefore, the division into multiple levels that deal with managing and processing the data and then fusing and using them to take automated decisions and for interaction with the user.

Within the model the integration with sensors and actuators networks is foreseen to detect additional parameters on the environment and the entity to monitor and act on it.

Starting from the lowest one the levels composing the model are characterized as follows:

Sensors: is the physical layer of the system, consisting of different types of sensors (tags, antennas, cameras etc.) detects the raw data about the environment and about the entity to be monitored. Each type of sensor is characterized by a model (the sensor model) that indicates the type of data collected and the bias. This model will be used by higher layers for subsequent processing of the data.

Measurement: is the level that takes care of processing the raw data received from the sensors to turn them into information useful for the higher levels. It is divided into blocks for the individual integrated subsystems: RTL, HP and WSAN.

> Measurement RTLS: through multilateration algorithms and, eventually, using pattern matching techniques with fingerprints previously capturedis possible to determine the position of the tags detected by sensors.
>
> Measurement CV: the images received from video cameras are processed with algorithms for extracting geometric and semantic information of the scene.

Measurement WSAN: performs calculations on the data received from different types of sensors. It filters incorrect values and transforms data into processable information from higher layers

Coordinator: The coordinator, according to certain environmental conditions, decides the modalities of work of the subsystems both at the individual level and as regards to the fusion strategies that will be adopted by the higher levels as explained in previous paragraphs. This level can give a feedback to lower layers for example moving cameras mounted in the region of interest.

Fusion Layer: is the level that is in charge of data fusion. Depending on the type of data received it is divided into Location Fusion Layer and Context Fusion Layer.

Location Fusion Layer: compares the data received from the detection systems in spatial (homography) and temporal (synchronization, interpolation) dimensions. Then the data are fused in order to get the location and identification of the entity with an error lower than that of the two systems.

Context Fusion Layer: deals with the fusion of context data detected by computer vision system, by the sensor network and by the RTL system.

Storage: the level of storage deals with the storage of data that can be used by other levels and for further analysis.

Decison Layer: it applies rules based on the information received from the lower levels and can take automatic decision or select alternatives for a human user..

User: deals with the interaction with the user. In particular deals with presenting the data in an appropriate manner, evaluating their importance and taking into account the impact of the user's workload. This level will also allow the user to change the status of the system and act on actuators when available.

In the present work we focused on the fusion between the computer vision system and a RFID UWB real time locating system to realize an hybrid people tracking system. The proposed system will be analysed further and experimental results will be presented.

5.3 System architecture

The system is composed of a total of 4 software applications that, collaboratively, allow the acquisition of data from sensors and their merging through fusion algorithms in order to generate events and/or alarms that can be managed from an operator’s workstation.

The system’s architecture is distributed, the needed functions where implemented in a modular and independent way, to preserve flexibility and ensure scalability. Communication between the different modules is guaranteed by an exchange of messages. These two choices allow to operate in a multi-language and multi-platform environment. As far as the architecture of the system is concerned, the previous works [24],[102],[32]were taken in account, which allowed to delineate a stratified system consistent of three levels:

• sources: this layer includes the acquisition of raw data from the sensors, and the software for their subsequent processing.

• fusion : in this level the incoming data are fused through prediction algorithms, allowing associations. Depending on the associations obtained, alarms or signals are generated. In this level of

data collection data are also stored to database allowing post-hoc analysis.

• visualization : all system clients belong to this level. Clients could be of various types, from smart surveillance systems to augmented reality.

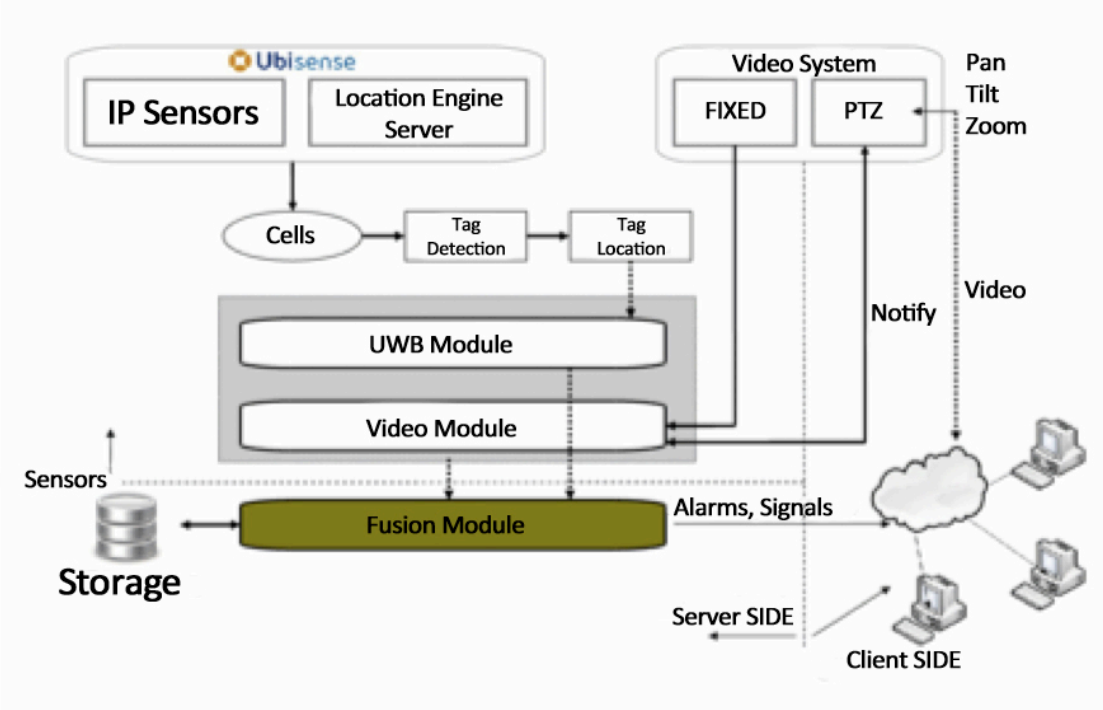

Figure 21 ***System architecture.***

According with the division into layers, 4 software modules have been implemented to cover separately each of the key roles of the project:

- The UWB and Video modules are part of the source level and are in charge to process the raw data coming from the respective sensors and send the obtained result to the merging module.
- The Fusion module occupies the same-named level, processing the data acquired from the sources in order to generate inferences. It can send signals/alarms to the clients. Received messages are stored on the database
- The clients represent the visualization layer, the point of contact with the user.

The system was installed in a rectangular indoor area of approximately 90 square meters (13.5 metres by 7 metres); the area, located in building 1 of the Sardinia Research Institute (Pula, Italy) is used as exhibition pavilion, and is an area overlooked by the laboratories operating on the floor.

The map below shows the preliminary installation schema and the coverage areas of the two sensor systems.

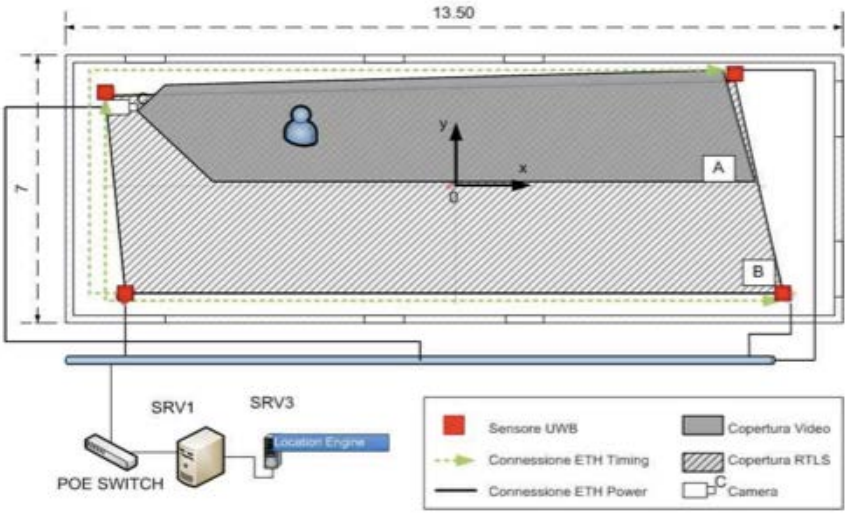

Figure 22 *Area of system installation*. The area marked with A and covered by both systems, the one marked with B and covered only by the RTLS.

As can be seen from Figure 23, the RTLS system forms a cell covering about 80% of the available area, installing the sensors on the wall with the maximum tilt possible. Instead, the video system is focused only on the left corridor of the area.
The installation was designed, to obtain, on one hand, the broadest coverage possible of the systems, and on the other hand, the determination of two kinds of areas, one where the two system overlap (A) and one covered only by the UWB system (B). In the following image it is possible to see one side of the area of the installation, with the indication of the video and UWB sensors.

Figure 23 *Area of system installation*. Focus on the "A" area.

5.3.1 System Specifics

The system modules have been developed using two different technologies: the Microsoft.Net framework, using the Visual Basic and C ++ language. For this reason, during the trials, the following configuration was used:

- the two .NET modules (UWB and Fusion) were installed on a machine (indicated as SRV1) with the operating system Microsoft Windows Vista 64.
- the two C++ modules (Video and RVId Client) were installed on a machine (indicated as SRV2) with the operating system Linux Kubuntu 64.

The two machines SRV1 and SRV2 belong to the same network segment and the first was statically addressed 192.168.0.1 to allow the use of a DHCP software on it. In the same portion of the network there is the IP video camera, connected in ETH, equipped with a small web Server for it's configuration. The camera provides MPEG-4 type video streams at 640 x 480 pixels via HTTP.
There is a virtual machine (indicated as SRV3) in the SRV1 machine. It has installed the Microsoft Windows XP operating system on which the proprietary Ubisense Location Engine Server of the RTLS-UWB system was installed. This machine is positioned on the same network segment of SRV1 and SRV2.
A POE Switch allows to supply power to the localization sensors and to attach them to the same network segment as the SRV3 machine.
In the SRV1 machine, as already anticipated, resides a DHCP server software with 192.168.0.x that assigns an IP address to all the entities described (SRV2, SRV3 machines and 4 sensors).
The SRV2 machine exposes an NTP server for synchronization; the SRV1 and SRV3 machines are able to synchronize with it using appropriate software that connects to SRV2.
The resulting system is rather complex and wanting to summarize the Hardware and Software that composes it, it is possible to list:

- Hardware:N° 4 Ubisense Real Time Localization Sensors; 1 Axis IP Camera; SRV1 Machine; SRV2 Machine.
- Software ◦ Ubisense Real Time Localization Software: *Location*
- *Engine Server*
- Software module for the processing of the data coming from the RTLS system: *UWB Dispatcher*
- Software module for the processing of the images acquired from the IP camera: *IFL*1
- Software module for the fusion of the data coming from the due sensors and for the generation of inferences: *Collector*.

5.3.2 IFL

The Video module used in the project, called Intelligent Face Logger, is in charge to acquire and process images from the cameras in order to extract the biometric information and spatial positions of individuals present in the scene. The development of the module, which was already available in a previous phase of the project, was mad on a Linux platform, with a lower level language with respect to VisualBasic.NET: C++, to allow the use of more efficient algorithms in the real-time analysis of the

acquired images. The camera used, produces a video stream which consists of frames whose size is 640 x 480 pixels at a rate of 30 frames per second; since the detection of biometric features in an image is directly proportional to the size of the source, only high-efficiency languages can ensure the processing of an image in less than 0.03 seconds (1/30).
The IFL tasks are in detail:

- -Blob detection in the video feed and the association of a unique identifier (per session).
- -Motion Detection
- -Tracking of the blobs in the video feed. This task allows the blob to maintain the same unique identifier between successive frames of a stream. In this way, the assignment of an identifier is based on the recognition of object.
- -Facial recognition. This task is carried out optimally by going to select the faces with better quality amongst those acquired by the video stream, which, otherwise, would be too many.
- -Application of homography techniques to perform a transposition of the position of a blob from the level of the image to that of the real-world reference.

Moreover the IFL moreover has to open a channel of communication with the collector to send the data on the location of the blob.

Figure 24 ***The IFL system.***

5.3.3 Ultra-wide band positioning

For the implementation of the wireless real time locating system a Ubisense UWB “Research Package” system was used .Ubisenseoffers a fairly new technology for accurate location sensing.

Generally, it achieves accuracy to within about six centimetres. However, it struggles with various obstacles as well. Ubisense relies on Ultra-wide band (UWB) Radio Frequency Identification.The system is constituted by tags, sensors and a server. Tags sends messages that are received by sensors that make a pre-processing on these data and sends the information to a server responsible for the processing.

Sensors used by the system allowto obtain information about the angle of arrival of the signal coming from the tag. This is possible because each sensor has an array of antennaseach capable of measuring the angle between the direction of arrival of the signal and the line parallel to the axis of the array and passing through the centre of the antenna.

Ubisense uses a bidirectional TDMA (Time Division Multiple Access) control channel for allocating timeslots to each tag. RFID tags transmit UWB signals to networked readers and are located using “angle of arrival” (AoA) and “difference of arrival”(TDoA) techniques. Ubisense claims an accuracy of about 15 centimetres in the three dimensionswith 95% confidence. The tags can be attached to various objects and people throughout the spaces, thus providing very useful information about any mobile object’s location.

Engineering issues currently restrict the scalability of the Ubisense system. Within a single cell, there is currently a trade-off between the number of tags within the cell and the number of updates per second that are achievable. Currently, our systems have been limited to only a small number of tags in order to provide faster updates. Shannon’s law would dictate that even with a significant increase in the number of tags, the wireless communication should permit a large number of updates. However, due to the TDMA controlling, the time slots must be sufficiently large to limit the possibility of overlapping transmissions and collisions.

Sensors used by the system allows to obtain, each individually, information about angle of arrival of the signal from the tag, this technique is possible because each sensor consists of an array of antennas, each of which can measure the angle between the direction of arrival of the signal and the line parallel to the axis of the array and passing through the centre of the antenna. This technique, however, is hampered by the presence of many reflected signals caused by the UWB characteristic bandwidth, this phenomenon is enhanced in indoor environments due to the phenomena of scattering by a number of objects in the environment.

In order to achieve a more precise localization, compared to using only the AoA, the system transmits a signal for time synchronization, which allows to use a multilateration based on TDoA, the difference of the times of arrival of two signals travelling between object and (at least) two reference nodes, to determine the location of the object within a hyperbola (having as focus the two reference nodes), as can be seen from the figure below.

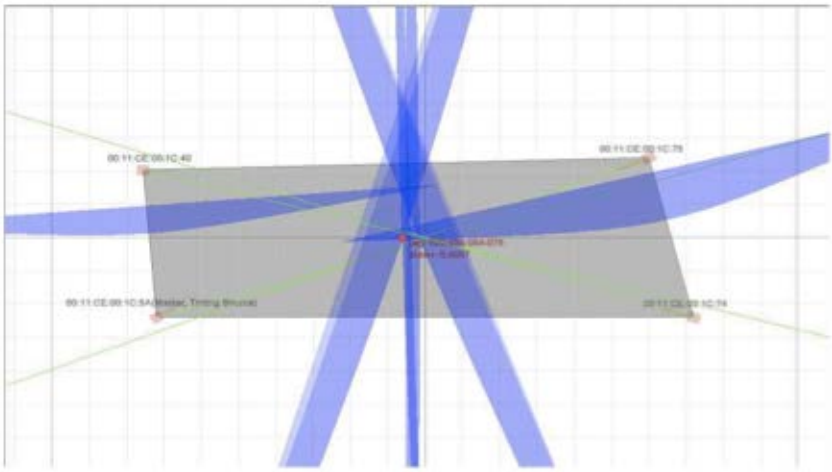

Figure 25 *TDOA localization of a tag.* You can see the hyperbole designed to locate the tag (red dot in the center of the image).

The tags used are sized 38 mm times 39 mm and transmit UWB signals at a frequency of 6-8 GHz. For normative compliancy and for monitoring a 2.4 GHz telemetry channel is present through which the sensors manage the behaviour of the tag, to indicate what will be the next time slot in which to transmit is present. The tags, in fact, emit pulses at a rate determined by the system that, generally, may be of 40 Hz or 160 Hz. In our case, the system works at 40 Hz, therefore it has a slot for the communication of 27,023 ms that can be used by the tags present in the cell. A tag, therefore, can send a signal with a maximum frequency of 1sample/0.109 s (4 timeslots) and a minimum of 1sample/0.221 s (32768 slots) depending on the number of tags simultaneously active in the cell. This frequency can be handled by the server via the implementation policies of QoS, aimed at better energy management of the tag. The QoS can be configured to reduce or increase the rate of the tag automatically in the event that a certain threshold speed is exceeded. The tags have, additionally, an on-board motion sensor that leads them to lower the rate of transmission whenever they are not in motion.

The RTLS system architecture foresees the connection of the network of sensors to a server that processes the data (the *Platform Server)* that hosts the Location Engine. The network of sensors follows a master-slave architecture, in which a sensor (one per cell) is used as a master, and therefore generates the reference timing signal, while the others take on the role of slaves. Each sensor is able to assume the role of master or slave as defined by the platform management software, which will send the correct firmware. The timing signal, required for the determination of the location with TDoA and AoA techniques, is transmitted on an ETH cable that connects in cascade or stars the slave sensors to the master.

The system requires a preliminary calibration that allows it to establish the correct location of the sensors, the delay, and the measurement errors; calibration is performed placing a tag in a known location. Sensors will “see” the tag and, based on the knowledge of the location of the tag (provided with the greatest precision possible), and of their own position, can determine their orientation in the reference system (the angles which form with the three axes of the system) and measure delays on timing. The calibration is fundamental to achieving maximum precision of the system, and in optimal conditions, it should be done with laser measuring tools.

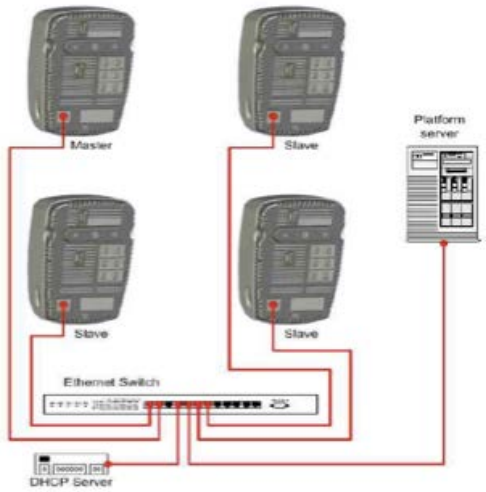

Figure 26 *Ubisense sensors* connected to the server.

5.3.4 Fusion Module

The most important process of the entire system resides in the Collector, and concerns the phase of data selection, fusion and comparison in order to obtain a representative score from the association; this phase is identified with the term "matching". The activities involved in this process took inspiration from the work ofCattoni and coll.[24]and Collins and coll.[32]from which the prediction schemes and the flow of processes were taken.
The matching process can be decomposed into the following phases:

- Data acquisition phase: the extraction of the data from their respective buffers. At this level the data is already filtered and *tagged,* , i.e. it is equipped with its own label, as it has both a *timestamp* that serves as a timing reference and a unique identifier.
- Prediction phase: the analysis of the buffers determines the "live" trajectories and predicts, using a prediction1 algorithm, the position of the trajectories according to referenced timestamp.
- Matching phase: the matching algorithm determines a distance between samples live trajectories, through a 1 to 1 comparison.

This approach to the problem of the sensor fusion comes from an adaptation of the work proposed by Collins [32], in which the use of the last two samples of a trajectory provides a prediction based on the linear law of velocity and on the statistical fusion of samples through the combination of two distributions, the prediction and the samples. The matching score calculated is proportional to the joint probability of the two positions (observation and object hypothesis). Given the experimental nature of the project, we used a simplified version of this approach, as it is done using only the linear prediction of speed leaving out the statistical information. Given the position of an incoming object, the algorithm predicts a new 2-D position hypothesis for every other know object in the system (assuming a constant velocity and a linear trajectory). Among the improvements planned for the future, the implementation of matching algorithms that take into account the statistical distribution of trajectories

will be considered, going to extend the indicative parameter of the degree of similarity to previous samples.
The life cycle of the data within the module could be summarized as follows:

- A thread listens for the arrival of new data that will transport, then to the other modules
- The incoming data are stored in a special buffer with a unique id. The unique id is determined by the succession of <Source Type><Source Number><Object ID>, thus it can deal with any strategy of ID allocation the sources may use;
- A copy of the incoming data is sent to a thread that will detect any crossing of the virtual fence;
- The crossing of the virtual fence triggers the process of analysis
- During the analysis process, at the arrival of data trajectories are selected in relation to their timestamp. Only the trajectories acquired in the 2 seconds previous the arrival of the new data are considered;
- -The selected trajectories are compared: for each trajectory, is determined a predicted position;
- A score corresponding to matching between the trajectories belonging to different sources is calculated. The score is determined considering the Euclidean distance between the trajectory in consideration and all the “alive” trajectories in the system.;
- The scores thus obtained are collected in a structure that is sent to the inference module. The structure will contain pairs of trajectories compared and the representative value of the comparison outcome;
- The inference module takes a decision based on the results of the comparison. The decision may be of match or non-matching. A classifier is used for the decision through the selection of the lowest score;
- Decisions resulting from the analysis of the comparisons are sent to active clients.

5.4 Tests

Test campaign have been undertaken to check the feasibility of automatic association of trajectories observed by the two subsystems and the reliability of high level information gained. This capability can be considered an essential function to build complex systems like video-surveillance o ambient intelligence applications. In this phase the locating accuracy is not evaluated, it will be subject of further investigations. Tests have been run inside the Ambient Intelligence Laboratory at Sardegna Ricerche (Sardinia, Italy), where the system is currently under development as described in previous paragraphs. As seen in Figure 28there is not a perfect matching between areas covered by the two subsystems. In this way is underlined how the two subsystems could be complementary.
Tests has been drawn up to check:

- Bias ;
- Distance between the trajectories achieved by the two subsystems;
- Matching of trajectories achieved by the two subsystems in different conditions;
- Crossing detection of a virtual fence;

An experimental dataset has been created logging the data achieved by the described system. People (that volunteered for the experiment) were asked to walk following different trajectories that have been designed in order to stress specific features of the system. Main trajectory types are:
The trajectories used are of different types:

1. Trajectory with three stops of 1 individual with tags. The stops must have a duration that exceeds 5 seconds, the movements between the stops must happen quickly;
2. Random trajectory of 1 individual with tags. Long duration trajectory (>40") with a stop in zone A;
3. Simple trajectory of 1 individual with tags, repeated twice from the door to the camera and return;
4. Simple trajectory of 1 individual with tags, repeated twice from the door to the camera and return. Repetitions with tag hanging around user's neck;
5. Simple trajectory of 1 individual without tag, repeated twice from the door to the camera and back. A tag is stationary. Repetitions with tags in two different points (P1 and P2);
6. Simple trajectory of 1 individual without tags. The individual follows part of the trajectory and then takes the tag;
7. Simple trajectory of 2 individuals, one with tag.
8. Simple trajectory of 2 individuals without tags. A tag is stationary. Repetitions with tags in two different points;
9. Simple trajectories of 2 individuals with tags;
10. Trajectories of 2 individuals with tags. The individuals will exchange the tags at the halfway point;
11. Trajectory of 3 individuals with tags. The third individual enters the scene at a later time.

As can be seen from the paths shown, we have always tried to follow trajectories that collect data from a transition between the two parts (A and B), starting and finishing outside of the joint coverage area (A).67 trajectories have been acquired and each one has been annotated indicating speed, direction and position of the tag (head or neck).

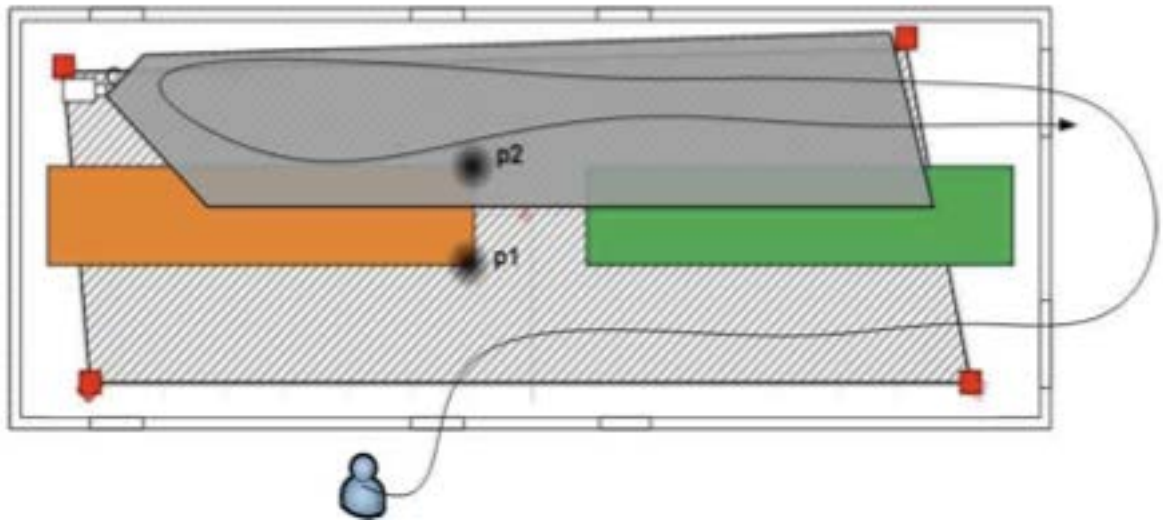

Figure 27 *Type 5 trajectory*. The user walks between the area where the two systems overlaps and in the area covered only by the UWB RTLS.

5.4.1 Bias

The phase of evaluation of individual subsystems requires a statistical characterization of the systematic error associated with it. To obtain this type of results, the experimental protocol has foreseen the acquisition of some trajectories of the dataset that help to detect the distinct phases of motion during the fulfilment of the path. For this reason the trajectories of type 1 were designed as the individual, in possession of the tag present on the scene, stopped for quite a significant period (>5 seconds) in known positions of the reference system. The transition from a position of "stop" to another was carried out with the greatest speed possible to highlight a sharp speed step compared to the static situation. The three stops occur approximately in positions of:

x = 350, y = 100
x = 0, y = 200
x = -350, y = 100

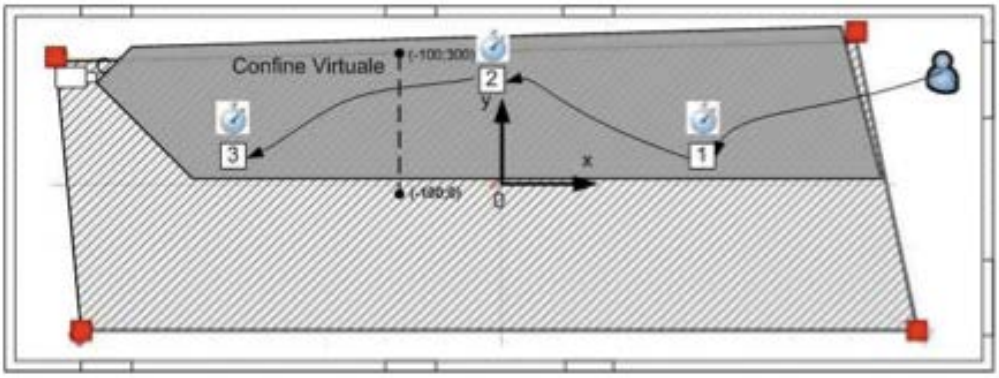

Figure 28 ***Trajectories used to test the bias.***

A filter has been applied with the purpose to isolate the trajectories points characterized by very low speeds, so that we can evaluate the static error. This analysis also helped to highlight any areas of joint coverage area (indicated by A in the reference system) in which the static error is greater than in others. Furthermore the trajectories of type 5 and 7 require the presence of a tag in one of two locations indicated as P1 and P2, with a simpler extraction of the static error of measurement of the static tags. During the measurements, given the absence of ground-truth, the mean value of the measurements was used to characterize the static error of measurement of the two systems. For this reason, for every source and for each stop area, the mean value of the points included in the associated temporal interval has been calculated; this value is then used to determine the bias in terms of average deviation from the mean value.

		IFL			UWB			
		x	y	Bias	x	y	Bias	Mean distance
Stop 1	Mean	273.18	207.27	6.2	351.77	153.7	13.62	97.05
	SD	50.1	5.79	5.72	25.88	24.28	16.19	
Stop 2	Mean	-114.22	268.33	2.71	-0.38	213.21	6.96	126.89
	SD	20.79	3.23	0.81	12.91	14.95	1.52	
Stop 3	Mean	-144.22	268.33	2.71	-0.38	213.21	6.96	126.89
	SD	20.79	3.23	0.81	12.91	14.95	1.52	

Table 3: Quantitative assessment of IFL and UWB bias. Values in centimetres.

Given these results we can assume that:

- IFL behaves in a not very repeatable manner, surely because of the small environmental variations that affect its performance, but behaves in a more stable way in measurements of positions, presenting data that, in most cases, has a very low bias. The system gives measurements which result compressed while reaching the horizon, from here the net localization difference with respect to the UWB system put in evidence in stop 1. Also, to confirm this the data relative to stop 1 is not specular to the one of stop 3, which should have been equal but of opposite sign.

- UWB is much more accurate in repeated trajectories, indicating a low influence of "environmental" variables, but proves little stability in the detection of positions, causing a more jagged data (albeit retaining a good precision) obvious thanks to an increasing bias compared to the IFL. The system behaves in an optimal manner in the area relating to stop 2 because of better coverage.

Results pointed out to scattered behaviour of the UWB subsystem, with a maximum accuracy of 15 cm. The CV subsystem presents an higher accuracy and lower variance, even if it can be affected by ambient conditions as environmental lights but also characteristic of the camera and of the Field of View (FoV) In fact the CV subsystem's accuracy decrease when the tracked person approaches the border of the image Moreover the test was useful to evaluate the different behaviour of the UWB system in relation to the different radio coverage of the areaFigure 29.

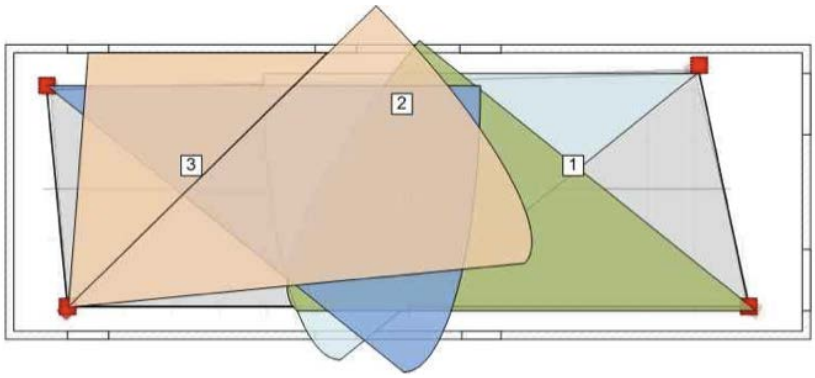

Figure 29 ***UWB sensors' coverage area.***

5.4.2 Crossing of a virtual fence

The system was tested to assess its ability to detect the crossings of a virtual fence, according to positions obtained from the sources. In this case the information fusion process is started by the event of a crossing. Linking the fusion process to an event is a solution to save computational resources as the system start looking for associations only when needed. In the tests, the chosen boundary is of a virtual type, in the sense that the system knows how to assess the position, but it is not associated with any barrier or obstacle really present in the scene. The use of the virtual boundaries allows to divide an area monitored by a camera in many sub-areas of different nature, without the need to use physical barriers. The virtual boundary used in testing is linear in shape and joins two points of the reference system whose coordinates are (-100.0) and (-100.300) as shown in Figure .

Going to evaluate the trajectories in which this phenomenon is more evident, it was realized that measurements obtained from the IFL system are close to the boundary and tend to vary the position in the direction of the x-axis causing the phenomenon which is well evident in illustration 5-22. For this reason the measurements carried out by the crossings executed by the IFL system have a number greater than those of the UWB system.

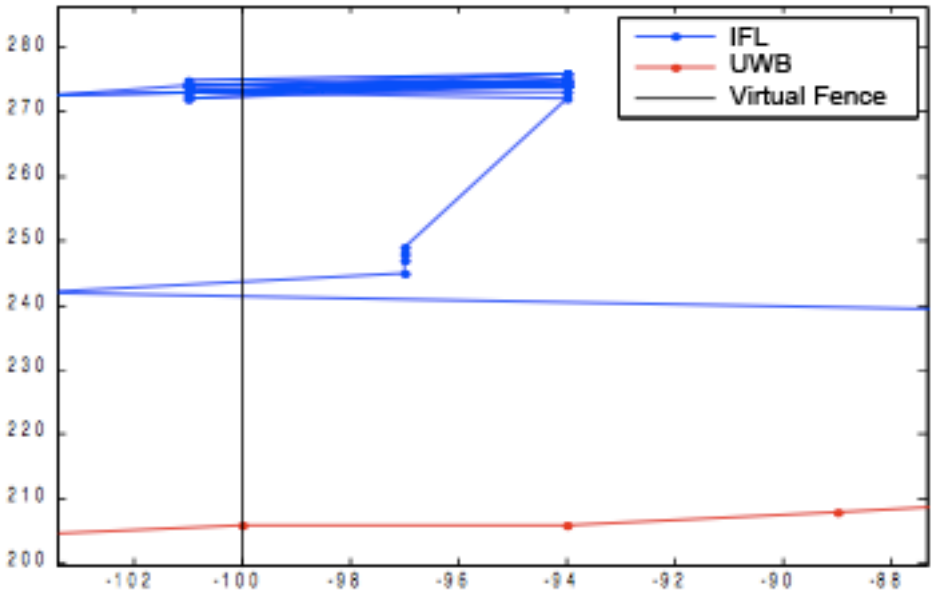

Figure 30 ***Crossing detection test.***

5.4.3 Matching

To measure the ability of the system to operate a correct matching between the trajectories acquire by the two subsystems we tested two different situations, characterized by different level of complexity:
(1) a person wearing a tag;
(2) two persons, one with the tag and one without it.

As in the first case both systems can be used to track the person, in the second case it is evident how the vision system can act as a backup of the wireless system ensuring the overall system reliability.
In the first case results pointed out that the two trajectories detected have an average distance of 120 cm with a standard deviation of 89 cm. Different threshold have been set to check the positive matching rate. At 50 cm it is 16%, at 80 cm it is 46%, at 150 cm it is 70% and at 180 cm it is 80%. This behaviour is consequence of the bias and of the different behaviour of the two subsystem, however, considering the dimension of a human body, this accuracy can be considered acceptable for people tracking.

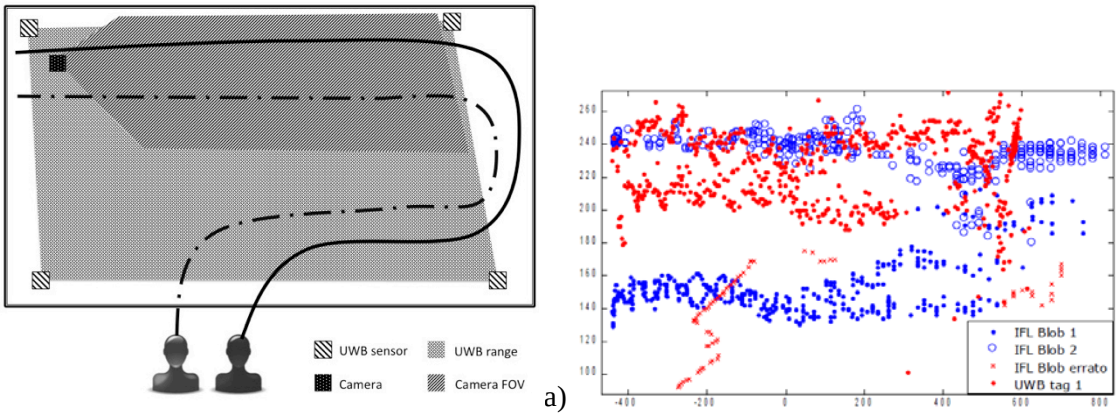

Figure 31 *The controlled area* (a) and an example of trajectories (b) correlated by fusing RFID UWB and computer vision systems information.

The second case presents another level of complexity. As described in 2.1 when multiple people are in the controlled area, the collector has to disambiguate data, by correctly associating the blob with the related TAG trajectory. Figure 31 b reports an example of two person with tag, detected by RFID and computer vision system correlated each other by the collector. The detected trajectory of the UWB tag appears irregular, as a consequence of the bias of the system. The distance between the two persons during the experiment was about 100 cm. The association between tag-person was based upon a proximity criterion. The implemented classifier showed a false positive matching rate around 60, even if the equal error rate (EER) is at 147 cm thus underlying a non-optimal accuracy in matching trajectories. In general results highlight a non-optimal performance of the system, however they show the ability of the system to operate a matching at a "coarse" level. This is related mainly to the bias of the two subsystems and to the use of a punctual association algorithm. It can be overcome operating on the physical setting of the experiment via an improvement, a fine-tuning calibration of the two subsystems and a more efficient matching algorithm.

5.5 Discussion

In this chapter, we presented an integrated system conceived for ambient intelligence applications where indoor areas must be under control of a human supervisor. The system is able to signal to track people and detect other events (i.e. crossing of a virtual fence) thanks to the fusion of information sent by two independent sources, namely, a UWB-RFID localization module and a computer vision module.

Results of tests highlight the different behaviour of the two subsystems. Indeed they show to have two different levels of accuracy and to be affected by different factors. The UWB system has shown to be more accurate even if with a scattered behaviour, requiring a filtering of not useful data. Moreover the accuracy is related to the position and the calibration of the antennas (Figure 29) and is affected by a source of radio interference at 2,4 GHz (a Wi-Fi network) . The IFL system has shown a lower accuracy but a more constant behaviour if compared with the UWB system. Moreover, as every

computer vision system, it is affected by changes in light and it becomes less accurate as the person to be tracked approaches the horizon of the image.
The ability of the system to match trajectories is not optimal. The trajectories detected by the two subsystems are usually at an average distance of 100 cm, thus the system may encounter errors especially in ambiguous cases i.e. there are two people one with tag and one without. However these performances could be improved with an intervention on the two subsystems, by adding more cameras to the IFL system and by a better calibration of the UWB system, and by using a more complex fusion algorithm. However the system is useful to show the possible integrations of the two subsystems. Indeed, we set up a mixed scenario, in which there are areas where the system overlaps, working in a redundant way, and areas where there is only one system, realizing a cooperative strategy. Moreover, as the two subsystems are affected by different factors they could act as a reciprocal backup.
In future developments the effectiveness of the system could be improved working on filtering and matching algorithms, as probabilistic models, especially to discriminate ambiguous cases. Further test campaigns will be undertaken evaluating also the locating accuracy compared to a ground truth system. Moreover the system could also be improved through the connection with other sensors, as a depth sensor, to have a better understanding of the ambient observed.

Chapter 6

Testing different Levels of Automation

6.1 Introduction

According to Parasuraman[120]automation does not merely supplant but changes human activity and can impose new coordination demands on the human operator. Automation can vary across a continuum, from a fully manual to a fully automated system and can be applied to different, replacing a function carried out by a human. To highlight where automation could meet humans activities,the author propose a simple four stage-view of human information processing, representing the way human achieve and process information to take a decision.

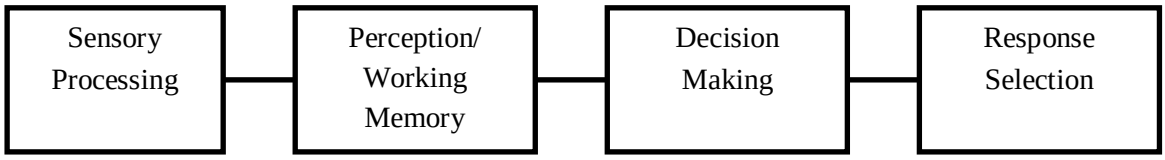

Figure 32 ***Four stage model of human information processing*** **[110].**

In the first stage there is the acquisition and registration of information achieved by multiple sources. It is a very basic step that may involve orienting and positioning sensory receptors (i.e. look at something), selective attention, pre-processing of data prior to full perception. In the second phase raw information achieved are consciously perceived and manipulated, keeping the process in the working memory. In this phase there could be also processes that may lead to the next decision phase (e.g. inference, rehearsal and integration).
In the third stage decision are taken according to the previous processes and, in the fourth, such decisions are implemented.
Even if it is a gross simplification it could help to understand the functions carried out by humans that automation could replace. Moreover in Wickens [80] and Gibson [64] some phases could overlap and be considered coordinated in a "*perception- action*"cycle.

Parasuraman uses these simple stages to draw a framework for automation design. To the four stages of (simplified) human information processing the author couples four classes of generic functions that could be automated in a system:

1) Information acquisition.
2) Information analysis.
3) Decision and action selection.
4) Action implementation.

Each of these functions could be automated at different degrees, realizing a number of combinations that may require more or less the intervention of humans. Representing different levels of automation in a continuum, these categories could be an alternative way to look at the levels of automation proposed by authors like Kaber and Endsley [87] and Sheridan and Verplank [134]. From these authors we can borrow some key points in the continuum that can be identified as:

- Massive presence of the user and automation used as a tool.
- Suggestion of the system to help the user.
- Automatic action of the system vetoed by the user.
- Autonomy of the automatic system.

For a better understanding of the model proposed by Parasuraman, that is taken as reference in the present work, we could better analyse the four classes of functions and how (and at which levels) automation could be applied.

The acquisition could be automated intervening on sensing and registration of input data. Analogous to the first state of human information processing step automating this level could lead to a better acquisition and organization of incoming data. Automation could be applied at a low level, driving sensor to achieve a better signal as, for example, moving PTZ cameras or adjusting focus for a better tracking of a moving subject. A moderate level of automation could consist in the organization of information acquired, according to some criteria. At this level some data could be highlighted (i.e. a priority list)but still presenting all the other raw data to the user. An higher level could consist in a filter applied to raw data, resulting in a presentation of a reduced list of salient element. This level of automation could also interact with the higher levels. Highlighting and filtering criteria could vary over time thanks to a feedback mechanism that, when higher level detect a certain situation or are in a certain state, change the rules applied to the first level. The effect of the automation in this level could be seen in a change of the mental workload as, simply highlighting or filtering data avoid the user to search long lists of useless data. Moreover there is an effect on level 2 of Situation Awareness,related to the better quality of the signal provided, enabling a better understanding i.e. a clear video feed helps an operator more than one that is out of focus.

Automating analysis functions require more complex reasoning as inferences. Data gained from the first level could be processed to make some predictions to help the user figure out future evolutions of a situation. Moreover an higher level of automationcould involve integration of data, combining different input in an higher level variable that can helps in making assumptions about a certain situation i.e. the speed of a car combined with the state of tyres and of the road could help in evaluating the space needed for an emergency brake. Even in this case prediction and integration algorithms could receive feedbacks from the higher levels, changing their behaviour depending on specific situations or system states. Automation in this level affects mainly perception and cognition of the user

The decision and the action selection imply the ability of the system to identify decision alternatives and eventually select one among them. The automation in this stage may result in different levels of user involvement. It is well represent by the scale proposed by Sheridan, points 3 to 7:
(3) Computer helps to determine options and suggests one, which human need not follow;
(4) Computer selects action and human may or may not do it;
(5) Computer selects action and implements it if human approves;
(6) Computer selects action, informs human in plenty of time to stop it;
(7) Computer does whole job and necessarily tells human what it did;
As can be seen automation could goes from suggesting a choice to the user to the full automation of the decision. A system could also switch from one level to another depending on the condition of the user or on some variable characterizing the situation. The theme of decision automation and its effect on the user have been investigated in literature and could be considered a quite controversial issue. In fact, as in Kaber[87],[89], Endsley [54], Parasuraman [120], when the user is not directly engaged in the process leading to a decision it may lose the Situation Awareness, going out-of-the-loop without being able to have a correct perception of what is happening. Indeed, when there is a system taking decisions, the user tends to rely only on the automation. Parasuraman calls this over-trust phenomenon "complacency". This could be very dangerous when the automatic system fails. A user that is out-of-the-loop is not able to recognize errors of the system. The complacency effect could happen also in other functional classes, however is more dangerous in the decision phase because it is joined with the risk of a low Situation Awareness. However also the amount of time the user has to take a decision is relevant to decide a proper level of automation. In situation in which the time taken by the user to decide could be too long to be effective, the system should decide autonomously (level 7 in the 10 LOAs of Sheridan and Verplank[134]) i.e. when the ABS system of a car automatically avoid the brakes locking. On the other side if the user has a reasonable amount of time a lower level of automation should be preferred. Also the reliability of the system and the complexity of the decision could affect the choice of the appropriate LOA. If a system is known to be perfectly reliable and the task is complex and prone to human errors could be reasonable to prefer a full automation. In nuclear power plants, when there are problems with the core, the first minutes are totally managed by an automated system, that is able also to implement actions.
Automating action implementation means "substituting the hands and the voice of the human" [120]. Even here there could be different levels of machine automation going from the opening-closing of a door to the piloting of remotely piloted aircraft (RPA).
As already explained all these levels could affect each other toward an adaptive automation that could take into account different condition to adapt the overall behaviour of the system.
Referring to the specific topic of Smart Surveillance we considered also the recommendations given by Endsley [56], that argues that system design should try to support and enhance situational awareness. She proposed a set of interface design criteria for enhancing situational awareness:

- Reduce the requirement for people to make calculations.
- Present data in a manner that makes level 2 SA (understanding) and level 3 SA (prediction) easier.
- Organise information in a manner that is consistent with the persons goals.
- Indicators of the current mode or status of the system can help the cue the appropriate situational awareness.
- Critical cues should be provided to capture attention during critical events.

- Global situational awareness is supported by providing an overview of the situation across the goals of the operator.
- System-generated support for projection of future events and states will support level 3 SA.
- System design should be multi-modal and present data from different sources together rather than sequentially in order to support parallel processing of information.

6.2 Experimental Design

According to this theoretical background we decide to test different level of automation of a smart surveillance system that could take advantage of the AmI system described in the previous chapter, to test the effect on an operator in terms of:

- Performance;
- Situation awareness;
- Workload.

Thank to the context capture system previously described it is possible to provide to the operator new sources of information that could be redundant or complementary to those normally acquired through the video feeds. Moreover some highlighting, filtering and calculations could be applied to this information affecting the first two levels of human information processing explained in the previous paragraph. These new information aremainly:

- Exact position of the people present in the observed environment;
- Identification of the people at a coarse level (role) or at a fine level (identity);
- Unexpected event detection (i.e. fall detection).

We decided to not implement automatic decision levels due to the risks in terms of out-of-the-loop and over-confidence on the system. Indeed experimental evaluation on the proposed system proposed in Chapter 5 underlined some reliability problems related to ambiguous situations that may requires a check of system's assertion.

According to the model proposed by Parasuraman [120] we decided to vary level of automation only of Information Acquisition and Information Analysis classes creating 3 different general levels of automation, corresponding to three different conditions tested:

- Manual (Condition 1): the user has only the bare video feeds. Only the fire alarm is automated as it was considered a so common technology that has to be included in a realistic system Figure 33.

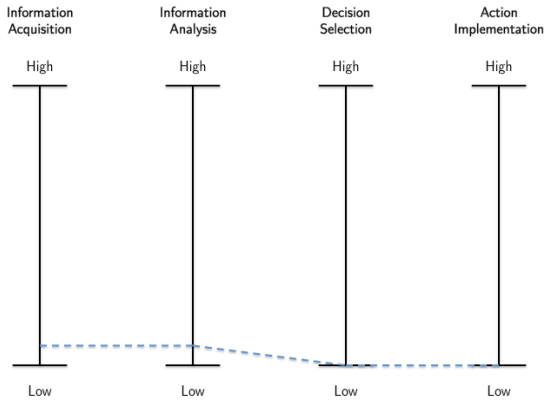

Figure 33 *Manual LOA* - Condition 1

- Low system assistance (Condition 2): the system gives a restricted set of cues to the user with some information highlighting. The system helps to identify areas where people are supposed to be. No information about identification is given. The user still has to do a lot of work to acquire information and analyse them. This level is represented in Figure 34.

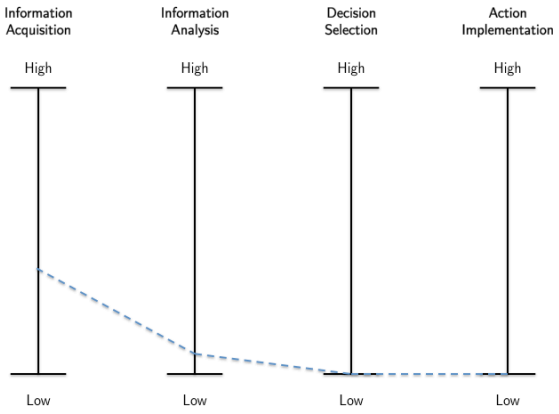

Figure 34 *Low sistem Assistance* - Condition 2

- High System Assistance (Condition 3): the system provides a bigger amount of information, with a real time people identification and tracking. Moreover a little analysis is made on data helping in the detection unusual events (i.e. fall detection) and making some calculation for the user (i.e. number or people in the scene).Figure 35: High System Assistance LOA - Condition 3

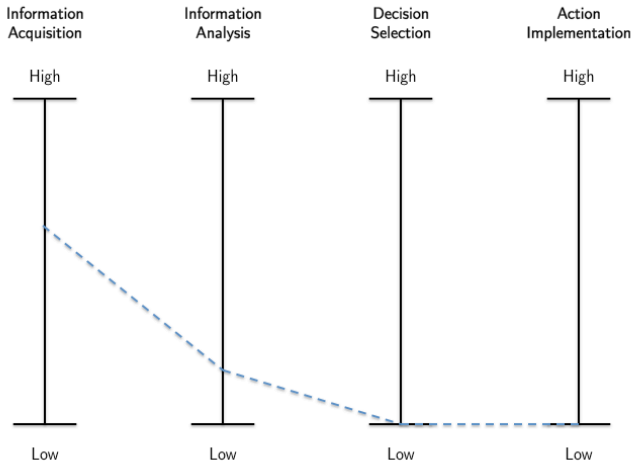

Figure 35 *High System Assistance LOA* - Condition 3

To test different levels of automation three simulator have been implemented. The simulators represent an interface of a video surveillance system. The use of a simulator, using recorded video, presents some limitation as emotional responsibility of operators, full communication, real work practices and shift- work cannot be fully addressed in a simulation [44]. Despite these weaknesses, simulation is still a powerful technique. Moreover the use of recorded videos allows us to simulate precise conditions and the repeatability of the experiment, with a fixed ground truth.

6.2.1 The simulators

Three simulators with different LOA were implemented to perform tests. To avoid effects related to the change of the interface among the three condition tested, the simulator have a have a common interface layout, similar to the one used by Girgensohn and coll. [67] . They are constituted by:

- a camera bank displaying 8 video feeds (200x200px) displaced according to a geographical criterion, considering the field of view of cameras, trying to reflect a natural mapping, to help the user get oriented. Each video feed has a label with the number of the room (or hallway) it refers to.
- a detail area (called Video Detail), that shows a larger view of a selected video feed (one at once) of 330x330 pixels
- a map representing the monitored floor, with placemarks indicating the cameras positions. It may have some interactive elements (further explained). In Every LOA when a fire is detected, a semi transparent red layer with an icon covers the room interested by the emergency.
- a control panel, used to fire alarms and call rescuers and where, according to the LOA implemented, several messages and information could be visualized.

In every LOA when a fire is detected an alarm blinks indicating also the room where the event is detected.
At every level of automation corresponds a different set of functionalities and a different kind of interaction with the user:

- Manual (Condition 1): The user sees the video streams in the camera bank. He/she can choose the video to enlarge by selecting it (by a click) inside the camera bank. The user can choose the camera also clicking on the placemarks on the map. Moreover when a video is in the detail window the corresponding one in the camera bank and the placemark on the map present a glowing effect to indicate the correspondence. The same effect occurs when the user rolls over a placemark or a video with his mouseFigure 36.
- Low system assistance (Condition 2): it has the same features of the previous condition plus the indication, through a semi transparent super imposed layer of the room where people presence is detected.

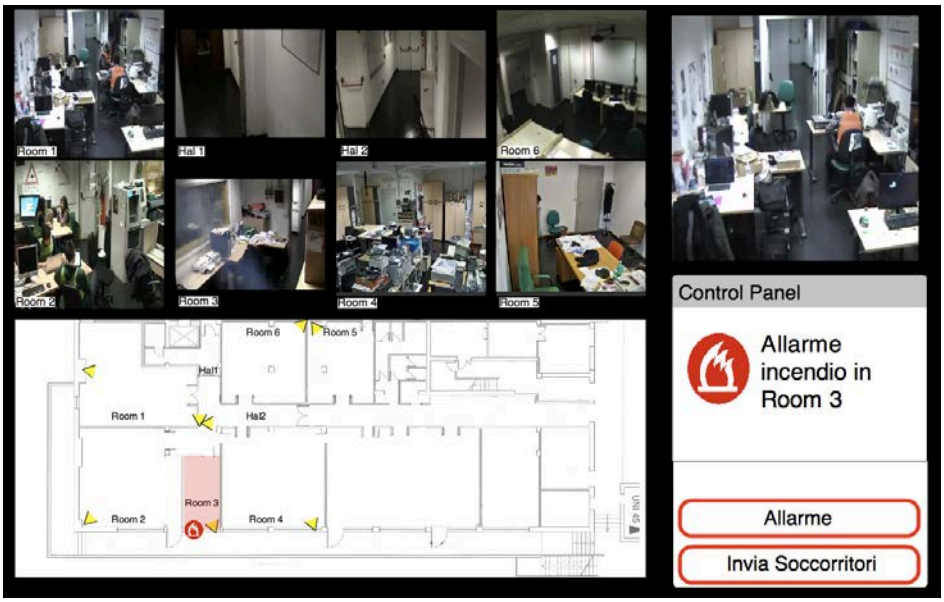

Figure 36 ***Interface layout of simulator in condition 1 in the event of fire.***

- High system assistance (Condition 3): it has the same features of the Manual simulator plus: (1) the indication of the exact position of people on the map, with a placemark of a colour indicating the role of the individual. Clicking on one of these placemark will make appear a box with the information about the individual; (2) the indication, on the control panel, of the exact number of workers and rescuers and, eventually, the number of rescuers to add. In case of incorrect balance between rescuers and workers a red blinking box appears around these information; (3) A drop down menu with the list of people that are in the observed area. Clicking on a name the corresponding placemark on the map will glow and a box will appear with information on role and identityFigure 37.

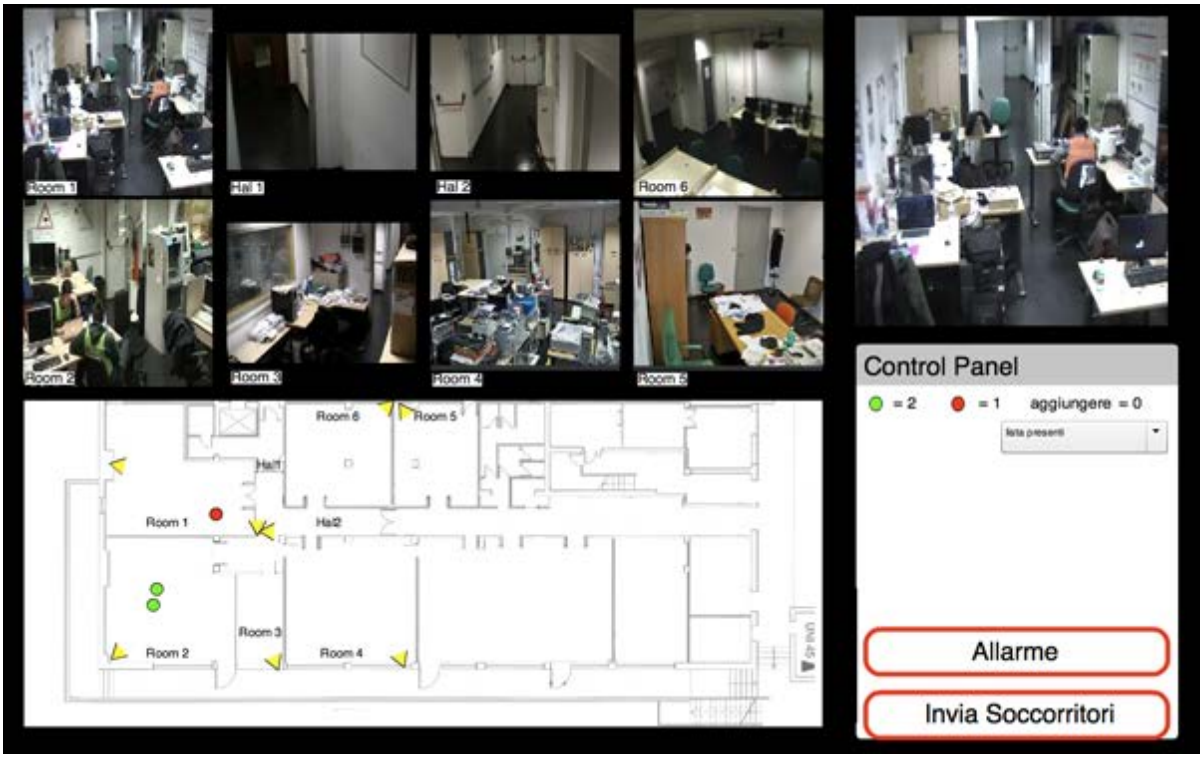

Figure 37 ***The simulator's interface layout in Condition 3.*** **Inside the control panel is possible to view the number of workers and rescuers present in the lab. The green and red dots on the map represent the location of individuals.**

All the alarms and important notifications were given in a textual way with blinking icons (2Hz) drawing the attention of the user. Moreover we decide to show only essential information on the map (as position and role of persons in high system assistance condition) to avoid visual cluttering.
Indeed a long monitoring task could lead to boredom and a degrading of attention generating Change Blindness (CB) and Inattentional Blindness (IB) that may affect the operator's SA. CB is the failure to detect a change (colour, position etc) in an observed phenomenon while IB is the failure to detect an unexpected stimulus even if looking at it. Manceroand colleagues [101], during a field test in a Police Control Room, observed how the CB rate was kept low thanks to a graphical interface drawing the attention of the operators on new events with colours and other visual cues. Moreover authors observed that operators avoid the use of a map, part of the Emergency Management System used, because it was heavily cluttered. A textual log was preferred instead (despite the lack of spatial information). Thus confirming the importance of the readability of information visualization and the use of visual cues to draw the users' attention on the right points.
The simulators were developed using Adobe Flash CS 5 and ActionScript 3. The interface area is of 1280x800 pixels. The simulators were able to recognize videos' cue point for the exact synchronization among video themselves and to recognize events. The simulators were able to log on a csv file all the interactions of the user with the interface. Particularly it counted the seconds between the occurrence of an event and the subsequent action of the user. The simulators were programmed to

have a training mode and a test mode, where the execution is “freezed” every 5 minutes and restarts at a specific command of the facilitator.

6.2.2 VideoDataset

A video surveillance dataset has been recorded to perform the test. The dataset includes 21 minutes of a variable number of persons (3 to 7) acting inside an indoor environment recorded by a network of 8 synchronized cameras. Videos were recorded at the resolution of 640x480 pixels at 30 fps. Video sequences were also manually annotated with ground truth. The dataset was recorded inside the Cattid laboratories (Sapienza University, Rome, Italy), the observed area is about 250 square meters, and the individuals involved are volunteer laboratory’s members.
In Figure 38is presented the plan of the monitored area indicating the cameras’ position.

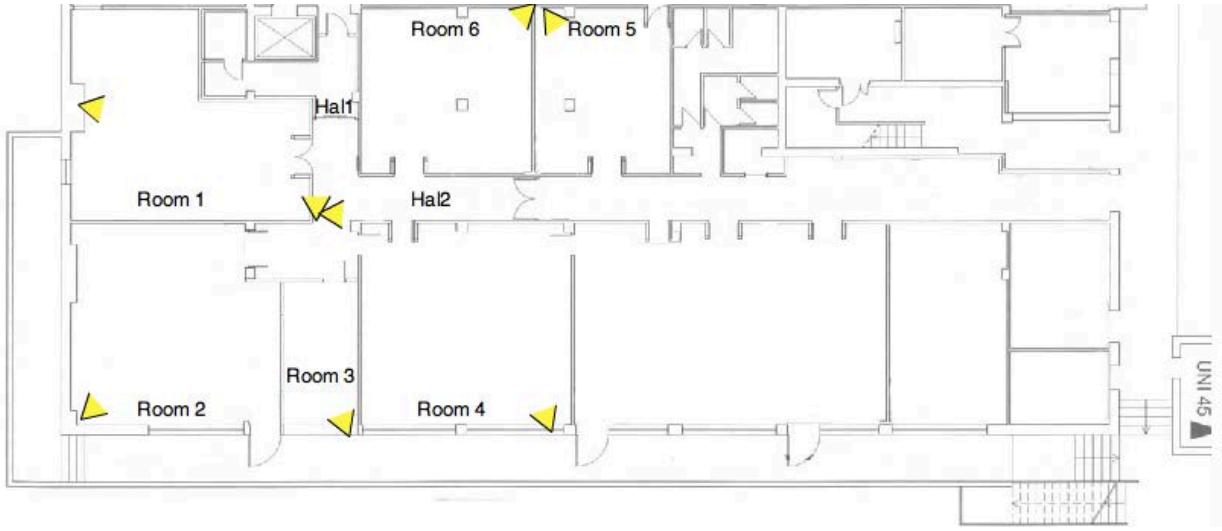

Figure 38 ***The monitored area with the position of cameras.***

Users’ activity and events have been scripted to create specific sequences that allows to test different aspects of a video-surveillance system and different task of the operator that is in charge to monitor the videos. These conditions created are:

- Individuals moving within the laboratory;
- Individuals divided in two groups dressed with jacket of different colours (green and red);
- Individuals entering/exiting the laboratory;
- Unusual events related to dangerous conditions.

The two groups of people correspond to a hypothetical team of workers and to a rescue team. As the test scenario is about a controlled environment i.e. an industrial plant, it is plausible that there are people wearing a uniform. They behave according certain rules related to their group i.e. when there is an emergency the rescue team has to intervene. The video operator knows these rules, and have to consider them while evaluating a scene. Individuals move inside the laboratory, they can also exit and new individuals enter. All the individuals move almost at the same speed, to limit the independent variables to be considered as, according to the work of Girgenersohnand colleagues [67], speed factor

may affect user's performance. The unusual events happen in rooms with nobody or only one person inside. This gives more responsibility to the video operator as he/she is the only one that can detect the event, nobody can call to fire an alarm. In relation to certain events the video operator is supposed to perform certain action, according to formal procedures explained to him/her. Thus the sequences are scripted as the operator behaves correctly so, even if he/she fails, after 45 seconds the sequence continue as if the correct command was given i.e. there is a fire: 45 seconds after the smoke is well visible a rescue team enters the room. The 45 seconds are considered (in the experimental setting) the sum of the time needed from the operator to understand the situation and give a command and the time for an individual inside the sequence to execute the command. Every sequence ends coming back to a "normal" condition with people behaving normally.

The two "unusual events" sequences are: (1) a dense smoke starts to fill an empty room, 2 rescuers came and stay there and the smoke begin to disappear; (2) a person, alone in a room, start to behave in a manner that should represent an illness and suddenly slumps on floor.

Moreover in one case a rescuer is in a zone partially not covered by cameras, and only the arms are visible (this will be useful to show how another locating system could act as a backup).

Other two challenging situation are simulated: (1)Few seconds after the fire alarm a worker enters the lab, requiring the user to send another rescuer while he/she is still paying attention to the fire; (2) after the ill worker slumps on floor three rescuers come in the room and take him outside requiring the user to call two of them back, to preserve the correct workers/rescuers balance.

Each video was annotated, highlighting important events, and as they occur, describing also the situation:

- Individuals in the scene and their group (workers/rescue);
- Location of the individuals;
- Entrance or exit of someone (and the coordinates of the point of entrance/exit);
- People entering and exiting from areas not covered by cameras;
- Unusual events;

Videos have been encoded using the flv format and cue points have been added to automatically trigger events in the simulator used to be sure of the exact synchronization.

All the annotations were merge into a human readable file describing the global situation, useful for the evaluators to evaluate the performance of the video operator. In this file the location of individuals were in a generic way (only the room or the hallway is indicated) and movement in hallways are not annotated, but is considered only the final destination of the individual (i.e. 1 red enters the lab from door 1 and goes in Room 5). Also a summary of the situation is presented indicating the number of people inside the scene for each room.

The main events annotated are:

- Entrance: one or more individuals enter the lab from outside;
- Exit: one or more individual exits the lab
- Moving: one or more individuals move among different place of the lab;
- Alarm: an unusual event is detected.

In Table 4we can see the events scripted in the dataset, the actions that the user is expected to do, and the SA and workload assessment.

Extra 5 minutes of video dataset, not annotated were made for training before preforming test

Time(mm.ss)	Event	Location	Subject	Description	TOT Green	TOT Red	Expected action
00.00	Start	Global	1 red; 2 green	The sequence begin	1	2	
01.50	Entrance	Room 5	2 green	2 green enter the lab from door 1 and go in Room 5	4	1	Ask 1 red to enter the lab.
02.50	Entrance	Room 1	1 red	1 red enters the lab from door 1 and goes in Room 1	4	2	
04.00	Moving	Room 5	2 green	2 green move from Room 5 to Room 2	4	2	
05.00	**first assessment**						
06.00	Alarm	Room 3		Smoke is detected in Room 3	4	2	Ask 2 red to enter in Room 3 to check alarm .
06.50	Moving	Room 3	2 red	2 red move from Room 1 to Room 3	4	2	
07.00	Entrance	Room 5	1 green	1 green enters the lab from door 2 and go to Room 5	5	2	Ask 1 red to enter the lab.
07.45	Entrance	Room5	1 red	1 red enters the lab from door 1 and goes in Room 5	5	3	
10.00	**second assessment**						
13.00	Entrance	Room1; Room2	2 green	2 green enter the lab from door 2. One goes to Room 1 and one to Room 2	7	3	Ask 1 red to enter the lab.
13.45	Entrance	Room 5	1 red	1 red enters the lab from door 1 and goes in Room 5	7	4	
14.00	Moving	Room 5	2 red	2 red goes from Room 3 to Room 5	7	4	
15.00	**third assessment**						
17.15	Alarm	Room 1		1 green slump on floor	7	4	Ask 3 red to enter the Room 1 to help the green.
18.00	Moving	Room 1	3 red	3 red go from Room 5 to Room 1	7	4	
18.15	Exit		3 red; 1 blu	3 red take 1 green out of the lab through door 1	6	1	Ask 2 red to enter the lab.
19.00	Entrance	Room 5	2 red;	2 red enter the lab from door 1 and goes in room 5	6	3	
20.00	**fourth assessment**						
21.00	End						

Table 4 ***Annotation of events occurring in the video dataset.***

6.2.3 Participants

Twenty fourparticipants (12 female) between the age of 25 and 35 (mean 29.5 SD= 3.36) volunteered the study. Participants had no prior experience of video surveillance tasks but they were confident in using computers and "platform" videogames. The latter was required as in many platform videogames the user has to get oriented using maps and has to quickly learn new scenarios, making relations between images seen on the screen (his/her point of view) and the map, with many similarities with a surveillance operator trying to understand in which room a camera is pointing. All participants reported to be right-handed, with normal hearing and normal or correct to normal vision.
To avoid learning effect each user tested only one condition.

6.2.4 Tasks

The controlled environment, that participants have to monitor, simulates a laboratory where there are individuals belonging to two roles: workers (with a green jacket) and rescuers (with a red jacket). The participants were asked to monitor the controlled environment looking in particular for some conditions and taking, as soon as possible, proper decisions and following actions.
The two conditions were:

(1) The proportion between workers and rescuers present in the environment should be fixed. There should be a rescuer for each two workers 2:1 (when workers are odd the subsequent even is considered). When some event changes the proportion (e.g. someone enters or leave the scene) and the rescuers are less than it should be, the operator should intervene balancing the scene. Through a button on the interface the user should call other rescuers specifying the exact number to reach the correct proportion. The interface ask also to specify the room where the rescuers should go, and the user has to choose the "random" room.

(2) The user has to look for unexpected events as (example was given in the task description):
- Intruder (not worker/rescuer);
- Fire;
- Accident/injury for some worker;
- Anything that may differ from the normal situation.

Even if the system may raise an alert the user has to fire an alarm (through a button on the interface) choosing the type from a list (fire, generic, injury, intruder). The user has also to call rescuers specifying the quantity and the room where they have to go (were the emergency is). The number of rescuers to be sent is related to the type of emergency:
- Generic: 1 rescuer;
- Intruder: 1 rescuer;
- Fire: 2 rescuers;
- Accident/injury: 3 rescuers.

6.2.5 Measures

A quali-quantitative approach has been used to measure various effects of different LOAs on users. Five different measures were taken during tests:

- Performances;
- Situation Awareness;
- Eye scanning pattern;
- Workload;
- Interface Usability.

The performances were evaluated considering the success rate (effectiveness) and the time (efficiency) spent by the user to accomplish to a certain task. The simulator was able tomeasure the difference between the time when a certain condition occurs and the time when the user performs an action and to evaluateif the chosen action was right, writing a log exported in csv format.

To determine if the tasks (for each event) were accomplished or not we considered the following criteria:

- a decision and a proper action must be taken in 35 seconds;
- the decision must be correct (i.e. the number of rescuers to be sent or the type of alarm)

We decided also to consider partially completed tasks. For example when a user recognize a lack of balance between workers and rescuers but sends a incorrect number of rescuers.

The user's Situation Awareness wasassessed through the combined use of a "freeze" technique and of a self-evaluation questionnaire.

The freeze technique is realized by obscuring the screen every 5 minutes (4 times in a test session) and asking to the user to draw, on a piece of paper representing the interface, the position of all the people detected in the videos. Users were asked to draw both on the map and on the camera bank, a symbol for workers and rescuers present on the scene.

The self-evaluation was made with a questionnaire the user had to fill for each freeze of the interface.

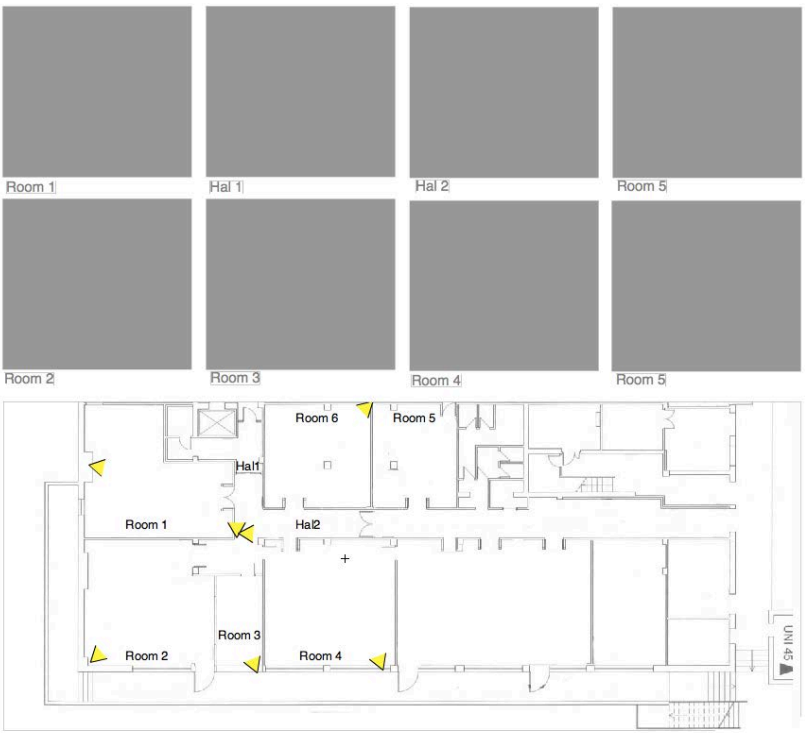

Figure 39 ***The schema used to assess the SA.***

A Tobii® ET17 remote eye-tracking system was used for recording ocular activity of participants. This system allows the collection of ocular data without using invasive and/or uncomfortable head-mounted instruments. It uses near infrared diodes to generate reflection patterns on the cornea of the eyes. A camera collects these reflection patterns, together with other visual information. Image processing algorithms identify relevant features, including the eyes and corneal reflection patterns. Three-dimensional position in space of each eyeball and the gaze point on the visual scene (2d) are then calculated. Sampling frequency was 30Hz. Fixations were then used to assess the time spent by the users looking at a certain area of the interface. From fixation data the scanning pattern (scanpath) was assessed. As reported by Camilli, Terenzi and Di Nocera [39]many works showed a relation between the scanpath and the mental workload of a user. Spatial statistic algorithms, used to study special patterns, could be used to assess the spatial distribution produced by a pattern of fixation. The authorsshowed a successful use of the Nearest Neighbour Index (NNI) to assess the scanpath. The NNI [30] is the ratio between the average of observed minimum distances between points and the random distance in a random distribution points (fixation points on the interface area, in this study). The NNI lies between 0 and 2.1491. Value major or equal to 1 indicates randomness while values below 1 indicates grouping. Visual scanning randomness (or entropy) was found to be related to workload. Other studies [23] showed that when the temporal demand is the most relevant factor contributing to the total workload the randomness increase. In other words, transitions of fixations between different areas of interest (AOIs) were reduced when mental workload was high, indicating attentional narrowing. NNI index was calculated using the ASTEF tool [22].
The workload was assessed also using the NASA-Task Load Index [76]. It is a multidimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales (mental demand, physical demand, temporal demand, effort, performance, and frustration). The assessment has been done through a questionnaire, easy and fast to be filled. Even if, as a subjective measure, could be affected by some context related variables, NASA-TLX is almost a de facto standard in Human Factors studies. The NASA-TLX questionnaire was given to the user at the end of the test.

Figure 40 ***The testing environment.***

Usability was assessed through a questionnaire. It focused mainly on the clearness of the interface and on its ability to support users during the tasks. It was very short to avoid overwhelming the participants. Moreover the users were asked to think aloud during all the test commenting their actions and expressing their thoughts about the interface and the tasks. The “thinking aloud” is a well-known technique to evaluate interfaces getting interesting qualitative data. The tests were recorded with a video camera at the back of the user to save his/her comments related to the actions on the interface.

6.2.6 Procedures

Before test session participants were trained on the interface performing sample actions (calling rescuers, firing alarms, zooming videos) and used it for 5 minutes (the same time interval used between each freeze). The participants were included in the sample only when they became able to perform all the operations needed for the tasks. Participants sat in front of the interface without other people in the room except for the facilitator and were asked to monitor the interface to accomplish the given tasks. The test lasted for 20 minutes and every 5 minutes the interface was masked and the user was asked to fill a SA questionnaire. At the end of the test participants compiled the NASA-TLX for the subjective assessment of mental workload and the questionnaire for the usability evaluation of the interface.

6.3 Data Analysis and Results

6.3.1 Performances

Performances were measured evaluating the correct completion of the tasks and the time needed. As already described, participants had 2 main task consisting of detecting and reacting in a certain way during the entire duration of the test (20 min). There are multiple occurrences of such events:

- 4 times there is a wrong balance between the workers and the rescuers that has to be compensated by the user;
- 2 times there are alarms that have to be signalled by the users. Moreover a proper action has to be done in relation to different kinds of alarms.

The sums of the proportions of tasks completed (6) were used as dependent variables in ANOVA statistic design using condition (Cond. 1 vs Cond.2 vs Cond. 3) as fixed factor. As can be seen from Figure 41there is statistical significance, $p<0.05, F(2,21)=5.2925$, among the different conditions.

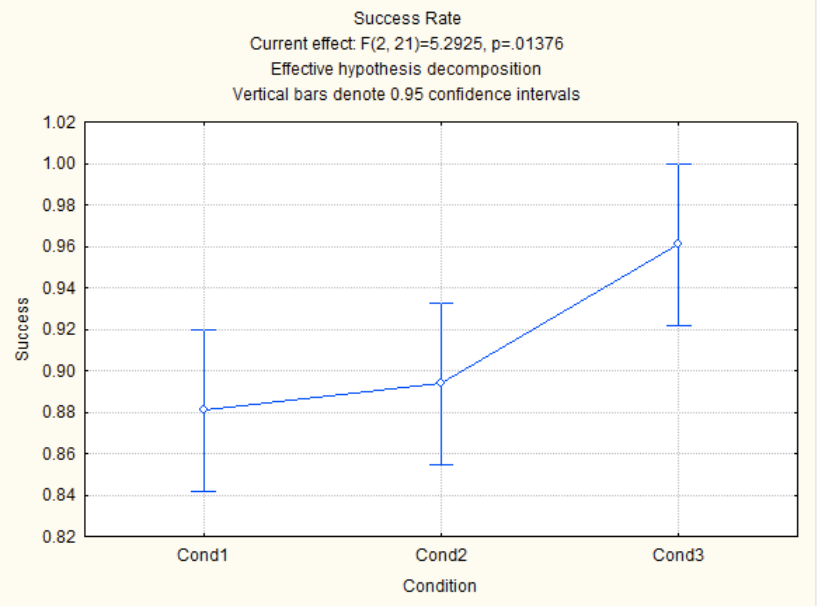

Figure 41 ***Tasks completion proportion among different conditions.***

Duncan post-hoc testing showed statistical significance ($p < 0.05$) between condition 1 and 3 and between condition 2 and 3. These results may be related to the fact that the higher LOA implemented in Condition 3 could better support users during the tasks.

To confirm this interpretation we could also consider qualitative evaluation made on two particular events related to the entrance of new workers. Indeed in one event a worker enters the scene just seconds after a fire starts in a room. In that moment all the attention of the user is focused on the fire loosing the control of the general situation. The users that failed in this task realized too late that in a room something changed, inferring that someone should have entered the lab. The other event is even more challenging. It happens as a consequence of the second alarm. A worker feels sick and slump on the floor. The user has to fire an alarm and send three rescuers to Room1. The rescuers take the workers away exiting the lab. In that moment the user is focused on these operation and, in the meantime has to realize that there is an incorrect balance between workers and rescuers.

While in condition 1 the user has to rely only on his/her perceptions and memory in condition 3 there is an evident cue given by the control panel, that blinks and clearly indicates the number of workers, rescuers and the number of rescuers missing. Moreover, for event 2, the user could also follow the worker entering by the new dot on the map.

The discussion about the time needed to complete a task (efficiency)Figure 42 is more complicated. Indeed, even if statistically significant, there is a small difference between mean values in different condition, not big enough to produce a real effect. Moreover difference between users among different condition is related to the different strategies adopted by the users. As emerged by the observation of the users during the test (registered by a video camera), they used different strategies according to the features of the interfaces proposed. For example in condition 3 some users, even if they noticed an event through the camera bank they check it on the information provided by the control panel, with an increase of the time but also with a decrease of the error rate.

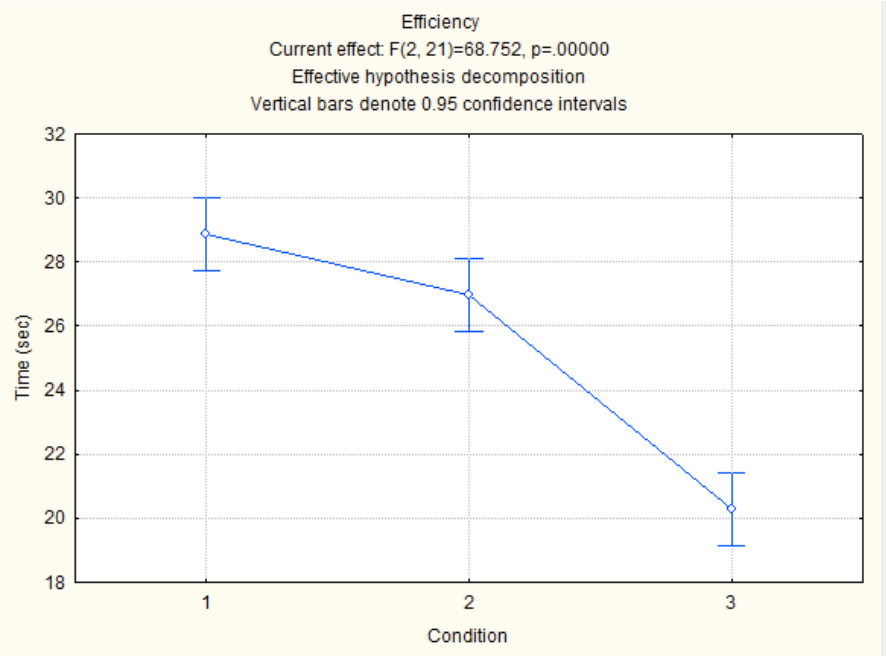

Figure 42 ***Time for completing the tasks.***

A qualitative evaluation made on the users' "thinking aloud" showed that users felt more confident about their choices in condition 3 as they could check their assumption with data reported in the control panel. This could have the double consequence, on one hand, of diminishing the stress level of the user, that can use a part of the system's situation awareness, on the other hand it could make the user rely more on the system, with a higher risk in case of system error.

6.3.2 Situation Awareness

Situation Awareness was assessed using the "freeze" technique. The interface was masked every 5 minutes (4 times during the whole test) and the user was asked to compile a questionnaire for each freeze. For each freeze we measured the number of details that the user were able to give then a proportion was calculated considering the maximum achievable level of detail.
The means of the SA values were used as dependent variables in ANOVA statistical design with the conditions as a fixed factor. The results are only tending toward statistical significance $F(2, 21)=2.6713$, $p=.09255$ Figure 43 .Duncan post-hoc test shown statistical significance between conditions 1 and 3 $p<.05$.

Figure 43 ***Situation Awareness assessment.***

The main difference noticed was of qualitative nature. In condition 3 the users were able to give more details with less uncertainty. They were asked to "think aloud" during the test and during the filling of the questionnaire. While in condition 1 and 2 they often tried to make inferences to remember the number and the position of the people in the scene, in condition 3 they were more self confident and fast in compiling the questionnaire. Indeed the time for compiling the questionnaire was compiled was measured. It could be not considered a reliable measure as it could be affected by subjective factors that weren't evaluated. However the higher self confidence could be a reason of the shorter time, even if without statistic significance, needed to fill the SA questionnaire (see Figure 44). A remark should be made for the second freeze. Indeed before the freeze a rescuers enters the lab and sits in room 6 in a zone that is only partially seen by the camera. In condition 1 and 2 almost half of the users had the "sensation" that someone has entered the scene but they weren't able to give other details. In condition 3 instead the map helps the user to correctly identify the position of the "hidden" rescuers. Moreover in condition 3 the users indicated the exact position of rescuers and workers, reproducing the dots on the map, while in the other condition they prefer to indicate only the number representing the sum of people in a room. Indeed, even for a trained user, it is not easy to translate the position of someone seen in a video on a map, while doing automatically this graphical representation could give at a glance a great quantity of information, helping also recognition and memory. Another remarkable condition is about the second alarm. In this case a worker feels sick and slumps on floor, disappearing from the scene because occluded by a table. Without any other cue some participants supposed that the worker exited the scene in a moment of distraction. In condition 2 and, especially, in condition 3 indications on the map and on the control panel (number of workers and a fall detection system) helped to correctly interpret the sudden absence of the worker.

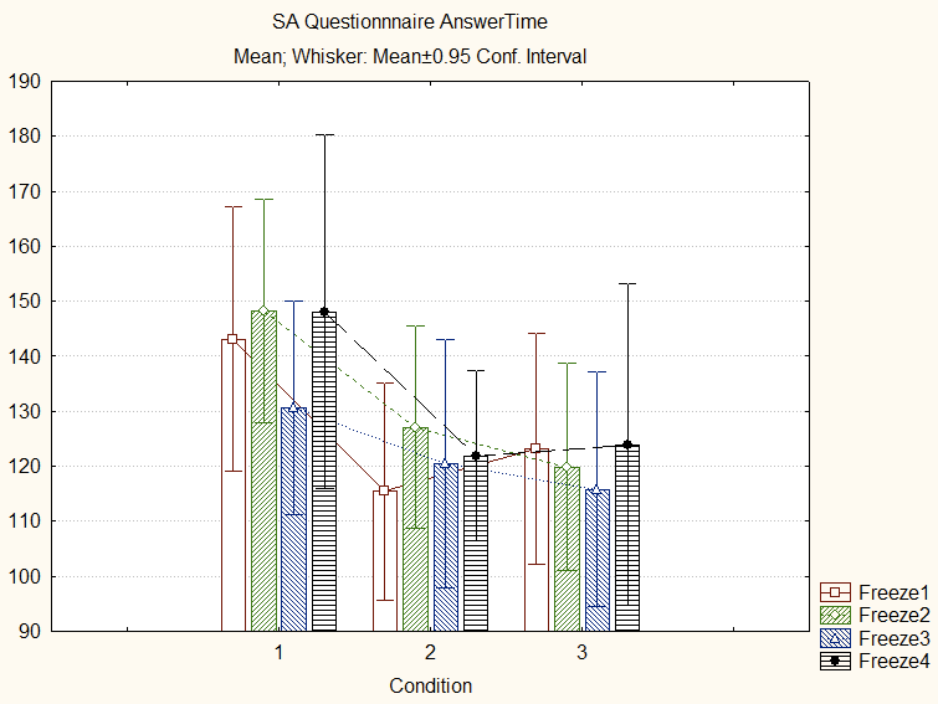

Figure 44 ***Time used for answer to SA questionnaire.***

6.3.4 Eye fixations

The ocular scanning behaviour of the user on the interface can be an important source of information. An interesting result is represented by the change in the area of the interface observed by the user. Fixations have been grouped into areas of interest (AOI) corresponding to the main elements of the interface:

- Camera bank (AOI 1).
- Map (AOI 2).
- Control panel (AOI 3).
- Video detail (AOI 4)

The means of fixation duration on each AOI for each minute have been calculated. This measure indicate the most used part of the interface and can highlight different strategies adopted by the users according to the tools and cues they could use. Of course this measure doesn't take into account the peripheral attention of the user that has shown to be attracted by sudden movements in video feeds or in the map or by alerts (blinking) in the control panel.

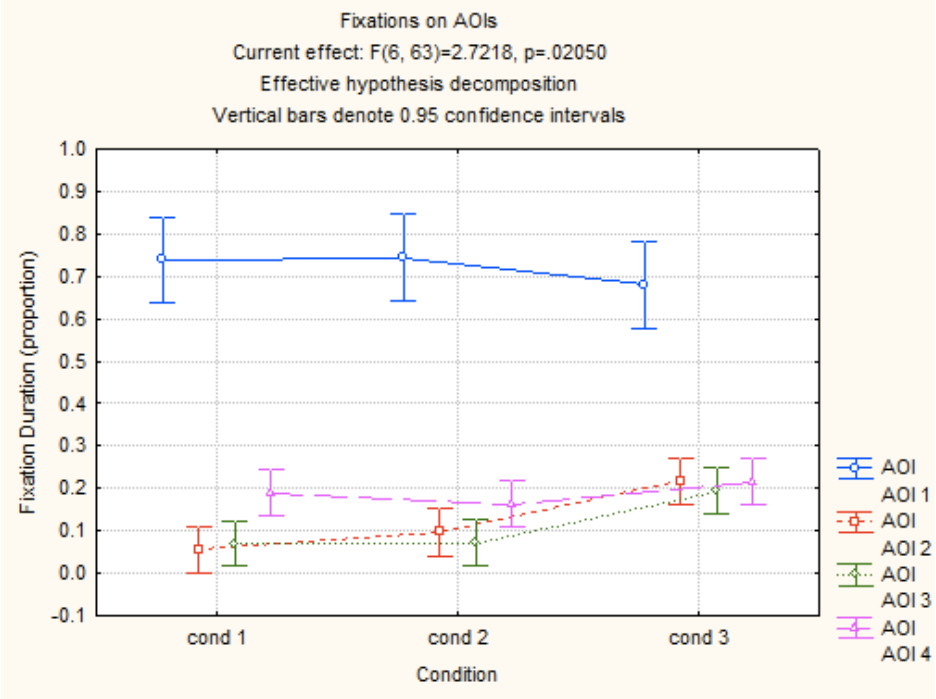

Figure 45: Fixation duration on different AOIs on the interface

Proportion of ocular fixations per minute on the interface AOI were used as dependent variable in ANOVA statistical design by using the Condition (Cond 1 vs Cond 2 vs Cond 3) as fixed factor. It highlighted a main effect, $p<.05$ $F(6,63)=2.7218$, of conditions on the user's scanpath. Duncan post-hoc test showed statistical significance between condition 1 and 3 for fixations duration on AOI 2(map) $p<.01$ and on AOI 3(control panel) $p<.05$. Another statistical significance has been found between condition 2 and condition 3 for fixations duration on AOI 2 $p<.05$ and on AOI 3 $p<.01$.
This could be related to the different use of these parts according to the different LOAs implemented. In conditions 2 and 3 map shows important information (presence and position of individuals) and in condition 3 the control panel shows the number of workers and rescuers and (eventually) the number of rescuers to call. This reflects the different strategies used by the participants to perform the assigned tasks.
This difference is evident looking at the heatmaps in Figure 46 representing the number of fixations on the interface's AOI in Condition 1 and in Condition 3. The red area indicates an higher concentration of fixation on map and on the control panel.

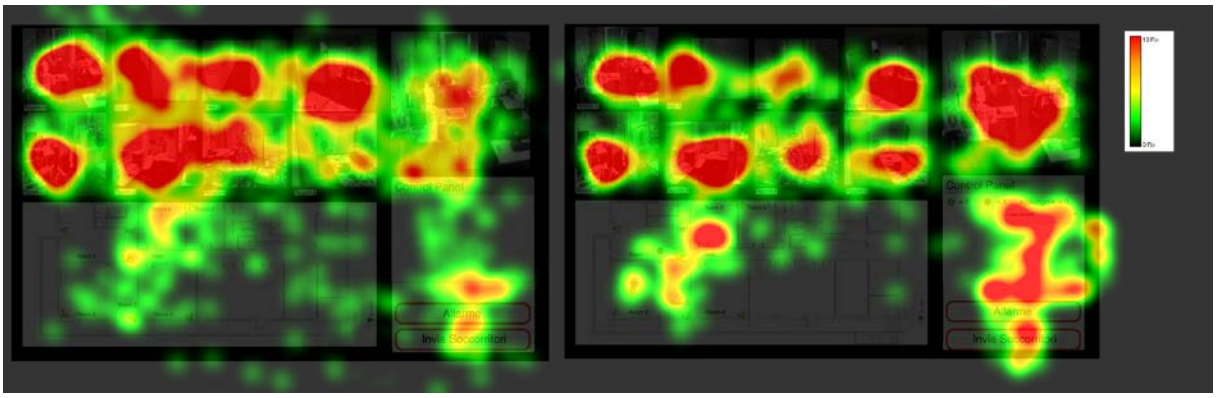

Figure 46 *The number of fixations* in condition 1(sx) and condition 3(dx).

In Figure 47are represented the sum of fixations duration for each minute of the test on AOI 2 (map).

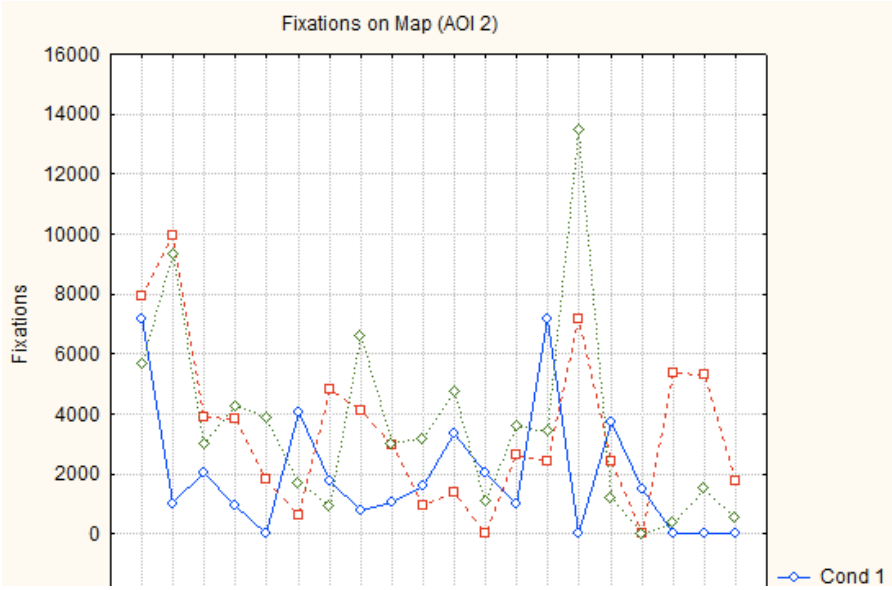

Figure 47 *Fixations per minutes on Map* (AOI 2)

Between minute 13 and 15 there is a peak in using map in condition 3. This event corresponds to a rescuer sitting in an area partially hidden to cameras. Almost all the users in every condition noticed that there is an hidden “presence” and, for condition 3 they used the map and the control panel (checking the number of people on the scene) to be sure of their suspect. In condition 1 and 2 this situation corresponds in a longer fixation duration caused by the clicking on the camera icons as from the “thinking aloud” emerged that this strategy was used by participants to have a better understanding of the correspondence between cameras FOV and the 2d space represented on the map.

6.3.5 Workload

NASA TLX questionnaire was used to assess the workload of participants. As expected, Mental Demand and Temporal Demand were the most influencing subscales concurring to the total workload assessment. Mental Demand could be defined as "The amount of mental and/or perceptual activity that was requires (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)" while the Temporal Demand could be defined as " The amount of pressure you felt due to the rate at which the task elements occurred. Was the task slow and leisurely or rapid and frantic?" [76].

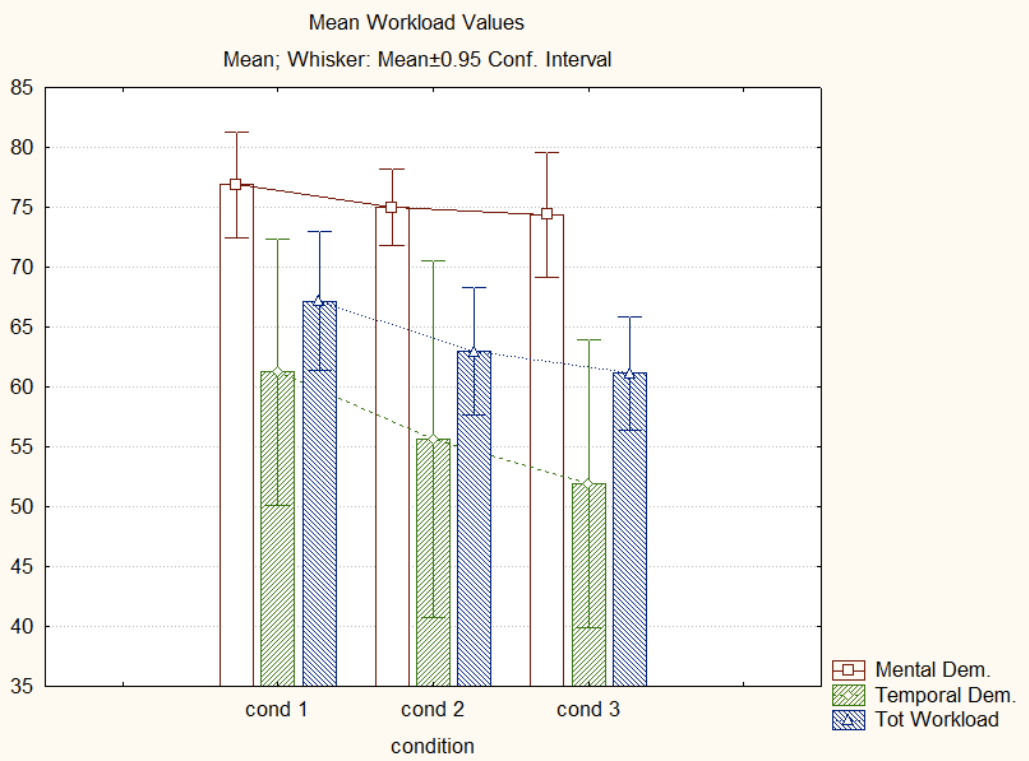

Figure 48 ***Values of Total Workload, Temporal Demand, Mental Demand for each condition.***

The values of Mental Demand, Temporal Demand and Total Workload assessment were used as dependent variables in ANOVA statistical design by using the Condition (Cond 1 vs Cond 2 vs Cond 3) as fixed factor.

There is no statistical significance of any variable in any condition however from the Figure 48 we could notice that the Mental Demand has quite no variations while the Temporal Demand has a little decrement thus there is a weak effect on the Total Workload. We could hypnotize that the Mental Demand is quite constant among conditions as the change is in the kind of task that the user should

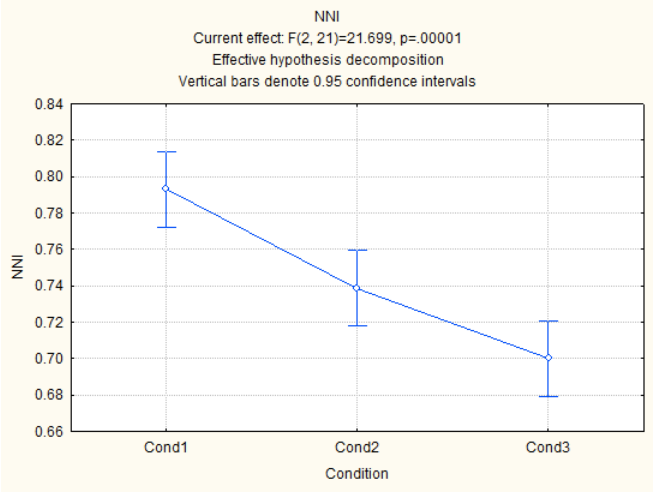

Figure 49 ***NNI in the three conditions.***

perform. Moreover even if in condition 2 and 3 the system gives more support it adds more information that have to be perceived and understood. The little difference in the Temporal Demand may be related to the higher support of the system that summarize information, giving to the user more time to accomplish the task.

Moreover also users' fixations could give information about the path followed by the user across the interface to have another indicator about workload .The spatial distribution of eye fixations was assessed calculating the NNI using ASTEF [22].

In the work of Camilli, Terenzi and Di Nocera [23], is shown that when Temporal Demand is the most loading component of the workload, the spatial distribution of eye fixation is more dispersed (i.e. more random, higher NNI values) respect to the fixations distribution recorded during easier task load conditions. Differently, when the most loading workload component is the Mental Demand, fixations spatial distribution is more grouped (i.e. less random, lower NNI values) respect to the fixations distribution associated with easier task load conditions.

The NNI mean values has been used as dependent variables in ANOVA statistical design using Conditions (Cond1 vs Cond2 vs Cond3) as fixed factor. The analysis showed a main effect, $p<.001$ $F(2,21)=21.699$, of condition on NNI. Duncan post-hoc test showed that the effect is present also within the conditions. The decreasing of NNI Figure 49 could be interpreted as a decrease of the Temporal Demand. Indeed, supported by the map and the control panel the users doesn't have to look around to be prompt for the arrival of new information or to have a comprehensive view of the scene but can focus on specific areas. This may confirm the assumption made observing the Workload assessment with the decrease of Temporal Demand.

6.3.6 Usability

The usability evaluation had the double objective to check if performances were eventually affected by poor interface usability and, as the layout of the interface was fixed, how the interaction with the interface changed when implementing different LOAs.
To assess the usability we asked the user to express their agreement to several assertions on a questionnaire using a 5-point Likert scale (where 1 is total disagreement and 5 total agreement). As the interface has no navigation the evaluation focused especially on labelling and on the intuitiveness and on the consistence of the interface. In all the conditions the average vote of these parameter were above 4 thus considered satisfying.
An interesting result was obtained asking users to express the perceived usefulness of the different UI elements on a 5 point Likert scale.

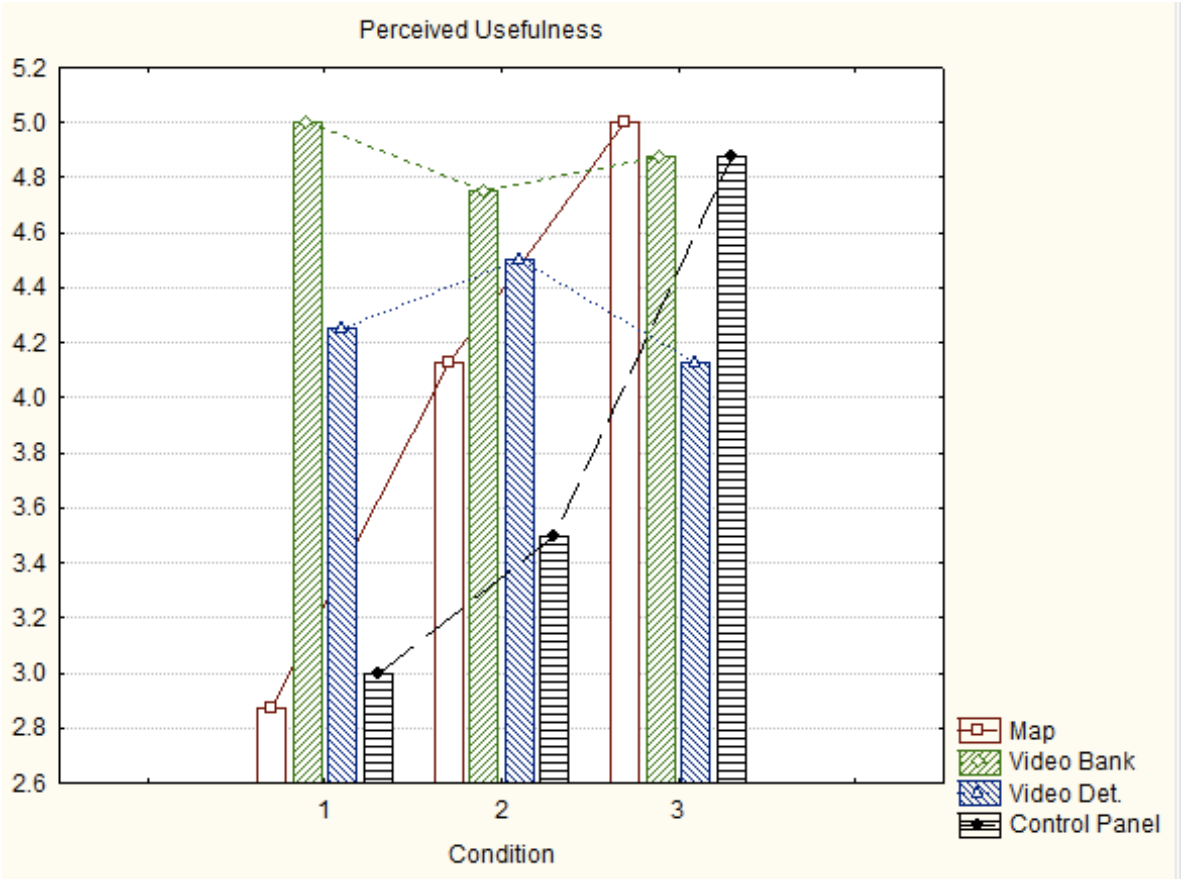

Figure 50 ***Perceived Utility of UI elements among different Conditions.***

The results, as visible in Figure 50, there is and increment perceived usefulness of the map and of the control panel. There is a strong increment on the map between conditions 1 and 2, 2 and 3 and 1 and 3. This is a reasonable effect of the implementation of new functions on the map that, especially in condition 3, becomes an important support to complete tasks. The effects on control panel are relevant mostly between condition 2 and 3. Even in this case the change is related to the implementation of new functions on this element that suddenly becomes one of the most important ones.

These results were expected and confirm previous assumptions and underlines how the different LOAs are able to deeply change the strategies adopted by users and their relation with the interface. The users were asked also if the quality and the size of the video feeds were adequate to properly complete the tasks. Even if there is no statistical significance in condition 1 the users complained about poor video quality and size while in condition 2 and 3 they judged it sufficient. This is confirmed also by the “thinking aloud” during the tests. Especially in condition 1, users complained about the contrast and the resolution of the video feeds.
There were two special conditions that made the users complain: (1) in room 6 the chairs were of a green similar to workers’ jacket, causing change blindness (2) there was a rescuer partially visible (a hand) but the detail was not so clear. A possible interpretation is that in condition 2 and 3 users rely also on other elements of the UI that are complementary and redundant, so the video, even if it is a primary input, can be integrated and temporary replaced by other information sources.

6.3.7 Discussion

Results showed that implementing different LOAs, without changing the interface layout, lead to the adoption of different strategies by the users, changing the action they perform to accomplish the given task. While the video feeds (camera bank and detail video) are the most used UI element in every condition, in condition 3 the map and the control panel, representing the visual output of the higher level of automation implemented, are used to have support to solve ambiguous cases, when information provided by video feeds were confused, due to many people moving though different rooms.
In condition 3 is observed an increment of maximum reaction time withan higher success rate.Indeed, observing users, we notice that they spend more time to look at cues provided by the interface (i.e. control panel) before taking their decision. Doing so users have a better confidence while choosing the option. Moreover in this conditions users add more details while answering to the Situation Awareness questionnaire. The visual cues, especially the blinking alarms, were able to attract the attention of the user. The most used strategy was to keep the control panel in the periphery of their attention thus, even if they miss some event, the cues in the control panel push them to look the video feeds and the map to verify if the event really occurred, following into an action.
Usually monitoring task are much longer so it very difficult to simulate the boredom.
In condition 1 users had to pay more attention to video feeds as even a little distraction may led to a loss of situation awareness (i.e. if a person disappears it could be exit or it could be in an area not covered by cameras). Moreover during this condition users complain about the dimension, contrast and resolution of video fees, underlining their tendency to rely only on that source. An interesting episode happened with two users that lost two subjects moving inside the scene. They keep searching for almost a minute to find where the persons were. In condition 2 there were several cases of loss of people but the cues provided by the map (highlighting the rooms with motion detected) helped to solve the ambiguity even if with a degree of uncertainty and effort of the user. In condition 3 the dots displaying people’s position on the map, the number and the list of workers/rescuers on the control panel was used as a reference when user lost the situation awareness looking only at video feeds.
There is no significant effect on Mental Workload as, probably, the new tools provided change the sub-tasks that the users have to accomplish. To a decrease of the memory and calculation demand corresponds and increase in interface complexity (more element to monitor). However a decrease in

Temporal Demand underlines how, increasing the cues provided to the users, the task becomes less frantic. Indeed the user can get the information he/she needs from specific cues on UI (in condition 3) without having to search among all video feeds (as in condition 1).
During the test on condition 3 the users showed to don't rely totally on cues provided by the system but they used them as a confirmation. As the choice was not pre-compiled but the interface provided only a cue the user felt like was their own responsibility to check the correctness of the choice. However this alone doesn't avoid the risk to accept a wrong cue, due to an incorrect calculation of the system.
As a synthesis from these results appears clearly that the main effect of different LOAs on user is the change of strategy used to accomplish the given task. The multiplication of information sources (redundant and complimentary) had effect on the quantity and quality of information that the users are able to achieve and remember, leading to a qualitative effect on the users performances.

Chapter 7

A proposal for a Smart Surveillance Natural User Interface

7.1 Designing Usable UIs

There are numerous studies regarding the design of user interface in control rooms or to support decision process. There is a complex relation between user interface levels and degrees of automation , levels of workload and operator performance.
The interface is the contact point between the system and the user, the gateway between the users intentions and the actions through the system's functionalities. It's evident how much an interface could influence user's performances. The ISO 9241-1 [47] defines Usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" where:

- effectiveness: the accuracy and completeness with which specified users can achieve specified goals in particular environments
- efficiency: the resources expended in relation to the accuracy and completeness of goals achieved
- satisfaction: the comfort and acceptability of the work system to its users and other people affected by its use.

This definition introduces three fundamentals independent variables in human computer interaction: the user, the task and the context of use. Usability is not an absolute concept but is related mainly to these three variables. The new version of the standard named "Ergonomics of Human System Interaction" introduces also general ergonomic principles which apply to the design of dialogues between humans and information systems: suitability for the task, suitability for learning, suitability for individualisation, conformity with user expectations, self descriptiveness, controllability, and error tolerance.
Even if expressed in few lines these concepts are very important in designing an interface for several key points:

- the interaction between a user and an interface is a dialogue. The user made itself a representation of the world, of the system and of the goal that he wants to achieve and performs and action. This is called by Norman[115]the “gulf of action”. Then the user perceives the (eventual) changes in the system’s state and gives it an interpretation according to the action performed and to his goal. This is called by Norman () the “gulf of evaluation”. Through a series of actions and evaluations the user has his dialogue with the system, adjusting his actions through constant feedbacks. An unexpected response of the system may cause a wrong assessment and a failure of the whole interaction.
- The interface maybe suitable for the task. As we have said using an artefact means changing the task and not the user’s abilities. So the new task may be properly designed taking into account the user’s goals, the user’s characteristics and the context.
- The interface maybe suitable for learning and has to be self-descriptive. Cooper[34], the father of Visual Basic, said that we are all intermediate users. In a short time we learn the basics of new interface and became quite good in performing average tasks. Then we can become expert users, using more fancy and hidden functionalities, but we tend to easily forget this knowledge as soon as we don’t use the interface. So the interface must support us to quickly learn the basics and become intermediate users but doesn’t have to rely upon our memory but be self explicative. As Nielsen [112] said “Recognition rather than recall”. So simple label names will help to recognize buttons, natural mapping (i.e. matching with reality of with processes like cause and effect) will help us to understand the system and to choose the correct action to perform to achieve our goals.
- The user must always be able to control the system, having constant feedback about what is happening (even when there is the “waiting” icon).
- The interface must prevent user errors and, in case of, be tolerant and allowing an undo (i.e. the recycle bin on the desktop).
- The Interface must be conform with user’s expectations. According to Cooper [34] expectations are relative not only to the primary (functional) goals, but they belong also to the emotional sphere. The user expects effectiveness, efficiency but also satisfaction of different level of goals. Indeed while the first goal may be merely functional (i.e. writing a document) the secondary goals may relate with the user experience (i.e. don’t waste time to look for function for text formatting, don’t feel that some function are too hard to be understood).

Even the aesthetical aspect of an interface may influence the satisfaction and the performance of the user. Norman found that a pleasant interface is able to change our mood and the way we perform a task. When we are stressed our brain is less capable of being creative and of evaluating alternatives, while when we are in a good mood our ability to concentrate increases.

7.2 The proposed Interface

This first interface designed for the testing of different LOAs was very simple and pointed more to functional aspects, to support the user in a video surveillance task according to the goals expressed in paragraph 4.1 which can be summarized in the main goal to keep a good level of situation awareness.

Based upon the results of the tests, and taking inspiration from several examples in literature, we designed also a proposal of interface that harvest the information obtained from the context capture system but that can also show how Ambient Intelligence paradigms could change the human computer interaction.

Gouin and Lavign [69] make a review of all the technologies that could be used in command and control (C2) room to help people interact with information in a more effective way. Indeed technologies could provide a huge amount of information but, to become a real advantage for the users, all these new data should be made easily accessible. The interface should be transparent and, at the same time, should help users to have a correct understanding of the information proposed and support decision-making and collaborative work. The authors show new types of human computer interfaces, from augmented reality to big multi-touch surfaces, highlighting how each one could bring an advantage in certain situation and operative scenarios. Also Iannizzotto and colleagues [82] recognize the need to overtake the old interfaces used for video surveillance (as joystick and keyboards) toward a more natural way to interact. Authors indeed propose a Perceptual User Interface that allows the user to interact with his/her bare hand on a surface, using gestures as commands.

From the literature is clear that the future doesn't consist in an unique interface able to satisfy all the users' needs. Instead a more plausible direction seems to have smart environments, filled with "transparent" interfaces, each one a little bit different from the others for scope or for the interaction paradigms used, but all working together to help the user achieve his/her goals.

With this in mind we decided to explore the concept of Ambient Display, described in the first part of this work, and decline it for a smart video surveillance system as a part of a larger smart control room. Indeed we are used to think to large, wall mounted screen as passive displays of information because the interaction could be difficult. In case of multi-touch display the user should stay close to the screen, losing the general view of the screen but also occluding the view to other observer. Moreover Gouin and Lavign[69] propose distance interaction trough gesture recognition. However, even if tracking technology is become more reliable, interaction through gestures could be tricky, especially if there is a need to indicate a point with a certain degree of accuracy. Moreover waving hands and arms could lead, in certain context to social acceptability problems.

Large displays are for their nature prone to be used by more than one user at once, to support collaborative decision making, and to contribute to keep a good level of Situation Awareness for all the team members. Endsley and Jones [53] have introduced the concept of "shared situational awareness" (SSA), considering that part of situational awareness of each member of a team operating on the same elements. Similarly, Klein [91] described the SSA as the level of environmental interpretation of events is common among all team members. According to these perspectives, the members of a team, despite having different objectives and responsibilities, must have a common understanding on the same elements of the situation. Other studies have developed the construct of "*distributed situation awareness*" (SA Distributed: DSA) to explain the situational awareness in the work of the teams (see, e.g. [128]). Generally, models consider the entire DSA collaborative system as the main unit of analysis and focus on the interaction and coordination between agents and sub-systems (see, e.g., [8]or [129]). In other words, these models consider all artifacts used by team members (e.g. computers, PDAs, paper documents and communication tools) as a resource to gain an awareness of the situation that is distributed among team members.

For these reason interaction with large displays should be limited to essential function that may harvest the capability of these devices to deliver a large quantity of information allowing a multi user interaction.

7.2.1 System Features

The interface designed, that we will call “Camera Wall”, is conceived as a part of a larger smart control room, and its role should be to provide information, alert the users attention on important events and allow multiple users to interact at a “coarse” level, to perform more complicated operations on other type of interface available in the room, as multi-touch panel, surface computers, tablet etc.
Indeed, as results of previous findings in this work, the “Camera Wall” implement some automated function related to the highlighting and filtering of information but leave the decision to the users.
A prototype of the system has been implemented to have a direct feedback of the design choices. For the prototype has been used the video dataset described in paragraph 3.1 .
The prototype was developed inside the Usability and Accessibility Lab of C.A.T.T.I.D., Sapienza University of Rome.
In the prototype we implemented only “live” functionalities without implementing searching among recorded video as it require a lower level of attention and situation awareness.
The main functions we thought for the Camera Wall are:

- view of all the video feeds;
- control of PTZ cameras;
- view of the location of individuals through a map;
- identification of individuals in the scene;
- setting of some basic rules (i.e. geofences).

Other functions could be added to the system according to specific case of use even if the criterion to follow should be to keep the interface simple, adding only the essential functions.
The Camera Wall is divided into three parts (as can be seen in Figure 51):

- a display of all the video streams acquired by cameras (called Camera Bank area);
- an interactive map where different kind of information could be visualized and accessed (MAP area);
- a text based report with information about general state of the systems, alarms etc (Message Board area);
- an area to see the enlarged video streams and other widgets with different kind of information (Detail area).

Figure 51The Interface of the "Camera Wall".

The dimension of the screen must be related to the number of video streams to be visualized and to the dimension of the map. For the user is important to see the video in a good quality and in a proper dimension because, as seen in paragraph 4.2.2 , the lack of these elements may lead to an excessive workload and to poor attention levels.
Implementing the prototype we decided to use a short-throw projector (60” projected area 1800x1600 pixels) In the prototype we used 8 video feeds from the dataset over mentioned representing different scenarios with different kind of activities. In a real scenario with more cameras the interface could be extended horizontally eventually adding projectors or screens.
Even if the system is able to automatically identify unexpected events the decision to show complete, the unfiltered view of all the video streams is useful to prevent errors deriving from system’s false-negative. It could be seen as a backup solution that allows to use it as a “traditional” video surveillance system, with the user choosing the video stream to focus and enlarge.

7.2.2 Interaction Style

To interact with the interface we choose a touchless paradigm using the Nintendo’s WiiMotetacking inspiration from the work of Bellucci and colleagues [11]. The hypothesis of interacting with a touch screen was not taken into account for the reasons already described and in orderto support the joint work of several users. The WiiMote is the controller of the game console Wii made by Nintendo. Despite it is usually seen under its leisure aspect the controller offers a variety of multimodal I/O

functionalities. As there are not official datasheets we refer to the work of Lee [98].The main features of the WiiMote that allows the development of cost effective multimodal, natural interfaces are:

- **infrared camera with object tracking:** at the top of the controller is an IR camera sensor with a resolution of 1,024 × 768 pixels, a 100 Hz refresh rate, and a 45 degree horizontal field of view. The camera chip integrates a multiobject tracking (MOT) engine, which provides high-resolution, high-speed tracking of up to four simultaneous IR light sources, transmitting data about position and size of tracked lights. This feature allows the easy development of camera-based IR tracking applications and the high refresh rate enable the tracking of fast movements.
- **Accelerometers:** the device mounts 3-axis linear accelerometer with a +/−3 g sensitivity range, 8 bits per axis, and a 100 Hz update rate.
- **Buttons:**the WiiMotes has 12 buttons. Four are arranged in a standard directionalpad layout. One button is on the bottom providing a trigger-like affordance for the index finger. The remaining seven buttons are intended to be used by the thumb. The remote design is symmetric, allowing use in either the left or right hand.
- **Vibration motor (haptic feedback):**a small vibration motor provides haptic feedback. Even if it has only a binary control (on/off) the intensity of the feedback could be controlled with PWM techniques.
- **Speaker (auditory feedback):**a small speaker in the remote's center supports in-game sound effects and user feedback. The audio data streams directly from the host with 4-bit, 4 KHz sound similar in quality to a telephonebut the overall volume and quality of sound pose limits to the use of this feature in certain context.

Moreover the device communicates with a Bluetooth connection being recognized as an Human Interface Device thus allowing an easy connection with computers.

In the absence of an official SDK the prototype was developed using WiiMoteLib a .NET managed library for using a Nintendo Wii Remote (WiiMote) and extension controllers from a .NET application. The interface was developed using Adobe Air (ActionScript3) that allows a fast prototyping of UI. A C# .NET application acts a gateway between WiiMote(s)and the Airapplication, allowing the latter to receive data from the controller and sending back commands (as the activation of the vibration motor). The C# .NET and the Air exchange data using a custom xml protocol over a socket connection. Under the projected area is placed the Nintendo "sensor bar". Despite the name it doesn't contain any sensor but is a plastic bar of 20cm containing 10 IR leds, 5 on each end. The WiiMote track these two ends as two blobs. As the distance between the led on the bar is known the library used is able to calculate both distance and, with a proper configuration, the *x,y* coordinates the controller is pointing on the screen.

The WiiMote is used as a pointer over the big projected screen. Users can point directly on the map without standing in a fixed position. When the controller is pointed toward the screen a rounded cursor appears. Using different colours for the cursor allows to multiple users to interact simultaneously. As the cursor enter in an interactive area as a button or a video the controller vibrates briefly, giving the user a haptic feedback. This is useful to avoid unintentional interaction. This kind of feedback is very important as give more consistence to the physical metaphor used.

Pointing the WiiMote and holding the A button on a sensible point a pop-up pie menu[21]showing the different choices available will appearFigure56. The pie menu is widely used in touch and touchless

interfaces instead of linear based menu because (1) all the choices are at the same distance from the initial cursor's position (2)can offer a bigger target size "target" space, and according to Fitt's law[60],[100] these two characteristics have a positive effect on the ability of the user to point a target. This is true especially when using input devices that give a lower pointing precision [62].
Moreover as in touch and touchless interfaces there is the coincidence between input and output device a pie menu can be used as a kind of "target" or "magnifying lens" giving the sensation of the exact correspondence of the menu with the selected area. When an item in the pie menu has a submenu it appears as another circle, external to the first one. This gives the benefit that all the choices of the previous submenus are available allowing a fast and easy correction of selection mistakes.

Figure 52 ***The WiiMote controller*****: front, side and rear view.**

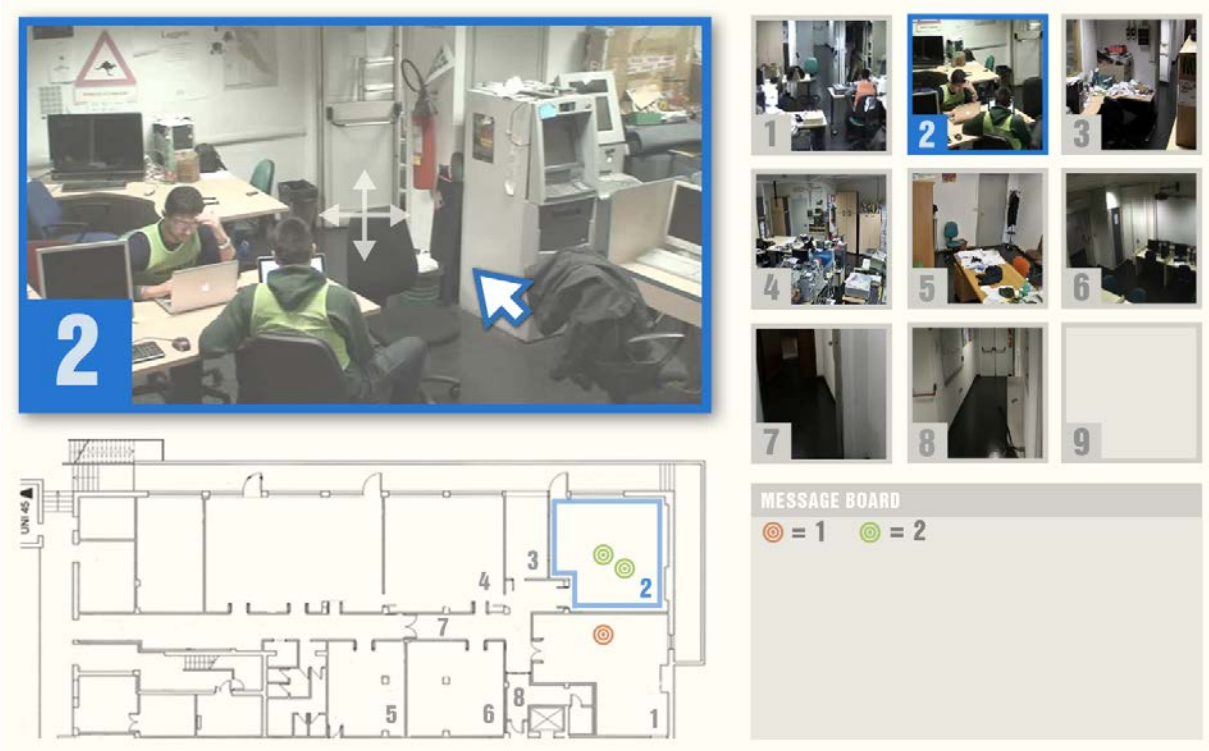

Figure 53 ***Camera Wall interface*****. The used can select a video feed and drag it in the detail area.**

In the Camera Bank area there are all the video feeds available. The user can select a video feed that will be enlarged in the Detail area. As seen Figure 53as the user clicks (with the A button) on a video feed it automatically enlarges and becomes draggable, a semi-transparent super-imposed layer appears to give feedback of the draggable status. The video could be dragged holding the A button, as it is released the video stops to be draggable and the super-imposed layer disappears. When a video feed is selected the corresponding box in the Camera Bank and the room the feed belongs to, are highlighted to give a feedback to the user.
The user could resize the video clicking on the bottom-right corner of the image or using the + and – button on the WiiMote control. In case of PTZ cameras + and – buttons cold be used for zooming and the directional pad could be used to move the camera.

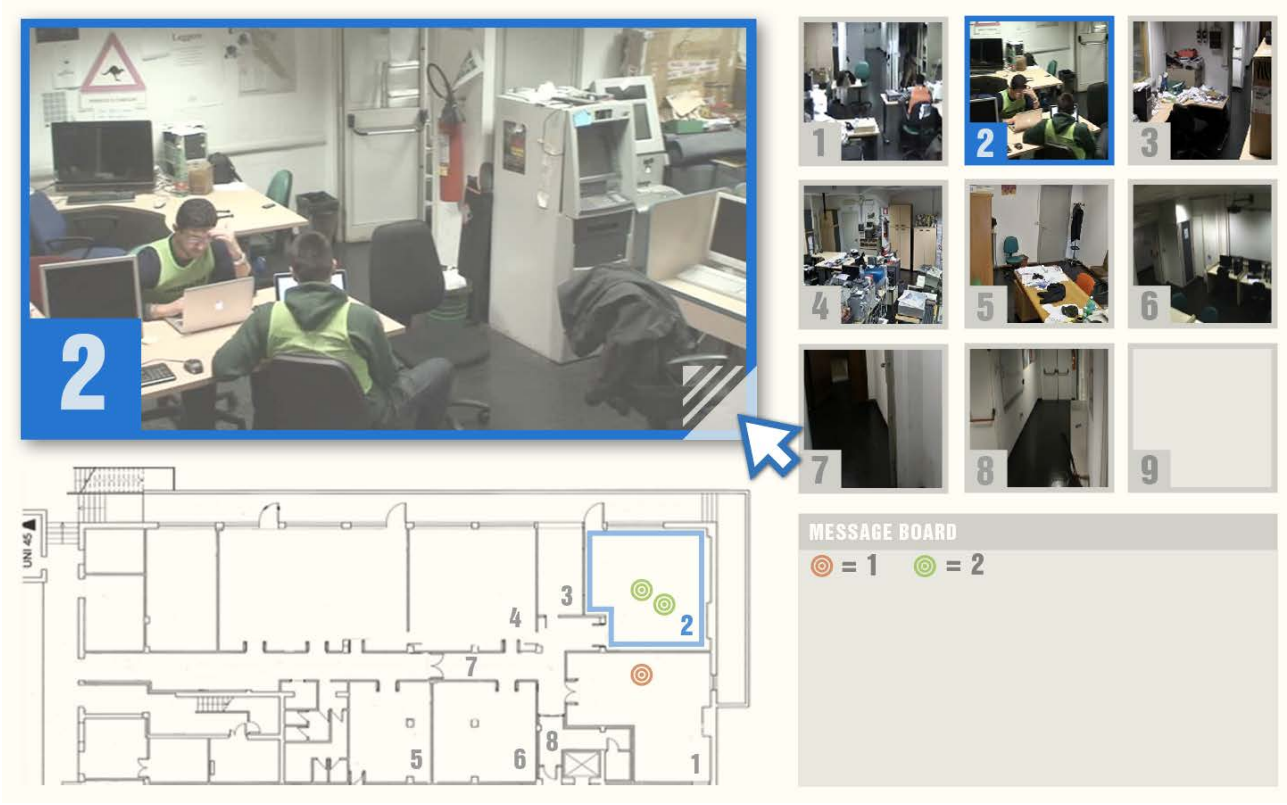

Figure 54 ***The user can resize the video using the bottom-right corner.***

The map visualizes the output of a RTL system. The dots represent people present in the observed environment. As in scenario represented in the video dataset used for the prototype there are two type of people, workers and rescuers, dressed with green and red jackets, the dots give, through colour, information about the role of the individuals localized. Pointing a dot and pressing the A button will make a widget appear with the information about the individual selected Figure 55.

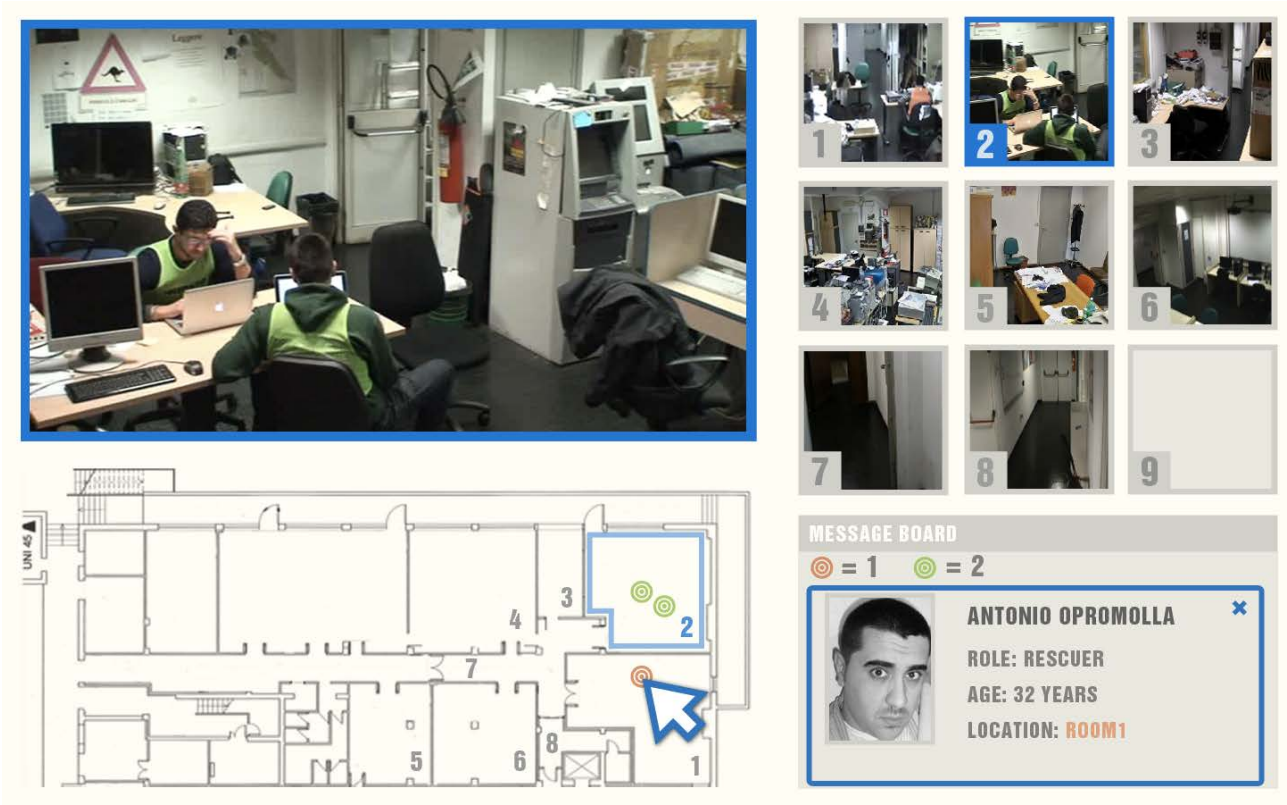

Figure 55 ***Information about individuals*** **are shown by clicking on the correspondent dot on the interface.**

Moreover on the map could be displayed other placemarks that may be used to access to other information related to a specific place eventually provided by a context capture system (e.g. temperature, pressure, light). Clicking on a room (or a hallway) the system shows in the detail area the best view of the selected place (if any), the perimeter of the area is highlighted to give the user an evident feedback of the room the video feed belongs to.

If the user holds the A button on the map, a pie menu appears, allowing to set some easy rule. We decide to implement only a geofencing rule, to allow or deny to a certain category of individuals the access to a certain area. The rationale is that this function could be use for a fast and coarse action while more detailed and articulated rule could be applied using other interface present in the smart room. As seen in Figure 56 the pie menu has two initial options: deny and allow. After one is selected another ring appears. As we have only two options (rescuers and workers) we decided to use a semicircle that is placed near the selected option. When the user select an option the menu disappears and he/she can move the pointer on the map to select the regions on which the rule should be applied. An overlay appears on the area to give as a feedback of the rule applied.

The message board area could be used to display information that may be the result of some automated filtering or highlighting. It could be used also to display personal information of individuals in the observed environment. Indeed, when a information appears in this area, it becomes draggable, becoming a widget around the interface. However we prefer to keep it in this corner to avoid occlusion of video feeds.

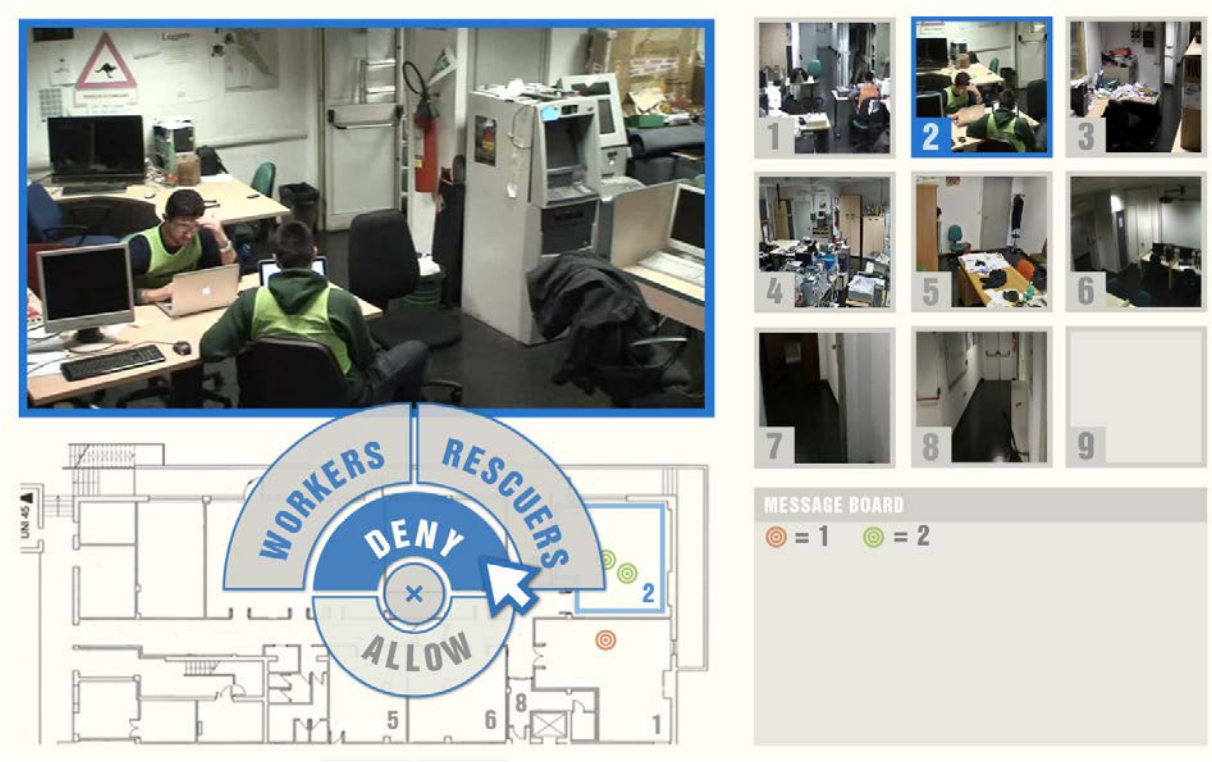

Figure 56 ***A pie menu is used to apply easy rules.***

7.3 Test

A test was made to assess the overall usability of the system using the described prototype. Indeed, as it is meant to part of a smart room, with other systems, contributing to the workload and to the team situation awareness, other measures could not be taken, as they would be partial or unreliable, due to the absence of a correct ecological context. A small sample of participant was used as the main interest was to get qualitative data about the user experience .

Participants

Seven participants (3 females; mean age = 26.77 years; SD = 3.01 years) volunteered in this test. All participants reported to be right- handed, with normal or correct to normal vision. The participants don't have prior experience in using a video surveillance system. At the beginning of the test, users were asked to fill a questionnaire where they reported about their use of computer and, particularly the game console Nintendo Wii. The latter implies a prior knowledge of the device used to interact with the system. For this reason we admit to test only participant with a minimum experience with the game console, to skip the training about the fundamental use of the device.

Tasks

The participants were asked to perform two simple tasks:

- During the first ten minutes of the test they have to follow a person moving through the observed area by selecting the most appropriate view;
- When a fire occurs they have to deny the access to the workers to the area.

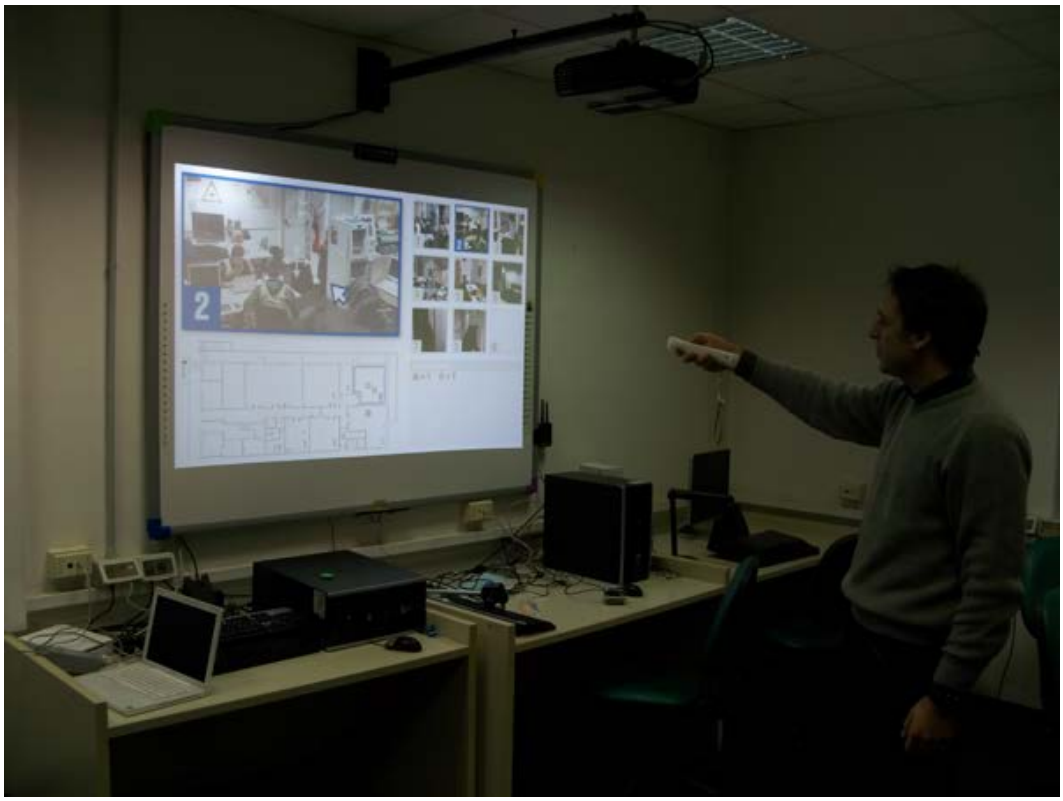

Figure 57 ***A user during test.***

Measures

Performances were assessed measuring the success rate accomplishing tasks. We considered also partially completed task when only a part of it was completed. The users were asked to "think aloud", describing their action and their thoughts about the interface, to provide qualitative information about their experience.

To assess the perceived usability of the system, the Us.E.questionnaire[38] was adapted to fit the specific context of use. Although this tool was originally developed for the evaluation of web sites, the items are sufficiently generalizable to assess a generic framework. The questionnaire consists of 19 statements to which participants had to report on their level of agreement / disagreement by using a five-point Likert scale. A recent study by Di Nocera and colleagues[40]showed the validity and reliability of Us.E. with respect to three dimensions of usability: Handling, Satisfaction and Attractiveness. Handling refers to simplicity of use and, in general, to the interaction with the "structural properties" of the technology. Satisfaction refers to the perceived users' satisfaction. This

factor can also be named “Perceived Utility”, since several contributing items describe the achievement of goals using technology. Attractiveness refers to the aesthetics features of the technology (e.g., use of colours and pictures); in this case study, this factor was strongly related to the overall appeal of the GUI.

Procedure

Participants received a 5 minutes training prior experimentation and were included in the sample only when they showed to be able to use the main functions. Participants stand in front of the projected interface, at 2.5 meters. They were left alone in the room except for the facilitator that is in charge to explain the task to the user.

7.3.1 Results

The success rate in executing the task was 85%. The problems occurred mostly when the users have to switch across different video feeds to follow someone in the scene. Moreover some problems occurred doing to pointing inaccuracy that lead to problems in using the pie-menu.
The Us.E questionnaire showed an overall positive resultFigure 58. All the values are over the middle point (2.5). The most appreciated factor is the Attractiveness. The interaction is engaging, far from the interface one would expect for a command and control system. In their comments the users reported that the large screen gives always the sensation to have the control of the situation having, at a glance, many information. The satisfaction (perceived utility) level is not very high as many users felt the need of having extra functions, as an automatic people tracking that dynamically show the best view on the tracked individual. This is due mainly to the prototypical state of the system, to the absence of a smart room environment that could support more elaborate functions but is related also to the explicit will to keep the interface as simple as possible.
Handiness was affected mainly by the accuracy of the pointing system (WiiMote controller). It resulted to have scattered behaviour in certain moments thus the users had problems in performing fine movements. However the problem could be solved at applicative level, applying some smoothing algorithms and snapping to UI elements.

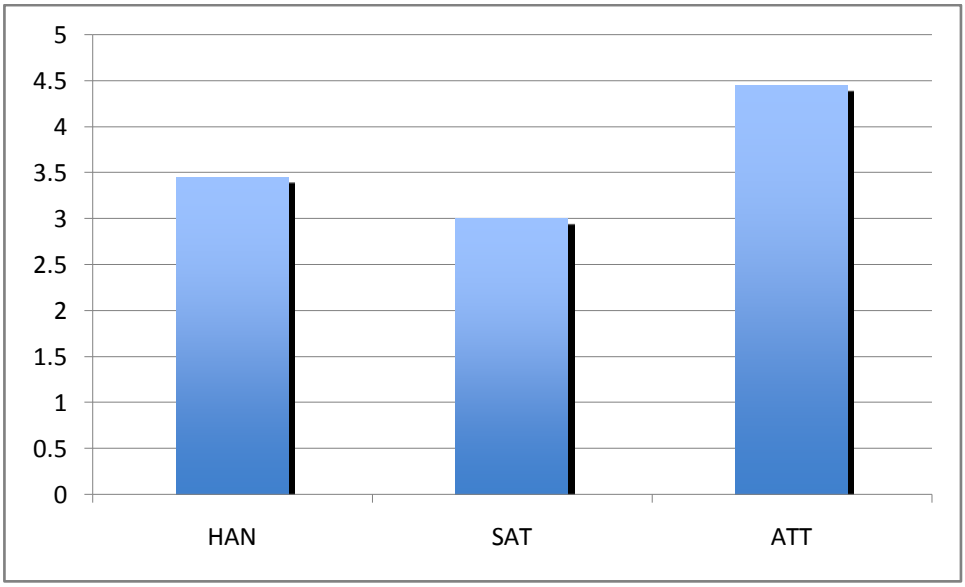

Figure 58 ***Usability assessment with Us.E. questionnaire*** **HAN= handiness, SAT= satisfaction, ATT=attractiveness.**

7.3 Discussion

The system described is only one of the possible interfaces that can constitute a smart room equipped for video surveillance. In it some AmI paradigms have been applied both in system and in user interface design, to propose a naïf way to interact with a video surveillance system with the aim to support collaborative work, decision making and shared/distributed situational awareness. Indeed the natural interaction paradigm and the LOA implemented could lead not only to a better efficiency and effectiveness but could also engage the user in a more attractive interaction (as emerged by usability tests) that might contrast the boredom and the attention degradation proper of video surveillance tasks.

Chapter 8

Concluding Remarks

In this dissertation we analysed different aspects of an Ambient Intelligence system. At a low technological level we propose a system based on sensor fusion techniques, exploring how the combined use of heterogeneous sensors could lead to better performance and especially to a better context understanding, related to the ability, thank to more detailed information, to make high level inferences. Context capture and understanding is recognized as a cornerstone of AmI, a fundamental source of information to allow to any AmI system to respond in an "intelligent" way to specific situations and to user's intentions.

But technology itself is not sufficient to bring a real advantage for people; indeed there is the need of and human-centered ambient intelligence design, which puts the final users (intended with all his/her complexity) at the centre of the entire design process.

As Ambient Intelligence intervene in many human activities by automating, through technology, certain tasks, is important to correctly choose the roles assigned to the system and to the user. Indeed the automation could affect different aspect of the task (i.e. perception, understanding, decision, action) leaving to the user different level of control on the system and on the situation. For this reason a special attention has been given to the different levels of automation that could be implemented in a system. Particularly through experiments made on a smart surveillance system we assess how different LOAs could change performances, Situation Awareness and workload. Results showed that the higher level of automation has a qualitative effect on users performance, with the result to change his/her behaviour and his/her attitude to the system. Results showed also how, through a correct design is possible to support users in certain tasks leaving them a certain level of control, to reduce the workload but, at the same time, avoiding "out of the loop" effects.

Moreover, evolving from results achieved, a Natural User Interface for a smart control room has been proposed. This user interface explores the concept of Ambient Displays showing how, not only LOAs are important for an effective user empowerment, but, according to AmI characteristics, is also important to design usable systems that considers not a single device but the entire "smart" environment.

As AmI is a very wide, multidisciplinary and continuously evolving field of study this dissertation should be read as a constant work in progress discourse about the relation between humans and technologies inside the AmI framework. Moreover methods and techniques used in the present work

could be generalized and hopefully contribute to the body of HCI and Human Factors research on Ambient Intelligence paradigms.

Bibliography

[1] Aarts, E. & de Ruyter, B. 2009, "New research perspectives on Ambient Intelligence", *Journal of Ambient Intelligence and Smart Environments,* vol. 1, no. 1, pp. 5-14.

[2] Aarts, E. & Marzano, S. 2003, *The new everyday view on ambient intelligence,* Uitgeverij 010 Publishers.

[3] Aarts, E.H.L. & Aarts, E. 2009, *True visions: The emergence of ambient intelligence,* Springer Verlag.

[4] Adelstein, F., Gupta, S.K.S., Richard, G. & Schwiebert, L. 2005, *Fundamentals of mobile and pervasive computing,* McGraw-Hill.

[5] Alvarez, G.A.& Franconeri, S.L. 2007, "How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism", *Journal of Vision,* vol. 7, no. 13.

[6] Anne, M., Crowley, J.L., Devin, V. & Privat, G. 2005, "Localisation intra-bâtiment multi-technologies: RFID, Wifi et vision", *Proceedings of the 2nd French-speaking conference on Mobility and ubiquity computing*ACM, , pp. 29.

[7] Ark, W.S.& Selker, T. 1999, "A look at human interaction with pervasive computers", *IBM Systems Journal,* vol. 38, no. 4, pp. 504-507.

[8] Artman, H.& Garbis, C. 1998, "Situation awareness as distributed cognition", *Proceedings of ECCE*.

[9] Ashton, K. 2009, "That 'Internet of Things' Thing", *RFID Journal,* www.rfidjournal.com/article/print/4986, .

[10] Augusto, J.C. 2007, "Ambient Intelligence: The Confluence of Ubiquitous/Pervasive Computing and Artificial Intelligence", *Intelligent Computing Everywhere,*, pp. 213-234.

[11] Bellucci, A., Malizia, A., Diaz, P. & Aedo, I. 2010, "Don't touch me: multi-user annotations on a map in large display environments", *Proceedings of the International Conference on Advanced Visual Interfaces*ACM, , pp. 391.

[12] Bick, M. & Kummer, T.F. 2008, "Ambient intelligence and ubiquitous computing", *Handbook on Information Technologies for Education and Training,* , pp. 79-100.

[13] Billings, C.E. 1997, *Aviation automation: The search for a human-centered approach,* Lawrence Erlbaum Associates.

[14] Bose, R. 2009, "Sensor networks motes, smart spaces, and beyond", *Pervasive Computing, IEEE,* vol. 8, no. 3, pp. 84-90.

[15] Botterman, M., (2009) Internet of Things: an early reality of the Future Internet. Workshop.

[16] Report, European Commission Information Society and Media.

[17] Boukraa, M. & Ando, S. 2002, "A computer vision system for knowledge-based 3D scene analysis using radio-frequency tags", *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*IEEE, , pp. 245.

[18] Boukraa, M. & Ando, S. 2002, "Tag-rased vision: assisting 3D scene analysis with radio-frequency tags", *Information Fusion, 2002. Proceedings of the Fifth International Conference on*IEEE, , pp. 412.

[19] Butz, A. 2010, "User Interfaces and HCI for Ambient Intelligence and Smart Environments", *Handbook of Ambient Intelligence and Smart Environments,*, pp. 535-558.

[20] By, E., WIENER, L. & RENWICK, E.C. 1980, "Flight-deck automation: Promises and problems", *Ergonomics,* vol. 23, no. 10, pp. 995-1011.

[21] Callahan, J., Hopkins, D., Weiser, M. & Shneiderman, B. 1988, "An empirical comparison of pie vs. linear menus", *Proceedings of the SIGCHI conference on Human factors in computing systems*ACM, , pp. 95.

[22] Camilli, M., Nacchia, R., Terenzi, M. & Di Nocera, F. 2008, "ASTEF: A simple tool for examining fixations", *Behavior research methods,* vol. 40, no. 2, pp. 373-382.

[23] Camilli, M., Terenzi, M. & Di Nocera, F. 2008, "Effects of Temporal and Spatial Demands on the Distribution of Eye Fixations", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*SAGE Publications, , pp. 1248.

[24] Cattoni, A., Dore, A. & Regazzoni, C. 2007, "Video-radio fusion approach for target tracking in smart spaces", *Information Fusion, 2007 10th International Conference on*IEEE, , pp. 1.

[25] CASAGRAS - Coordination and Support Action (CSA) for Global RFID-related Activities and Standardisation, 2009 Final report, Available from http://www.rfidglobal.eu

[26] Cavanagh, P. & Alvarez, G.A. 2005, "Tracking multiple targets with multifocal attention", *Trends in cognitive sciences,* vol. 9, no. 7, pp. 349-354.

[27] Ceipidor, U.B., Dibitonto, M., D'Ascenzo, L. & Medaglia, C.M. 2010, "Localization Issues in a ZigBee Based Internet of Things Scenario", *The Internet of Things,* , pp. 157-165.

[28] Cerrada, C., Salamanca, S., Pérez, E., Cerrada, J. & Abad, I. 2007, "Fusion of 3D vision techniques and RFID technology for object recognition in complex scenes", *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*IEEE, , pp. 1.

[29] Chae, H.& Han, K. 2005, "*Combination of RFID and vision for mobile robot localization*", *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on*IEEE, , pp. 75.

[30] Clark, P.J.& Evans, F.C. 1954, "*Distance to nearest neighbor as a measure of spatial relationships in populations*", *Ecology,* vol. 35, no. 4, pp. 445-453.

[31] Cohen, P.R. & McGee, D.R. 2004, "Tangible multimodal interfaces for safety-critical applications", *Communications of the ACM,* vol. 47, no. 1, pp. 41-46.

[32] Collins, R.T., Lipton, A.J., Fujiyoshi, H. & Kanade, T. 2001, "Algorithms for cooperative multisensor surveillance", *Proceedings of the IEEE,* vol. 89, no. 10, pp. 1456-1477.

[33] Cook, D.J., Augusto, J.C. & Jakkula, V.R. 2009, "*Ambient intelligence: Technologies, applications, and opportunities*", *Pervasive and Mobile Computing,* vol. 5, no. 4, pp. 277-298.

[34] Cooper, A., Reimann, R. & Cronin, D. 2007, *About face 3: the essentials of interaction design,* Wiley-India.

[35] Coutaz, J., Crowley, J.L., Dobson, S. & Garlan, D. 2005, "Context is key", *Communications of the ACM,* vol. 48, no. 3, pp. 49-53.

[36] Dee, H.M.& Velastin, S.A. 2008, "How close are we to solving the problem of automated visual surveillance?", Machine Vision and Applications, vol. 19, no. 5, pp. 329-343.

[37] Dey, A.K. 2001, "Understanding and using context", *Personal and ubiquitous computing,* vol. 5, no. 1, pp. 4-7.

[38] Di Nocera, F., Ferlazzo, F. & Renzi, P. 2003, "L'usabilità a quattro dimensioni", *Ricerche di Psicologia,* vol. 26, no. 4, pp. 83-104.

[39] Di Nocera, F., Terenzi, M. & Camilli, M. 2006, "Another look at scanpath: distance to nearest neighbour as a measure of mental workload", *Developments in human factors in transportation, design, and evaluation,* , pp. 295-303.

[40] Di Nocera, F., Terenzi, M. & Ferlazzo, F. 2007, "Misurare l'Information Handling: un contributo sperimentale sulla sensibilità della Scala" Maneggevolezza" di Us. E. 1.0", *Giornale italiano di psicologia,* vol. 34, no. 1, pp. 223-236.

[41] Dibitonto, M., Buonaiuto, A., Marcialis, G., Muntoni, D., Medaglia, C. & Roli, F. 2011, "Fusion of Radio and Video Localization for People Tracking", *Ambient Intelligence,*, pp. 258-263.

[42] D'Orazio, T., Leo, M., Guaragnella, C. & Distante, A. 2007, "A visual approach for driver inattention detection", *Pattern Recognition,* vol. 40, no. 8, pp. 2341-2355.

[43] Donald, F.M. 2008, "The classification of vigilance tasks in the real world", *Ergonomics,* vol. 51, no. 11, pp. 1643-1655.

[44] dos Santos, I.J.A.L., Teixeira, D.V., Ferraz, F.T. & Carvalho, P.V.R. 2008, "The use of a simulator to include human factors issues in the interface design of a nuclear power plant control room", *Journal of Loss Prevention in the Process Industries,* vol. 21, no. 3, pp. 227-238.

[45] Dourish, P. 2004, *Where the action is: the foundations of embodied interaction,* The MIT Press.

[46] Ducatel, K.& Comisión Europea 2001, *Istag: Scenarios for ambient intelligence in 2010,* Office for Official Publications of the European Communities.

[47] Dul, J., De Vlaming, P. & Munnik, M. 1996, "A review of ISO and CEN standards on ergonomics", *International Journal of Industrial Ergonomics,* vol. 17, no. 3, pp. 291-297.

[48] Durso, F.T. 1999, *Situation Awareness As a Predictor of Performance in En Route Air Traffic Controllers.*,.

[49] Edlund, J., Gustafson, J., Heldner, M. & Hjalmarsson, A. 2008, "Towards human-like spoken dialogue systems", *Speech Communication,* vol. 50, no. 8-9, pp. 630-645.

[50] El-Zabadani, H., Helal, S. & SCHMALZ, M. 2006, "PerVision: An integrated pervasive computing/computer vision approach to tracking objects in a self-sensing space", *Smart Homes And Beyond: Icost 2006, 4th International Conference on Smart Homes and Health Telematics*IOS Press, , pp. 315.

[51] Emerson, T., Reising, J. & Britten-Austin, H. 1988, "Workload and situation awareness in future aircraft", *Aerospace Behavioral Engineering Technology Conference, 6 th, Long Beach, CA*, pp. 107.

[52] Endsley, M.R. 2000, "Situation models: An avenue to the modeling of mental models", Proceedings of the Human Factors and Ergonomics Society Annual MeetingSAGE Publications, , pp. 61.

[53] Endsley, M. & Jones, W. 1997, "Situation awareness, information warfare and information dominance", *Endsley Consulting, Belmont, MA, Tech.Rep,*, pp. 97-01.

[54] Endsley, M.R. 1996, "Automation and situation awareness", *Automation and human performance: Theory and applications,* , pp. 163-181.

[55] Endsley, M.R. 1995, "Measurement of situation awareness in dynamic systems", *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, no. 1, pp. 65-84.

[56] Endsley, M.R. 1995, "Toward a theory of situation awareness in dynamic systems", *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, no. 1, pp. 32-64.

[57] Endsley, M.R. 1988, "Design and evaluation for situation awareness enhancement", *Human Factors and Ergonomics Society Annual Meeting Proceedings*Human Factors and Ergonomics Society, , pp. 97.

[58] Endsley, M.R.& Kiris, E.O. 1995, "The out-of-the-loop performance problem and level of control in automation", *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, no. 2, pp. 381-394.

[59] Flach, J.M. 1995, "Situation awareness: Proceed with caution", Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, pp. 149-157.

[60] Fitts, P.M. 1954, "The information capacity of the human motor system in controlling the amplitude of movement.",*Journal of experimental psychology,* vol. 47, no. 6, pp. 381.

[61] Fitzmaurice, G.W., Ishii, H. & Buxton, W.A.S. 1995, "Bricks: laying the foundations for graspable user interfaces", *Proceedings of the SIGCHI conference on Human factors in computing systems*ACM Press/Addison-Wesley Publishing Co., , pp. 442.

[62] Forlines, C., Wigdor, D., Shen, C. & Balakrishnan, R. 2007, "Direct-touch vs. mouse input for tabletop displays", *Proceedings of the SIGCHI conference on Human factors in computing systems*ACM, , pp. 647.

[63] Gershenfeld, N., Krikorian, R. & Cohen, D. 2004, "The Internet of Things.",*Scientific American,* vol. 291, no. 4, pp. 76-81.

[64] Gibson, J.J. 1986, *The ecological approach to visual perception,* Lawrence Erlbaum.

[65] Giles, J. 2010, "Inside the race to hack the Kinect", *The New Scientist,* vol. 208, no. 2789, pp. 22-23.

[66] Girgensohn, A., Kimber, D., Vaughan, J., Yang, T., Shipman, F., Turner, T., Rieffel, E., Wilcox, L., Chen, F. & Dunnigan, T. 2007, "DOTS: support for effective video surveillance", *Proceedings of the 15th international conference on Multimedia*ACM, , pp. 423.

[67] Girgensohn, A., Shipman, F., Turner, T.& Wilcox, L. 2007, "Effects of presenting geographic context on tracking activity between cameras", *Proceedings of the SIGCHI conference on Human factors in computing systems*ACM, , pp. 1167.

[68] Gordon, E.M. 1965, "Cramming more components onto integrated circuits", *Electronics Magazine,* vol. 4.

[69] Gouin, D. 2011, *Using Large Group Displays to Support Intensive Team Activities in C2,* .

[70] Gugerty, L.J. 1997, "Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory.",*Journal of Experimental Psychology: Applied,* vol. 3, no. 1, pp. 42.

[71] Hall, D.L. & Llinas, J. 2001, *Handbook of multisensor data fusion,* CRC Pr I Llc.

[72] Haering, N., Venetianer, P.L.& Lipton, A. 2008, "The evolution of video surveillance: an overview", Machine Vision and Applications, vol. 19, no. 5, pp. 279-290.

[73] Haller, S., The Things in the Internet of Things, Proceedings of Internet of Things Conference 2010, Tokyo, 2010.

[74] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H. & Pankanti, S. 2005, "Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking", Signal Processing Magazine, IEEE, vol. 22, no. 2, pp. 38-51.

[75] Harper, R. 2008, *Being human: human-computer interaction in the year 2020,* Microsoft Research.

[76] Hart, S.G.& Staveland, L.E. 1988, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research", *Human mental workload,* vol. 1, pp. 139-183.

[77] Hauland, G. 2002, *Measuring team situation awareness in training of en route air traffic control.Process Oriented Measures for Experimental Studies,* .

[78] Hauss, Y.& Eyferth, K. 2003, "Securing future ATM-concepts' safety by measuring situation awareness in ATC", *Aerospace science and technology,* vol. 7, no. 6, pp. 417-427.

[79] Hightower, J., Brumitt, B. & Borriello, G. 2002, "The location stack: A layered model for location in ubiquitous computing", *Mobile Computing Systems and Applications, 2002. Proceedings Fourth IEEE Workshop on*IEEE, , pp. 22.

[80] Hollands, J.G.& Wickens, C.D. 1999, *Engineering psychology and human performance.* Prentice Hall New Jersey.

[81] Hsu, H.H., Cheng, Z., Huang, T. & Han, Q. 2006, "Behavior analysis with combined RFID and video information", *Ubiquitous Intelligence and Computing,* , pp. 176-181.

[82] Iannizzotto, G., Costanzo, C., La Rosa, F. & Lanzafame, P. 2005, "A multimodal perceptual user interface for video-surveillance environments", *Proceedings of the 7th international conference on Multimodal interfaces*ACM, , pp. 45.

[83] Ishii, H. 2008, "Tangible bits: beyond pixels", *Proceedings of the 2nd international conference on Tangible and embedded interaction*ACM, , pp. xv.

[84] Jaimes, A. & Sebe, N. 2007, "Multimodal human-computer interaction: A survey", *Computer Vision and Image Understanding,* vol. 108, no. 1-2, pp. 116-134.

[85] Jansen, Y., Karrer, T. & Borchers, J. 2010, "MudPad: tactile feedback and haptic texture overlay for touch surfaces", *ACM International Conference on Interactive Tabletops and Surfaces*ACM, , pp. 11.

[86] Jia, S., Sheng, J., Chugo, D. & Takase, K. 2007, "Human recognition using RFID technology and sterero vision", *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*IEEE, , pp. 1488.

[87] Kaber, D.B.& Endsley, M.R. 2004, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task", *Theoretical Issues in Ergonomics Science,* vol. 5, no. 2, pp. 113-153.

[88] Kaber, D.B.& Endsley, M.R. 1997, "Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety", *Process Safety Progress,* vol. 16, no. 3, pp. 126-131.

[89] Kaber, D.B., Omal, E. & Endsley, M.R. 1999, "Level of automation effects on telerobot performance and human operator situation awareness and subjective workload", *Automation technology and human performance: Current research and trends,* , pp. 165-170.

[90] Karray, F., Alemzadeh, M., Saleh, J.A.& Arab, M.N. 2008, "Human-computer interaction: Overview on state of the art", *International Journal on Smart Sensing and Intelligent Systems,* vol. 1, no. 1, pp. 137-159.

[91] Keval, H.& Sasse, M. 2006, "Man or gorilla? Performance issues with CCTV technology in security control rooms", *16th World Congress on Ergonomics Conference, International Ergonomics Association*, pp. 10.

[92] Keval, H.U.& Sasse, M.A. 2008, "Can we ID from CCTV? Image quality in digital CCTV and face identification performance", *Proc. SPIE*.

[93] Kinney, P. 2003, "Zigbee technology: Wireless control that simply works", *Communications design conference*.

[94] Krahnstoever, N., Rittscher, J., Tu, P., Chean, K. & Tomlinson, T. 2005, "Activity recognition using visual tracking and rfid", *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*IEEE, , pp. 494.

[95] Kuniavsky, M. 2010, *Smart things: ubiquitous computing user experience design,* Morgan Kaufmann.

[96] Kurt, T.E. 2007, *Hacking roomba,* Wiley Pub.

[97] Lakoff, G. & Johnson, M. 1980, *Metaphors we live by,* Chicago London.

[98] Lee, J.C. 2008, "Hacking the nintendo wii remote", *Pervasive Computing, IEEE,* vol. 7, no. 3, pp. 39-45.

[99] Lee, J.S., Su, Y.W. & Shen, C.C. 2007, "A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi", *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*Ieee, , pp. 46.

[100] MacKenzie, I.S. 1992, "Fitts' law as a research and design tool in human-computer interaction", *Human-Computer Interaction,* vol. 7, no. 1, pp. 91-139.

[101] Mancero, G., Wong, W. & Loomes, M. 2009, "Change blindness and situation awareness in a police C2 environment", *European Conference on Cognitive Ergonomics: Designing beyond the Product---Understanding Activity and User Experience in Ubiquitous Environments*VTT Technical Research Centre of Finland, , pp. 5.

[102] Marchesotti, L., Singh, R. & Regazzoni, C. 2004, "Extraction of aligned video and radio information for identity and location estimation in surveillance systems", *Int. Conf. of Information Fusion, Stockholm, Sweden*Citeseer, .

[103] McDaniel, T.L.& Panchanathan, S. 2006, "A visio-haptic wearable system for assisting individuals who are blind", *ACM SIGACCESS Accessibility and Computing,* , no. 86, pp. 12-15.

[104] McTear, M.F. 2002, "Spoken dialogue technology: enabling the conversational user interface", *ACM Computing Surveys (CSUR),* vol. 34, no. 1, pp. 90-169.

[105] Merk, L., Nicklous, M., Stober, T.& Hansmann, U. 2001, "Pervasive Computing Handbook", .

[106] Michael L., A. 2003, "Embodied Cognition: A field guide", *Artificial Intelligence,* vol. 149, no. 1, pp. 91-130.

[107] Mitchell, H.B. 2007, *Multi-sensor data fusion: an introduction,* Springer.

[108] Moore, G.E. 1998, "Cramming more components onto integrated circuits", *Proceedings of the IEEE,* vol. 86, no. 1, pp. 82-85.

[109] Nakamura, E.F., Loureiro, A.A.F.& Frery, A.C. 2007, "Information fusion for wireless sensor networks: Methods, models, and classifications", *ACM Computing Surveys (CSUR),* vol. 39, no. 3, pp. 9.

[110] Nakashima, H., Aghajan, H. & Augusto, J.C. 2009, *Handbook of Ambient Intelligence and Smart Environments,* Springer-Verlag New York Inc.

[111] Nardi, B.A. 1996, "Studying context: A comparison of activity theory, situated action models, and distributed cognition", *Context and consciousness: Activity theory and human-computer interaction,*, pp. 69-102.

[112] Nielsen, J. 2004, *Designing web usability,* Pearson Education.

[113] Nielsen, J. 1992, "Evaluating the thinking-aloud technique for use by computer scientists", .

[114] Norman, D.A. 2003, *Emotional design: Why we love (or hate) everyday things,* Basic Books New York.

[115] Norman, D.A. 1998, "The psychopathology of everyday things", .

[116] Norman, D.A. 1993, *Things that make us smart: Defending human attributes in the age of the machine,* Basic Books.

[117] Norman, D.A. 1990, "The'problem'with automation: inappropriate feedback and interaction, not'over-automation'", *Philosophical Transactions of the Royal Society of London.B, Biological Sciences,* vol. 327, no. 1241, pp. 585-593.

[118] Norman, D.A.& University of California, San Diego. Dept. of Cognitive Science 1990, *Cognitive artifacts,* Dept. of Cognitive Science, University of California, San Diego.

[119] Oviatt, S.L. 2003, "Multimodal interfaces", *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications,*, pp. 286-304.

[120] Parasuraman, R., Sheridan, T.B.& Wickens, C.D. 2000, "A model for types and levels of human interaction with automation", *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 30, no. 3, pp. 286-297.

[121] Patwari, N., Ash, J.N., Kyperountas, S., Hero III, A.O., Moses, R.L. & Correal, N.S. 2005, "Locating the nodes: cooperative localization in wireless sensor networks", *Signal Processing Magazine, IEEE,* vol. 22, no. 4, pp. 54-69.

[122] Plouznikoff, N., Plouznikoff, A. & Robert, J.M. 2005, "Object augmentation through ecological human-wearable computer interactions", *Wireless And Mobile Computing, Networking And Communications, 2005.(WiMob'2005), IEEE International Conference on*IEEE, , pp. 159.

[123] Prati, A., Vezzani, R., Benini, L., Farella, E. &Zappi, P. 2005, "An integrated multi-modal sensor network for video surveillance", Proceedings of the third ACM international workshop on Video surveillance & sensor networksACM, , pp. 95.

[124] Power, D.J.& Sharda, R. 2009, "Decision support systems", *Springer Handbook of Automation,* , pp. 1539-1548.

[125] Pylyshyn, Z.W.& Storm, R.W. 1988, "Tracking multiple independent targets: Evidence for a parallel tracking mechanism*", *Spatial vision,* vol. 3, no. 3, pp. 179-197.

[126] Raja, Y., Gong, S. & Xiang, T. 2011, "User-assisted visual search and tracking across distributed multi-camera networks", *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 20.

[127] Regazzoni, C.S., Cavallaro, A., Wu, Y., Konrad, J. & Hampapur, A. 2010, "Video Analytics for Surveillance: Theory and Practice [From the Guest Editors]", Signal Processing Magazine, IEEE, vol. 27, no. 5, pp. 16-17.

[128] Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R. & Young, M.S. 2008, "What really is going on? Review of situation awareness models for individuals and teams", *Theoretical Issues in Ergonomics Science,* vol. 9, no. 4, pp. 297-323.

[129] Salmon, P., Stanton, N., Walker, G. & Green, D. 2006, "Situation awareness measurement: A review of applicability for C4i environments", *Applied Ergonomics,* vol. 37, no. 2, pp. 225-238.

[130] Sarter, N.B., & Woods, D.D. (1991). Situation awareness: A critical but ill-defined phenomenon. International Journal of Aviation Psychology, 1, 45-57.

[131] Schilit, B., Adams, N. & Want, R. 1994, "Context-aware computing applications", *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*IEEE, , pp. 85.

[132] Scholl, B.J., Pylyshyn, Z.W. & Feldman, J. 2001, "What is a visual object? Evidence from target merging in multiple object tracking", *Cognition,* vol. 80, no. 1-2, pp. 159-177.

[133] Sheridan, T.B. 1997, "Supervisory control", *Handbook of human factors and ergonomics,* , pp. 1025-1052.

[134] Sheridan, T.B. 1978, *Human and computer control of undersea teleoperators,* .

[135] Simon, D., Cifuentes, C., Cleal, D., Daniels, J.& White, D. 2006, "Java™ on the bare metal of wireless sensor devices: the squawk Java virtual machine", *Proceedings of the 2nd international conference on Virtual execution environments*ACM, , pp. 78.

[136] Smith, R.B. 2007, "SPOTWorld and the Sun SPOT", *Proceedings of the 6th international conference on Information processing in sensor networks*ACM, , pp. 565.

[137] Smith, K.& Hancock, P. 1995, "Situation awareness is adaptive, externally directed consciousness", Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, pp. 137-148.

[138] Snidaro, L., Micheloni, C. & Chiavedale, C. 2005, "Video security for ambient intelligence", *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 35, no. 1, pp. 133-144.

[139] Snidaro, L., Visentini, I. & Foresti, G. 2010, *Sensor Management: A New Paradigm for Automatic Video Surveillance*, Springer Berlin / Heidelberg.

[140] Stedmon, A. 2011, "The camera never lies, or does it? The dangers of taking CCTV surveillance at face value and the importance of human factors", *Surveillance & Society,* vol. 8, no. 4, pp. 527-534.

[141] Stedmon, A.W., Harris, S. & Wilson, J.R. 2011, "Simulated multiplexed CCTV: The effects of screen layout and task complexity on user performance and strategies", *Security Journal,* vol. 24, no. 4, pp. 344-356.

[142] Steinberg, A.N. 1999, *Revisions to the JDL data fusion model,* .

[143] Streitz, N.A. 2007, "From human-computer interaction to human-environment interaction: ambient intelligence and the disappearing computer", *Lecture Notes in Computer Science,* vol. 4397, pp. 3.

[144] Streitz, N.A., Rocker, C., Prante, T., van Alphen, D., Stenzel, R. & Magerkurth, C. 2005, "Designing smart artifacts for smart environments", *Computer,* vol. 38, no. 3, pp. 41-49.

[145] Streitz, N. & Nixon, P. 2005, "Introduction", *Commun.ACM,* vol. 48, no. 3, pp. 32-35.

[146] Taylor, R. 1990, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design", *the Situational Awareness in Aerospace Operations AGARDCP478,*, no. Situational Awareness in Aerospace Operations.

[147] Tian, Y., Brown, L., Hampapur, A., Lu, M., Senior, A. & Shu, C. 2008, "IBM smart surveillance system (S3): event based video surveillance system with an open and extensible framework", Machine Vision and Applications, vol. 19, no. 5, pp. 315-327.

[148] Ullmer, B. & Ishii, H. 2000, "Emerging frameworks for tangible user interfaces", *IBM Systems Journal,* vol. 39, no. 3.4, pp. 915-931.

[149] Valli, A. 2005, *Notes on natural interaction,* .

[150] Vural, U. & Akgul, Y.S. "Operator Attention Based Video Surveillance", .

[151] Wallace, E. & Diffley, C. 1988, "CCTV control room ergonomics", *Published by Police Scientific Development Branch of the Home Office, Publication,* , no. 14/98.

[152] Wan, K. 2009, "A Brief History of Context", *Arxiv preprint arXiv:0912.1838,* .

[153] Want, R., Schilit, B.N., Adams, N.I., Gold, R., Petersen, K., Goldberg, D., Ellis, J.R. & Weiser, M. 1995, "An overview of the PARCTAB ubiquitous computing experiment", *Personal Communications, IEEE,* vol. 2, no. 6, pp. 28-43.

[154] Warneke, B., Last, M., Liebowitz, B. & Pister, K.S.J. 2001, "Smart dust: Communicating with a cubic-millimeter computer", *Computer,* vol. 34, no. 1, pp. 44-51.

[155] Weiser, M. 1991, "The computer for the 21st century", *Scientific American,* vol. 265, no. 3, pp. 94-104.

[156] Weiser, M. & Brown, J.S. 1996, "The coming age of calm Technology [1]", *Xerox PARC.Retrieved July,* vol. 8, pp. 2007.

[157] Wilson, R.A. & Keil, F.C. 2001, *The MIT encyclopedia of the cognitive sciences,* The MIT Press.

[158] Wobbrock, J.O., Aung, H.H., Rothrock, B. & Myers, B.A. 2005, "Maximizing the guessability of symbolic input", *CHI'05 extended abstracts on Human factors in computing systems*ACM, , pp. 1869.

[159] S. Yu, I. Aedo, P. Diaz, P. Acuna, T. Onorati 2011, "iNeres: Personalized Mobile Emergency Notification and Evacuation Routes System in Indoor Environment", .

[160] Zhang, J., Orlik, P.V., Sahinoglu, Z., Molisch, A.F. & Kinney, P. 2009, "UWB systems for wireless sensor networks", *Proceedings of the IEEE,* vol. 97, no. 2, pp. 313-331.

[161] Zhong, H., Shi, J. & Visontai, M. "Detecting unusual activity in video", *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*IEEE, , pp. II.

[162] Zimmermann, A., Lorenz, A. & Oppermann, R. 2007, "An operational definition of context", *Proceedings of the 6th international and interdisciplinary conference on Modeling and using context*Springer-Verlag, , pp. 558.

List of Works Related to the Thesis

International Conference Papers

- Ceipidor, U.B., Dibitonto, M., D'Ascenzo, L. & Medaglia, C.M. 2010, "Localization Issues in a ZigBee Based Internet of Things Scenario", The Internet of Things, , pp. 157-165.
  (Related to section 2.1.3)

- Dibitonto, M., Buonaiuto, A., Marcialis, G., Muntoni, D., Medaglia, C. & Roli, F. 2011, "Fusion of Radio and Video Localization for People Tracking", Ambient Intelligence, , pp. 258-263.
  (Related to section 5)