Chapter 4

VIDEO AND RADIO ATTRIBUTES EXTRACTION FOR HETEROGENEOUS LOCATION ESTIMATION

A Context-based Ambient Intelligence Architecture

S.Piva, R.Singh, M.Gandetto and C.S.Regazzoni

Department of Biophysical and Electronical Engineering, University of Genova, Italy

piva@dibe.unige.it

Keywords: AmI architecture, context awareness, multimodal Location feature extraction, radio-based location, video-based location

1. Introduction

Thanks to advances in a range of heterogeneous fields, from video analysis and understanding to software agents, from multi-sensor context assessment to pervasive communications, research on Ambient Intelligence systems is making significant progress towards the implementation of Smart Spaces (SS) where users are provided with an ensemble of ubiquitous virtual services [1][2][3][4]. ISTAG (Information Society Technology Advisory Group) gives a more formal definition of Ambient Intelligence (AmI) that points out how it should provide technologies to surround users with intelligent sensors and interfaces and to support human interactions [5]. While the specific services being provided can vary significantly according to the application domain, a common requirement for many AmI systems is that of supporting their users with guidance on available services, safety warnings and navigation aids. Providing effective assistance to users complex scenarios requires giving to the system a certain degree of awareness of the user's general preferences, current activities and real time condition [2]. In a word the system needs knowledge about the context in which user and system itself are acting. Context perception and understanding involves all the issues related to intelligent sensors and the management of heterogeneous information sources. As a human being perceives through his senses a huge amount of information and exploits them to decide for his actions, an artificial intelligent system based on contextual sensing must be provided with the appropriate sensors set and with the not easily achievable ability to extract interesting data at a higher abstraction level [6]. The central concept of context awareness represents the possibility for the system of biasing itself and its reactions to the environment [7][8]. A key role among many useful features is represented by the detection of the user position inside the monitored environment because many applications are location dependent. The successful location of objects of potential interest and their identification are still open issues in the state of the art; anyhow different positioning approaches are available based on: Video [9] and Radio Signals [10]. The wide diffusion of radio devices on the market has given the opportunity to improve the functionality of SSs as most of the moving devices are equipped with such systems. The identification of users is not a problem for Radio-Based Systems (RBS) while is a challenge for Video-Based ones (VBS). The easier implementation of location tasks for RBSs has the price of less accurate localization data measurements. A tradeoff solution is then a system able to integrate heterogeneous sensors to perform localization exploiting the features of these two systems. Pursuing this aim, an architecture exploiting CCD-Video Cameras and 802.11 Wireless LAN positioning methodologies is presented; in particular the paper specifically addresses one of the first step towards fusion that is, according to a popular Data Fusion model [10], Data Alignment. It is easy to understand that the development of such a system is challenged by a large number of open problems, not last the design of an architecture assembling the fundamental functions of a Smart Space: Perception, Analysis, Decision, and Action (including Communication with the user). The aim is then the one to define a general structure able to take into account the main issues related to the management of the many different heterogeneous technologies characterizing an Ambient Intelligence environment. First of all in this work we propose a logical architecture we think has the generality and the completeness to allow the analysis of many of the possible problems raised by a Smart Space. An interesting inspiration point to deal with this kind of issue is to study the principles at the bases of the human brain working. A good inspiration can be found in [11] where a model for self consciousness is reported and motivated on neuro-physiological bases. This chapter is organized as follows: after this introduction, section 2 reports the most known ambient intelligence system implementations in the state of the art; section 3 defines the main issues and functionalities characterizing an ambient intelligence system; the proposed architecture and its inspiring model is described in section 4 while some concepts about the central topic of context awareness can be found in section 5. Some results for the location sensing task are depicted in section 6 and finally in 7 we draw some conclusions about the work.

2. Related work

Ambient intelligence is at the moment an open research field, still not bounded for what concerns topics of interest and related issues. In addition, systems which would be addressed as commercially appealing by the market are still developed at prototypal level. Technology is not enough mature to consider Ambient Intelligence a reality for users or a commercial product: many researches often deal with a single aspect (or a few) of a whole complex system. Although this, a certain number of integrated projects are in progress to show the potentiality of this technology in actual test beds. Among the others, it is mandatory to remember the first Ambient Intelligence system presented by Trivedi, Huang e Mikic. It uses several cameras and microphones to acquire information on the surrounding environment [6]. According to concepts developed of the Oxygen project at M.I.T. (Massachusetts Institute of Technology), in next future computational power will be freely available everywhere and it will not be necessary to carry computers or personal communication terminals to access it [12]. More recently at the Artificial Intelligence Lab of the M.I.T. an Intelligent Room has been designed and developed. Research has been focused on natural and reliable vocal interfaces, context-specific user-transparent system reactions, dynamic resource allocation and natural cooperation among different contiguous environments [13][14]. On the private companies side, a strong interest in this research field can be found in the Dutch Philips where one of the most important examples of this kind is represented by the Phenom project [15]. Phenom is a home-oriented long-term research project, which aims at creating an environment that shows a context aware behaviour. At the Georgia Institute of Technology a long term research group is developing a project called The Aware Home Research Initiative (AHRI). It is an interdisciplinary research endeavour aimed at addressing the fundamental technical, design, and social challenges presented by the creation of context aware home environment [16][17]. Because of the topical argument, several works are still ongoing, such as a project funded by the Dutch Ministry of Economic Affairs: Ambience. The project involves academic partners as well as research divisions of private companies. The goal of the Ambience project is to jointly create networked context aware environments [18]. Another important work aiming at defining and realizing a complete and complex test bed is represented by VICOM (Virtual Immersive COMmunication) [19]. This Italian three-year project involves many academic and private partners and is developing strategies for context data extraction and management as well as multiple sensor systems and ad-hoc networking communication technologies [20]. To expand the vision about the related ongoing researches on specific topics the reader can also refer to [21][22][23] and for the central issue explored in this paper, context extraction, good sources can be found in the works by Neerincx and Streefkerk [24], Caarls and Persa for a multisensor fusion-based data extraction [25]. An example of association of mobility and context awareness can be found in [26]. Many publications on these topics are results of the work of Prof. James Crowley of the Institute National Polytechnique de Grenoble who directs the Prima project [27] about perception and integration for smart spaces [28][29]. Another specific reference for context based system design is finally represented by [30].

3. Main tasks of Ambient Intelligence systems

In Figure 4.1 the idea of the intelligent environment surrounding the user is depicted. In the user-centered approach the system is designed to be the less intrusive possible: the idea is the one to allow the use and the enjoyment of additional functionalities and facilities without the need of any training to learn the way to interact with new technologies. The design has to start from the user needs and preferences and to give the smart space the ability to automatically adapt to them. The functionalities cooperating to this aim are something that tries to resemble an intelligent behavior in which a sensing task has the duty to collect data about the environment, data and information that becoming the input of the analysis task are processed to extract useful and concise higher level metadata the decision logics employs to choose the best action to undertake relatively to its metrics. Then the chosen action is realized by the action/communication oriented smart space function in order to apply an influence on the user itself. Consequences of these actions are again captured by the sensing tasks to close the loop and to judge the result of the processed in optimizing the interaction between the system and the actors interacting with its domain. In this work we go deeper inside the sensing task to explore a method to associate different techniques in obtaining contextual data, in particular for what concerns the position of the interacting objects.

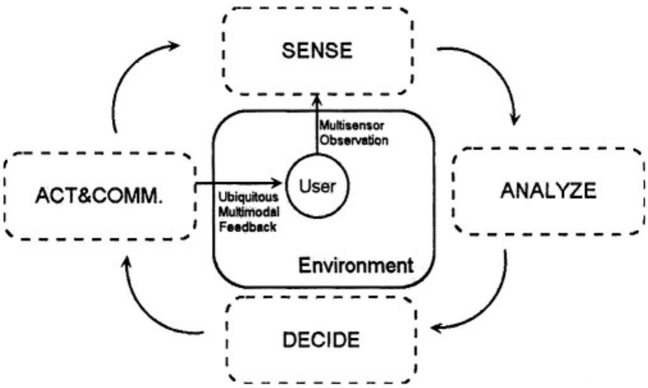

Figure 4.1. User centered ambient intelligence closed loop

4. Architecture design

4.1 Inspiration

Dealing with the complex issues related to Ambient Intelligence systems rises the need for a coherent and complete to allow a proper control on designing actions. In order to validate the design of a general structure to manage the main issues related to a Smart Space, our approach is the one to take inspiration from a theoretical human conscience model. The desired inspiration can be found in the work of Antonio Damasio [11] where a model for brain and conscious reasoning is described and motivated on neuro-physiological bases. This approach results particularly useful in the design of a complex distributed systems characterized by high-level analysis and interaction capabilities. Systems of this kind needs to be aware of the behavior of all objects acting in their scope of sensing as well as of its own components' status and reactions. In this sense, awareness of the context, in which the system acts, is not enough; the system has also in some way to be conscious of its own internal state. In this section we thus introduce in details the cited model in order to use it as the root for our architecture design. In the description of Damasio, the acquisition and the behavior of what we call self consciousness comes from the interaction of two components: the organism and the objects and in terms of the relationships they hold in the course of their natural interactions. For our purposes the organism in question is the AmI system whereas the object is any entity that gets to be known by the system; the relationships between organism and object are the contents of the knowledge we call consciousness. Seen in this perspective, consciousness consists of constructing knowledge about two facts:

- the organism is involved in relating to some object;
- perceived objects cause changes in the organism [11].

While the human perceiving system changes dutifully at the mercy of the objects it interacts with, a number of brain regions whose job is to regulate the life process does not change at all in terms of the kind of object they represent. The degree of change occurring in the object –the body– is quite small. This is because only a narrow range of body states is compatible with life and the organism is genetically designed to maintain that narrow range and equipped to seek it. The body's internal state must be relatively stable by comparison to the environment surrounding it. So the deep roots for the 'self' are to be found in the ensemble of brain devices that continuously and non-consciously maintain the body state within the narrow range and relative stability required for survival. Following the Damasio model, the state of activity within the ensemble of such devices is defined Proto-Self, whereas the "non-conscious" part of self is represented by the Core self and Autobiographical Self. The following are the definitions of the three components concurring to form the human consciousness:

The Proto-Self is an interconnected and temporarily coherent collection of neural patterns which represent the state of the organism, moment by moment, at multiple levels of the brain. We are not conscious of the proto-self.

The Autobiographical Self is a conscious part of self, based on autobiographical memory which is constituted by implicit memories of multiple instances of individual experience of the past and of the anticipated future. This memory grows continuously with life experience but can be partly remodeled to reflect new experiences.

The Core Self: is generated for any object that provokes the core-consciousness mechanism that is to say the continuous environment awareness due to the analysis of stimulating objects. Because of the permanent availability of provoking objects, it is continuously generated and thus appears continuous in time. The mechanism of Core Self requires the presence of Proto-Self. Core Self can be triggered by any object. The mechanism of production of Core Self undergoes minimal across a lifetime. We are conscious of the Core Self.

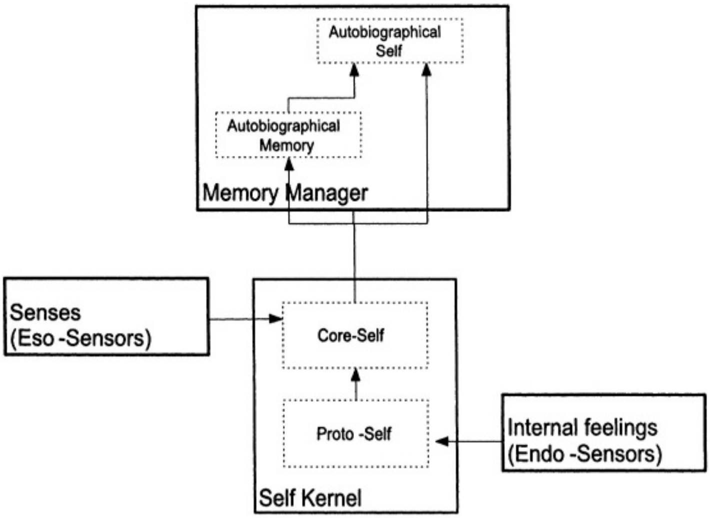

Figure 4.2. Damasio's Human Conscience Model. The described components are depicted in the diagram along with their connections. The Proto-Self and Core-Self are linked to senses and to internal organs by means of nerves connections and electric messages.

In Figure 4.2 Damasio's model is depicted, as it can be seen Proto Self constitutes the basis for the Core Self whereas the Autobiographical Self has got inferences coming from the Core Self and the Autobiographical Memory. This means the Autobiographical Self, that can be seen as the short term memory, is continuously composed by recalled elements coming from the long term memory (Autobiographical Memory).

4.2 Mapping the Model into an AmI Architecture

As previously stated our aim is to try and exploit the abstract concepts proposed by Damasio and translate them in the architecture of an artificial system able to be aware of the context it is working in. Proto-Self functionality can be realized through a system's internal state context analysis: the non-conscious human monitoring on the inner parts of the organism can be translated into the use of internal sensors. In both cases the aim to control critical variables and to analyze the relationship between environmental (external) events and internal state conditions. Core-Self is instead represented in its artificial intelligence counterpart, as the physical context representation, namely the collection and association of the continuous observation data coming from the senses (for the human brain) or the external sensors (for what concerns a context awareness based system). Autobiographical self, based on the use of autobiographical memory represents the conscious part of self. The two concepts are strictly connected and can be seen as the short-term memory and the long-term memory respectively in the model of the human brain. They both constitute the idea of growing experience that can be implemented with associative techniques working on a knowledge base.

4.3 Artificial Sensing

Sensors own a key role in the definition and in the input of both Proto-self and Core-self: in an intelligent system they represent the bridge between the 'brain', namely the intelligent core of the organism, and the world, being it internal or external reality. Sensors (or receptors) are used by the system to keep contact to the interesting data and variables of the working environment (External World) as well as of its own internal state (Internal World). Considering this kind of distinction we can classify these devices or software agents into two groups that do not differ for technological aspects but for the aim of their observation, that is to say their observed domain. This means we can distinguish between Endo-receptors and Eso-receptors and associate them respectively to Proto-self and Core-Self (Figure 4.2). In particular, as suggested by their name, Endo-receptors are devoted to the observation of the internal state of the system: that is to say devices fit for analyzing internal components or variables proper of devices making part of the whole organism (system). Instances of sensors belonging to this class are: computational units (i.e.: Desktop PCs present in the environment available for users), devices' status sensors, thermal sensors, safety-oriented sensors (smoke, gas, fire, water infiltration, etc.), lighting sensors and so on. Endo-receptors are what is needed to realize the concept of the Proto-Self, as described in the introductory part. With Eso-receptors we refer to all the devices used by the structure to keep track of the events occurring in the observed domain and to collect data about the target of the analysis, being humans or other external interacting objects. Eso-receptors are the counterpart of the human senses, in this category fall sensors such as video sensors (working in visible or infra-red wave length fields), radio sensors (i.e. WLAN, Aps, BlueTooth, Global Positioning Systems (GPS)), standard or directional microphones, weight sensors, fingerprints readers, electro-magnetic waves emission scanners, photoelectric cells, etc. The focus of this work on external sensing tasks will be described in next sections.

4.4 Proposed structure

The complete structure of an AmI system here proposed (Figure 4.3) is defined in terms of interconnected logical modules. Each module implements a particular functionality such as sensing (i.e.: Endo/Eso receptors), analyzing (i.e.: Context Manager and Memory Manager), deciding (i.e.: Decision Manager) and acting (i.e.: Actuators and Multimode Communicator). In this sense, the system steps through the aforementioned functionalities with inferences on an Internal World (i.e.: the "artificial organism", the system itself) and on an External World (i.e.: the "objects", users). The latter environment represents those elements not under the direct control of the system. This means the system is aware of the objects interacting in its domain, but it cannot apply a direct inference on them. For example, in the realized test-bed, we apply the interaction towards the external world (the users) by adapting message communication.

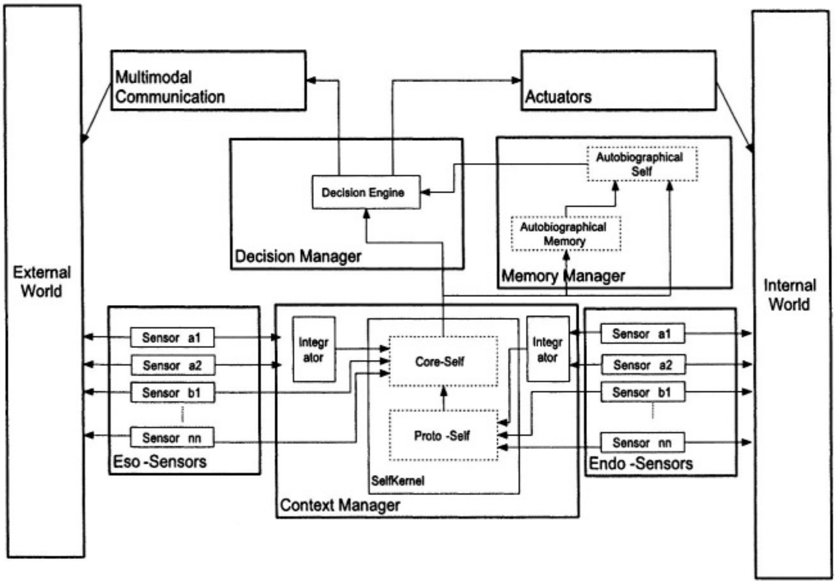

Figure 4.3. The proposed AmI logical architecture. Internal and External World represent the system itself and the objects and users interacting with the Smart Space, respectively. System adaptivity is explicitly expressed in the Decision Manager and in the Multimodal Communication modules

The only bridge keeping a connection between the intelligence of the system and the external world is expressed, as previously stated, by the pro-active functionalities it is provided with. Referring to the model, this is what represents the source of the continuous stimuli received by the Core-self. On the other hand, the Internal World is every thing the system can directly control: physical devices, internal software parameters but also physical components the system can apply its inference on. As it can be seen in Figure 4.3 the core part of the architecture that will be further described in next paragraph, takes origin from the Self Kernel sub module introduced in the description of Damasio's model. In particular Proto-self and Core-self are encapsulated in a Context Manager (CM) devoted to the analysis of heterogeneous data and to the generation of contextual information. The ensemble of internal and external state information is encoded by the Context Manager a hierarchical representation, where a superstate (e.g. ARRIVAL, meaning a new moving object entering the scene) gives the highest-level summarization of the context, while lower-level attributes

Table 4.1. **Hierarchical representation of the context**

	Level	Description
1	Event level	super-states
2	Behaviour level	trajectories of users
3	Object level	identity of users
4	Feature level	number and position of users, system components status
5	Signal level	raw, non processed data

describe the position and behavior of individual people, objects and resources [31]. The reference hierarchy is summarized in Table 4.1.

In such a structure, lower levels contribute to the estimation of the higher ones, a super state (an event) can be evaluated by considering objects and related features(i.e., the object's position). Event and Behaviour levels define the context in which objects and users act, whereas the remaining levels contribute to the definition of more specific and personalized description for objects. Autobiographical Self with Autobiographical Memory, respectively embody system's long and short-term memory; they constitute the Memory Manager module (MM) devoted to the storage under the form of symbolic metadata of events produced by Context Manager. In addition it forms a Knowledge–Base for higher level modules. In an artificial system this concepts represents a static memory in which all the information needed to take decisions are stored and a buffer keeping trace of the instantaneous variables used to manage data processing. The Decision Manager depicted in the diagram uses the context awareness represented by the Self Kernel output to make the system reacting and have its influence on the AmI domain, as either internal or external world. To provide its results, the decision logic also relies on the direct observations of the receptors and on the past data provided by the Memory Manager. The decision can affect the Internal and External Worlds in different ways:

- Internal world: directly through some physical or software Actuator Modules (i.e.: Actuators in the diagram) such as mechanical tools which close or open windows and doors, as well as thermostatic devices affecting the rooms' temperature or in general controlling and modifying parts of the system etc.
- External world: through Multimodal Communication channels: the decision concerns the choice of the proper information to be delivered and the choice of the best communication channel to address the user.

In the scope of this work we further describe in the following sections the issues related to sensing tasks, particularly dealing with video and radio-based location sensors.

5. Context aware systems

The ability of an Ambient Intelligence system to adapt its behavior to the peculiar features of the user and of the environment, as previously introduced, is related to sensing and to context information extraction. But what do we mean with context? To give a formal definition, we can say contextual information can be defined as an ordered multilevel set of declarative information concerning events occurring both within the sensing domain of a Smart Space (SS) and within the communication and actions domains of the SS itself. Relations among events are also included (explicitly or implicitly) within contextual information [32]. An event can be defined as the occurrence of some fact that can be perceived by or be communicated to the SS; an event is characterized by attributes that basically answer questions about where (position) and when (time) the event occurred. Other attributes involve what (core) consists the event of, who (identity) is involved in the event, and possibly why (reason) the event occurred [33]. Events can be used to represent any information that can characterize the situation of an interacting user as well as of a Smart Space component, i.e. an entity. An entity can be a person, a place, or an object that is relevant to the interaction between a user and an application. The user and the SS parts themselves are entities. The multilevel nature of contextual information is related to the possibility of detecting and representing events at multiple abstraction levels. Context-awareness refers to the property of a SS to internally represent in terms of events the state of its users, of their surroundings and of SS parts, as well as to be provided with rules that make it possible to adapt its behaviour accordingly. Therefore, in a context aware SS, contextual information can be either used by itself, i.e. it can have its own value for the interacting user, or it can be used to select which services can be provided to the user and which interaction modalities are more suited to let him/her access such services. In brief, an estimate of the user context can be used to optimize the adaptation and personalization process in such a way to maximize the service value. Pascoe introduced a set of four context-aware capabilities that applications can support [34]:

- Contextual sensing: a system detects the context and simply presents it to the user, augmenting the user's sensory system.
- Contextual adaptation: a system uses the context to adapt its behaviour instead of providing a uniform interface in all situations.
- Contextual resource discovery: a system can locate and use resources which share part or all of its context.
- Contextual augmentation: a system augments the environment with additional information, associating digital data with the current context.

This distinction results particularly useful in the design of a complex distributed systems characterized by a high-level analysis and interaction capabilities. In the example we present we particularly take into account capability of Context Adaptation with the consideration that many user oriented application are dependent upon his location inside the monitored environment.

5.1 Location feature

As previously stated in this section we go inside the sensing functionalities of a Smart Space and choose to deal with the issues related to context aware information of the active objects, in particular its positioning, into the system domain. So which are those context features, that give the maximum information related to the object, should be used in the Ambient Intelligence system? What are those context aware information that are peculiar to the object, how it can benefit the AmI functional capability and which are those inferences that can be drawn based on selected context information, are the general questions that a system designer has to come across while designing the efficient AmI system. While looking for these answers, it is but obvious to look forward towards location as a context feature because firstly it gives the particular information about the object's position, Secondly, almost all the location sensing devices momentarily identify the object, thirdly, based on the location information other inferences such as object's intention, its mood, orientation, general behavior, track, identity and many important contextual data can be extracted. The location is defined as the spatial position of the object of interest in the known environment. The spatial position is given with respect to some particular reference position. This reference position is function of those sensor devices which momentarily identifies and extract the location information. Location data has been a very important information in navigation, military purposes, exploring earth, finding oneself in an unknown environment, in crime detection and prevention, controlling mentally challenge people's activity, civil applications, controlling the activity in the museum and nation's important security places etc.. Due to the need of the automatic services and heavy user's demand for location based services, and the presence of multi- wireless devices in the market, the research to implement and provide easy location services is continuously going on. The common reason of such interest is that automatically extracted knowledge of position in space and time of objects in an environment is the basis of multiple possible forms of intelligent cooperation with the object themselves, varying from pure perception, to active interaction, to personalized remote or short range communications. Having realised that knowledge of location is an important context aware information in the AmI system, we are trying to explore the possibility of combing and using multi-location sensors existing in the same environment. In order to validate the possibility of fusing heterogeneous sensors, we have incorporated two sensors namely, Video and Radio based location system. The idea is to extract the video and radio attributes and combine them efficiently to obtain as much inference as possible. As a step towards this, in this chapter, the attributes extraction methodologies, i.e. position, for the above two mentioned sensors are explored and described below. The effort is in the direction of proving the fact that the above two sensors can be combined and aligned to obtain the context aware information such as identity and accurate position.

5.2 The formalism

The formalism here-in-after used to describe the logical functional overview of the proposed systems assumes that a set of heterogeneous sensors $S = \bar{S}^c : c = 1, ...N_s$ is divided in, N_s, different classes

$\bar{S}^c = S_i^c : i = 1, ...N_{\bar{S}^c}$ where $N_{\bar{S}^c}$ is equal to the number of sensors in class (c^{th}) . Each sensor is directly connected to a dedicated Computational Units (i.e.: CU) belonging to the set $U = u_l : l = 1, ..., N_\mu$ with N_μ equal to the total number of corresponding sensors. Each CU acquires data providing Object Reports (OR) $\bar{r}_{i,m}^c(k)$ for each object m found at time k from i^{th} sensors in c^{th} class. OR is represented as a multidimensional vector composed by different features related to the detected object:

$$\bar{r}_{i,m}^c(k) = [\bar{f}_1^i(k), \ldots, \bar{f}_{N_r}^i(k)] \tag{4.1}$$

with N_r the total number of features $\{hatf$ in the report. For each detected physical object tracks are instantiated and updated:

$$T_m(k) = \hat{r}_m(K - k) : k = 0, ..., K \tag{4.2}$$

where K is the current time, m = detected object. Tracks are sequences of estimated reports $\hat{r}_m(i)$ derived from integration of heterogeneous ORs:

$$\hat{r}_{i,m}(k) = [\hat{f}_1^i(k), \ldots, \hat{f}_{N_r}^i(k)] \tag{4.3}$$

In the presented case, only two classes of Sensors have been included in the

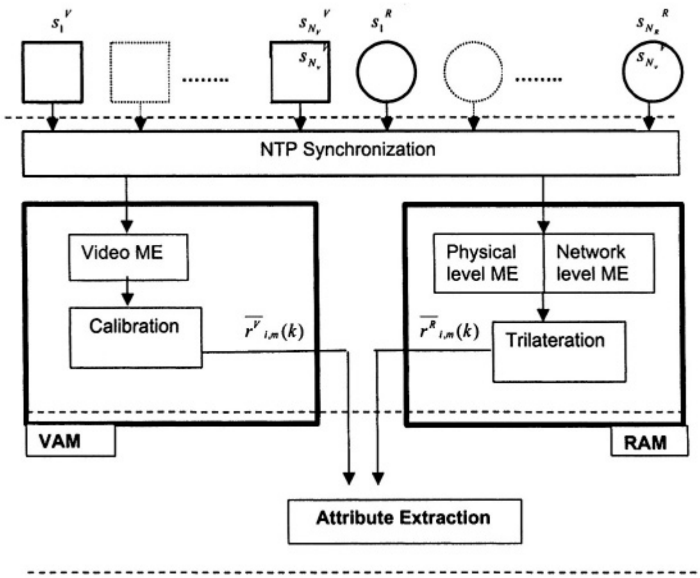

Figure 4.4. Functional architecture of the extraction strategy

architecture: Static CCD Video Cameras and 802.11 WLAN Base Stations (i.e.: classes $c = R, V$). In particular, a Video Analisys Module (i.e.: VAM) takes care of extracting metadata from video sources whereas the Radio Analisys Module (i.e.: RAM) does the same with Base Stations. Output Object Reports are respectively addressed as Video Object Reports (VOR) and Radio Object Reports (ROR)$\bar{r}_{i,m}^{c=V,R}(k)$ shown in Figure 4.4.

5.3 Alignment and Extraction of Video and Radio Object Reports

Introduction. As it can be seen from Figure 4.4, dedicated sub-modules (i.e.: Video ME, Physical Level ME and Network Level ME) are specifically devoted to the extraction of metadata that is coded under the form of an Object Report:

$$\bar{r}_{i,m}^{c}(k) = [\bar{p}^i(k), \bar{id}^i(k), \bar{c}^i(k)] \tag{4.4}$$

where $\bar{p}^i(k), \bar{id}^i(k), \bar{c}^i(k)$, respectively indicates position, id and class (e.g.: pedestrian, vehicle, others) of detected objects moving in the monitored environment.

Video object extraction. Video Objects Reports $\bar{r}_{i,m}^{V}(k)$ are evaluated by Video Metadata Extractors (i.e.: VME) at each timestamp . VME takes as input raw video frames from synchronized grabbers; typically, chain of logical tasks can be assembled in order to process Video data [9], the first step is however a Dynamic Change Detection (see Figure 4.2 right) performing the difference between the current image and a reference one (i.e. background). Each moving area (called Blob) detected in the scene is bounded by a rectangle to which a numerical label is assigned (Figure 4.2 left). Thanks to the detection of temporal correspondences among bounding boxes, a graph-based temporal representation of the dynamics of the image primitives can be built. The core part of such systems is however represented by tracker algorithm, that outputs to the Calibration submodule OR ($\bar{r}_{i,m}^{V}(k)$) with features:

$$\begin{aligned} \bar{f}_1^i(k) &= \bar{p}_1^i(k) = [x_I, y_I] \\ \bar{f}_2^i(k) &= \bar{id}_1^i(k) = [id] \\ \bar{f}_3^i(k) &= \bar{c}_1^i(k) = [c] \end{aligned} \tag{4.5}$$

where (x_I, y_I) are the coordinates (in pixels) of the center of mass in the Image Plane for the m-th object (i.e.: blob) at time k detected by i^{th} sensor whereas the scalars id and c respectively indicates the tracked id (progressive integer number, e.g.: 1,2..) and class of the object (integer number, e.g.: 1=human, 2=vehicle, 3=others).

Video Object Spatial Alignment. Spatial alignment for Video ORs is achieved through Camera Calibration. Camera calibration [40][41] is the process by which optical and geometric features of Video Cameras can be determined (Figure 4.5 shows detection of features). Generally, these features are addressed as intrinsic and extrinsic parameters and they allow estimation of a correspondence between coordinates in the Image Plane (x_I, y_I) and in the $3D$ Real World Space (x_W, y_W, z_W). After the 3-D conversion the last step is represented by the projection on $2D$ Map Plane (x_M, y_M). Various methods

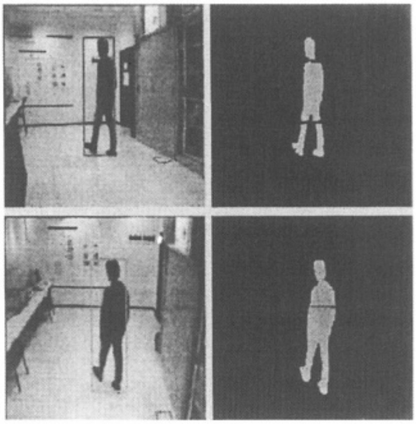

Figure 4.5. Right: Binary Change Detection images showing moving pixels in the current scene. Left: Two views for the available Video Cameras, moving object is highlighted by Bounding Boxes.

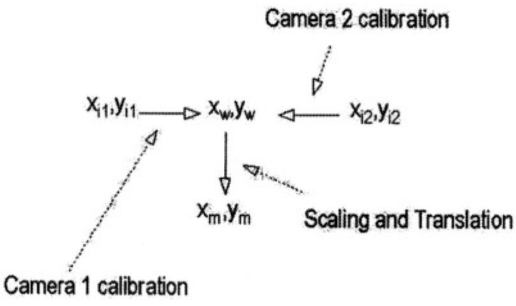

Figure 4.6. Joint calibration methodology.

have been proposed to perform calibration: some uses non-linear optimisation techniques [42], others systems of linear equations. Camera calibration we use is based on classic Tsai method [41]. In the presented system, all video sensors have been calibrated with a common calibration strategy. In Figure 4.6, the chosen approach is outlined. First of all cameras are calibrated on reference images in a unique map then a common reference point has to be found in order to make the system able to switch between the different reference systems. Origin of World Space (x^0_W, y^0_W, z^0_W) represents a good choice because it is common to the all the cameras. The alignment algorithm can be decomposed in the following steps:

1 ImageCoordinates (x_I, y_I) are converted in WorldCoordinates (x_W, y_W, z_W) through calibration.

2 (x_W, y_W, z_W) are converted in Map coordinated (x_M, y_M) with translation and scaling transformations.

3 OR is rewritten as :

$$\bar{r}^V_{i,m}(k) = [\bar{p}^i(k), \bar{id}^i(k), \bar{c}^i(k)] = [x_M, y_M, id, c] \tag{4.6}$$

It is important to note that, in this case,id and c features which should be stationary (i.e.: identity and class of an object do not change over time) values are subject to variations over time due to induced errors in the Change Detection and Tracking steps.

Radio object extraction. The Radio Objects can be defined as those objects which possess electronic wireless communication facilities (e.g.: Bluetooth or WLAN cards). The wireless communication network as a part of the video surveillance system shares the information related to the objects. The Radio Objects Reports (i.e. **ROR**)$\bar{r}^R_{i,m}(k)$ are evaluated by receiving the signals sent by objects device to three base stations (BS). BS as a part of the wireless system are able to communicate with the radio object via receivers and they recognize RORs via their network ID. The system is based on a path-loss model of the signal power transmitted from/to APs and receivers. The observed power is converted into distances using path-loss equations, eq.8 [42][45].

$$S = S_o - 10\,\alpha\,log\frac{d}{d_o} \tag{4.7}$$

Where S is the received power in dB, p_o is the received power at a reference distance (d=1 meter), d is the distance between transmitter and receiver and a is the path-loss exponent. Unfortunately, due to the presence of multi-path fading and noise interference in the environment, the received power is not only dependent on the path loss. Therefore, Equation 4.7 can be represented as Equation 4.8 where the observed power $S + N$, due to fading and path-loss is

$$(S + N) = S_o - 10\,\alpha\,log\frac{d}{d_o} + X_\sigma = S + X_\sigma \tag{4.8}$$

Where random variable X_σ represents the medium-scale fading in the channel and is typically reported to be Gaussian random variable with zero-mean (in dB) and variance σ^2, also represented as $N(0, \sigma)$ [45][48]. The Probability Density Function (p.d.f.) of the received power in eq. (9) is $N(\bar{S}, \sigma)$ with $\bar{S}$ mean and Standard Deviation, σ.

Having estimated *(S + N)* and X_σ, it is possible to compute the distance between transmitter and receiver using Equation 4.8. The distances obtained by the transceiver are tri-laterated [36] to estimate the position (X_M, Y_M) in the common Map Plane and to fill an OR following the policy applied to the video case:

$$\bar{r}^R_{i,m}(k) = [\bar{p}^i(k), \bar{id}^i(k), \bar{c}^i(k)] = [x_M, y_M, id, c] \tag{4.9}$$

In this case the unstable feature is expected to be position whereas identity and class are constant over time.

Radio Object Reports Spatial Alignment. Figure 4.7 shows the logical functioning architecture of the WLAN network for estimation of the position $\bar{p}^i(k)$ using received signal strength (RSS) features. The $RSS_{1,2,3}$ are the received signal strength of the signal at the object device end. The Path-Loss Equation 4.8 uses this information to evaluate the distance d_{123} that is considered to be the distance between the BS and the object. Given the problem of presence of multi-path fading and noise in received signal, and their negative effect on position accuracy, it is desired to enhance the accuracy in two steps:

- At signal level using Pre-Post Cursor Multi-path Mitigator [37].
- At feature level using Feature Function (FF) which is created in the offline phase of the spatial alignment [46].

Spatial alignment and projection in the common 2-D map is performed in two phases:

1 Offline phase

2 Online phase

The offline phase consists of signal strength data collection at several predefined positions in the test site. Based on the RSSobserved (i.e.: collected power strength) at known positions and RSStheoratical,(i.e. the theoretically computed power strength for known distances) the RSSmultipath+noise can be computed as the difference between the above two signal values. The ratio, δ, in Equation 4.10, is obtained for each known distance between transmitter and set of different position. The polynomial fit function, which is FF, is computed based on the collected signal measurements, where:

$$\delta = (RSS_{multipath+noise}/RSS_{observed}) \tag{4.10}$$

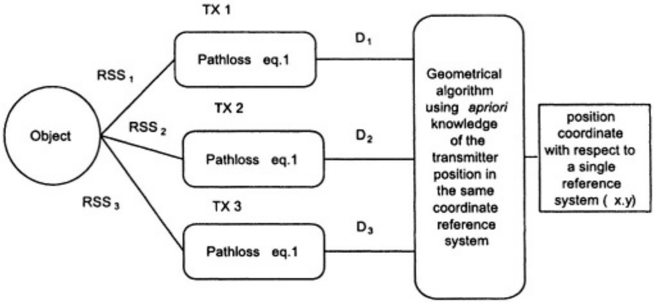

Figure 4.7. Logical functioning architecture of the WLAN Positioning network using RSS features.

The function FF is utilized in an online phase. During online phase, user's signal are collected by each transmitter and 2D position is computed as explained in last section. The details about position method using wlan 802.11b system can be found in [46]. In Figure 4.8, the view of the experimental site is shown. The BS are localized within the map and identified by dots. Three circles are centred in transmitters with radius equal to estimated transmitter distance (i.e.: d_{123}). The overlapping region of the three circles represent the most probable region where the target has to be located.

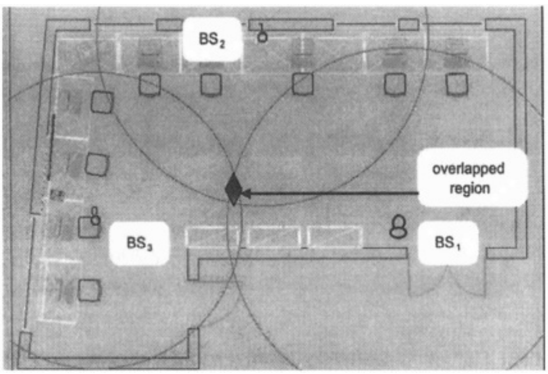

Figure 4.8. View of the experimental site with Base Stations and overlapping region.

6. Results

6.1 The environment

The environment in which the tests took place can be divided in two topologies:

- Experimental test bed from which video object reports have been extracted;
- Simulated test bed from which radio object reports have been extracted.

Once the two results set have been gathered, the heterogeneous location information are sent to a processing unit to start the alignment process.

6.2 Results for video object extraction

The following mock-up architecture has been set-up in the laboratory of Biophysical and Electronic Engineering Department at University of Genoa: two CCD-Video Cameras with 352x288 pixels resolution, 10 fps frame rate and partially overlapped fields of view are used in a $9x7$ meters room. The two cameras are connected to a processing unit which performs the video object extraction. In Figure 4.5 the output screenshots of the Video Metadata Extractor during the tests are reported for the two fields of view. In Figure 4.9 it is possible to see the video object position on the room map. Tracks come from the VME after calibration. As it can be seen two different tracks are available: the first one comes from camera one and the other from the second video sensor.

6.3 Results for radio object extraction

In this part of the result section, the simulation environment to examine the proposed radio object extraction system is described. The system has been developed and simulated in Mathworks Matlab 6.0 environment. The transmitter used in the simulator is using the CCK Coding and DQPSK modulation for the transmitted signal at 11Mbps and it satisfies all the requirements specified in [47]. The multipath fading present in the channel for indoor Line of Sight (LOS) is Rice distributed with AWGN and delay spread of 100 ns [48]. The effect of Doppler spread has been neglected as the terminal is considered to be almost static in indoor ambience. The room used for Video Experimental Trial is simulated. The parameter used in path-loss are n= 1,6-1,8 P0= -20 dBm ± 1 dB. The three transmitters' position are known on the map as shown in Figure 4.5. The signals from three transmitters are generated for a given positions, multipath fading level and noise effect, that is to say a random attenuation with fixed delays, is added to it. At the terminal end, the proposed method at signal and feature level is applied on the radio signals. The results in terms of relative positioning error is obtained as shown in Figure 4.10, moreover as in Figure 4.9, in Figure 4.12 the tracks of user position are plotted for the four different methodologies.

Figure 4.10 shows the error obtained in five simulated positions. The positions are estimated in four cases: without any mitigator (track called crude in

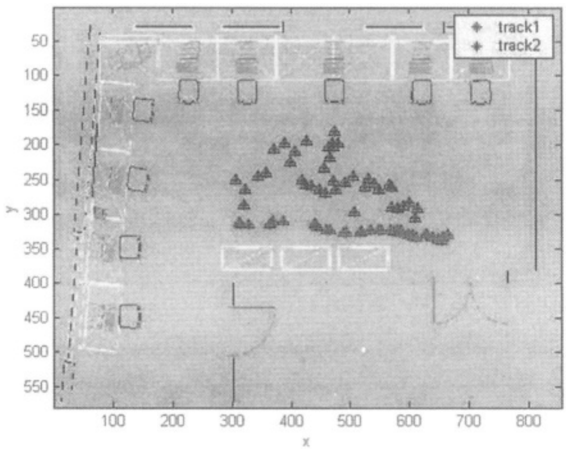

Figure 4.9. 2-D map of the site in which tests took place. Two tracks are aligned and projected for Video Object Reports.

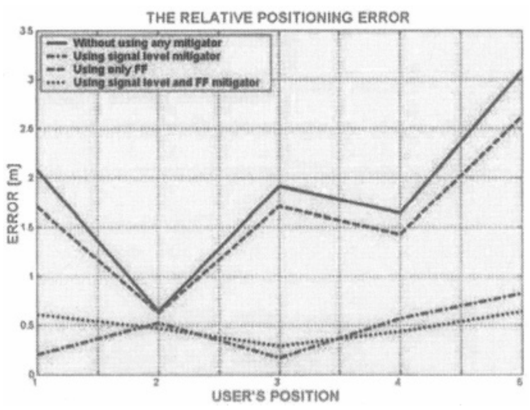

Figure 4.10. The simulation results after utilizing both signal level and FF.

the figure), using feature level mitigator (FF) (called features), using pre-post cursor signal level mitigator (called pre-post) and using both mitigator together (called feature+prepost). As it can be seen in Figure 4.10 that system without using any mitigation is affected by a big positioning error, whereas pre-post cursor mitigation reduces this error to 33%. The FF algorithm is prominently improving the accuracy by 75%. The best positioning accuracy recovered is, however, with the use of both level of mitigator indicated by a 79% result.

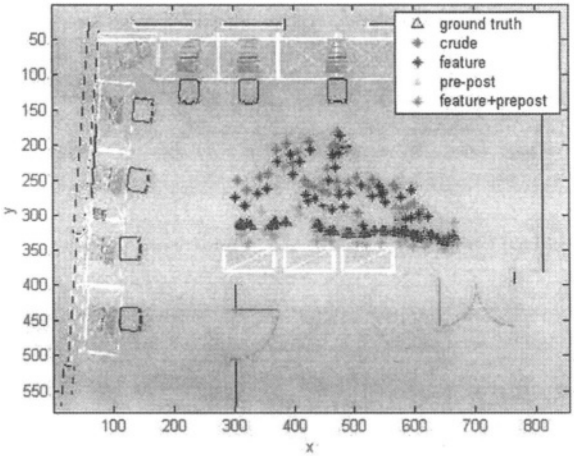

Figure 4.11. The track for Radio Object Reports

6.4 Alignment results

Presented qualitative results are a preliminary attempt to show how Radio and Video data can be aligned in a common ground plane in order to be fused. In addition, error in localization, especially in Radio sensor (in case of both mitigator) and Video sensor is encouraging and it will allow association between the heterogeneous tracks. In Figure 4.12 the results of two heterogeneous sensors are plotted together.

7. Conclusions

Ambient Environment provides the intelligent environment around the object. The efficiency of the AmI is highly dependent upon the context aware information. The context aware information is provided by the multiple sensors co-existing in the environment. The functionality of the AmI is structured into four main task. First, sensing task has the duty to collect data about the environment, data and information that becoming the input of the analysis task which are processed to extract useful and concise higher level metadata, the decision logics employs to choose the best action to undertake relatively to its metrics. Then the chosen action is realized by the action/communication

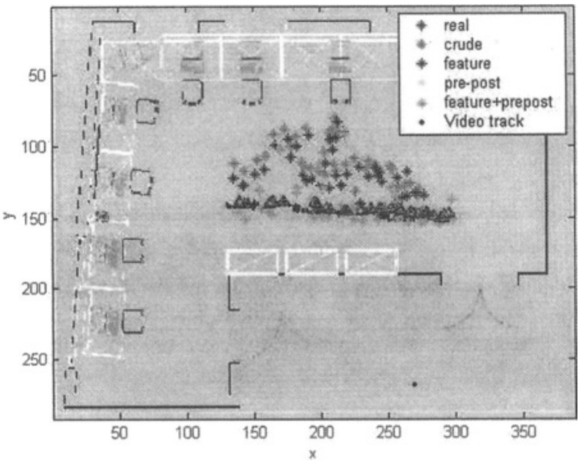

Figure 4.12. 2-D map of the site in which tests took place.

oriented smart space function in order to apply an influence on the user itself. As the context aware information is extracted through sensing task, the context aware system has been described deeply. The location information as a very important context aware information has been explained and described with special attention on the fusion aspect between multi-sensor. The multi-sensors in particular has been considered to be video and radio sensors. The attributes extraction in case of both the sensor has been described and explained in detail. Then based on the extracted feature along with the position information, task of alignment is performed. The results are explained that shows that based on the extracted location attributes it is positioning to do the association and estimation of the precise positioning. The effort is mainly on the shown proof that it is possible to align multi-positioning sensors by extracting the track obtained by each sensors. As a step towards their fusion, this provides the sufficient ground to efficiently fuse the multi-location sensors.

8. Acknowledgments

This work was performed under co-financing of the MIUR within the project FIRB-VICOM and the project Per2.

References

[1] D. Garlan, D. Siewiorek, A. Smailagic, and P. Steenkiste, "Project Aura: Toward Distraction-Free Pervasive Computing", IEEE Pervasive Computing,

April-June 2002.

[2] L. Marchesotti, C. Bonamico, C. Regazzoni, and F. Lavagetto, "Video Processing and Understanding Tools for Augmented Multisensor Perception and Mobile User Interaction in Smart Spaces", International Journal of Image and Graphics, 2004 (to appear).

[3] A. Nijholt, "Disappearing Computers, Social Actors and Embodied Agents", Proceedings of 2003 International Conference on CYBERWORLDS, Singapore, pp. 128-134, 2003.

[4] F. Vatalaro, "Dalle Telecomunicazioni alla Telepresenza Immersiva", Notiziario Tecnico Telecom Italia, anno 11, n°3, Dicembre 2002.

[5] ISTAG, "Scenarios for Ambient Intelligence in 2010". http://www.cordis.lu/istag.htm

[6] M. Trivedi, K. Huang and I. Mikic, "Intelligent Environments and Active Camera Networks", IEEE Transactions on Systems Man and Cybernetics, October 2000.

[7] VICOM project, "Virtual Immersive COMmunications", 2002-2005. http://www.vicom-project.it

[8] L. Marchesotti, L. Marcenaro and C. Regazzoni, "Heterogeneous Data Collection and Representation within a Distributed Smart Space Architecture", Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems), Ghent, Belgium, 2002.

[9] L. Marcenaro, F. Oberti, G.L. Foresti and C.S. Regazzoni, "Distributed architectures and logical task decomposition in multimedia surveillance systems", Proceedings of the IEEE, Vol.89, N.10, Oct.. 2001, pp. 1355 -1367.

[10] E. Waltz and J. Llinas, "Multisensor data fusion" , ISBN 0-89006-277-3, 1990 Artech House, Inc.

[11] A. R. Damasio, "The Feeling of What Happens-Body, Emotion and the Making of Consciousness", Harvest Books, September 2000.

[12] http://oxygen.lcs.mit.edu/

[13] Brooks, R. A. with contributions from M. Coen, D. Dang, J. DeBonet, J. Kramer, T. Lozano-Perez, J. Mellor, P. Pook, C. Stauffer, L. Stein, M. Torrance and M. Wessler, "The Intelligent Room Project", Proceedings of the Second International Cognitive Technology Conference (CT'97), Aizu, Japan, August 1997.

[14] Lorigo, L.M., R.A. Brooks and W.E.L. Grimson, "Visually-Guided Obstacle Avoidance in Unstructured Environments", Proceedings of IROS '97, Grenoble, France, September 1997, pp. 373-379.

[15] http://www.research.philips.com/InformationCenter /Global/FArticleSummary.asp?lNodeId=712/

[16] http://www.cc.gatech.edu/fce/ahri/

[17] Kidd, et al. The Aware Home: A Living Laboratory for Ubiquitous Computing Research, Proceedings of the Second International Workshop on Cooperative Buildings, Position paper, October 1999.

[18] http://carol.wins.uva.nl/porta/ambience/index_en.html/

[19] http://vicom-project.it

[20] F. Vatalaro Dalle Telecomunicazioni alla Telepresenza Immersiva, Notiziario Tecnico Telecom Italia, anno 11, n°3, Dicembre 2002.

[21] http://www.ubiq.com/hypertext/weiser/UbiHome.html

[22] A. Pentland. Looking at people: Sensing for Ubiquitous and Wearable Computing, IEEE Trans. On PAMI, Vol. 22, n. 1, pp. 107-119, January 2000.

[23] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, A System for Video Surveillance and Monitoring, tech. report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

[24] M.A. Neerincx and J.W. Streefkerk, Interacting in Desktop and Mobile Context: Emotion, Trust and Task Performance, European Symposium on Ambient Intelligence, Eindhoven, Netherlands, November 2003.

[25] Jurjen Caarls, Pieter Jonker, Stelian Persa, Sensor Fusion for Augmented Reality, European Symposium on Ambient Intelligence, Eindhoven, Netherlands, November 2003.

[26] Heikki Ailisto, Ville Haataja, Vesa Kyllönen and Mikko Lindholm, Wearable Context Aware Terminal for Maintenance Personnel, European Symposium on Ambient Intelligence, Eindhoven, Netherlands, November 2003.

[27] http://www-prima.inrialpes.fr/Prima/

[28] J. L. Crowley, J. Coutaz, G. Rey and P. Reignier, "Perceptual Components for Context Aware Computing", UBICOMP 2002, International Conference on Ubiquitous Computing, Goteborg, Sweden, September 2002.

[29] J. L. Crowley, "Context Aware Observation of Human Activities", IEEE International Conference on Multimedia and Expo, ICME-2002, Lausanne, Aug 2002.

[30] Bauer, T. and Leake D., WordSieve: A Method for Real-Time Context Extraction. Modeling and Using Context: Proceedings of the Third International and Interdisciplinary Conference, Context 2001, Springer Verlag, 2001, pp. 30-44.

[31] L. Marchesotti, S. Piva, C.S. Regazzoni, "Structured Context Analysis Techniques in a Biologically Inspired Ambient Intelligence System", Proceedings of the IEEE on System Man and Cybernetics, Special Issue on Ambient Intelligence, 2004 (to appear).

[32] J. L. Crowley, "Context Aware Observation of Human Activities", Proceedings of the IEEE International Conference on Multimedia and Expo, ICME-2002, Lausanne, Aug 2002.

[33] L. Marchesotti, G. Scotti and C. Regazzoni "Issues in Multi-camera Dynamic Metadata Information Extraction and interpretation for Ambient Intelligence", Nato Asi 2003, NAREK center of Yerevan University, Tsakhkadzor, Armenia 18-29 August 2003

[34] J. Pascoe, “Adding Generic Contextual Capabilities to Wearable Computers”, in Proceedings of 2nd International Symposium on Wearable Computers, October 1998, pp. 92-99.

[35] E. Waltz and J. Llinas, “Multisensor data fusion” , ISBN 0-89006-277-3, 1990 Artech House, Inc.

[36] Jeffrey Hightower and Gaetano Borriello, “Location Systems for Ubiquitous Computing,” Computer, vol. 34, no. 8, pp. 57-66, IEEE Computer Society Press, Aug. 2001.

[37] K. Pahlavan and P. Krishnamurthy, Principles of Wireless Networks: A Unified Approach, Prentice Hall PTR, 2002.

[38] P Shan E. J. King “Cancel Multipath Interference In Spread Spectrum Communications” Wireless System Design, March 2001 Pg 49-52.

[39] Siddhartha Saha, Kamalika Chaudhuri, Dheeraj Sanghi, Pravin Bhagwat, “Location Determination of a Mobile Device Using IEEE 802.11b Access Point Signals,” IEEE Wireless Communications and Networking Conference (WCNC) 2003 New Orleans, Louisiana, March 16-20, 2003.

[40] Tsai, Roger Y. (1986) “An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, 1986, pp. 364-374.

[41] Tsai, Roger Y. (1987) “A Versatile CameraCalibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses,” IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, August 1987, pp. 323-344.

[42] J.Renno, J. Orwell, G.A. Jones.“ Towards Plug-and-Play Visual Surveillance: Learning Tracking Models”. ICIP02 (III: 453-456) September 2002 Rochester New York.

[43] L. Marcenaro, F. Oberti, G.L. Foresti and C.S. Regazzoni, “Distributed architectures and logical task decomposition in multimedia surveillance systems”, Proceedings of the IEEE, Vol.89, N.10, Oct.. 2001, pp. 1355 -1367.

[44] Luca Marchesotti, Giuliano Scotti and Carlo Regazzoni, “Issues in multicamera dynamic metadata information extraction and interpretation for ambient intelligence”, in press, Yerevan Armenia, NATO ASI 2003.

[45] B. Sklar, “Rayleigh Fading Channels in Mobile Digital Communication Systems Part1: Characterization,” IEEE Communication Magazine, July 1997, page 90-100.

[46] R. Singh, M. Gandetto, M. Guainazzo, C.S. Regazzoni, “A Novel positioning system for static location estimation employing WLAN in indoor environment,” accepted in the 15th IEEE Symposium on Personal, Indoor and Mobile Radio Communications PIMRC, IEEE, April. 2004.

[47] LAN MAN Standards Committee of the IEEE Computer Society: P802.11b, Supplement to IEEE Std 802.11-Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification-Higher Speed

Physical Layer (PHY) Extension in the 2.4GHz band, IEEE Standard Department, 1999.

[48] G. L. Stüber, "Principle of Mobile Communication", Norwell, MA: Kluwer, 1996.