

FROM MULTI-SENSOR SURVEILLANCE TOWARDS SMART INTERACTIVE SPACES

M. Gandetto, L. Marchesotti, S. Sciutto, D. Negroni and C.S. Regazzoni

carlo@dibe.unige.it

DIBE, University of Genoa, Via dell'Opera Pia 11a GENOVA

ABSTRACT

This paper proposes a novel architecture for multisensor data fusion in the context of Ambient Intelligence (AmI). The proposed system integrates a heterogeneous network of sensors with CCD cameras and computational units working together in a LAN. Activities of humans interacting in the monitored area are detected and classified by combining sensors data output with a neural method.

1. INTRODUCTION

Actual trends in Machine Vision go on the direction of systems capable not only to analyze data coming from multiple and heterogeneous sensors, but also somehow to react to appropriate "stimula" in order to show some degree of "intelligence". Systems of this kind, which can understand the state of a particular environment and appropriately react by instantiating a customized communication with people populating the monitored environment go under the name of Ambient Intelligence Systems. ISTAG gives in [1] a more formal definition of Ambient Intelligence (AmI) that points out as it should provide technologies to support human interactions and to surround users with intelligent sensors and interfaces. According to this, Brooks in [2] states that an Intelligent Environment has to make computation "ready-at-hand", putting the computers out into the real world of people more than people into the virtual world. In [3] Starner focuses a key issue of Wearable computing Systems which plays a fundamental role also in Ambient Intelligence: the capability of context sensing. Rhodes in [4] states that given a user and a set of goals, context is represented by a set of features related to the environment which are not created explicitly to be input to the system. According to [5] at least four different areas are involved in the design of AmI architectures: person identification, surveillance/monitoring, 3D methods, and smart rooms/perceptual user interfaces. Examples of this kind of systems are found in literature only at prototypal level, however many projects have been started on developing applications in the Ambient Intelligence environment [8] as well as technologies and skills to integrate "awareness"

(identification and tracking), "intelligence" (adaptivity), and natural interaction. Our vision defines AmI systems as a set of virtual entities that owns three fundamental capabilities: analysis, awareness, interaction [6]. The proposed system is therefore able to process and analyse simultaneously data coming from an heterogeneous network of sensor (virtual sensors, cdd cameras, probes, etc.) combining them in a unique and symbolic representation of what happens in the monitored environment. The added value to a classic data fusion system can be found in the capacity of the system to perceive the context in which it operates, and also to keep track of its current state. From this assumption ones can derive the need of awareness the system has to have. In the proposed case this is achieved exploiting a paradigm (explained in terms of logical architecture in the following paragraphs) based on a neural fusion method that merges observations coming from "state" and "observation" sensors. The last feature of the system, the capacity of interact has to be pursued through the use of pervasive, non-invasive interfaces which have to be spread everywhere into the monitored environment, ideally reaching a state in which they become "transparent" to final users. The paper outline is the following: in section 2-3 an overview of system architecture is given in term of logical and physical architecture whereas in section 4 the fusion method is outlined. In section 5-6 preliminary results are proposed and in section 7 conclusions on presented work are dealt.

2. THE LOGICAL ARCHITECTURE

The logical structure of a AmI system is defined by logical tasks which are necessary for system implementation and their interconnections [6]. The architecture shown in figure 1 is based on the idea of modeling the environment in which the system has to work as an "External world". This external world is for example an intelligent building. In this case the ambient can be characterized by some parameters on the inner state such as temperature of the rooms, the status of the doors and windows (if open or close), the status of the lights, computers in the building (if on or off), the

network loading of the computers, etc. To understand the status of the ambient the system needs sensors able to observe the inner state.

In the building are also present people (interacting with the environment) which can, e.g., open windows or doors switch on or off computers, etc. Sensorial observations bring to the system static and dynamic information about objects interacting or which will interact with system. All this information which arrive from these two different types of sensors (state sensors and observation sensors) have to be fused in order to be able to determine which decision must to be taken by the system. This decision can affect the ambient in two different way:

- Directly with the ambient through some actuator device such as mechanical tool that close or open windows and doors, etc.
- With the people which interact with the system through personal or common communications such as information that appears on the PDA or mobile phone, if the people have got them or through directional speaker, monitors, etc.

The associative memory becomes in this scenario one of the most important part of the system. The scope of this block is to select features in order to reduce the dimensionality of the input feature vector and make the decision process computationally feasible. Furthermore the associative memory has also to considerate the previous decisions (communication and state changing) in order to produce a dynamic behavior of the system by leaving users to perceive an augmented environment focused on their objectives.

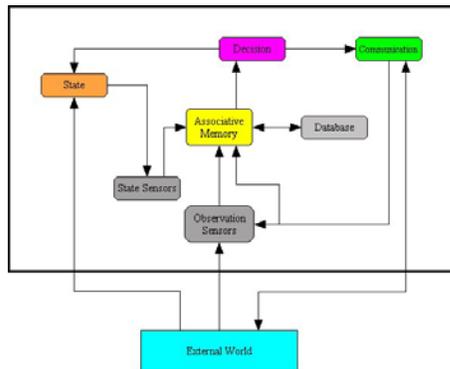


Figure 1. Logical Architecture

3. PHYSICAL ARCHITECTURE AND COMMUNICATION STRUCTURE

As shown in figure 2, physical architecture is made by a sum of sensors communicating together through an agents based network [10]. Four kinds of sensors have been used:

- Video-cameras: using to evaluate the people distribution inside a monitored laboratory.
- PC access: returns the percentage use of laboratory processing capabilities obtained from a simple application that monitors users logins.
- PC network loading: returns the percentage use of laboratory networking capabilities obtained from a simple application monitoring network.
- PC process loading: returns the percentage use of laboratory PC processor capabilities obtained from a simple application.

As already explained, the sum of sensors is divided in two subgroups: “Observation Sensors” and “State Sensors”. The first ones monitor people interacting with the external environment, whereas the second ones realize the system state changes as related to users actions.

In the current implementation the observation sensors are the two cameras, all linked to software agents.

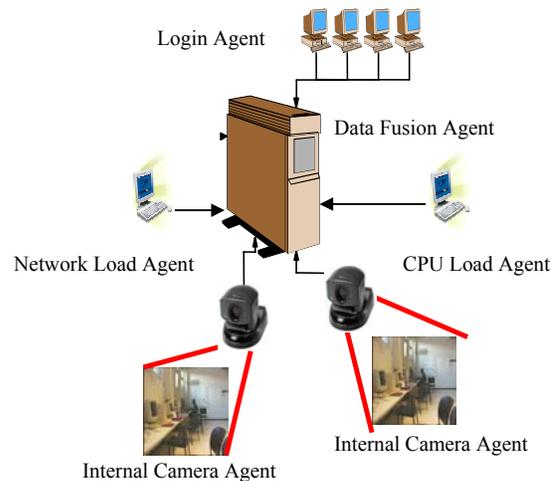


Figure 2. Physical Architecture

The state sensors are agents which allow one to evaluate the CPU load of the computers in the lab, the data load produced by them as well as the one of the entire network. The number of active logins in each computer is also monitored.

The various logical sensors are linked to the central unit, localized in a PC with Data Fusion Agent. This central processing unit is able to collect and store data acquired from the sensors: in particular it receives data from remote processing units and sensors through a standard local area network. The communication channel is characterized by UDP and TCP sockets: in particular, the login controller use a multicast UDP and the network load agent and cameras a TCP line.

4. FEATURE MAPPING AND FEATURES VECTOR

Self-Organizing Maps (SOM) are one of the most important architecture of the neural network type. A SOM performs a spatial organization process of the input features called *Feature Mapping*. This method is realized by a unsupervised learning technique (input and output are not mapped by an external supervisor). The input vector x is compared with the parametric reference vector m_i and the best match is defined as “response”: the input is mapped onto his location. To define the *winner* node c which x belongs to, the smallest of the Euclidean distances $\|x - m_i\|$ is evaluated :

$$\|x - m_c\| = \min(\|x - m_i\|)$$

In our work all the operations can be divided in five steps according to [9]:

1. Map Initialisation: reference vectors are initialised to random values bounded from the minimum and the maximum of the learning set.
2. Map Training: a first phase called *ordering phase* in which the reference vector of the map unit are ordered and a second one where a fine-tuning is performed.
3. Evaluation of Quantization Error using different features vectors.

$$err = \|x - m_c\|$$

4. Map Calibration: the map units are calibrated using know input data sample defined by a label.
5. Map Visualization.
6. Real time work.

The features vector, x_{FV} , used in the previous steps, is defined as below:

$$x_{FV} = [x_P, x_{PC}]$$

where x_P and x_{PC} includes the output of the sensors processed upon the following considerations.

4.1 People distribution

People distribution is a feature extracted by the internal video-camera. The used method is able to track the movements of people which are detected in the monitored scene. Tracking technique used in this case is similar to the one described in [7] and is able to maintain object identities during occlusions by using Kalman filtering and model matching.

The calibration of the camera allows one to obtain the position of any detected object. The features are composed by four float number, f_j , for each camera. The

laboratory is divided in four zones and any number indicates the number of people present in a zone in the laboratory by any camera weighted by the *time of presence* in that zone:

$$f_j = \sum_{i=0}^N k_i n_{ij}$$

where k_i is equal to one if person i is present in the j -th area and n_{ij} is the time (in percentage) in which person i is in the j -th area; N is the total number of people in j -th area and $0 \leq j \leq n$, where n is the number of region in which the laboratory has been divided. The features obtained by the two camera are:

$$x_P = [f_0^1, f_1^1, f_2^1, f_3^1, f_0^2, f_1^2, f_2^2, f_3^2]$$

4.2 PCs usage

This variable takes into account the number of people which after entering the laboratory, actually log into the available PCs, the network and the processor loading for each PC. Two of these ones are integer number, one x_{AL} that indicate the number of login active and the other $x_{\Delta AL}$ to indicate the difference between the number of active login in the previous feature and in the current one. In this feature there are also present three float number $x_{NL}^i, x_{CPU,L}^i, x_{ON,OFF}$ per each PC monitored, to indicate the network loading, the percentage of processor usage, and PC state (switched on or off)

$$x_{PC} = [x_{AL}, x_{\Delta AL}, x_{NL}^i, x_{CPU,L}^i, x_{ON,OFF}]$$

5. TRAINING AND CALIBRATION PROCEDURES

Training session has been performed by collecting 275 a features vectors in an offline mode. The data collection has been performed recording information from sensors for a total period of ten hours; this period has been divided into three subsets to get a sequence in early morning, half day and evening. This procedure has been performed in order to present to the Aml system all the meaningful situations. The whole sequence has been reiterated to have 5000 input in the ordering phase and 20000 for the fine-tuned one.

The next step has been the map calibration in which 680 features vectors, derived from some particular situations, have been collected. In this case only 275 have been manually selected and presented to the system.

6. RESULTS

After the map training and calibration (figure 3) the Aml system infrastructure has been tested in real conditions. The figure 4 represents the U-matrix after the learning phase and it's possible to identify the classes. To represent an event or a sum of events seven super state have been identified and for each one a percentage of false alarm P_{FA} and correct decision P_D has shown in table 1.

Super-State	Real World	P_D	P_{FA}
WHF	Low Human Work	90%	8%
WHL	High Human Work	92%	12%
WAF	Low Machine Work	92%	8%
WAL	High Machine Work	85%	7%
ARRIVE	Laboratory Incomes	85%	10%
Intrusion	Not-Authorized Presence	85%	10%
Empty	Everything Stopped	95%	3%

Table 1.

7. CONCLUSIONS

An Aml system has been described. In particular a general logical structure has been considered for realizing a system able to acquire information from different intelligent sensors and to integrate this information using an associative memory module.

Section devoted to physical architecture description shows the realized Aml system with special attention towards the observation and state sensor and the communication infrastructure. Then the SOM method and the used features have been described and tested.

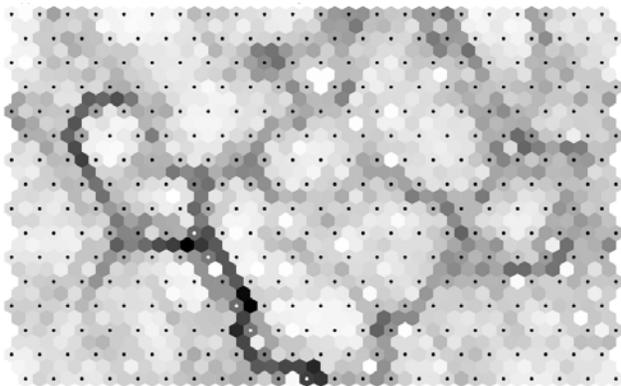


Figure 3. Self Organized Map Layout after training and calibration.

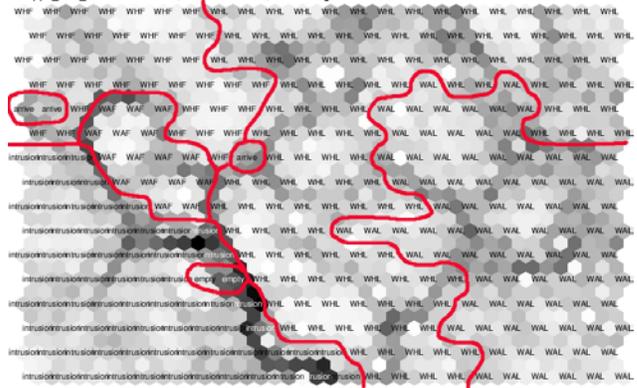


Figure 4. Self Organized Map Layout with Label and Cluster.

8. REFERENCES

- [1] ISTAG Scenarios for Ambient Intelligence in 2010 <http://www.cordis.lu/istag.htm>
- [2] Michael Coen, Brenton Phillips, Nimrod Warshawsky, Luke Weis-man, Stephen Peters and Peter Finin. "Meeting the Computational Needs of Intelligent Environments: The Metaglugue System". In *Pro-ceedings of MANSE '99*. Dublin, Ireland. 1999.
- [3] Thad Eugene Starner, "Wearable Computing and Contextual Awareness" Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MIT June 1999.
- [4] Rhodes, "Personal communication", MIT Media Laboratory, 1999.
- [5] Alex Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2000.
- [6] Luca Marchesotti, Lucio Marcenaro and Carlo Ragazzoni, "Heterogeneous Data Collection and Representation within a Distributed Smart Space Architecture" in *Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems)*, Ghent, Belgium, 2002.
- [7] I. Haritaoglu, D. Hardwood and L.S. Davis, "W4S: a real-time sys-tem or detecting and tracking people in 21/2d", in *Europ. Conf. On Computer Vision*, pp 877-892, 1998.
- [8] Ambience Project www.extra.research.philips.com/euprojects/ambience/
- [9] T. Kohonen, J.Hynninen, J.Kangas, J.Laaksonen, Self-Organizing Map Program Package, University of Technology – Laboratory of Computer and Information Science, Helsinki, Finland, April 1995 .
- [10] L. Marchesotti, P. Remagnino and G.A. Jones, "Evaluating Centralized and Distributed Multi-Agent Architectures for Visual Surveillance Applications", in *Eusipco 2000*, Tampere, Finland, 2000.