Università degli Studi di Bologna

FACOLTÀ DI INGEGNERIA

DOTTORATO DI RICERCA IN INGEGNERIA ELETTRONICA,
INFORMATICA E DELLE TELECOMUNICAZIONI

Ciclo XXII

SISTEMI RICONFIGURABILI A BASSO CONSUMO PER APPLICAZIONI DI MONITORAGGIO DISTRIBUITO

Tesi di Dottorato di:

MICHELE MAGNO

Relatore :

Chiar. mo Prof. Ing. **LUCA BENINI**

Coordinatore:

Chiar. mo Prof. Ing. **PAOLA MELLO**

Settore Scientifico Disciplinare: ING/INF01 Elettronica

Anno Accademico 2009/2010

SISTEMI RICONFIGURABILI A BASSO CONSUMO PER APPLICAZIONI DI MONITORAGGIO DISTRIBUITO

A dissertation submitted to the
DEPARTEMENT OF ELECTRONICS, COMPUTER SCIENCE AND SYSTEMS
OF UNIVERSITY OF BOLOGNA

for the degree of Doctor of Philosophy

presented by
MICHELE MAGNO
born December 9, 1976

March 2010

Keywords

Ambient Intelligence

Wireless Sensor Networks

Video Surveillance Systems

Power Management

Multimodal Surveillance

Acknowledgements

First and foremost, I'd like to thank Professor Luca Benini, whose guidance was essential in my achieving this great undertaking. In our years of working together, he instilled in me a love for research and a passion for learning and striving to better myself. I'd also like to deeply thank Davide Brunelli, whose friendship, ideas, and support made this experience truly fantastic. Thank you to both. Without you two, I probably wouldn't have arrived at this point.

Thanks to all the Micrellab guys, with whom I spent time both inside and outside the lab. Luciano, Martino, Piero, Marco, Betta, Augusto, and all the rest: thank you.

I'd like to acknowledge professor Lothar Thiele, who gave me the chance to work in his group at ETH in Zurichand who also allowed me to meet wonderful people like Kai, Federico, Roman, "Musty", Andreas,Iuliana, Wolfgang, Clemens and Matthias.Also, thank you to Regula, my "crazy" roommate who made my days in Zurich truly amazing. And of course, to all the friends I made there: Martin, Francis, Angelica, Daniel, Madelaine, Elin, Carol, and everyone else.

Infinite thanks to my three muses, who, in these part years, truly warmed my heart. In chronological order: Silvia, who made me a better man; Valeria, who, with your love, swept me away with your warmth; Maria, for your excitement and for your dreams.

Last but certainly not least, I'd like to thank my parents, who always supported the decisions I made and who made sure that I always felt close to home even when I was kilometers awayand to my sister, who, with her children, always brightened up my vacations.

I'm also extremely grateful for all my friends, with whom it was easy to forget those hard workweeks and replace them with wonderful weekends. And also, thanks to my little Sky, who made it possible for both my thesis and these acknowledgments to be written in correct English.

Thank you to all.

Contents

List of Figures

List of Tables

Introduction

The term Ambient Intelligence (AmI) refers to a vision on the future of the information society where smart, electronic environment are sensitive and responsive to the presence of people and their activities (Context awareness). In an ambient intelligence world, devices work in concert to support people in carrying out their everyday life activities, tasks and rituals in an easy, natural way using information and intelligence that is hidden in the network connecting these devices. This promotes the creation of pervasive environments improving the quality of life of the occupants and enhancing the human experience. AmI stems from the convergence of three key technologies: *ubiquitous computing*, *ubiquitous communication* and *natural interfaces*.

Ambient intelligent systems are heterogeneous and require an excellent cooperation between several hardware/software technologies and disciplines, including signal processing, networking and protocols, embedded systems, information management, and distributed algorithms.

Since a large amount of fixed and mobile sensors embedded is deployed into the environment, the *Wireless Sensor Networks* is one of the most relevant enabling technologies for AmI. WSN are complex systems made up of a number of sensor nodes which can be deployed in a target area to sense physical phenomena and communicate with other nodes and base stations. These simple devices typically embed a low power computational unit (microcontrollers, FPGAs etc.), a wireless communication unit, one or more sensors and a some form of energy supply (either batteries or energy scavenger modules). WNS promises of revolutionizing the interactions between the real physical worlds and human beings. Low-cost, low-computational power, low energy consumption and small size are characteristics that must be taken into consideration when designing and dealing with WSNs.

To fully exploit the potential of distributed sensing approaches, a set of challenges must be addressed. Sensor nodes are inherently resource-constrained systems with very low power consumption and small size requirements which enables than to reduce the interference on the physical phenomena sensed and to allow easy and low-cost deployment. They have limited processing speed, storage capacity and communication bandwidth that must be efficiently used to increase the degree of local ”understanding” of the observed phenomena.

A particular case of sensor nodes are video sensors [88, 98]. This topic holds strong interest for a wide range of contexts such as military, security, robotics and most recently consumer applications. Vision sensors are extremely effective for medium to long-range sensing because vision provides rich information to human operators. However, image sensors generate a huge amount of data, which must be heavily processed before it is transmitted due to the scarce bandwidth capability of radio interfaces. In particular, in video-surveillance [49], it has been shown that source-side compression is mandatory due to limited bandwidth and delay constraints. Moreover, there is an ample opportunity for performing higher-level processing functions, such as object recognition that has the potential to drastically reduce the required bandwidth (e.g. by transmitting compressed images only when something ‘interesting‘ is detected) [94]. The energy cost of image processing must however be carefully minimized.

Imaging could play and plays an important role in sensing devices for ambient intelligence. Computer vision can for instance be used for recognising persons and objects and recognising behaviour such as illness and rioting. Having a wireless camera as a camera mote opens the way for distributed scene analysis. More eyes see more than one and a camera system that can observe a scene from multiple directions would be able to overcome occlusion problems and could describe objects in their true 3D appearance. In real-time, these approaches are a recently opened field of research .

In this thesis we pay attention to the realities of hardware/software technologies and the design needed to realize systems for distributed monitoring, attempting to propose solutions on open issues and filling the gap between AmI scenarios and hardware reality. The physical implementation of an individual wireless node is constrained by three important metrics which are outlined below.

Despite that the design of the sensor network and its sensor nodes is strictly application dependent, a number of constraints should almost always be considered. Among them:

- Small form factor to reduce nodes intrusiveness.
- Low power consumption to reduce battery size and to extend nodes lifetime.
- Low cost for a widespread diffusion.

These limitations typically result in the adoption of low power, low cost devices such as low power microcontrollers with few kilobytes of RAM and tenth of kilobytes of program memory with whom only simple data processing algorithms can be implemented. However the overall computational power of the WNS can be very large since the network presents a high degree of parallelism that can be exploited through the adoption of ad-hoc techniques. Furthermore through the fusion of information from the dense mesh of sensors even complex phenomena can be monitored.

In this dissertation we present our results in building several AmI applications suitable for a WSN implementation. The work can be divided into two main areas:*Low Power Video Sensor Node and Video Processing Alghoritm* and *Multimodal Surveillance* .

Low Power Video Sensor Nodes and Video Processing Alghoritms
In comparison to scalar sensors, such as temperature, pressure, humidity, velocity, and acceleration sensors, vision sensors generate much higher bandwidth data due to the two-dimensional nature of their pixel array. We have tackled all the constraints listed above and have proposed solutions to overcome the current WSN limits for Video sensor node. We have designed and developed wireless video sensor nodes focusing on the small size and the flexibility of reuse in different applications. The video nodes target a different design point: the portability (on-board power supply, wireless communication), a scanty power budget (500mW), while still providing a prominent level of intelligence, namely sophisticated classification algorithm and high level of reconfigurability. We developed two different video sensor node: The device architecture of the first one is based on a low-cost low-power FPGA+microcontroller system-on-chip. The second one is based on ARM9 processor. Both systems designed within the above mentioned power envelope could operate in a continuous fashion with Li-Polymer battery pack and solar panel. Novel low power low cost video sensor nodes which, in contrast to sensors that just watch the world, are capable of comprehending the perceived information in order to interpret it locally, are presented. Featuring such intelligence, these nodes would be able to cope with such tasks as recognition of unattended bags in airports, persons carrying potentially dangerous objects, etc., which normally require a human operator. Vision algorithms for object detection, acquisition like human detection with Support Vector Machine (SVM) classification and abandoned/removed object detection are implemented, described and illustrated on real world data.

Multimodal surveillance
In several setup the use of wired video cameras may not be possible. For this reason building an energy efficient wireless vision network for mon-

itoring and surveillance is one of the major efforts in the sensor network community. Energy efficiency for wireless smart camera networks is one of the major efforts in distributed monitoring and surveillance community. For this reason, building an energy efficient wireless vision network for monitoring and surveillance is one of the major efforts in the sensor network community. The Pyroelectric Infra-Red (PIR) sensors have been used to extend the lifetime of a solar-powered video sensor node by providing an energy level dependent trigger to the video camera and the wireless module. Such approach has shown to be able to extend node lifetime and possibly result in continuous operation of the node.Being low-cost, passive (thus low-power) and presenting a limited form factor, PIR sensors are well suited for WSN applications. Moreover techniques to have aggressive power management policies are essential for achieving long-term operating on standalone distributed cameras needed to improve the power consumption. We have used an adaptive controller like Model Predictive Control (MPC) to help the system to improve the performances outperforming naive power management policies.

0.1 Thesis Organization Outline

The reminder of the dissertation is organized as follows.

Chapter 1 introduces the basic concepts of Ambient Intelligence (AmI). It provides a general definition of the main building blocks and defines the critical factors common to AmI applications. Several example AmI projects are presented to provide an insight into the current research in this field.

Chapter 2 describes WSNs. This chapter highlights the characteristics of WSN and the main application scenarios. A more detailed description of the building block of a WSN, the *Wireless Sensor Node*, is provided together with an overview of the state of the art of such devices.

Chapter 3 afterwards a Support Vector Machine (SVM) overview, we present our work in developing video sensor notes and video processing techniques, finally a little introduction on the low-cost, low-power Pyroelectric InfraRed (PIR) .

Chapter 4 demonstrates how such sensors can be integrated within a video surveillance network to augment its performance and to overcome some limitations of the video systems. Moreover, the chapter describes how PIR sensors can be used in conjunction with Wireless Video Sensor Nodes (WVSN) and photovoltaic energy harvesting modules to extend node lifetime using power management policies and Model Predictive Control.

Conclusions conclude the dissertation summarizing the results presented in this thesis.

Chapter 1

Ambient Intelligence

1.1 Ambient Intelligence: general definitions

In the AmI vision, humans will be surrounded by smart devices embedded in everyday objects such as furniture, clothes, vehicles, roads and smart materials. Devices are aware of human presence and activities, take care of his needs and are capable of responding intelligently to spoken or gestured indications of desire. Furthermore they are unobtrusive, often invisible: nowhere unless we need them. Interaction should be relaxing and enjoyable for the citizen, and not involve a steep learning curve [157].

The ISTAG (Information Society Technology Advisory Group) is a team that has been set up to advise the European Commission on the overall strategy to be followed in carrying out the IST thematic priority under the European framework programme for research. The ISTAG reflects and advises on the definition and implementation of a coherent policy for research in ICT in Europe. This policy should ensure the mastering of technology and its applications, and should help strengthen industrial competitiveness and address the main European societal challenges [78].

The first ISTAG meeting took place in 1999 and defined the objective of the group as

> *start creating an ambient intelligence landscape (for seamless delivery of services and applications) in Europe relying also upon testbeds and open source software, develop user-friendliness, and develop and converge the networking infrastructure in Europe to world-class*
>
> — ISTAG, "Orientations for Workprogramme 2000 and beyond"

The ISTAG promotes the creation of pervasive environment improving the quality of life of the occupants and enhancing the human experience. Such smart, electronic environment are proactive to the presence of people and their activities. *Context awareness* is a key factor of this vision. Computer react based on their environment. Devices collect information about the circumstances under which they operate and react accordingly [131, 132].

Ambient Intelligence stems from the convergence of three key technologies:

Ubiquitous Computing

The vision of ubiquitous computing emerged in the late 80s at Xerox Palo Alto Research Center (PARC) when a heterogeneous group of researcher developed a novel paradigm of interaction between human and computers [156]. The term *ubiquitous computing* has been forged by Mark Weiser few years later [154] and refers to omnipresent computers that serve people in their everyday lives at home and at work, functioning invisibly and unobtrusively in the background and freeing people to a large extent from tedious routine tasks. Ubiquitous computing has as its goal the enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user [155]. The technology required for ubiquitous computing is threefold: cheap, low-power electronic devices, a network that ties them all together, and software systems implementing ubiquitous applications. Human-smart environment interaction is possible through hand held devices that collect information from the environment or context aware services that are aware of people presence, understand their activities and react in a proactive manner. Some people say that ubiquitous computing is the Third Wave of Computing, where the First Wave was many people, one computer (mainframe), the Second Wave is the era of one person, many computers (Personal Computers). The Third Wave will be the era of many computers per person [16] (see figure 1.1).

Ubiquitous Communication

An important factor to fully exploit the power of ubiquitous system and to provide information everywhere it is needed is the presence of a rich wired and wireless communication infrastructure. Wireless communication is well suited for dynamic environment where the users moves within smart ambients. In order to realize demands for ubiquitous communication and pervasive computing, a change from the traditional approach of centralized, planned wireless communication networks such as GSM, toward an adaptive, self-organizing, multi-user, multi-system distributed wireless communications platform is essential [119] (see figure 1.2). To implement wireless technology on a wide level, however,

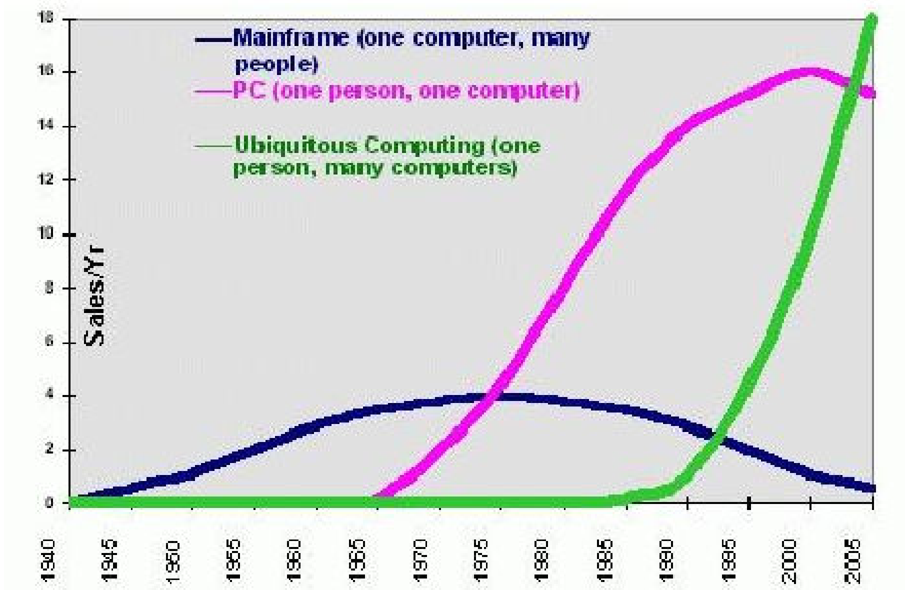


Figure 1.1: Trends in computing

the wireless hardware itself must meet several criteria on the one hand, while easy integration and administration as well as security of the network must be ensured on the other. Some of the unique features that the ambient intelligence scenario presents and that must be considered are: very large networks (hundred or thousands of nodes), both mobile and fixed nodes, node failure must be kept in mind, small battery size (for easier integration) and data centric communication (i.e. redundant data can be aggregated, compressed, dropped etc.). Incorporating these unique features into protocol design is important in order to efficiently utilize the resources of the environment [117].

Intelligent User Friendly Interfaces

Intelligent user interface have a fundamental role in ambient intelligence. These interfaces go beyond the traditional keyboard, mouse, and display paradigm to improve human computer interaction by making it more intuitive, efficient, and secure. Thus, Ubiquitous computing inspires application development that is off the desktop. In addition to suggesting a freedom from well-defined spaces, this vision assumes that physical interaction between humans and computation will be more like the way humans interact with the physical world. Input has moved beyond the explicit nature of textual input (keyboards) and selection (pointing devices) to a greater variety of data types. This has resulted in not only a greater variety of input technologies but also a shift from explicit means

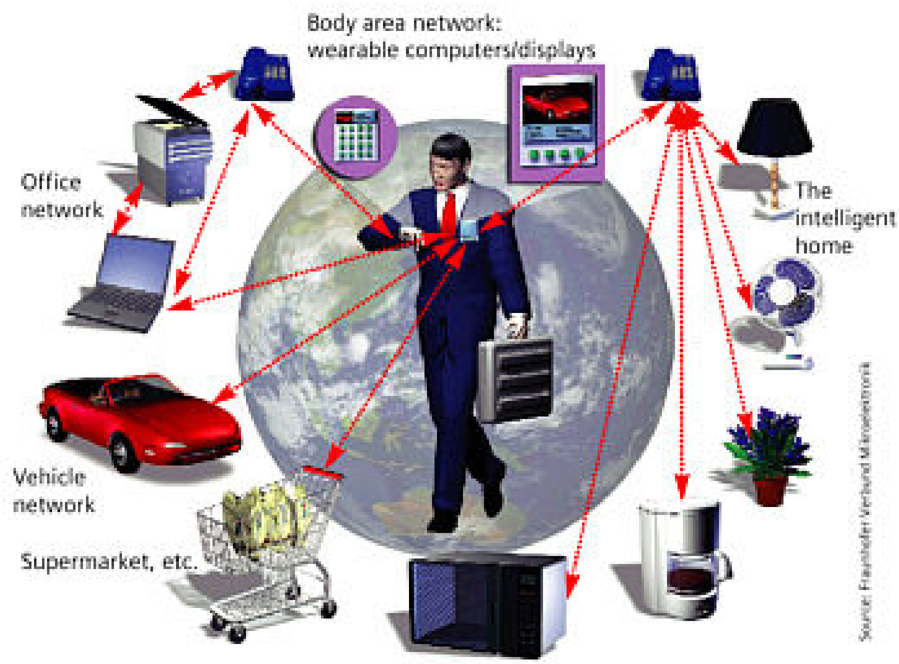


Figure 1.2: Distributed communication network

of human input to more implicit forms of input. Computer interfaces that support more natural human forms of communication (such as handwriting, speech, and gestures) are beginning to supplement traditional interfaces. Intelligent human computer interaction promises to support more sophisticated and natural input and output, to enable users to perform potentially complex tasks more quickly, with greater accuracy, and to improve user satisfaction.

In 2001, two years later the first meeting, the ISTAG group has published a final report where four scenarios are described in order to offer provocative glimpses of futures that can be realized [52]. Each scenario contains positive and negative aspects that allow for a composite, even contrasted, picture of the future.

The analysis of these scenarios allow to identify the critical factors in building AmI systems. The factors are divided into 3 main topics.

Socio-political factors AmI should facilitate human contact and be oriented toward community and cultural enhancement. However to be acceptable AmI should inspire trust and confidence and thus needs to be driven by humanistic concerns, not technological ones since people do not accept everything that is technologically possible and available [114]. A major criticism came from the observation that being immersive, personalized,

Figure 1.3: Intelligent Natural Interfaces (*Photo:Philips*)

context-aware and anticipatory it brings up social, political and cultural concerns about the loss of privacy, the power concentration in large private companies and fear for an increasingly individualized, fragmented society [159]. This criticism should be kept in mind for a widespread acceptance of this new technology.
AmI also should exploit its great potential to enhance education and learning. Everyday life skills will grow because of rising opportunities and means of personal expression and interaction [51].

Business and industrial models Economic aspects of AmI are a fundamental factor for the diffusion of this technology. The most important questions are related to how translate technological and social changes into potential business models. However a number of elements emerged from the scenario that highlight several potentialities of AmI. Among them: enhancements in the productivity and the quality of products and services, comprehensive methods of monitoring and extracting information on real-world, reducing reaction times in unforeseen circumstances, new products and new services.

Technology requirements Five main technology requirements emerge from the analysis of the scenarios [52]:

1. Very unobtrusive hardware. Miniaturization is necessary to achieve dense dissemination of devices and to develop new sensors and

smart materials. In addition self-generating power and micro-power usage will be necessary due to poor scaling capability of batteries technology and new displays and smart surfaces should be developed to provide satisfactory interaction with the environment.

2. A seamless mobile/fixed communications infrastructure. Complex heterogeneous networks need to function and to communicate in a seamless and interoperable way. This implies a complete integration of mobile and fixed and radio and wired networks. Advanced techniques for dynamic network management will be necessary.
3. Dynamic and massively distributed device networks. A huge amount of sensors will be spread in the environment. This networks should be self configurable according to its specific, dynamic status and the current task with variable actors and components. Databases should be accessible on demand from anywhere in the system.
4. Natural feeling human interfaces. The design of novel multimodal, multi-user, and multi purpose interface for speech, gesture, and pattern recognition adaptive to user requirements is required.
5. Dependability and security. Technology should be safe for user both from the physical and psychological point of view. Thus technology should be tested and both hardware and software should be robust. For this reason there is likely to be an emerging emphasis on self-testing and self-organizing systems.

Ambient Intelligence will be brought to us with the promise of an enhanced and more satisfying lifestyle. However, its social benefits cannot be realized unless a number of requirements regarding socio political-issues, business model and technology development have been met. Several field of research will be involved in this change and furthermore novel interdisciplinary approaches will be necessary. Issues such as environmental and social sustainability, privacy, social robustness and fault tolerance will determine the take up of AmI.

1.2 Ambient Intelligence projects

A number of leading technological organizations are exploring pervasive computing apart from Xeroxs Palo Alto Research Center (PARC).

The Laboratory for Computer Science (LCS), the Artificial Intelligence Laboratory (AIL) at the Massachusetts Institute of Technology (MIT) together with several industrial partner have started the project *Oxygen* [108]. The mission of the project is to *bring an abundance of computation and communication within easy reach of humans through natural perceptual interfaces of speech and vision so* *computation blends into peoples lives enabling them to easily do tasks they want to do collaborate, access knowledge, automate routine tasks and their environment*. The project focus on *network technologies* to connect dynamically changing configurations of self-identifying mobile and stationary devices to form collaborative regions, on *software technologies* to develop software systems able to adapt to users, to the environment, to change and to failure with minimal user intervention and without interruption to the services they provide, on *perceptual technologies* to build multimodal interaction with the electronic environment, and on *user technologies* for user support.

IBM created a living laboratory, called Planet Blue, to understand how people will interact with the emerging world of the wireless Internet [76]. The applications developed within this laboratory aim at highlight the requirements of the underlying infrastructure needed to support workers. The objective of Planet Blue is to define the future of post-PC personal computing and drive IBM's research in information access devices. The project focus on the development of dynamic personal portals, enhanced Personal Information Management (PIM) and smart meetings.

Carnegie Mellon University has started *Project Aura* that focuses on user attention [34]. The project motivation come from the observation that also user attention is a (limited) resource in a computer system. Aura's goal is to provide each user with an invisible halo of computing and information services that persists regardless of location and support it. Aura's related project includes: distributed real-time object system and interactive media, mobile file access, application-aware networking, wearable computers and cognitive assistance for everyday computing.

Chapter 2

Wireless Sensor Networks

2.1 Wireless Sensor Networks overview

Advances in the fields of micro electronics, wireless communication, embedded microprocessors and micro-fabrication allowed the the birth of one of the most rapidly evolving research and development fields: *Wireless Sensor Networks* (WSN) [44, 167]. WSN are complex system consisting of spatially distributed autonomous devices, called *Sensor Nodes*, that collaborate to monitor physical or environmental conditions at different locations. Design, implementation, and deployment of a WSN involves a wide range of disciplines and considerations for numerous application-specific constraints [20]. In the last five years, significant progress has been made in the development of WSNs, and some WSN-based commercial products have already appeared on the market.

Even if WSN are strictly application dependent, it is possible to define a list of basic features [77].

- Self-organizing capabilities.
- Short-range broadcast communication and multihop routing.
- Dense deployment and cooperative effort of sensor nodes.
- Frequently changing topology due to fading and node failures.
- Limitations in energy, transmit power, memory, and computing power.

These characteristics make WSN different from other wireless systems and make them one of the most important enabling technologies for several applications.

2.1.1 Wireless Sensor Network applications

Historically WSNs were developed for military applications [36], however there has been a significant interest also in several other fields of human activities [122]. Following a list of application is discussed.

Military

Being capable of self organization a large number of sensor nodes could be rapidly deployed along defensive perimeter or into battlefields (for example by dropping them from a helicopter as shown in figure 2.1). Once on the field they would establish an ad hoc network and monitor for hostile military units. For example in [105] a wireless network of many low-cost acoustic sensors is used to determine both a snipers's location and the bullet's trajectory. Furthermore even if the loss of some sensors is likely to happen the ability to adapt to a changing topology will not prevent a redundant network to work properly. Clearly, fusing the information from a heterogeneous set of sensors can improve the precision and the number of inferences about the activity going on [72].

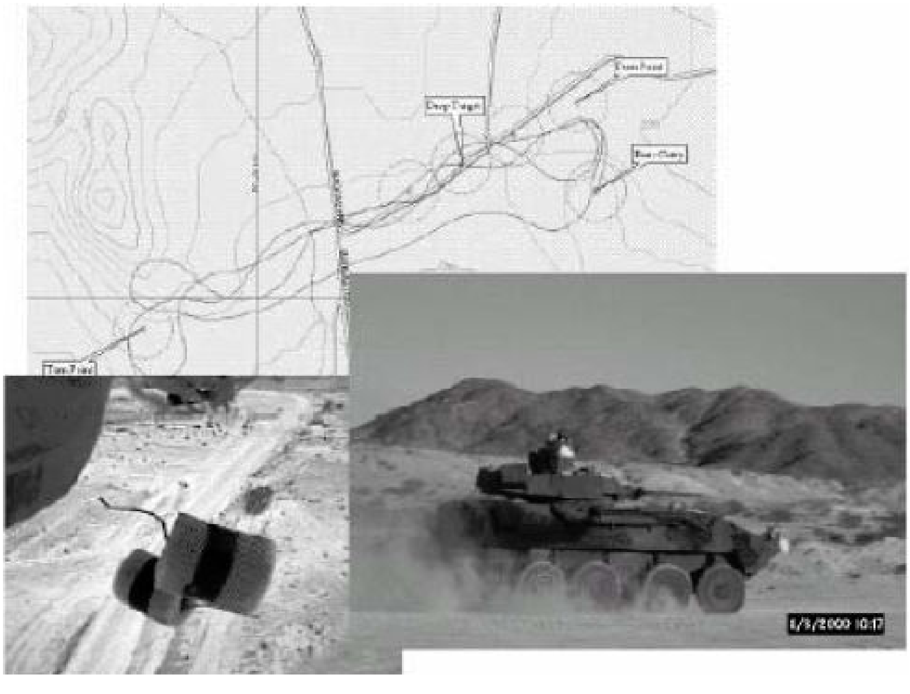


Figure 2.1: WSN Application on battlefield

Environmental and habitat monitoring

WSN have shown to provide an effective means to monitor geographically remote areas. Thanks to the ability of transmit collected data to a data repository on a server, WSNs have been a great improvement in traditional monitoring systems where data required manual downloading by a maintenance team [106]. Some applications of environment monitoring through WSN include the the Environmental Observation and Forecasting Systems (EOFS) project which is large-scale distributed system designed to monitor, model, and forecast wide-area physical processes such as river systems like the Columbia river estuary [142] and the Sensor Web Project [113] which is a systems used to implement a global surveillance program to study volcanoes. The system uses a network of sensors linked by software and the internet to a satellite and has been designed with a flexible, modular, architecture to facilitate expansion in sensors, customization of trigger conditions, and customization of responses. Examples of WSNs applications for habitat monitoring include the Berkeleys habitat modeling at Great Duck Island [143] (see figure 2.2).

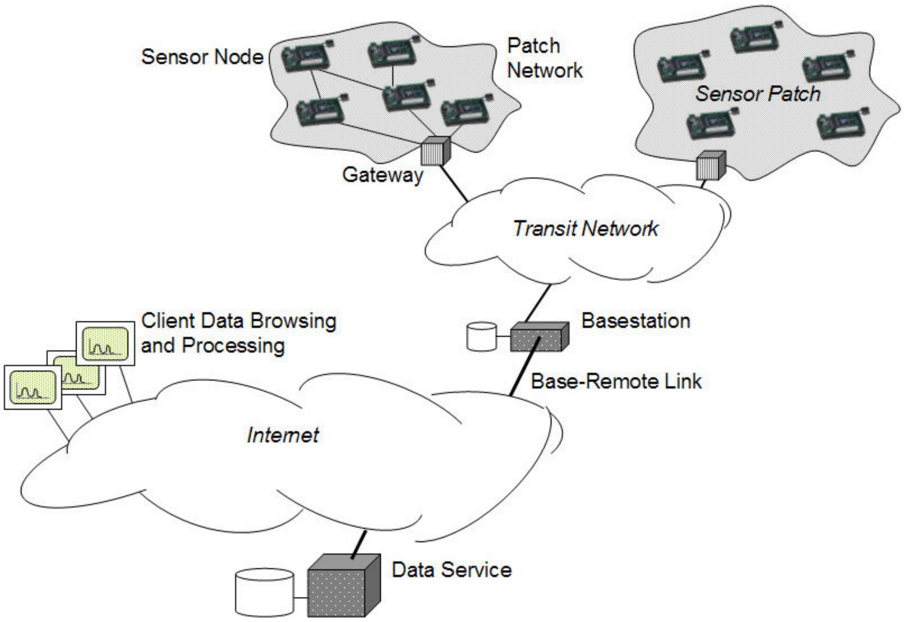


Figure 2.2: Structure of the WSN for habitat monitoring on Great Duck Island

Health care

Patient monitoring systems can be used to collect patient physical status related data at home and, in some cases, in outdoor scenarios, facilitate disease management, diagnosis, prediction and follow-up. Use of WSN can bring great benefit to this activity since the monitoring of people in their natural environments is not practical when it is necessary to use cables to connect the sensors with the processing and communication units [109]. Some example application includes elderly care [25], post stroke rehabilitation [115] and support of people who suffer of physical disability in order to provide imminent feedbacks when occurs [29] (see 2.3).

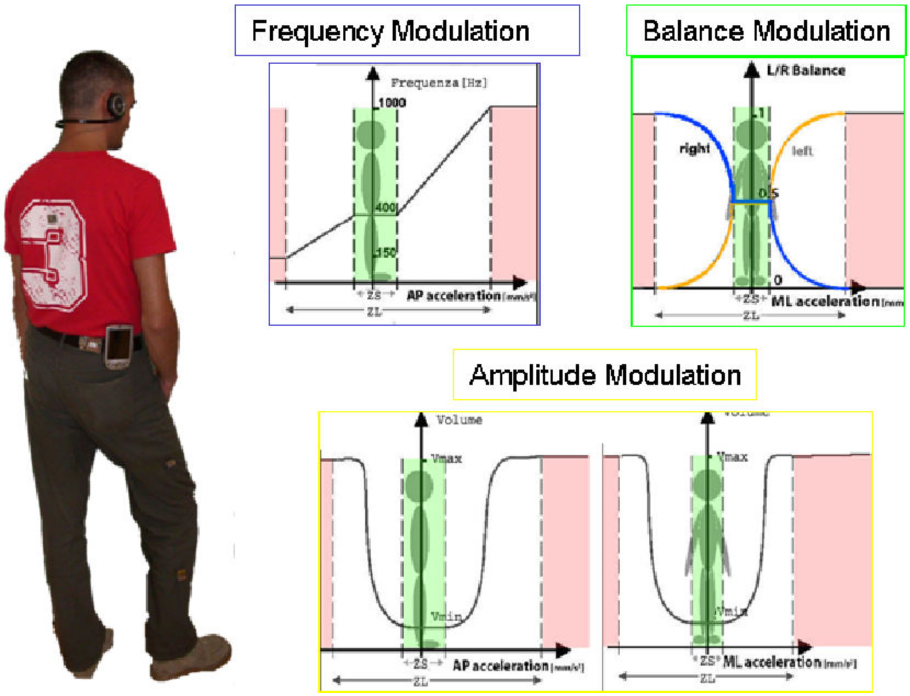


Figure 2.3: Audio bio-feedback for impaired people support

Domotic

Home automation is a field within building automation that focus on the application of automatic techniques for the comfort and security of home residents. The possibility to embed a large number of sensors into everyday objects allow the continuous monitoring of the home status. This results in a more efficient tuning of systems such as the heating, ventilating, and air conditioning (HVAC) and the easy and natural interface with electronic devices [120].

Logistic

Tracking of goods is one of the most important aspect for modern companies. In a globalized world, production process is distributed among several country and many actors take part of it. WSN provide opportunities for the control and management of transport and logistics processes, since sensor nodes can be associates with goods and track their path, who used them and eventually report misuse. An overview of issues and possible approaches can be found in [57].

Figure 2.4: WSN can be used for logistic support

Surveillance

As for military application WSN can be used to monitor the access to building, restricted areas and other critical infrastructure such as power and telecoms grids or roads and motorway. Heterogeneous systems that comprise lower-cost sensors, such as presence or acoustic sensors, can support more bulky and expensive sensors such as imagers, in order to provide cost effective and efficient systems. The use of this setup is even more effective if we consider that it is rather difficult for security guards to continuously watch a set of video monitors when most of the time nothing occurs is considered. Thus low-cost sensor can help to focus their attention only where it is necessary [168].

2.2 Wireless Sensor Nodes

WSN basic building blocks are called *Wireless sensor nodes* or *sensor nodes*. A sensor node is a device capable to collect data from one or more sensors, perform some sort of computation with it, than (wirelessly) send this data to other nodes or system for further analysis.

The major characteristics and requirements of a sensor node can be listed in the following [128]:

Low cost

WSN may consist of hundred or thousand of sensor nodes, thus single sensor node cost should be kept low. Also, it is likely that sensor node will be embedded into everyday object, therefore, for a widespread diffusion of sensor network, their cost should not be excessive.

Low cost requirement results in the adoption of low level components such as low power microcontrollers with limited amount of data and program memory available. As a consequence, even if, due to the high number of nodes working in parallel within the network, the overall computational power and memory available to the network can be quite high, single node capabilities are strictly limited. Thus, application for WSN should be made up of many simple tasks done in parallel by the nodes of the network.

Limited size

Sensor nodes will be embedded into the surroundings, into object and even into user garments. For this reason, unobtrusiveness is a critical point in order not to impair normal activities. A consequence of miniaturization is the evolution of sensor nodes from dedicated embedded devices where commercial off the shelf components with emphasis on small form factor, low-power processing and communication, share a common board to system on chip sensor nodes where on a common die coexist an MCU, a wireless transceiver and sensors.

Low power

Power consumption is one of the biggest issues in the design of WSNs. Nodes, typically, are equipped with batteries, thus they have a limited amount of available energy. Often a frequent change of batteries can be unfeasible, specially in large WSN, or can not be possible when, for example, nodes are placed in harsh environment. In many application scenarios, the target node lifetime should be several years long. This imposes drastic constraints on power consumption that can drop down to an average of few tenth of microwatts.

Limited power consumption usually is achieved using low power hardware or performing several trade off between the energy consumption and other network characteristics such as: quality of service, latency, sensing accuracy, reactiveness to changes in topology, node size (since batteries do not scale as quickly as integrated circuits).

Another approach is to rely on energy scavenging systems to extend node lifetime. However energy harvesting, typically, provide a non constant amount of energy that must be carefully managed to assure the desired service.

Wireless

Wireless is a key factor for many applications that rely on mobile nodes, and in order to reduce WSN cost. In fact, sensor nodes, even if fixed, may be placed in environment where communication infrastructure are not present. In this situation the cost of wiring sensor nodes can be too high and result in sensor network rejection.

Scalability and self organization

Wireless sensor nodes should be able to autonomously organize themself and to adapt to changes in their setup and number. This characteristic is fundamental since often WSN are deployed without a precise control of nodes position (for example, when dropped on battle field) and also because, due to the low cost hardware used, nodes failure can be rather common. For this reason sensor network should be able to provide a graceful degradation as the number of nodes decrease. Furthermore, self organization is necessary where mobile nodes move within different regions and interact with a multitude of different other nodes.

Figure 2.5 presents the system architecture of a generic sensor node which, typically, is made up of four basic building blocks.

- Sensing Unit.
- Computational Unit.
- Communication Unit.
- Power Unit.

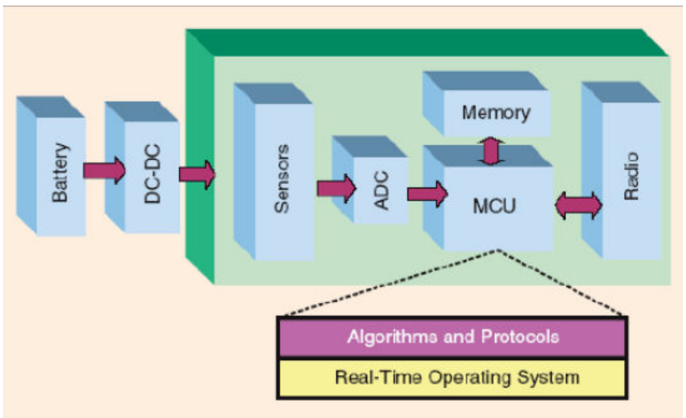


Figure 2.5: Generic architecture of a sensor node

An example of wireless sensor node is presented in figure 2.6 [60].

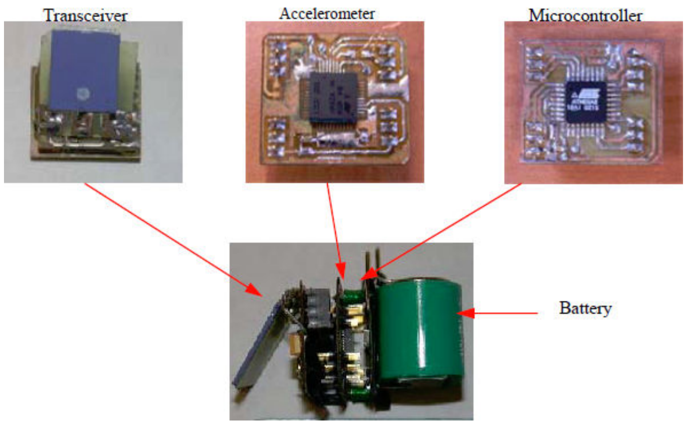


Figure 2.6: WiMoCA wireless sensor node

2.2.1 Computational Unit

Sensor nodes should collect data from the environment, process it and communicate. For this reason a central processing unit is needed. The CPU should be able to manage the sensor node activity while meeting the energy consumption, size and cost constraints. There are a large number of available microcontroller, microprocessors and FPGA that can be integrated within sensor nodes, which allow a high degree of flexibility [150, 15].

Microcontrollers

Nowadays, microcontroller includes a sufficient amount of memory and enough computational power to iterate with sensors and communication devices such as short-range radio to compose a sensor node. Furthermore they provide non-volatile memory for data storage and several other devices such as: ADC, UART, SPI, counters and timers.

There are many types of microcontrollers, ranging from 4 to 32 bits, varying the number of timers, bits of ADC and power consumption. In particular they provide several different operating modes that allow to save energy when the sensor node is idle.

FPGA

Field Programmable Gate Array (FPGA) presents some disadvantages with respect to microcontrollers. The most important is related to power consumption, which is not as low as microcontrollers one. However the development of ultra low power FPGA can make these devices a suitable solution for sensor node.

2.2.2 Sensor and Actuator Unit

A *sensor* is a device that converts a physical phenomenon into an electrical signal. On the other hand, an *actuator* convert an electrical signal into physical phenomena. The first decade of the 21^{st} century has been called as the "Sensor Decade" for the dramatic increase in sensor R&D over the past years [158]. Sensors are used to measure various physical properties sch as temperature, force, pressure, flow, position, light intensity, acceleration, incident infrared radiation, etc. [134].

Sensors may be classified in a number of ways. One useful way is to classify sensors either as active or passive. The former require an external source of power, thus they consume power even when nothing is detected. The latter generate their electrical output signal without requiring external voltage or current. A list of popular sensors is presented in table 2.1.

Most sensors require an output conditioning circuit to amplify and filter their output in order to be processed by a microcontroller. Typical sensor conditioning circuits include amplifier, filtering, level translation, impedance transformation.

Property	Sensor	Active/Passive	Output
Temperature	Thermocouple Silicon Thermistor	Passive Active Active	Voltage Voltage/Current Resistance
Force/Pressure	Strain Gage Piezoelectric	Active Passive	Resistance Voltage
Accelerometer	Accelerometer	Active	Capacitance
Infrared radiation	Pyroelectric InfraRed	Passive	Voltage/Current
light intensity	Photodiode	Passive	Current

Table 2.1: Popular sensors and their output.

2.2.3 Communication Unit

The wireless communication channel enables to transfer signals from sensors to exterior world, and also an internal mechanism of communication to establish and maintain of WSN. This medium needs to be bidirectional, to be energy-efficient, and have relatively slow date rate. Two basic techniques are used: optical communication and radio frequency communication [151].

Optical communication

Two main technologies are available for optical communication: laser and infrared.

Laser communication consumes less energy than RF over larger range, is secure, since upon interception the signal is interrupted, and do not

need antennas. However it requires line of sight and alignment between transmitter and receiver and this is a major drawback since several applications presents randomly deployed nodes.
Also infrared is directional and requires line of sight between 2 communicating nodes. It allows only short range (less than 10 meters), but do not require antennas. An interesting solution is presented with the PushPin project [91] in order to achieve omni-directional ifrared communication on a single plane.

Radio frequency communication
Based on electromagnetic waves, one of the most important challenges for this typology of communication is antenna design and size. However RF communication present several advantages. It is easy to use, to integrate and it is a well established technology. Power consumption of RF communication is affected by type of modulation, data rate and transmission power. An important aspect to consider when working with RF transceiver is that idle state (radio active but not transmitting, nether receiving) drawn as much current as receive mode. Thus wireless protocols must reduce as much as possible this waste of energy.

2.2.4 Power Unit

Power supply unit usually consists of a battery and a dc-dc converter. Thus, the power needs of large wireless sensors network (maybe deployed in harsh environment) is the current biggest impediment that keeps them from becoming completely autonomous, forcing them to be either connected to an external power source or have lifecycles that are curtailed by batteries. Furthermore, in some application like gesture recognition, where sensor are embedded into user garments, battery size is the most relevant factor when seeking unobtrusiveness since battery technology tends to be a limiting factor in miniaturization [116].

For this reason in the last years, energy harvesting has emerged as one alternative to provide perpetual power solution to sensor network.

2.3 State of the art

In this section a we present a series of commercial and academic solutions of wireless sensor nodes and their main features.

2.3.1 Smart Dust

The goal of the Smart Dust project, founded by DARPA (Defense Advanced Research Projects Agency), is to demonstrate that a complete sensor communication system can be integrated into a cubic millimeter package. This involves both evolutionary and revolutionary advances in miniaturization, integration, and energy management [22, 153]. A conceptual diagram of a Smart Dust mote is presented in figure 2.7.

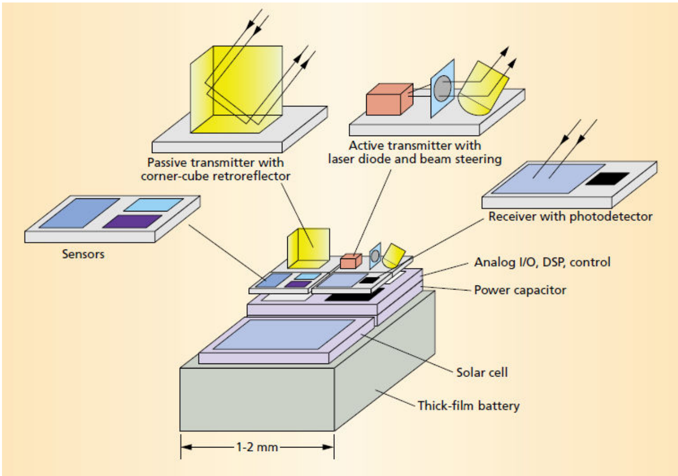


Figure 2.7: A diagram of the Smart Dust mote

Many sensors, including temperature, pressure, and acceleration sensors, from MEMS and CMOS processes can be attached to a mote. In contrast to typical computing systems, in an autonomous cubic-millimeter package computation must focus on minimizing a given tasks energy consumption. This is achieved through frequency and voltage scaling, since the computation requirement for this motes are limited. Communication is possible by means of two approaches: passive reflective systems between nodes and the base stations and active steered laser systems between motes. The power system consists either of a thick-film battery, or a solar cell with a charge-integrating capacitor for periods of darkness, or both.

2.3.2 Intel mote

The Intel Mote is a new sensor node platform motivated by several design goals: increased CPU performance for data compression as well as initial classification and analysis, improved radio bandwidth and reliability, and the usage of commercial off-the-shelf components in order to maintain cost-effectiveness. An important aspect of the platform design was to increase performance while preserve battery life. To satisfy these requirements, Intel chose a system on chip from Zeevo Inc. including a CMOS Bluetooth radio and an ARM7TDMI core operating at 12MHz and with 64KB SRAM and 512KB FLASH [112].

The Intel Mote is built on a 3×3 cm circuit board that integrates the Zeevo module, a surface-mount 2.4GHz antenna, various digital I/O options using stackable connectors and a multi-color status LED (see figure 2.8).

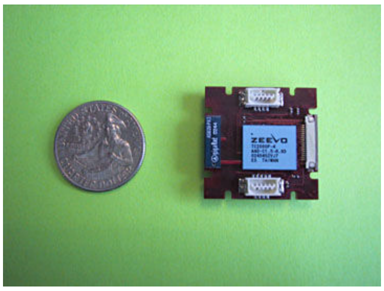


Figure 2.8: The intel mote

Intel second generation of sensor nodes are the Intel Mote 2. This motes are based on an Intel PXA270 XScale CPU with 32 MB of flash and 32 MB of SDRAM resulting in high performance processing capabilities. The processor integrates a DSP co processor, a security co processor and an expanded set of I/O interfaces. The platform also provides an on-board 802.15.4 radio and the option to add other wireless standards such as Bluetooth and 802.11b via an SDIO interface. The complete platform is hosted on a single 36×48 mm printed circuit board [87, 136](see figure 2.9).

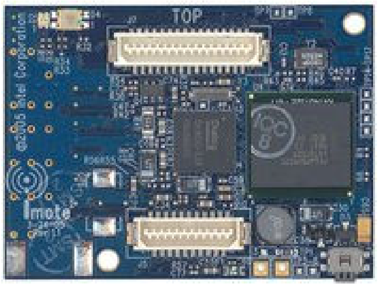


Figure 2.9: The intel mote 2

2.3.3 Mica Mote

MICA Motes (see figure 2.10), developed by UC Berkeley research group on wireless sensors, is a mote module used for research and development of low power, wireless, sensor networks. The motes measures 3.16×6.35 cm and is created using off-the-shelf hardware, but the architecture and its capabilities could be implemented in just a few square millimeters of custom silicon. The main microcontroller is an Atmel ATMEGA128 running at 4MHz with 128kB of FLASH and 4kB of RAM. The radio module is based on an RF TR1000 transceiver operating at 916.5 MHz. Several sensor extension board can be connected to the base board, such as: thermal temperature, barometric pressure, magnetic fields, light, passive infrared, acceleration, vibration, and acoustics [74].

Figure 2.10: The Mica mote

An evolution of the Mica motes are the Mica2 mote [40] and the the MICAz [41] mote from Crossbow [42]. The latter, in particular, is a 2.4 GHz, IEEE 802.15.4/ZigBee, board used for low-power, wireless, sensor networks.

2.3.4 Tmote Sky

Tmote Sky [135] is an ultra low power wireless module for use in sensor networks, monitoring applications, and rapid application prototyping. On a single $3,22 \times 6.55$ cm board it integrates an ultra low power microcontrolloer (MSP430 from TI), sensors (Humidity, temperature and light sensors), a Zigbee compliant radio (CC2420 from Chipcon), antenna and programming capabilities (see figure 2.11). Tmotesky offers a robust solution with hardware protected external 1MB flash, in the event of a malfunctioning program, the module loads a protected image from flash to restore proper operation.

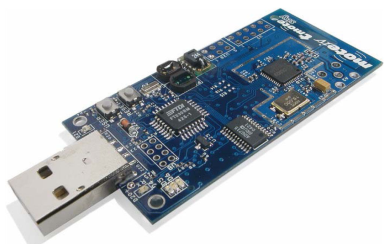


Figure 2.11: The Tmote Sky mote

2.3.5 BT Node

The BTnode (see figure 2.12) is an autonomous wireless sensor platform developed at ETH Zurich by the Computer Engineering and Networks Laboratory (TIK) and the Research Group for Distributed Systems [56]. The mote is based on a Bluetooth radio and a microcontroller. It serves as a demonstration platform for research in mobile and ad-hoc connected networks (MANETs) and distributed sensor networks. Currently the latest version is revision 3 which includes a core CPU Atmel ATmega128L with 4kByte EEPROM, 64kByte SRAM, 128kByte Flash and a dual radio device composed of a Zeevo ZV4002 Bluetooth radio and a low power Chipcon CC1000 radio operating at 868 MHz. The BTnode rev3 is compatible to the old BTnode rev2 and the Berkeley Motes. This twin device can operate both radios simultaneously or shut them down independently when not in use.

Figure 2.12: The BTnode mote

2.3.6 System on chip

One of the main limitations of the platforms presented in the previous sections is that they are built using commodity chips, which themselves are not specifically designed for wireless sensor network applications. As a result, they suffer several inefficiencies that lead to limited functional capabilities, high power consumption, and limited operational lifetimes [59]. A breakthrough innovation happened when the whole sensor node has been integrated on a single chip. In the following sections we present the solutions proposed by 2 Original Equipment Manufacturers (OEM).

Freescale solutions

With the mission of making the world a smarter place with leading embedded semiconductor solutions for cars, mobile phones, networks and many more, Freescale is a leading company that develops and produces electronic devices for many applications: automotive, computer networks, communications infrastructure, office buildings, factories, industrial equipment, tools, mobile phones, home appliances and everyday consumer products. Freescale has joined the Zigbee alliance in 2004 as a promoter and, since then it has develop several solution for Zigbee.

In particular 2 system on chips have been developed for WSN.

MC1322x Platform in a Package (PiP)
The MC1322xV [67] is Freescales third-generation ZigBee platform which incorporate a complete, low power, 2.4 GHz radio frequency transceiver, 32-bit ARM7 core based MCU, hardware acceleration for both the IEEE 802.15.4 MAC and AES security, and a full set of MCU peripherals into a 9.5×9.5mm Platform-in-Package (PiP). The MC13224V solution includes a fully functional 32-bit TDMI ARM7 processor, 128KB FLASH, 96 KB RAM and, 80K ROM containing boot code, all device drivers and fully compliant IEEE 802.15.4 MAC. Typical power consumption is 21mA in Rx mode and 29mA in Tx mode and drops to less than 1μA in stop mode. This device can be used for wireless applications ranging from simple proprietary point-to-point connectivity to complete ZigBee mesh networking in order to provide a highly integrated, total solution, with premier processing capabilities and very low power consumption.

MC1321x System in Package (SiP)
The MC1321x family is Freescales second-generation ZigBee platform which incorporates an 8 bit MCU (MC9S08GT) with a Zigbee compliant transceiver (MC1320x) into a single 9×9mm package [67]. The MC13213 provides 60 K Flash memory and 4 K of RAM and can operate at up to 40MHz. It consumes 35mA in Tx mode and 42mA in Rx mode when the MCU operates at 16MHz. By using the IEEE 802.15.4 Compliant MAC, or BeeStack ZigBee Protocol Stack, the MC1321x solution can be used for wireless applications from simple proprietary point-to-point connectivity to a complete ZigBee mesh network.

Ember solutions

Ember's mission is to be the leading provider of wireless sensor and control network technologies that enable dramatic energy efficiency improvements for businesses, homes, and the utilities that serve them. For this reason Ember joined the Zigbee Alliance in 2003 as a promoter and developed several devices and tools to develop Zigbee based applications [54].

Since 2005 ember produces the SN250, system on chip for Zigbee based WSN. The EM250 combines a 2.4GHz IEEE 802.15.4 compliant radio transceiver with a 16-bit microprocessor with 128kB Flash and 5kB RAM in a 7$\times$7mm package. Requiring 28mA in RX mode and 24 in TX mode and being able to drop power consumption down to 1μA, it is optimized for designs requiring long battery life and low external component count.

Chapter 3

Low Power Video Sensor Nodes and Video Processing Alghoritms

3.1 Overview

Due to the rapid evolution of semiconductor technology, on-die computing capacity becomes exponentially smaller and cheaper. Small, low-power processing elements as well as low-power radio interfaces and microfabricated sensors have been recently exploited to build low-cost and low-power miniaturized wireless sensor nodes [45, 61, 125]. These nodes can be deployed in a target area to sense physical phenomena and communicate with other nodes and base stations.

This work is dedicated to a particular case of such nodes: video sensors [88, 98]. This topic holds strong interest for a wide range of contexts such as military, security, robotics and recently also consumer applications.

Vision sensors are extremely effective for medium- to long-range sensing because vision provides rich information to human operators. However, image sensors generate a huge amount of data, which must be heavily processed before transmission due to the scarce bandwidth capability of radio interfaces. In particular, it has been shown that in case of video-surveillance source-side compression is mandatory due to limited bandwidth and delay constraints [49].

Moreover, there is an ample opportunity for performing higher-level processing functions, such as object recognition, that has the potential to drastically reduce the required bandwidth (e.g. by transmitting compressed images only when something relevant is detected) [94]. In contrast to sensors that

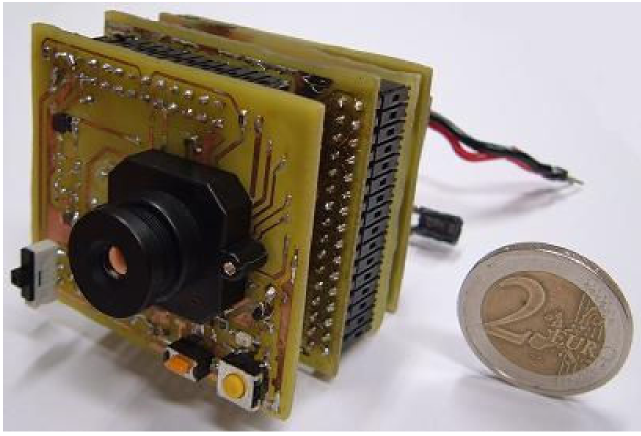


Figure 3.1: MicrelEye node

just "watch" the world, todays research is aimed at developing intelligent devices capable of comprehending the perceived information in order to interpret it locally. Featuring such intelligence, these nodes would be able to cope with such tasks as recognition of unattended bags in airports, persons carrying potentially dangerous objects, etc., which normally require a human operator [103, 86] . The energy cost of image processing must, however, be carefully minimized.

The aforementioned energy and performance requirements emphasize the potential benefits of exploiting hybrid architectures (i.e. microprocessor + FPGA /DSP /ASIC) to enable efficient processing before transmission [123]. In this case coprocessors must be coupled with software partitioning strategies for specific applications. For example, complex but highly parallel motion estimation algorithms commonly used by video coders can be replaced by efficient hardware implementations running on FPGA [127, 97].

3.2 A low-power wireless video sensor node for distributed human detection

This chapter presents the design of a novel video sensor node architecture based on a low-cost low-power FPGA + microcontroller system-on-chip (SoC). The node features dynamic reconfiguration capabilities and supports low-power local processing and wireless communication using various proprietary standards (e.g. Bluetooth). This new architecture addresses the bandwidth bottleneck by performing on-board image processing (e.g. pixel threshold analysis, image recognition and classification, etc) and offers considerable flexibility in exploring the trade-offs between processing and communication.

The rest of the section is organized as follows. In the next subsection we review related work. Section 3.2.2 presents the targeted video surveillance scenario. The hardware architecture of the node is described in Section 3.2.4. Video processing algorithm is presented in Section 3.2.10. In order to perform object detection, we propose a new Support Vector Machine (Section 3.2.3) - based (SVM-based) [147, 133] approximated algorithm suitable for embedded systems implementation (called ERSVM). SVM-like algorithms are considered here due to their good robustness and sparsity properties. The results obtained on a specific case-study (people detection) are reported in Section 3.2.15. Timing analysis of different modes and energy requirements are discussed along with obtained classification accuracies. Section 3.2.20 concludes the section.

3.2.1 Related work

A number of devices are available on the market for distributed video processing. These devices contain DSPs [1, 2, 3], large FPGAs [4, 5, 6], dedicated video processing engines [7, 8, 9, 10, 11, 12] and feature MIPS ratings from tens to thousands of MIPS. The supported image resolutions range from QVGA to 2.0 Mpixels and the frame rate ranges from 15 fps to more than 200 fps. On the other hand, the power consumption of these devices ranges from a few Watts to tens of Watts. Hence, they typically require either connection to the power grid, or massive rechargeable batteries and large solar panels for battery recharging.

MicrelEye, the node proposed in this section, targets a different design point respect to the ones above mentioned, in fact it targets a power budget of 500mW instead of a few Watts or more, while still supporting 15 fps (presence/absence) person detection at QVGA (320×240) image resolution. A system designed within this power envelope could be operated in a continuous fashion with a $5 \times 5 \times 3$ cm Li-Polymer battery pack (3.6V, 5x850mAh) and $4 \times 5 \times 5$ cm solar panels (1W in the sun). So MicrelEye is ideal for outdoor or temporary application where the power supply is a critical constriction, in fact there are many situations in which vast and inaccessible areas should be visually monitored to detect unusual events or to acquire environmental data over long periods. Examples include natural environments such as forests, deserts, and even planetary exploration as well as temporary market or stand. In this cases MicrelEye thanks to low power and to on board wireless transceiver is a optimal choice. On the other hand the high-end solutions based on commercial video processors haven't a so limited power budget and they are more powerful processor and can use for a more computationally expensive as object tracking or face recognition. So where is available a wired power and communication and a complex algorithm is needed the best choice is a more powerful and power expensive smart camera.as for example airport surveillance

Obviously, dealing with such a resource-constrained platform represents a challenging task. Firstly, suitable classification methods must be designed in order to fit the available resources; secondly, the designed classification methods must be implemented on the processing unit characterized by low power, limited processing speed, low capacity memory.

In the field of wireless sensor networks, a few video sensor nodes have been reported. All these nodes are based on commercial off-the shelf components to meet the tight cost constraints typical of distributed sensing applications. An early node prototype, Panoptes, is a camera device equipped with an Intel StrongARM processor, a Logitech USB camera, and Linux OS [62]. Panoptes features a StrongArm processor and a WiFi 802.11 network interface and consumes more than 5W. Meerkats [104] is a device which maintains the same class as Panoptes, but makes use of more recent components (e.g. XScale processor replaces StrongARM).With respect to Panoptes and Panoptes, MicrelEye has a much lower power consumption but on the other hand the MicrelEye exploit less powerful devices. Cyclops is a much lower power device that features a Xilinx CPLD and Atmel mocrocontroller unit (MCU) ATmega128L [124]. Even though the authors do not quote an overall power consumption figure (e.g. including DC/DC converter losses), it should be roughly 1/2 of that of MicrelEye. However, it is a much lower-end device in that it achieves a frame rate lower than 4 fps for basic (presence/absence) object detection task on a small image (128×128). Similarly, the wireless node proposed by Ferrigno et al. [65] is equipped with the Microchip PIC16LF877 microcontroller without any HW acceleration and performs software image compression at low frame rate (less than 1 fps). In respect to Cyclops and the last one node MicrelEye is more computationally intelligent, capable to perform on board objects classification locally as person detection with a good velocity, 15 fps, for a good locally distributed base video surveillance, while this feature is not possible with the node proposed by Ferrigno cause absence of on board intelligence and is too slow with Cyclope. So this system are better for ultra low power applications.

3.2.2 Scenario

In this section we describe the scenario targeted by the detection framework proposed below. Consider some area such as, for example, many-storied building, each entrance/exit being equipped with a powerful wired computer and a video sensor (*door-keeper station*).

Intelligent wireless video nodes with moderate computational capacity and self-contained power supply are installed throughout the building at key points such as doors connecting different halls, elevators, etc. The door-keeper station can collect images of people entering the building. Optionally it can have access to database of suspicious persons in order to check whether the person that enters into the building belongs to the group of suspicious ones or not.

What is important here is the fact that the door-keeper station can collect or select the images of two classes of objects to be later distinguished by the nodes. For example, these classes can be two different persons, or persons with rucksack versus persons without it. In some other scenarios, like agricultural holding, "person versus animal" classification can be required, etc.

After the images of two classes have been selected, the process of training the corresponding classifier, optionally foregone by feature generation/selection/reduction phase, is initiated by the door-keeper. As this training phase is done, the parameters describing the classifier can be sent to the base station, which in its turn can distribute them among the nodes in order to perform classification locally.

When motion or presence is detected by a node, the check-up of the motion source is carried out. In case of positive outcome, the node transmits the image of the motion source to the base station, where, for example, a human operator can make final recognition and take some actions if needed.

In conclusion, it is worth mentioning that the roles of door-keeper station and base station can be united.

3.2.3 Support Vector Machines

Support Vector Machines (SVM) is a supervised classifiers belonging to the class of linear discriminant classifiers. Such classifiers build discriminant functions that are a combination (either linear or not linear) of the input vectors' components. Geometrically, a discriminant function defines an hyperplane that separates two classes [53]. Several solution have been proposed to deal also with non-separable data.

The original idea about SVM has been developed since 1979 by Vladimir Vapnik [146, 148, 149]. Recently there has been an explosion in the number of research papers on the topic of SVM. SVMs have been successfully applied to a number of applications ranging from particle identification, face identification, and text categorization to engine knock detection, bioinformatics, and database marketing [21].

The simplest case deal with 2 classes linearly separable data. If we call $\mathbf{x}_i$ the vector with the features, and $y_i = \pm 1$ the label of each input vector. A discriminant function that is a linear combination of the components of x can be written as:

$$\begin{aligned} \mathbf{x}_i \cdot \mathbf{w} + b > +1 \quad & y_i = 1 \\ \mathbf{x}_i \cdot \mathbf{w} + b < -1 \quad & y_i = -1 \end{aligned} \tag{3.1}$$

Where **w** is a weight vector that determines the orientation of the separating hyperplane and b is a bias that indicate the distance from the origin of the separating hyperplane see figure 3.2.

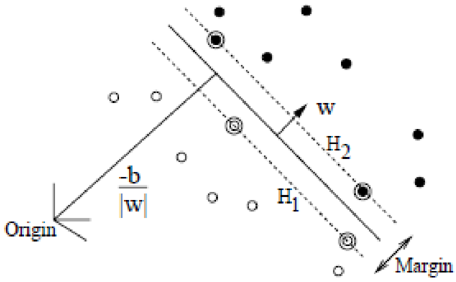


Figure 3.2: Best separating hyperplane in the separable case (feature space = 2)

It is clear that infinite planes can be defined to separate the two sets of samples. A smart choice is to select the one that presents higher margin. The hyperplane with higher margin can be found if ve consider the points where the equality in equation 3.1 holds. Such points lay on 2 hyperplanes ($H1$, $H2$) that share the same normal vector **w** and relative distance (margin) equal to $\frac{1}{\|w\|}$. Thus we can find the optimal hyperplane (the one with maximum margin) by minimizing $\|w\|^2$ subject to constraints 3.1.

Note how the only points needed to build the separating hyperplane are the one that lay on $H1$ and $H2$. Such points are called *support vectors*.

In a more complex case, where we have to distinguish between more than 2 classes, 2 solutions are possible: build an hyperplane that separate each class from all the other, build an hyperplane for each couple of classes (see figure 3.3).

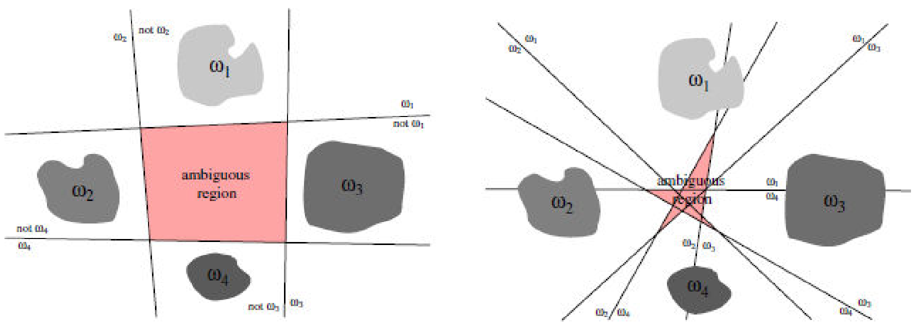


Figure 3.3: Two options for building a set of separating hyperplanes in the multiple class example

This approach can be extended to handle non separable data. The idea is to relax the constraints in equation 3.1, but only when necessary. In order to do it we introduce a further cost called *slack variables*, ξ_i.

$$\begin{array}{ll} \mathbf{x}_i \cdot \mathbf{w} + b > +1 - \xi_i & y_i = 1 \\ \mathbf{x}_i \cdot \mathbf{w} + b < -1 + \xi_i & y_i = -1 \\ \xi_i \leq 0 & \forall i \end{array} \tag{3.2}$$

In equation 3.2 for an error to occur ξ_i must be greater than 1, hence $\sum \xi_i$ is an upper bound of the training error. We can take this contribution into account by changing the objective function to be minimized to $\frac{\|w\|^2}{2} + C(\sum \xi_i)^k$ [32], where C is a user defined constant. The higher is C the higher is the penalty assigned to errors. A graphical representation of the use of slack variables is presented in figure 3.4.

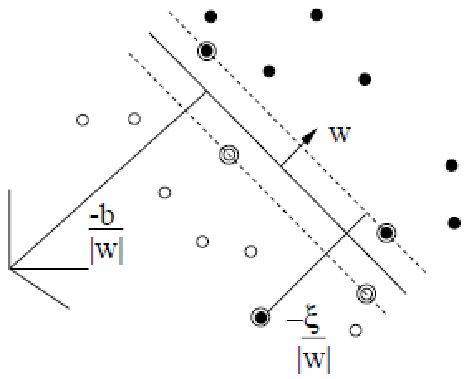


Figure 3.4: Separating hyperplanes in case of non separable data.

The concept above can be further extended to non linear hyperplanes. The basic idea is to map the input feature vector into a space with much higher dimensionality ($n \gg m$) where they can be easily separated.

$$\Phi : R^m \rightarrow R^n \tag{3.3}$$

It can be shown that in the training steps the vector of features appears always as a product of vectors ($\mathbf{x}_i \cdot \mathbf{x}_j$), thus if we are able to find a *Kernel function* $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_i)$ we will use only such functions and we do not even need to know Φ.

Some example of kernel are presented in 3.4.

$$
\begin{aligned}
K(\mathbf{x}, \mathbf{y}) =& (\mathbf{x} \cdot \mathbf{y} + 1)^p \\
K(\mathbf{x}, \mathbf{y}) =& e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}} \\
K(\mathbf{x}, \mathbf{y}) =& \tanh(k\mathbf{x} \cdot \mathbf{y} - \delta)
\end{aligned}
\tag{3.4}
$$

3.2.4 Hardware architecture

The main objectives in MicrelEye's hardware design can be summarized as follows:

- Low power consumption (suitable for wireless sensor networks)
- Local ”intelligence” (on-board image processing capability)
- A reconfigurable architecture to achieve flexibility
- Wireless connectivity

In order to satisfy the above requirements we decided to use a SoC (System on Chip) which includes an FPGA and an MCU. This architecture is fully programmable and supports FPGA acceleration of computationally demanding image processing tasks, which would vastly exceed the MCU's capabilities. On the other hand, higher-end solutions based on commercial video processors would not fit our limited power budget, while an ASIC solution would be way too expensive in terms of non-recurring design costs and prototype fabrication.

ATMEL FPSLIC SoC [13] was chosen because it provides lower power consumption than other FPGA and CPU solutions. It includes MCU, 40K gates FPGA and SRAM on the same chip, thus reducing power consumption by eliminating capacitive loading associated with inter-device PCB connections. The FPSLIC architecture is optimal for our application target because it offers the advantages of a hybrid processor+FPGA architecture without the high power consumption and cost of higher-end system FPGAs. An external SRAM has been added to provide necessary memory resources for computation.

In this work wireless communication is based on a Bluetooth transceiver. Bluetooth was chosen because of its sufficiently high bandwidth and the ease to interface MicrelEye with a host device (i.e. a personal computer, a PDA or any other Bluetooth Serial Port Profile device). However, ZigBee radio interface is also supported.

The whole system is designed to achieve low power consumption. Each device provides a power saving mode to lower consumption when it is not used.

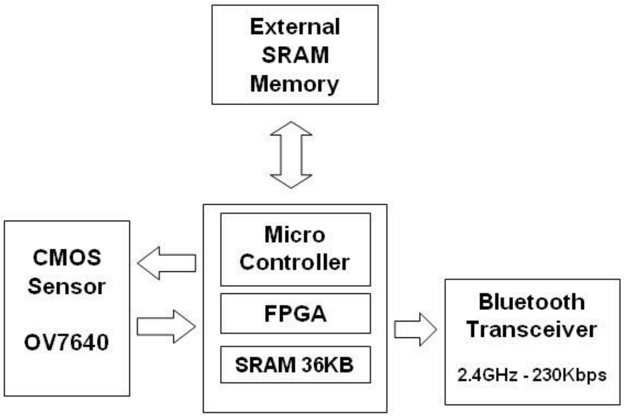


Figure 3.5: MicrelEye hardware architecture

The MCU is used to configure the image sensor at boot by setting its internal registers. It also implements the last steps of the object recognition processing flow.

The FPGA provides high-speed interface logic needed to capture images from the sensor. It is used to manage the access to the SRAM memory and for managing the synchronization with the continuous stream of data coming from the image sensor. Furthermore, it is used to perform almost all image processing needed for the detection operation. The block that manages interfacing between FPSLIC and the Bluetooth transceiver is also mapped on FPGA. Finally the finite state machine (FSM) that governs the overall working of the system runs on the FPGA as well.

The external SRAM is used to extend the limited amount of internal memory embedded on the FPSLIC chip and therefore to provide the memory for image storage and subsequent image processing operations. As a result, the embedded FPSLIC memory is not used for these operations and is all available as shared memory for data exchange between FPGA and MCU.

3.2.5 CMOS image sensor

The sensing device is OV7640 from Omnivision, which supports 30 fps frame rate in color mode and 60 fps in black-white (BW) mode. It operates at 2.5 V for the internal core and 2.7V for external I/O. The power consumption is 40 mW when operating at 30 fps and only 30 μW when in standby. It is a 640x480 capable device, but we chose to use a 320x240 (QVGA) resolution in order to reduce the amount of data that need to be stored and processed. Though the sensor can work with clock frequencies up to 24 MHz, we set its clock to 12 MHz in order to satisfy the memory access time when saving a frame. BW mode has been used with 30 fps frame rate.

The device has several internal registers whose values can be set through a two-wire SCCB (Serial Camera Control Bus). The registers control image settings and provide the possibility to set a variety of output formats.

3.2.6 Processing core

The digital processing device is an FPSLIC produced by Atmel. It combines an AT40K, a 40K gates FPGA and a high performance AVR 8-bit RISC microcontroller. A small cut of onboard SRAM is also available: 36KB are available, of which a maximum of 16KB can be used as data memory with the remaining 20KB being reserved for MCU program storage. The data SRAM is accessible to both MCU and FPGA.

Having MCU and FPGA on the same chip enables one to avoid off-chip communications when exchanging data between the two devices in order to reduce the overall power consumption. Furthermore, resources on 8-bit MCU are not sufficient enough to interface with an image sensor that provides a significant flow of data. For this reason, having also a small amount of programmable interface logic is the best choice. The core is clocked at 14.74 MHz so it can achieve about 14 MIPS executing powerful instructions in a cycle. The typical power consumption for MCU is about 2-3 mA per MHz.

3.2.7 External SRAM

In order to store both the frames acquired from the sensor and the processed images, MicrelEye needs external memory. A static CMOS RAM (BS62LV8001 from BSI), which provides 1M x 8-bit, has been added to the system. The device has a wide Vcc operation voltage that ranges from 2.4V to 5.5V. Typical standby current is 1.5 μA at 3 V/25 °C, with a maximum access time of 55 ns at 3.0 V/85 °C (therefore two clock cycles are required at system speed of 24 MHz). Moreover, the chip has an automatic power down feature that significantly reduces power consumption.

3.2.8 Bluetooth transceiver

Wireless capabilities of the node are provided by the integration of LMX9820A Bluetooth transceiver, which was chosen for its low power consumption. This device is a highly integrated radio, baseband controller and memory device. It is a complete solution that includes hardware and firmware from antenna and lower layers of the Bluetooth stack up to the application layers including several connection profiles. Our application exploits the Serial Port Profile

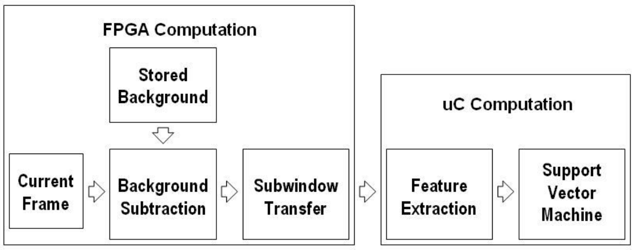


Figure 3.6: Video processing algorithm flow

that allows us to establish a link between the transceiver and a remote device through a virtual serial port. LMX9820A features a small form factor, ideal for our goals. Internal working is based on a processor and the Digital Smart Radio technology. The firmware supplied includes a Bluetooth stack v 1.1. Data rates up to 704 kbps can be reached over RFComm. 230400 bps data rate has been chosen for our design. In fact, even if the whole computation is done onboard so that the amount of data to be transmitted is reduced to an minimum, it may be needed to transmit a complete frame (e.g. to verify the scene when an object has been recognized).

Power consumption for the transceiver with 2.7 V supply is only 2mA when in idle mode and about 30 mA when a connection is established in continuous transmit mode.

3.2.9 Power supply

The power supply section of the node includes a 4.28V battery and two DC/DC converters. The first one is used to generate 2.7 V voltage reference for all components except the image sensor core. The second converter generates the 2.5 V voltage reference for CMOS sensor core. Both components have been chosen because of a low power consumption (max 225 μA at maximum output current) and low drop-out voltage (about 120mV).

3.2.10 Video processing algorithm

The algorithm follows the block diagram in Figure 3.6. It has been split between FPGA and MCU to exploit parallelism thus reducing overall computation time. The first processing steps are done on the FPGA because of requirements in terms of speed and computational power to store and manipulate images. In this section we describe in detail the algorithm steps as well as hardware and software algorithm implementation.

3.2.11 FPGA computation

The first hardware processing step performed on FPGA is frame acquisition from the image sensor. Data coming from the sensor is in YUV 4:2:2 format: for each pixel only two bytes are transmitted, one for the luminance component and the other for solely one of the chrominance components (i.e. a typical output bytes sequence is YUYVYUYV). Since we are interested in 8-bit grayscale image, we use only "Y" bytes. The hardware block that interfaces FPGA with video sensor automatically discards chrominance components, sending only luminance bytes to the memory store block.

Each frame is first stored in external memory, the subsequent operation being a pixel-by-pixel subtraction of a fixed background from the acquired frame (the absolute value of the difference are computed). This fixed background represents the reference scene and is acquired at boot. It can be updated at regular time intervals (e.g., to adapt to slow or permanent background changes). The background can also be updated through an explicit command sent over Bluetooth interface (e.g., for collaborative operation).

After background subtraction the region-of-interest (ROI), 128x64 subimage, is extracted starting from a position which can be changed on a frame-by-frame basis. This variable subimage extraction phase allows tracking of moving objects across different frames. Alternatively, the position of the subframe can be set to a stationary value, for instance when monitoring a fixed space region such as an entrance.

ROI is stored into on-chip memory in FPSLIC because the remaining processing steps are performed by the microcontroller, which does not have a direct access to external SRAM. The dual-RAM architecture is very useful because it enables parallelized computation between hardware and software. In fact, while MCU is computing features and performing recognition on ROI stored on internal memory, FPGA can acquire the next frame from CMOS sensor and compute background subtraction because this operations only involves external memory.

Once ROI of the subtracted image is transferred into internal memory, microcontroller computations start. This phase consists of the following two steps: feature extraction and classification.

3.2.12 Feature extraction

In order to form the feature vector we calculated the average values of gray for each column and row of ROI and normalized them to [0,1] range. Such calculations normally require addition and division operations. However, the dimensions of ROI have been chosen in such a way that both number of rows and number of columns are powers of two. Thus, iterative division algorithm can be replaced by a simple shifting operation. As a result, the dimensionality of data to be fed to the classifier is reduced from 8192 elements of ROI to 192 elements of the feature vector, first 128 elements being rows averages followed by the averages of 64 columns. Undoubtedly both smart ROI size and feature extraction contribute significantly to resource sparing.

3.2.13 Classification algorithm (ERSVM)

As regards the classification step, a SVM-like hardware-oriented algorithm has been developed and implemented.

Like Artificial Neural Networks (ANNs), SVMs [147, 133] are aimed at recovering unknown dependencies on the basis of available data. They have been introduced by V. Vapnik and colleagues in 1990s, and have been primarily applied to optical character recognition [38]. Proven to be efficient classifiers, nowadays their area of application is spread from electricity load prediction and biomedical engineering to face detection and face recognition. Some reasons of such a success are reported below. First of all, SVMs are based on the results of *Statistical Learning Theory* (*Structural Risk Minimization* principle, *VC dimension* complexity measure) [147]. Secondly, SVMs reduce training phase to solving Constrained Quadratic Optimization problem, which, in contrast to ANNs training, does not suffer from local minima. In addition, it is worth noticing that when applied to high-dimensional data, SVMs (in contrast to other classifiers like NNs) do not suffer from curse of dimensionality. A detailed description of SVM algorithm would require a considerable amount of space and goes beyond the scope of this section. Below only key points of non-linear SVM classifier are given. For a brief introduction we refer the readers to [33].

Being a "learning from examples" technique, SVM is firstly trained on a set of available data known as *training set*. Such a training phase is normally performed offline and results in constructing the classification function, which is then used online during the forward, or prediction phase. So, the computationally expensive training phase can be performed by a powerful base station, and the evaluated classification function is sent to the nodes, where the prediction phase is run in order to classify the patterns under observation.

In fact, binary SVM is a linear classifier, i.e. patterns are discriminated by a hyperplane, which is represented as a linear combination of a subgroup of the training set patterns. However, linear algorithms are limited in their capabilities since normally real-world data are not linearly separable. This obstacle is overcome using the so-called *kernel trick* [133], which consists in implicit mapping the patterns onto higher-dimensional space, where data is much more likely to be separable by a hyperplane. Explicitly this leads to the following form of classification function:

$$y(\mathbf{x}) = \sum_{i=1}^{N_{sv}} \beta_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{3.5}$$

where N_{sv} is the number of relevant patterns, *Support Vectors* $\mathbf{x}_i$ from the training set, and $K(\mathbf{u}, \mathbf{v})$ is a kernel function. One of the most used, known, and studied kernels is the so-called *Gaussian* kernel $K(\mathbf{u}, \mathbf{v}) = e^{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2}$.

As one can see, the complexity of the forward phase is proportional to N_{sv} and depends on the complexity associated with calculating $K(\mathbf{u}, \mathbf{v})$. So, implementing SVM classification on resource-limited platforms gives rise to the following two issues: reduction of N_{sv} and building hardware-oriented kernels. Recently, both of them have been touched upon by scientific communities.

Firstly, various approaches have been proposed for the reduction of N_{sv} [93, 160, 161, 82]. In particular, in [160] SVs are considered as variables to be optimized, and it has been demonstrated that it is enough to use much less "optimized" vectors in the classification function (3.5) in order to reach almost the same accuracy. We used the modified version of the algorithm proposed in [160]. As the result, the new classification function can be written as

$$y(\mathbf{x}) = \sum_{i=1}^{N_{ev}} \beta_i K(\mathbf{x}_i^{ev}, \mathbf{x}) + b \tag{3.6}$$

where N_{ev} is the number of optimized vectors $\mathbf{x}_i^{ev}$ called *Expansion Vectors*. In our experiments $\mathbf{x}_i^{sv}$ has been reduced by the order of magnitude without any significant loss of the classifier's accuracy.

Secondly, we implemented a new kind of kernel function recently proposed in [17], whose calculation implies only shifting and addition operations and avoids more computationally-expensive multiplications. The kernel is as follows:

$$K(\mathbf{u}, \mathbf{v}) = 2^{-\gamma|\mathbf{x}_i - \mathbf{x}_j|_1} . \tag{3.7}$$

As compared to the Gaussian kernel, here the base of 2 is used instead of e. Moreover, the distance is calculated using L_1-norm rather than the Euclidean one. Also, the parameter γ must be a power of two. All these modifications lead to better use of microcontroller resources and lowering the overall computational complexity. As reported in [17], using this kernel does not affect SVM accuracy.

Once computation has been completed, the detection final result is a binary information. Alternatively, the user can send a remote command to reconfigure

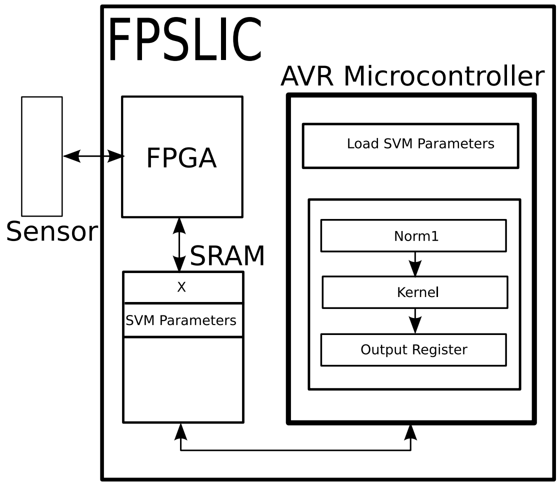


Figure 3.7: Microcontroller implementation of ERSVM: architecture

the system to force it to send the whole image containing the recognized object instead of the simple binary detection flag.

3.2.14 ERSVM: microcontroller implementation

Being designed for low-power applications, embedded AVR 8-bit RISC microcontroller provides several low-power operating modes. Normally this means small amount of available resources, which in turn implies that an accurate optimization of the code is required. Indeed, in FPSLIC there are 4-16 KB of Data RAM and 16-32 KB (depending on the configuration) of Program RAM, while the external memory can not be directly accessed. Therefore both program and data have to meet this strong limitation. The Data RAM is both accessible from FPGA side and AVR data memory bus, thus it can be used as an interface between the programmable logic and the microcontroller.

The macro blocks of the presented architecture are depicted in Fig. 3.7. The estimation function of SVM is divided in two main blocks: the first one is used to load all the SVM parameters in the memory. This process is executed at the very beginning (power on) or whenever one wants to dynamically reconfigure the device with a new set of machine parameters (e.g. to detect some other kind of object).

The second function starts each time a new vector has to be classified. It reads the previously generated feature vector and provides a classification. For the sake of simplicity this function has been divided into three subparts: $norm1$, $kernel$, and $output\ register$. In $norm1$ L_1 norm is computed:

$$norm1_i = \sum_{j=1}^{d} \left| x_{ij} - x_{ij}^{ev} \right|$$

As regards the $kernel$ unit, the following expression is computed through a CORDIC-based algorithm:

$$kernel_i = \alpha_i 2^{-\gamma \cdot norm1_i}$$

In this way, as suggested in [17], the convergence of the algorithm is guaranteed.

The last module adds or subtracts all $kernel_i$ results and the bias according to the related label y_i.

The code has been written in C and then compiled using avr-gcc [14]. The amount of memory required for instructions is 622 bytes. As regards the memory used for data, its amount (in bytes) depends on number of EVs (N_{ev}) and on number of features (d) as follows:

$$M_{data} = d + d \cdot N_{ev} + N_{ev} + 5 \tag{3.8}$$

where the first term corresponds to the vector to be classified, the second term is the memory needed to store EVs, then N_{ev} bytes are used to store all the α_i, and the other bytes are used for bias, γ, and temporary variables such as iterators.

The following approximate equation, which correlates the number of clock cycles to perform a classification with d and N_{ev}, has been empirically obtained by averaging over different trials:

$$N_{clk} = c_1 + c_2 \cdot N_{ev} + c_3 \cdot d \cdot N_{ev} \tag{3.9}$$

where on the average $c_1 = 120$, $c_2 = 880$, $c_3 = 88$. Here the first term represents the time needed for function call/return and variables initialization. The second term designates the clock cycles used upon running $kernel$ and $output\ register$ blocks. Finally the third term concerns the time spent for computing $norm1$.

3.2.15 Experimental results

We firstly focus on MicrelEye power and performance. Then, in Section 3.2.19 we describe the case study and report the obtained classification accuracies.

We compare three different implementations of the object recognition algorithm. In the first implementation (serial implementation) hardware and

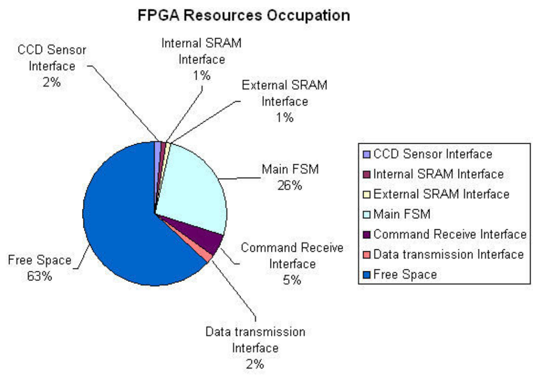


Figure 3.8: Resources utilization on FPGA

software processing have been serialized. When subimage transfer is completed, FPGA sends an interrupt to MCU which signals that feature extraction can start and then halt, while waiting for MCU processing to end. In the second implementation (parallelized implementation) hardware and software processing have been parallelized so that hardware runs continuously. When subimage transfer finishes, the FPGA sends an interrupt to MCU but instead of halting it immediately starts acquiring next frame from CMOS sensor. In the last implementation (optimized implementation) the same parallelized approach has been used together with a more optimized version of the memory access mechanism. This has been obtained by reducing the complexity of the memory access request-acknowledge protocol between central FSM and memory interface block. In the non-optimized version every access starts with a request from FSM, which then halts waiting for an acknowledge signal. In the optimized version of the protocol waiting has been eliminated. By doing this, each memory access is one cycle shorter than before (clearly, this reduced flexibility requires constant memory access time). In our system this translates into a significant reduction of the whole processing time, and this is due to the large number of memory access required to store the frame and compute background subtraction.

For each implementation several indicators have been measured to evaluate performance in terms of execution time, power consumption and energy per frame.

Power consumption for each component has been measured and the results are reported in Figure 3.9. As one can see, power consumption of the DC-DC converter is small, peripheral components (i.e. CMOS sensor, SRAM and Bluetooth transceiver) have comparable consumption, while the most consuming component, as expected, is the FPSLIC (the processing device). Resources occupation on FPGA has also been measured in order to evaluate the possibility

Table 3.1: Current for serial implementation steps

Processing Step	Current
Frame Acquisition	156 mA
Background Subtraction	168 mA
Subimage Transfer	167 mA
MCU computation	152 mA

to add new features to the system. Figure 3.8 demonstrates that even if we are working on a resource constrained device (only 40K gate FPGA), we have about 63% of the device resources free.

3.2.16 Serial implementation

The overall time needed to complete serial implementation execution is 174 ms (sum of frame acquisition, background subtraction, subimage transfer, feature extraction, and SVM computation times). In particular, frame acquisition takes about 63 ms. This is because, with the non-optimized version of the memory access protocol each memory access takes 4 cycles to complete.

Background subtraction is completed in 42 ms. Each pixel subtraction takes 8 cycles to load two pixels (one from the background and one from the current frame), a cycle to compute the difference and 4 cycles to store the result. This operation must be repeated 77280 times (76800 pixels of a 320x240 frame plus two additional row control pixels added at the end of each line).

Subimage transfer takes 3.3 ms, and it is the shortest step involving 8192 pixels (i.e. pixel contained in the 128x64 window extracted from the complete frame).

Finally, microcontroller processing takes 66 ms. It consists of the loops required in order to compute row and column mean values for the features and to compute SVM result.

For each of these steps, power consumption has been measured, and the corresponding results are reported in Table 3.1. The maximum object recognition frame rate reached with the abovementioned timings is about 5 fps. Taking into consideration the values reported in Table 3.1 and 2.7 V voltage of the power supply, we obtain 74.14 mJ per frame energy dissipation and a average power consumption of about 0.43W.

3.2.17 Parallelized implementation

Hardware and software execution can be overlapped, exploiting the presence of two distinct processing devices in the system. By doing this, overall time necessary for frame processing (i.e. hardware execution time) is reduced to

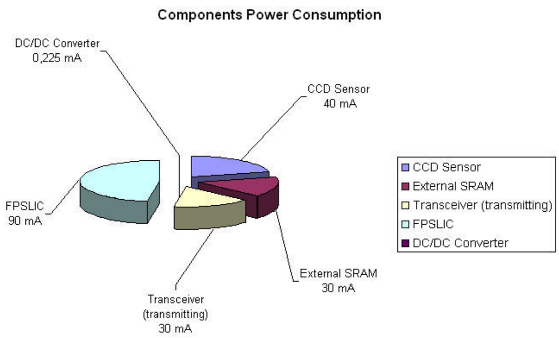


Figure 3.9: Power consumption for each component

Table 3.2: Current for parallel implementation steps

Processing Step	Current
Frame Acquisition	179 mA
Background Subtraction	192 mA
Subimage Transfer	184 mA

108 ms, which is the real benefit of this implementation. As regards power consumption, it remains almost equal to that in serial implementation.

Because of the different processing overlaps, single step consumptions cannot be precisely measured. For parallel implementation we measured consumption within hardware processing steps, while taking into account that MCU consumption is also included in this measure. The results are reported in Table 3.2. Thus, even if power consumption is greater, about 0.5W, time reduction brings to a value of 54 mJ for the energy per frame dissipation. The maximum object recognition frame rate achievable within this implementation is about 9 fps.

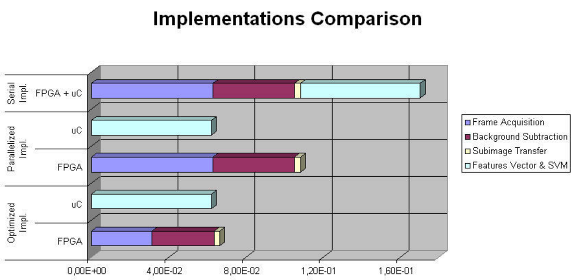


Figure 3.10: Hardware and software timings comparison

3.2.18 Optimized implementation

The optimized version of the parallelized implementation differs significantly from the previous one. This is due to the memory access time reduction. This change not only affects overall time by requiring 3 clock cycles per access instead of 4, but also reduces the time needed for access (particularly to write memory) and it is possible to increase the frame rate of images coming from the sensor.

Therefore in this new implementation CMOS sensor has been configured for a 30 fps output and frame acquisition time has been reduced to 33 ms. Memory access time reduction also affects background subtraction: the time needed for this operation with new protocol is about 32 ms. Finally, the time needed to transfer subimage to internal memory is also reduced and is equal to 2.68 ms. The overall time is therefore 68 ms, leading to about 2.5 times speed-up with respect to the first solution. Therefore, the object recognition frame rate is 15 fps. In terms of energy efficiency, this new solution needs 35 mJ per frame, thus providing 53% reduction of energy consumption with respect to the first solution and a 37% with respect to the second one. While the power consumption is about the same of second one.

3.2.19 Classification accuracy

Firstly, the presented classification algorithm has been extensively validated on multiple well-known standard data sets. The obtained accuracy values were close to to the ones obtained with standard SVM, whereas the number of support vectors has been reduces by an order of magnitude. More details can be found in [83]. Below the results obtained for the case study of people detection are presented.

The initial images have been acquired during 4 different sessions. The sessions differ by place, time, and lighting conditions. For example, some places simultaneously had two different kinds of illumination sources: artificial (daylight lamps) and natural (windows). So the people passing behind provoked soft shadows in the camera field of view, and partial cloudiness added a slight brightness fluctuation. Therefore the resulting data sets are characterized by sufficiently high level of heterogeneity and soundness. As the result, 219 positive samples have been generated. Negative objects (like boxes, hall trees, etc.) were less numerous. In order to create balanced data, additional negative samples have been generated. To this aim poorly scaled or centered images of people have been used (e.g. people located too close or too far). In total, 438 samples have been obtained. 140 randomly chosen ones have been preserved for test set, whereas the rest 298 ones have been used for training.

Table 3.3: Accuracy for person detection problem, Gaussian kernel (floating point) and hardware-friendly kernel (HFK, floating point and fixed point). l denotes the number of training samples. Fixed point: data width is 16 bit, fractional part width is 8 bit

Algorithm	Gaussian kernel	HFK	HFK, fixed point
SVM	91.4%	94.3%	95.7%
ERSVM, $N_{ev}/l = 2\%$	94.3%	94.3%	93.6%
ERSVM, $N_{ev}/l = 4\%$	92.9%	94.3%	96.4%

Five-fold cross validation has been used for model selection. The summary of results obtained for both kernels is presented in Table 3.3. Besides, ERSVM accuracy remained almost the same when N_{ev} has been decreased (this time from 11 to 5), and this accuracy is compatible with that provided by classic SVM, whereas N_{ev} is much lower than N_{sv}: 5 versus 145 for the Gaussian kernel, and 5 versus 221 for the hardware-friendly kernel.

In order to estimate minimal number of SVs that can be delivered by standard SVM, an alternative model selection has been also performed: 5-fold cross validation has been used, but the number of SVs (instead of accuracy) has been considered as the criterion. The best models were characterized by 46 SVs (Gaussian kernel) and 87 SVs (hardware-friendly kernel), both providing 92.9% accuracy. Summing up, even in this case the number of SVs is 9-17 times higher than that provided by ERSVM, whereas the accuracy is lower.

The most frequently misclassified samples for both SVM and ERSVM cases are presented in Fig. 3.11. As regards false negatives (two leftmost images), the first sample corresponds to the situation when the contrast between the person's clothes and the background is low. Background subtraction procedure led to a very "dark" sample, which has been misclassified. The second image represents a tricky situation: a person with extended arms, when fitting the entire person led to non-standard scale. As regards false positives (two rightmost images), the first image has been considered as a negative sample because of poor person centering and additional object presence (at the left). The object in the second image is a hall tree with a pullover and a backpack put on, which is undoubtedly a tricky and controversial negative objects, and SVM could permit itself to be mistaken in this case. So, the major part of these samples can be considered as outliers.

3.2.20 Conclusions

In this work we presented MicrelEye, a low-power and low-cost, yet computationally intelligent video wireless sensor node. The device architecture is based

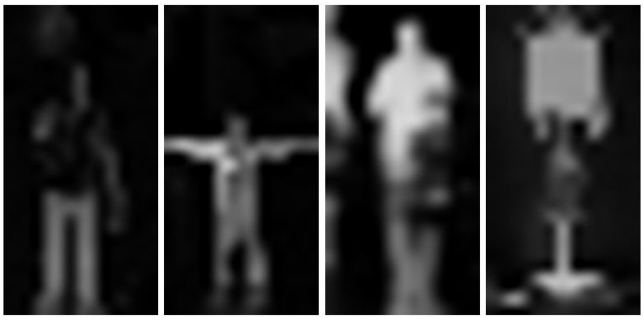


Figure 3.11: The most frequently misclassified samples. Two leftmost: false negative misclassifications, two rightmost: false positive misclassifications

Table 3.4: Total Power Consumption

Implementation	Power Consumption
Sequential	430 mW
Parallel	500 mA
Optimized	500 mA

on low-cost off-the-shelf "FPGA + microcontroller" system-on-chip from Atmel. It performs object classification locally on QVGA BW images at 15 fps with an average power of 0.5W, thus achieving 35 mJ per frame power consumption. This performance and power figures are already compatible with battery powered operation and can be further improved by aggressive power management, eventually moving to a low-voltage reconfigurable SoC architecture.

As regards classification, we proposed and implemented a new SVM-like hardware-oriented algorithm. As compared to standard SVM, the implementation of recently proposed hardware-friendly kernel along with significant reduction of the number of support vectors have led to at least one order of magnitude reduction of classification phase complexity with typical classification time being several milliseconds.

The case study considered in this work is people detection. The heterogeneous data sets have been generated. The obtained results, along with possibility to perform classification at as high rates as 15 fps with a average power of 500 mW, suggest that the present technology allows for the design of simple intelligent video nodes capable of performing local classification tasks, thus incrementing the notion of invisible and pervasive computing.

3.3 Abandoned/removed object detection for low power video surveillance systems

The demand for reliable surveillance systems is increasing, especially for mass transit and public areas such as airports, railway and subway stations, sport and concert event venues. For this reason, video surveillance systems that, through the analysis of video sequences, perform automatic detection of security-related events or aid human personnel at monitoring a place are gaining increasing interest. A key aspect for current video surveillance systems is the capability of reliably detecting common events such as abandoned and removed object within the scene. Typical scenarios are, e.g., detection of unattended packages in a railway station or in an airport [95, 92], and detection of stolen objects in a museum [64]. Nevertheless many proposals have recently addressed this specific task [130, 64, 95, 92, 24, 121, 141, 144, 140, 23, 129, 26], none of them are based on an embedded and unobtrusive architecture able to be long-term operating, to execute surveillance algorithms completely locally and to rise alarms wirelessly only when suspicious events happen.

We aim at filling this gap by proposing a multi-modal video surveillance system, characterized by low power consumption and low cost, and based on a CMOS video sensor and a Pyroelectric InfraRed (PIR) sensor. The use of the PIR sensor can notably reduce the overall power consumption of the system in absence of events, as shown in [102], where an embedded video system has been designed to detect structural effectively and rapidly changes in the monitored scene by jointly exploiting camera and PIR. The objective of this work is to propose a more advanced video analysis framework that, based on similar low-cost and low-power architecture, is able to detect events such as abandoned or removed objects.

Recently, applications which exploit Low-power Video Wireless Networks (LP-VWN) consisting of networks of low-cost video sensors connected by low-rate wireless channels and constrained by low-power budget, have gained increasing attention. LP-VWNs, in fact, represent a strategic enabling technology for a number of applications in surveillance, environmental monitoring, entertainment and health care. Designing a distributed video system within the tight power budget typical of mobile devices and wireless sensor networks is a very challenging task. Typical applications are in the domain of object detection or tracking.

When an event is detected, if the full image is not essential for the particular application, the system may transmit only some very limited amount of information, such as number of objects, size, position, trajectory, etc. saving a large amount of energy in wireless transmission and extending the autonomy of the batteries. Clearly, nothing should be done from the point of view of data transmission and power consumption if the targeted object is not detected because simple raw cameras are exploited. In this case, the detection of abandoned or removed objects can be performed only after the collection of continuous video streams transmitted to cumbersome power-unconstrained base station. Of course this approach would be extremely energy and bandwidth inefficient, difficult to port on stand-alone mobile embedded systems and ultimately not scalable in a network. Smart wireless video networks architectures are possible only if they are based on devices with an adequate trade-off between power consumption and processing capabilities, thus the key challenges we addressed are the development of energy-efficient algorithms and low-power architectures which can support vision-related tasks.

Research on low-cost video node design has been very active in the last years and a number of node prototypes have been designed [47, 37, 90, 166, 63, 66, 68]. We can classify these approaches in three categories: (i) low-cost nodes with wired interface (e.g., the node designed by Corely et al. at CMU [37]), (ii) wireless nodes with significant power consumption (e.g., the Panoptes nodes designed by Feng et al. [63]), (iii) application specific single ultra-low power single chip solution (e.g., the chip designed by Zhang et al. [166]). Nodes in the first category obviously do not satisfy the basic requirement of being wireless, while nodes in the second category consume roughly 10x more power than typical nodes in a wireless sensor networks. Finally, the single-chip solution have extremely low power consumption, but it is not programmable nor configurable in field. One important common point in current video wireless nodes of the first and second category is that the digital signal processing subsystem is the main power bottleneck. This is due to the fact that the high data rate of CMOS image sensor imposes the selection of fast processors and memories with high power consumption. Hence, the main open challenge in this area is to synergically develop algorithms and architectures for energy-efficient image processing without giving up the flexibility of in-field configuration.

Energy autonomy and efficiency of the implemented algorithms are undoubtedly the primary design challenges to be addressed on systems subject to low computational capabilities and memory constraints. Both issues are addressed by the integration of multi-modal information using additional ultra-low power PIR sensors which increases energy efficiency because the camera is triggered only when necessary and, in the same time, reduce considerably the average power consumption of the wireless video node because camera is in shutdown state in absence of events.

Other work presented a combination of video sensor with other low-cost and low-level sensors, which are used mainly for triggering the camera at the right time and not to promote a reduction of the system energy requirements. A distributed network of motes equipped with PIR, acoustic and magnetic sensors with adjustable sensitivity have been proposed in [72], stealthiness and effectiveness in a military surveillance applications. A network of IR sensors and cameras are used also in [126] to balance privacy and security in surveillance applications.

We present a video sensor architecture designed for low-power and low-cost video surveillance centered around a STR912F from ST-Microelectronics equipped with an ARM966E 16/32-bit RISC, 96 MHz operating frequency, 96 KB SRAM and several interfaces. We implemented an algorithm for detecting abandoned and removed objects within the scene which is optimized for low-power architectures constrained by limited computational capabilities. The main constraints when developing algorithms for such architectures characterized by small available memory is efficiency and timing performance. Furthermore optimizations have to be implemented taking into account that a floating-point unit is unavailable. However, experimental results demonstrate the quality of our multi-modal ARM-based approach. Moreover we analyze different configurations and characterize the system in terms of runtime execution and power consumption, comparing the results of efficiency with floating point implementations on personal computers.

The remainder of the section is organized as follows. In the next section we present the system architecture focusing on the constraints of energy budget, memory and computational capability offered by an ARM-based solution. The developed system and the description of the several power modes used by the application is also discussed.Section 3.3.3 depicts the algorithm implemented for the detection of abandoned/removed objects. In particular we discuss constraints and requirements of implementation on limited platform when optimizations are necessary. Experimental measurements and achieved performance are the focus of Section 3.3.4. Finally, Section 3.3.5 draws conclusions.

3.3.1 System architecture

The developed smart camera is showed in Figure 3.13 and it consists of three modules: an multi-sensor layer (MSL) equipped with an image sensor and a pyroelectric sensor, a processing unit(PU) based on ARM9 architecture, and a wireless communication unit (WCU), as shown in Figure 3.12.

The MSL includes a small PCB with 1 megapixel color CMOS imager VS6624.It supports up to 15 fps SXGA with progressive scan and up to 30 fps with VGA format with a typical power consumption of $120\,mW$ when active, while it

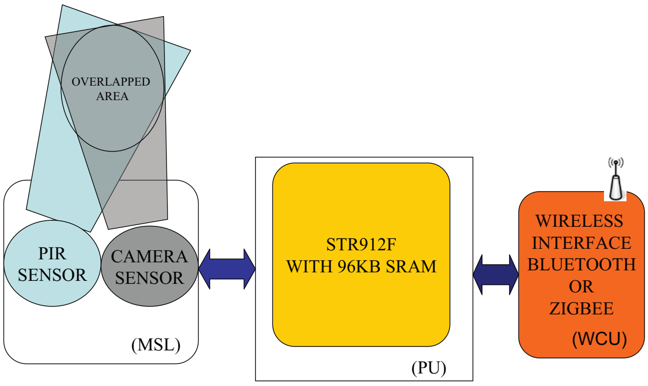


Figure 3.12: Video sensor node architecture.

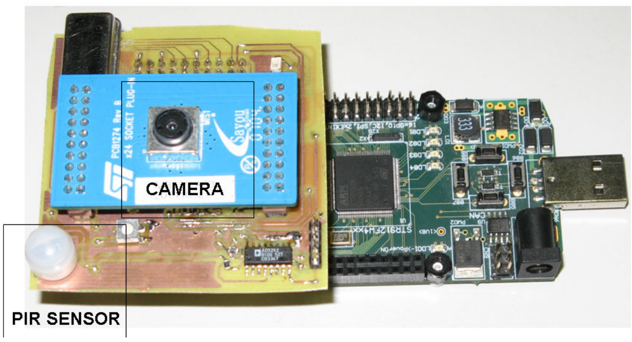


Figure 3.13: Developed prototype of the video sensor node.

decreases down to $23\,mW$ in stand-by mode. The system exploits PIR Sensor typically used in surveillance to provide simple, but reliable, digital presence/absence signals. The video sensor and the PIR sensor are built to cover the same field of view, in this way the PIR sensor can be aware of the the movements in the scene triggering the detection algorithm. The MSL is directly fitted into a PU board which is employed for digital image processing using single-cycle DSP instructions with configurable and flexible power management control. For example the typical current consumption for this microcontroller is about $1,7\,mA/MHz$ in RUN mode and only a few mA in SLEEP mode which is an attracting feature for wireless sensor networks design where the power consumption is a major constraint. Finally wireless communication is guaranteed by a Bluetooth transceiver adopted because of the bandwidth and the easy interface to host devices (i.e. PC, PDA). However, ZigBee radio interface is also supported.

The main goal of our system is to perform automatic detection of events such as the presence of abandoned and/or removed objects in the scene using non unobtrusive embedded platforms. Other specifications concern the need for low power consumption, the use of a PIR sensor to reduce the presence of false positives, and the possibility of sending an alarm to a remote host wirelessly. To satisfy the requirements, the information coming from the PIR sensor is used to "wake up" the system in occurrence of specific events, as well as to evaluate when to start the video analysis stage. In fact, if the PIR sensor does not identify any event, the camera is switched off and the microcontroller is set to SLEEP mode minimizing the power consumption.

Figure 3.14 shows the flow chart of the application. When triggered by an event from the PIR sensor, the system switches to RUN mode the ARM core, which runs full speed and all clocks are on, while the camera is kept off until movements in the field of view disappear. Then the camera is activated and takes a picture of the environment which is processed by the detection application, described in Section 3.3.3, then the system switches back into SLEEP mode where the power consumption decreases up to 90% since only the PIR module operates as reported in next sections. This way the number of false positives is minimized beacuse the system processes the frames only in absence of moving objects in the monitored area enhancing robustness and autonomy. Finally, when an object is recognized as abandoned or removed, the system sends wirelessly alarms containing the number of objects, the regions of interest, size... and the full picture if requested by the host. In power characterization presented in this work, we considered a Bluetooth interface and we decided to send the full content of the image in order to estimate the autonomy of the platform.

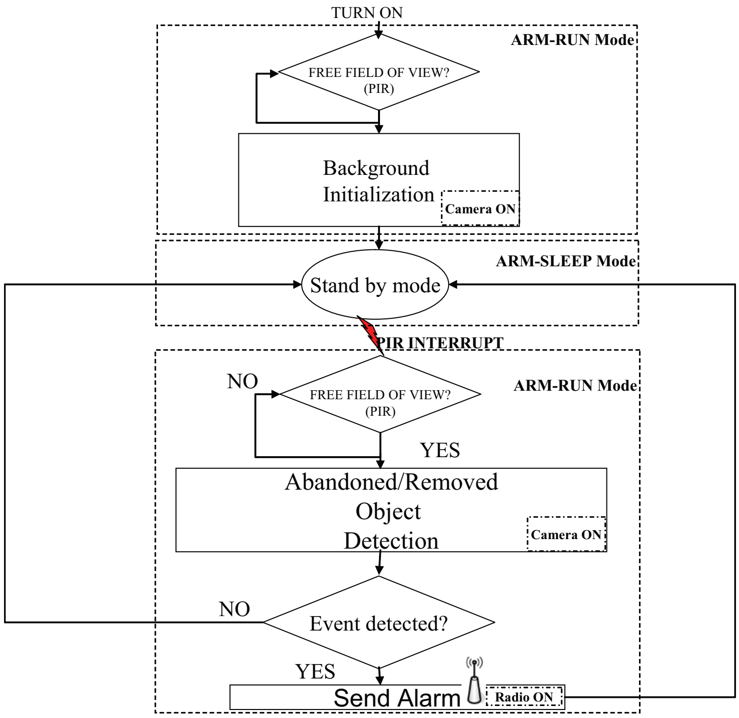


Figure 3.14: Flow chart of the application.

3.3.2 Pyroelectric InfraRed (PIR) Sensor Nodes

Pyroelectric InfraRed (PIR) sensors are devices able to transduce changes in their temperature, due to incident infrared radiation, into an electric signal. A pyroelectric element behaves like a polarized planar capacitor whose charge varies according to $\Delta Q = A \cdot p \cdot \Delta T$ (where A is the area of the sensing element and p is the material specific pyroelectric coefficient). Typical PIR sensors embed 2 elements placed in series with opposite polarization. As a consequence when a body moves in front of the sensor 2 peaks, one positive and one negative, are produced (see figure 3.15).

PIR sensors are used in conjunction with an array of Fresnel lenses used to shape the sensor Field of View.

Our prototype PIR sensor board has been designed using Commercial Off-the-Shelf (COTS) components. The detector is Murata IRA E710 [111] and the signal conditioning circuit is a double stage amplifier, which achieves a total gain of about 1400 and operates as a band-pass filter between 0.57Hz and 11Hz. This is a suitable range for detecting moving people [118]. Furthermore, it biases the output voltage at $\frac{V_{dd}}{2}$ when no movements are detected. The conditioning circuit board includes also a low power voltage regulator used to decouple power supply lines from the transceiver ones and a comparator used to generate a wake up signal when the board is in a low power state. The sen-

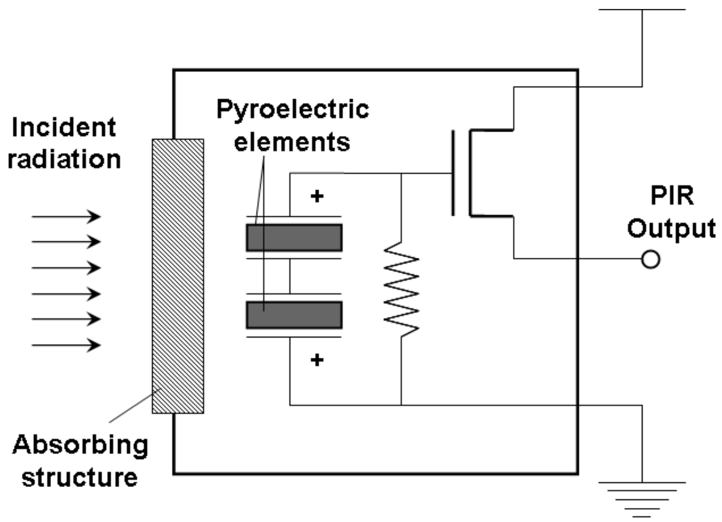


Figure 3.15: PIR schematics and output when passages in the two directions (left to right and right to left) occur.

sor and its conditioning circuits are hosted in the package of a PIR presence detector, IS-215T [75].

3.3.3 The video analysis algorithm

This section describes the video analysis algorithm which is applied every time the intrusion detection block based on the PIR sensor detects absence of movements in the monitored scene and captures a new image from the scene activating the camera. By means of the PIR sensor, we can assume that all visible changes appearing in the scene in absence of movements have to be considered possible instances of removed or abandoned objects. Hence, a first stage of the algorithm consists in a background subtraction approach aimed at detecting visible changes in the scene background. Then, a labeling algorithm is implemented to enumerate and locate the areas of the image, or Regions-of-Interest (ROIs), where a stationary change of the background has taken place. Finally, a blob analysis stage provides the classification of each ROI between abandoned and removed object. All stages of the proposed video analysis algorithm have to be particularly memory efficient and need to avoid the use of floating point instructions given their implementation on the embedded architecture. Figure 3.16 shows the flow diagram of the algorithm.

Background subtraction To detect stationary visible changes in the scene, we adopt a typical *background subtraction* approach, that is we compare the current frame captured from the camera, F, with a model of the background of the scene, B, computed at initialization time. To do this, each pixel at coordinates

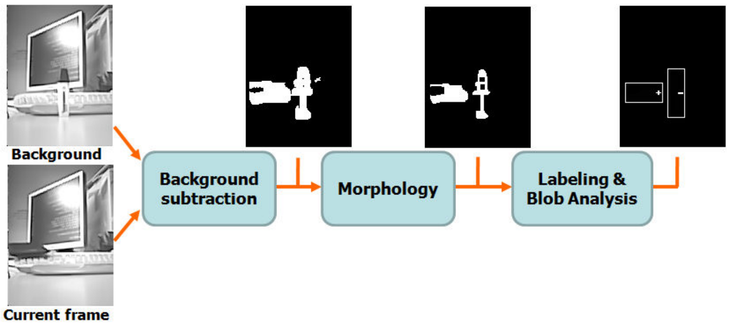


Figure 3.16: The flow diagram of the proposed change detection algorithm.

(x, y) in the current frame is compared with its homologous in the background model by means of a function aimed at measuring the similarity between the two image points.

To deal with illumination changes and photometric distortions that typically occur in real working conditions and may easily be misinterpreted as structural changes, we compute the Normalized Cross-Correlation (NCC) [145], which is invariant to linear photometric transformations between corresponding windows on F and B, on a squared neighborhood (i.e. a window of radius r) centered on the pixel under evaluation:

$$NCC(x, y) = \frac{F(x, y) \circ B(x, y)}{\| F(x, y) \|_2 \cdot \| B(x, y) \|_2} \tag{3.10}$$

where the term at numerator is the dot product between $B(x, y)$ and $F(x, y)$, and the two terms at denominator represent the L_2 norms of $F(x, y)$ and $B(x, y)$, respectively.

Then, the NCC function is thresholded yielding a binary image, referred to as *change mask*, C, which highlights those parts of the current frame which have been subject to a change with respect to the background model:

$$C(x, y) = \begin{cases} changed, & NCC(x, y) < \tau_{NCC} \\ unchanged, & otherwise \end{cases} \tag{3.11}$$

The use of the NCC is motivated by the fact that the system ought to be robust toward these kinds of distortions which can typically be found since the background model is computed once at initialization. On the other side, the implementation of the NCC function is particularly simple compared to more advanced approaches, and this aspect is particularly relevant since the algorithm has to be implemented on an ARM-based embedded architecture using a fixed point approach to maximize performance. In particular, to perform the square root and division operations of (3.10) a fixed-point square root function for ARM and a integer division have been utilized.

A typical effect of the use of the NCC over a window is that the segmentation of the foreground in the change mask becomes less accurate along the borders of the objects. In particular, there's a typical *fattening* effect, that is the object appears bigger since its borders are increased by a number of pixels proportional to r. To deal with this effect, a simple binary morphology operator of erosion is applied on the change mask as many times as the chosen value of r.

Labeling After the background subtraction stage, a labeling algorithm is applied to group together connected components of the change mask. In this case, we use the algorithm proposed in [48], which is an efficient algorithm with low memory requirements for the labeling of binary images. In particular, the algorithm only requires two image scans and it has a memory complexity of $O(1)$. Once the labeling is performed, another image scan is deployed to compute the ROI coordinate of each connected component. Then a simple area-closing approach is performed to eliminate spurious components that might have been generated by noise.

Blob analysis In the last stage of the algorithm, each valid ROI is classified either as an abandoned or removed object. The key idea beyond the adopted classification algorithm is that if an object is abandoned on the background, in F the number of edges along the borders of the corresponding connected component should increase compared to B. Conversely, if an object is removed from the background model, then F should display much less edges along the borders of the area where the object was initially located compared to B.

Hence, the approach relies on the estimation of the number of edges that appear on F along the borders of the connected component we want to classify. First of all, we detect all *contour* points within the ROI as those points that belong to the foreground and have at least one of their 8-connected neighbors set as background. On each contour point of coordinate (x, y), we compute the horizontal and vertical derivatives D_x, D_y of point $F(x, y)$ by means of the Sobel operator Then, we approximate the magnitude of the gradient in (x, y) as:

$$|G(x,y)| = max\left(|D_x(x,y)|, |D_y(x,y)|\right) \tag{3.12}$$

A threshold is used to classify the contour point as being or not in presence of an edge in F. Then, the number of contour points associated with edges, N_{CE} is computed and thresholded:

$$Class(x,y) = \begin{cases} removed, & N_{CE} < \tau_C \\ abandoned, & otherwise \end{cases} \tag{3.13}$$

to yield final classification of the ROI.

3.3.4 Experimental results

The above-mentioned application was fully implemented in ARM9 firmware. In the following we will focus on video sensor node power and performance. Since for this work we used only the internal 96KB SRAM, the camera is set to grab a 160x120 pixel (QCIF) gray scale image in YCbCr 4:0:0 format. The amount of byte for one image in this format is only 19200bytes, since each pixel uses only a byte. The abandoned/remove algorithm needs at least 3 images to work properly. In fact we need a stored background to achieve the NCC background subtraction and two images to store the change mask and the eroded image. For this reason, the total amount of RAM to stored all the required images grows up 76800bytes.

Power consumption is reported in Table 3.5a) while Table 3.5b) depicts also the processing time necessary to discriminate if objects are abandoned or removed from the environment. The time to elaborate the blob analysis depends on the number and size of ROIs. So it will be zero if the system does not detect any blob and about 100 ms for three ROIs 16x16. These results show how the power consumed by the whole system in SLEEP mode is less than 10% of power requirements of a fully active node. So without the information of a low-cost PIR sensors, the systems would waste the 90% of its energy, in the worst case. Moreover through PIR sensor information, the platform is able to switch on the camera as late as possible, reducing the camera power consumption again of around 20%. Moreover, the power consumption of wireless communication is minimized because of higher accuracy of the detection reduces the number of false positive.

To perform a quantitative evaluation of the abandoned/removed object detection algorithm, a dataset of images was acquired under real conditions within two sessions which differ by location and illumination conditions. A total of 50 images has been collected, each one showing different objects and simulating the frame collected by the system when the camera is switched on. In particular, each image includes a number of abandoned/removed objects that varies between one and three. Different tests with different backgrounds, chosen among the images of the dataset, have been performed, for a total of 141 cases of abandoned/removed objects tested (70 abandoned objects, 71 removed objects). Figure 3.17 shows a subset of the dataset.

In terms of change detection, our algorithm detected a total of 162 objects. In particular, it was always able to detect the presence of objects placed in the scene, with a percentage of false negatives (missed detections) equal to 0%.

(a) Power consumption of the video sensor node.

Component	Power $[mW]$
ARM9 mode (RUN / IDLE / SLEEP)	450 / 49,5 / 15
Video sensor mode (ON / IDLE)	165 / 23
TX/RX mode (ACTIVE / IDLE)	98 / 10
PIR sensor	1,5
Video Node	
Active with/without video sensor	626,5 / 484,5
Alarms Transmission	572,5
SLEEP, only PIR is Active	51

(b) Energy requirement of each task.

Task	Energy $[mJ]$	Time $[ms]$
Frame Acquisition	58,5	93,5
NCC Background Subtraction	455,8	940
Labeling	29	60
Blob Analysis	0 - 48,6	0 - 95
Image 160x120 Transmission	601,1	1050

Table 3.5: Energy requirements of the low-power video system.

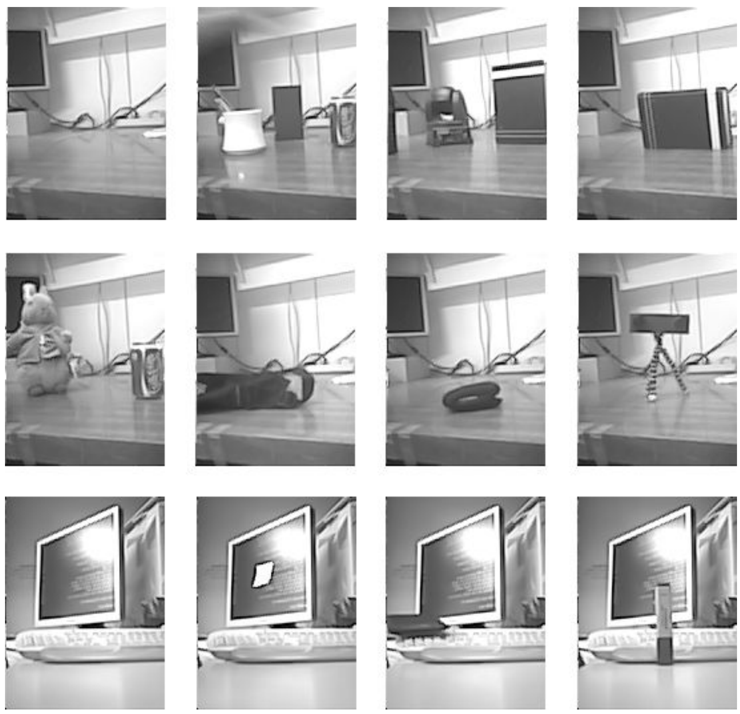


Figure 3.17: Subset of the dataset used for the experimental evaluation.

Instead, there's a number of false positives (false alarms) equal to 13% of the total number of detected objects.

To evaluate the fixed point approach we used the same datasets of images to compare the changed mask obtained from a NCC implementation on a floating-point Pentium4 architecture and on the presented fixed-point ARM-based solution. The difference concerns only 1% of the number of the pixels pointed out from the NCC implementation on a PC. However, after the morphology operator of erosion, the accuracy of ROI detection on fixed-point ARM is not degraded with respect to the implementation on a Pentium4.

As for the performance reported by the classification algorithm, it yielded a number of misclassified objects equal to 7.8%. In particular, the percentage of correct detection for the removed object class is 98.6%, while the percentage of correct detection for the abandoned object class is 85.7%.

3.3.5 Conclusions

The interest in low-cost and small size video surveillance systems able to collaborate in networks of detection systems has been increasing over the last years. In this section we have presented a multi-modal video sensor node designed for low-power and low-cost video surveillance which is able to detect objects abandoned or removed in the environment. The system is multi-modal and a PIR sensor assists a CMOS video camera to increase the efficiency of the algorithm and to extend the life time of the system. We addressed different configurations and characterized the system in terms of runtime execution, power consumption and efficiency.

Chapter 4

Multimodal surveillance

4.1 Overview

Video surveillance and other security-related applications have gained many credits due to the terroristic threats of the last years. Several industrial and academic projects have recently started to increase the accuracy of (semi) automatic surveillance systems. In addition, the abatement of hardware costs allows the deployment of thousands of cameras for surveillance purposes at a reasonable cost.

The ever-increasing demand of security and the low cost of cameras contributed to the diffusion of the research in distributed multi-camera surveillance systems. Multiple cameras enable the surveillance of wider areas and the exploitation of redundant information (provided by the different viewpoints) might solve classical limitations of single-camera systems, such as occlusions.

Moreover energy efficiency for wireless smart camera networks is one of the major efforts in the distributed monitoring and surveillance community. If video cameras are equipped with circuits that receive and convert energy from regenerative sources such as solar cells, an effective power management becomes essential for the design of small sized and perpetually powered devices, which can be deployed unattended for years and feature smart vision applications.

Pyroelectric InfraRed (PIR) detectors take advantage of pyroelectricity, which is the electrical response of a polar, dielectric material to a change in its temperature, to detect a body at thermal disequilibrium with the surrounding environment. These sensors are typically used in commercial applications to detect presence of individuals to trigger alarms and can be used to have a multimodal surveillance system.

PIR sensors can be integrated within a video surveillance network also to increase the lifetime of Wireless Video Sensor Nodes (WVSN). Low-cost video surveillance systems based on wireless sensor networks will hit the market with the promise of flexibility, quickly deployment and providing accurate real-time visual data. However, many technical problems have to be still overcome for a widespread diffusion of such a technology. For instance, even if research continues to develop higher energy-density batteries, capacity constraints limit the lifespan of common wireless sensor nodes. For this reason, energy-aware design and maximization of the sensor network lifetime become the major key research challenges for WVSN and their applications.

To enhance vision sensor networks, two successful strategies can be adopted:

1. exploiting alternative power sources which increase the autonomy of the nodes considerably;

2. exploring multi-modal sensor integration which can save on-board power consumption

Recently, several researchers have proposed alternative power sources and Energy Scavenging techniques to extract and convert power from the surrounding environment and to replenish energy buffers like batteries or supercapacitors. In particular, photovoltaic (PV) harvesters are the most promising to enable perpetual operation of WSNs [28, 139]. Unfortunately if the power consumption of a device can be estimated, the power generated by a PV module changes non-linearly under varying temperature or solar irradiance and techniques which automatically tune the operating point of the solar cell should be considered to provide the maximum output power.

From the sensor capability point of view, CMOS imagers are generally high-power consuming devices and accuracy of the information increases the required power. Therefore they should be activated very carefully in order to save energy and their functions could be replaced by low-power low-level vision devices during the idle intervals, when the density of the events or the energy stored is low. Being able to detect variations of incident infrared radiation, due to movement of bodies not at thermal equilibrium compared the environment, the use of a network of PIR may lead to the extraction of more complex data such as object direction of movements, speed, distance from sensor and other characteristics [137]. The combination of several vision devices with heterogeneous features allows the development of multimodal surveillance applications with efficient energy policies. In fact, video would still provide high-level information when required, and PIR sensors would assure a continuous monitoring service triggering the CMOS camera when an event is detected.

In this chapter we present the design, implementation and characterization of a self powered video sensor node, able to detect people and supported by PIR sensors to enhance energy efficiency [101, 100]. Moreover we show a simple but optimal power management tailored for multi-modal video sensor nodes and based on model predictive controller (MPC)[99]. Finally we propose a cooperative policy to manage power consumption of a WVN powered by solar scavengers and supported by a network of PIR sensors that perform a coarse classification of movements.

4.2 A solar-powered video sensor node for energy efficient multimodal surveillance

Building an energy efficient wireless vision network for monitoring and surveillance is one of the major efforts in the sensor network community. In this section we describe an application for people detection, which exploits both network architecture flexibility and on-board processing capabilities. The application, based on support vector machine engine (SVM), is able to detect events (e.g. when the environment is changed due to the movement of subject in the scene), and distinguishes the presence of people or human bodies rather than objects or animals in the field of view before generating alarms or sending information through the wireless link. We focus on the design, implementation and characterization of a self-powered video sensor node, able to detect people and supported by PIR sensors to enhance energy efficiency.

The video sensor node is designed to support flexibility in terms of distribution of the processing tasks across the network and is powered by a solar scavenger using a 70 cm^2 photovoltaic panel. Keeping the nodes constantly active is clearly impracticable, because of the power consumption of components such as imager, transceiver and microprocessor. Therefore the proposed architecture follows a hardware/software hierarchical design with three layers which can be separately activated, as showed in figure 4.1.

The figure considers a hypothetical surveillance scenario where events occupy the 4% of the time and only 20% of them results in an alarm to report. The objective is to wake up the video acquisition only in presence of people and to reduce the number of not-interesting events in order to guarantee longer lifetime while the system is recharged by a fluctuating and unpredictable energy source. Once the video is waken-up, the node locally classifies input images and wirelessly sends to a base station only relevant ones, thus saving energy by reducing the amount of transmitted data.

We developed a novel method to modulate the status of each layer by ex-

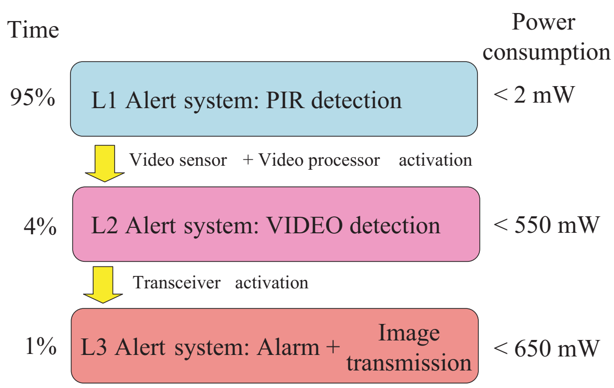


Figure 4.1: Hierarchical design of the video sensor node, with three different layers for the alert system.

ploiting a PIR based wake-up circuit and local image processing. The sensitivity of the trigger signal from the PIR detector is adjusted dynamically according to the available energy in the reservoirs, the average contrast of the images taken from the scene and the probability of seeing a person in the camera FOV.

4.2.1 Related work

Recent years witness a rapid growing of research and development of surveillance and multimodal applications using multiple sensors, including video and other kind of sensors. The aim of such systems is both to overcome some points of failure of a particular kind of sensor and to balance different parameters fixed by the application among which power consumption plays a central role.

Power management is a critical issue when dealing with wireless sensor networks and it is well known that batteries does not scale as much as electronic device [116] thus posing a severe limitation in the achievable unobtrusiveness. Also the cost of batteries often exceeds the one of nodes. At last, in some application, it may be not possible to reach the sensors (i.e. due to dangerous environment, like battlefields) in order to replace batteries.

In [70] the authors attempt to formalize and analyze the trade-off between power conservation and quality of surveillance in target tracking sensor networks. In [165] a dynamic sensor selection is applied to efficiently use available sensor energy and extend overall network life. Another attempt to extend network life by capitalizing on low power states of its node can be found in [19]. In this work the amount of data collected by the system is tuned in order to minimize power consumption while achieving high accuracy. Finally in [72], a distribute network of motes equipped with acoustic and magnetic sensors have been deployed in order to achieve longevity, adjustable sensitivity, stealthiness and effectiveness in a military surveillance application. Since in this paper the authors aim at achieving longevity through sensor selection techniques, they use a high number of low power nodes with low resolution (magnetic field detector) and network life extension is obtained by reducing number of active sensors when any activity is detected and successively wake them up. In contrast we have a unique sensor, which provides much more information and we modulate its activity through the use of another low power sensor.

In contrast to the work presented in this session none of the cited works attempted to reduce the node power consumption except using low power hardware, and they either do not consider a stochastic source of energy as the one provided by an energy scavenging system.

4.2.2 System architecture

The hardware architecture of the solar-powered video sensor is displayed in section 3.3.1 and consists of several modules: the solar harvesting unit, the vision board which hosts both the CMOS imager and the PIR sensor with a common area under monitoring, the wireless module, the microprocessor and other peripherals.

Computational unit and CMOS imager

The core of the video node consists of an STR91xF microprocessor from STMicroelectronics with an ARM966E 16/32-bit RISC architecture, 96 MHz operating frequency, 96 KB SRAM and several peripheral interfaces that can be disabled if not used. The microprocessor provides the high-speed logic interface necessary to capture images from the camera and processing data for people detection or object classification, it also offers configurable and flexible power management control through operative frequency scaling.

The vision module includes a SXGA CMOS color digital camera targeted for mobile applications featuring low-size and low-power consumption and a Pyroelectric Infrared Detectors, which detection area is overlapped with the field of view of the video sensor.

Wireless communication capabilities have been supported through a suitable interface for both Zigbee and Bluetooth compliant transceiver. The module has a stackable design as the sensor node, hence the wireless layer is easy to replace. We implement hardware and software interfaces in order to host different wireless standard used in wireless sensor network community such as Zigbee and Bluetooth or proprietary protocols. All the performance and measurements discussed in this section are referred to the version with Bluetooth capability.

Figure 3.13 shows the developed prototype, the whole system is designed with low power consumption as the primary goal. The system is powered by an energy management module which hosts solar harvesting capability. The solar cell used to replenish the energy reservoirs has a nominal output power of $500\,mW$ under full outdoor irradiance and a harvesting circuit extracts the maximum power available from the solar cell following the optimal operating point at the minimum energy cost.

4.2.3 Energy harvesting unit

Energy harvesting is a low cost-effective operation, in term of energy harnessed, device size and efficiency. One of the primary issues to address is minimizing the power consumed by the harvester itself. Less power will require the circuit, faster will be the growth of the harvested energy in the accumulator.

The *I-V* characteristic of a *PV* module is given by the following equation:

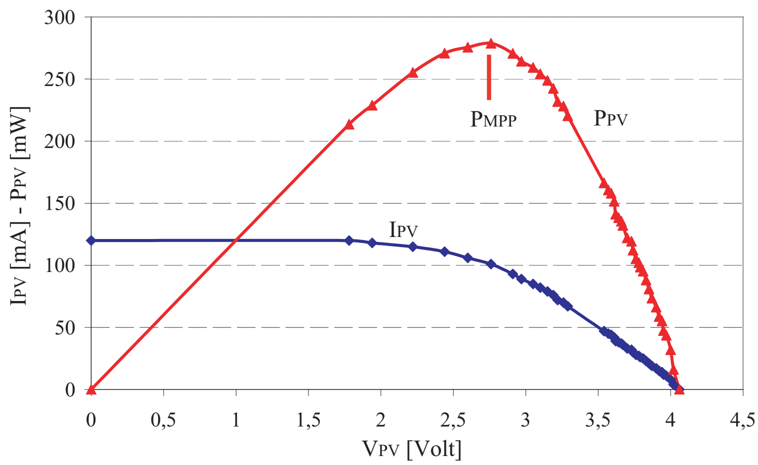
$$I_o = I_g - I_{sat}\left\{e^{\frac{q}{AKT}(V_o+IoR_s)} - 1\right\} \tag{4.1}$$

where I_g is the generated current, I_{sat} is the reverse saturation current, q is the electronic charge, A is a dimensional factor, K is the Boltzmann constant, T the temperature in degree Kelvin, R_s the series resistance of the cell. The internal shunt resistance is neglected in this model. The plot of the PV module adopted in our solar harvester is shown in figure 4.2(a).

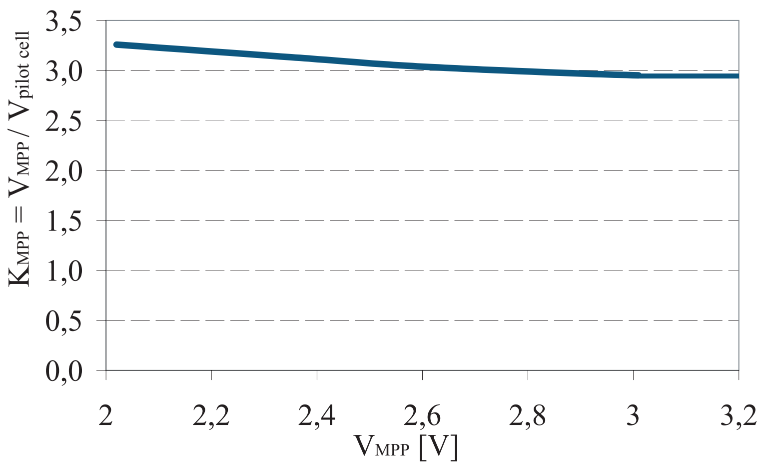
One key design challenge is how to optimize the efficiency of solar energy collection under non stationary light conditions and therefore maximum power point tracking techniques (MPPT) aim to automatically find the operating point (V_{PV}, I_{PV}) at which a *PV* module should operate to provide the maximum output power following it when light intensity changes. There are several methods and algorithms to track the MPP [55], we adopt one based on Fractional Open-Circuit Voltage (FOCV) which is the most used and cost-effective in medium and small-scale solar harvester. This method exploits the nearly linear proportional relationship between the operating voltage at MPP (V_{MPP}) of the main photovoltaic module and the open circuit voltage of a small additional PV array used as pilot-cell ($V_{pilot\,cell}$) under the same light L and temperature T conditions (4.2).

$$V_{MPP}\left(T,L\right) \approx K_{MPP} \cdot V_{pilot\,cell}\left(T,L\right) \tag{4.2}$$

We adopt the CPC1824 from Clare, Inc. [107] for the pilot-cell. It is a mono-



(a) I-V and P-V plots of the used photovoltaic module.



(b) Near linear relation between V_{MPP} and the pilot cell.

Figure 4.2: Characteristic of the photovoltaic module.

lithic photovoltaic module of only $9\ mm^2$, and it works as irradiance sensor providing feedback information to the harvester. The pilot cell follows almost linearly the behavior of the main PV module during light variations. As shown in figure 4.2(b), the ratio between the operating voltage at the MPP of the main module and $V_{pilot\,cell}$ is almost constant under several solar intensities.

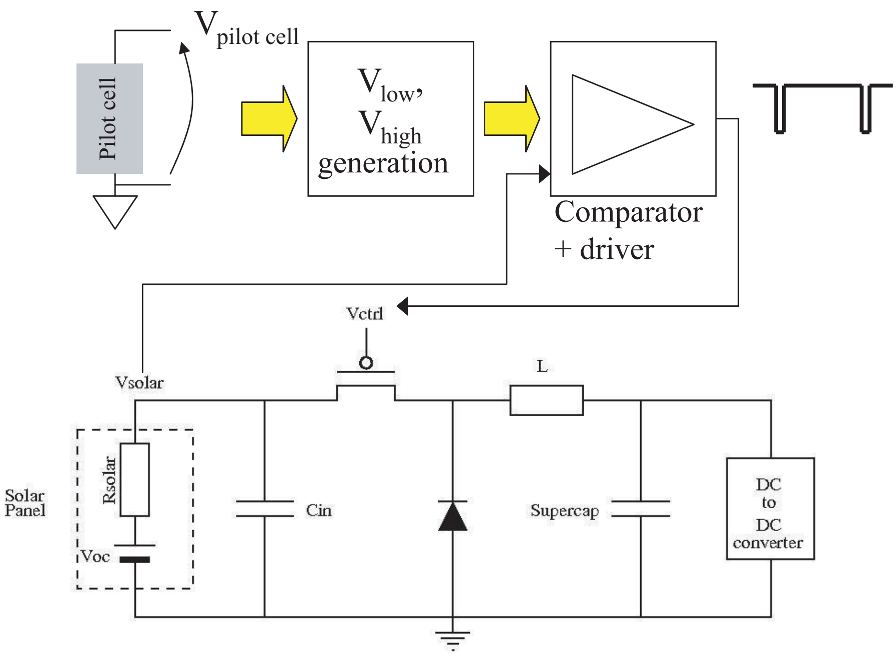


Figure 4.3: Conceptual schematic of solar harvester: buck power converter and MPP tracker.

Figure 4.3 depicts the schematic of the solar scavenging circuit for the video sensor node. By measuring the pilot-cell voltage the circuit estimates the MPP of the main module generating a lower and an upper threshold around its value. Then an ultra-low power comparator continuously checks the operating point of the main cell to the thresholds adjusting dynamically the duty cycle and the frequency of the control signal which drives the power converter circuit. Solar energy harvesters usually exploit buck configuration because the voltage level of the energy reservoirs is lower than the nominal operating voltage of the solar cell. In our implementations we exploit supercapacitors as energy storage devices, since they overcome many drawbacks of batteries that are critical in WSN applications and for long-live maintenance-free embedded systems. The harvester achieves an efficiency of the 80% and depending on solar irradiation can provide a maximum output power of about $500\,mW$ while the power consumed by energy harvesting process is less than $1\,mW$.

4.2.4 PIR Model analysis

Figure 4.4 shows the PIR output as a function of distance.

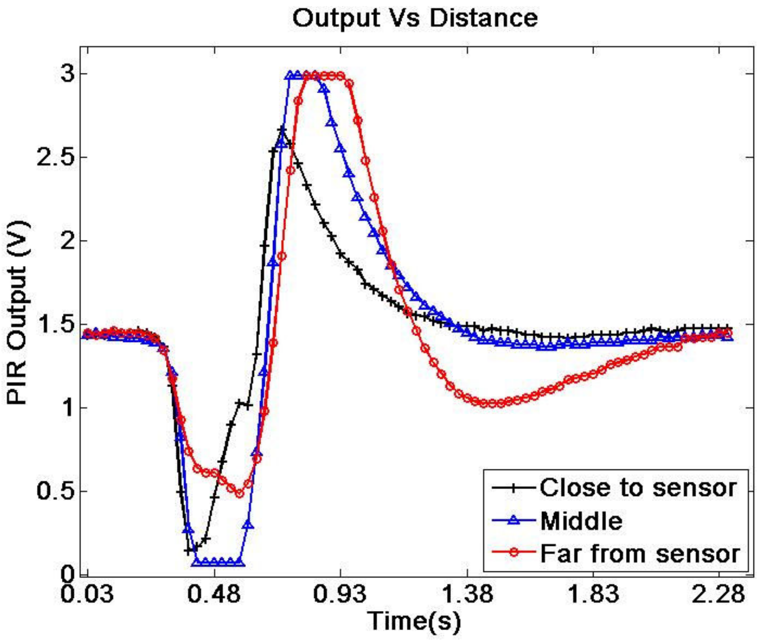


Figure 4.4: Output of a PIR sensor in case of passages at different distances.

From this plot, we can see how signal duration increases with distance while signal amplitude is at a maximum for passages in the middle position.

Signal duration increase is due to the FoV conic shape. In fact, a PIR is mostly sensitive to entrances and exits from its FoV and these two instants are more distant when a person walks far from the sensors.

Output peak-to-peak amplitude decreases with distance because far bodies result in a smaller change in the incident radiation. Amplitude reduction for closer passages is due to the interaction of the two sensitive elements. In figure 4.5 we highlighted each elements' FoV. In proximity of the sensor the two FoVs are overlapped, thus compensating each other.

In case of isolated people, each passage can be easily segmented using two thresholds above and below $\frac{Vdd}{2}$. The starting of the passage is detected when one of the threshold is broken, the end when the PIR output remains between the threshold for a certain time T. According to results from previous work of our group [164], we placed the thresholds at $\frac{Vdd}{2} \pm 300mV$ and $T = 1sec$.

When a passage is detected, each sensor extracts its duration and the PIR output amplitude. These two features are wirelessly sent to a central unit in order to evaluate the distance of passage, thus reducing the power consumption related to wireless communication and the bandwidth required. The central unit calculates the ratio between homogeneous features (duration and ampli-

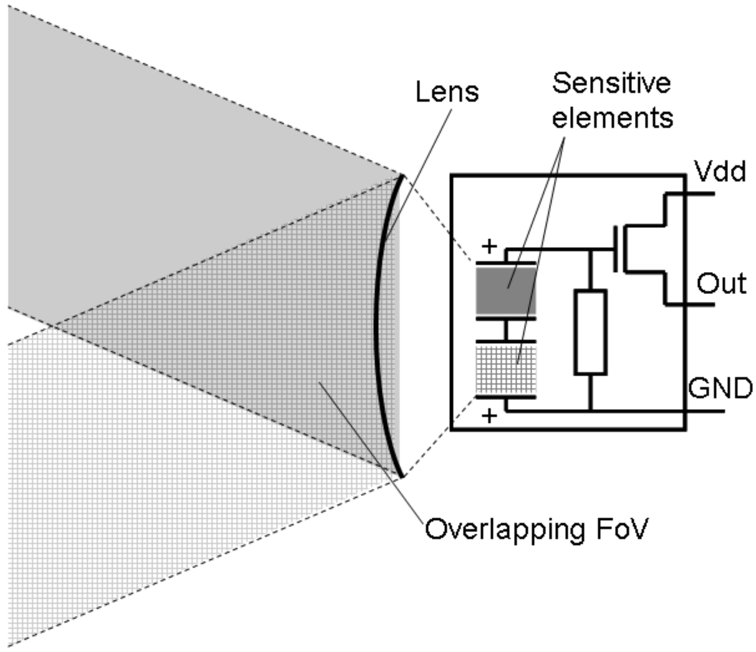


Figure 4.5: Schematic of a typical C.O.T.S. PIR. Two sensing elements are used in series with opposite polarization, the output is pre-amplified through a built in MOS transistor. Highlighted with shading, the FoV of each sensing element. Notice how, in proximity of the device, the two FoVs are overlapped.

tude). Therefore each passage results in a two-elements vector of features (relative duration and relative amplitude) with whom we estimate the position of the person ((see figure 4.6)).

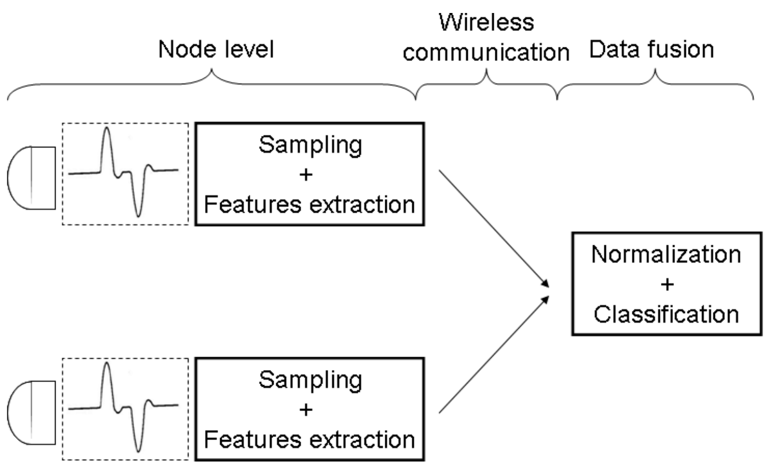


Figure 4.6: Task allocation for distance detection.

In figure 4.7, we plotted such vectors for a subset of samples from passages at different distances. As can be seen from this figure, it is not possible to define well separated region of the space for each distance of passage, so we decided to rely on a classifier in order to estimate it.

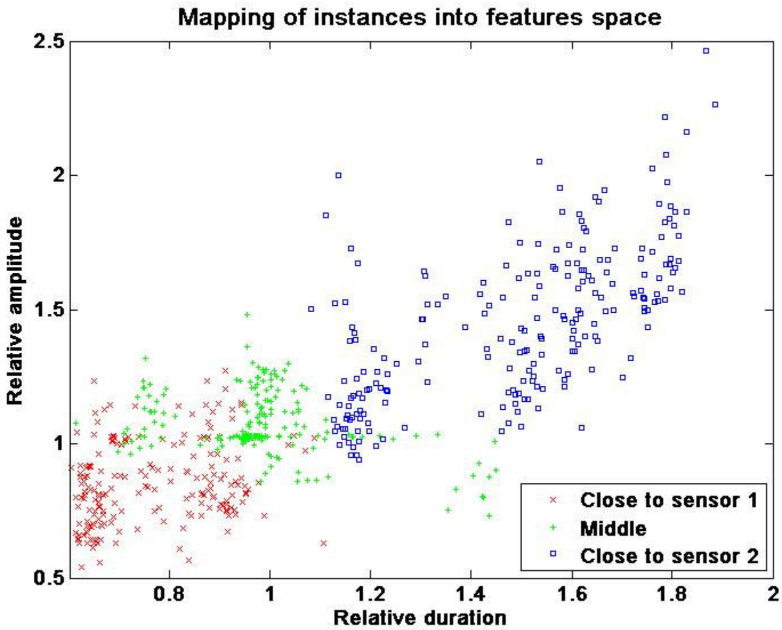


Figure 4.7: Mapping of input vector in the two dimensional feature space. The three classes are located into partially overlapped areas of the space.

4.2.5 PIR sensors wake-up unit

As in the other works presented in this section (see sections 3.3.2) we used a commercial PIR detector that includes 2 sensitive elements placed in series with opposite polarization. The details of this device have been presented earlier in section 3.3.2, a schematic of this device is presented in figure 3.15.

In particular in this work we are interested in the amplitude of the output signal which, outside the area where the FoV of the 2 elements is overlapped (see figure 4.5), is inversely proportional to the distance from the detector as can be seen in figure 4.8.

The sensor output signal is conditioned as in 3.3.2

In addition to the amplifier we designed a trigger with adjustable threshold. The schematic of the circuit is presented in figure 4.9. Here the series of R1, R2 (where R1=R2) and the digital potentiometer produces the 2 thresholds which are symmetrical to $\frac{V_{dd}}{2}$ and their reciprocal distance increases with the resistance of the digital potentiometer. When the amplified output breaks one threshold it generate an interrupt for the Video node core. Thus, by on-line programming the potentiometer we can adjust the sensitivity of the wake-up signal.

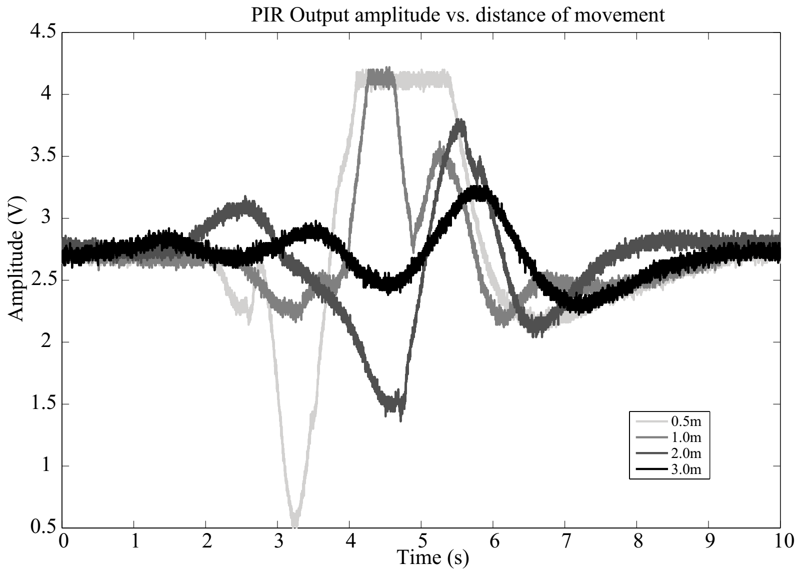


Figure 4.8: Output of a PIR sensor when a person moves at different distances

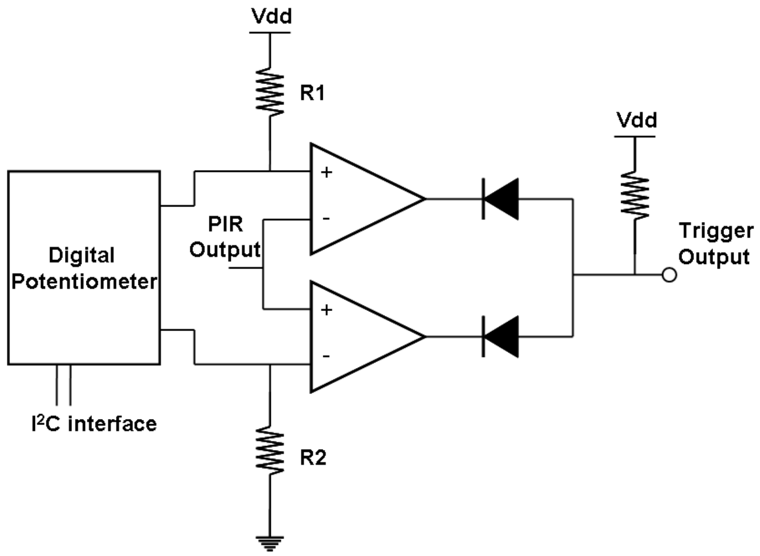


Figure 4.9: Schematics for trigger generation using PIR output signal.

4.2.6 System analisys

4.2.7 Sensor node characterization

The ARM microprocessor STR91xFoffers configurable and flexible power management control which allows dynamic power consumption reduction. It supports three global power control modes: RUN, IDLE and SLEEP. SLEEP mode is used by the video sensor node when no events are registered in the filed of view. When triggered by an event from the PIR sensor, the system switches into

Component	Power $[mW]$
ARM9 (RUN mode)	450
ARM9 (IDLE mode)	49,5
ARM9 (SLEEP mode)	15
Video sensor (ON mode)	165
Video sensor (IDLE mode)	23
TX/RX module (ACTIVE mode)	98
TX/RX module (IDLE mode)	10
PIR sensor	1,5
Solar Harvester	0,98
Video Node (Active)	650
Video Node (Sleep)	50

Table 4.1: Power consumption of the video sensor node.

RUN mode starting the detection application until the PIR trigger events or regions of interest are discovered in the current image, then the system switches back into SLEEP mode where the power consumption decreases up to 90% since only the PIR module operates. Power consumptions are reported in table 4.1.

4.2.8 Human detection application

Figure 4.10 presents the main steps of the implemented algorithm for human body detection. After triggered by the PIR sensor, all the system wakes up and the CMOS imager acquires and sends a frame to the microprocessor with YCbCr 4:0:0, grayscale, 8-bit format. In order to isolate a 128×64 region-of-interest (ROI) of the event we initially perform a background subtraction using the three-frame algorithm sub-image [80]. A pixel-by-pixel subtraction is performed using the first and second frame stored in the memory, then another pixel-by-pixel subtraction uses the second and third frame. Finally the two results pass in a logical AND to have a difference-image that allows to detect and track moving objects across different frames.

This new image is stored in SRAM and we use it to search and isolate region of interests (ROI) in a 128×64 sub-image. To obtain the vector of feature for the following classification step, we calculate the average values of gray for each column and row in ROI (which is equivalent to project the ROI image onto horizontal and vertical axes). Thus the size of the input vector for the classifier is reduced from 8192 to 192 elements. Undoubtedly both smart ROI size and efficient feature extraction algorithm contribute significantly to save energy and time processing.

Regarding the classification function, a highly tuned SVM-like hardware oriented algorithm has been implemented for the STR91xF [85]. A detailed de-

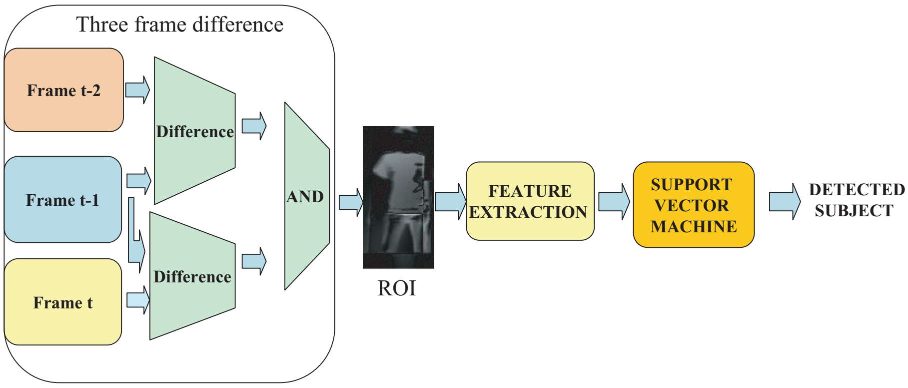


Figure 4.10: Flow chart of the human detection application.

Task	Energy $[mJ]$	time $[ms]$
Three Frame Difference	440	720
ROI Extraction	12,2	20
Feature Extraction	9,6	16
SVM	21,21	35

Table 4.2: Energy requirement

scription of this algorithm and its performance in people recognition can be found in [84]. Being a ”learning from examples” technique, SVM [148, 133] it is firstly trained on a set of available data known as *training set*. Such a computationally expensive training phase is performed off-line by a powerful base station, then the classification function are loaded to the nodes to classify the patterns under observation.

Thanks to background subtraction the training set is independent from the node position and orientation, thus all SVM can be trained at once using the same training set.

The output of the classification can be simply binary report of the presence of the human body in the field of view, or an image of the region of interest with the detected subject. This result can be sent via wireless to a controller unit.

4.2.9 Autonomy of the system

We considered a typical application scenario of an outdoor surveillance. Assuming a rate of events as presented at the beginning of this section we estimated the capacity necessary to perform a complete and effective service during the night using the energy harvested and saved during the day. Experimental results using different size supercapacitors without solar harvesting capabilities, show that the system can achieve autonomy of several hours (figure 4.11). Increasing the capacity up to $500\,F$ it is possible to operate for about

8 hours, till the next morning.

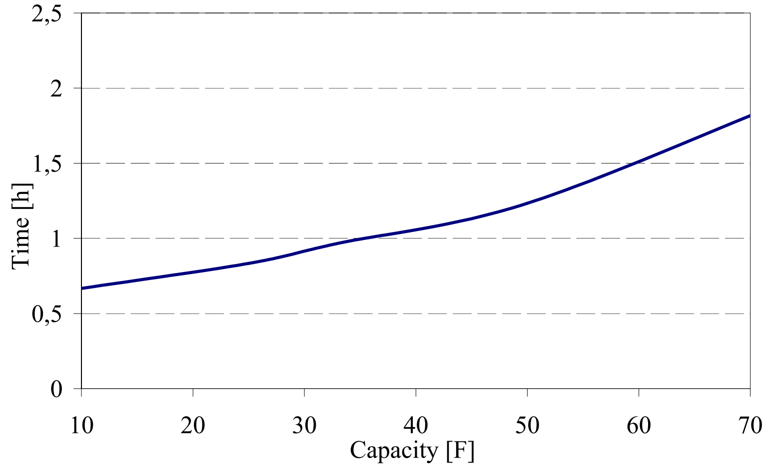


Figure 4.11: Autonomy of the system varying the capacity of the reservoirs without environmental harvested energy

4.2.10 Dynamic adjustment of the detection area

In a distributed vision network several nodes cooperate for an efficient surveillance service and the area under monitoring is covered by multiple nodes deployed in the environment and the whose projections of camera field of views are usually overlapped. For this reason it is possible to develop distributed policies for smart dynamic coverage of the region under surveillance. For instance when a node is lacking of energy it could reduce its detection area and consequently its activity while other cooperative nodes compensate augmenting PIR sensitivity for longer distance events. In such a cooperative vision, a dynamic adjustment of the detection area on each single video is necessary.

Figure 4.12 shows the amplitude of the PIR signal as a function of the distance of the detected object. This result highlights how is possible to modulate the detection area by adjusting the thresholds used to generate a wake-up signal for the video node.

If we assume a uniform probability that a person moves in a certain point of the area of interest, by increasing the threshold we reduce the sensitivity of the trigger and the area covered by the PIR and consequently the probability to activate the camera.

For this reason the threshold (4.3) is regulated as a function of the following parameters:

- contrast of the image, C;
- the energy available in the supercapacitor, E_{CAP};

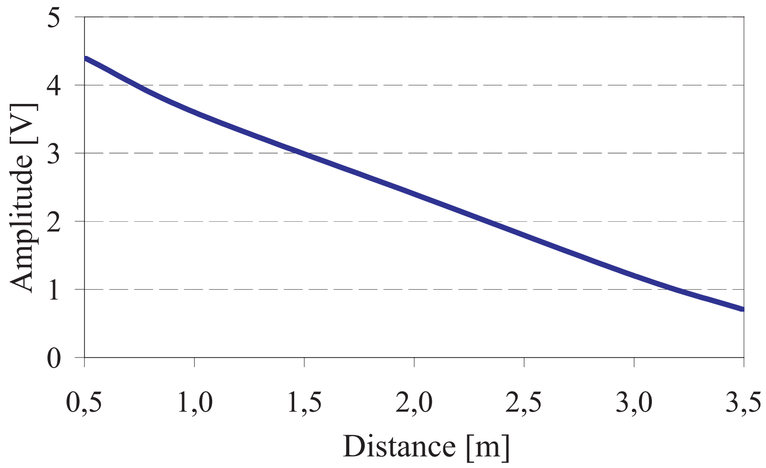


Figure 4.12: Amplitude of the PIR Output signal as function of the distance of the object.

- the probability of seeing a person moving in a certain point at a certain time, p.

$$V_{threshold} = \alpha \frac{p}{E_{CAP}} + \beta C \tag{4.3}$$

Images with low contrast C may result in a loss of accuracy of the SVM algorithm. Thus, it is better to suspend the vision algorithm saving energy when the contrast of the image is lower than a defined value $C < C_{th}$. Concurrently, when the contrast of the images is low, the threshold of the PIR could be reduced in order to extend the area under monitoring and sending alarms relying only on PIR detection. The value of the threshold should be inversely proportional to the energy available in the supercapacitor and directly proportional to the probability density of a people moving in the field of view. In fact when more energy is available a higher number of detection can be tolerated. On the other hand, if the probability of detecting a person is higher, lack of energy in the accumulator forces a higher reduction of the field of view of the PIR if we want to extend the lifetime.

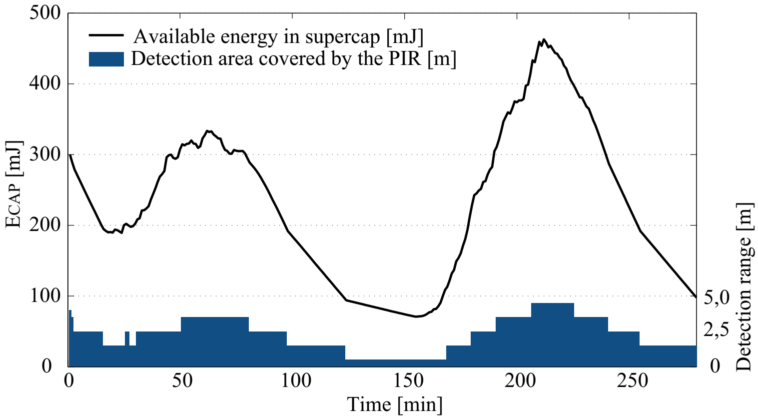
A simulation to verify the performance of the proposed dynamic threshold is depicted in figure 4.13(a). The energy harnessed from the solar cell is powering the sensor node and replenishing the energy storage E_{CAP} with the exceeding energy. When the energy in the storage is enough to sustain the desired quality of service, the detection area covered by PIR sensor increases (up to $4\,m$ in our scenario). Similarly, as soon as the available energy decreases due to a reduction of the harvesting supplying, the threshold switches diminishing the area covered by PIR and consequently the rate of activation of the camera. The simulation covers about five hours of operation of the sensor node, and the threshold function is approximated using discrete values. It worth to notice that simulations are performed using energy storage devices with limited capacitance of $33\,F$ and a constant contrast C of the images higher than the threshold C_{th}. To prove the effectiveness of the dynamic adjustment of the monitored area, figure 4.13(b) illustrates the behavior of the node with different configurations. The plot compares the energy stored in the supercapacitor in the same operating condition of figure 4.13(a) with the situation when the threshold of PIR sensor is fixed with a constant size of the area under monitoring of $3\,m$ (dashed plot). Using a fixed threshold the trade-off between energy and sensitivity is off-line design parameter and wide detection areas increase the probability to be out of service because of the empty energy accumulator, as happens in the figure during the interval $I_{OFF}\,[111, 168]$. The plot shows also the performance of the video node without solar harvester and when no environmental energy is stored in the accumulator. Obviously in this case the video node has a limited lifetime as for all battery-operated systems.

4.2.11 Conclusion

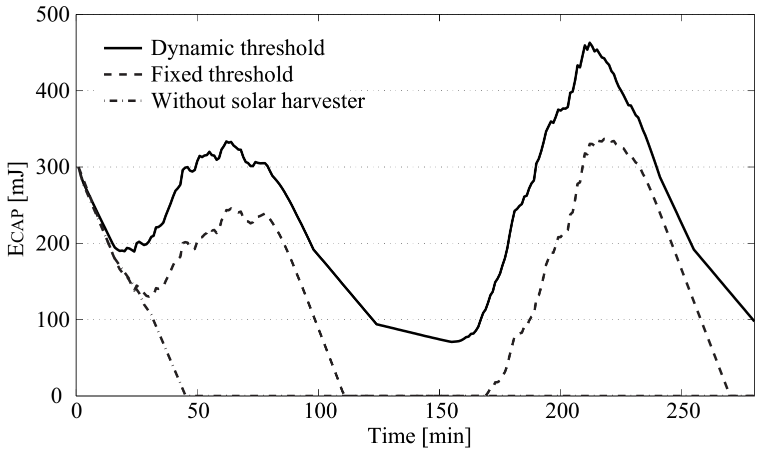
An integrated self-powered video sensor node for energy efficient surveillance has been proposed. The adoption of a solar harvester for supplying the node leads to several benefits such as the possibility to extend the lifetime of the vision sensor network. However since the amount of energy provided by the photovoltaic module cannot be predicted the status of the system must be dynamically adjusted. A multimodal platform equipped with different family of vision sensor with heterogeneous features of power consumption and resolution permits to adopt very effective energy management techniques reducing considerably the activation of the camera, the microprocessor and other power consuming devices. In the proposed system the sensitivity of a low power PIR based wake-up circuit is adjusted dynamically according to the available energy on-board, to the contrast and the probability of moving subjects enter the video node field of view. With such a technique, under a hypothetical surveillance scenario, we estimated that using a 500F super capacitor the wireless video node is able to operate for about 8 hours during nighttime.

4.3 Adaptive Power Control for Solar Harvesting Multimodal Wireless Smart Camera

The interest on distributed, smart and reliable surveillance systems based on Wireless Sensor Networks (WSN) has recently gained momentum.Mass tran-



(a) Variation of the area under monitoring as function of the stored energy.



(b) Comparison of the energy efficiency in different solution: with dynamic variation of the PIR sensitivity threshold, with fixed threshold and without solar harvester.

Figure 4.13: Simulation results of the energy efficiency using a dynamic PIR sensitivity threshold.

sit, public areas, sport and event venues are the places where flexible and low-cost smart cameras could be the breakthrough for the next decade and could gain large part of the security and surveillance marketplace, provided that they have the capability of operating immediately after the deployment without running out of energy for years. The importance of deploying cameras in unobtrusive locations, forces the installation in areas which are hard-to-wire or where there is no pre-established infrastructure. Therefore, the autonomy of the system becomes one of the primary design constraints.

For this reason, in contrast to cameras that just *watch* the world or *understand* what happens around by performing some simple algorithms locally, we aim at developing intelligent devices that are capable of taking care of themselves and that perform actions autonomously which serve to extend the system lifetime. At the same time, the cooperating devices should guarantee adequate accuracy and quality of service of the surveillance application.

Developing energy efficient wireless video nodes and aggressive power management policies are essential for achieving a long-term operation of a distributed system of standalone cameras. Recently, video nodes have been also enhanced with energy harvesting circuits, able to receive and convert energy from regenerative sources in the environment using transducers such as photovoltaic modules [101, 58]. A smooth optimum control of the power consumption is a striking concept for reducing the size of the solar cell, for coping with the unpredictable profile of the energy intake and developing smart cameras which can be deployed unattended for years.

We present an optimal feedback controller for power management and performance optimization tailored for multi-modal video surveillance applications. The objective is to achieve continuous operation, dynamically tuning operation modes according to the status of the system and the amount of energy in the storage, while maximizing the monitoring performance. The controller adapts the size of the detection area, to achieve a the maximum autonomy and lifetime of the system and to guarantee, at the same time, an adequate accuracy of the detection application fixed by the end-user. So the main contribution of our work is an approach for dynamic control, which allows tradeoffs between energy-efficiency and system performance by adjusting the sensitivity of the system.

The application used as case study is a human and body image recognition and classification. Computation intensive tasks for video elaboration are triggered by Pyroelectric Infrared Sensors (PIR) and employ state-of-the-art Support Vector Machine (SVM) technology, highly tuned for low power consumption [86]. The combination of several vision devices with heterogeneous features allows the development of multimodal surveillance applications with efficient energy policies. In fact, video would still provide high-level information when required, and PIR sensors would assure a continuous monitoring service triggering the CMOS camera when an important event is detected.

Our dynamic management problem has been formulated as a discrete-time optimal control problem and has been solved using the theory and computational tools developed in the field of model-predictive control (MPC) [18, 96]. The optimization process is by taking into account the power consumption of the node during the execution of the application, the amount of the energy collected by the solar cell, the predicted amount of energy intake estimated in the near future, the accuracy requirements and the size of the area under monitoring.

The remainder of the section is organized as follows. Related work is reviewed in the next section. Section 4.3.2 describes the current implementation of our video system for wireless sensor networks and illustrates the design of the Model Predictive Controller. Section 4.3.5 shows the performance of a system with the proposed controller, highlighting the difference with previous heuristic approaches. Finally Section 4.3.10 concludes the section.

4.3.1 Related Work

Research on low-cost video nodes constrained by low-power budgets has been very active in the last years and a number of node prototypes have been presented [47, 90, 166, 63, 66, 68] confirming that video wireless networks represent a strategic enabling technology for a number of applications in surveillance, environmental monitoring, entertainment and health care.

Low-cost video systems with wired interfaces [37] represent the first generation of stand-alone nodes. To satisfy the basic requirement of being wireless, prototypes such as the Panoptes-class [63] consumes roughly ten times more power than typical nodes in a wireless sensor networks. Rather than sending the raw sensor data through the network for processing, a recent approach [35] focused on mote platform to perform computer vision problems through in-network processing of sensory data. This allows the node to processes the data and sends only key data elements through the network to a central server, saving the energy for large image transmission. In this work we push towards the power reduction adapting the behavior of the node to the available energy.

Distributed networks of motes equipped with video sensors have been proposed in [73] to guarantee stealthiness in military surveillance applications. A network of IR sensors and cameras is presented also in [126] to balance privacy and security in surveillance applications. One important issue in current video wireless nodes is that the digital signal processing subsystem is the main consumer of energy. This is due to the fact that the high data rate of CMOS image sensor imposes the selection of fast processors and memories with high power consumption. Hence, the main open challenge in this area is to synergetically develop algorithms and architectures for energy-efficient image processing without giving up the flexibility of in-field configuration within tight power budgets typical for WSNs.

Power management is a critical issue when dealing with surveillance systems which could be long-term operating and unobtrusive. The size of batteries often exceeds that of nodes themselves. Thus accurate evaluation of the trade-off between power conservation and quality of surveillance in target tracking sensor networks has been presented in [71]. A first step towards power consumption reduction has been introduced by multi-modal systems with the combination of vision devices with heterogeneous features. Video

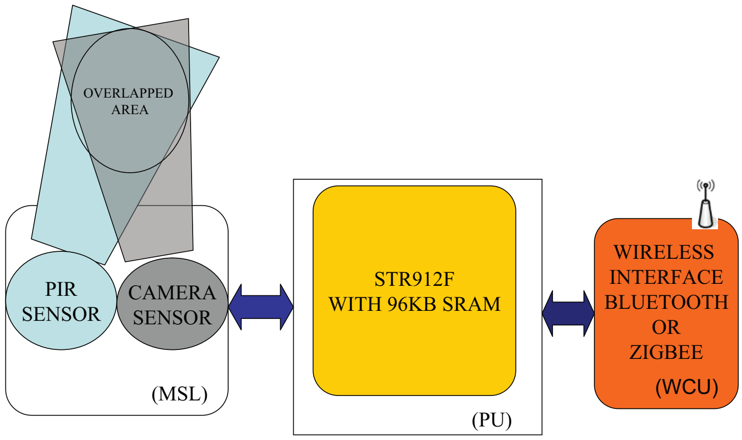


Figure 4.14: Video sensor node architecture.

sensors still provide high-level information when required, while low-cost and low-power sensors such as PIR sensors assure a continuous monitoring service. The design objective of multi-modal surveillance systems is to overcome some points of failure of a particular kind of sensor as well as to balance various parameters such as power consumption vs. surveillance quality. For instance, PIR sensor capabilities of detecting both presence and direction of movement have been exploited in [43] to enhance a video surveillance system, while [31] presents a camera for remote surveillance which is equipped with a PIR sensor. The PIR provides triggers for a light during night time that illuminates the scene in presence of moving animals. Finally, another attempt to minimize power consumption while guaranteeing accuracy is presented in [19] where network lifetime is extended by capitalizing on low power states of multi-modal video nodes.

4.3.2 System Architecture

We optimized the performance of the system using a vision application which performs human and people detection in camera snapshots and a video node similar to [101]. The sensor node is based on a wireless smart camera for sensor networks which is equipped with an ARM9 core.

Figure 4.14 shows the hardware architecture of the smart camera. It is a multilayer system with reconfigurable features. The multi-sensor layer is equipped with a SXGA CMOS color digital camera and a PIR sensor, the processing unit is based on a STR912F ARM9 microprocessor from STMicroelectronics, operating at 96Mhz and with 96KB SRAM on die, while wireless communication has been implemented supporting both Zigbee and Bluetooth pro-

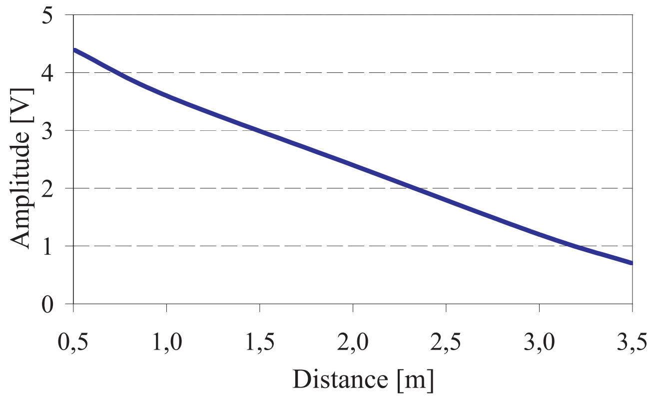


Figure 4.15: Amplitude of the PIR Output signal as function of the distance of the object.

tocols. The system is equipped with an energy harvester unit to provide power supply. The regenerative source is the solar energy collected by a small photovoltaic module. The scavenging circuit adjusts dynamically the operating point of the photovoltaic panel (V_{panel}, I_{panel}) to obtain a fixed output power under steady environmental conditions (e.g. light irradiance, temperature). If the collected power is the maximum achievable, such a technique is called Maximum Power Point Tracking (MPPT).

Our power management approach is general, as it could be adopted to systems where the available regenerative energy from the environment can be predicted to some extent. To guarantee long lifetime we dynamically adjust the threshold level which forces the pyroelectric sensor to trigger an event and to wake-up the camera. In a monitored area several video nodes cooperate and a dynamic reduction of the detection area of a video node can be compensated by others, if necessary.

Figure 4.15 shows the amplitude of the PIR signal as a function of the distance of the detected object. This result highlights how it is possible to modulate the detection area by adjusting the thresholds used to generate a wake-up signal for the video node. If we assume a uniform probability that a person moves in a certain point of the area of interest, y increasing the threshold, we reduce the sensitivity of the trigger and as a result, we also reduce the area covered by the PIR device and the probability to activate the camera, as illustrated in Figure 4.16. The PIR threshold is directly regulated by the Model Predictive Controller which determines a viable trade-off between the quality of service of the monitoring application in terms of area covered by the people detection algorithm, and the long-term autonomy of the surveillance system.

Figure 4.17 shows the model of the proposed approach. High level Model Predictive Control (MPC) [18] is exploited to adjust the sensitivity of the pyroelectric infrared sensor. MPC aims at improving the performance of the system using predicted values of input or output variables under specified restrictions

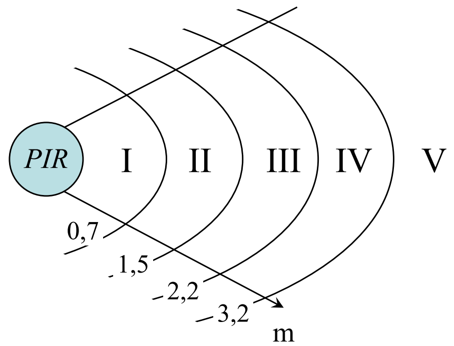


Figure 4.16: Zones detected by varying PIR triggering threshold.

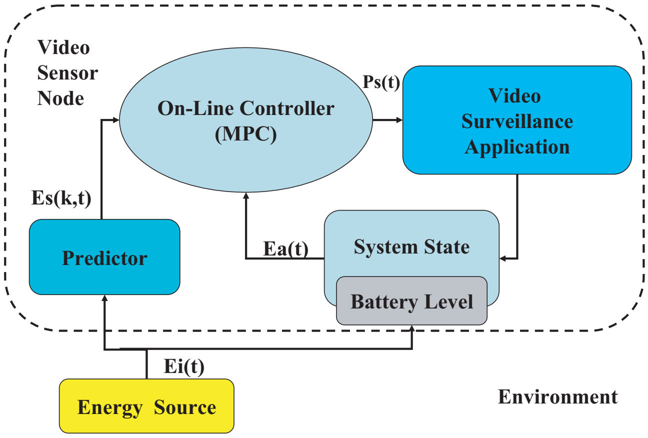


Figure 4.17: System model for PIR threshold control

of some features. We assume that the harvester provides to the video node an amount energy $E_i(t)$ within the unit time interval starting at time t. This energy is stored in a storage device, e.g. a battery. In the same interval the system can use energy from the battery. The available energy at time interval t is denoted as $E_a(t)$. The energy intake form the harvester $E_i(t)$ is also used by the predictor module for delivering estimations $E_s(k,t)$ of the future expected energy according to the length of the selected horizon. The energy consumption of video processing in the unit time interval starting at t depends on the sensitivity threshold of PIR trigger $P_s(t)$ in the same interval. In this condition, the controller dynamically adjusts the PIR sensitivity at regular time intervals. If the controller increases the sensitivity of PIR, the events detected will be more frequent and the video application will detect and classify more persons, but consequently the power consumption will increase.

4.3.3 Model Predictive Controller Design

Model Predictive Control [18, 96] is an advanced control technique used extensively in industrial process control applications, which aims at achieving defined system performance under specified restrictions on input and output variables. Its major advantage is that it can deal with *multi-input-multi-output* control problems where the system performance depends on the correlation among several parameters. The basic idea of such an approach is to optimize an appropriate objective function defined over a time interval in the future. A model of the system is used to predict the behavior over N prediction intervals where each one has a length of L unit intervals. The total length $N * L$ is called the prediction horizon, as depicted in Figure 4.18. The solution of the optimization problem is computed by selecting an input trajectory which includes the control inputs in the following N prediction intervals and which maximizes the objective function while satisfying the constraints. Once the solution is computed over the whole control horizon, only the first feedback control action which is related to the first prediction interval, is applied to the system. Then the solution is computed again at the beginning of the next prediction interval. In this way, Model Predictive Control provides performance prediction, optimization, constraint satisfaction, and feedback control within a single algorithm.

A model of the system is used to predict the behavior over N prediction intervals and all of them is called the prediction horizon and the solution of the optimization problem is computed by selecting an input trajectory which includes the control inputs in the following N predicted interval periods and which maximize the objective function while satisfying the constraints, Figure 4.18. Once the solution is computed over the whole control horizon, only the first feedback control action is applied to the system, and the solution is computed again at the end of sampling period. In this way, Model Predictive Control provides performance prediction, optimization, constraint satisfaction, and feedback control into a single algorithm.

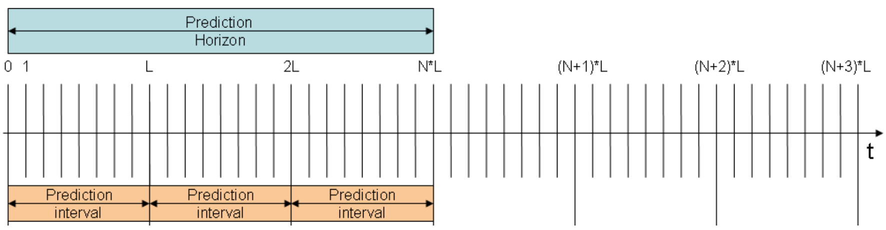
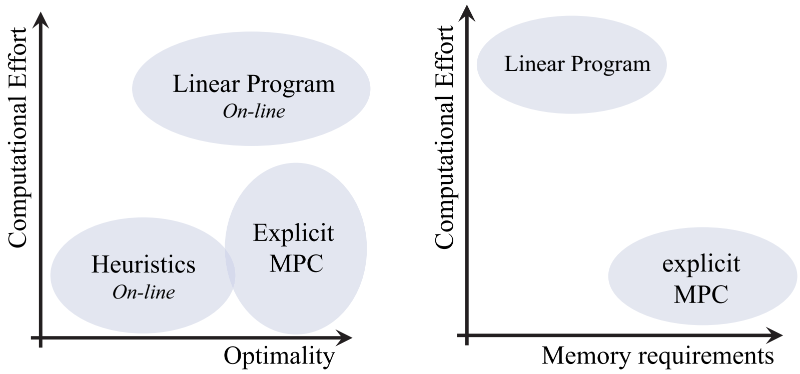


Figure 4.18: Predicted horizon for MPC when $t = 0$.

The algorithm has the same complexity as solving a Linear Program (LP). It can be implemented as an implicit solver which requires LP solutions under



(a) Computational Effort vs Optimality.

(b) Computational Effort vs Memory Requirements.

Figure 4.19: Trade-offs using different algorithms.

real-time constraints. This causes an increment of the computational effort as depicted in Figure 4.19(a). Another solution is to implement the Model Predictive Controller as an explicit approach. In this way we pre-compute off-line a lookup table of linear control actions [18]. As a result, the computation of the optimization problem is translated into a linear combination of input parameters according to gain and offset coefficients that depend on input parameters. In this way, as illustrated in Figure 4.19(b), we shift the effort of solving a LP on-line to an increase in memory requirements, due to the storage of pre-solved control laws that are computed off-line. Moreover, with a certain order of approximation, most of the control laws can be clustered reducing the memory occupancy [110].

In our approach, the predictor uses tuples $(t, E_i(t))$for all times $t \geq 1$ and delivers N predictions, i.e. for the energy production of the energy source within one of the next N prediction intervals. Following well known prediction equations based on Exponentially Weighted Moving Average (EWMA) [39, 110, 81], the predictor produces estimations $E_s(k, t)$ where $1 \leq k \leq N$ denotes the prediction interval, see Figure 4.18.

The problem of adjusting the PIR sensitivity has been formulated as linear program (LP), and the performance objective is to maximize the monitoring area and thus, to guarantee higher QoS by maximizing the maximum number of processed events (e.g. the higher number of detected people which cross the monitored area). For this objective, a previous work [101] proposes an heuristic algorithm where the size of the detection area depends linearly on the energy stored on the on-board reservoirs (e.g. battery or supercapacitor), which consequently attempts to decrease the monitored area, by a reduction of the PIR sensitivity, when the scavenged energy is low (e.g. at night) and increase the area when scavenged energy is high (e.g. during the day). In contrast to the work shown in last section, we propose to compute the optimal solution using MPC. The LP presented in this section models a large variety of application scenarios, constraints and optimization objectives.

4.3.4 Linear Program Specification

In this section we will show the linear problem specification. Before formulating the problem we introduce the following equation:

$$E_a(t + k \cdot L) = E_a(t) - k \cdot L \cdot E_p + \sum_{j=0}^{k-1} E_s(t, j) + -L \cdot E_v \cdot P_s(t + j \cdot L)$$

It shows the expected content of the energy storage at times $t + kL$ for $1 \leq k \leq N$. The E_p is the energy consumption, independent from time, of our system when it is in sleep mode and only the PIR sensor is used by the system. E_v is the power consumption of image processing when an event happens and it is a constant value. Figure 4.18 shows the meaning of k, N, and L. Finally we can write the linear program which optimize the behavior of the multi-modal video node:

maximize λ subject to:

$$P_s(t + k \cdot L) \geq \lambda \qquad \forall 0 \leq k < N$$

$$E_a(t + k \cdot L) = E_a(t) - k \cdot L \cdot E_p +$$

$$+ \textstyle\sum_{j=0}^{k-1} \left(E_s(t, j) - L \cdot P_s(t + j \cdot L) \cdot E_v\right) \geq 0 \qquad \forall 1 \leq k \leq N$$

$$E_a(t + N \cdot L) \geq E_a(t) - 100$$

The first inequality states that the threshold of the PIR should be regulated to maximize the monitored area. The second inequality gives the energy balance of the system, taking into account the power consumption when the video algorithm is activated and the energy intake form the solar harvester. Finally the last inequality is used to guarantee a stable behavior of the system, constraining the controller not to plan the exploitation of all the energy before the end of the prediction interval.

Component	Power $[mW]$
ARM9 (RUN mode)	450
ARM9 (SLEEP mode)	15
Video sensor (ON mode)	165
Video sensor (SLEEP mode)	0
TX/RX module (ACTIVE mode)	98
TX/RX module (SLEEP mode)	0
PIR sensor	1,5
Solar Harvester	0,98
Video Node (Active)	650
Video Node (Sleep)	17,5

Table 4.3: Power consumption of the video sensor node.

4.3.5 Experimental Results

We have compared our adaptive management approach with three other heuristic solutions. The first is proposed in [101] where a simple controller adjusts the sensitivity of the PIR sensor according only to the amount of energy in the battery. The others are extensions of the same controller where the information form the predictor and future state of system are added as available knowledge of the controller. In this way we can identify the contribution of the MPC and how it outperforms controllers with the same input variables.

4.3.6 People detection application

The vision algorithm implemented on the video sensor node is a human detection application based on a SVM-like classification for embedded systems [85]. It forces the system to the sleep mode until the PIR sensor detects an event. Then the camera is activated to acquire a frame and the microprocessor begins the analysis of the captured image. The video processing lasts for about $3s$. The power consumption of all power modes and components is given in Table 4.3.

We considered a typical application scenario of outdoor surveillance. In our simulations we assume to take a rate of events which represent people passing in the field of view of the video node. We considered the main entrance of our School and we collected the profile of the events during 10 consecutive days and the number of people who entered the building. In the same way, we measured the energy intake from the energy harvesters and the solar light intensity during the same period. All the information was stored in files used as input to our simulations.

Notice that the controller does not know in advance the number of persons which enter the monitored area, generating an event. Clearly the system have to count this contribution and the simulation will use the following equation for the state of the system:

$$E_a(t) = E_a(t-1) - E_p + E_i(t) - E_v \cdot P_s(t) \cdot N_p(t) \tag{4.4}$$

where $N_p(t)$ is the number of events for unit of sensitivity depicted in Figure 4.20. In spite the MPC gives only continuous values of P_s the system discretizes the optimal solution to have just five values even if this solution will be sub-optimal. We assume that the controller gives just five values of P_s (1..5) and the number of detected people is linear with PIR sensitivity.

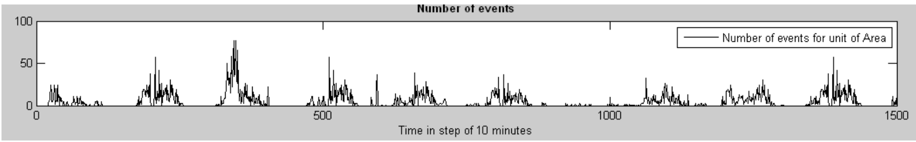


Figure 4.20: Profile of the people entered in the monitored area and detected by PIR sensor in $10 days$.

4.3.7 Adaptive controller vs. heuristic algorithm with and without energy prediction

Under the above mentioned assumption, the first comparison shows a system without an adaptive controller. To dynamically adjust the PIR sensitivity we used only the energy stored in the battery. So when the stored energy is at the maximum value the system can set the PIR sensitivity to highest value to detect all the events in its field of view (up to 5 m in our scenario) and guarantee the best performance. Instead if the energy is decreasing, due to a reduction of the harvesting processing, the area covered by video node decreases. Consequently the rate of activation of the camera is shortened because the PIR sensitivity is lower. For our simulations we assume to have only five areas like Figure 4.16 shows. Moreover we assume a linear distribution of events in the area. Under this assumption if the PIR covers the III area, it detects 3 times the value of events of I. The battery level was divided in five portions and each portion has an associated level of PIR sensitivity. The lowest level of sensitivity corresponds to the lowest battery level, and vice versa.

Figure 4.21 shows how the adaptive controller is able to maintain PIR sensitivity always at a higher level, in contrast to the heuristic controller which has to follow the energy level. By exploiting the prediction of incoming energy, the adaptive controller maximizes the monitored area. Note that PIR sensitivity of the adaptive controller is higher than the heuristic approach even if the available energy is lower, in fact when it knows that the system will have enough

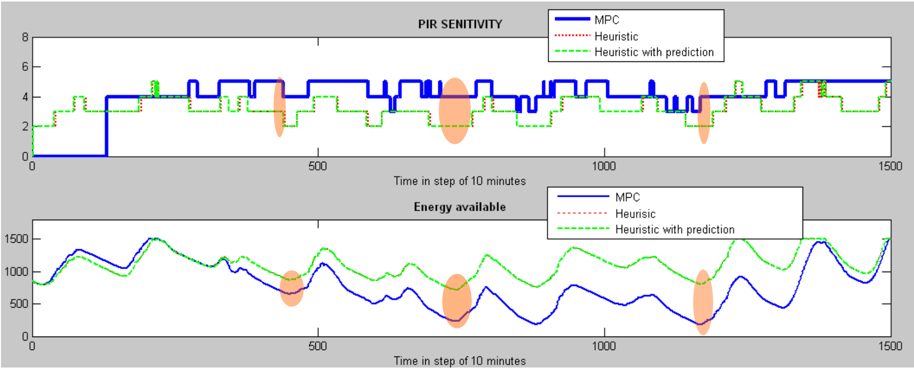


Figure 4.21: Comparison between adaptive controller and dynamic heuristic controller with and without prediction. The elliptical areas show that the MPC controller keeps a higher monitored area also when a battery low situation happened.

energy in future, the controller speculates on the detection of more events. The figure shows also an optimization of heuristic dynamic controller. In this situation the system knows the prediction of the energy intake. The effect of this second version is that the sensitivity of PIR is more reactive to the future values of energy. The Figure 4.22 shows how the heuristic controller with prediction is faster than the first one, but still worse than a controller based on MPC, even though we used the exact value of future energy at the time $t+1$ as estimations.

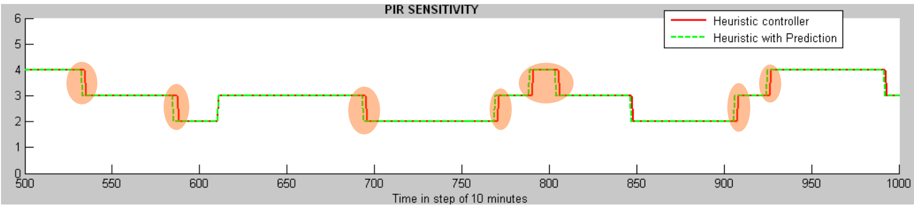


Figure 4.22: Comparison of heuristic controllers. The elliptical areas show how the controller with prediction is faster to increase or decrease the monitored area taking advantage from prediction.

4.3.8 Adaptive controller vs. advanced heuristic

The last simulation uses both the prediction and a new way to elaborate the maximal area to cover as equation (3) shows. This controller finds the maximal value of P_s(t) that satisfies equation (3), hence at least $X = 20$ detection for unit of PIR sensitivity in the next period of time. $E_s(t)$ is the exact value of future energy at time t and E_m is a fixed value of battery to guarantee the survival of the node, in our model E_m is the 10% of the maximal level of battery. Figure 4.23 shows the comparison of our adaptive model with this approach. The main feature of the advanced heuristic controller is the capability to increase the monitored area when the energy intake is plentiful, but as soon as the battery is discharging the controller decreases remarkably more than other approaches.

$$E_a(t) = E_a(t-1) - E_p + E_s(t) - E_v \cdot P_s(t) \cdot X \geq Em \tag{4.5}$$

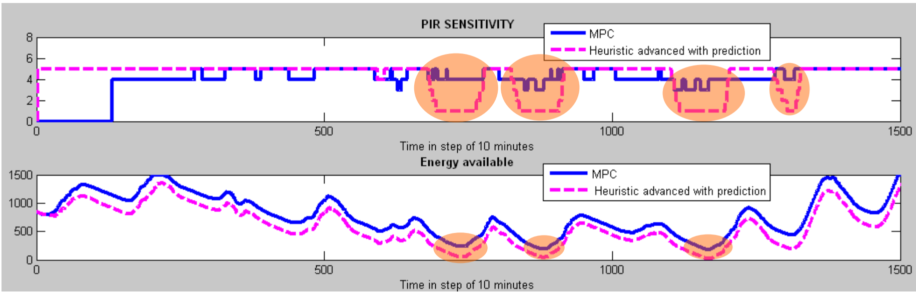


Figure 4.23: Comparison between MPC based controller and heuristic advanced controller. In this comparison in the elliptical areas you can see how the heuristic controller brings the system in a critical battery low situation and consequently to decrease the monitored area

4.3.9 Comparison

The simulations have demonstrated that an intelligent control with MPC ensures that the minimum area under monitoring is always maximized compared to other algorithms. Heuristic algorithms which depends only on the knowledge of the energy are greedy and quickly reach lower area when the available energy is low.

The controller based on MPC also increasing the area under surveillance on average. Comparing the rate of the coverage, measured as the integral of the covered area during the time, MPC keep the area larger than other algorithms as shown in Figure 4.24.

4.3.10 Conclusion

A simple but optimal feedback controller for power management based on model predictive controller (MPC) has been presented for achieving performance maximization under defined system constraints. The controller adapts parameters of the application, such as the size of the detection area adjusting the sensitivity of a low power PIR based wake-up circuit, in order to guarantee the maximum lifetime of the system while keeping high accuracy of the surveillance application. Simulation results and measurements on the video sensor node demonstrate that our approach outperforms naive power management policies, while improving performance.

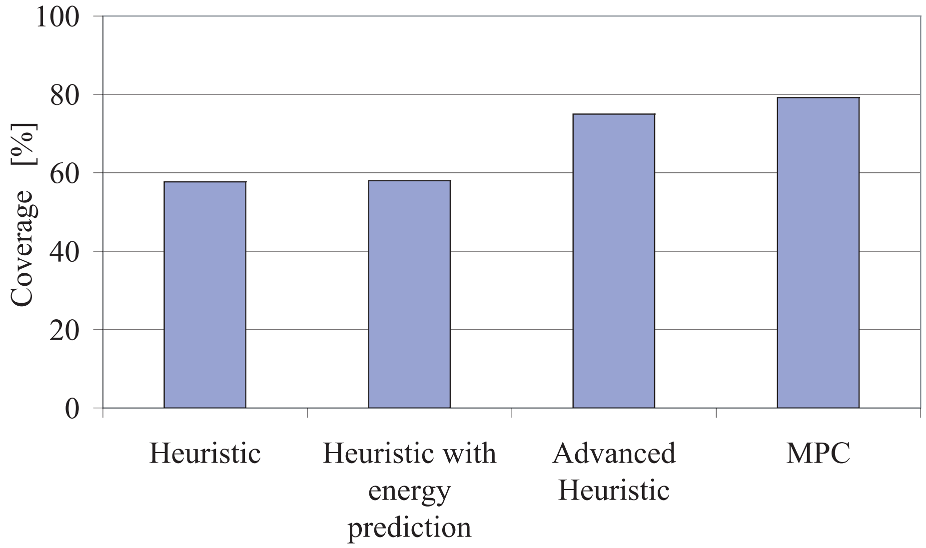


Figure 4.24: Comparison of the global coverage during the simulation.

4.4 Energy Efficient Cooperative Multimodal Ambient Monitoring

Recent years are witnessing the ever-growing demand for security in both public and private spaces. This feeling has pushed for the development of video surveillance and other security-related applications. As new generation of low-power, low-cost devices hits the market it is likely that new scenarios were a large number of cameras are embedded in the environment will emerge.

Thanks to their flexibility and ability to provide accurate real-time visual data *Wireless Video Sensor Networks* (WVN) are gaining many credits. A WVN is made up of several wireless *Video Sensor Nodes* (VSNs) and each of them embeds a low-power imaging sensors, processors, and communication units to survey the Area of Interest (AoI). Power-aware design and maximization of the sensor network lifetime becomes one of the main objective [50, 69].

Typical approaches for energy consumption reduction in Wireless Sensor Networks (WSN) include: selection of low-power components [74], use of improved wireless protocols [138] and adapting parameters such as clock rate [89] or sample rate [79].

Exploiting renewable energy resources in the devices's surrounding is an alternative solution to increase nodes lifetime [116]. In particular, photovoltaic (PV) harvesters are good candidates to achieve perpetual operation of WSN [30]. Unfortunately if the power consumption of a device can be estimated runtime for a certain interval in the future, the power generated by a PV module changes nonlinearly under varying temperature or solar irradiance. Techniques which automatically tune the operating point of the solar cell should be considered to provide the maximum output power, since they lead to several benefits such as: the possibility to use smaller PV modules, to reduce the capacity of the energy reservoir, or to allow higher power consumption operations onto a sensor node.

The major constraint when dealing with circuits for high efficient energy harvesting is that implementing maximum power point tracking techniques using small-size *PV* modules is practicable only if the power consumed by the additional hardware is considerably lower than the amount of output power that it gains. Thus the area of the deployment and the availability of the environmental energy need particular attention at design time when an estimation of the energy intake during the day or along a year is fundamental.

These approaches try to address the power consumption issue by extending single nodes lifetime. However, WVNs prompt for the development of high level *Power Management* (PM) policies. For example in [162] the author proposes two scalable and flexible techniques for WVNs power management by considering the content of the video data sensed both locally and by other video nodes within the network.

If we consider the energy issue at a network level we can exploit the use of heterogeneous network. Surveillance, as well as target tracking and classification, are classical applications which require global information of a certain spatial-temporal region and exploiting Multi modal-sensors is a promising approach to increase effectiveness of such systems [46]. On the other hand, generally sensor nodes only has a local view and spreading global information increases the communication traffic and the overall power consumption.

In addition, the redundancy provided by a mesh of heterogeneous nodes can be used to perform power-performances trade-off. The typical approach here is to support high-power CMOS imagers with a mesh of low-power, low-cost sensors densely spread in the environment [72]. While the former are kept into a low power state, the latter operate as a trigger to provide continuous area monitoring. In Boettcher et al. [27] low-power acoustic sensors are used to detect position of moving vehicle through a time-difference of arrival technique. This information is used in conjunction to an imager used to take an image of the vehicle and send it to the base station. Another example can be found in the work of Wang et al. [152] where a WVN is supported by a network of microphones. The latter are used to provide an indication of the distance of a vehicle from the video sensor node. This information, together with the recognition accuracy of the video sensor node estimated at training time and the actual energy of the video sensor node, is used to evaluate a cost function used by a cluster head to select which video sensor node should be turned on.

In this section we present a combination of PM techniques optimized for multi modal surveillance systems:

- a set of wireless VSN are used in conjunction with a network of low-cost, low-power Pyroelectric InfraRed (PIR) sensors to detect presence of people moving along a path in an outside area;
- the architecture is scalable and can be extended to an ambient of any size and shape. The minimum cluster of node is composed by 2 VSNs used in conjunction with a network of low-cost, low-power Pyroelectric InfraRed (PIR) sensors to detect presence of people moving along a path in an outside area;
- particular surveillance applications are interested in distinguishing the presence of human bodies rather than objects or animals in the field of view before generating alarms or sending information through the wireless link. Furthermore we implement algorithms capable to process images, to detect the particular target and finally to send only the image that shows the face of the person. It is a particular needs for unobtrusive video-surveillance solutions which has to handle both security and privacy issues, guaranteeing to not process or record private data, while still detecting and identifying potential threats;
- the video sensor nodes are powered by a solar scavenger using a 70 cm^2 photovoltaic panel, to guarantee the maximum energy autonomy of the systems and flexibility for the reuse of the system or the adjustment of the deployment.
- to guarantee a balanced energy usage a trade-off between energy availability and quality of the service is adopted and VSNs are activated only when they can provide a useful contribution at the minimum energy expenditure. Thus a bidding-like protocol is engaged to select the most suitable offer to perform the image analysis.

The PM techniques is performed in a distribute manner, since obviously nodes constantly active makes unfeasible to meet the energy requirements. We use the information from the PIR sensor network to activate only a subset of wireless VSNs enhancing an efficient collaborative approach. In fact, the pir-network is capable to detect and to estimate the position and direction of movement of the people along the track and therefore it identifies which VSN faces the persons. When the PIR sensor network detects a body moving in the area of interests it broadcasts a message to the VSNs with an indication of the body presence and direction of movement. According to this information and the available energy provided by the harvesting system, each VSN calculates a cost function and broadcast this value to the other VSNs. Each node compares its own cost function with the ones received from the other nodes of the network. The node with higher cost function wakes up and monitor the area of interest. This PM policies guarantees that for any passage at least one node wakes up even in presence of message losses. At the same time we can keep the majority of VSNs into a low power state to preserve or replenish their energy storages.

Our approach is similar to the one presented in [152]. However our distributed policy is more robust to nodes failure and messages loss. In fact in the work of Wang et al. if the cluster head fails the whole cluster is not able to operate until maintenance is performed. Furthermore, if node communication is compromised at a certain time, the nodes of the network can not be waken up to classify object passing by. In our work, instead, all nodes check locally if they are the one that should be turned on, therefore even if communication among nodes is not possible, in the worse case, all VSNs will wake up and analyze the image.

The rest of the section is organized as follow. Sections 4.4.1 and 3.3.2 presents the WVN and the PIR sensor network that compose our system. Section 4.4.3 explains how the two sensor networks are used in conjucntio and the distributed power management techniques. Section 4.4.5 describes our network simulation and compares our approach with the case presented in [101]. Finally Section 4.4.7 concludes the section providing further comments on the results we achieved and comparison with the state of the art.

4.4.1 Video Sensor Nodes Description

The hardware architecture of the solar-powered video sensor is composed of four main modules (see figure 4.14) and designed to achieve low power consumption of the overall system. Each module can autonomously operate in different states to save energy when its contribution is not needed, as you can see in previous section.

As in the other works presented in this section (see sections 3.3.2, 4.2.5) we used a commercial PIR detector that includes 2 sensitive elements placed in series with opposite polarization. To form a PIR sensor node, this board is connected to a Zigbee module that provides wireless connectivity with the other nodes of the network.

An overview of the node power consumption in different operating states is reported in Table 4.4.

Component	Power $[mW]$
Active Radio TX	48
Active Radio RX	37
Active Radio Off	13
Sleep	0.6
PIR board prototype	0.2

Table 4.4: Power consumption of the PIR sensor node.

4.4.2 PIR based people tracking.

In this work the PIR sensor network should provide to the video system a coarse estimation of people position and direction of movement. For this reason we adopt the solution presented in [163]. In this setup the area of interest is covered by an array of PIR sensor nodes that are organized in small clusters. Each cluster is an autonomous network basic block and is made up of two sensors facing each other that locally detect body position and direction of movement through the classification of simple features (signal duration and peak to peak amplitude) extracted from PIR output. One of the two nodes act as a block manager. It receives the features from the other node and perform the classification step. This information is used in conjunction with the simple detection of the first peak direction (either positive or negative) that indicates the direction of movement (see figure 3.15).

A linear Support Vector Machine (SVM) classifier has been used to classify people position into three classes according to their distance from the two PIRs (see figure 4.25). According to the results presented in [163] linear SVM presents a good trade-off between correct position detection (86.06 %) and computational and memory cost (respectively, 6 multiplication, 6 sums and 2 max, and 6 bytes of Flash) and can be efficiently implemented by the low-cost low-power microcontroller that manage the PIR sensor nodes. The confusion matrix for the linear SVM classifier is presented in table 4.5.

Table 4.5: Support Vector Machines classifier's confusion matrix

	classified as		
	close to 1	middle	close to 2
close to 1	166	32	0
middle	14	181	12
close to 2	0	29	190

As can be seen from this table, this classifiers present limited uncertainty since passages in proximity of one PIR sensor are never confused with passages close to the other one.

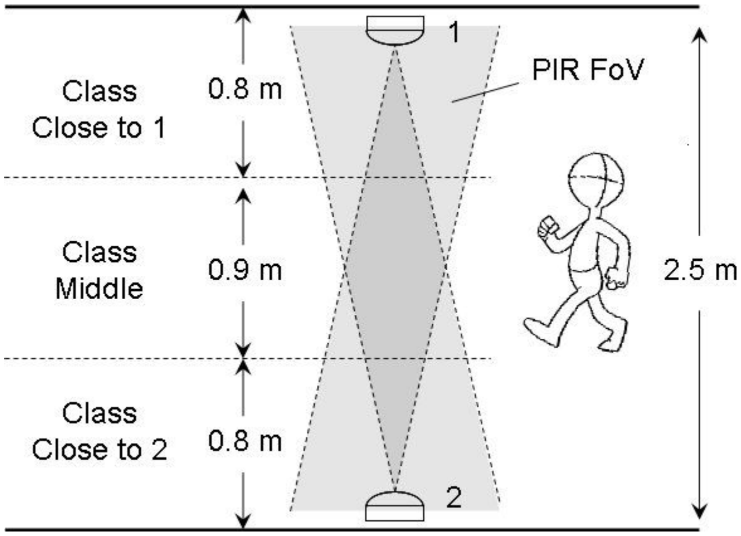


Figure 4.25: Basic configuration used to estimate people position and distance.

4.4.3 Cooperative Ambient Monitoring

In this work we address the scenario where an outdoor, isolated Area of Interest (AoI) is covered by a heterogeneous network made up of our VSNs and PIR sensor nodes. While PIR sensors provide a coarse but continuous coverage of the AoI, VSN are used to identify and report relevant events such as people passages.

The nodes of the network are organized in clusters made up of 2 or 4 VSN and 2 PIR sensor nodes each arranged as presented in figures 4.26 and 4.27. The PIR sensor nodes are in the configuration presented in figure 4.25, while the VSNs point toward the PIR sensors. Each cluster monitors a small part of the AoI and is used in conjunction with other identical ones to cover bigger areas. Each cluster works independently from the others and, with the exception of the final people recognition result, wireless communication is performed only locally among the nodes of a cluster. In our scenario we assume that people move only along three passages, namely *Zone 1*, *Middle Zone* and *Zone 2*.

4.4.4 Multimodal Distributed Power Management

When no transit occurs, the sensor nodes of the network are kept into a low power state. Periodically the VSNs wake-up and poll the PIR manager for synchronization and indication of passages. As the PIR motes detect a transit, the direction of movement is evaluated as well as the zone the body is moving. This information is broadcast to the VSNs of the cluster.

Based on body direction and position and available energy, each VSN computes a cost function that represent the ability offered to wake-up the camera

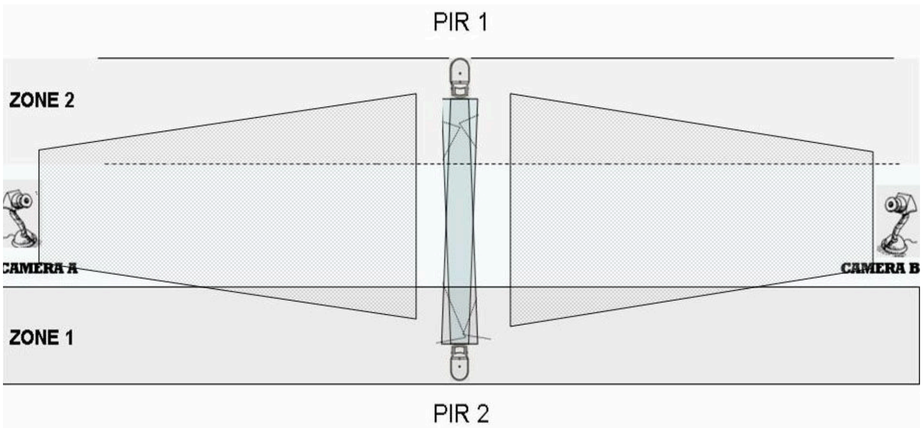


Figure 4.26: Cluster with 2 VSN.

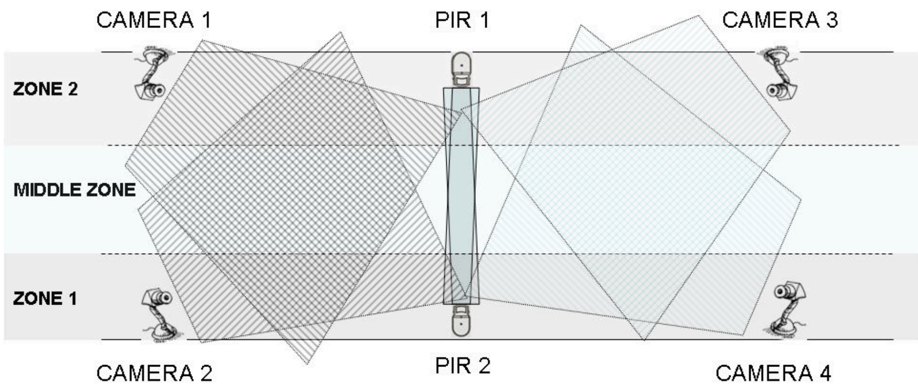


Figure 4.27: Cluster with 4 VSN.

Direction	VSN A	VSN B
Left to Right	20	1
Right to Left	1	20

Table 4.6: Values for the D factor for the 2 VSNs cluster

Direction	VSN 1	VSN 2	VSN 3	VSN 4
Left to Right	20	20	1	1
Right to Left	1	1	20	20

Table 4.7: Values for the D factor for the 2 VSNs cluster

and to correctly identify the body. For this reason the cost function returns a higher value when the body is moving toward the VSN and in its field of view center.

The cost function has the following expression:

$$CF = \frac{E}{C} \cdot \frac{1}{P \cdot D} \cdot \gamma(E, C) \tag{4.6}$$

Where E denotes the actual energy available of the node, C the max capacity of battery, P and D are weights factors depending on body position and direction and $\gamma(E, C)$ is a non linear factor used to decrease the weight of nodes with low energy. The ratio of the energy used in comparison to the accumulator capacity represents an important parameters to trade off with the accuracy of the calssification and the selection of the best camera.

Position and direction influence the CF and the performance during simulations. Tables 4.6, 4.7 and 4.8 present the optimal values.

As can be seen from tables 4.6 and 4.7 when a body is moving toward a VSN the value of D is much smaller. As a consequence the VSNs that see the face of the person are selected even if the others have much more available energy. When the two VSNs facing the front of the person are close to run out of energy the others result in a higher CF and can still provide some information on the people passing.

The value of the P factor is used only in the 4 VSNs cluster (in the 2 VSNs cluster is always 1) in order to distinguish between VSNs that face the front of the body and select the one that better points toward it.

Finally, the value of γ helps to reduce the probability that a node with low available energy is activated. This parameter assumes the following values:

Zone	VSN 1	VSN 2	VSN 3	VSN 4
Zone 1	1	3	1	3
Zone Middle	1	1	1	1
Zone 2	3	1	3	1

Table 4.8: Values for the P factor (used only in the 4 VSNs cluster)

$$if \quad \frac{E}{C} \quad \geq TH \quad \gamma = 1 \tag{4.7}$$

$$else \quad \gamma = 0.5 \tag{4.8}$$

Once the nodes compute their own CF they broadcast it to the other VSNs. A timeout is used in order not to stuck at if any other nodes message is lost. Than locally, each VSN check if its own CF is higher than the others. If any of the CF from the other VSNs is higher than the local one the VSN switches on is imager and start processing the image.

Since the camera mote knows at least its own CF, when the timeout expires, if any message has been received it consider itself the best VSN and starts acquiring the image. Therefore, if a transit is detected by PIR sensors at least one VSN turns on. Such approach is robust and guarantees that every event detections will be served. In fact, if some bidding messages does not arrive, the camera deems to have the highest CF providing an activation. In the worst case, more cameras will be activated after a single event. This guarantees to not miss any events, but on the other hand there is an overhead form the power consumption point of view.

To show the influence of the chosen parameters figure 4.28 presents the energy level of the four nodes in the hypothetical case where passages happen only from right to left in Zone 2 and no energy is harvested. In this case, if all VSNs have the same amount of energy, VSN 1 is the best candidate to detect the body. However as its energy decreases at a certain point VSN 2 will result in a higher CF and starts detecting transits. After a while also VSN 3 and 4 start processing images even if they can not see the face of the person since also VSN 2 energy is depletehd and its CF is lower than the one of VSN 3 and 4.

4.4.5 System Lifetime Evolution

We compared the two variants of the proposed approach with the case where the area of interest is covered by 4 VSNs equipped with a PIR sensor that produce a wake-up signal in presence of bodies [101]. In the latter case the camera and the ARM9 microcontroller are active when a person enter in the field of

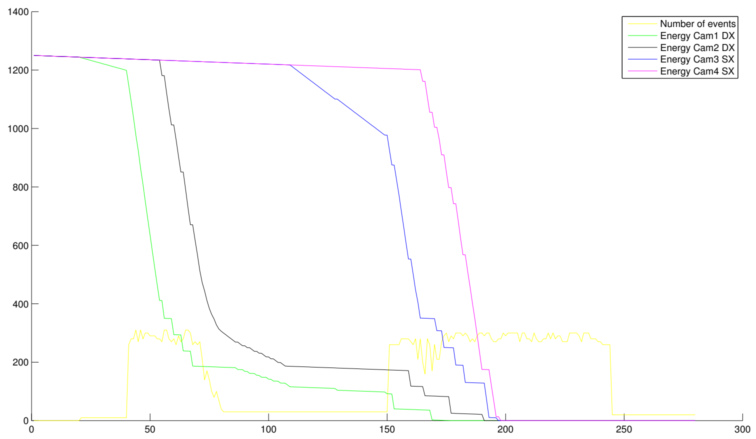


Figure 4.28: Simulation of VSNs energy level when passages occur always in the same position.

view of the imager. Once the node is awake, it processes the image from the CMOS camera in the same manner as described above. The size of the VSN field of view is modulated by changing the threshold above which the PIR sensor produces a wake-up trigger.

In this case VSNs do not have to broadcast the value of their cost function, thus they can save energy. However, the system do not use efficiently its resources. In fact the presence of other VSN that cover the same field of view is not taken into account and when a body moves in proximity of multiple VSNs all of them wake up. Moreover the work presented [101] do not consider people direction of movement in order to select which camera can better identify the subject.

4.4.6 Experimental result vs Camera with PIR

To evaluate the effectiveness of our approach, we simulated how the energy of the VSNs evolved as people passed across the PIR sensors.

People passages are modeled according to a profile of events that describes passages during 2 consecutive days in front our our lab (see figure 4.29). The energy intake from the energy harvesters has been modeled by measuring the incident solar light intensity measured during the same period (see figure 4.30). Figure 4.31 compares VSNs over the 2 days period time. All VSNs are equipped with a 40F supercapacitor to store energy from the solar harvester.

At simulation start, when no events are detected and no energy is harvested we see how the solution proposed in [101] presents less power consumption. This is related to the fact that no wireless messages are sent for synchronization and can be seen comparing the nodes energy levels on the box on the left part

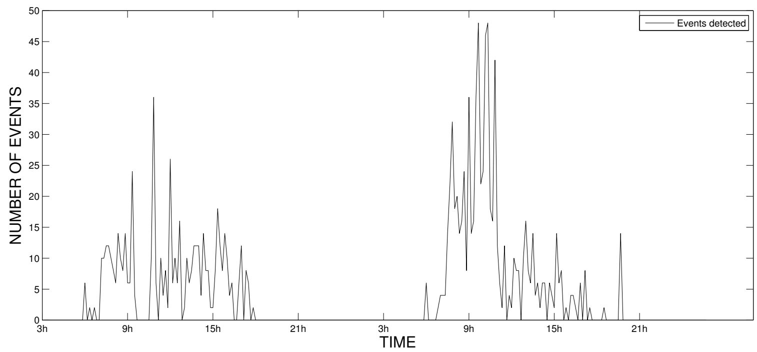


Figure 4.29: Number of events detected from PIR sensors in 2 consecutive days

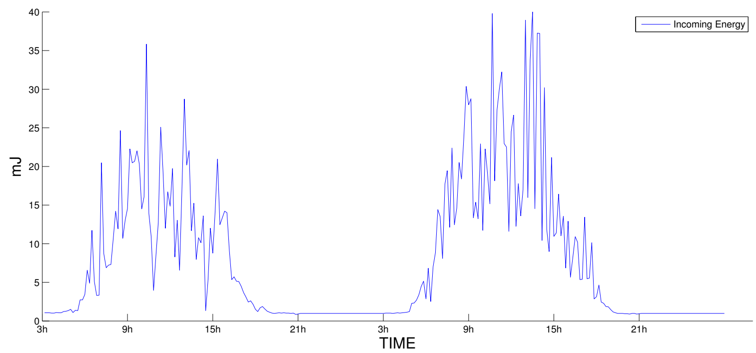


Figure 4.30: Energy incoming from energy harvesters during 2 consecutive days.

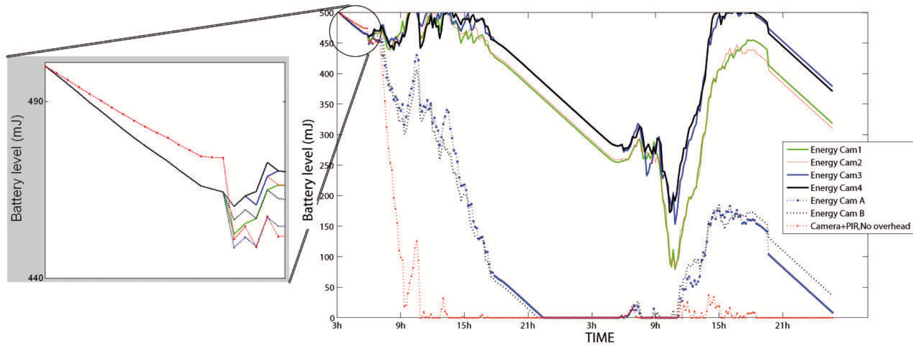


Figure 4.31: Simulation 4 cameras against a camera with PIR and 2 cameras in 2 sides

of figure 4.31.

As people start passing the solution with on board PIR (dotted line) quickly consumes its energy since all four VSNs are waken-up at every passage, despite person position and direction of movements. As a results after few hours the node has already exhausted his energy and can not monitor the area of interest anymore. To perform continuous operation in the proposed scenario, each sensor node should be equipped with a 300F supercapacitor.

The solution with 2 VSNs performs better than the previous one. As we can see from figure 4.31 computation is balanced among the two VSNs, thus the system is able to monitor the area of interest until evening when the gates of the building close. However, before the sun set the VSNs do not collect enough energy to operate all night long. Furthermore, as the second day starts, they need time to replenish their energy, so this system is not able to continuously operate during the second morning. To perform continuous operation each VSN should be equipped with a 102F supercapacitor.

Finally the solution with 4 VSNs is able to operate continuously with a 40F supercapacitor.

4.4.7 Conclusion

Wireless Video Sensor Networks (WVN) made up of a large number of Wireless Video Sensor Nodes (VSN) are gaining popularity as a flexible mean to monitor remote areas.

For this kind of systems, power-aware design is crucial since battery replacement is often unfeasible or too expensive.

Low power hardware that can operate in low power states when no events occur is a standard choice when dealing with power-aware design. Furthermore, a network of low power sensors (i.e. passive infrared sensors) may provide trigger capabilities in order to keep the system into a low power state as long as no events are detected. Solar harvesting capabilities may further extend nodes lifetime in an outdoor scenario. However since solar irradiance is not predictable, careful power management is still necessary.

Further power saving policies can be defined when considering the network as a whole. In this case redundancies can be exploited to balance works among the nodes of the network and relax the constraints on the harvested energy.

In this section we presented a multimodal ambient monitoring system where all system design steps are optimized for low-power consumption. This systems stems from the conjunction of 2 sensor network: a low power, low cost PIR based sensor network and a WVN. The former is responsible to provide a coarse, yet continuous monitoring, the latter is activated only when events are detected and aim at a better classification of the event itself. Nodes are organized in clusters made up of 2 or 4 VSN and 2 PIR sensors.

We proposed a distributed policy where each VSN, on the basis of the information from the PIR sensor network and its available energy, computes a cost function that is broadcast to the other nodes of the network. By comparing its own cost with the one received from the other nodes a VSN understands if it must monitor the event or it can stay into low power state. In the former case the VSN CMOS imager is turned on, an image is acquired and classified in order to understand if the event was generated by a person.

This is a robust policy, since for every event at least one VSN is activated, despite some messages may be lost. In fact each VSN has at least its own cost, therefore, if any other message is received it turns on. Furthermore, since this policy is distributed among the nodes of the network we do not have single point of failure for the whole system.

We compared our solutions with the one proposed in a previous work shown in 4.2. We showed that with our approaches we can achieve continuous operation with a 40F or 102F supercapacitor (4 or 2 VSN respectively) which are respectively 7.5 and 3 times smaller than the one needed for continuous operation of the system described in the previous work (300F).

Conclusions

Ambient Intelligence promotes pervasive and distributed technologies that are not intrusive and always present. Wireless Sensor Network (WSN) is certainly the most important of these technologies and allows an environment (such as a room, a building, a park) to be user-interactive and to be aware of the intentions of the users.

The dependence on a large amount of fixed and mobile sensors embedded into the environment makes *Wireless Sensor Networks* (WSNs) one of the most relevant enabling technologies for AmI. WSNs are complex systems made up of a number of sensor nodes, simple devices that typically embed a low power computational unit (microcontrollers, FPGAs etc.), a wireless communication unit, one or more sensors and a some form of energy supply (either batteries or energy scavenger modules). Low-cost, low-computational power, low energy consumption and small size are characteristics that must be taken into consideration when designing and dealing with WSNs.

Low-cost and low-power video surveillance systems based on networks of wireless video sensors will soon enter the marketplace with the promise of flexibility, quick deployment and providing accurate and real-time visual data. Energy autonomy and efficiency of the implemented algorithms are undoubtedly the primary design challenges to be addressed on systems subject to low computational capabilities and memory constraints.

In this thesis we have discussed our results about the hardware/software design of monitoring systems that can receive their energy from regenerative sources such as solar cells. We started with the design of video sensor nodes, suitable also for wearable computing applications and we have continued with the analysis of embedded video processing algorithm to achieve an intelligent node and use the wireless communication just when it is needed.

We then focused on the problem of extension of battery lifetime of a wireless surveillance system. We shown how we can extend the lifetime of a wireless video node powered by a solar scavenger using a PIR sensor and a tunable wake-up threshold. Moreover we show how the design of a multimodal platform equipped with different family of vision sensors with heterogeneous features of power consumption and resolution permits us to adopt very effective energy management techniques reducing considerably the activation of the camera and other power consuming devices.

In conclusion, Ambient Intelligence will have a major impact on software and embedded systems design. It will introduce many new media applications and new user interface concepts, bringing innovations in several fields of human activity. In this thesis we have contributed tackling some of the numerous open research challenges in the sensor networks domain.

Publications

During the Phd several papers are published for some journals and international conferences. Below the list:

Journal

A. Kerhet, M. Magno, F. Leonardi, A. Boni, and L. Benini. *A low-power wireless video sensor node for distributed object detection*. Journal of Real-Time Image Processing, Springer, Vol.2(4):331342, 2007

Alex E. Susu, Michele Magno, Andrea Acquaviva, David Atienza, Giovanni De Micheli. *Exploration of Reconfiguration Strategies for Environmentally Powered Devices.* Transactions on HiPEAC-1, Lecture Notes in Computer Science (LNCS), Springer-Verlag Berling Heidelberg New York. 2007

Conferences

A. Kerhet, F. Leonardi, A. Boni, P. Lombardo, M. Magno, and L. Benini. *Distributed video surveillance using hardware-friendly sparse large margin classifiers.* In AVSS 2007: Proceedings of the 2007 IEEE International Conference on Advanced Video and Signal based Surveillance, 2007.

M. Magno, L. Benini *A Low-Power Configurable Wireless Video Sensor Node for Distributed Vision Applications.* In Proc. International Conference on Distributed Smart Cameras. 2007

M. Magno, D. Brunelli, P. Zappi, and L. Benini.*A self-powered video node triggered by pir sensors*. In 5th European Conference on Wireless Sensor Networks (EWSN), Ganuary 2008.

M. Magno, D. Brunelli, P. Zappi, and L. Benini. *A solar-powered video sensor node for energy efficient multimodal surveillance*. In DSD 08: Proceedings of the 2008 11th EUROMICRO Conference on Digital System Design Architectures,Methods and Tools, pages 512519, Washington, DC, USA, 2008. IEEE Computer Society.

M. Magno, F. Tombari, D. Brunelli, L. Di Stefano, and L. Benini. *Multi-modal video surveillance aided by pyroelectric infrared sensors*. In Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion, Algorithms and Applications (M2SFA2), 2008.

M. Magno, D. Brunelli, L. Benini *Detection of abandoned/removed objects with a video sensor node aided by Infrared Sensor.* In Proc. 6th European Conference on Wireless Sensor Networks. 2009

M. Magno; F. Tombari; D. Brunelli; L. Di Stefano; L. Benini *Multimodal Abandoned/Removed Object Detection for Low Power Video Surveillance Systems.* In Proc. Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova 2009

M. Magno, D. Brunelli, L. Thiele and L. Benini *Adaptive Power Control for Solar Harvesting Multimodal Wireless Smart Camera.* In Proc. Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009)

Submitted

Magno M., Zappi P., Brunelli D., L. BENINI *Energy Efficient Cooperative Multimodal Ambient Monitoring.*

Magno M., Lanza A. Brunelli D., Di Stefano L., Benini L. *Energy aware multimodal video surveillance embedded system.*

Magno M., Brunelli D., Benini L. *Resource manager for video surveillance embedded system.*

Bibliography

[1] VC 4465 www.vision-components.com.

[2] EyeSpector? 4400 www.eyespector.com.

[3] VETECH VE 262 www.canpolar.com/vetech/.

[4] Bi-i V301 www.analogic-computers.com.

[5] LW-1.3-G-1394 www.isgchips.com.

[6] Elphel 333 www3.elphel.com/en/products.

[7] Pixel Velocity www.pixelvelocity.com/.

[8] Iris P-Series www.matrox.com/imaging/products/.

[9] Tag Plus www.tattile.com/new.site/start.htm.

[10] Dica21 www.apptech.philips.com/industrialvision.

[11] Imapcar www.eu.necel.com/applications/automotive.

[12] XC1-XS1 www.sonybiz.net/biz/.

[13] Atmel FPSLIC www.atmel.com/products/FPSLIC/.

[14] Gnu project http://www.gnu.org/.

[15] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. *Computer Networks*, 38:393–422, 2002.

[16] Alan Daniels. Ubiquitous computing. http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html, 1997.

[17] D. Anguita, S. Pischiutta, S. Ridella, and D. Sterpi. Feed-forward support vector machine without multipliers. *IEEE Trans. on Neural Networks*, 17(5):1328–1331, 2006.

[18] A. Bemporad, F. Borrelli, and M. Morari. Model Predictive Control Based on Linear Programming - The Explicit Solution. *IEEE Transactions on Automatic Control*, 47(12):1974–1985, Dec. 2002.

[19] A. Y. Benbasat and J. A. Paradiso. Groggy wakeup - automated generation of power-efficient detection hierarchies for wearable sensors. In *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) March 26 28, 2007 RWTH Aachen University, Germany*, 2007.

[20] L. Benini, E. Farella, and C. Guiducci. Wireless sensor networks: Enabling technology for ambient intelligence. *Microelectron. J.*, 37(12):1639–1649, 2006.

[21] K. P. Bennett. Support vector machines: Hype or hallelujah? *SIGKDD Explorations*, 2:2000, 2000.

[22] Berkley. http://robotics.eecs.berkeley.edu/ pister/SmartDust/.

[23] M. Beynon, D. Van Hook, M. Seibert, A. Peacock, and D. Dudgeon. Detecting abandoned packages in a multi-camera video surveillance system. In *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS 03)*, pages 221–228, 2003.

[24] M. Bhargava, C. Chen, M. Ryoo, and J. Aggarwal. Detection of abandoned objects in crowded environments. In *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS07)*, pages 271–276, 2007.

[25] D. Bhatia, L. Estevez, and S. Rao. Energy efficient contextual sensing for elderly care. *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 4052–4055, Aug. 2007.

[26] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos. Real time, online detection of abandoned objects in public areas. In *Proc. IEEE Conf. on Robotics and Automation (ICRA06)*, pages 3775 – 3780, 2006.

[27] P. W. Boettcher and G. A. Shaw. Energy-constrained collaborative processing for target detection, tracking, and geolocation. In *IPSN*, pages 254–268, 2003.

[28] D. Brunelli, L. Benini, C. Moser, and L. Thiele. An efficient solar energy harvester for wireless sensor nodes. In *DATE '08: Proceedings of the conference on Design, automation and test in Europe*, pages 104–109, New York, NY, USA, 2008. ACM.

[29] D. Brunelli, E. Farella, L. Rocchi, M. Dozza, L. Chiari, and L. Benini. Bio-feedback system for rehabilitation based on a wireless body area network. *Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on*, pages 5 pp.–531, March 2006.

[30] D. Brunelli, C. Moser, L. Benini, and L. Thiele. An efficient solar energy harvester for wireless sensor nodes. In *DATE '08: Proceedings of the conference on Design, automation and test in Europe*, 2008.

[31] P. Bryant and H. W. Braun. Some applications of a motion detecting camera in remote environments. Technical report, University of California San Diego, HPWREN, 2003.

[32] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[33] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, Kluwer Academic*, 2(2):121–167, 1998.

[34] Carnegie Mellon University. Aura project overview.
http://www-2.cs.cmu.edu/ aura/.

[35] P. Chen, P. Ahammad, C. Boyer, H. Shih-I, L. Leon, E. Lobaton, M. Meingast, O. Songhwai, S. Wang, Y. Posu, A. Yang, Y. Chuohao, C. Lung-Chung, J. Tygar, and S. Sastry. Citric: A low-bandwidth wireless camera network platform. In *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, pages 1–10, Sept. 2008.

[36] C.-Y. Chong and S. Kumar. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91(8):1247–1256, Aug. 2003.

[37] D. Corley and E. Jovanov. A low power intelligent video-processing sensor. In *Proc. Thirty-Fourth Southeastern Symposium on System Theory*, pages 176–178, 2002.

[38] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20, 1995.

[39] D. R. Cox. Prediction by exponentially weighted moving average. pages 414–422, 1961.

[40] CrossBow. http://www.xbow.com/Products/productdetails.aspx?sid=174.

[41] CrossBow. http://www.xbow.com/Products/productdetails.aspx?sid=164.

[42] CrossBow. http://www.xbow.com/index.aspx.

[43] R. Cucchiara, A. Prati, R. Vezzani, L. Benini, E. Farella, and P. Zappi. Using a wireless sensor network to enhance video surveillance. *Journal of Ubiquitous Computing and Intelligence (JUCI)*, 1:1–11, 2006.

[44] D. Culler, D. Estrin, and M. Srivastava. Overview of sensor networks. *IEEE Computer*, 37(8):41–49, 2004.

[45] D. Culler, D. Estrin, and M. Srivastava. Overview of sensor networks. *IEEE Computer*, August 2004.

[46] D. Datcu, Z. Yang, and L. J. M. Rothkrantz. *Multimodal Surveillance: Sensors, Algorithms, and Systems*, chapter Multimodal, pages 311–338. Artech House Publishers, July 2007.

[47] P. de la Hamette et al. Architecture and applications of the fingermouse: a smart stereo camera for wearable computing hci. *Personal Ubiquitous Comput.*, 12(2):97–110, 2008.

[48] L. Di Stefano and A. Bulgarelli. A simple and efficient connected components labeling algorithm. In *Proc. Int. Conf. on Image Analysis and Processing (ICIAP 99)*, pages 322–327, 1999.

[49] C. P. Diehl. *Toward Efficient Collaborative Classification For Distributed Video Surveillance*. PhD thesis, Carnegie Mellon University, 2000.

[50] Q. Dong. Maximizing system lifetime in wireless sensor networks. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 3, Piscataway, NJ, USA, 2005. IEEE Press.

[51] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J. Burgelman. Thats what friends are for ambient intelligence (ami) and the is in 2010. In *Innovations for an e-society*, 2001.

[52] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J. C. Burgelman. Scenarios for ambient intelligence in 2010. Technical report, IST Advisory Group, February 2001.

[53] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2001.

[54] Ember. http://www.ember.com/index.html.

[55] T. Esram and P. Chapman. Comparison of photovoltaic array maximum power point tracking techniques. *Energy Conversion, IEEE Transaction on*, 2:439–449, 2007.

[56] ETH Zurich. http://www.btnode.ethz.ch/.

[57] L. Evers, M. Bijl, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga. Wireless sensor networks and beyond: A case study on transport and logistics, 2005.

[58] X. Fan, W. Shaw, and I. Lee. Layered clustering for solar powered wireless visual sensor networks. In *ISM '07: Proceedings of the Ninth IEEE International Symposium on Multimedia*, pages 237–244, Washington, DC, USA, 2007. IEEE Computer Society.

[59] W.-C. Fang, S. Kedar, S. Owen, G.-Y. Wei, D. Brooks, and J. Lees. System-on-chip architecture design for intelligent sensor networks. In *IIH-MSP '06: Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia*, pages 579–582, Washington, DC, USA, 2006. IEEE Computer Society.

[60] E. Farella, A. Pieracci, and A. Acquaviva. Design and implementation of wimoca node for a body area wireless sensor network. *Systems Communications, 2005. Proceedings*, pages 342–347, Aug. 2005.

[61] J. Feng, F. Koushanfar, and M. Potkonjak. System-architectures for sensor networks issues, alternatives, and directions. *IEEE International Conference on Computer Design (ICCD)*, 2002.

[62] W.-C. Feng, B. Code, E. Kaiser, M. Shea, and L. Bavoil. Panoptes: scalable low-power video sensor networking technologies. In *Proceedings of the ACM International Multimedia Conference*, pages 562–571, Berkeley, Calif, USA, November 2003. TeX Users Group.

[63] W.-c. Feng, B. Code, E. Kaiser, M. Shea, W.-c. Feng, and L. Bavoil. Panoptes: scalable low-power video sensor networking technologies. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 562–571, New York, NY, USA, 2003. ACM.

[64] S. Ferrando, G. Gera, and C. Regazzoni. Classification of unattended and stolen objects in video-surveillance system. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS 06)*, 2006.

[65] L. Ferrigno, S. Marano, V. Paciello, and A. Pietrosanto. Balancing computational and transmission power consumption in wireless image sensor networks. In *IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems*, 2005.

[66] L. Ferrigno and A. Pietrosanto. A low cost visual sensor node for bluetooth based measurement networks. In *Proc. 21st IEEE Instrumentation and Measurement Technology Conference IMTC 04*, volume 2, pages 895–900 Vol.2, 2004.

[67] Freescale. http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=MC13224V&nodeId=0106B9869925657103.

[68] P. Garda, O. Romain, B. Granado, A. Pinna, D. Faura, and K. Hachicha. Architecture of an intelligent beacon for wireless sensor networks. In *Proc. NNSP'03 Neural Networks for Signal Processing 2003 IEEE 13th Workshop on*, pages 151–158, 2003.

[69] A. Giridhar and P. R. Kumar. Maximizing the functional lifetime of sensor networks. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 2, Piscataway, NJ, USA, 2005. IEEE Press.

[70] C. Gui and P. Mohapatra. Power conservation and quality of surveillance in target tracking sensor networks. In *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 129–143, New York, NY, USA, 2004. ACM.

[71] C. Gui and P. Mohapatra. Power conservation and quality of surveillance in target tracking sensor networks. In *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 129–143, New York, NY, USA, 2004. ACM.

[72] T. He, S. Krishnamurthy, L. Luo, T. Yan, L. Gu, R. Stoleru, G. Zhou, Q. Cao, P. Vicaire, J. A. Stankovic, T. F. Abdelzaher, J. Hui, and B. Krogh. Vigilnet: An integrated sensor network system for energy-efficient surveillance. *ACM Trans. Sen. Netw.*, 2(1):1–38, 2006.

[73] T. He, S. Krishnamurthy, L. Luo, T. Yan, L. Gu, R. Stoleru, G. Zhou, Q. Cao, P. Vicaire, J. A. Stankovic, T. F. Abdelzaher, J. Hui, and B. Krogh. Vigilnet: An integrated sensor network system for energy-efficient surveillance. *ACM Trans. Sen. Netw.*, 2(1):1–38, February 2006.

[74] Hill. Mica: a wireless platform for deeply embedded networks. *Micro, IEEE*, 22(6):12–24, Nov/Dec 2002.

[75] Honeywell security and custom electronic. Is-215t datasheet, 2008.

[76] IBM. Planet blue project overview. http://www.research.ibm.com/compsci/planetblue.html.

[77] M. Ilyas, I. Mahgoub, and L. Kelly. *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. CRC Press, Inc., Boca Raton, FL, USA, 2004.

[78] Information Society Technology. Istag mission. http://cordis.europa.eu/ist/istag.htm, 1999.

[79] H. Junker, P. Lukowicz, and G. Tröster. Sampling frequency, signal resolution and the accuracy of wearable context recognition systems. *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, 1:176–177, 31 Oct.-3 Nov. 2004.

[80] M. M. K. Yoshinari and. A human motion estimation method using 3-successive video frames. In *Proc. of Int. Conf. on Virtual Systems and Multimedia GIFU*, pages 135–140, 1996.

[81] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava. Power management in energy harvesting sensor networks. In *ACM Trans. Embedd. Comput. Syst*, 2007.

[82] S. Keerthi, O. Chapelle, and D. DeCoste. Building support vector machines with reduced classifier complexity. *Journal of MachineLearning Research*, 7:1493–1515, 2006.

[83] A. Kerhet, M. Hu, F. Leonardi, A. Boni, and D. Petri. Svm-like algorithms and architectures for embedded computational intelligence. Technical report, DIT, University of Trento, 2007.

[84] A. Kerhet, M. Hu, F. Leonardi, A. Boni, and D. Petri. Svm-like algorithms and architectures for embedded computational intelligence. 2007.

[85] A. Kerhet, F. Leonardi, A. Boni, P. Lombardo, M. Magno, and L. Benini. Distributed video surveillance using hardware-friendly sparse large margin classifiers. In *AVSS 2007: Proceedings of the 2007 IEEE International Conference on Advanced Video and Signal based Surveillance*, 2007.

[86] A. Kerhet, M. Magno, F. Leonardi, A. Boni, and L. Benini. A low-power wireless video sensor node for distributed object detection. *Journal of Real-Time Image Processing, Springer*, Vol.2(4):331–342, 2007.

[87] R. Kling. Intel motes: advanced sensor network platforms and applications. *Microwave Symposium Digest, 2005 IEEE MTT-S International*, pages 4 pp.–, June 2005.

[88] G. Koguta, M. Blackburn, and H.R.Everett. Using video sensor networks to command and control unmanned ground vehicles. *AUVSI Unmanned Systems in International Security(USIS)*, 2003.

[89] Y.-H. Lee and C. Krishna. Voltage-clock scaling for low energy consumption in real-time embedded systems. pages 272–279, 1999.

[90] D. Li, Y. Jiang, and G. Chen. A low cost embedded color vision system based on sx52. In *Proc. IEEE International Conference on Information Acquisition*, pages 883–887, 2006.

[91] J. Lifton, D. Seetharam, M. Broxton, and J. Paradiso. Pushpin computing system overview: a platform for distributed, embedded, ubiquitous sensor networks. In *in F. Mattern and M. Naghshineh (eds): Pervasive 2002, Proceedings of the Pervasive Computing Conference*, pages 139–151. Springer Verlag, 2002.

[92] S. Lim and L. Davis. A one-threshold algorithm for detecting abandoned packages under severe occlusions using a single camera. Technical Report CS-TR-4784, CS Dept., University of Maryland, 2006.

[93] K.-M. Lin and C.-J. Lin. A study on reduced support vector machines. *IEEE Transactions on Neural Networks*, 14(6):1449–1459, 2003.

[94] A. J. Lipton, C. H. Heartwell, N. Haering, and D.Madden. Automated video protection, monitoring and detection. *IEEE Aerospace and Electronic Systems Magazine*, 18:3–18, 2003.

[95] S. Lu, J. Zhang, and D. Feng. A knowledge-based approach for detecting unattended packages in surveillance video. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS 06)*, 2006.

[96] J. Maciejowski. *Predictive Control with Constraints*. Prentice-Hall, 2001.

[97] W. MacLean. An evaluation of the suitability fo fpgas for embedded vision systems. In *IEEE Computer Vision and Pattern Recognition Conference*, 2005.

[98] E. Magli, M. Mancin, and L. Merello. Low-complexity video compression for wireless sensor networks. *IEEE International Conference on Multimedia and Expo (ICME)*, 2003.

[99] M. Magno, D. Brunelli, L. Thiele, and L. Benini. Adaptive power control for solar harvesting multimodal wireless smart camera. In *ICDSC*. IEEE Computer Society, 2009.

[100] M. Magno, D. Brunelli, P. Zappi, and L. Benini. A self-powered video node triggered by pir sensors. In *5th European Conference on Wireless Sensor Networks (EWSN)*, Ganuary 2008.

[101] M. Magno, D. Brunelli, P. Zappi, and L. Benini. A solar-powered video sensor node for energy efficient multimodal surveillance. In *DSD '08: Proceedings of the 2008 11th EUROMICRO Conference on Digital System Design Architectures, Methods and Tools*, pages 512–519, Washington, DC, USA, 2008. IEEE Computer Society.

[102] M. Magno, F. Tombari, D. Brunelli, L. Di Stefano, and L. Benini. Multi-modal video surveillance aided by pyroelectric infrared sensors. In *Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion, Algorithms and Applications (M2SFA2)*, 2008.

[103] M. Magno, F. Tombari, D. Brunelli, L. di Stefano, and L. Benini. Multimodal abandoned/removed object detection for low power video surveillance systems. In S. Tubaro and J.-L. Dugelay, editors, *AVSS*, pages 188–193. IEEE Computer Society, 2009.

[104] C. B. Margi, X. Lu, G. Zhang, G. Stanek, R. Manduchi, and K. Obraczka. Meerkats: A power-aware, self-managing wireless camera network for wide area monitoring. In *Workshop on Distributed Smart Cameras (DSC'06)*, October 2006.

[105] M. Maroti, G. Simon, A. Ledeczi, and J. Sztipanovits. Shooter localization in urban terrain. *Computer*, 37(8):60–61, Aug. 2004.

[106] K. Martinez, J. K. Hart, and R. Ong. Environmental sensor networks. *Computer*, 37(8):50–56, 2004.

[107] Clare, Inc. Clare solar cells - cpc series data sheet http://www.clare.com/products/solarcell.htm. August, 2005.

[108] MIT. Oxygen project overview. http://www.oxygen.lcs.mit.edu/.

[109] E. Monton, J. Hernandez, J. Blasco, T. Herve, J. Micallef, I. Grech, A. Brincat, and V. Traver. Body area network for wireless patient monitoring. *Communications, IET*, 2(2):215–222, February 2008.

[110] C. Moser, L. Thiele, D. Brunelli, and L. Benini. Adaptive power management in energy harvesting systems. In *Design, Automation and Test in Europe Conference and Exhibition, 2007. DATE '07*, pages 1–6, April 2007.

[111] Murata manufacturing. Pyroelectric infrared sensors, 2005.

[112] L. Nachman, R. Kling, R. Adler, J. Huang, and V. Hummel. The intel®mote platform: a bluetooth-based sensor network for industrial monitoring. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 61, Piscataway, NJ, USA, 2005. IEEE Press.

[113] NASA. http://sensorwebs.jpl.nasa.gov/.

[114] Ognjen Obucina. Ambient intelligence: The vision of information society. http://www.dke.univie.ac.at/extern/bi_ws20012002/ss2002/Ambient-Intelligence.pdf, 2002.

[115] C. Otto, E. Jovanov, and A. Milenkovic. A wban-based system for health monitoring at home. *Medical Devices and Biosensors, 2006. 3rd IEEE/EMBS International Summer School on*, pages 20–23, Sept. 2006.

[116] J. Paradiso and T. Starner. Energy scavenging for mobile and wireless electronics. *Pervasive Computing, IEEE*, 4(1):18–27, Jan.-March 2005.

[117] M. Perillo and W. Heinzelman. *Fundamental Algorithms and Protocols for Wireless and Mobile Networks*. CRC Hall, 2005.

[118] Perking Elmer. Frequency range for pyroelectric detectors. http://www.perkinelmer.com.

[119] Pesch, Dirk and Barrett, John and O'Reilly, Fergus. Ubiquitous communication through adaptive wireless systems. http://www.irishscientist.ie/2001/01toc1.htm, 2001.

[120] Philips. http://www.research.philips.com/technologies/projects/ambintel.html.

[121] F. Porikli, Y. Ivanov, and T. Haga. Robust abandoned object detection using dual foregrounds. *EURASIP Journal on Advances in Signal Processing*, 2008(1), 2008.

[122] D. Puccinelli and M. Haenggi. Wireless sensor networks: applications and challenges of ubiquitous sensing. *Circuits and Systems Magazine, IEEE*, 5(3):19–31, 2005.

[123] J. Rabaey, M. Hammer, and J. D. Silva. Picoradio supports ad-hoc ultra low-power wireless networking. *IEEE Computer*, 2000.

[124] M. Rahimi, R. Baer, O. Iroezi, J. Garcia, J. Warrior, D. Estrin, and M. Srivastava. Cyclops: in situ image sensing and interpretation in wireless sensor networks. In *the ACM Conf. on Embedded Networked Sensor Systems (SenSys)*, San Diego, CA, November 2005.

[125] K. Raja, I. Daskalopoulos, H. Diall, S. Hailes, and T. Torfs. Sensor cubes: A modular, ultra-compact, power-aware platform for sensor networks. *International Conference on Information Processing in Sensor Networks (IPSN SPOTS)*, April 2006.

[126] A. Rajgarhia, F. Stann, and J. Heidemann. Privacy-sensitive monitoring with a mix of IR sensors and cameras. In *Proceedings of the Second International Workshop on Sensor and Actor Network Protocols and Applications*, pages 21–29, August 2004.

[127] N. K. Ratha and A. K. Jain. Computer vision algorithms on reconfigurable logic arrays. *IEEE Computer*, 10(1):29–43, Jan. 1999.

[128] K. Romer and F. Mattern. The design space of wireless sensor networks. *Wireless Communications, IEEE [see also IEEE Personal Communications]*, 11(6):54–61, Dec. 2004.

[129] C. Sacchi and C. Regazzoni. A distributed surveillance system for detection of abandoned objects in unmanned railway environments. *IEEE Trans. Vehicular Technology*, 49(5), 2000.

[130] J. San Miguel and J. Mart?nez. Robust unattended and stolen object detection by fusing simple algorithms. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS 08)*, pages 18–25, 2008.

[131] B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *In Proceedings of the Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE Computer Society, 1994.

[132] B. N. Schilit and M. M. Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8:22–32, 1994.

[133] B. Schölkopf and A. Smola. *Learning with Kernels*. The MIT Press, 2002.

[134] http://www.sensorsportal.com/.

[135] Sentilla. http://www.sentilla.com/moteiv-transition.html.

[136] R. C. Shah, L. Nachman, and C.-y. Wan. On the performance of bluetooth and ieee 802.15.4 radios in a body area network. In *BodyNets '08: Proceedings of the ICST 3rd international conference on Body area networks*, pages 1–9, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[137] M. Shankar, J. B. Burchett, Q. Hao, B. D. Guenther, and D. J. Brady. Human-tracking systems using pyroelectric infrared detectors. *Optical Engineering*, 45(10):106401 1–10, 2006.

[138] C.-N. Shyi, W.-F. Weng, and H.-T. Wu. Low power platform for wireless sensor network. volume 2, pages 417–421, July 2005.

[139] F. Simjee, D. Sharma, and P. H. Chou. Everlast: long-life, supercapacitor-operated wireless sensor node. In *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems*, pages 315–315, New York, NY, USA, 2005. ACM.

[140] P. Spagnolo, A. Caroppo, M. Leo, T. Martifiggiano, and T. D'orazio. An abandoned/removed objects detection algorithm and its evaluation on pets datasets. In *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS 06)*, page 17, 2006.

[141] M. Spengler and B. Schiele. Automatic Detection and Tracking of Abandoned Objects. In *Proc. IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003.

[142] D. C. Steere, A. Baptista, D. McNamee, C. Pu, and J. Walpole. Research challenges in environmental observation and forecasting systems. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 292–299. ACM, 2000.

[143] R. Szewczyk, J. Polastre, A. Mainwaring, and D. Culler. Lessons from a sensor network expedition. In *EWSN '00: Proceedings of the 1st European Workshop on Wireless Sensor Networks*, pages 307–322, 2004.

[144] Y. e. a. Tian. Robust and efficient foreground analysis for real-time video surveillance. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 05)*, pages 1182–1187, 2005.

[145] F. Tombari, L. Di Stefano, and S. Mattoccia. A robust measure for visual correspondence. In *Proc. Int. Conf. on Image Analysis and Processing*, pages 376–381, 2007.

[146] V. Vapnik. *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1982.

[147] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[148] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[149] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, September 1998.

[150] M. Vieira, J. Coelho, C.N., J. da Silva, D.C., and J. da Mata. Survey on wireless sensor network devices. *Emerging Technologies and Factory Automation, 2003. Proceedings. ETFA '03. IEEE Conference*, 1:537–544 vol.1, Sept. 2003.

[151] M. A. M. Vieira, A. B. da Cunha, and D. C. J. da Silva. Designing wireless sensor nodes. In *SAMOS*, pages 99–108, 2006.

[152] X. Wang, D. Bi, L. Ding, and S. Wang. Bidding and voting strategy for energy efficient collaborative target classification in wireless sensor networks. In *CISP '08: Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 4*, pages 23–27, Washington, DC, USA, 2008. IEEE Computer Society.

[153] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister. Smart dust: Communicating with a cubic-millimeter computer. *Computer*, 34(1):44–51, 2001.

[154] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, September 1991.

[155] M. Weiser. Hot Topics: Ubiquitous Computing. *IEEE Computer*, Oct. 1993.

[156] M. Weiser, R. Gold, and J. S. Brown. The origins of ubiquitous computing research at parc in the late 1980s. *IBM Syst. J.*, 38(4):693–696, 1999.

[157] C. Weyrich. Orientations for workprogramme 2000 and beyond. Technical report, IST Advisory Group, September 1999.

[158] J. S. Wilson. *Sensor Technology Handbook*. Elsevier, 2005.

[159] D. Wright, S. Gutwirth, M. Friedewald, E. Vildjiounaite, and Y. Punic. *Safeguards in a World of Ambient Intelligence*. Springer, 2008.

[160] M. Wu, B. Schölkopf, and G. Bakir. Building sparse large margin classifiers. In *Proceedings of the 22nd International Conferenceon Machine Learning*, pages 1001–1008, 2005.

[161] M. Wu, B. Schölkopf, and G. Bakir. A direct method for building sparse kernel learning algorithms. *Journal of Machine Learning Research*, 7:603–624, 2006.

[162] N. Zamora, J.-C. Kao, and R. Marculescu. Distributed power-management techniques for wireless network video systems. pages 1–6, April 2007.

[163] P. Zappi, , E. Farella, and L. Benini. Pyroelectric infrared sensors based distance estimation. *to be published in Sensor 2008*, 2008.

[164] P. Zappi, E. Farella, and L. Benini. Enhancing the spatial resolution of presence detection in a pir based wireless surveillance network. *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 295–300, Sept. 2007.

[165] P. Zappi, C. Lombriser, E. Farella, D. Roggen, L. Benini, and G. Tröster. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *5th European Conference on Wireless Sensor Networks (EWSN)*, pages 17–33, January 2008.

[166] G. Zhang, T. Yang, S. Gregori, J. Liu, and F. Maloberti. Ultra-low power motion-triggered image sensor for distributed wireless sensor network. In *Proc. IEEE Sensors*, volume 2, pages 1141–1146 Vol.2, 2003.

[167] F. Zhao and L. Guibas. *Wireless Sensor Networks: An Information Processing Approach (The Morgan Kaufmann Series in Networking)*. Morgan Kaufmann, July 2004.

[168] X. Zhu, S. Ying, and L. Ling. Multimedia sensor networks design for smart home surveillance. *Control and Decision Conference, 2008. CCDC 2008. Chinese*, pages 431–435, July 2008.