Robotics and Autonomous Systems 62 (2014) 1316–1328

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

A distributed perception infrastructure for robot assisted living

Stefano Ghidoni [a,*], Salvatore M. Anzalone [b], Matteo Munaro [a], Stefano Michieletto [a], Emanuele Menegatti [a]

[a] *Intelligent Autonomous Systems Laboratory, University of Padova, Padova, Italy*

[b] *Institut des Systèmes Intelligents et de Robotique, University Pierre et Marie Curie, Paris, France*

HIGHLIGHTS

- We developed a novel distributed infrastructure based on open-source middleware.
- We realized an ambient intelligence system for assisted living.
- The system does not require people cooperation.
- Cooperation of the audio and video sensors increases performance and reliability.
- Integration with a robot for performing actions on the environment.

ARTICLE INFO

Article history:
Available online 19 April 2014

Keywords:
Ambient intelligence
Intelligent autonomous systems
Assisted living
Camera network
Distributed sensing
Autonomous robots

ABSTRACT

This paper presents an ambient intelligence system designed for assisted living. The system processes the audio and video data acquired from multiple sensors spread in the environment to automatically detect dangerous events and generate automatic warning messages. The paper presents the distributed perception infrastructure that has been implemented by means of an open-source software middleware called NMM. Different processing nodes have been developed which can cooperate to extract high level information about the environment. Examples of implemented nodes running algorithms for people detection or face recognition are presented. Experiments on novel algorithms for people fall detection and sound classification and localization are discussed. Eventually, we present successful experiments in two test bed scenarios.

1. Introduction

Distributing intelligence among agents and objects that share the same environment is the main idea of Ambient Intelligence (AmI) [1–3]. Such systems are based on sensors and actuators spread across the environment to capture relevant information and to act on the environment accordingly.

Research in this area focused on the extraction of relevant information from the raw data streams collected by sensors [4–6] and on user profiling [7] (i.e. the intelligent environment can adapt its behaviors according to the people who are inside it). In recent years, particular attention has been given to ambient intelligence applications aimed at helping elderly people living alone and maintaining their independence, while assuring a high level of assistance when needed [8,9].

In this paper, an ambient intelligence system meant for assisted living and surveillance applications is described. Our system is composed of several audio and video sensors spread in the environment that let the system process images and sounds from multiple sources, and acquire also panoramic views of the room. The system presented in this paper offers two main advantages over the works presented in the literature: first, it includes both the audio and video event detection, that is identified as being highly beneficial also by previous works [10]. Second, unlike most of the works in this field [11], it does not rely on accelerometers or any other technology that requires the cooperation from the people living in the intelligent environment. This is the first time a “passive” system handling both audio and video data is presented in the literature.

The capability of an AmI system to work without needing wearable sensors is particularly important when assisting elderly people [10]. In fact, it is well known from clinical studies [12] and by companies in the business of *assisted and monitored living for elderly people* that any system based on the cooperation of the monitored persons is doomed to fail, even if this “cooperation” means only to

* Corresponding author.
E-mail address: ghidoni@dei.unipd.it (S. Ghidoni).

http://dx.doi.org/10.1016/j.robot.2014.03.022
0921-8890/

wear a small intelligent device (e.g. a bracelet with an accelerometer or a more intrusive necklace with an emergency button on it). Therefore, our system is designed to be non intrusive and to work with non-cooperating people.

An additional feature of our system is the use of the innovative Omnidome® [13] dual-camera sensor together with standard perspective and Pan–Tilt–Zoom (PTZ) cameras. The Omnidome sensor is particularly suited for video-surveillance applications, as it couples the 360° panoramic view of an omnidirectional camera with the high resolution offered by a PTZ unit; the cross-calibration of the two sensors represents a strong plus, that makes it possible to consider it as a single, high-resolution omnidirectional sensor. This paper presents the first experiments employing this sensor in an assisted living application, and shows which are the benefits in such context.

The system focuses on the detection of the audio and video salient events in order to guarantee a safe and secure environment for elderly people living at home. Several papers in the literature describe which are the main tasks to be accomplished for assisting elderly people living at home [1,14]. The scenario outlined in these papers is rather complex, because the events to be detected can belong to the audio, video or inertial sensory field. Systems for assisted living are therefore equipped with very heterogeneous sensors, in order to gather a high amount of data about the environment, and better knowledge about the context. The ultimate goal of assisted living is to provide a set of user-centered services [1,15].

The structure of an assisted living system is usually based on a set of applications working at the same time [16,17]. Such applications are quite independent, because they analyze data provided by diverse sensors, and aim at detecting very different events: for this reason a data fusion stage is often missing, since observations made by different sensors are unrelated.

Our system aims at detecting high level knowledge about the environment, and exploits several processing algorithms: some of them, like fall detection, are specific for assisted living, and were expressly developed for this system, while others (like face detection) are more general, therefore state-of-the-art solutions have been employed. A Pioneer P3-AT robot is also part of the system to extend the range of functionalities of a typical AmI setup: the integration of a mobile platform represents a strong advantage, since it enables the system to become active on the environment [18].

A key advantage of the system we propose is that it is built upon a framework that handles communications among different nodes that process the video streams acquired by the cameras. In other words, we exploited an existing framework meant for multi-media applications to implement the distributed architectures described in the literature [16]. This is the only way to ensure fast realization and scalability of this kind of systems that are spread across the environment. Since the system is multimodal i.e. it analyzes the audio and video data, we chose a multi-media framework called NMM (Network-integrated Multi-media Middleware) [19].

2. System infrastructure

Our system is part of the "Safe Home" project, and was designed to provide increased safety to elderly people living at home. This includes a wide variety of services:

- a home security system,
- a home safety system,
- a domotic system for automating doors, windows and air conditioning,
- a video communication framework for connecting the assisted person with friends and relatives,
- a telemedicine system,
- a data collection and recording infrastructure,
- a front-end for handling the interface from and to external systems (e.g. for controlling the domotic system from outside the assisted living environment).

Regarding the first two points, it should be noted that security means the protection of the assisted people against possible intrusions from outside (e.g. a thief), while safety refers to the protection of the people against unwanted dangers caused inside the apartment (e.g. a fall).

The whole structure of the Safe Home system is shown in Fig. 1. As it can be seen, the system includes sensors, actuators and an interface to a remote control room that is used by human operators to react to the alarms issued by the system. Such system has been developed in the context of a cooperation project among university and industry. Our task was to develop computer vision and audio processing algorithms to assist elderly people: we therefore decided to focus on fall detection, that is the most dangerous event in this context [12], and on sound classification, focused on events related to security.

As in the case of several AmI systems proposed in the literature [16], our system relies on multiple sensors (like microphones and cameras). A distributed processing is therefore desirable for scalability purposes. In our approach, each sensor is connected to a separate computer that processes its data stream to extract high-level information. This model lets the system communicate effectively, because only compact high-level information is exchanged, while heavy data streams are processed locally. A centralized approach would impose strong limitations on the total number of sensors, because only few of them can be connected to the same computer, and cables have a maximum length in the orders of a few meters.

To realize the distributed perception infrastructure described above, we exploited the open-source Network-integrated Multimedia Middleware (NMM) by Motama [19], that provides communication facilities over the network. In NMM, the basic processing unit is a node, and every processing algorithm should be placed inside a node in order to interact with other entities in the middleware.

Nodes are organized by means of graphs, that describe how nodes are connected and exchange data. One of the main advantages of using NMM is that a single graph can connect nodes that reside in different machines on the same network, without any overhead for handling the communication over the network. A complex perception system can be synthesized by implementing the system functionalities inside nodes in the appropriate machines, and connecting them using a distributed graph.

NMM is not designed to control active devices, thus for controlling mobile robots that can act in the environment a different system has to be used. The control of the mobile robot in the environment is achieved through the Robot Operating System (ROS) framework [20], that provides an easy access to both high level behavior and low level control of the robot. We created a communication link between NMM and ROS to let the robot and the camera network cooperate effectively. In the following sections, we present the different processing modules and how we have implemented the cooperation among them, in order to achieve the desired system functionalities.

An example: the click & move functionality

To show the typical structure of a NMM graph and the set of processing nodes involved, the simple "click & move" application is considered. This was developed for sending commands from the camera network to the robot. It presents a graphical user interface in which all video flows collected by the different cameras are