

Doctoral School on Engineering Sciences Università Politecnica delle Marche

Extended summary

Robots as Cyber-Physical Systems: new perspectives for the design of autonomous robots

Curriculum: Ingegneria Informatica, Gestionale e dell'Automazione

Author:

Andrea Claudi

Tutor(s):

Prof. Aldo Franco Dragoni

Date: 30-01-2013

Abstract. Today's control systems are made of an embedded system connected to the system to be controlled through a feedback loop; thus the physical process influence the computation and vice versa. The term "cyber-physical system" defines a control system developed according to a new design methodology, which models both the systems taking into account their reciprocal interactions.

In the current scientific literature, the design of control systems for autonomous robots is a problem to be addressed separately from the application of real-time techniques or artificial intelligence techniques on them. However, choices made to solve one of the two problems may have consequences on the other. Therefore, it seems that it is not sufficient to separately understand both the problems: it is important to understand the interaction between them.

This work proposes an alternative approach to the design of intelligent autonomous robots, derived from the cyber-physical approach to control systems, which consists in tackling jointly both the problem of controlling the robotic system and the problem of deploying an artificial intelligent application on it.

The first part of the work describes various modelling abstractions, focusing on hybrid modelling techniques to integrate continuous and discrete models in a single model. Hybrid modelling techniques are subsequently used to describe concurrent computations.

The second part of the work is devoted to embedded and real-time systems. This part focuses on embedded hardware and architectures, particularly describing various I/O and interrupts mechanism, which are responsible for the interaction between the cyber and the physical part of the system. Finally, to take time into account, the real-time operating system abstraction is described.

Finally, the third part of the thesis apply theories and methodologies introduced in the first two part to the design of an intelligent autonomous robot system to be used in the contexts of video-surveillance and ambient-assisted living.

Keywords. Cyber-Physical Systems, Real-Time Operating Systems, Autonomous Robotics, Computer Vision, Intelligent Systems

1 Problem statement and objectives

In good engineering practice, systems are made up of components. In order for the composition - and thus the system - to be well understood, we need to fully understand the components, and then to fully understand the interactions between the individual components.

A cyber-physical system (CPS) is an integration of computation with physical processes [1].

Today's control systems are made up of an embedded system connected to the controlled system through a feedback loop; thus the physical processes influence the computations and vice versa. The term "cyber-physical system" defines a control system developed according to a new modelling methodology, which models both the physical system and the computation system taking into account their reciprocal interactions. This mutual influence is the focus of cyber-physical systems. A CPS is not the union, but the intersection of the physical and the cyber part of the system.

This means that it is not sufficient to separately understand physical and computational components. It is necessary to understand their interactions. The design of such systems, therefore, requires to understand the joint dynamics of computing, software, networks, and physical processes.

Nowadays, CPS are widespread in several applications domain, like communication, infrastructure, health care, manufacturing, military and transportation.

1.1 Motivation

CPS systems are playing a major role in applications of paramount importance. Many of these applications make extensive use of Artificial Intelligence methods, to provide a better integration of the system with the human component, like in human-in-the-loop systems. For example, consider the following application.

Ambient Assisted Living (AAL) defines actions and policies needed to promote and improve living conditions for elderly and disabled people. The goal of AAL is to foster autonomy, safety and social inclusion within domestic spaces.

Consider an AAL system able not only to monitor the living environment of the elderly, but also to act on it [2]. Such a system could interact with the elderly, facilitating normal activities of daily living (opening and closing windows, cupboards and drawers, moving heavy objects, etc.), and monitoring his health. Moreover, the system could be proactive: it could inform and advise the elderly, it could engage him in discussions, ask for his help, etc. The system could automatically answer phone call, if the elderly is not at home, or is not in condition to answer. In emergency situations, the system could contact directly the hospital or the police, if needed.



The system will collect and aggregate data from different sensors. Some of them will provide informations about the environment, some other will be body-area sensors, to monitor health parameters such as temperature and heartbeat. Informations provided by the sensors will be analysed and used for complex logical-symbolical reasoning; employing advanced artificial intelligence methods, the system can deduce facts, make plans and act on the environment. Probably it will need an autonomous robot system to interact with the elderly, and it will need to provide a psychologically acceptable human interface.

Such a system is a typical example of a CPS with a human-in-the-loop, with complex and tight requirements, regarding even juridical, ethical and social issues. This system requires also a strong integration of artificial intelligence methodologies. Our opinion is that such a level of integration could be reached only jointly modelling and developing the control system and the artificial intelligent part of the system, taking into account the timeliness of both the systems.

1.2 Objectives

In the current literature, the design of control systems for intelligent autonomous robots is a problem to be addressed separately from the application of real-time and artificial intelligence techniques on them. However, choices made to solve one of the problems involved in the design of the robot may have consequences on the other problems. Hence, it seems that it is not sufficient to separately understand the problems: it is important to understand the interaction between them.

This work proposes an alternative approach to the design of intelligent autonomous robots, derived from the cyber-physical approach to control systems, which consists in addressing jointly the problems of modelling and interacting with the physical robotic system, controlling it, and using an artificial intelligence application to guide its behaviour.

In particular this work focus on modelling and developing an intelligent autonomous robot controlled by visual information, to be possibly used in contexts such as videosurveillance and ambient assisted living. The proof-of-concept proposed in this work is an autonomous robot whose main task is to follow a frontal face moving in the environment. Visual informations are provided through a smart sensor (e.g. a smartphone camera), while the control system is implemented using a real-time operating system.

2 Research planning and activities

2.1 Requirements

2.1.1 Face-detection and face-recognition

The system has to analyse a video stream to detect and identify human faces and their locations. If an identified face matches with a face present in a database, the systems



tries to keep that face in the centre of the frame, moving the robot.

To satisfy these requirements, we choose to employ a smartphone as a smart sensor. A smart sensor is a sensor which can provide an amount of processing power. It can be used to do some pre-processing on acquired data.

2.1.2 Identifying face position

When the system starts to track a face, it must maintain it in the centre of the video frame. So the robot is required to compensate for the movements of the face, using as references the centre of the frame and the height of the face. The smartphone application is appropriately structured to satisfy these requirements, and positioning errors are calculated basing on the references.

2.1.3 Robot movements

The system must control the movements of the robot. The control system has to determine the appropriate control effort on each actuator, and apply it through output signals. Movements that do not conflict each other can be performed at the same time. Moreover, the control system is supposed to take into account the timeliness of the video processing application, synchronizing with it.

To satisfy the requirement about timeliness, we choose to use a real-time operating system. We choose the ERIKA real-time kernel [3] because it can work on dsPIC architectures, which are good in control and signal processing architectures. Requirements about concurrency were satisfied using hybrid modelling.

2.1.4 Operating modes

It is required that in the event of a failure, the robot stops immediately. If the failure is a temporary condition, after a pre-determined amount of time the system should start to search for faces in the following environment, like previously described. If the system does not detect any face, after a pre-determined amount of time it will return on the stop condition, and so on.

These requirements are about the operating modes of the system. We can distinguish three different operating modes:

- stop;
- search:
- follow.

When the system is in follow mode, it operates tracking an identified face. When the system is in search mode, it observes the surrounding environment, detecting faces and trying to match them with faces stored in the internal database. Finally, when it is in stop mode, the system completely stop. A model of the system, with the appropriate transitions is proposed to satisfy this requirement.



3 Analysis and discussion of main results

3.1 Modelling the autonomous robot



Figure 3.1: A timed automaton depicting robot's three modes of operation

The model in fig. 3.1 describes three different operation modes for the robot.

The system will start in **stop** mode. After a period of time, the system will transition to **search** mode, and start to search for faces. Once a face is identified, the **face** input is set to "true", and the system goes to **follow** mode; otherwise, the system will go back



to stop mode. The face signal is set to "false" in the event of a failure, thus triggering the transition to stop mode.

This model is actually a hierarchical model. The follow mode, indeed, is the result of the parallel composition of three finite state machines, which we describe in the following.



Figure 3.2: An hybrid model describing the control of left / right movements of the robot.

The figure 3.2 depicts a finite state machine describing the control of the left/right degree of freedom of the robot. The system has one input, e_x , and two outputs, l and r, which are two digital PWM signals that are in input, respectively, to the left and the right motor of the robot. To prevent chattering, we use a particular form of state-dependent behaviour, called hysteresis. The same considerations apply to models describing the other degrees of freedom of the robot.

3.2 Hardware and software architecture

Robot's hardware architecture is made up of three main components: an embedded board equipped with a real-time kernel, a smartphone and an USB to UART converter to connect the devices.

The FLEX embedded board is based on a Microchip DSP micro-controller, and hosts the Erika real-time kernel. Erika [3] is a minimal real-time operating system fully compliant to the OSEK/VDX automotive standards. UART connectors are used to interface the board with a USB connector.

The Htc Evo 3D smartphone, equipped with Android 4.0 kernel, is used to capture a video stream of the environment through the rear camera sensor. The smartphone is connected to the FLEX board using a USB converter.



The movements of the robot are controlled by three servomotors. Two of them are used to control the wheels, while the third is used to control the inclination of the smartphone with respect to the ground.

The software architecture is determined by the interconnection of the hardware components of the robot. Two distinct applications are used for video recording and to control the movements of the robot.

The real-time control application runs on the Erika real-time kernel, and is made up of 4 different tasks, scheduled with a fixed-priority algorithm. Three of them are used to control the servomotors via a PWM output from the dsPIC; the fourth task, at an higher priority, is the core of the control application. It reads error data from the smartphone and implements a PID controller for the robot.

The smartphone application uses the OpenCV computer vision library to obtain results comparable with similar systems, as it constitutes a reference implementation for facedetection and face-recognition algorithms. Moreover, OpenCV has a strong focus on computational efficiency to support real-time applications.

When a new face is detected in the video stream, the application compares the center of the frame with the center of the face, and the height of the face with a fixed reference height. Error values are transmitted to the control board which, using a simple PID controller, computes the appropriate control action to correct the error.

3.3 Experimental results

Several experiments were performed under the condition of a well-defined reference scenario, to test the real-time performances of the autonomous robot in terms of speed and accuracy, using various face-detection and face-recognition algorithms. Experiments on face-detection algorithms were performed using Haar Cascades [4] and Local Binary Pattern (LBP) [5], while experiments on face-recognition were conducted on Scale-Invariant Feature Transform (SIFT) [6], Speeded Up Robust Feature (SURF) [7] and Oriented and Rotated BRIEF (ORB) [8].

From graphs in fig. 3.3, it is clear that the LBP algorithm is the best choice for realtime face-detection. Although some configurations seems not to be suitable for a real-time control of the robot, it is possible to remain below the desired control period maintaining an acceptable level of accuracy, as reported in table 3.1. Here the accuracy percentage is in terms of faces detected on the number of frames in which a face is present.

A second set of experiments evaluates real-time performances for face-recognition algorithms. Results for this experiments are reported in fig. 3.4.

SIFT and SURF performances are greatly above the upper bound for the control period, and are clearly unsuitable in the proposed reference scenario. The recently proposed ORB algorithm provides better results, with an execution time compatible with the desired control period for the robot.



Andrea Claudi Robots as Cyber-Physical Systems: new perspectives for the design of autonomous robots



(c) LBP results for the first experiment. (d) LBP results for the second experiment.

Figure 3.3: Experimental results.

4 Conclusions

This work focused on modelling and developing an intelligent autonomous robot controlled by visual information. This robot was conceived to be used in video-surveillance systems, in particular to patrol along perimeters of sensitive areas, to move from room to room in a building, or working as a semi-mobile camera in sensitive sites like stations, airport, etc. The goal of the robot is to observe the surrounding environment, identify human faces and try to recognize people. If the identified person is present in a database, the robot start to follow his face, trying to keep it in the centre of the frame. This system could be adapted to serve in an ambient assisted living context, providing it some mean to interact with the assisted person and with the domotic system.

Hybrid techniques are used to model the system. Hardware architecture of the robot is based on requirements from the modelling phase. Appropriate hardware to execute



Algorithm	Window (px)	Scale	Accuracy (%)
haar_frontalface_default	90x90	1.3	69.09
haar_frontalface_alt	90x90	1.3	71.79
haar_frontalface_alt2	90x90	1.3	73,33
haar_frontalface_alt_tree	90x90	1.3	53.66
LBP_frontalface	100x100	1.3	65.60
LBP_frontalface	100x100	1.2	70.44
LBP_frontalface	90x90	1.3	66.76
LBP_frontalface	90x90	1.2	71.74
LBP_frontalface	80x80	1.3	67.05
LBP_frontalface	80x80	1.2	72.49
LBP_frontalface	70x70	1.3	69.64
LBP_frontalface	70x70	1.2	73.62

Table 3.1: Accuracy of face-detection algorithms



ORB/0.5 ORB/0.75 SURF/0.5SURF/0.75 SIFT/0.5 SIFT/0.75

Figure 3.4: Face-recognition algorithm time performances



complex Artificial Intelligence methods is chosen, to achieve a good trade-off between reactivity and accuracy. An Android smartphone is used to capture and process a video stream of the monitored environment, while an embedded real-time operating system is used to control robot's movements. The control system is implemented as a task executing in the real-time operating system, and using a simple PID algorithm to control the movements of the robot. Several experiments were performed to test the real-time performances of the autonomous robot in terms of speed and accuracy, and to identify the best algorithm for computer vision to be used in the context of the deployed system. A combination of LBP [5] (for face detection) and ORB [8] turns out to be the more reliable in terms of accuracy and timeliness [9, 10].

As future works, we plan to build a new prototype, which will make use of advanced and smart sensors, and of a linear actuator to adapt the height of the robot to the height of the person to be followed. Also, the robot will provide a series of led to show the actual operation mode. Additionally, some works is already done to integrate a bluetooth module into the control system, to communicate with the smartphone without the need to be connected with it. The addition of microphones and speakers to the system will potentially allow a vocal communication with people around the robot; to make this possible we are working on an open source implementation of the VoiceXML W3C protocol [11] which, in conjunction with a Text-To-Speech (TTS) software and with an Automatic Speech Recognition (ASR), will give the robot a limited ability to dialogue with humans about simple topics. Other classical AI applications in robotics, which will be interesting to test on the robot, will be path planning in a known environment and mapping an unknown environment.

References

- [1] E. A. Lee and S. A. Seshia, *Introduction to Embedded Systems A Cyber-Physical Systems Approach*, 2011. [Online]. Available: http://leeseshia.org
- [2] A. F. Dragoni and D. Regno, "Virtual caregiver assistant for ambient assisted living," in Proceedings of 6th Euro-Mediterranean Conference on Medical Informatics and Telemedicine, 2010.
- [3] P. Gai, E. Bini, G. Lipari, M. Di Natale, and L. Abeni, "Architecture for a Portable Open Source Real-Time Kernel Environment," in *Proc. of the 2nd Real-Time Linux* Workshop (RTLWS), 2000.
- [4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1. IEEE, 2001, pp. 511–518.
- [5] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.



- [6] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition*, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 2. IEEE, 2004, pp. II-506.
- [7] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Proc.* of the European Conference on Computer Vision (ECCV), pp. 404–417, 2006.
- [8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in Proc. of the International Conference on Computer Vision (ICCV), 2011, pp. 2564–2571.
- [9] S. Saraceni, A. Claudi, and A. F. Dragoni, "An active monitoring system for real-time face-tracking based on mobile sensors," in *Proc. of ELMAR-2012*, Zadar, Croatia, 2012, pp. 53–56.
- [10] A. Claudi, F. Di Benedetto, G. Dolcini, L. Palazzo, and A. F. Dragoni, "Marvin: Mobile autonomous robot for video surveillance networks," in *Computer Modeling* and Simulation (EMS), 2012 Sixth UKSim/AMSS European Symposium on, nov. 2012, pp. 21–26.
- [11] J. Larson, "VoiceXML and the W3C speech interface framework," *IEEE Multimedia*, vol. 10, no. 4, pp. 91–93, Oct. 2003.

