Visual-based interface for robots and intelligent environments

Introduction and summary of the project

The introduction of computers has had a tremendous impact on task solving methods in all areas. However, although they arrive to help us in our work, some people claim that computers have reduced work effectiveness. For those people, computers are not truly user-friendly because users must engage in an explicit, directed dialogue with the computer, instead of using more natural ways of interaction. Mice and graphical displays are everywhere, showing that there is a frontier between the computer world and the real one. The existence of these two different worlds has two inherent problems: it is necessary to spend a lot of time transferring information between them and it is also compulsory to learn a new interaction protocol to establish the human-computer communication.

The effort to share the information could be reduced by better integrating the virtual world of the computer with the real world of the user and vice versa. To achieve this integration, several dialogue techniques were developed in the past and are now in use. Command languages, function keys, menu selections, icons or windows are some of the traditional dialogue techniques. However, in all these traditional interaction techniques, the user cannot combine both worlds within the same interface space. Thus, by example, they do not incorporate the enormous potential for interaction of the human hands. If traditional interfaces are justified for those scenarios where people are obliged to work sitting in front of a desk, there are other places where these same interfaces can lose their efficiency. **Particularly, they are not the friendliest interfaces in the field of mobile robotics or intelligent environments, where the user usually stands in front of the robot or a whiteboard trying to run augmented commands using natural gestures.** Besides, in these situations, the computer-based system is usually equipped with visual sensors which could permit it to capture interesting interaction aspects like the user's **hand gestures**, or the **emotional state that his/her face or his/her global gestural pose is showing**. A natural interface could use this information to ease the communication with the user. This particular goal is actually driving an important part of the research efforts of two research groups at University of Coimbra and University of Málaga. Thus, both research teams are very interested in generating in-depth knowledge of human-oriented interfaces which will perceive these visual aspects to develop a more natural human-computer interaction. The aim of this project is to provide us with a dedicated forum to share information and experiences.

Detailed methodology and work programme of both research teams

We believe that next generation of interfaces for intelligent environments and robots require low-cost, easily configurable visual-based human tracking systems to provide perceptual awareness of users. Most approaches to vision-based perceptual interfaces have so far relied mainly on a single visual modality to detect, track or recognize human users. Our view is that these new interfaces must combine several visual modalities. If the system must simultaneously run several modules, the visual perception could not be understood as a passive reconstructive activity, but as one intimately related to action. Animate vision is based on two principles:

- ✓ *Active vision*. Vision must operate continuously, furnishing results within a fixed delay. Rather than trying to obtain a maximum of information from any one image, the camera is an active sensor that interacts with the environment, sometimes even moving around it, picking information selectively.
- ✓ *Task-oriented techniques*. Since visual sensing is performed with limited resources, visual strategies must be planned so that only necessary information will be obtained.

The generation of the appropriate visual strategy entails knowing what information to extract, where to get it, and how to get it. This is facilitated by the knowledge of the task.

Thus, the visual system interacts deliberately with the environment by controlling the gaze and moving the focus of attention. In animate vision, the data acquisition and the information extraction processes are closely related and they depend on the current task. Therefore, the perception process becomes an active mechanism that extracts the most relevant information from the huge amount of input data depending on the application. This selection or pre-attention mechanism allows to efficiently exploit the available computational resources either by dedicating all of them to a specific perceptual task or by sharing them among a small set of tasks. Therefore, a central outcome of this project is to encourage both research groups to share their advances in low-level image processing and on the development of human-oriented attention mechanisms, which will be capable of handling dynamic environments, detecting human faces and hands and tracking objects of interest which the human may be interacting.

Other very important task that the visual perception system must solve is the capturing of human motion. It will be especially interesting to track the human hands. In addition to **facial expressions**, **non-verbal communication is often conveyed through gestures and body movement.** Human hands are highly deformable articulated objects with many degrees of freedom and can, through different postures and motions, be used to express information. In this project, people of both groups could analyse and discuss about different vision-based approaches to human motion capture. As the application of these approaches is for real-time systems, the analysis must be especially focus in fast solutions, based on hierarchical tracking of regions (hands, head, or torso) and edges (human silhouette).

The project schedule is divided **into three main modules**, which are grouped into **low level visual perception**, **high level visual perception** and **application framework**. These are research areas on which both teams are currently working. Although the main goals to achieve in the different sections are presented below, it must be noted that the main objective of this project is to establish a shared forum of discussion. Therefore, the timetable of the project schedule will be organised around meetings and researcher exchanges. A set of deliverables, associated to each of the different meetings, will be suggested to be the outcome of the different modules. That is, the first deliverable will be provided after the first meeting, and it will consist of several reports, which will be classified according to the research area they are related to.

A. Low level image processing

This module includes all early image processing cues such as region segmentation and tracking. Mainly, these visual tasks require high processing speed. Tasks which are included in this module are:

- Perceptual grouping for mid-level segmentation using concepts like colour contrast, region shape, skin colour and depth.
- An attentional mechanism for human activity tracking
- Tracking of regions of interest.
- Tracking and fixation camera movements

The team, which will work in this module, will consist of Dr. Juan A. Rodríguez, Dr. Rebeca Marfil and Ricardo Vázquez-Martín. All from Dep. Tecnología Electrónica at University of Málaga.

Outcomes: In the first meetings, researchers of both teams will analyse the image features which must be involved in a perception-based segmentation approach. It would be taken into account that, at this level, these features must allow detection and tracking of the human activity.

B. High level image processing

High level image processing has two main objectives. The first is related to **estimation of the human movements.** The underlying question is: "How can a robot classify a perceived action by observing the movements of an actor?" The curves of the movements have a **similar trace** but **distinct qualitative** aspects. Human observers seem to arrive to right conclusion by perceiving these qualitative aspects. Our solution is to use descriptors taken from Laban Movement Analysis (LMA) to describe the expressive content of the action. The second is the recognition of objects in the environment, which can be necessary to understand certain gestures or commands. This set of objects can be defined a priori.

The different tasks to carry out are described below:

- Model-based human motion capture
- Analysis of learning-based approaches to human motion capture
- PCA-based object recognition
- Parameterisation of human movements
- Acquisition, registration of human movement trajectories by tracking devices and making them available in a database for studies on human motion analysis. The database will be available by internet for the community
- Theoretical framework for human motion analysis by Laban Movement Analysis.
- Translation of the framework of Laban Movement Analysis into the domain of probabilistic inference (Bayesian models)
- Evaluation of the performance of classification for expressive actions, gestures and grasping actions

Dr. Luis Molina, Juan P. Bandera, Álvaro Carmona, Rocío López, Dr. Jorge Dias, Joerg Rett, Paulo Menezes, Diego Faria and Carlos Simplicio will unfold their activities in this research area. Dr. Antonio Bandera, Dr. Juan A. Rodríguez and Dr. Rebeca Marfil develop part of their activities in this area.

Outcomes: This is the module where both research groups are showing a major interest. Besides, it is the research area where they can share more experiences. Therefore, this issue would be addressed in all meetings. We expect a theoretical framework for human trajectories support by a database of human movements and the application and evaluation of Laban Movement Analysis as a high-level descriptor for human movements.

C. Application framework

Two different application frameworks are taken into account in this project: **mobile robotics** and **intelligent environments**. The emerging field of Human-Robot Interaction (HRI) represents an interdisciplinary effort that addresses the need to integrate social informatics, human factors, cognitive science and usability concepts into the design and development of robotic technology. Recently a number of projects are aiming to provide also a systematic assessment of the ethical issues involved in the Robotics R&D. Recent findings in Roboethics are presented through the official Roboethics website **http://www.roboethics.org.** The aim is the development of a **social robot**, which can be defined as an embodied agent that is part of a heterogeneous society of robots or humans.

As the physical capabilities of robot improve, the reality of using them in everyday environments such as offices, factories, homes and hospitals is quickly becoming more feasible. In these environments, the socially interactive robot is able to recognize each other and engage in social interactions. It must be capable of communicating and interacting **with humans and other social robots**, understand and even **relate to humans, in a personal way**. To interact with humans, social robots must simultaneously perceive a great variety of natural social cues from visual and auditory channels, and to deliver social signals.

On the other hand, intelligent environments are designed to facilitate computer use by making computers aware of humans and enabling voice and gesture commands. Smart houses are considered a good alternative for the independent life of older people and people with disabilities. Numerous intelligent devices, embedded into the home environment, can provide the resident with both movement assistance and 24-h health monitoring. Modern home-installed systems tend to be not only physically versatile in functionality but also emotionally human-friendly, i.e., they may be able to perform their functions without disturbing the user and without causing him/her any pain, inconvenience, or movement restriction, instead possibly providing him/her with comfort and pleasure. We think intelligent environments will spread to various application areas such as video surveillance, health care, or city environments. We think about interfaces that consist of ceiling-mounted video cameras that detect the hand of the user. The user controls the TV, VCR, and curtains by pointing at them. Some output signals confirm the user's selection. After choosing the desired device, the user makes a desired action by predefined hand gestures. A voice-generated message confirms the recognized gesture command before its execution.

Outcomes: Both groups will first discuss their own architectures for future smart environments based on the experience and results of their systems. Related but as topic of its own the groups will discuss and formulate a statement addressing the ethical issues of man-machine relationships. Besides, both research groups have proposed and tested in the last years visual-based modules which have been applied to the development of a social robot. The aim is to continue this research line and to employ similar concepts in the intelligent environment's framework. The analysis and design of new interfaces to use in the described scenarios will be the responsability of Dr. Antonio Bandera, Prof. Francisco Sandoval, Álvaro Carmona, Pedro Núñez, José M. Pérez-Lorenzo, Dr. Jorge Dias, Joerg Rett, Paulo Menezes, Diego Faria and Carlos Simplicio.

d. Meetings

Four meetings will be organised to share information and experiences. In these meetings, the main advances of both groups will be presented and discussed. The tentative dates for these meetings will be:

- February-2008
    - Place: Coimbra
- September-2008
    - Place: Málaga
- June-2009
    - Place: Coimbra
- December-2009
    - Place: Málaga

e. Researcher exchanges

It will be also desirable to encourage researcher exchange between both groups. A tentative timetable of these stays could be:

- May-June-July 2008: Rocío López at Coimbra
- May-June-July 2009: Álvaro Carmona at Coimbra
- May-June-July 2008: Jörg Rett and Diego Faria at Malaga
- May-June-July 2009: Carlos Simplicio and Jörg Rett at Malaga

Project schedule

ABR	Antonio Bandera Rubio	LMT	Luis Molina Tanco	JR	Joerg Rett
FSH	Francisco Sandoval Hernández	PNT	Pedro M. Núñez Trujillo	PM	Paulo Menezes
JARF	Juan Antonio Rodríguez Fernández	RVM	Ricardo Vázquez Martín	DF	Diego Faria
JBR	Juan P. Bandera Rubio	RLG	Rocío López Garrido	JD	Jorge Dias
ACC	Álvaro Carmona Calvo	JPL	José M. Pérez Lorenzo	CS	Carlos Simplicio
RMR	Rebeca Marfil Robles				

Tasks	Team and coordinators	Expected outcomes	First year	Second year
A Low level image processing	JARF, RMR, RVM	A visual system for autonomously detecting and tracking detected ROIs		
A1. Segmentation based on colour contrast, skin colour and depth				
A.2. Attentional mechanism				
A.3. Regions of interest (ROIs) tracking				
A.4. Tracking and fixation movements				
b.4. High level image processing	LMT, JBR, RLG, ACC, JD, PM, DF, JR, CS	Development of a human motion capture system and a object recognition system. Human Move ements Data Base. Laban Analysis systems		
B.1. Person detection and model initialization			RLG at UC	
B.2. Teacher motion racking				
B.3. Object recognition				
B.4. Data Base of Human Movements			JR,DF at UMA	
B.5. Laban Analysis of Human Movements			JR,DF at UMA	
C.2. Application framework	ABR, FSH, PNT, JPL, ACC, JD, PM, DF, JR, CS	Definition of the application frameworks. Analysis of the ethical considerations about the human-machine interaction.		
C.1. Control architecture to support natural interfaces				ACC at UC
C.2. Robotics for Intelligent Environments				JR,CS at UMA
C.3. Smart House & Robot Ethics				JR,CS at UMA

Project goals

This project will permit to establish an open forum of discussion between two research teams. Four different meetings will be celebrated at Málaga and Coimbra, and different researcher stays have been programmed.

Thus, the main objective of this project will be to establish a scenario where two research groups from University of Coimbra and University of Málaga can discuss and analyse the development of a visual-based interface for human-computer interaction. A goal for interface developers should be the creation of interfaces that are "humane". Raskin defines the humane interface: "An interface is humane if it is responsive to human needs and considerate of human frailties" (J. Raskin, "The Humane Interface - New Directions for Designing Interactive Systems," ACM Press, New York, 2000). He also indicates that in order to develop a humane interface the designers need "an understanding of the relevant information on how both humans and machines operate". Particularly, this interface will be developed to permit a human to interact with a social robot or with an intelligent environment. In both cases, the internal computer should be invisible, not demanding human user adaptation, but offering the benefits of its data-processing power. These scenarios should not require a new way of working: it should simply augment current working modes (M. Weiser, "The World Is Not a Desktop," Interactions, vol. 1, no. 1, 1994, pp. 7–8).

Particular topics of discussion include:

- Development of a perceptual-grouping segmentation algorithm
- Testing of an attention mechanism for human head and hands detection and tracking
- Development of visual modules for gesture and face recognition
- Testing of a model-based human motion capture
- Analysis of learning-based approaches to human motion capture
- PCA-based object recognition
- Acquisition, registration of human movement trajectories by tracking devices and make them available in a data-base for studies on human motion analysis. The data-base will be available by internet for the community
- Theoretical framework for human motion analysis by Laban Movement Analysis.
- Translation of the framework of Laban Movement Analysis into the domain of probabilistic inference (Bayesian models)
- Evaluation of the performance of classification for expressive actions, gestures and grasping actions
- Formulation of a statement addressing the ethical issues of man-machine relationships