Jornada de Seguimiento de Proyectos, 2010
Programa Nacional de Tecnologías Informáticas

PINES: Prototipos de Interacción Natural mediante Interfaces *Enactivas* basadas en entradas visuales TIN2007-67896

Javier Varona *
Universitat de les Illes Balears (UIB)

Abstract

Multimodal human-computer interfaces integrate input and output modalities, such as audio, speech, vision, gesture and touch. Enactive interfaces are related to an advance of the multimodal paradigm, a fundamental "natural interaction" concept which is not yet fully exploited in most of the existing human-computer interfaces. Enactive knowledge represents the kind of knowledge "learning by doing", based on the experience of perceptual responses to action, acquired by demonstration and sharpened by practice. Even if up to now HCI technologies have not fully exploited the potential of enactive knowledge, recent technological advances have created the possibility to significantly enhance naturally the interface perception by means of visual inputs, the so called Vision-Based Interaction (VBI). Computer Vision technology applied to the human-computer interface has notable success to date. However, its usability and technological transference stills more research work. In this project, our goal is the design and development of real-time robust vision algorithms for user body and face motion recognition for building enactive interfaces. Besides, we plan to make a set of natural interaction prototypes (NIPs) for testing the algorithm's usability in real cases. Therefore, we could test the possible applications by the final users that would use them.
Keywords: Advanced Interfaces, Computer Vision, Human-Computer Interaction.

1 Objectives

As stated by Bruner [5], the traditional interaction with the information mediated by a computer is mostly based on symbolic or iconic knowledge, and not on enactive knowledge. Enactive Knowledge is information gained through perception-action interaction in the environment. In many aspects the Enactive Knowledge is more natural than the other forms both in terms of the learning process and in the way it is applied in the world. Enactive Interfaces are new types of Human-Computer Interface that allow to express and transmit the Enactive knowledge by integrating different sensory aspects. The driving concept of Enactive Interfaces is then the

*Email: xavi.varona@uib.es

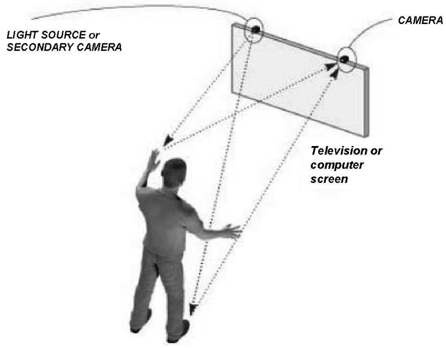

Figure 1: Proposed natural interaction prototypes.

fundamental role of action for storing and acquiring knowledge (action driven interfaces). Enactive Interfaces are then capable of conveying and understanding gestures of the user, in order to provide an adequate response in perceptual terms.

In order to recover the user's body motion, computer vision algorithms are the methods used in current state-of-the-art projects in gestural interfaces [6]. Mainly, for its capabilities to sense without touching or using any attached device. Vision-based interaction (VBI) provides a wider and more expressive range of input capabilities by using computer vision techniques to process images from one or more cameras in real-time, in order to reliably estimate relevant visual information about the user. Specifically, we are interested in gestural interfaces where the gestures do not require the user to touch or handle directly the interface, i.e., interfaces based on the recognition of user's body motion. Figure 1 shows how the proposed interface works.

However, interactive applications pose particular challenges [7]. The response time should be very fast. Users should sense no appreciable delay between when they make a gesture or motion and when the computer responds. Computer vision algorithms should be reliable, work for different people, and work against unpredictable backgrounds. In addition, all interaction interfaces should go through a usability evaluation in order to study how usable the device actually is. Hornbaek identifies in his work [9] those papers that measure usability and how they measure it: he does a review of papers published in core HCI journals or proceedings between 1999 and 2002 where he identifies 180 articles that report the use of measures of usability. However, few of the papers analyzed by Hornbaek or in the literature review are about vision-based interfaces. Therefore there is a lack of usability evaluations and recommendations in this domain. For these reasons, we proposed the next objectives:

1. Review and develop robust and adaptive algorithms for real-time recognition of human motion. For the facial motions' particularity, the study will be divided in:
   (a) facial motions, and
   (b) body motions.

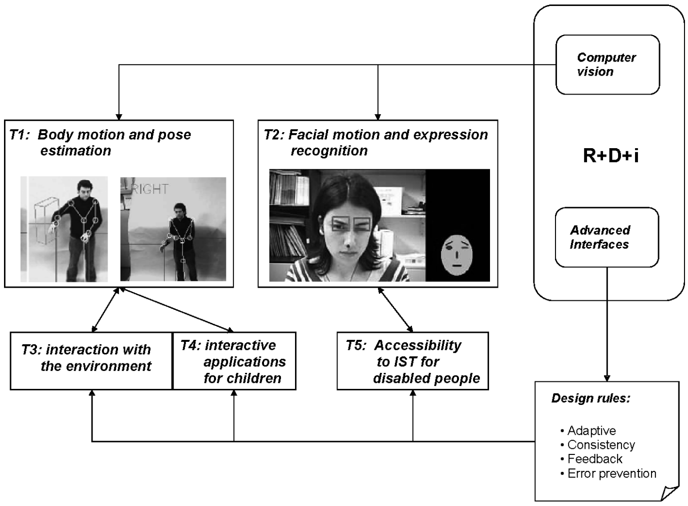

Figure 2: Project architecture.

2. To perform a set of natural interaction prototypes to validate the usability of the developed methods in real scenarios:
   (a) interaction with the environment in virtual reality applications,
   (b) interactive applications for children, and
   (c) accessibility to the information society for disabled people.

As seen in Fig. 2, the project were divided into five main tasks that reflect the objectives and clearly show the research and development relationship as prototypes rely on scientific achievements, but on the other hand, computer vision methods should be designed according to user requirements. That is why we propose a scheme that joint research and development. In that dual relationship between the computer vision methods for the creation of enactive interfaces, and the usability evaluation through prototypes, is where much of the difficulty of this project.

2 Progress

2.1 Performed activities

As stated in previous section, the project has two main objectives, which could be divided into research and development. The first objective is devoted to research on computer vision and is the project basis. For this reason, its implementation is planned throughout practically the entire project. During the first two years we have done an intensive work on this objective in order to obtain the computational methods that allow the development of the planned prototypes. In this sense, we have developed several algorithms for body and facial motion recognition, which could be used in natural interaction prototypes.

The first task, body motion recognition, has been divided into pose recovery and gesture recognition. The idea of this division arises from our previous work, which we concluded with the assumption that the gesture recognition is simplified if previously is recovered the user's pose (i.e., the 3D position of its joints). The starting point is that it is possible to recover the full posture from local movement of terminal parts of the body (called *end-effectors*) such as the hands. Our approach is based on the combination of computer vision algorithms for tracking the end-effector's [18], and an advanced method of inverse kinematics, Priority Inverse Kinematics [2], to estimate overall body posture from the known positions of the end-effectors [3, 4]. Therefore, our initial hypothesis is that from the estimated position of the user is possible to address the recognition of gestures.

From the estimated user's posture, we have developed two new gesture representations that solve the main problems of these systems. These new gesture representations are based on histograms of the joints involved in the gesture performances. The detailed explanation of a study of these representations can be found in [10]. The recognition system was supplemented by classification techniques that allow recognition despite the variability of styles between different people performing the same gesture. It should be taken into account that has been achieved that the gesture recognition system performs in real time, something essential to its application in the proposed interaction prototypes. The first complete system of gesture recognition was recently published in the *Interacting with Computers* journal [21]. In addition, we are conducting a new approach with the aim of improving the initial results. This approach is based on a feature of the PIK algorithm, which allows including priorities within the space of solutions to reduce the search space by the objective of the application. For the integration of computer vision methods within this framework, we propose the definition of image-based constraints, which prioritize the estimated poses according to their assessment in the images. This work is based on the definition of a functional, which minimizes the error between the estimated and actual pose projected onto the image [12].

In parallel, we worked on a second approach based on stochastic methods by using particle filters [19]. Specifically, it was defined a human action space, which we implemented through Bayesian Filtering, which can address the problem of pose estimation in a computationally efficient way [17]. Within this line of research we began collaborating with Dr. Courty, of the University of Bretagne Sud (France), which defined a new approach to inverse kinematics with particle filters.

From the results of the body motion recognition task, we have developed a first prototype with interactive applications for children (task 4). This prototype is installed at the *Maria Antonia Salvà* school of Palma, and it is used in the psychomotor classroom by the teachers of four and five years-old children. We are currently expecting the teachers' comments in order to improve the prototype and to perform its usability evaluation.

In the facial motion task, we based on our previous work on face detection and tracking for human-machine interfaces [13]. From the obtained results in this work we concluded that the application of the computer vision methods developed in human-computer interaction require that the facial motion recognition algorithms be fully automatic (especially the initial detection of the user), robust (including the automatic fault recovery) and in real-time. During the project and by an agreement of collaboration in the project SINA (Advanced Natural Interaction Systems, agreement with the Government of the Balearic Islands in the *Plan Avanza* framework) we have developed a set of algorithms that meet all these objectives for facial motion recognition [20]. In addition, we have completed the evaluation of the usability of the accessibility to the information society for disabled people prototype. This prototype is a hands-free interface based on computer vision techniques for motion impaired users. This interface does not require the user to use his upper body limbs as other input devices demand. Therefore, users with motion difficulties can take advantage of this kind of system when standard devices are not suitable for them [14]. In order to evaluate the prototype's usability, the approach that we have taken is to integrate usability evaluation at relevant points of the software development. We have presented a development process which follows a prototyping system with multiple evaluations with end-users [16]. These evaluations involve users' observation and objective tests. We showed how their feed-back has helped to improve greatly the quality of the developed software and we want to contribute with a possible framework to follow when implementing vision-based interfaces.

2.2 Difficulties found

In the comments of the project's reviewers about the financing proposal, they noted the difficulty of the tasks of prototypes' usability evaluation due to the inexperience of the research team on this issue. Therefore, the funding proposal rejected the part of personnel devoted to the prototypes' development. Regarding these comments, noting that we have established contacts and collaborations with usability experts to acquire the skills needed to perform the task correctly. In fact, as we sated before, we have evaluated one of the prototypes from the point of view of Usability Engineering and People-Centered Design. The biggest difficulty is that the lack of resources for contracting personnel requires overstrain of the research team in development works. As a solution, we seek partnerships in technology transfer projects to find the necessary funding to complement the development of prototypes.

3 Results indicators

3.1 Goal achievement

Currently, we consider that three of the planned five tasks are practically completed. The other two tasks are started and they are planned to finish in time. However, the prototype for interaction with the environment in virtual reality applications has been reoriented to a prototype for Telerehabilitation. The main reason is that we have started collaboration with a public Foundation in order to develop interaction methods based on body motion recognition and we could obtain extra founds in order to contract personnel for helping in development tasks. In addition, we consider that the objective of this prototype coincides with the objective of this task, which is to evaluate the usability of the computer vision methods for body motion recognition.

3.2 Scientific production and relevance of results

The PINES project has produced 7 journal articles with impact index [1, 8, 12, 17, 19, 20, 21] and 2 PhD theses [11, 15]. Note that this includes two articles into the highly relevant Pattern Recognition and Interacting with Computers journals. On the other hand, the scientific production also includes presentations in international conferences such as ASSETS, AMDO, HSI and in the Gesture Workshop. In addition, 2 papers have been submitted to indexed journals.

3.3 Utility of results and their role in economic and social development

The achieved results are directly applicable to the main research areas of our group and their associated technology transfer projects. Our most important active project of this kind is SINA (2007-2010): a technology project within the *Plan Avanza* by means of an agreement with the Balearic Islands government. SINA focuses on the creation and support of hands-free interfaces to access the computer by disabled people (`http://sina.uib.es`).Recently, we have signed a contract with the *Fundación iBit* (`http://www.ibit.org`)in order to make the technological transference of the body motion recognition results for Telerehabilitation applications

3.4 Human resource development

As we stated before, 2 group members finished their PhD thesis within PINES.

3.5 Collaborations with other research groups

Our group maintains active collaborations with national and international research groups. Specifically, we collaborate with the next national groups: the GIGA group at the Universidad de Zaragoza (Prof. Francisco Serón), the ISE Laboratory at the Computer Vision Center of the Universitat Autònoma de Barcelona (Dr. J. Gonzàlez), and the 4all-Lab at the Universitat Politècnica de Catalunya (Dr. Pere Ponsa). We also collaborate with international research groups by means of internships of the members of our group in the Machine Vision Unit, University of Edimburgh (Prof. Robert Fisher), the Computer Graphics and Multimedia Lab at Adetti (Prof. Miguel Sales Dias), and the SAMSARA group (VALORIA laboratory) at the University of Bretagne-Sud (Dr. Nicolas Courty). In addition, we maintain the collaboration with the VRLab at the Swiss Federal Institute of Technology (Prof. R. Boulic); specifically by means of our acceptation as member of interest in the ENACTIVE Network of Excellence funded by the European Union (see http://www.enactive.org).

We have also to remark the strong relation with project ITADA (TIN2007-67993) because the majority of the researchers of both projects belong to the *Unidad de Gráficos, Visión e Inteligencia Artificial* at the Universitat de les Illes Balears. The present project represents the group's research line devoted to vision-based interfaces for Human-Computer Interaction.

3.6 Project coordination, development and management

We are using a web-based project management system, including all results and demos in the website `http://dmi.uib.es/~ugiv/pines/`,SVN repositories and e-mail lists. However, no strict rules are being applied with respect to document formats, and we simply comply with the acknowledgement normative on project publications and results.

References

[1] Arellano, D., Varona, J., and Perales, F. J. *Generation and visualization of emotional states in virtual characters.* Comput. Animat. Virtual Worlds 19, 3-4 (Sep. 2008), 259-270.

[2] Baerlocher, P. and Boulic, R. *An inverse kinematics architecture enforcing an arbitrary number of strict priority levels.* Vis. Comput. 20, 6 (Aug. 2004), 402-417.

[3] Boulic, R., Varona, J., Unzueta, L., Peinado, M., Suescun, A., and Perales, F.J. *Real-Time IK Body Movement Recovery from Partial Vision Input.* In Proc. of the 2nd International Conference on Enactive Interfaces (ENACTIVE05), Genova, Italy, November 2005. **1st ENACTIVE Best Collaborative Paper Award.**

[4] Boulic, R., Varona, J., Unzueta, L., Peinado, M., Suescun, A., and Perales, F.J. *Evaluation of on-line analytic and numeric inverse kinematics approaches driven by partial vision input.* Virtual Reality 10, 1 (May. 2006), 48-61.

[5] Bruner, J. *Acción, pensamiento y lenguaje. Compilación de J. Linaza.* Madrid: Alianza Editorial. 1995.

[6] Canny, J. *The Future of Human-Computer Interaction.* Queue 4, 6 (Jul. 2006), 24-32.

[7] Freeman, W. T., Anderson, D. B., Beardsley, P. A., Dodge, C. N., Roth, M., Weissman, C. D., Yerazunis, W. S., Kage, H., Kyuma, K., Miyake, Y., and Tanaka, K. *Computer Vision for Interactive Computer Graphics.* IEEE Comput. Graph. Appl. 18, 3 (May. 1998), 42-53.

[8] Gonzàlez, J., Rowe, D., Varona, J., and Roca, F.X. *Understanding dynamic scenes based on human sequence evaluation.* Image Vision Comput. 27, 10 (Sep. 2009), 1433-1444.

[9] Hornbaek, K. *Current practice in measuring usability: Challenges to usability studies and research.* International Journal of Human-Computer Studies, 64(2) (Feb. 2006), 79-102.

[10] Jaume-i-Capó, A., Varona, J., and Perales, F. J. *Representation of Human Postures for Vision-Based Gesture Recognition in Real-Time.* In Gesture-Based Human-Computer interaction and Simulation: 7th international Gesture Workshop, GW 2007, Lisbon, Portugal, May 23-25, 2007, Revised Selected Papers, M. Sales Dias, S. Gibet, M. M. Wanderley, and R. Bastos, Eds. Lecture Notes In Artificial Intelligence, vol. 5085. Springer-Verlag, Berlin, Heidelberg, 102-107.

[11] Jaume-i-Capó, A. *Reconocimiento visual del movimiento humano en tiempo real para la interacción natural persona-ordenador.* PhD thesis, Universitat de les Illes Balears (UIB), July 2009.

[12] Jaume-i-Capó, A., Varona, J., González-Hidalgo, M., and Perales, F. J. *Adding Image Constraints to Inverse Kinematics for Human Motion Capture.* EURASIP Journal on Advances in Signal Processing, vol. 2010, Article ID 142354, 13 pages.

[13] Manresa-Yee, C., Varona, J., and Perales, F.J. *Towards hands-free interfaces based on real-time robust facial gesture recognition.* Lecture Notes in Computer Science, vol. 4069. Springer-Verlag, Berlin, Heidelberg, 2006, 504-513.

[14] Manresa-Yee, C., Varona, J., Perales, F. J., Negre, F., and Muntaner, J. J. *Experiences using a hands-free interface.* In Proceedings of the 10th international ACM SIGACCESS Conference on Computers and Accessibility (Halifax, Nova Scotia, Canada, October 13 - 15, 2008). Assets '08. ACM, New York, NY, 261-262.

[15] Manresa-Yee, C. *Advanced and natural interaction system for motion impaired users.* PhD thesis, Universitat de les Illes Balears (UIB), July 2009.

[16] Manresa-Yee, C., Ponsa, P., Varona, J., Perales, F. J. *Improving the usability of a hands-free vision-based interface.* Interacting with Computers (in review), 2010.

[17] Rius, I., Gonzàlez, J., Varona, J., and Roca, F.X. *Action-specific motion prior for efficient Bayesian 3D human body tracking.* Pattern Recognition 42, 11 (Nov. 2009), 2907-2921.

[18] Varona, J., Buades, J.M., and Perales, F. J. 2005. *Hands and face tracking for VR applications.* Computers & Graphics. 29, 2 (Apr. 2005), 179-187.

[19] Varona, J., Gonzàlez, J., Rius, I., and Villanueva, J. J. *Importance of detection for video surveillance applications.* Opt. Eng. 47, 087201 (Aug. 2008).

[20] Varona, J., Manresa-Yee, C., and Perales, F. J. *Hands-free vision-based interface for computer accessibility.* J. Netw. Comput. Appl. 31, 4 (Nov. 2008), 357-374.

[21] Varona, J., Jaume-i-Capó, A., Gonzàlez, J., and Perales, F. J. *Toward natural interaction through visual recognition of body gestures in real-time.* Interact. Comput. 21, 1-2 (Jan. 2009), 3-10.