

# Computer Vision interaction for Virtual Reality

Homero V. Ríos Figueroa and Joaquín Peña Acevedo

Laboratorio Nacional de Informática Avanzada, A.C.  
Rébsamen 80, c.p. 91090, Xalapa, Veracruz, México  
Email: [hrios,jpena]@xalapa.lania.mx

**Abstract.** As virtual reality evolves towards more natural interfaces, new contact less interaction based on gesture recognition is unfolding. This interaction is supported on geometric, dynamic and cognitive modelling of gestures. As well as other branches of artificial intelligence, computer vision plays an important role in this modelling.

The purpose of this paper is to describe how computer vision is helping to develop virtual reality and present some interfaces developed in our laboratory.

**Keywords:** computer vision, gesture recognition, virtual reality, human-computer interaction.

# Computer Vision interaction for Virtual Reality

## 1 Introduction

During 1945, Vannevar Bush conceived the use of computers beyond calculation and thought about them as a fundamental tool for transforming human thought and creative activity [2]. He anticipated the use of computers for multimedia processing.

Since Bush's time many technological breakthroughs have occurred, but computers are still limited in their multimedia understanding [18]. This means that we have increased the effective bandwidth of information from computers to humans, by sending audio, images, audio, graphics, haptic data, but the same rate of improvement has not happened in computer understanding. Most computers still receive input from low bandwidth devices like keyboards or mouse. Only few interfaces are able to understand application related domains of audio, visual or haptic information [7].

Several researchers have identified this unbalance and are working on more intuitive interfaces like virtual reality, speech recognition, image understanding and multimodal interfaces [3, 18]. In the rest of this paper we will concentrate on image understanding techniques which are relevant for virtual reality.

The main conceptual components of a virtual reality system are: a) *immersion*, the ability to experience a 3D world as a reality, b) *viewpoint*, refers to the point of observation of the user, c) *navigation*, which allows to change viewpoint and d) *manipulation*, the capacity to interact and change the relative position of objects in the environment [14].

Another important component of virtual reality (VR) is *tracking*, since it helps to establish the correct position of the user in the environment, and therefore provides the appropriate visualization and interaction [36]. There are many technologies which help in tracking body part and they are based on electric, magnetic, ultrasonic, infrared, optic or image analysis principles [3]. In this paper, we are interested in this last type of tracking.

Since visual perception allows many organisms to interact successfully with their surroundings it is natural to think that computer vision (CV) might help to bring closer people and computers. This interaction would be based on the interpretation of body language, through *gesture recognition*. This new research topic is bringing together people with backgrounds on computer vision, human computer interaction, psychology, and artificial intelligence.

In the rest of this paper we will concentrate on gesture recognition as it applies to human computer interaction in general, and virtual reality in particular. Also, we will present a general methodology which emerges from related works worldwide. Finally, we will describe the main interfaces developed in our lab, and how they compare to other similar systems.

## 2 Computer vision and virtual reality

At first glance, computer vision and virtual reality do not seem related, in fact they look opposite. CV attempts to reconstruct surfaces, recognize objects and provide motion information of objects from images. That is, it goes from images to objects. In contrast, computer graphics and VR work from object models to images. So in certain sense, they are complementary [22]. However, the boundary between these areas is becoming thinner as many people realize that the solution to key problems in both areas involve a close interaction between the physics of image formation, geometric and mechanical modelling of object shape, deformation and motion [33, 34, 8]. As well as cognitive modelling of actions and behaviour to express agents or organism responses to visual stimuli [9].

Other problems that are of special interest to both fields but with different points of view are stereoscopic vision and object tracking. In the case of CV the problem of stereopsis is object reconstruction from two images of the object. For VR, the problem is how to obtain two different images of a scene composed of objects, so they are appropriate for visualization in devices like head mounted displays. The common link is that of stereoscopic visualization as achieved by the human brain [16].

The problem of object tracking is very important for both areas, because in the case of VR, helps to track body parts to provide the appropriate feedback of change in position of the user or objects in the virtual environments. Also, real time tracking is vital for proper visualization and reducing the so called “lag time” [36].

In the case of computer vision, object tracking is also very important in the context of time varying imagery, real time vision systems and robotic systems with visual capacity [8]. In addition, object tracking is very important for organism to follow predators and prey.

Recently, CV and VR have come closer as many people realize that more natural interfaces can be built by using CV to provide object tracking. That is following body parts or providing gesture recognition, without having to wear any special equipment. This has the advantage to provide more freedom of movement and makes the computer to adapt to human, rather than the other way around.

Another close interaction between CV and VR has come from what is called *augmented reality* [1]. The idea is to register in a common environment, real and virtual objects. That is, in these systems it is possible to see the real world augmented with computer generated objects. For example, a surgeon during an operation will be able to see an MRI volume superimposed with the images of the actual patient [13].

CV is also helping to animate VR specially when the environment incorporates autonomous agents. Some simulation programs of artificial life use synthetic vision as a sensor for organisms, so they can react to the presence of other beings, for instance approaching or receding [30].

## 2.1 Related works

We are interested in research works which use computer vision as a general human computer interface. The more general context in which we find this concept is in what is called *smart rooms* and *smart clothes* [23]. This idea is to have many cameras and computers in a network which are continuously analysing the images of people. These cameras can be in different places in a room, street, or they can be attached to human clothes. As a result of the analysis, people can be recognized, tracked, or communicate with computers and decisions can be made.

Other works are more specific and involve face recognition [31], emotions recognition [23], sign language understanding [32], teleoperation in virtual reality robotics environments [20], tracking of the whole human body for surveillance applications [5], hand tracking [10, 25, 15], iris tracking and recognition [41, 17, 27, 24, 40] and head tracking for general human computer interaction [26, 4, 39].

## 2.2 Framework to relate VR and CV

From the analysis of works which involve CV and human computer interaction, we have obtained a framework which helps to understand previous work, and also as a guide to develop new interfaces (Figure 1).

In computer graphics and virtual reality we have a database of 3-D models which are use for rendering and visualization. In these systems, the user navigate or interacts with the environment through a graphical user interface (GUI), or directly special hardware (data glove, HMD, etc.). In any case, the actions of the user modify the database of graphics objects and this in turn changes the visualization.

What is relatively new in VR, is telemanipulation and augmented reality where the graphic objects in the database have some correspondence with actual objects in the real world. In this case, by moving one graphics object, for instance a 3-D model of a robot arm, it is possible to move a real robot, possibly in a distant location [20]. Also, in augmented reality the correspondence and registration is more critical, since virtual and real object have to appear as sharing the same space [1]. Vision techniques have a role here in helping to provide proper registration and in maintaing the consistency and correspondence of real and virtual worlds [20, 13].

In the case of direct, contactless, human computer interaction through gesture recognition, the vision systems plays a more active role, since it allows the user to communicate with the computer, without having to wear any special hardware [23]. Since the system must have a representation of the gesture to recognize or track, we can think of it has geometric and dynamic model information that fetched by the model based vision module to guide its search through the image data. This model can be implicit in the system or explicitly represented and could be stored also in the 2-D, 3-D model database, but not necessarily.

The images are analyzed and features and external forces are extracted which drive model matching, and the vision module updates the model parameters and dynamics for the best fit [6, 33, 34, 8].

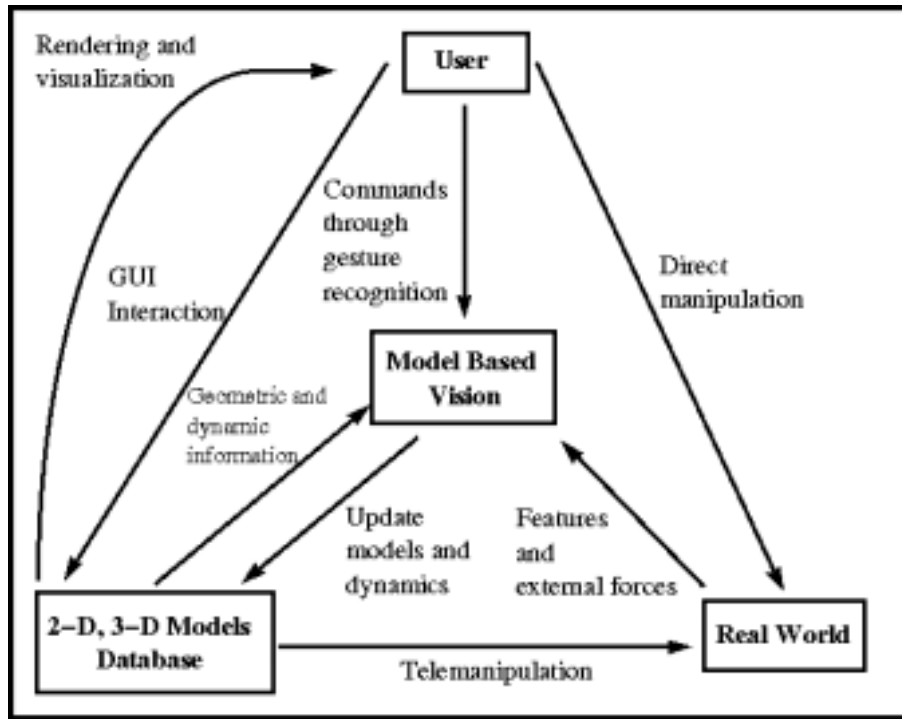


Fig. 1. Model based vision can be an important part of a virtual reality system.

### 3 Trackers developed in our laboratory

Our work has been concentrated on developing human computer interfaces through computer vision in general, and in particular for virtual reality systems (Figure 2). Our graphics interfaces are original and also some of the tracking methods that we have developed. In the next subsections we will describe the features of our iris, hand and glasses tracker, and the GUIs that have been implemented for manipulation and navigation in virtual reality systems.

#### 3.1 Iris tracker

The novelty of our iris tracker is that we do not need to attach the camera to the user's head, or provide the initial position of the head as in other works [17, 41]. We do not provide the eyes rotation, but we are able to obtain the center, and radius of a pair of circles which are fitted to each iris [27], as as is illustrated in Figure 3. Taking the mid point of the segment that joints the centers of the circles, we define a visual cursor that can be employed for activating buttons in GUIs. By using visual fixation, and if both iris remain in the same position in several frames (5 in our implementation), we interpret this behaviour as the



**Fig. 2.** Work environment

location of a point of interest, and we can associate with it the “click” of a button.

Our iris detector works by applying edge detection to each frame, followed by thinning and search of circles using a Hough transform. The most promising pair of circles are selected by applying heuristics [27]. The performance of this tracker is between 3-5 frames/second running on a Indy, SGI workstation.



**Fig. 3.** Iris detector. (a) Edges on the image. (b) The mark represents the visual cursor which is defined by the position of the eyes.

### 3.2 Hand tracker

Our approach for the hand tracker is not as powerful and general as in other works [10, 25, 15], but we provide simpler solutions for tracking a hand with extended fingers [24, 28].

Our method uses edge detection and frame differences to extract moving features with high gradients. Then, we apply a thinning algorithm, followed by a Hough transform to detect lines [12]. Finally, we extract the most meaningful segments which correspond with the fingers to obtain a global centroid from all the segments. This centroid provides a reference position for a cursor (Figure 4). The changing size of the segments provides cues about the proximity (approaching or receding) of the hand from the camera. The performance of the hand trackers is also between 3-5 frames/second on the same platform as described above.



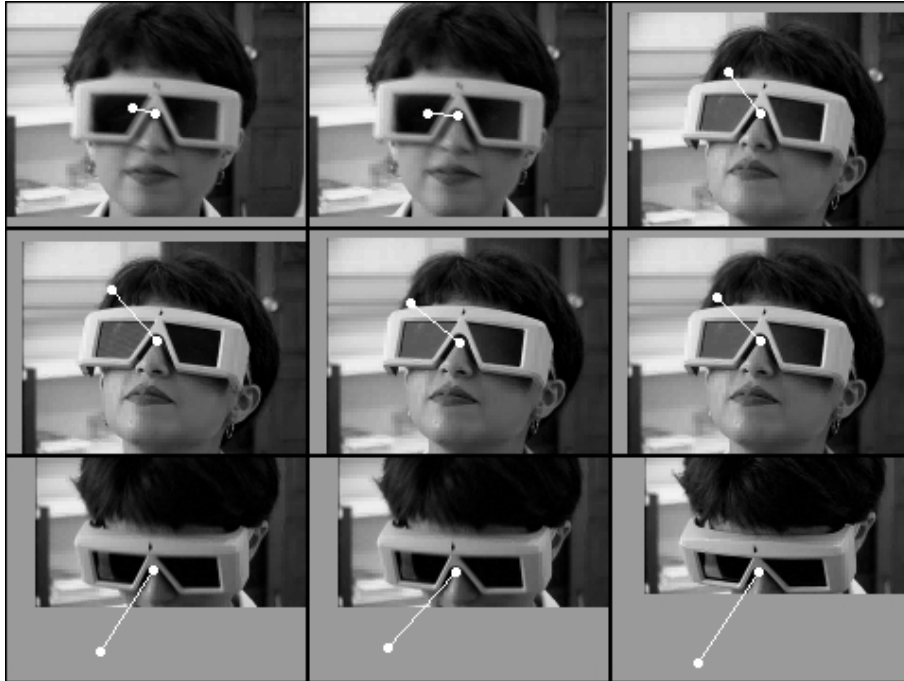
**Fig. 4.** Hand detector. This images show the way to establish the communication with the computer and the cursor defined by the fingers.

### 3.3 Glasses tracker

Some types of stereoscopic glasses do not provide head tracking, so we decided to develop one using computer vision. Since the model that we have (Crystal eyes, Stereoscopic corp.) has a rectangular frame, the idea was to fit parallelograms to the different images. Since a rectangular shape can even deform into a trapezoidal shape under perspective transformation, in this case we use an affine approximation, an under these conditions a rectangular deforms into a parallelogram. For this method to work, the perspective effect should remain small, so it is important that the glasses do not get too close to the camera during side view.

For each frame the method searches for high gradient points, and from them, a contour following algorithm is applied [12]. Only closed contours and those that satisfy a shape criteria are preserved. If two contours remain that approximate the shape of the glasses, these are fitted with an affine transformation using the

method of normalization, provided with the DIAS software [35, 37]. From the affine deformation of the rectangle into a parallelogram, it is possible to work out the 3-D rotation of the glasses in space [39]. The lines shown in Figure 5 are the projections of a spacial line segment in direction of the line of sight.



**Fig. 5.** Determination of the line of sight from the parallelogram that fits the glasses. The image point corresponding to the center of the glasses is shifted into the center of the image.

#### 4 Graphical User interfaces developed in our laboratory

The graphical user interfaces that we have developed allow the manipulation of 2-D and 3-D objects with hand or eyes movements. It is possible to select different objects, change their orientation in space, bring closer or move away the object, navigate in a 3-D environment or change the relative position of objects [24]. The first one is used to manipulate objects so that they move themselves in accordance with the movements of the user head (Figure 6).

Another graphical user interface works as an examiner of 3-D objects. It uses the cursor defined by the eyes detector to activate some buttons which serve to select, rotate, or bring near or far an object (Figure 7). The iris must be detected at least five times to activate a button.

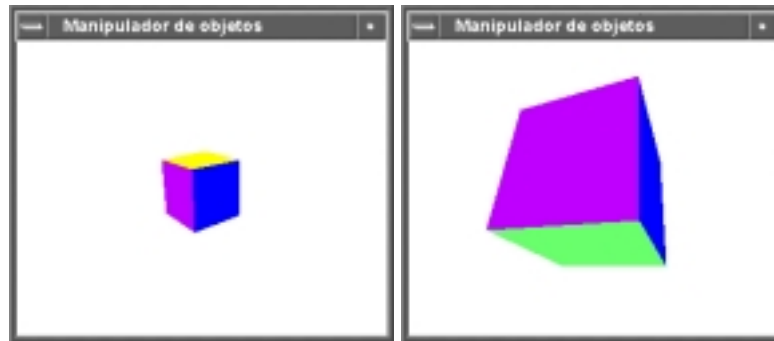


Fig. 6. The cube changes its position and orientation while the user head is moving itself

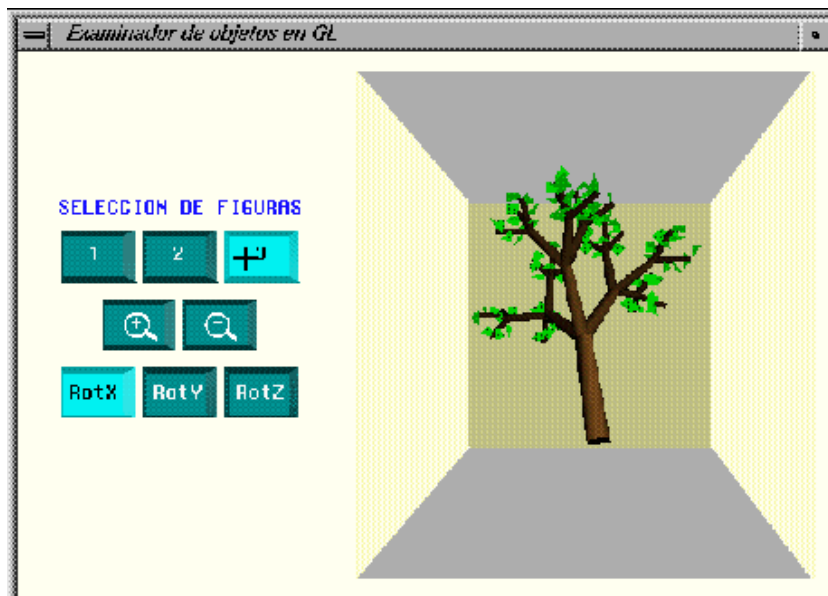


Fig. 7. Examiner of 3-D objects. The button number 3 is selected.

Figure 8 shows another interface. We can move in the virtual environment only by translations. They are in correspondence with the translations of the defined cursor with regard to its initial position.

Besides of navigation, we implemented the selection of objects contained in the virtual environment.

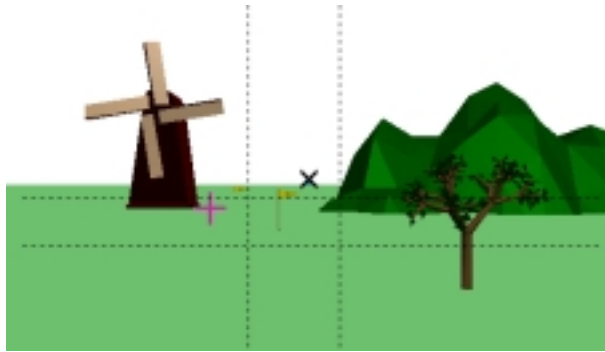


Fig. 8. Navigation in a virtual environment by using the iris tracker.

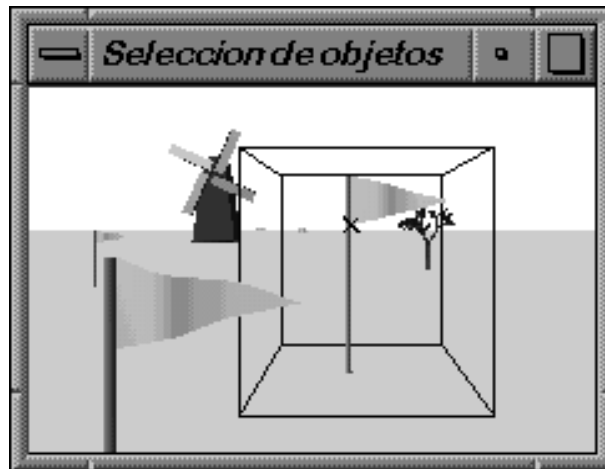


Fig. 9. The frame box around the flag indicates this object is selected.

## 5 Conclusion and future work

We have analyzed works in the literature that relate computer vision and virtual reality and as a result we have proposed a framework to understand present and future works. Also, we presented the contribution of our group in developing eyes, hand and glasses trackers and their graphical user interfaces.

Our ongoing work in collaboration with the National Autonomous University of Mexico (UNAM) and the University of Houston-Downtown (UH-D) involves the creation of cooperative virtual environments which gesture recognition interfaces and autonomous, intelligent agents, and will be reported in future academic events.

## 6 Acknowledgements

This work has been funded by the Mexican National Council of Science and Technology (CONACYT) as project ref. C098-A “Distributed simulation of gesture recognition interfaces and intelligent agents for virtual environments”.

## References

1. Aliaga, D.G. (1997). “Virtual objects in the real world”. *Communications of the ACM*, Vol. 40, No. 3, pp. 49-54.
2. Baecker, R.M. and Buxton, W.A.S. (1987). “A historical and intellectual perspective of human-computer interaction”. In *Readings in Human-Computer Interaction, a multidisciplinary approach*, Baecker, R.M. and Buxton, W.A.S. (Eds.), Morgan Kaufmann.
3. Barfield, W. and Furnes, T.A. (1995). *Virtual Environments and Advanced Interface Design*. Oxford University Press.
4. Basu, S., Essa, I. and Pentland, A. (1996). “Motion regularization for model-based head tracking”. *MIT Media Laboratory Perceptual Computing Section*, Technical Report No. 362.
5. Baumberg, A.M. and Hogg, D.C. (1993). “Learning flexible models from image sequences”. Report 93.36, Research Report Series, School of Computer Studies, University of Leeds, U.K.
6. Blake, A. and A. Yuille (Eds.). (1992). *Active Vision*, MIT Press.
7. Bolt, R.A. (1985). “Conversing with computers”. *Technology Review*, 88 (2), 35-43.
8. Brown, C.M. and Terzopoulos (Eds.) (1994). *Real-time Computer Vision*.
9. Buxton, H. (1997). “Visual interpretation and understanding”. Cognitive Science Research Paper 452, School of Cognitive and Computing Sciences, University of Sussex, U.K.
10. Cipolla, R., Okamoto, Y. and Kuno, Y. (1993). “Robust structure from motion using motion parallax”. Proceedings of the International Conference on Computer Vision, IEEE Press.
11. Faugeras, O. (1993). *Three-Dimensional computer vision, a geometric viewpoint*. MIT Press.
12. Gonzalez, R.C. and R.E. Woods. (1992). *Digital Image Processing (3rd ed.)*, Addison-Wesley.
13. Grimson, W.E.L. (1995). “Medical applications of image understanding”. *IEEE Expert, Intelligent systems and their applications*, Vol. 10, No. 5, pp. 18-28.
14. Hayward, T. (1993). *Adventures in Virtual Reality*. QUE corp.
15. Heap, T. and Samaria, F. (1995). “Real-time hand tracking and gesture recognition using smart snakes”. Technical Report, Olivetti Research Limited, U.K.
16. Hubel, D.H. (1988). *Eye, Brain and Vision*, Scientific American Library
17. Lam, K.M. and Yang, H. (1996). “Locating and extracting the eye in human face images”. *Pattern Recognition*, Vol. 29, No. 5, pp. 771-779.
18. Maybury, M.T., (Ed.). (1993). *Intelligent Multimedia Interfaces*. MIT Press.
19. Mundy, J.L. and Zisserman, A. (Eds.). (1992). *Geometric Invariance in Machine Vision*. MIT Press.
20. Natonek, E., Zimmerman, T., Fluckiger, L, “Model based vision as feedback for virtual reality robotics environments”. *Virtual Reality, Annual International Symposium '95*, IEEE

21. Ohzu, H. and Habara, K. (1996). "Behind the scenes of virtual reality: vision and motion". *Proceedings of the IEEE*, Vol. 84, No. 5, pp. 782-798.
22. Page, I. (1988). "The disputer: a dual-paradigm parallel processor for graphics and vision". In Page, I. (Ed.), *Parallel Architectures for Computer Vision*, Oxford University Press.
23. Pentland, A.P. (1996). "Smart rooms". *Scientific American*, April 1996, pp. 54-62.
24. Peña, J., Ríos, H.V., y Barradas, P. (1997). "Interacción con escenarios 3D por medio de ademanes y movimientos oculares", *Memorias del Congreso Computación Visual 97*, pp. 213-219, Facultad de Ciencias, UNAM, México.
25. Rehg, J.M. and Kanade, T. (1993). "DigitEyes: vision-based human hand tracking", School of Computer Science, Carnegie Mellon University, Technical report number: CMU-CS-93-220.
26. Rekimoto, J. (1995). "A vision-based head tracker for fish tank virtual reality" *Virtual Reality, Annual International Symposium '95*, IEEE
27. Ríos, H.V. y Barradas, P.D. (1996). "Interacción Hombre-Máquina por medio de movimientos oculares". *Memorias del V Congreso Iberoamericano de Inteligencia Artificial*, pp. 492-501.
28. Ríos, H.V., Figueroa, J.M. y Barradas, P.D. (1997). "Visión por computadora en interfaces Hombre-Máquina", *Soluciones Avanzadas*, No. 42, febrero 1997, pp. 51-56.
29. Rothe, I., Suesse, H. and Voss, K. (1996). "The method of normalization to determine invariants", *Pattern Analysis and Machine Intelligence*, Vol. 18, No. 4, pp. 366-376.
30. Winter, S. y Rudomin, I. (1997). "Synthetic computer vision for autonomous agents in distributed partitioned environments". *Memorias del Congreso Computación Visual 97*, pp.157-166, Facultad de Ciencias, Universidad Nacional Autónoma de México.
31. Schwartz, E.I. (1995). "A face of one's own". *Discover the world of Science*, vol. 16, no.2, December 1995, pp. 78-87.
32. Starner, T. and Pentland, A. (1996). "Real-time American sign language recognition from video using hidden Markov models". *MIT Media Laboratory Perceptual Computing Section*, Technical Report No. 375.
33. Terzopoulos, D., Witkin, A. & Kass, M., "Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion", *Artificial Intelligence*, Vol. 36 (1988), pp. 91 - 123.
34. Terzopoulos D. (1991). "Visual Modelling", *Proceedings of the British Machine Vision Conference*, BMVA.
35. Towersoft. (1995). *DIAS, Dialog and Programming System for Digital Image Analysis*. User reference manual, version 4.0, Towersoft, Berlin, Germany.
36. Vince, J. (1995). *Virtual Reality Systems*. Addison-Wesley.
37. Voss, K. and Suesse, H. (1997). "Invariant fitting of planar objects by primitives". *Pattern Analysis and Machine Intelligence*, Vol. 19, No. 1, pp. 80-84.
38. Voss, K. (1993). *Discrete Images, Objects and Funciones in Z<sup>n</sup>*. Algorithms and Combinatorics 11, Springer-Verlag.
39. Voss, K., Ríos, H.V. and Peña, J. (1998). "Head tracking by glasses detection". *Computación y Sistemas*, Revista Iberoamericana de Computación, No. 4, CIC, IPN, México (accepted paper, to appear).
40. Wildes, R.P. (1997). "Iris recognition: an emerging biometric technology". *Proceedings of the IEEE*, Vol. 85, No. 9, pp. 1348-1364.

41. Young, D., Tunley, H. and Samuels, R. (1995). "Specialized Hough transform and active contour methods for real-time eye tracking". Cognitive Science Research Paper, No. 386, University of Sussex, England, U.K.