International Journal of Information Technology and Knowledge Management
July-December 2009, Volume 2, No. 2, pp. 405-410

A REVIEW OF VISION BASED HAND GESTURES RECOGNITION

G. R. S. Murthy & R. S. Jadon

With the ever-increasing diffusion of computers into the society, it is widely believed that present popular mode of interactions with computers (mouse and keyboard) will become a bottleneck in the effective utilization of information flow between the computers and the human. Vision based Gesture recognition has the potential to be a natural and powerful tool supporting efficient and intuitive interaction between the human and the computer. Visual interpretation of hand gestures can help in achieving the ease and naturalness desired for Human Computer Interaction (HCI). This has motivated many researchers in computer vision-based analysis and interpretation of hand gestures as a very active research area. We surveyed the literature on visual interpretation of hand gestures in the context of its role in HCI and various seminal works of researchers are emphasized. The purpose of this review is to introduce the field of gesture recognition as a mechanism for interaction with computers.

***Keywords**:* Hand Gesture Recognition, Human Computer Interaction, Computer Vision.

1. Introduction

With the development of information technology in our society, we can expect that computer systems to a larger extent will be embedded into our environment. These environments will impose needs for new types of human-computer-interaction, with interfaces that are natural and easy to use.

The user interface (UI) of the personal computer has evolved from a text-based command line to a graphical interface with keyboard and mouse inputs. However, they are inconvenient and unnatural. The use of hand gestures provides an attractive alternative to these cumbersome interface devices for human-computer interaction (HCI). User's generally use hand gestures for expression of their feelings and notifications of their thoughts. In particular, visual interpretation of hand gestures can help in achieving the ease and naturalness desired for HCI. Vision has the potential of carrying a wealth of information in a non-intrusive manner and at a low cost, therefore it constitutes a very attractive sensing modality for developing hand gestures recognition. Recent researches [1, 2] in computer vision have established the importance of gesture recognition systems for the purpose of human computer interaction.

The primary goal of gesture recognition research is to create a system which can identify specific human gestures and use them to convey information or for device control. A gesture may be defined as a physical movement of the hands, arms, face, and body with the intent to convey information or meaning. Gesture recognition, then, consists not only of the tracking of human movement, but also the interpretation of that movement as semantically meaningful commands. Two approaches are commonly used to interpret gestures for Human Computer interaction. They are

(a) *Methods Which Use Data Gloves:* This method employs sensors (mechanical or optical) attached to a glove that transduces finger flexions into electrical signals for determining the hand posture. This approach forces the user to carry a load of cables which are connected to the computer and hinders the ease and naturalness of the user interaction.

(b) *Methods Which are Vision Based:* Computer vision based techniques are non invasive and based on the way human beings perceive information about their surroundings. Although it is difficult to design a vision based interface for generic usage, yet it is feasible to design such an interface for a controlled environment [3].

1.2 Application Domains

In this section, as the gesture recognition can be used in many more areas, we present an overview of the some of the application domains that employ gesture interactions.

(a) *Virtual Reality:* Gestures for virtual and augmented reality applications have experienced one of the greatest levels of uptake in computing. Virtual reality interactions use gestures to enable realistic manipulations of virtual objects using ones hands, for 3D display interactions [4] or 2D displays that simulate 3D interactions [5].

* Department of Computer Applications, MITS, Gwalior, (M.P.), INDIA. *E-mail: murthy.grs@gmail.com, rsj_mits@yahoo.com*

(b) *Robotics and Telepresence:* Telepresence and telerobotic applications are typically situated within the domain of space exploration and military-based research projects. The gestures used to interact with and control robots are similar to fully-immersed virtual reality interactions, however the worlds are often real, presenting the operator with video feed from cameras located on the robot [6]. Here, gestures can control a robots hand and arm movements to reach for and manipulate actual objects, as well its movement through the world.

(c) *Desktop and Tablet PC Applications:* In desktop computing applications, gestures can provide an alternative interaction to the mouse and keyboard [7]. Many gestures for desktop computing tasks involve manipulating graphics, or annotating and editing documents using pen-based gestures [8].

(d) *Games:* When, we look at gestures for computer games. Freeman *et al.* [9] tracked a player's hand or body position to control movement and orientation of interactive game objects such as cars. Konrad *et al.* [10] used gestures to control the movement of avatars in a virtual world, and Play Station 2 has introduced the Eye Toy, a camera that tracks hand movements for interactive games [11].

(e) *Sign Language:* Sign language is an important case of *communicative gestures*. Since sign languages are highly structural, they are very suitable as test-beds for vision algorithms [12]. At the same time, they can also be a good way to help the disabled to interact with computers. Sign language for the deaf (e.g. American Sign Language) is an example that has received significant attention in the gesture literature [13, 14, 15 and 16].

2. VISION BASED GESTURE RECOGNITION

Vision-based interaction is a challenging interdisciplinary research area, which involves computer Vision and graphics, image processing, machine learning, bio-informatics, and psychology. To make a successful working system, there are some requirements which the system should have:

(a) ***Robustness:*** In the real-world, visual information could be very rich, noisy, and incomplete, due to changing illumination, clutter and dynamic backgrounds, occlusion, etc. Vision-based systems should be user independent and robust against all these factors.

(b) ***Computational Efficiency****:* Generally, Vision-based interaction often requires real-time systems. The vision and learning techniques/algorithms used in Vision-based interaction should be effective as well as cost efficient.

(c) ***User's Tolerance****:* The malfunctions or mistakes of Vision-based interaction should be tolerated. When a mistake is made, it should not incur much loss. Users can be asked to repeat some actions, instead of letting the computer make more wrong decisions.

(d) ***Scalability:*** The Vision-based interaction system should be easily adapted to different scales of applications. Fox eg. the core of Vision-based interaction should be the same for desktop environments, Sign Language Recognition, robot navigation and also for VE.

Most of the systems reviewed rely on the simple idea of detecting and segmenting the gesturing hand from the background using motion detection or skin color. According to Wacs *et al.* [17] proper selection of features or clues, and their combination with sophisticated recognition algorithms, can affect the success or failure of any existing and future work in the field of Human Computer interaction using hand gestures.

2.1 Features for Gesture Recognition

Selecting features is crucial to gesture recognition, since hand gestures are very rich in shape variation, motion and textures. For static hand posture recognition, although it is possible to recognize hand posture by extracting some geometric features such as fingertips, finger directions and hand contours, such features are not always available and reliable due to self-occlusion and lighting conditions. There are also many other non-geometric features such as color, Silhouette and textures, however, they are inadequate in recognition. Since it is not easy to specify features explicitly, the whole image or transformed image is taken as the input and features are selected implicitly and automatically by the recognizer. Hand features can be derived using the following three approaches:

(a) *Model based Approaches (Kinematic Model):* Model based approaches attempt to infer the pose of the palm and the joint angles [18, 19]. Such an approach would be ideal for realistic interactions in virtual environments. Generally, the approach consists of searching for the kinematic parameters that brings the 2D projection of a 3D model of hand into correspondence with an edge-based image of a hand [18]. A common problem with the model-based approaches is the problem of the feature extraction (i.e. edges). The human hand itself is rather texture less and does not provide many reliable edges internally. The edges that are extracted are usually extracted from the occluding boundaries. In order to facilitate extraction and unambiguous correspondence of edges with model edges the approaches require homogeneous backgrounds and high contrast backgrounds relative to the hand.

(b) *View based Approaches:* Due the above mentioned fitting difficulties associated with kinematic model based approaches, many have sought alternative representations of the hand. An alternative approach that has garnered significant focus in recent years is view-based approach [20]. View-based approaches, also referred to as appearance-based approaches, model the hand by a collection of 2D intensity images. In turn, gestures are modeled as a sequence of views.

(c) *Low Level Features based Approaches:* In many gesture applications though all that is required is a mapping between input video and gesture. Therefore, many have argued that the full reconstruction of the hand is not essential for gesture recognition. Instead many approaches have utilized the extraction of low-level image measurements that are fairly robust to noise and can be extracted quickly. Low-level features that have been proposed in the literature include: the centroid of the hand region [21], principle axes defining an elliptical bounding region of the hand [14], and the optical flow/affine flow [22] of the hand region in a scene.

2.2 Classification of Hand Gestures

Hand Gestures can be classified using the following two approaches.

(a) *Rule based Approaches:* Rule-based approaches consist of a set of manually encoded rules between feature inputs. Given an input gesture a set of features are extracted and compared to the encoded rules, the rule that matches the input is outputted as the gesture. As an example, in [23, 24] predicates related to low-level features of the motion of the hands are defined for each of the actions under consideration. When a predicate of a gesture is satisfied over a fixed number of consecutive frames the gesture is returned. A major problem with rule-based approaches is that they rely on the ability of a human to encode rules.

(b) *Machine Learning based Approaches:* A popular machine learning approach is to treat a gesture as the out-put of a stochastic process. Of this class of approach Hidden Markov Models (HMMs) [25, 26, 27] by far have received the most attention in the literature for classifying gestures.

2.3 Gestures Taxonomies

Several taxonomies have been suggested in the literature that deals with psychological aspects of gestures. They vary from author to author. Kendon [28] distinguishes "autonomous gestures" which occurs independently of speech from "gesticulation" that occurs in association of speech. McNeill [29] recognize three groups of gestures: "iconic" and "metaphoric gestures" and "beats". The taxonomy that seems most appropriate for HCI purposes (Figure 1) was developed by Quek [30, 31].

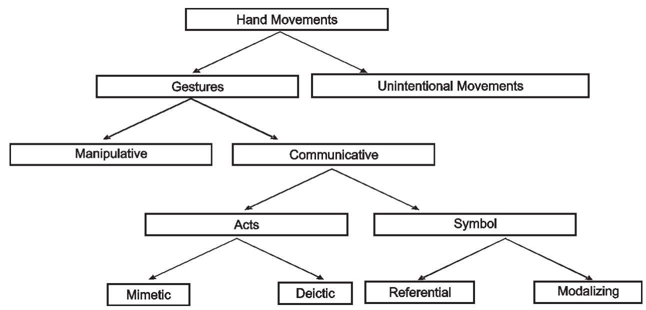

Figure 1: Taxonomy of Hand Gestures

3. RELATED WORKS

Recognizing gestures is a complex task which involves many aspects such as motion modeling, motion analysis, pattern recognition and machine learning, even psycholinguistic studies. The enormous potential for sophisticated and natural human computer interaction using gesture has motivated work as long ago as 1980 with systems such as Bolt's seminal "Put-That-There" [32].

Whilst "Put-That-There" used a data glove as input, video has been used in more recent systems. Starner and Pentland's video-based American Sign Language recognition system is a worthwhile and impressive achievement [33]. A vision-based system able to recognize 14 gestures in real time to manipulate windows and objects within a graphical interface was developed by Ng *et al.* in [34]. Abe *et al.* [35] describe a system that recognizes hand gestures through the detection of the bending of the hand's five fingers, based on image-property analysis. Hasanuzzaman *et al.* [36] presented a real-time hand gesture recognition system using skin color segmentation and multiple-feature based template-matching techniques. In their method, the three largest skin-like regions are segmented from the input images by skin color segmentation technique from YIQ color space and they are compared for feature based template matching using a combination of two features: correlation coefficient and minimum (Manhattan distance) distance qualifier. In their experiment, they have recognized ten gestures out of which two are dynamic facial gestures. These Gesture commands are being sent to their pet robots AIBO through Software Platform for Agent and Knowledge Management (SPAK) [37] and their actions are being accomplished according to users pre defined action for that gesture.

Edge-based techniques to extract image parameters from simple silhouettes are used in [38]. The work presented by Kortenkamp *et al.* [39] shows a system able to recognize six distinct gestures using a coarse 3-D model of a human. Nielsen *et al*. [40] proposed a real time vision system which uses a fast segmentation process to obtain the moving hand from the whole image and a recognition process that identifies the hand posture from the temporal sequence of segmented hands. They have used Hausdorff distance approach for robust shape comparison. Their system recognitions 26 hand postures and achieved a 90% recognition average rate.

Yang *et al.* [22] employ two-dimensional (2-D) motion trajectories, and a time-delay neural network to recognize 40 dynamic American Sign Language (ASL) gestures. Yin and Xie [54] create a fast and robust system that segments using color and recognizes hand gestures for human–robot interaction using a neural network. A system intended to be particularly robust, in real world environments, is presented by Triesch and Malsburg [41]. The strength of the recognition is based on color features extracted from a cadre of training images, and an elastic-graph-match approach.

Alon *et al.* [42] applied DSTW to the simultaneous localization and recognition of dynamic hand gestures. They implemented a hand-signed digit recognition system and their algorithm can recognize gestures using a fairly simple hand detection module that yields multiple candidates. The system does not break down in the presence of a cluttered background, multiple moving objects, multiple skin-colored image regions, and users wearing short sleeves shirts.

Ramamoorthy *et.al.* [43] employ HMM based real time dynamic gesture recognition system which uses both the temporal and shape characteristics of the gesture for recognition. Use of both hand shape and motion pattern is a novel feature of this work. Chen *et al.* [44] introduce a hand gesture recognition system to recognize continuous gestures before stationary background. The system consists of four modules: real time hand tracking and extraction, feature extraction, Hidden Markov model (HMM) training, and gesture recognition.

Yin Xiaoming *et al*. [45] use an RCE neural network based color segmentation algorithm for hand segmentation, extract edge points of fingers as points of interest and match them based on the topological features of the hand, such as the center of the palm. Xiong *et al*. [46] investigated the utility of motion symmetries of a speaker's hands when they are both engaged in communication. For computing hand motion symmetries they have used an approach based on the correlation computations by deploying a two-stage algorithm of window-based correlation and 'hole-filling'. Local symmetries are detected using a windowing operation and demonstrated that the selection of a smaller window size results in better sensitivity to local symmetries at the expense of noise in the form of spurious symmetries and 'symmetry dropoffs'.

Elena *et al*. [47] approached the problem of hand recognition through a matching process in which the segmented hand is compared with all the postures in the system's memory using the Hausdorff distance. The system's visual memory stores all the recognizable postures, their distance transform, their edge map and morphologic information. Joshua R. New *et al*. [21] present a real-time gesture recognition system which can track hand movement, define orientation, and determine the number of fingers being held up in order to allow control of an underlying application.

4. CONCLUSION

The importance of gesture recognition lies in building efficient human–machine interaction. Its applications range from sign language recognition through medical rehabilitation to virtual reality. Given the amount of literature on the problem of gesture recognition and the promising recognition rates reported, one would be led to believe that the problem is nearly solved. Sadly this is not so. A main problem hampering most approaches is that they rely on several underlying assumptions that may be suitable in a controlled lab setting but do not generalize to arbitrary settings. Several common assumptions include: assuming high contrast stationary backgrounds and ambient lighting conditions. Also, recognition results presented in the literature are based on each author's own collection of data, making comparisons of approaches impossible and also raising suspicion on the general applicability. To ameliorate these problems there is a need for the establishment of a standard database for the evaluation and comparison of techniques, similar in spirit to the FERET database [48].

In summary, a review of vision-based hand gesture recognition methods has been presented. Considering the relative infancy of research related to vision-based gesture recognition, remarkable progress has been made. To continue this momentum, it is clear that further research in the areas of feature extraction, classification methods and gesture representation are required, to realize the ultimate goal of humans interfacing with machines on their own natural terms

REFERENCES

[1] Konstantinos G. Derpanis. "A Review of Vision-Based Hand Gestures" (2004).

[2] Sushmita Mitra, and Tinku Acharya, "Gesture Recognition: A Survey", *IEEE Transactions on Systems, Man and Cybernetics*–Part C: Applications and Reviews, **37**(3) (2007).

[3] T. S. Hunang and V. I. Pavloic, "Hand Gesture Modeling, Analysis, and Synthesis," *Proc*. of Int. Workshop *on*

Automatic *Face and Gesture Recognition*, Zurich, (1995), 73-79.

[4] Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z., Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W. "Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists". In: *Proc. of ICPR'96* **II** (1996), 964-968.

[5] Gandy, M., Starner, T., Auxier, J. & Ashbrook, D. "The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring". Proc. of *IEEE Int. Symposium* on Wearable Computers. (2000), 87-94.

[6] Goza, S. M., Ambrose, R. O., Diftler, M. A. & Spain, I. M. "Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform". In: Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, (2004) 623–629.

[7] Stotts, D., Smith, J. M. & Gyllstrom, K. "Facespace: Endo-and Exo-Spatial Hypermedia in the Transparent Video Facetop". In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia. ACM Press, (2004) 48–57.

[8] Smith, G. M. & Schraefel. M. C. "*The Radial Scroll Tool*: Scrolling Support for Stylus-or Touch-Based Document Navigation". In Proc. 17th ACM Symposium on User Interface Software and Technology. ACM Press, (2004) 53–56.

[9] Freeman, W., Tanaka, K., Ohta, J. & Kyuma, K. "Computer Vision for Computer Games". Tech. Rep. and International Conference on Automatic Face and Gesture Recognition, (1996).

[10] Konrad, T., Demirdjian, D. & Darrell, T. "Gesture + Play: Full-Body Interaction for Virtual Environments". In: CHI '03 Extended Abstracts on *Human Factors in Computing Systems.* ACM Press, (2003) 620–621.

[11] Website: University of Chicago: Mcneill Lab for Gesture and Speech Research. *Electronic Resource*, (2006).

[12] Valli, C. & Lucas, C. "Linguistics of American Sign Language: An Introduction". Washington, D. C.: Gallaudet University Press, (2000).

[13] Martinez, A., Wilbur, B., Shay, R. & Kak, A. "Purdue RVL-SLLL ASL Database for Automatic Recognition of ASL". In *IEEE Int. Conf.* on Multimodal Interfaces, (2002) 167–172.

[14] Starner, T., Weaver, J. & Pentland, A. "Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video". *PAMI*, **20**(12) (1998) 1371–1375.

[15] Vogler, C. & Metaxas, D. "A Framework for Recognizing the Simultaneous Aspects of American Sign Language". *Comp. Vision and Image Understanding*, **81**(3) (2001) 358–384.

[16] Waldron, M. "Isolated ASL Sign Recognition System for Deaf Persons". *IEEE Transactions on Rehabilitation Engineering*, **3**(3) (1995) 261–271.

[17] Juan P. Wachs, Helman Stern, Yael Edan, "Automated Setup of a Hand-Gesture Recognition System" IEEE Transactions on Systems, Man, and Cybernetics-Part A: *Systems and Humans*, **35**(6) (2005) 932-944.

[18] Stenger, B., Mendonca, P. & Cipolla, R. "Model-Based 3D Tracking of an Articulated Hand". In IEEE Conference on *Computer Vision and Pattern Recognition*, (2001) 310–315.

[19] Wu, Y., Lin, J. and Huang, T. "Capturing Natural Hand Articulation". In IEEE International Conference on Computer Vision **II** (2001) 426–432.

[20] Gupta, N., Mittal, P., Dutta Roy, S., Chaudhury, S. & Banerjee, S. "Developing a Gesture-Based Interface". *IETE Journal of Research*, **48**(3) (2002) 237–244.

[21] New, J. R., Hasanbelliu, E. and Aguilar, M. "Facilitating User Interaction with Complex Systems via Hand Gesture Recognition." In Proc. of Southeastern ACM Conf., Savannah, (2003).

[22] M. H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2-D Motion Trajectories and its Application to Hand Gesture Recognition," in *PAMI.*, **29**(8) (2002) 1062–1074.

[23] Cutler, R. & Turk, M. "View-Based Interpretation of Real-Time Optical Flow for Gesture Recognition". In IEEE Int. Conf. on *Automatic Face and Gesture Recognition*, (1998) 416–421.

[24] Mu-Chun Su, "A Fuzzy Rule-Based Approach to Spatio-Temporal Hand Gesture Recognition", IEEE Trans. on Systems, Man and Cybernetics, Appl. and Review, **30** (2000).

[25] Starner, T., Weaver, J. & Pentland, A. "Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video". IEEE Trans. on *PAMI*, (1998) 1371–1375.

[26] Wilson. A. & Bobick. A. "Parametric Hidden Markov Models for Gesture Recognition". IEEE Transactions on *Pattern Analysis and Machine Intelligence*, **21**(9) (1999) 884–900.

[27] Lee, H. & Kim, J. "An HMM-Based Threshold Model Approach for Gesture Recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence, **21**(10) (1999) 961–973.

[28] Kendon. "Some Relationships between body motion and speech." In A. W. Siegman and B. Pope (eds.), Studies in Dyadic Communication. New York: Pergamon Press (1972).

[29] McNeill. D. "Hand and Mind: What Gestures Reveal about Thought. "*Chicago*: University of Chicago Press (1992).

[30] F. K. H.Quek, "Toward a Vision-Based Hand Gesture Interface," in Virtual Reality Software and Technology Conference, (1994) 17-31.

[31] F. K. H. Quek, "Eyes in the Interface," Image and Vison Computing, **13** (1995).

[32] R. A. Bolt, "Put-that-there: Voice and Gesture at the Graphics Interface," *Proc. SIGGRAPH80*, (1980).

[33] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**, (1998) 1371–1375.

[34] C. W. Ng and S. Ranganath, "Real-Time Gesture Recognition System and Application," *Image Vis. Comput.*, **20**(13–14) (2002) 993–1007.

[35] K. Abe, H. Saito, and S. Ozawa, "Virtual 3-D Interface System via Hand Motion Recognition from Two Cameras," *IEEE Trans. Syst., Man, Cybern. A*, **32**(4) (2002) 536–540.

[36] Md. Hasanuzzaman, V. Ampornaramveth, Tao Zhang, M. A. Bhuiyan , Y. Shirai and H. Ueno, "Real-Time Vision-Based Gesture Recognition for Human Robot Interaction", in the Proc. of IEEE Int. Conf. on Robotics and Biomimetics, *Shenyang China*, (2004).

[37] Vuthichai Ampornaramveth and Haruki Ueno, "Software Platform for Symbiotic Operations of Human and Networked Robots", *NII Journal*, **3** (2001) 73-81.

[38] J. Segen, "Controlling Computers with Gloveless Gestures," in *Proc. Virtual Reality Systems Conf.*, New York, (1993) 66–78.

[39] D. Kortenkamp, E. Huber, and R. P. Bonasso, "Recognizing and Interpreting Gestures on a Mobile Robot," in *Proc. Artificial Intelligence (AAAI) Workshop*, Portland, (1996) 915–921.

[40] Elena Sanchez-Nielsen, Luis Anton-Canalís and, Mario Hernandez-Tejera , "Hand Gesture Recognition for Human-Machine Interaction", *Journal of WSCG*, **12**(1-3) (2003).

[41] J. Triesch and C. Von Der Malsburg, "A System for Person-Independent Hand Posture Recognition Against Complex Backgrounds," *IEEE Trans. PAMI*, **23** (2001) 1449–1453.

[42] Jonathan Alon, Vassilis Athitsos, Quan Yuan, and Stan Sclaroff, "Simultaneous Localization and Recognition of Dynamic Hand Gestures" *Proc. IEEE Workshop on Motion and Video Computing*, (2005).

[43] Aditya Ramamoorthy, Namrata Vaswani, Santanu Chaudhury, Subhashis Banerjee, "Recognition of Dynamic Hand Gestures", *Pattern Recognition* **36** (2003) 2069-2081.

[44 F. S. Chen, C. M. Fu and C. L. Huang, "Hand Gesture Recognition using a Real-Time Tracking Method and Hidden Markov Models", Image Vision Computer, (2003) 745-758.

[45] Y. Xiaoming and X. Ming, "Estimation of the Fundamental Matrix from Uncalibrated Stereo Hand Images for 3D Hand Gesture Recognition", *Pattern Recognition*, **36** (2003) 567-584.

[46] Yingen Xiong, Francis Quek & David McNeill, "Hand Gesture Symmetric Behavior Detection and Analysis in Natural Conversation", in the Proc. *IEEE Int. Con. ICMI'02*.

[47] Elena Sánchez-Nielsen, Luis Antón-Canalís, Mario Hernández-Tejera, "Hand Gesture Recognition for Human Machine Interaction", in the *Journal of WSCG 04*, **12** (2003).

[48] Phillips, P., Moon, H., Rizvi, S. & Rauss, P. "The FERET Evaluation Methodology for Face-Recognition Algorithms". PAMI, **22**(10) (2000) 1090–1104.