Chapter excerpt from:

B. Kisacanin, V. Pavlovic, and T. Huang (eds.), **Real-Time Vision for Human-Computer Interaction**, Springer, August 2005.

http://www.springer.com/0-387-27697-1

RTV4HCI: A Historical Overview

Matthew Turk

University of California, Santa Barbara
mturk@cs.ucsb.edu

1 Introduction

Real-time vision for human-computer interaction (RTV4HCI) has come a long way in a relatively short period of time. When I first worked in a computer vision lab, as an undergraduate in 1982, I naively tried to write a program to load a complete image into memory, process it, and display it on the lab's special color image display monitor (assuming no one else was using the display at the time). Of course, we didn't actually have a camera and digitizer, so I had to read in one of the handful of available stored image files we had on the lab's modern VAX computer. I soon found out that it was a foolish thing to try and load a whole image – all 512×512 pixel values – into memory all at once, since the machine didn't have that much memory. When the image was finally processed and ready to display, I watched it slowly (very slowly!) appear on the color display monitor, a line at a time, until finally the whole image was visible. It was a painstakingly slow and frustrating process, and this was in a state of the art image processing and computer vision lab.

Only a few years later, I rode inside a large instrumented vehicle – an eight-wheel, diesel-powered, hydrostatically driven all-terrain undercarriage with a fiberglass shell, about the size of a large van, with sensors mounted on the outside and several computers inside – the first time it successfully drove along a private road outside of Denver, Colorado completely autonomously, with no human control. The vehicle, "Alvin," which was part of the DARPA-sponsored Autonomous Land Vehicle (ALV) project at Martin Marietta Aerospace, had a computer onboard that grabbed live images from a color video camera mounted on top of the vehicle, aimed at the road ahead (or alternatively from a laser range scanner that produced depth images of the scene in front of the vehicle). The ALV vision system processed input images to find the road boundaries, which were passed onto a navigation module that figured out where to direct the vehicle so that it drove along the road. Surprisingly, much of the time it actually accomplished this. A complete cycle of the vi-

sion system, including image capture, processing, and display, took about two seconds.

A few years after this, as a PhD student at MIT, I worked on a vision system that detected and tracked a person in an otherwise static scene, located the head, and attempted to recognize the person's face, in "interactive time" – i.e., not at frame-rate, but at a rate fast enough to work in the intended interactive application [24]. This was my first experience in pointing the camera at a person and trying to compute something useful about the person, rather than about the general scene, or some particular inanimate object in the scene. I became enthusiastic about the possibilities for real-time (or interactive-time) computer vision systems that perceived people and their actions and used this information not only in security and surveillance (the primary context of my thesis work) but in interactive systems in general. In other words, real-time vision for HCI. I was not the only one, of course: a number of researchers were beginning to think this could be a fruitful endeavor, and that this area could become another driving application area for the field of computer vision, along with the other applications that motivated the field over the years, such as robotics, modeling of human vision, medical imaging, aerial image interpretation, and industrial machine vision.

Although there had been several research projects over the years directed at recognizing human faces or some other human activity (most notably the work of Bledsoe [3], Kelly [11], Kanade [12], Goldstein and Harmon [9]; see also [18, 15, 29]), it was not until the late 1980s that such tasks began to seem feasible. Hardware progress driven by Moore's Law improvements, coupled with advances in computer vision software and hardware (e.g., [5, 1]) and the availability of affordable cameras, digitizers, full-color bitmapped displays, and other special-purpose image processing hardware, made interactive-time computer vision methods interesting, and processing images of people (yourself, your colleagues, your friends) seemed more attractive to many than processing more images of houses, widgets, and aerial views of tanks.

After a few notable successes, there was an explosion of research activity in real-time computer vision and in "looking at people" projects – face detection and tracking, face recognition, gesture recognition, activity analysis, facial expression analysis, body tracking and modeling – in the 1990s. A quick subjective perusal of the proceedings of some of the major computer vision conferences shows that about 2% of the papers (3 out of 146 papers) in CVPR 1991 covered some aspect of "looking at people." Six years later, in CVPR 1997, this had jumped to about 17% (30 out of 172) of the papers. A decade after the first check, the ICCV 2001 conference was steady at about 17% (36 out of 209 papers) – but by this point there were a number of established venues for such work in addition to the general conferences, including the Automatic Face and Gesture Recognition Conference, the Conference on Audio and Video Based Biometric Person Authentication, the Auditory-Visual Speech Processing Workshops, and the Perceptual User Interface workshops (later merged with the International Conference on Multimodal Interfaces).

It appears to be clear that the interest level in this area of computer vision soared in the 1990s, and it continues to be a topic of great interest within the research community.

Funding and technology evaluation activities are further evidence of the importance and significance of these activities. The Face Recognition Technology (FERET) program [17], sponsored by the U.S. Department of Defense, held its first competition/evaluation in August 1994, with a second evaluation in March 1995, and a final evaluation in September 1996. This program represents a significant milestone in the computer vision field in general, as perhaps the first widely publicized combination of sponsored research, significant data collection, and well-defined competition in the field. The Face Recognition Vendor Tests of 2000 and 2002 [10] continued where the FERET program left off, including evaluations of both face recognition performance and product usability. A new Face Recognition Vendor Test is planned for late 2005, conducted by the National Institute of Standards and Technology (NITS) and sponsored by several U.S. government agencies.

In addition, NIST has also begun to direct and manage a Face Recognition Grand Challenge (FRGC), also sponsored by several U.S. government agencies, which has the goal of bringing about an order of magnitude improvement in performance of face recognition systems through a series of increasingly difficult challenge problems. Data collection will be much more extensive than previous efforts, and various image sources will be tested, included high resolution images, 3D images, and multiple images of a person. More information on the FERET and FRVT activities, including reports and detailed results, as well as information on the FRGC, can be found on the web at `http://www.frvt.org`.

DARPA sponsored a program to develop Visual Surveillance and Monitoring (VSAM) technologies, to enable a single operator to monitor human activities over a large area using a distributed network of active video sensors. Research under this program included efforts in real-time object detection and tracking (from stationary and moving cameras), human and object recognition, human gait analysis, and multi-agent activity analysis.

DARPA's HumanID at a Distance program funded several groups to conduct research in accurate and reliable identification of humans at a distance. This included multiple information sources and techniques, including face, iris, and gait recognition.

These are but a few examples (albeit some of the most high profile ones) of recent research funding in areas related to "looking at people." There are many others, including industry research and funding, as well as European, Japanese, and other government efforts to further progress in these areas. One such example is the recent European Union project entitled Computers in the Human Interaction Loop (CHIL). The aim of this project is to create environments in which computers serve humans by unobtrusively observing them and identifying the states of their activities and intentions, providing helpful assistance with a minimum of human attention or distraction.

Security concerns, especially following the world-changing events of September 2001, have driven many of the efforts to spur progress in this area – particularly those with person identification as their ultimate goal – but the same or similar technologies may be applied in other contexts. Hence, though RTV4HCI is not primarily focused on security and surveillance applications, the two areas can immensely benefit each other.

2 What is RTV4HCI?

The goal of research in real-time vision for human-computer interaction is to develop algorithms and systems that sense and perceive humans and human activity, in order to enable more natural, powerful, and effective computer interfaces. Intuitively, the visual aspects that matter when communicating with another person in a face-to-face conversation (determining identity, age, direction of gaze, facial expression, gestures, etc.) may also be useful in communicating with computers, whether stand-alone or hidden and embedded in some environment. The broader context of RTV4HCI is what many refer to as *perceptual interfaces* [27], *multimodal interfaces* [16], or *post-WIMP interfaces* [28] central to which is the integration of multiple perceptual modalities such as vision, speech, gesture, and touch (haptics). The major motivating factor of these thrusts is the desire to move beyond graphical user interfaces (GUIs) and the ubiquitous mouse, keyboard, and monitor combination – not only for better and more compelling desktop interfaces, but also to better fit the huge variety and range of future computing environments.

Since the early days of computing, only a few major user interface paradigms have dominated the scene. In the earliest days of computing, there was no conceptual model of interaction; data was entered into a computer via switches or punched cards and the output was produced, some time later, via punched cards or lights. The first conceptual model or paradigm of user interface began with the arrival of command-line interfaces in perhaps the early 1960s, with teletype terminals and later text-based monitors. This "typewriter" model (type the input command, hit carriage return, and wait for the typed output) was spurred on by the development of timesharing systems and continued with the popular Unix and DOS operating systems.

In the 1970s and 80s the graphical user interface and its associated desktop metaphor arrived, and the GUI has dominated the marketplace and HCI research for over two decades. This has been a very positive development for computing: WIMP-based GUIs have provided a standard set of direct manipulation techniques that primarily rely on recognition, rather than recall, making the interface appealing to novice users, easy to remember for occasional users, and fast and efficient for frequent users [21]. The GUI/direct manipulation style of interaction has been a great match with the office productivity and information access applications that have so far been the "killer apps" of the computing industry.

However, computers are no longer just desktop machines used for word processing, spreadsheet manipulation, or even information browsing; rather, computing is becoming something that permeates daily life, rather than something that people do only at distinct times and places. New computing environments are appearing, and will continue to proliferate, with a wide range of form factors, uses, and interaction scenarios, for which the desktop metaphor and WIMP (windows, icons, menus, pointer) model are not well suited. Examples include virtual reality, augmented reality, ubiquitous computing, and wearable computing environments, with a multitude of applications in communications, medicine, search and rescue, accessibility, and smart homes and environments, to name a few.

New computing scenarios, such as in automobiles and other mobile environments, rule out many of the traditional approaches to human-computer interaction and demand new and different interaction techniques. Interfaces that leverage natural human capabilities to communicate via speech, gesture, expression, touch, etc., will complement (not entirely replace) existing interaction styles and enable new functionality not otherwise possible or convenient. Despite technical advances in areas such as speech recognition and synthesis, artificial intelligence, and computer vision, computers are still mostly deaf, dumb, and blind. Many have noted the irony of public restrooms that are "smarter" than computers because they can sense when people come and go and act accordingly, while a computer may wait indefinitely for input from a user who is no longer there or decide to do irrelevant (but CPU intensive) work when a user is frantically working on a fast approaching deadline [25].

This concept of *user awareness* is almost completely lacking in most modern interfaces, which are primarily focused on the notion of *control*, where the user explicitly does something (moves a mouse, clicks a button) to initiate action on behalf of the computer. The ability to see users and respond appropriately to visual identity, location, expression, gesture, etc. – whether via implicit user awareness or explicit user control – is a compelling possibility, and it is the core thrust of RTV4HCI.

Human-computer interaction (HCI) – the study of people, computer technology, and the ways they influence each other – involves the design, evaluation, and implementation of interactive computing systems for human use. HCI is a very broad interdisciplinary field with involvement from computer science, psychology, cognitive science, human factors, and several other disciplines, and it involves the design, implementation, and evaluation of interactive computer systems in the context of the work or tasks in which a user is engaged [7]. The user interface – the software and devices that implement a particular model (or set of models) of HCI – is what people routinely experience in their computer usage, but in many ways it is only the tip of the iceberg. "User experience" is a term that has become popular in recent years to emphasize that the complete experience of the user – not an isolated interface technique or technology – is the final criterion by which to measure the utility of any HCI technology. To be truly effective as an HCI technology,

computer vision technologies must not only work according to the criteria of vision researchers (accuracy, robustness, etc.), but they must be useful and appropriate for the tasks at hand. They must ultimately deliver a better user experience.

To improve the user experience, either by modifying existing user interfaces or by providing new and different interface technologies, researchers must focus on a range of issues. Shneiderman [21] described five human factors objectives that should guide designers and evaluators of user interfaces: time to learn, speed of performance, user error rates, retention over time, and subjective satisfaction. Researchers in RTV4HCI must keep these in mind – it's not just about the technology, but about how the technology can deliver a better user experience.

3 Looking at People

The primary task of computer vision in RTV4HCI is to detect, recognize, and model meaningful communication cues – that is, to "look at the user" and report relevant information such as the user's location, expressions, gestures, hand and finger pose, etc. Although these may be inferred using other sensor modalities (such as optical or magnetic trackers), there are clear benefits in most environments to the unobtrusive and unencumbering nature of computer vision. Requiring a user to don a body suit, to put markers on the face or body, or to wear various tracking devices, is unacceptable or impractical for most anticipated applications of RTV4HCI.

Visually perceivable human activity includes a wide range of possibilities. Key aspects of "looking at people" include the detection, recognition, and modeling of the following elements [26]:

- Presence and location – Is someone there? How many people? Where are they (in 2D or 3D)? [Face and body detection, head and body tracking]
- Identity – Who are they? [Face recognition, gait recognition]
- Expression – Is a person smiling, frowning, laughing, speaking . . . ? [Facial feature tracking, expression modeling and analysis]
- Focus of attention – Where is a person looking? [Head/face tracking, eye gaze tracking]
- Body posture and movement – What is the overall pose and motion of the person? [Body modeling and tracking]
- Gesture – What are the semantically meaningful movements of the head, hands, body? [Gesture recognition, hand tracking]
- Activity – What is the person doing? [Analysis of body movement]

The computer vision problems of modeling, detecting, tracking, recognizing, and analyzing various aspects of human activity are quite difficult. It's hard enough to reliably recognize a rigid mechanical widget resting on a table, as image noise, changes in lighting and camera pose, and other issues

contribute to the general difficulty of solving a problem that is fundamentally ill-posed. When humans are the objects of interest, these problems are magnified due to the complexity of human bodies (kinematics, non-rigid musculature and skin), and the things people do – wear clothing, change hairstyles, grow facial hair, wear glasses, get sunburned, age, apply makeup, change facial expression – that in general make life difficult for computer vision algorithms. Due to the wide variation in possible imaging conditions and human appearance, robustness is the primary issue that limits practical progress in the area.

There have been notable successes in various "looking at people" technologies over the years. One of the first complete systems that used computer vision in a real-time interactive setting was the system developed by Myron Krueger, a computer scientist and artist who first developed the VIDEOPLACE responsive environment around 1970. VIDEOPLACE [13] was a full body interactive experience. It displayed the user's silhouette on a large screen (viewed by the user as a sort of mirror) and incorporated a number of interesting transformations, including letting the user hold, move, and interact with 2D objects (such as a miniature version of the user's silhouette) in real-time. The system let the user do finger painting and many other interactive activities. Although the computer vision was relatively simple, the complete system was quite compelling, and it was quite revolutionary for its time. A more recent system in a similar spirit was the "Magic Morphin Mirror / Mass Hallucinations" by Darrell et al. [6], an interactive art installation that allowed users to see modified versions of themselves in a mirror-like display. The system used computer vision to detect and track faces via a combination of stereo, color, and grayscale pattern detection.

The first computer programs to recognize human faces appeared in the late 1960s and early 1970s, but only in the past decade have computers become fast enough to support real-time face recognition. A number of computational models have been developed for this task, based on feature locations, face shape, face texture, and combinations thereof; these include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Gabor Wavelet Networks (GWNs), and Active Appearance Models (AAMs). Several companies, such as Identix Inc., Viisage Technology Inc., and Cognitec Systems, now develop and market face recognition technologies for access, security, and surveillance applications. Systems have been deployed in public locations such as airports and city squares, as well as in private, restricted access environments. For a comprehensive survey of face recognition research, see [34].

The MIT Media Lab was a hotbed of activity in computer vision research applied to human-computer interaction in the 1990s, with notable work in face recognition, body tracking, gesture recognition, facial expression modeling, and action recognition. The ALIVE system [14] used vision-based tracking (including the Pfinder system [31]) to extract a user's head, hand, and foot positions and gestures to enable the user to interact with computer-generated autonomous characters in a large-screen video mirror environment. Another compelling example of vision technology used effectively in an interactive en-

vironment was the Media Lab's KidsRoom project [4]. The KidsRoom was an interactive, narrative play space. Using computer vision to detect the locations of users and to recognize their actions helped to deliver a rich interactive experience for the participants. There have been many other compelling prototype systems developed at universities and research labs, some of which are in the initial stages of being brought to market. A system to recognize a limited vocabulary of American Sign Language (ASL) was developed, one of the first instances of real-time vision-based gesture recognition using Hidden Markov Models (HMMs).

Other notable research progress in important areas includes work in hand modeling and tracking [19, 32], gesture recognition [30, 22], facial expression analysis [33, 2], and applications to computer games [8].

In addition to technical progress in computer vision – better modeling of bodies, faces, skin, dynamics, movement, gestures, and activity, faster and more robust algorithms, better and larger databases being collected and shared, the increased focus on learning and probabilistic approaches – there must be an increased focus on the HCI aspects of RTV4HCI. Some of the critical issues include a deeper understanding of the semantics (e.g., when is a gesture a gesture, how is contextual information properly used?), clear policies on required accuracy and robustness of vision modules, and sufficient creativity in design and thorough user testing to ensure that the suggested solution actually benefits real users in real scenarios. Having technical solutions does not guarantee, by any means, that we know how to apply them more appropriately – intuition may be severely misleading. Hence, the research agenda for RTV4HCI must include both development of individual technology components (such as body tracking or gesture recognition) and the integration of these components into real systems with lots and lots of user testing.

Of course, there has been great research in various areas of real-time vision-based interfaces at many universities and labs around the world. The University of Illinois at Urbana-Champaign, Carnegie Mellon University, Georgia Tech, Microsoft Research, IBM Research, Mitsubishi Electric Research Laboratories, the University of Maryland, Boston University, ATR, ETL, the University of Southampton, the University of Manchester, INRIA, and the University of Bielefeld are but a few of the places where this research has flourished. Fortunately, the barrier to entry in this area is relatively low; a PC, a digital camera, and an interest in computer vision and human-computer interaction are all that is necessary to start working on the next major breakthrough in the field. There is much work to be done.

4 Final Thoughts

Computer vision has made significant progress through the years (and especially since my first experience with it in the early 1980s). There have been notable advances in all aspects of the field, with steady improvements in the

performance and robustness of methods for low-level vision, stereo, motion, object representation and recognition, etc. The field has adopted more appropriate and effective computational methods, and now includes quite a wide range of application areas. Moore's Law improvements in hardware, advancements in camera technology, and the availability of useful software tools (such as Intel's OpenCV library[1]) have led to small, flexible, and affordable vision systems that are available to most researchers. Still, a rough back-of-the-envelope calculation reveals that we may have to wait some time before we really have the needed capabilities to perform very computationally intensive vision problems well in real time. Assuming relatively high speed images (100 frames per second) in order to capture the temporal information needed for humans moving at normal speeds, relatively high resolution images (1000 $\times$ 1000 pixels) in order to capture the needed spatial resolution, and an estimated 40k operations per pixel in order to do the complex processing required by advanced algorithms, we are left needing a machine that delivers 4×10^{12} operations per second [20]. If Moore's Law holds up, it's conceivable that we could get there within a (human) generation. More challenging will be figuring out what algorithms to run on all those cycles! We are still more limited by our lack of knowledge than our lack of cycles. But the progress in both areas is encouraging.

RTV4HCI is still a nascent field, with growing interest and awareness from researchers in computer vision and in human-computer interaction. Due to how the field has progressed, companies are springing up to commercialize computer vision technology in new areas, including consumer applications. Progress has been steadily moving forward in understanding fundamental issues and algorithms in the field, as evidenced by the primary conferences and journals. Useful large datasets have been collected and widely distributed, leading to more rapid and focused progress in some areas. An apparent "killer app" for the field has not yet arisen, and in fact may never arrive; it may be the accumulation of many new and useful abilities, rather than one particular application, that finally validates the importance of the field. In all of these areas, significant speed and robustness issues remain; real-time approaches tend to be brittle, while more principled and thorough approaches tend to be excruciatingly slow. Compared to speech recognition technology, which has seen years of commercial viability and has been improving steadily for decades, RTV4HCI is still in the Stone Age.

At the same time, there is an increased amount of cross-pollination between people in the computer vision and HCI communities. Quite a few conferences and workshops have appeared in recent years devoted to intersections of the two fields. If the past provides an accurate trajectory with which to anticipate the future, we have much to look forward to in this interesting and challenging endeavor.

[1] `http://www.intel.com/research/mrl/research/opencv`

References

1. M. Annaratone, E. Arnould, T. Gross, H. Kung, and J.Webb, "The Warp computer: architecture, implementation and performance," IEEE Trans on Computers, C-36(12), pp. 1523-1538, 1987.
2. M. Black, and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," Proceedings of the International Conference on Computer Vision, pp. 374-381, Cambridge, MA, 1995.
3. W.W. Bledsoe, "Man-machine facial recognition," Technical Report PRI 22, Panoramic Research Inc., Palo Alto, CA, August 1966.
4. A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schütte, and A. Wilson, "The KidsRoom: a perceptually-based interactive and immersive story environment," PRESENCE: Teleoperators and Virtual Environments, 8(4), pp. 367-391, August 1999.
5. P. J. Burt, "Smart sensing with a pyramid vision machine," Proceedings of the IEEE, Vol. 76, pp. 1006-1015, 1988.
6. T. Darrell, G. Gordon, W. Woodfill, and H. Baker, "A Magic Morphin Mirror," SIGGRAPH '97 Visual Proceedings, ACM Press. 1997.
7. A. Dix, J. Finlay, G. Abowd, and R. Beale, Human-Computer Interaction, Second Edition, Prentice Hall Europe, 1998.
8. W. Freeman, K. Tanaka, J. Ohta, and K. Kyuma, "Computer vision for computer games," Proc. Second International Conference on Automatic Face and Gesture Recognition. Killington, VT, 1996.
9. A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identi?cation of human faces," Proc. IEEE, Vol. 59, pp. 748-760, 1971.
10. P. J. Grother, R. J. Micheals and P. J. Phillips, "Face Recognition Vendor Test 2002 Performance Metrics," Proceedings 4th International Conference on Audio Visual Based Person Authentication, 2003.
11. M. D. Kelly, "Visual identification of people by computer," Stanford Artificial Intelligence Project Memo AI-130, July 1970.
12. T. Kanade, "Picture processing system by computer complex and recognition of human faces," Dept. of Information Science, Kyoto University, Nov. 1973.
13. M. W. Krueger, Artificial Reality II, Addison-Wesley, Reading, MA, 1991.
14. P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The ALIVE system: wireless, full-body interaction with autonomous agents," ACM Multimedia Systems, Special Issue on Multimedia and Multisensory Virtual Worlds, Spring 1996.
15. J. O'Rourke and N. Badler, "Model-based image analysis of human motion using constraint propagation," IEEE Transactions on PAMI, vol.2, no.6, pp.522-536, 1980.
16. S. Oviatt, T. Darrell, and M. Flickner, "Multimodal interfaces that flex, adapt, and persist," Communications of the ACM, Vol. 47, No. 1, pp. 30-33, January 2004.
17. P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, October 2000.
18. R.F. Rashid, "Towards a system for the interpretation of Moving Light Displays," IEEE Transactions on PAMI, vol.2, no.6, pp.574-581, Nov. 1980.

19. J. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: an application to human hand tracking," Proceedings of the 3rd European Conference on Computer Vision (ECCV '94), Volume II, pp. 35-46, May 1994.
20. S. Shafer, personal communication, 1998.
21. B. Shneiderman, Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison Wesley, 3rd edition, March 1998.
22. M. Stark and M. Kohler, "Video based gesture recognition for human computer interaction," in W. D. Fellner (Ed.), Modeling - Virtual Worlds - Distributed Graphics, 1995.
23. M. Turk, "Computer vision in the interface," Communications of the ACM, Vol. 47, No. 1, pp. 60-67, January 2004.
24. M. Turk, "Interactive-time vision: face recognition as a visual behavior," Ph.D. Thesis, MIT Media Lab, September 1991.
25. M. Turk, "Perceptive media: machine perception and human computer interaction," Chinese Computing Journal, 2001.
26. M. Turk and M. Kölsch, "Perceptual Interfaces," G. Medioni and S.B. Kang (eds.), Emerging Topics in Computer Vision, Prentice Hall, 2004.
27. M. Turk and G. Robertson, "Perceptual User Interfaces," Communications of the ACM, Vol. 43, No. 3, pp. 33-34, March 2000.
28. A. van Dam, "Post-wimp user interfaces," Communications of the ACM, 40(2):63-67, 1997.
29. J.A. Webb and J. K. Aggarwal, "Structure from motion of rigid and jointed objects," Artificial Intelligence, vol.19, pp.107-130, 1982.
30. C. Vogler and D. Metaxas, "Adapting Hidden Markov models for ASL recognition by using three dimensional computer vision methods," IEEE International Conference on Systems, Man and Cybernetics, pp. 156-161, October 1997.
31. C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780-785, July 1997.
32. Y. Wu and T. S. Huang, "Hand modeling, analysis, and recognition," IEEE Signal Processing Magazine, May 2001.
33. A. Zelinsky and J. Heinzmann, "Real-time visual recognition of facial gestures for human-computer interaction," Proc. Second International Conference on Automatic Face and Gesture Recognition. Killington, VT, 1996.
34. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458, 2003.