Human-computer interaction through computer vision

Homero V. Rios
Faculty of Physics and Artificial Intelligence
Universidad Veracruzana
Sebastián Camacho 5
Xalapa, Veracruz 91000, Mexico
+52 28 172957
hrios@mia.uv.mx

ABSTRACT

In this paper, we describe the work of our group on the use of computer vision to complement existing modes of human-interaction. Our main achievements have been the development of an iris tracker and a hand tracker for HCI similar to a “visual mouse”, and a deformable model of a generic face for facial feature extraction, for emotion recognition.

Keywords

Computer vision, human-computer interaction, gesture recognition, emotion recognition.

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. He anticipated the use of computers for multimedia processing.

Since Bush's time many technological breakthroughs have occured, but computers are still limited in their multimedia understanding [4]. 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 [1].

Several researchers have identified this unbalance and are working on more intuitive interfaces like virtual reality, speech recognition, image understanding and multimodal interfaces [4].

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 and the systems that we have developed in our lab.

COMPUTER VISION AND HUMAN-COMPUTER INTERACTION

We are interested in research work that uses 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* [5]. The 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, emotions recognition [5], sign language understanding, teleoperation in virtual reality robotics environments, tracking of the whole human body for surveillance application, hand tracking, iris tracking and recognition and head tracking for general human computer interaction.

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 [3]. 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.

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 location of a point of interest, and we can associate it with 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. The performance of this tracker is between 3-5 frames/second running on a Indy, SGI workstation. For our algorithms to work properly, it is important to have a good illumination that generates good contrast in the images and the edges appear well defined. We have found this experimental condition as the most critical for our implementation.

HAND TRACKER

Our approach for the hand tracker is not as powerful and general as in other works [6], but we provide simpler solutions for tracking a hand with extended fingers. 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. 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. 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 tracker is also between 3-5 frames/second on the same platform as described above.

FACIAL FEATURE DETECTION

Since faces are non-rigid objects the best way to model them is through the use of deformable models which should have enough parameters to accommodate most of the variations in shape. Active contours have been used to detect and track facial features like head contour, lips, eyebrows and eyes [8]. The problem is that since “snakes” can adapt any shape, sometimes they take nonvalid shapes. One solution is the use of trainable deformable models from examples, like the point distribution model proposed by Cootes [2].

In our case, we have used this approach to detect facial features and interpret facial gestures. The difference between this work and ours, is that we have built a database of faces which shows more clearly different gestures. Also,

our model has been expanded to include landmarks which enhances the identification of facial gestures.

WORK IN PROGRESS AND CONCLUSION

Our work in human-computer interaction through computer vision continues in tracking people in digital video and also in the interpretation of their behaviour.

REFERENCES

1. Bolt, R.A. Conversing with computers. *Technology Review*, 88(2), 35-43.
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models--Their training and applications. *Computer Vision, Graphics and Image Understanding* 61 (1995) 38-59.
3. Lam, K.M. and Yang, H. Locating and extracting the eye in human face images. *Pattern Recognition*, Vol. 29, No. 5, 1996, 771-779.
4. Maybury, M.T. (ed.). *Intelligent Multimedia Interfaces.* MIT Press, 1993.
5. Pentland, A.P. Smart rooms. *Scientific American.* April 1996, 54-62.
6. Rehg, J.M. and Kanade, T. DigitEyes: vision-based human hand tracking,. School of Computer Science, Carnegie Mellon University, Technical report number:CMU-CS-93-220.
7. Rios, H.V. y Barradas, P.D. Human-computer interaction through ocular motion. Proceedings V Latinamerican conference in artificial intelligence, Limusa, 1996, 492-501. (In Spanish).
8. Terzopoulos, D., Szeliski, R.: Tracking with Kalman Snakes. In: Blake, A., Yuille, A. (eds.): *Active Vision*. MIT Press (1992), 3-20