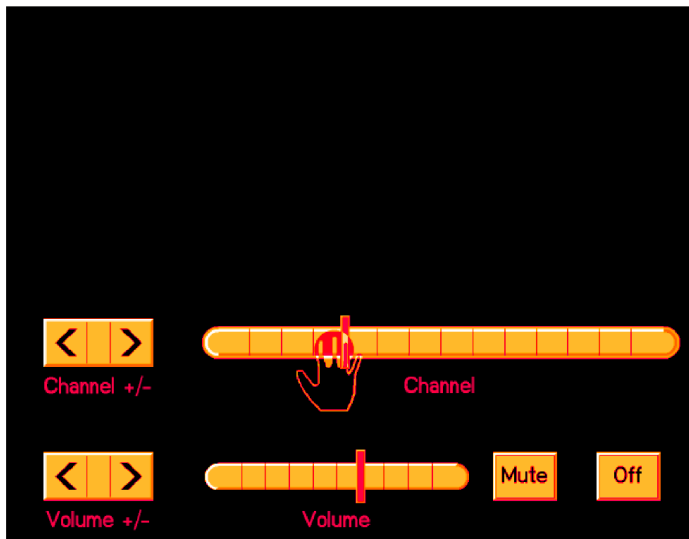# Television Set Controlled By Hand Gestures

We made a prototype of a television set that the viewer could control by hand gestures, instead of a remote control.

A small video camera on the television set scans the viewer for the appropriate hand signal controls. The user uses only one gesture: the open hand, facing the camera. To turn the TV set on, the viewer raises his open hand. The computer recognizes that gesture, and turns on the television screen. Overlaid on the screen is a hand icon, and some graphical controls. The hand icon tracks the position of the viewer's hand. The viewer can then move his own hand to adjust the various controls for channel, volume, etc. When the viewer closes his hand, the on-screen graphical overlay disappears, and he watches television as usual.

Our prototype was a two-screen system, using a computer workstation and a television. We were able to evaluate the feel of controlling a television set by hand signals. Controlling a television set remotely through hand gestures seemed to be exciting for the people who tried the prototype. However, "channel-surfing" would tire the arm; a more relaxed hand signal needs to be developed for extended viewing.

Background and objectives: Infra-red-based television remoted controls are a popular and successful television interface, which viewers value. We sought to explore other interface methods which might be complementary with the television remote control. The work addressed two fundamental issues of gesture--based human--computer interaction: (1) How can one communicate a rich set of commands without extensive user training and memorization of gestures? (2) How can the computer recognize the commands in a complicated visual environment?

Technical discussion: Our solution to both the command vocabulary and the command recognition issues relies on the visual feedback provided by the television display. The user only needs to memorize a single hand signal (the open hand), which forms a distinct patten for the computer to recognize.

To track the hand, we used normalized correlation of the image with a template for the hand. The normalized correlation between two vectors is the cosine of the angle between

them. The template pixels form the entries of one vector and the corresponding pixels at some offset position in the image form the second vector. The normalized correlation will be highest at the position of the image patch most closely matching the template pixel values. Usually, this is the position of the hand. We find the hand position to sub--pixel accuracy by modeling the correlation surface as a quadratic polynomial and finding the position of maximum correlation.

For efficiency, we did not process stationary objects. We maintain a running average of the scene, and remove this stationary background from the incoming image.

There is a tradeoff between the system response time and field--of--view. To obtain reasonable response, we limited the field of view to 25 degrees during search for the trigger gesture, and 15 degrees during tracking of the hand.

Computer Vision for Computer Interaction
SIGGRAPH Computer Graphics magazine, November 1999
W. T. Freeman, P. A. Beardsley, H. Kage, K. Tanaka, K. Kyuma, C. D. Weissman.
Also available as MERL-TR99-36.

Computer Vision for Interactive Computer Graphics
IEEE Computer Graphics and Applications, Vol. 18, No. 3, May-June 1998
W. T. Freeman, D. B. Anderson, P. A. Beardsley, C. N. Dodge, M. Roth, C. D. Weissman, W. S. Yerazunis, H. Kage, K. Kyuma, Y. Miyake, and K. Tanaka.
Also available as MERL-TR99-02.

Computer Vision for Computer Games
In 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA. IEEE.
W. T. Freeman, K. Tanaka, J. Ohta, K. Kyuma
Also available as MERL-TR96-35.