Deictic Gestures With A Time-of-Flight Camera

Martin Haker, Martin Böhme, Thomas Martinetz, and Erhardt Barth

Institute for Neuro- and Bioinformatics, University of Lübeck, Germany

Keywords: Deictic gestures, gesture disambiguation, time-of-flight camera

Fig. 1. Segmented image of the user with the detected locations of the head and hand marked by crosses. The time-of-flight camera measures the three-dimensional positions of these points, which are then used to compute the pointing direction.

We use a novel type of sensor, the time-of-flight (TOF) camera, to implement simple and robust gesture recognition. The TOF camera [1] provides a range map that is perfectly registered with an intensity image at 20 frames per second or more, depending on the integration time. The camera works by emitting infrared light and measuring the time taken by the light to travel to a point in the scene and back to the camera; the time taken is proportional to the distance of the point from the camera, allowing a range measurement to be made at each pixel.

In this paper, we use gestures recognized using the TOF camera to control a slideshow presentation, similar to [2]. Another idea we adapt from [2] is to recognize only gestures made towards an "active area"; valid gestures made with the hand pointing elsewhere are ignored. This solves the problem (also known as the "immersion syndrome") that unintentional hand movements or gestures made towards other people may erroneously be interpreted as commands.

We expand this idea by allowing the same gesture to mean different things when made towards different active areas. Specifically, the slideshow is controlled

The ARTTS project is funded by the European Commission (contract no. IST-34107) within the Information Society Technologies (IST) priority of the 6th Framework Programme. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

in the following way: To go to the next slide, point to the right of the screen and make a flick of the hand; to go to the previous slide, point to the left of the screen and flick your hand. Point at the screen and a dot appears at the location you are pointing to, allowing you to highlight certain elements of the slide.

Our algorithm works as follows: We segment the user using an adaptive range threshold (see Figure 1). As Nickel and Stiefelhagen [3] show, the line connecting the head and hand is a good estimate for the pointing direction. We find the head and hand using a simple but effective heuristic: The initial guess for the hand is the topmost pixel in the leftmost pixel column of the silhouette; the head is the topmost pixel in the tallest pixel column. We define rectangular regions of interest (ROIs) around the initial guesses and compute the centroids of the foreground pixels in the ROIs to find the centres of the head and hand blobs; these positions are marked by crosses in Figure 1. Because the TOF camera allows us to determine the position of the head and hand in space, we directly obtain a 3D vector (in camera coordinates) for the pointing direction. A Kalman filter tracks the head and hand from frame to frame.

To determine where the user is pointing on the screen, we need to know its position relative to the camera. This is determined in a calibration procedure where the user points at the four corners of the screen from two different locations; this information is sufficient to compute the position of the screen.

The "flick" gesture is recognized when the hand remains approximately stationary and the difference between consecutive frames in the hand region is above a threshold for a certain number of frames. This very simple technique is sufficient because hand flicks are only recognized when the user is pointing at one of the two active regions; when pointing elsewhere, the user need not be concerned that hand movements might be misinterpreted as gestures.

We believe pointing is a powerful way to determine whether a gesture is intended for the system at all and to assign different meanings to the same gesture depending on where the user is pointing. A simple gesture with a simple recognition procedure is sufficient for our application because the meaning of the gesture is strongly tied to the direction it is made in. In addition, we have shown how the TOF camera makes it easy to compute a 3D vector for the direction the user is pointing in because it directly measures the spatial position of objects. The range map also makes segmentation easier than with a conventional camera.

References

1. Oggier, T., Büttgen, B., Lustenberger, F., Becker, G., Rüegg, B., Hodac, A.: SwissRanger^{TM} SR3000 and first experiences based on miniaturized 3D-TOF cameras. In Ingensand, K., ed.: Proc. 1^{st} Range Imaging Research Day, Zurich (2005) 97–108
2. Baudel, T., Beaudouin-Lafon, M.: CHARADE: Remote control of objects using free-hand gestures. Communications of the ACM **36** (1993) 28–35
3. Nickel, K., Stiefelhagen, R.: Pointing gesture recognition based on 3D-tracking of face, hands and head orientation. In: International Conference on Multimodal Interfaces. (2003) 140–146