In Proceedings of the 14th International Conference on Pattern Recognition, Brisbane, Australia, Aug. 1998, Vol. 2, pages 1244–1246

Free-Hand Pointer by Use of an Active Stereo Vision System

Yi-Ping Hung†, Yao-Strong Yang‡, Yong-Sheng Chen†, Ing-Bor Hsieh‡, and Chiou-Shann Fuh‡
† Institute of Information Science, Academia Sinica, Taipei, Taiwan
‡ Deptartment of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan

E-mail: `hung@iis.sinica.edu.tw`

Abstract

In this work, we have developed a system of free-hand pointer by tracking the finger of the speaker with an active stereo vision system and then computing the projection direction in either of the following two modes: the finger-orientation mode and the eye-to-fingertip mode. We prefer the latter for its robustness. In order to allow the speaker to move around in a wider 3D space without reducing the pointing resolution, we utilize a well-calibrated active stereo vision system which has a relatively small field of view but can control its stereo cameras to fixate at the moving finger. Our experiments have successfully demonstrated the feasibility of developing a free-hand pointer using an active stereo vision system.

1. Introduction

Human hands play an important role in human-computer interface. They usually function through keyboard, mouse, digitizer, touch panel, and so on. The input capability of human hand is extended through the help of some special equipments, e.g. data-glove [1] and colored glove [5]. Although wearing these gloves can simplify the segmentation and feature extraction task, it will be most desirable if the user is allowed to use his/her free hands to communicate with the computer without holding or touching any physical tool.

There are many related works of human-computer interaction which utilize free hand/finger tracking: interface in VR environment [6], television control [2], and finger pointer [3]. In this paper, we focus on applying finger tracking to free-hand pointer for presentation. Such a system will allow the speaker to use his/her finger (e.g., index finger), instead of a rod or a laser pointer, to indicate the area on the projection screen where he/she wants the audience to pay attention to. We use stereo vision to compute 3D finger position and direction without building a complicated 3D hand model [7]. The methods used to detect the finger position and to compute the finger direction are simple and efficient. A well-calibrated active stereo vision system—the IIS (Institute of Information Science) head—helps to track and fixate the moving finger in a wider 3D space.

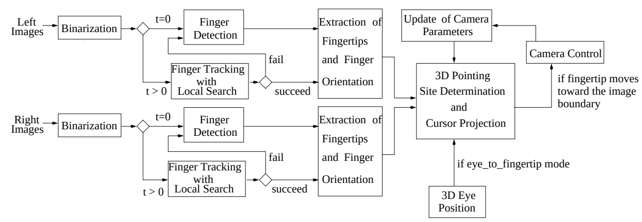

Figure 1. The block diagram of our free-hand pointer.

In the following, we briefly describe our free-hand pointer whose block diagram is shown in Figure 1. First, the left and right images taken with our stereo vision system are binarized. In the initial stage, we use mathematical morphology to detect and segment the finger with a global search. In the subsequent images, local search using prediction is used to efficiently track the finger. Time-consuming global search of finger will be invoked again only if the system loses track of the finger. Once the finger is detected and segmented, we use a simple and efficient algorithm to extract the fingertip and the finger orientation. Then, with the calibrated camera parameters, 3D fingertip position and finger direction can be computed by using the 2D fingertip position and finger direction obtained from the left and right images. We propose two ways for computing the 3D projection site on a given projection plane. Once the 3D projection site is determined, a cursor (or a mark) can be projected onto that site accordingly. If the detected fingertip moves toward the image boundary, our system can move the cameras to fixate at the fingertip so that the fingertip can remain within the field of view.

2. Finger Tracking

In the initial stage or whenever we lose track of the finger with local search, a global search is performed to detect the finger in stereo images. In this work, we assume the background is much darker than the hand. Otsu's method[4] is used to determine the binarization threshold automatically. Then, morphological closing is used to remove the noise. Since the finger is the thin part compared with the palm, it can be extracted with the operation: $I' = I - I \circ K$ followed by a largest connected component labeling, where I' is the resulted image, I is the original image, and K is the structuring element used for opening.

The fingertip is initially chosen to be the farthest point along the contour of finger from the palm. The finger direction is initially chosen to be the vector pointing from the centroid of finger to the fingertip. Both the fingertip and finger direction can be refined using a simple cylindrical model for finger.

Once the finger is detected, a local search window centered at the centroid of the finger region can be used to speed up the detection of finger in the next image frame. If the speaker's finger does not move dramatically, his/her finger can usually be tracked more efficiently with local search; otherwise, the system will invoke global search for finger detection to ensure robustness.

3. Determination of Projection Site

Once the 2D position of the fingertip and the 2D orientation of the finger in the left and right images are obtained, we can compute the 3D position of the fingertip and the 3D orientation of the finger using the known camera parameters of the stereo cameras. In this paper, we propose two approaches for determining the pointing direction of the finger, which can then be used to compute the projection site of the cursor (or the mark) by intersecting the pointing ray with the projection screen.

1. Finger-Orientation Mode:

step 1: Select several points on the mid-line of the finger in the left image.

step 2: Compute their corresponding points by finding the intersection points of their epipolar lines and the mid-line of the finger in the right image.

step 3: Compute the 3D positions of these selected points (camera parameters are known) and fit a 3D line to them.

2. Eye-to-Fingertip Mode:

step 1: Move the fingertip to any two points between the projection site and the right eye (or the left eye) and compute their 3D finger positions. Hence, we can get the direction vector between the eye and the projection site.

step 2: With the knowledge of the average distance between the eye and hand, we can have a rough estimate of the 3D eye position.

step 3: While the speaker is moving his/her fingertip, the system keeps on detecting the 3D position of the fingertip and computing the intersection point of the eye-to-fingertip vector and the projection screen.

In the *finger-orientation* mode, the resolution of the finger pointer is not very high. Also, its performance is quite sensitive to noise. Since the distance between the fingertip and the eye is much longer than the length of the finger, the pointing performance of the *eye-to-fingertip* mode is less sensitive to noise. A minor disadvantage is that the user has to raise up his/her hand high enough such that the finger is between the eye and the projection screen.

4. Finger Tracking in a Wider Space

For visual tracking systems using fixed cameras, the field of view of the camera limits the 3D space where the target can be tracked. This limitation can be removed if the camera can be controlled to fixate at the moving target (i.e., if using an active vision system). Another approach to remove this space limitation is to use a lens (or a camera system) with a wider field of view. However, this approach will make the target look smaller in the images, thus reduce the resolution and accuracy in extracting the 3D position of the target. In the application of free-hand pointer for presentation, the pointing resolution is quite important. Hence, we adopt the approach of using an active vision system.

In the experiments, we use an active vision system, the IIS head shown on the left side of Figure 2, to move the cameras to fixate at the finger when the finger is approaching the image boundary. Our IIS head has been calibrated [8] very accurately such that the prediction error is about one pixel and the epipolar error is about 0.2 pixel, even when all its eight joints are moving simultaneously. Therefore, it is easy and straightforward for us to use the IIS head to track and fixate at the finger of the speaker in a much wider 3D space.

5. Experiments

We will show some experimental results for the two pointing modes. Figure 2 illustrates the operation of our free-hand pointer in the *eye-to-fingertip* mode. A mark is projected on the projection screen at the site computed from the fingertip position, pre-determined eye position, and the given position

Figure 2. This figure illustrates the operation of our free-hand pointer. The speaker is pointing his index finger to the projection screen (under the word "3D"). The IIS head on the left takes the stereo image pairs to be processed. According to the projection site determined from the stereo image pairs, the computer then controls its cursor to the word "3D".

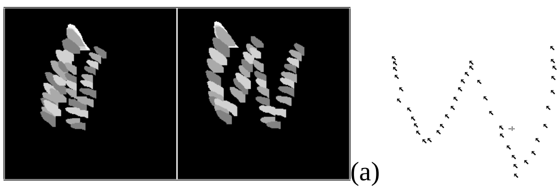
(a) (b)

Figure 3. (a) The image trajectory of the finger in the stereo image pairs when using the *eye-to-fingertip* mode to write a "W". (b) The pointing trajectory on the projection screen.

and orientation of the projection screen. Figures 3 shows the result of writing a "W".

Figures 4 shows the result of using the *finger-orientation* mode to write a "Y". In this mode, user's finger direction should be as perpendicular as possible with the baseline of the left and right cameras. If the direction of the finger in the image is almost parallel with the epipolar line, it will cause some difficulty in determining 3D finger direction. There is no such limitation for the *eye-to-fingertip* mode because only the fingertip is used to compute the 3D finger direction by use of a pre-determined eye position.

6. Conclusions

In this work, we have developed a system for human-computer interaction, namely, the free-hand pointer, by using the finger tracking technique based on an active stereo vision system. We have proposed two modes when implementing the free-hand pointer—the *finger-orientation* mode and the *eye-to-fingertip* mode, and have found that the latter is preferred in most cases. The space limitation of the finger movement is removed by using a well-calibrated active stereo vision system. Our experiments have successfully demonstrated the feasibility of implementing a free-hand pointer for presentation using an active stereo vision system.

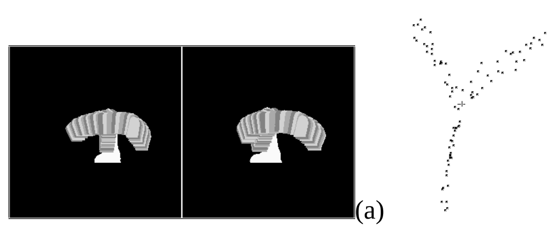
(a) (b)

Figure 4. (a) The image trajectory of the finger in the stereo image pairs using the *finger-orientation* mode to write a "Y". (b) The pointing trajectory on the projection screen.

Acknowledgements

This work was partially supported by the National Science Council under grant NSC85-2213-E-001-016 and NSC86-2745-E-001-007.

References

[1] T. Baudel and M. Beaudouin-Lafon. Charade: Remote control of objects using free-hand gesture. *Communications of the ACM*, 36:28–37, 1993.

[2] W. T. Freeman and C. D. Weissman. Television control by hand gestures. In *Proceedings of International Workshop on Automatic Face-and-Gesture-Recognition, Zurich*, pages 179–183, 1995.

[3] M. Fukumoto, Y. Suenaga, and M. K. "finger-pointer": Pointing interface by image processing. *Computer and Graphics*, 18:633–642, 1994.

[4] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, Reading, MA, 1992.

[5] Y. Iwai, K. Watanabe, Y. Yagi, and M. Yachida. Gesture recognition using colored gloves. In *Proceedings of the International Conference on Pattern Recognition, Vienna, Austria*, pages 662–666, 1996.

[6] V. I. Pavlovic, T. Sharma, and T. S. Huang. Gestural interface to a visual computing environment for molecular biologists. In *Proceedings of 2nd International Conference on Automatic Face- and Gesture-Recognition (ICAFGR), Killington, VT*, pages 30–35, 1996.

[7] J. M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *Proceedings of European Conference on Computer Vision, Heidelberg*, pages 35–46, 1994.

[8] S. W. Shih, Y. P. Hung, and W. S. Lin. Calibration of an active binocular head. *IEEE Transactions on Systems, Man and Cybernetics*, 28A(4), July 1998.