Interaction with Large Screen Displays Using Computer Vision and Depth Information

Tak Wai Derek Chan *Supervisor: Masahiro Takatsuka*
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

1. Introduction

Motivation

- Displays are becoming larger and more common, such as high definition televisions (HDTVs)
- Interaction using traditional keyboard and mice limit the capability of large displays.
- The most suitable modality varies with type of computer system, physical arrangement of displays and environment.
- A trend in current interaction technology is motion tracking e.g. the Kinect is used for interaction using body movements.

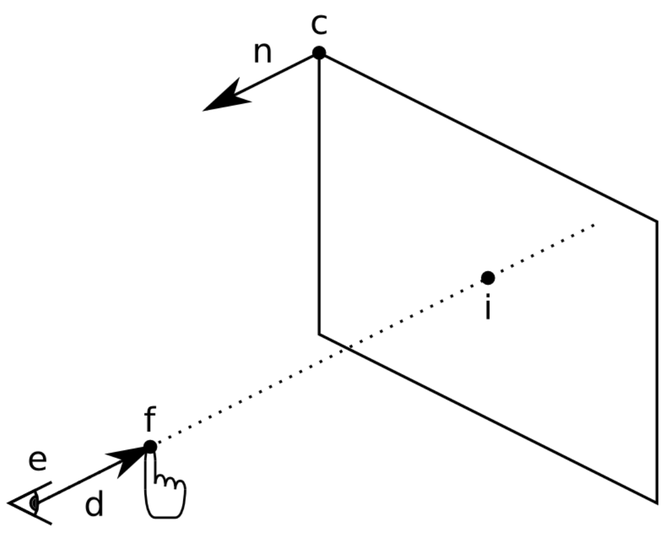
Aim

- How can video and depth data be used to accurately determine the location a person is pointing at?

2. Background

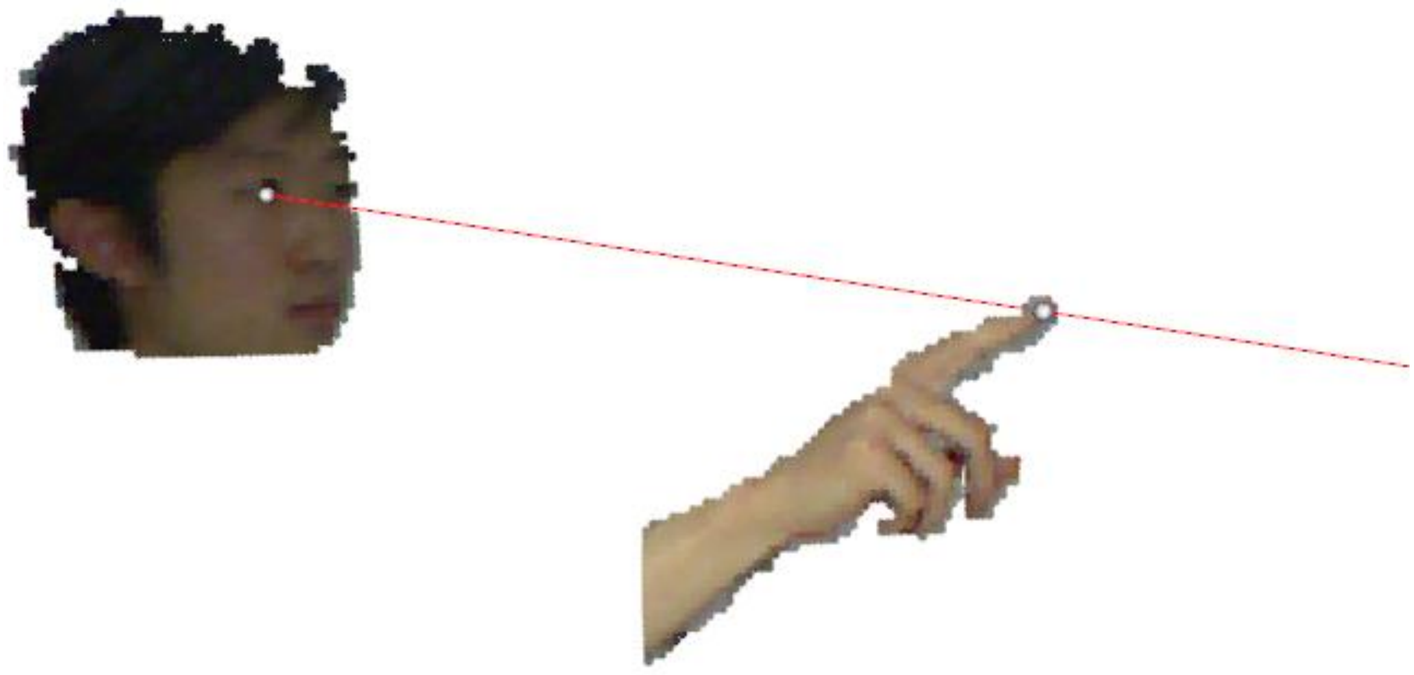
Image-Plane Pointing

- Image-plane pointing is a form of ray pointing technique, where the location a user is pointing at is determined by the intersection of a ray and an object, such as the plane of a display.
- The ray is defined by the position of the eye and fingertip.
- The fingertip will overlap the position of the screen where the event is generated, similar to using a touchscreen. This is a form of direct interaction, where the input space is the same as the display space.



Previous Work

- Cheng [1] used monocular computer vision to implement image-plane pointing. A head and eye detector was used to detect the position of the eye. Skin color segmentation was used to determine the position of the fingertip. The 3D positions of these points were approximated using environmental constraints, such as using the width of the face to determine depth.
- Gallo et al. [2] used the Kinect to implement image-plane pointing. The eye position was approximated using the position of the head, and the fingertip was approximated using the position of the head.
- The pointing location is not exactly collinear with the eye and fingertip in these approaches.

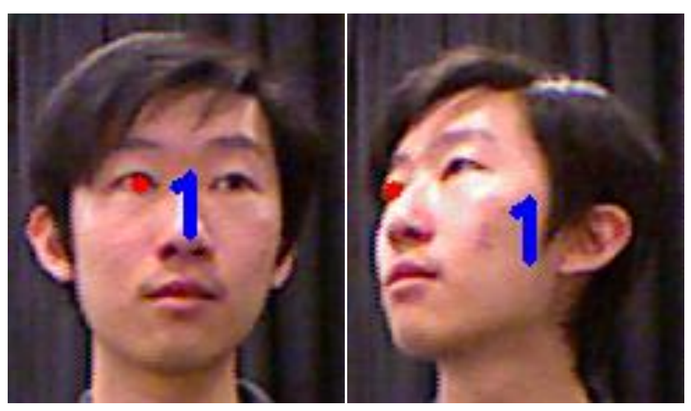


3. Fingertip Tracking

- We define the fingertip as the point on the hand furthest away from the elbow.
- This definition will find the most extreme point on the finger in normal use cases.

4. Head Tracking

- Head tracking is performed to track the position of the eye.
- A model of the user's head is aligned with the user's current head.
- Using optical flow and 3D information, the relative position and orientation of the head can be tracked continuously.
- Iterative closest point (ICP) is then used to correct the error in tracking.
- The position of the eye can then be selected and tracked even when it is not visible.

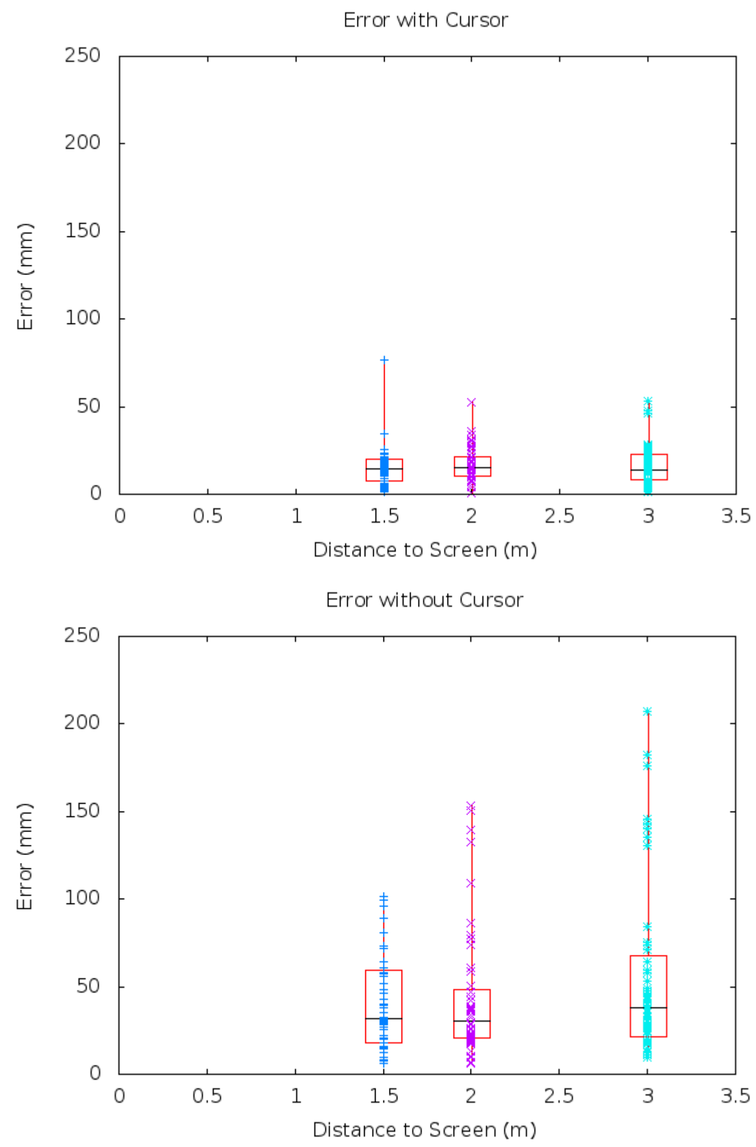


5. Head Model

- Multiple views of the head are registered to create a 3D model.
- User rotates head to create 3D model of front half of head.

6. Results

- The error is the difference between the system's measurement of the pointing location and the actual pointing location.



7. Conclusions and Future Work

- We successfully demonstrated the use of depth information for determining pointing location.
- Fingertip and eye position an be accurately tracked.
- The accuracy while using a cursor is still much better than without a cursor, suggesting the use of better calibration

Future Work

- Use of multiple depth sensors for increased usable area and automatic screen calibration.

References

1. K. Cheng. *Direct interaction with large displays through monocular computer vision*. PhD thesis, University of Sydney., 2008.
2. L. Gallo, A. P. Placitelli, and M. Ciampi. Controller-free exploration of medical image data: Experiencing the kinect. In Computer-Based Medical Systems (CBMS, 2011 24th International Symposium on, pages 1-6, June 2011.