

News Explorer

Using Kinect to explore how human body can be used as a pointing device

Anand Doshi, anandpd@umich.edu

University of Michigan, Ann Arbor, MI 48104

Abstract

This paper explores the use of Kinect 3D visual sensor to translate natural movements of a human body to move a cursor on a 2D screen.

Keywords

Kinect, 3D visual sensing, visual motion tracking

Introduction

Anyone who has used a computer is familiar with the pointing device known as a mouse or in many cases a trackpad. The mouse was invented by Douglas Engelbart's Augmentation Research Center at the Stanford Research Institute in the 1960s [1]. But before Apple commercialized the mouse and the graphical user interface, the mouse was a revolutionary yet untapped interaction device for the Xerox Alto [2]. The mouse transformed physical hand motion into virtual pointer motion. At a time when computers were primarily controlled using commands typed on a keyboard, the mouse opened up new ways of interaction.

In the same vein as the mouse, the Kinect represents one of the first commercial forays into marker-less visual motion tracking, and its technology could become as ubiquitous as the mouse. When compared to a mouse that senses motion in 2-dimensions, Kinect uses a camera and an infra-red depth sensor to capture the 2-dimensional pixel data, as well as depth information for each pixel, effectively giving us 3-dimensional co-ordinates for each pixel. Unlike the mouse, motion tracking is only one of the applications of Kinect. Also, Kinect is not limited to a small interaction space like the mouse. Depending on the sophistication of algorithms used to consume the sensor data, Kinect has the potential to make anything it "sees" as an interaction device.

The acquisition of PrimeSense by Apple [3], the Kickstarter campaign of Structure Sensor [4], Google's Project Tango [5], Leap Motion [6] and Microsoft's HoloLens [7] represent continued interest in 3D sensing technologies. 2D interaction devices like mouse, trackpad or joystick work well when the interface itself is 2D, but reach their limitations when used with virtual (VR) or augmented reality (AR) applications. However, 3D sensing technologies are not limited to 3D interfaces and interactions, but can also be used in interacting with 2D interfaces. Sooner or later, your laptop's web-camera and your smartphone's camera will be replaced with a Kinect like 3D

visual sensor, opening up a huge space for using natural interactions afforded by the human body, as well as the surrounding objects.

Anticipating this future, this paper explores the use of Kinect, a 3D visual sensor, to use a human body as a pointing device for a 2D interface.

Related Work

Pino et al. (2013) evaluated the performance of Kinect for 2D and 3D pointing tasks using Fitts' Law and concluded that Kinect performed marginally better than the mouse for 3D pointing tasks, while it performed poorly compared to a mouse for 2D pointing tasks [8]. However, performance alone does not determine whether a technology should or should not be used for interaction. Moving one's body to interact has experiential benefits, which are often lacking when we use a mouse either while sitting or standing.

Nebeling et. al (2014) used Kinect based gesture recognition and implemented 10 common browser features in Kinect Browser. It displayed two cursors on a large projected screen by independently tracking both hands [9]. The various gestures were based on pointing, movement of both hands relative to each other, speed of movement, and the open or closed state of each hand. However, these gestures involve large movements, and turn out to be less efficient and often uncomfortable compared to using a mouse. Leap Motion limits the range of motion, but allows the use of fingers to perform similar gestures more efficiently. At the same time, it still consumes more energy compared to a mouse, and tires the user after a short time. Instead of defining new gestures to control existing interactions, I see more merit in using gestures we are used to performing in our daily lives to be mapped to new interactions.

News Explorer

We often think of hand gestures, when we think of gestures in general. Most of the desired gestures for interaction require users to learn them. But, if we think about gestures that could possibly be sensed by a 3D sensor, it would include walking, standing, sitting, smiling, frowning, turning your head, shoulder, wrist and waist and body posture. We perform these gestures naturally, without thought. In the right context, like sitting in front of the TV and yawning, or walking in front of a large screen, or sitting in the living room adopting a thinking pose, these gestures could be used for interaction. While some of us may be wary of having multiple sensors in our vicinity, pervasive technology has become an accepted fact. We can thus take advantage of these natural, accustomed gestures to make our lives easier in specific cases.

News Explorer uses the location of a human body as a virtual pointing device. It uses the location data from Kinect to map the x-coordinate of the body to the pointer's x-coordinate and the distance of the body from Kinect to map to the pointer's y-coordinate. As you walk in Kinect's field of view, the black cursor on the screen moves across a world-map, highlighting countries as the cursor passes over them. When you stop walking and wait, and if the cursor has highlighted a

specific country, it triggers a search for news videos on Youtube related to that specific country, and starts playing one of the videos from the search results. The cursor stops moving if the video is being played. You can move around a bit while watching the video. However, the video stops as soon as you step out of Kinect's field of view. This has been used to resume body tracking after you are done watching the video.

The video feed from Kinect's camera, combined with the depth information from the infra-red sensor, can be used to identify objects, gestures and body postures, as well as tracking a human body abstracted as a skeleton. The position of various joints give us discrete data points making them effective for gesture recognition and position detection. However, the algorithm to generate a skeleton out of video and depth data is non-trivial. Microsoft provides an implementation of such an algorithm, but it is limited to use on Windows. PrimeSense, the company that helped make the Kinect, used to provide OpenNI and NiTE, which are a set of drivers and algorithms for making sense of Kinect's raw data and was compatible across many platforms. However, after its acquisition by Apple, these softwares remained largely unmaintained. Thus, skeleton data is not readily available when using Kinect on the latest version of MacOS. As a workaround, I have used the Average Point Tracking algorithm, that is available as part of Open Kinect for Processing library [10].



In News Explorer, the XY coordinates and depth information is normalized and sent to a NodeJS web-server via Open Sound Control (OSC) protocol. The NodeJS web-server efficiently transmits this information to a web page using the WebSocket protocol, ensuring a lag-free experience. The web page consists of a world map and a circular cursor, rendered using D3.js JavaScript library. As

you move in front of the Kinect, your position is transmitted to the web page, which is then scaled to the screen's resolution and used to decide the next position of the cursor. A country gets selected when you and consequently the cursor stops moving. The country name is sent to the NodeJS server via the WebSocket. The server then uses the country name to search for news on Youtube using its search API. On receiving the search results, it randomly selects one of the results and sends the VideoID or the PlaylistID back to the web page via WebSocket. The web page then starts playing the video in an overlay on top of the world map. If you step out of Kinect's field of view, the overlay and the video are removed from the view.

Result, Reflection and Future

When using a mouse, its small movements are translated to larger movements of the pointer on the screen. However, in the case of News Explorer, large movements of the human body are translated to small movements of the pointer on the screen. This is counterintuitive to what we are used to and does not make for a good user experience. However, this could make for a good experience when used on a large screen or in virtual reality.

It was also difficult as a user to select countries that were smaller than the cursor. This was not due to the cursor size, but the minimum speed of the cursor when mapped to the movement of one's body. This could also be attributed to the lack of precision due to the use of Average Point Tracking algorithm instead of using skeleton data. The use of skeleton data could have provided more area for movement and tracking instead of relying on depth thresholding, and could have improved the precision. It would also afford additional gestures and helped to implement zooming and panning of the world map. This could make for a richer interaction and allow for displaying additional details on the screen.

With more companies investing in 3D sensing, there could emerge a technology which provides precise tracking of small body movements like that of fingers, eyes and orientation of the head. Until then, inertial sensors can be used to track such fine movements. Another possibility is the additional use of short range video and depth sensing to track finer movements, as seen in HoloLens [7]. Project Soli by Google [11] is also taking an interesting approach for tracking finer motion using radar technology. Sensors will be a prevalent part of lives and despite the concerns for privacy, they have the potential to augment our lives. As Golden Krishna said, "The Best Interface is No Interface" [12] and 3D sensing is one way of achieving it.

Conclusion

Kinect-like 3D sensing systems are likely to replace the camera we know today, and could be used as effective interaction devices, that use our natural body movements and gestures we are accustomed to, to create intuitive interactions and better user experience. Just like the touch-screens replacing the mouse in many use cases, 3D visual sensing might replace the touch-screens of today and will be an integral part of ubiquitous computing. Their use in 2D interfaces has been

proven inefficient. But inefficiency need not always lead to poor user experience, especially when there is no additional learning involved.

References

1. A. and M. R. W. Connected, “Father of the Mouse’0,” Mouse - Doug Engelbart Institute. [Online]. Available: <http://www.dougenelbart.org/firsts/mouse.html>. [Accessed: 21-Nov-2016].
2. M. Gladwell, “Creation Myth,” The New Yorker, Sep-2011. [Online]. Available: <http://www.newyorker.com/magazine/2011/05/16/creation-myth>. [Accessed: 21-Nov-2016].
3. C. Velazco, “Done Deal: Apple Confirms It Acquired Israeli 3D Sensor Company PrimeSense,” TechCrunch, 2013. [Online]. Available: <https://techcrunch.com/2013/11/24/apple-primesense-acquisition-confirmed/>. [Accessed: 21-Nov-2016].
4. “Structure Sensor: Capture the World in 3D,” Kickstarter. [Online]. Available: <https://www.kickstarter.com/projects/occipital/structure-sensor-capture-the-world-in-3d>. [Accessed: 21-Nov-2016].
5. “Tango,” Tango. [Online]. Available: <http://get.google.com/tango/>. [Accessed: 21-Nov-2016].
6. L. Motion, “Leap Motion,” Leap Motion. [Online]. Available: <https://www.leapmotion.com/>. [Accessed: 21-Nov-2016].
7. @hololens, “Microsoft HoloLens,” Microsoft HoloLens. [Online]. Available: <http://www.microsoft.com/microsoft-hololens/en-us>. [Accessed: 21-Nov-2016].
8. A. Pino, E. Tzemis, N. Ioannou, and G. Kouroupetroglou, “Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation,” Human-Computer Interaction. Interaction Modalities and Techniques Lecture Notes in Computer Science, pp. 358–367, 2013.
9. M. Nebeling, A. Huber, D. Ott, and M. C. Norrie, “Web on the Wall Reloaded,” Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces - ITS '14, 2014.
10. “Getting Started with Kinect and Processing,” shiffman.net. [Online]. Available: <http://shiffman.net/p5/kinect/>. [Accessed: 21-Nov-2016].
11. “Project Soli,” Project Soli. [Online]. Available: <https://atap.google.com/soli/>. [Accessed: 21-Nov-2016].
12. G. Krishna, The best interface is no interface: the simple path to brilliant technology. Berkeley, CA: New Riders, 2015.