Embodied interaction essay

Touching another world

Michael Persson

Abstract

In this paper is presented an account of the current depth-map camera based gesture interfaces, and perceived issues with their implementations in mobile systems. Instead, a digital hand anchored to nodes in a physical hand is suggested as a way to enable mobile gesture based control over a digital overlay of the real world.

Introduction

"Technological advancements and a better understanding of the psychological and social aspects of HCI have led to a recent explosion of post-WIMP interaction styles, in which many novel input devices draw on user skills in interacting with the real world." (Hornecker, 2011)

Last year, I wrote an essay in the field of human-computer interaction presenting a rather vague concept of an mixed reality interfacing system that would serve to translate the physical movements of the user into the digital world. I talked briefly about what kind of technology I expected this system to consist of, but being an early draft of a concept it was not my intent to more closely analyze the current state of technology or whether or not the contemporary trends in motion tracking and gesture recognition would serve as a good platform to develop an inherently mobile concept.

There is a number of current research projects in close proximity to this field, among others BeThere (Sodhi et al., 2013), OmniTouch (Harrison et al., 2011), and Leap Motion, all of which will be more thoroughly documented in the background section. The obvious problem with all of them, however, is their innate lack of mobility. Being (as so many other current iterations of gesture tracking) based on camera technology, there is a seemingly unbreakable requirement of forcing the hardware into the periphery of the user. This needs not be such a big problem in situations where the user stays in the same place, but the second he decides to move around in the physical world, the technology either needs to be everywhere, or in such a shape that it can be brought along without adding a burden.

In an environment with increasingly mobile requirements, it seems like a strange thing to focus so much time and effort on developing solutions which presuppose unwieldy,

expensive and sensitive hardware to function, when so much of the current human activity, social or otherwise, is positioned in the real world, on the go.

This paper intends to understand the current technologies and analyze their strengths and weaknesses, in order to arrive at a conclusion about whether or not the depth camera paradigm lends itself well to mobile use. In lieu of the current technologies, how useful would a non-camera based movement and gesture tracking system be in a real world implementation? How far-fetched is the idea of a mixed reality digital overlay on the user, mirroring his or her moves in the digital sphere?

Background

Mixed realities is a collective term for a number of simulation technologies, such as augmented reality and augmented virtuality. An augmented reality is a physical reality strengthened by digital means, while an augmented virtuality is the opposite, that is, a digital simulation strengthened by physical means. Ubiquitous computing is the idea of dispersing technology into all aspects of the physical world. Our homes, the nature, and even the clothes we wear will be (or are already) augmented with different kinds of computers in the purpose of a more interconnected and interactive environment. There are many different views on how ubiquitous computing should be implemented, but they all have one thing in common - the concept of information and interaction permeated in an information space. (Benyon, 2005)

A key component in the field of embodied interaction is arriving at a shared meaning, that is, to find ways to make the technology understand the modalities of its user. Where we have traditionally worked with technology in a linear fashion due to single-core processing units, computer mice and input markers have worked in a sufficient fashion to funnel our attention to one task at a time, at one location at a time. As the field of computing progresses, however, the overall power of the new hardware can to a greater extent support and exploit the higher complexity of multimodal input available, and we can see that a single cursor is a poor replacement for the rich variety of interactions that can be had with contemporary technology.

In order to define and analyze the most commonly used technologies, I have studied a

number of current iterations of the embodied interaction paradigm.

Kinect

The Kinect technology is the commercial standard for motion and gesture tracking. With software developed by a subsidiary of Microsoft for a range camera technology by Israeli developer PrimeSense, the Kinect was originally designed for the Xbox 360, but its potential for interaction design projects was quickly recognized and as such its source development kit was released for Windows computers half a year after its initial release. The Kinect is presented by a physical, webcam-like sensor which uses an infrared light grid to map a depth field independent of ambient lightning requirements. This depth field is subsequently used to track what is happening in the physical space, enabling the use of gesture recognition. The Kinect also has the capability for voice recognition and as such can be controlled by vocal means. With the advent of the new Xbox One, Microsoft has developed a new version of the Kinect hardware with improved resolution.

BeThere

BeThere (Sodhi et al., 2013) is labelled as a proof-of-concept augmented reality solution for ”3D mobile collaboration with spatial input”. In essence, it is a device which allows a remote user to generate and control a digital hand marker in a persistent autogenerated space to assist the local user in a physical environment through the viewport of a smartphone. It works through capturing the 3D shape of objects through depth sensors, giving the users a shared picture of the geometry of the work space. The prototype is enabled through the use of Kinect hardware, combined with a shorter range depth sensor called Optimra DS311. According to the designers, some perceived limitations with the design is the unwieldy size and weight of the depth sensors.

OmniTouch

OmniTouch (Harrison et al., 2011) is a wearable multitouch interaction device which is basically a shoulder mounted projector which projects digital interfaces on nearby surfaces such as papers, walls, tables or even the extremities of the user. The prototype has two parts; the first being the custom depth camera (since the Kinect was not able to properly assess objects and planes closer than 50 centimeters) which is used to determine and track

the space and the gesture interactions, and the second being the projector from which the digital interface is projected. The interaction is accomplished by using multi-touch hand gestures across the projections, where the depth cameras will recognize them and use them for input. The OmniTouch system is strapped on top of the shoulder of the user, but is also stabilized with straps around the waist.

Leap Motion

The Leap Motion is a soon-to-be-available commercial gesture based PC interface where a small, usb-connected box generates a spherical depth map which tracks objects inside of it. This lets the computer "see" the hands of the user, which makes it possible to translate the gestures and positions of the hands to corresponding actions and movements inside the computer interface. The depth map is generated by two cameras and three infrared light markers. While it is basically a small Kinect, the resolution is higher and the scan area is lower, making it better suited to track precise movements of hands, where the Kinect is built to track geometry on a bigger scale, for example an entire room.

Google Glass

Google Glass is a wearable computer currently under development by Google. It is an eyeglass frame with a small screen, controlled by a combination of voice recognition and a side mounted touch pad.

Discussion

In the background, I listed a number of current (or at the very least recent) iterations of the gesture recognition interfacing paradigm, all of which are using a kind of Kinect-style depth camera for the purpose of input. Intuitively, this feels like a good way to track and translate human movement into a language that a computer will understand, since it's a hands-free solution which does not prerequisite external artifacts to be worn on the body, and anyone in the allocated "area of effect" can use it largely without having to customize the system to new users.

The most glaring weakness of the current technology seems to be the fact that since it is mostly based on depth camera technology: it's fragile, unwieldy, and expensive. With the

most mobile iteration of the depth camera technology, the OmniTouch, you're actually walking around with something akin to an exoskeleton, requiring waist and shoulder straps just to lock it in place! While it is obvious that the prototype presents a functional proof-of-concept, how far away is a version that you can actually use without a harness? With an increasingly mobile social activity setting in the contemporary world, is it reasonable to design mobile technology that restricts your movement?

Cumbersome and fragile technology is not the only problem, however. An innate limitation of optical projections is the fact that we can't actually get tactile feedback from it. While the OmniTouch makes sure that we always get to touch a physical surface of some kind to interact with the technology, the other three mentioned artifacts does not. The Leap Motion has a blurb that goes *"So you can do everything without touching anything"*, implying that it is somehow a good thing to omit the sense of touch, while in fact, it seems to be a bad thing: *"Embodied skills depend on a tight coupling between perception and action."* (Dourish, 2004). In this quote, Dourish is referring to the work of Polanyi, where there is made a connection between the proximal and the distal, where the distal is experienced through tacit, proximal feedback. Basically, if we can't feel what we are doing, we are losing a big part of the experience, and we may (for a lack of a better word) feel "blind", or detached from the context. Of course the connection between proximal and distal doesn't necessarily need to be in the form of physical sensations, but in the case of interfaces it seems more reasonable to adopt tactile elements than to exclude them through incompatible technology.

So, if depth camera technology proves to be unsuitable for mobile use due to restrictive traits such as fragility, cost, and inability to generate tactile feedback, what kind of interactions could we use instead?

One important topic of discussion within embodied interaction is the idea of effort, or more precisely, the way effort is used to communicate importance. In the words of Lyons et al., *"When represented in a way that can be felt by a player, the cost of action or inaction can itself become the lesson."* (2012) it is implied that effort plays a pivotal role in how we choose to experience reality. With the advent of modern computing, or more specifically the internet, this choice seems to have become less pervasive in our everyday lives. The interaction with the internet doesn't actually seem to have an opportunity cost, which makes it increasingly hard to make informed judgments about which parts of the

information flow to actually focus on. By moving the technology back into the real world, could it be possible to reconnect the effort (and its benefits) to the quality of the interaction, rather than being overrun with the unbridled flow of information as we may experience it today?

Prototype

"Interactions with physical artifacts, as has been explored, often also implies a reaching through *those artifacts to a symbolic realm beyond."* (Dourish, 2004)

With the risk of taking Dourish a little less metaphorically than he may have intended, this section is intended to briefly present the interaction concept I have in mind. As the base of this prototype, I imagine three main components from which the system will form: a display unit, a control unit, and the digital overlay. The display unit would likely be a lightweight, worn monitor, such as the Google glass framework, to be worn whenever the user wants to be able to visualize the digital overlay. If we assume that a Google glasses-esque system would be used as the main computing unit, they would naturally also be used to contain all the required software for the client-side system to work. The control unit would be some form of gloves or bracelets that would be able to track the positions of the users hands with their fingers and joints. These joints would then be anchored to digital representations of the users hands, serving to translate their movements into digital space (imagine the standard motion capture technology used for cinema). The digital overlay is a colloquial term for everything that can be interacted with outside of the physical world, or the realm which the "digital hands" of the users would inhabit. The general idea is to generate the digital overlay as a basic physical simulation shaped as a basic simulacrum of the physical world. These three components would combine to form a mobile interface platform which could have a number of different application areas.

The main idea is to be able to interact with the digital world, knowingly or unknowingly, through interacting with the physical world. If you see a physical door handle, you could have a digital door handle at the same place, and by using the real door handle, you also use the digital door handle (since your "digital hands" always are situated in the same space as your physical ones). This could be used as a natural lock mechanism, where doors wouldn't open unless both handles were used simultaneously, and as such could replace traditional security measures such as keys. In this example is demonstrated the fact that

you do not have to actually see the digital door handle to be able to interact with it, and the tactility of the interaction would not be an issue since you are actually interacting with the physical door handle as well. Of course, the system would also allow for less opaque interactions, such as unlocking physical doors from an exclusively digital terminal.

With this digital overlay you could also play digital games (maybe throwing a digital frisbee, or having a digital chess set which you move around a field), but it should also be possible to augment physical games, such as tracking golf balls with electronic tags, or even seeing targeting reticules or live ball trajectory predictions during the swing.

While it would be rather time-consuming to populate this digital world with content, it could be something of a collaborative effort, where everyone can participate in its creation with free source development kits, using geographical markers to pinpoint the locations of the installations. Commercial companies could use the digital "canvas" to customize their establishments with whatever content they wanted.

The point of most technologies investigated in this paper is to create a multimodal rich interface through which to interact with technology, and for the purpose of such an interface, this prototype should have no problems to replicate their current functionality. Since the main modality is about tracking hand gestures in a three-dimensional space, it is not functionally different from the way you use depth cameras, at least not in the contexts of contemporary systems. There are some glaring technical issues to iron out, such as how to properly represent tactile sensations, and how to stay seamlessly connected to the digital overlay, the standards through which interaction with legacy technology would be enabled, and a number of other equally important considerations. At this early concept stage, however, it could be argued that the possibilities are more important than the limitations.

References

Benyon, D., Turner, P., & Turner, S. (2005). *Designing interactive systems: People, activities, contexts, technologies*. Addison-Wesley Longman.

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. The MIT Press.

Harrison, C., Benko, H., & Wilson, A. D. (2011, October). OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 441-450). ACM.

Hornecker, E. (2011). The role of physicality in tangible and embodied interactions. *interactions*, *18*(2), 19-23.

Lyons, L., Slattery, B., Jimenez, P., Lopez, B., & Moher, T. (2012, February). Don't forget about the sweat: effortful embodied interaction in support of learning. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (pp. 77-84). ACM.

Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., & Maciocci, G. (2013, April). BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 179-188). ACM.

Google Glass (n.d.). Retrieved June 9, 2013 from Wikipedia: http://en.wikipedia.org/wiki/Google_Glass

Kinect (n.d.). Retrieved June 9, 2013 from Wikipedia: http://en.wikipedia.org/wiki/Kinect

Leap Motion (n.d.). Retrieved June 9, 2013 from Wikipedia: http://en.wikipedia.org/wiki/Leap_Motion