Virtual Sensors: Rapid Prototyping of Ubiquitous Interaction with a Mobile Phone and a Kinect

Lauren Norrie
School of Computing Science
University of Glasgow
norrielm@dcs.gla.ac.uk

Roderick Murray-Smith
School of Computing Science
University of Glasgow
rod@dcs.gla.ac.uk

ABSTRACT
The Microsoft Kinect sensor can be combined with a modern mobile phone to rapidly create digitally augmented environments. This can be used either directly as a form of ubiquitous computing environment or indirectly as framework for rapidly prototyping ubicomp environments that are otherwise implemented using conventional sensors. We describe an Android mobile application that supports rapid prototyping of spacial interaction by using 3D position data from the Kinect to simulate a proximity sensor. This allows a developer, or end user, to easily associate content or services on the device with surfaces or regions of a room. The accuracy of the hotspot marking was tested in an experiment where users selected points marked on a whiteboard using a mobile phone. The distribution of the sample points were analysed and showed that the bulk of the selections were within about 13cm of the target and the distributions were characteristically skewed depending on whether the user came to the target from the left or right. This range is sufficient for prototyping many common ubicomp scenarios based on proximity in a room. To illustrate this approach, we describe the design of a novel mobile application that associates a virtual book library with a region of a room, integrating the additional sensors and actuators of a smartphone with the position sensing of the Kinect. We highlight limitations of this approach and suggest areas for future work.

Author Keywords
Rapid prototyping, virtual sensors, Kinect, mobile interaction

ACM Classification Keywords
D.2.2 Design Tools and Techniques: User interfaces

General Terms
Design, Experimentation, Human Factors

MobileHCI 2011, Aug 30-Sept 2, 2011, Stockholm, Sweden.

INTRODUCTION
Mobile interaction designers are encouraged to explore novel interactions that account for contextual information about a user, such as the position of the user in a room, in order to make interaction with devices more manageable. However, sensing equipment is costly and its configuration is expensive in time and effort. For spacial interactions, embedding sensors in a room can be infeasible or disruptive, making it difficult for researchers to test interactions 'in the wild'. Rapid prototyping is the process of simulating software design ideas quickly [9]. We propose a rapid prototyping framework that uses the Microsoft Kinect sensor to simulate spacial interaction in a room. The Microsoft Kinect provides a cheap and robust 3D sensor suitable for tracking humans in indoor settings. Using the Kinect to sense user behaviour is beneficial as it requires minimal equipment to sense many interactions and, as such, it is low cost and causes little disruption to the working environment. The sensor was released in 2010 and is being used by millions of consumers. Since the hardware and development environments are widely available, any system developed to work with the Kinect can be shared with many other users.

Virtual Sensor	Kinect Data
Proximity sensor	3D position
Accelerometer	Second derivative of 3D position
Pose sensor	Skeleton tracking
Occupancy sensor	User detection
Motion sensor	Skeleton tracking
Light sensor	Camera
Sound meter	Microphone array

Table 1. Examples of Virtual Sensors using Kinect data.

Our virtual bookshelf example below illustrates the simulation of a proximity sensor using 3D position data provided by the Kinect. Like a physical proximity sensor, the application detects when a user is standing in proximity of a target position and uses this to trigger a book library on the mobile phone. Simulating with virtual sensors allows for many interaction ideas to be explored early in the design stage. Limitations of this approach include jitter in the sensor readings and occlusion caused by humans and objects in the room; the fixed range of the camera also restricts its application. Therefore, a more reliable system would detect the position of a user using a physical sensor and the data from the Kinect could be combined for added contextual reliability. Table 1 suggests other examples of virtual sensors that would be possible with this system. An interesting advantage of connecting mobile devices to the framework is the potential to perform real-time user identification by synchronising device acceleration with hand position. The fine grained control of the device may compliment the coarse contextual data of the Kinect and allow for more interesting interactions to be explored.

The paper is structured as follows: firstly, we look at related work in the area of ubiquitous interaction; we then explain our virtual bookshelf mobile application that illustrates spacial interaction design using the rapid prototyping framework; then we describe an initial experiment for determining the practical value of the 3D position data provided by the Kinect for the purpose of this application, including the equipment required to run the experiment and analysis of the results; and finally, we discuss our conclusions and areas for future work.

RELATED WORK

Location-aware applications use the Global Positioning System (GPS) to track the position of a user; however, these applications are limited to outdoor conditions due to the nature of GPS [2]. Applications of indoor positioning are less applied despite users spending long durations in buildings, such as homes and offices, where may interesting interactions take place. Reliable positioning of a user within a room is difficult without embedding sensors in the environment or performing complex computations [3]. Vision techniques for user tracking can be used to detect the position of a user more easily and this is possible with the Kinect.

In our experiment, we interpret the hand position of a user as the position of a mobile phone. Hand tracking has previously been explored in LED light tracking with a web-camera [11] and also with the Wii controller, another example of a static position-based system that has been successful in the design of interaction [6], which uses Infra-red to track hand movement. However, these methods require a user to hold a controller in order to register hand movement. The advantage of the Kinect system is that multiple hands can be tracked without additional equipment. Furthermore, as the Kinect uses an Infra-red camera to perform depth sensing, the requirement of strict lighting conditions is avoided; a limitation of other camera-based methods [7].

Our system combines the Kinect data with an Android mobile phone to explore ubiquitous interaction. Rapid prototyping frameworks have already been developed for ubicomp using both Android and the Arduino microcontroller [4]; however, designers may require multiple microcontrollers to simulate interactions. GAIM [1] is a prototyping framework that abstracts input from video games controllers to ease the development of multi-platform games and which allows for a system to adapt to the dynamic availability of input hardware. Similarly, our framework for ubiquitous interaction could allow mobile designers to develop interactions with abstracted hardware choices, such as the implementation of spacial interaction, and may allow users to engage with a system using the hardware available to hand.

PROTOTYPE

To illustrate the process of rapidly prototyping spacial interaction using the Kinect, we describe the design of a virtual bookshelf application, as pictured in Figure 1. The application allows a user to place a virtual book library in a region of a room and explore this with a mobile phone acting as an augmented reality, peephole display [10]. In the final system, it is intended that the application will be triggered by a proximity sensor embedded into a physical bookshelf in the users' home. As this requires specialised hardware and is too disruptive for testing, the idea is prototyped using a virtual proximity sensor.

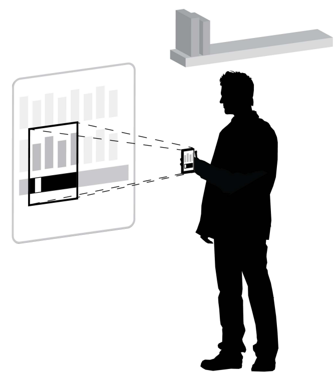

Figure 1. Virtual bookshelf prototype. The user stands in the position of a virtual proximity sensor which launches a virtual bookshelf application on a mobile phone.

The application has two modules: one receives data from the Kinect, including the current position of the user, and the other responds to user interaction with the mobile device. The Kinect application manages a table of virtual sensors that are added using a mobile phone. A mobile application can place a virtual sensor in the room by requesting to store an action at the current position of the user: to place the virtual bookshelf, we request to store the command 'bookshelf'. When the action 'bookshelf' is received, we know that the user is in proximity of our virtual sensor and launch the virtual bookshelf application.

This application would perform reliably when the user is in view of the Kinect. When the design stage is complete, the virtual sensor may be replaced or combined with a real sensor to increase performance. The sensors available in a mobile device could also be considered, for example, coupling the position with a bearing-based direction [8] from a magnetometer could be used to determine when the user is facing the wall.

Our simple message-based system for registering virtual sensors allows for many sensors and mobile applications to be registered using a single Kinect. The designer of the virtual bookshelf can quickly integrate spacial interaction into the mobile application and gain feedback on its use before committing to this part of the system. We gained feedback from mobile phone users about the virtual bookshelf prototype and received promising comments, though the idea of the application itself did not appeal to all. Some users did not see spacial interaction as an advantage for accessing their personal content. However, the potential of coupling the Kinect with a mobile device was highly inspiring and users were readily able to imagine their own ideas of applications: One user travelled often and wanted the application as a method of home security by being aware of the number of users that had been detected. Another user worked late and wanted to use the system as a social tool by revealing surprise messages to their parter in the home. In a business perspective, a user was interested in real-time user identification and how it could allow both shops and visitors to manage their visits automatically and build on the 'check-in' facility of foursquare, a location-based social network [5]. The system allows developers to integrate spacial interaction into their own mobile applications by using widely available hardware that would allow them to share their ideas with others.

EXPERIMENT

Design

The aim of this experiment was to determine the use of Kinect position data in practice by investigating the accuracy of users who were asked to select target positions with a mobile device. We ran the experiment with 3 participants of varied age, gender and height and we generated 120 sample positions for 2 fixed points.

Figure 2. Experiment setup. The participant is pointing the mobile device in front of Point 1 in the forward facing condition and is being tracked by the Kinect.

The setup of the experiment is shown in Figure 2. The following equipment was used in the experiment: a Microsoft Kinect sensor, an Android mobile phone, a laptop installed with the Primesense OpenNI software and a wireless network. The laptop was attached to the Kinect via USB and communicated with the mobile device over a wireless network. The Kinect sensor was an approximate distance of 3 meters from the whiteboard and was placed in a location such that its view was of the testing area. The whiteboard was marked with two points: Point 1 located at the intersection of the optical axis of the Kinect and the whiteboard, and Point 2 measured 1 meter left of Point 1. The distance of the Kinect scaled the position data to approximately 1px:2cm. To start the tracking process, participants were required to perform the OpenNI PSI pose, a stance where the arms are held at a 90 ° towards the Kinect, in order to calibrate the skeleton data.

Participants were asked to reach the mobile device in front of a point and press a button on the device; when the button was pressed, the position of the right hand was recorded from the Kinect skeleton data. Participants received vibrotactile feedback to confirm this selection. The experiment investigated two conditions: facing forwards towards the Kinect and facing towards the whiteboard. Using the right hand forced movement direction to change and this is noticeable in the results. Each task was repeated 10 times consecutively for each condition and both points.

Analysis

The graphs in Figure 3 display all point samples and show the separation between Point 1 and 2, indicating that these two points could be classified uniquely. These graphs also reveal noticeable differences between the forward and backward facing conditions: the direction of the hand movement caused the samples to be characteristically skewed. This effect was due to a communication delay between the phone and the laptop and has been improved with a more reliable communication protocol; we could analyse this further by synchronising the sampled positions with the accelerometer movement of the mobile phone.

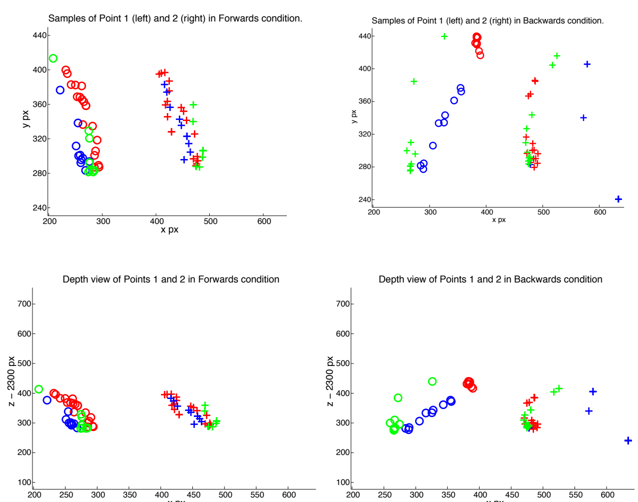

Figure 3. Each graph shows sample positions of Point 1 and Point 2. Top: (x,y), bottom: (x,z); Forwards condition: left, backwards condition: right. Each point is represented consistently by symbol and participants are represented by colour. Positions ranged between (0,0,0) and (640, 480, 3200) pixels. The graphs have been scaled to focus on the results.

The top graphs in Figure 3 show the distribution of sample points in the y-axis. The height of the user and the rate at which they perform the marking are factors of this variation. The bottom graphs illustrate hand movement in the z-axis. There were no significant effects of placing Point 2 at an angle from the Kinect as the two points can be classified uniquely.

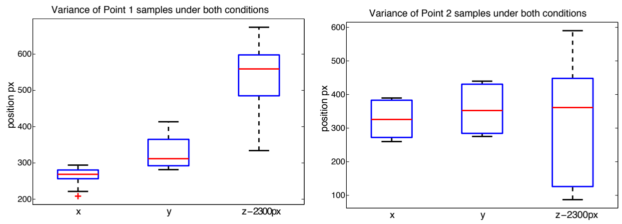

Figure 4. Variance between all samples of Point 1 (top) and Point 2 (bottom) under both forwards and backwards conditions. Samples are less varied for Point 1, which was central in the view of the Kinect.

The boxplots in Figure 4 display the variation of the 3D sample positions for each point. Samples were measured to be within a maximum distance of 13cm, suggesting that a range should be set in order to identify a position uniquely. This range would restrict the use of position data to applications requiring a point separation greater than this range.

Samples of Point 1 are noticeably less dispersed than those of Point 2. This was due to the central position of Point 1 as it allowed users to perform the marking comfortably while being tracked by the Kinect. Point 2 was at the edge of the Kinect's field of view and this forced users adjust their position in order to stay inside its viewing area. As a user exits the Kinect's field of view, the skeletal tracking becomes more uncertain about the location of each body part. The requirement of the user to have their hand in view was a restriction that could be avoided by using center of gravity data in place of the tracking process. This would gain a more robust approach of determining the position of a user but would lose the additional information of skeletal tracking about the position of each body part.

CONCLUSIONS

We demonstrated that the Kinect can be used to simulate a proximity sensor and we tested the variability of sampled positions when users targeted a point on the wall. The results indicate that the system is reliable enough to create compelling augmented reality systems, where the combination of mobile phone and Kinect are used to emulate a range of virtual sensors. With the framework, users can turn any surface or object in a room into digitally augmented ones, where the physical object or area acts as a mnemonic device for the content or service, as illustrated with our virtual bookshelf example. The ease of placing virtual sensors allows users to immediately customise a new room, such as a hotel or meeting room, to have augmented content that is accessible from any mobile device.

The limitations on this approach are common to most vision-based techniques. The static sensing area of the Kinect forces interactions to be bound to a fixed location and there are limitations on the range of depth sensing. Additionally, there can be problems with obstructions in the environment. Combining the inertial sensors of a mobile device would allow us to overcome some of these issues.

One way in which the framework could be extended is to allow interactions with multiple users and multiple devices, which allows much more complex and interesting interactions to be rapidly prototyped. Correlating the acceleration of each device with the motion of the Virskeletons sensed by the Kinect would allow immediate identification of users. Virtual sensors other than position-based sensors could also be implemented. Our framework will become available to developers to encourage the design of mobile interaction applications based on the low cost standard components involved and the sharing of novel interaction ideas.

ACKNOWLEDGEMENTS

We are grateful for funding from the EPSRC.

REFERENCES

1. Brehmer, M., Graham, T. C. N., and Stach, T. Activate your gaim: a toolkit for input in active games. Futureplay '10 (2010), 151–158.

2. Cheok, A. D., and Li, Y. Ubiquitous interaction with positioning and navigation using a novel light sensor-based information transmission system. *Personal Ubiquitous Comput.* (2008), 445–458.

3. Cohn, G., Morris, D., Patel, S. N., and Tan, D. S. Your noise is my command: sensing gestures using the body as an antenna. In *Proc. of Human factors* (2011), 791–800.

4. Kaufmann, B., and Buechley, L. Amarino: a toolkit for the rapid prototyping of mobile ubiquitous computing. MobileHCI '10 (2010), 291–298.

5. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In *Proc. of 2011 Human factors*, CHI '11 (2011), 2409–2418.

6. Schreiber, M., Wilamowitz-Moellendorff, M., and Bruder, R. New interaction concepts by using the wii remote. In *Proc. of Human-Computer Interaction.* (2009), 261–270.

7. Shoemaker, G., Tsukitani, T., Kitamura, Y., and Booth, K. S. Body-centric interaction techniques for very large wall displays. NordiCHI '10 (2010), 463–472.

8. Strachan, S., and Murray-Smith, R. Bearing-based selection in mobile spatial interaction. *Personal Ubiquitous Comput.* (2009), 265–280.

9. Tang, L., Yu, Z., Zhou, X., Wang, H., and Becker, C. Supporting rapid design and evaluation of pervasive applications: challenges and solutions. *Personal Ubiquitous Comput.* (2011), 253–269.

10. Veas, E. E., and Kruijff, E. Handheld devices for mobile augmented reality. MUM '10 (2010), 3:1–3:10.

11. Wachs, J. P., Kölsch, M., Stern, H., and Edan, Y. Vision-based hand-gesture applications. *Commun. ACM* (2011), 60–71.