Selecting and Controlling Physical Devices Through Head-Mounted Infrared Targeting

Anonymous for submission

ABSTRACT
Insert abstract

Author Keywords
smart devices; universal remote control; wearable computing; glass

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Figure 1. Teaser figure. Needs a caption.

INTRODUCTION
BH: What's the overall motivation for the project? What problem does it address? Increasingly, devices and services in our built environment are networked and can be controlled remotely. The proliferation of smart, controllable devices such as intelligent lighting, AV equipment, HVAC systems, or kitchen appliances raises the question of how to best interact with them. Prior work has introduced several approaches that use handheld mobile devices as *universal remote controls* to select and control appliances in the proximate environment- e.g., through pointing at such devices to select them, and through touch-screen or button interfaces to send control commands [2, 15].

BH: what are the drawbacks of this existing work? Some drawbacks of using handheld devices are that the device first has to be retrieved (e.g., from a pocket) and aimed; that two hands may be necessary for operation (one to hold the device, one to operate the touch screen); and that the user's visual attention is split between looking down at a screen and out at the device to-be-controlled. **BH: these all sound kind of weak. SC: When the number of smart devices increases, it also takes time to traverse within a list of dozens of items. In addition, mapping the name of a device to its real physical presence is not always intuitive. For example, it is hard to identify where is "light in area E" in an office building. Moreover, in a shared space, the person trying to control the device might not be the one that named it.** These general limitations have led researchers to investigate "always-available" physiological input modalities for phones on one hand [9], and wearable computing devices on the other hand.

In this paper, we introduce a novel method for selecting and controlling smart appliances in physical spaces though the use of a head-worn computing device with near-eye display and wireless communication. We augment Google Glass[1] with custom hardware for this purpose. Users first look in the direction of the device they wish to control to initiate interaction (e.g., at a lamp to control lighting, or at a speaker to change music playback volume). If multiple devices fall within communication range, an on-screen disambiguation dialog lets users select a single target. **SC: as well as led indicators on the devices?** Once acquired, a device specific control UI shown on the head-mounted display enables adjustment of discrete and continuous parameters through a touchpad interface (see Figure 1).

Our hardware relies on infrared communication between Glass and target devices to establish a connection; and on wireless ZigBee radio communication to exchange control messages. Glass is augmented with a narrow IR emitter and a ZigBee radio. Target devices similarly have IR receivers and ZigBee radios. These choices offer selection-through-head-orientation to initiate interaction; once initiated, users can look away from the target device while issuing control commands.

To understand the system performance, usability, and user experience of head-orientation targeting and head-mounted displays for device control, we first report measurements of the target range and accuracy of our device. We then conduct a comparative study of device acquisition time and error of different interface variants. We find that target acquisition through head orientation

[1] `http://www.google.com/glass/start/`

is faster than selecting items from a list, given the constraints of linear input using a head-worn touch controller. Finally, we report high-level feedback from *n* users who use our system for home automation tasks.

RELATED WORK

Relevant prior work exists in the areas of remote control of physical appliances, evaluations of pointing in physical space and augmented reality applications. We discuss each in turn.

Remote Control of Physical Appliances

Universal device control applications for smart phones are commercially available (e.g., Belkin WeMo[2]), though they usually do not offer spatial selection of target devices, forcing users to browse through lists of configured devices instead. Rukzio found that users strongly preferred either touching a mobile device to a target appliance or pointing at a distance to list browsing [13].

Several approaches to spatial selection with handheld devices exist to control appliances [2, 15, 17, 14] or to exchange information with smart infrastructure sensor networks [7, 8, 4]. Key design decisions are the method by which a target device is selected; and the method by which it is then later controlled or configured.

In several techniques, users select objects of interest with laser pointers. The laser dot provides immediate visual feedback to the user what is being selected. Beigl's early AIDA handheld combines laser pointing with IR communication to exchange commands [2]. Patel extends this technique by modulating the laser light to communicate the controllers' identity [15] to initiate radio communication. These proofs-of-concept do not include thorough evaluations. Kemp et al. use a laser pointer to indicate to robots which item to pick up in a room [6].

The XWand [17] determines its absolute position and orientation and uses a virtual room model to select target devices. Position is determined through two ceiling-mounted cameras; orientation is determined using a built-in IMU. Users can employ physical gestures or utter speech commands to control selected devices. This technique requires room instrumentation and an up-to-date virtual model of device locations. The Tricorder [7] uses IMU orientation coupled with room-localization based on received signal strength indicators (RSSI) to estimate what a user is pointing at.

Handheld projectors can both display a user interface in space and communicate control information optically, e.g., by encoding information temporally (using Gray codes in Picontrol [14] and RFIG [12]) or spatially (using QR codes in the infrared spectrum in SideBySide [16]). Printed tags like QR codes can also be affixed to devices and read by cameras. Common tagging systems are optimized to be read from a close distance, though it is possible to redesign codes that can be read further away (by encoding less information) [5].

Standard infrared remote controls for televisions and AV equipment are only meant to control a single device and it is unnecessary to have the user point precisely. These controllers tend to use wide-angle infrared LEDs.

Our main area of differentiation is that we employ head orientation as the selection mechanism instead of pointing - the user looks at the target device to initiate interaction. Selection techniques with very small selectors such as laser dots are less appropriate for head-mounted applications as head orientation only indicates a general area of visual interest. It does not necessarily match gaze orientation as extra-ocular muscles can move the eyes. We therefore select a source with a wider angle of illumination (an IR LED), but restrict its angle to be narrower than in general purpose IR applications.

Evaluation of room-scale selection

Pausch et al.'s early investigation of head-mounted displays compared head-tracking to handheld orientation control for a target acquisition task in a virtual reality room shown on a head-mounted display [11]. They found a clear performance benefit for head-tracking.

On the other hand, Card et al. experimentally determined that the bandwidth of neck muscles is much lower than that of arm, wrist or finger muscle groups [3], which limits the performance of any head orientation-based interaction scheme. However, many other factors such as device characteristics and device acquisition time (e.g., pulling a phone out of one's pocket) contribute to overall performance and preference of different selection techniques. **SC: Would like to add that we are only using head gesture for connecting, since we believe a user's attention is often drawn to the objects he intend to interact with. For further interactions, we still use hand gestures with heads-up-display.** Compared to a screen where every pixel is a potential target, the required accuracy for physical device selection in a room is much lower, and head orientation may provide sufficient accuracy.

Myers et al compared different methods of interacting with displays at a distance [10] and quantified selection time and jitter or position error when using remote handheld pointing. Various techniques outperformed laser pointers.

Our work is complementary as it provides concrete performance data on using head orientation to select targets in a physical environment.

Augmented Reality Interfaces

Agumented reality applications overlay digital information and graphics on the real world, e.g., through head-mounted displays [1] or other wearable devices. Our work is somewhat orthogonal to the research focus of this field as our device's graphics are shown in the

[2]http://www.belkin.com/us/wemo

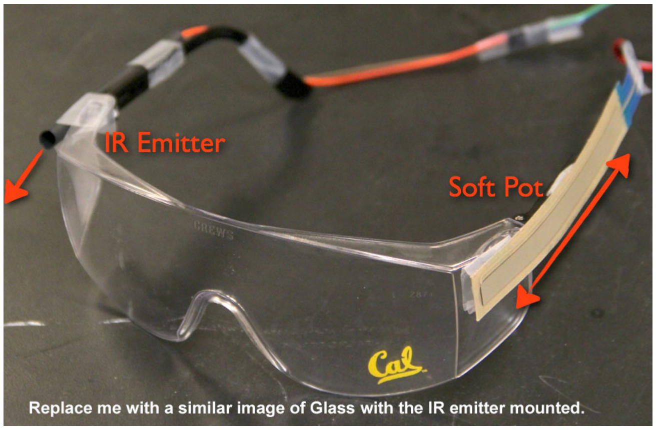

Figure 2. Our augmented Glass prototype has a frame-mounted infrared emitter.

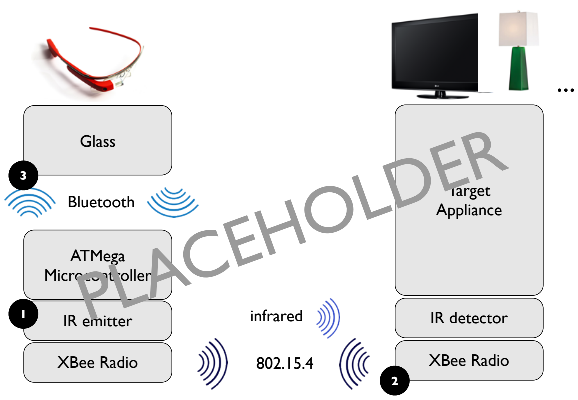

Figure 3. System architecture.

visual periphery; they are not referenced to particular objects in the world.

HARDWARE DEVICE AND INTERACTION

Our prototype consists of a Google Glass Explorer Edition head-worn computing device, augmented with an infrared emitter that is mounted on the frame, pointing out in the direction of the wearer's view (Figure 2). The IR emitter LED is mounted in an opaque hollow tube, that restricts the outgoing angle of illumination.

Interaction Flow

Look: Users select a target device by looking in its general direction. Glass periodically sends a device id through its IR emitter analogous to Patel's approach [15]. Target appliances have IR receivers and offer immediate visual feedback by toggling an LED whenever a valid id is received. They also respond by initiating an XBee radio connection and sending their id. This allows the Glass application to show which devices are currently targeted on the display. **SC: Currently the led is implemented but clients don't response with their IDs and display on Glass at this stage. We can add that. But we felt like the users would use led as the primary indicator and not to distract them with frequently changing the content on the screen?**

Initiate: Users confirm to initiate interaction by tapping on the Glass touchpad. The next section on disambiguation deals with cases in which multiple devices received valid IR signals. At this point, all further communication switches over to the XBee wireless network so that line of sight to the target is no longer needed (otherwise restricting head movement can become straining).

Control: Glass displays a control interface for parameters of the chosen device. **BH: Describe the interaction scheme for navigating and adjusting parameters here**. Control commands are sent over XBee radios.

Disengagement: **BH: how do you end an interaction? Is there a timeout if the user forgets to manually back out of the control screen? SC: currently the only way is to swipe back. A timeout mechanism can be added easily. I guess the design decision lies in what is an appropriate duration?**

Disambiguation

BH: describe why disambiguation is necessary - combination of the relative inaccuracy of head orientation and the spread of our IR signal. Then describe how to overcome it.

Prototype Implementation

In our prototype, Glass communicates over Bluetooth to an additional microcontroller board the user has to wear (Atmel ATMega256). This board marshals XBee to Bluetooth messages in both directions and also controls the IR LED mounted on the Glass frame (Figure 3). This architecture was mostly chosen for reasons of expediency. We selected XBee 802.15.4 radios to avoid the latency associated with connecting and disconnecting to Bluetooth devices **BH: Is this true or not? SC: Origainlly we didn't know we'd work on Google Glass, therefore BT didn't have a higher priority. We learned that bluetooth can concurrently communicate with up to 7 devices (while up to 255 can be paired but inactive), giving us more complexity when scaling up. In addition ZigBee has shorter wakeup delay and lower power consumption. (referenced here: http://superuser.com/questions/332767/limit-to-the-number-of-devices-that-can-be-paired-with-a-bluetooth-device)** Future head-mounted devices could clearly integrate IR emitters; the choice of local wireless technology could also change. In particular, one could substitute WiFi modules.

DEVICE CHARACTERIZATION

We determined the usable range and accuracy empirically with one IR emitter and two IR receivers. The IR emitter constantly sent out an id signal. The receivers that correctly received the signal turn their LED on for 300 ms.

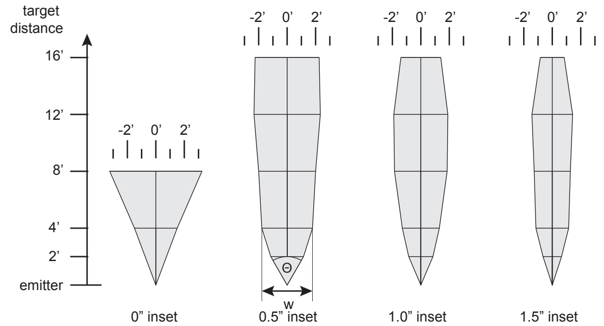

Figure 4. Our augmented Glass prototype has a frame-mounted infrared emitter.

distance/ depth	2′	4′	8′	12′	16′
0″	74°	78°	N/A	N/A	N/A
0.5″	60°	48°	28°	22°	16°
1.0″	46°	36°	26°	18°	10°
1.5″	36°	32°	18°	14°	6°

Table 1. Measured IR coverage angles Θ at different target distances and different depths of IR emitter inside shielding tube.

We placed all three devices at the same height with clear line of sight. The IR emitter is first places 2 feet away from the receivers. The receivers were moved sideways apart from each other until they could no longer receive stable signals. We then recorded the distance of the two receivers for the calculation of coverage angles. The steps are repeated for IR emitters in different distances (as shown in Table 1). We then repeated measurements with the emitter placed at various depths in the tube (see Figure 4).

In summary, **BH: describe what the measured results mean in practice.**

TARGET ACQUISITION EXPERIMENT

BH: text from Google Doc goes here

EXAMPLE SCENARIOS

BH: Here we should either describe a couple of example scenarios we've built, or get some students to run through realistic control tasks based on those scenarios - e.g.: "Dim the light to 20% and then start the movie, turn up the volume to 80%."

CONCLUSION AND FUTURE WORK

Our prototype does not yet have a general way of communicating device descriptions and capabilities - this means our prototype cannot yet control new, unknown devices - they first have to be modeled. We may extend the PUC work [**?**] to address this issue.

More importantly though, the main limitation is that extra hardware for IR communication is needed for the HMD and each controllable device. One potential approach would be to combine the growing availability of high-resolution indoor maps with live data from the HMD's camera to determine what a user is looking at without IR data exchange.

BH: add some general summary

REFERENCES

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. Recent advances in augmented reality. *IEEE Comput. Graph. Appl. 21*, 6 (Nov. 2001), 3447.

2. Beigl, M. Point & click-interaction in smart environments. In *Handheld and Ubiquitous Computing*, H.-W. Gellersen, Ed., no. 1707 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 1999, 311–313.

3. Card, S. K., Mackinlay, J. D., and Robertson, G. G. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst. 9*, 2 (Apr. 1991), 99122.

4. Costanza, E., Panchard, J., Zufferey, G., Nembrini, J., Freudiger, J., Huang, J., and Hubaux, J.-P. SensorTune: a mobile auditory interface for DIY wireless sensor networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, ACM (New York, NY, USA, 2010), 23172326.

5. Cross, A., Cutrell, E., and Thies, W. Low-cost audience polling using computer vision. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, ACM (New York, NY, USA, 2012), 4554.

6. Kemp, C. C., Anderson, C. D., Nguyen, H., Trevor, A. J., and Xu, Z. A point-and-click interface for the real world: laser designation of objects for mobile manipulation. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, HRI '08, ACM (New York, NY, USA, 2008), 241248.

7. Lifton, J., Mittal, M., Lapinski, M., and Paradiso, J. A. Tricorder: A mobile sensor network browser. In *Proceedings of the ACM CHI 2007 Conference-Mobile Spatial Interaction Workshop* (2007).

8. Mittal, M., and Paradiso, J. Ubicorder: A mobile device for situated interactions with sensor networks. *Sensors Journal, IEEE 11*, 3 (2011), 818–828.

9. Morris, D., Saponas, T. S., and Tan, D. Emerging input technologies for always-available mobile interaction. *Found. Trends Hum.-Comput. Interact. 4*, 4 (Apr. 2011), 245316.

10. Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., and Long, A. C. Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, CHI '02, ACM (New York, NY, USA, 2002), 3340.

11. Pausch, R., Shackelford, M. A., and Proffitt, D. A user study comparing head-mounted and stationary displays. In *Virtual Reality, 1993. Proceedings., IEEE 1993 Symposium on Research Frontiers in* (1993), 4145.
12. Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., and Willwacher, T. RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, ACM (New York, NY, USA, 2004), 406415.
13. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., and Chin, J. An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *UbiComp 2006: Ubiquitous Computing*. Springer, 2006, 87104.
14. Schmidt, D., Molyneaux, D., and Cao, X. PICOntrol: using a handheld projector for direct control of physical devices through visible light. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, ACM (New York, NY, USA, 2012), 379388.
15. Shwetak N. Patel, G. D. A. A 2-way laser-assisted selection scheme for handhelds in a physical environment. 200–207.
16. Willis, K. D., Poupyrev, I., Hudson, S. E., and Mahler, M. SideBySide: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, ACM (New York, NY, USA, 2011), 431440.
17. Wilson, A., and Shafer, S. XWand: UI for intelligent spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, ACM (New York, NY, USA, 2003), 545552.