Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces

Barrett Ens[1], Ahmad Byagowi[1], Teng Han[1], Juan David Hincapié-Ramos[2], Pourang Irani[1]

[1]University of Manitoba
Winnipeg, Canada
{bens, hanteng, irani}@cs.umanitoba.ca, byagowi@umanitoba.ca

[2]Lenovo R&T
Beijing, China
jramos4@lenovo.com

ABSTRACT

Current wearable interfaces are designed to support short-duration tasks known as micro-interactions. To support productive interfaces for everyday analytic tasks, designers can leverage natural input methods such as direct manipulation and pointing. Such natural methods are now available in virtual, mobile environments thanks to miniature depth cameras mounted on head-worn displays (HWDs). However, these techniques have drawbacks, such as fatigue and limited precision. To overcome these limitations, we explore combined input: hand tracking data from a head-mounted depth camera, and input from a small ring device. We demonstrate how a variety of input techniques can be implemented using this novel combination of devices. We harness these techniques for use with Spatial Analytic Interfaces: multi-application, spatial UIs for in-situ, analytic taskwork on wearable devices. This research demonstrates how combined input from multiple wearable devices holds promise for supporting high-precision, low-fatigue interaction techniques, to support Spatial Analytic Interfaces on HWDs.

Keywords

Head-worn display; HWD; HMD; naturalism; augmented reality; spatial interaction; analytic task; wearables

1. INTRODUCTION

Wearable technologies, such as watches, rings and head-worn displays (HWDs) are becoming commonplace, but their current interfaces are primarily designed to support micro-interactions: short bursts of activity, such as setting reminders or receiving notifications. *Spatial Analytic Interfaces* (SAIs) [4] have been proposed for moving beyond micro-interactions, to support *everyday analytic tasks* on HWDs. SAIs distribute information among multiple spatially situated information displays. These spatial layouts support efficient task switching via head motion [5], and their wearable platform allows tasks to be performed in-situ, when required for a particular situation. Supporting such everyday analytic tasks will require effective techniques for common operations such as selecting data and manipulating filter controls.

To support effective interaction in SAIs, we can draw from techniques developed for interacting in 3D and virtual environments [1, 16]. However, unlike in a lab setting where many such techniques are designed, a wearable interface must be made practical for mobile, *in-situ* use; users must not be encumbered by bulky input devices. One option for eliminating hand-held devices is to track the user's hands using a depth camera embedded in the HWD [7], to detect intuitive grasping and pointing gestures. Such natural techniques have been shown to provide advantages in many tasks [2], but are also prone to fatigue [8] and lack of precision [17]. Another wearable form factor recently gaining attention is the ring device [3, 9, 14, 20]. Finger-worn devices can provide subtle and precise input, and may be used with the arm in a resting position to reduce fatigue, but these lack the naturalism of intuitive gestures that hand-tracking provides.

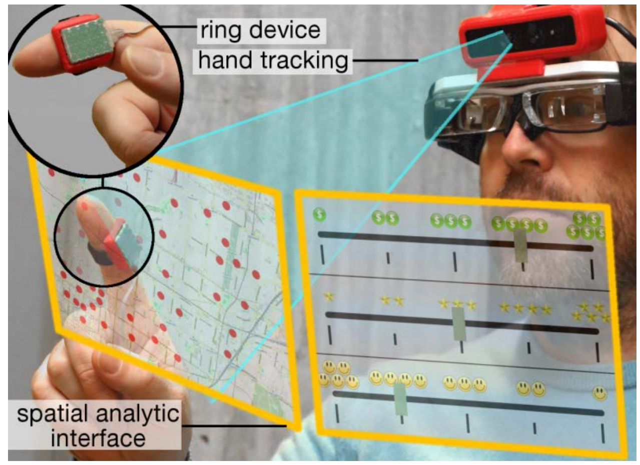

Figure 1. We combine input from a ring device with hand tracking by a head-worn depth camera to support interaction with Spatial Analytic Interfaces, which support everyday analytic tasks with multiple 2D views situated in 3D space.

In this paper, we propose the combined use of wearable, optical hand tracking with input from a ring device to support in-situ interaction with SAIs (Figure 1). We show how the combined benefits of these input devices (Table 1) can be used to support naturalism, while also allowing precise selection and manipulation with reduced fatigue. We also aim to support *dual-tier* interaction for SAIs [4]. Dual-tier interaction supports two levels: manipulation of window layouts, for instance to place two views side-by-side for comparison, and interaction with window contents, such as selection of small data points. With both hand tracking and a ring device, users can seamlessly combine large, coarse gestures to manipulate multi-window layouts, with fine-grained input for

Table 1: A summary of the tradeoffs between hand tracking input and ring device input.

	benefits	drawbacks
hand tracking (grasping, pointing)	intuitive; fast, coarse motions	fatigue; limited precision
ring device-indirect (rotation, tapping, swiping)	precise; fast repetitions	requires device;

SUI'16, October 15–16, 2016, Tokyo, Japan.

 ISBN 978-1-4503-4068-7/16/10. $15.00.

DOI: http://dx.doi.org/10.1145/2983310.2985757

precise control of window contents. We envision this mixture of input methods to be used much like interaction with current hybrid, touchscreen laptops; direct input gestures may be used periodically for fast, convenient manipulation, while indirect input is applied over a longer duration, for precision and minimal fatigue.

2. HYBRID INTERACTION TECHNIQUES

Combining multiple input modes is a strategy that has been employed by researchers to overcome various difficulties in a number of interaction domains. For instance, conventional desktop input devices such as mice do not adapt well to large displays; moving the cursor over large distances requires either excessive clutching, or reduced precision due to high control-display (CD) gain [13]. HybridPointing [6] and ARC-Pad [8] use absolute, direct pointing to select a coarse region, followed by indirect, relative motion to acquire a target. An alternative technique, proposed by Nancel et al. [13], attempts to minimize mode-switching costs, by using head motion for coarse selection, followed by direct hand motion for precise interaction.

Hybrid techniques have also been applied in 3D environments. For example, to disambiguate the intended target of a raycast selection in dense 3D environments, the DepthRay and LockRay techniques [18] place a cursor at a discrete point on the ray, which can be controlled by moving the pointer along the ray axis. For the DepthRay technique, pointing and depth manipulations occur simultaneously, whereas the LockRay technique requires the pointer position to be 'locked' in place before the depth cursor can be manipulated. This temporal separation of the modes increases precision at a cost of time. Researchers have also explored techniques that combine natural hand gesture input with voice input [10, 15]. In virtual environments, the voice commands are a useful way to trigger mode switching, such as translation and rotation, while the hands are used for object manipulation.

These hybrid techniques, and many others, may described according to various dimensions, such as the number of input devices, and the nature of the different modalities (e.g. absolute vs. relative, direct vs. indirect, voice vs. hand). Another useful dimension to explore is the temporal relationship between modalities, as described in the framework of Vernier and Nigay [19]. Five potential relationships they described are shown in Figure 2. We use similar diagrams in the figures below to help depict the relationships of the various modalities used in our prototype implementations. Our goal is to show how a wide variety of rich interaction techniques can be created by combining ring device input with hand-tracking on a HWD. The techniques we demonstrate are aimed at providing natural input for use with SAIs, while improving precision and minimizing fatigue.

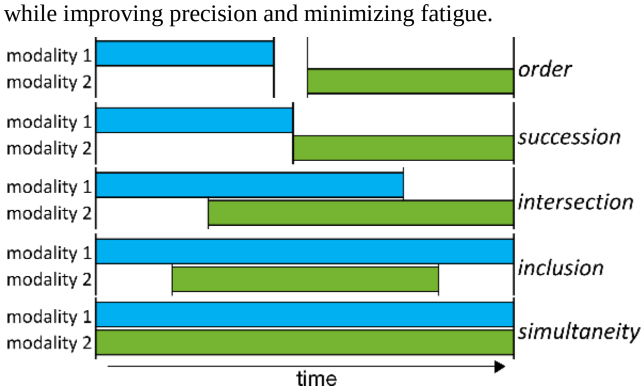

Figure 2. We apply the framework of Vernier and Nigay [19], which characterizes various ways to combine input modalities.

3. IMPLEMANTATION

To explore interactions that combine natural, direct input and precise, indirect input for SAIs, we built a prototype system that sends hand tracking and ring input data to a HWD. For hand tracking, we mounted a Softkinetic DS-325 depth camera on a Moverio BT-200 HWD (Figure 1). The camera input is processed by SoftKinetic's iisu middleware on a desktop computer, which connects to the Moverio via UDP. Transferred data include the hand centroid position, finger and thumb positions, pointing tip position, and hand state (open/closed). This setup is similar to upcoming commercial devices that contain depth cameras[1].

To supplement direct hand input, we use a small ring device (Figure 3) capable of basic and well-known touch gestures, such as tapping and flicking. The device contains a small capacitive touch sensor and a nine-axis inertial measurement unit (IMU). The capacitive touch sensor is composed of an array of capacitors arranged in a 3×4 grid on a surface measuring 12×16 mm. The capacitors are connected to a Microchip MTCH6102 controller, which sends position data with a resolution of 384×567, along with detected trackpad gestures. A Bosch BNO055 IMU module contains an Amtel ARM Cortex-M0 processor, and provides absolute pitch, roll and yaw. Both the touch control and IMU are interfaced using an I^2C bus, requiring only four wires for connection.

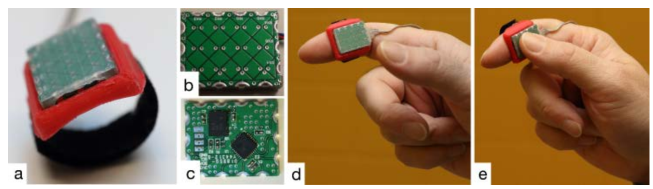

Figure 3: Our ring device (a) contains a miniature trackpad composed of an array of capacitive sensors (b) and a nine-axis IMU (c). When worn on the index finger (d), this small device supports simple techniques such as tapping (e) and swiping.

The unit is attached to a 3D-printed base, and affixed comfortably to the wearer's finger by a hook-and-loop fastener strip (Figure 3d). Data are relayed by Bluetooth to the HWD through a tethered Arduino microcontroller, worn on the wrist. Data filtering and all other processing are done onboard the Moverio unit, which runs Android 4.0. We developed the HWD program using Unity 3D.

4. SAI INTERACTIONS

We implemented several interaction techniques that leverage the benefits of direct and indirect input, drawn from a large body of available literature. We aim to show how these techniques can be supported in a wearable form factor to allow effective interaction with SAIs. We demonstrate these implementations using the novel combination of hand-tracking with a ring device, as described in the previous section.

For demonstration, we use a hotel search scenario as an example of an everyday analytic task. Imagine a traveler who arrives in a new city and needs to make a hotel booking. Given her immediate need for a room, she performs this search in-situ, while exploring the city centre. To assist in her search, she opens three windows containing a map, a filter panel, and a hotel preview panel. Viewing these through her HWD allows her to switch between views using head motion, more efficiently switching views on a smartphone. How might we also support effective dual-tier interaction with these multiple virtual window?

[1] For example, Microsoft HoloLens and Meta 2 track hand gestures: http://www.microsoft.com/microsoft-hololens/en-us http://www.metavision.com/

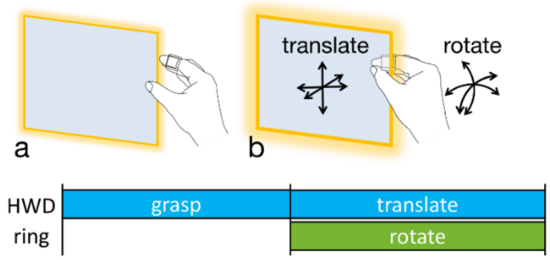

Figure 4. Direct manipulation is useful for infrequent, coarse-grained gestures. a) A user 'grabs' a window using a grasp gesture. b) Combining depth camera and ring IMU date allows the window to be translated and rotated freely in 6 DoF.

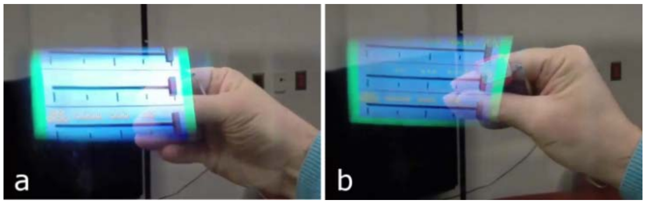

Figure 5. Our implementation, as viewed through the Moverio HWD showing direct manipulation of a window in 6 DoF.

4.1 6 DoF Direct Manipulation

Given a depth channel, human hands can be robustly segmented and the positions of centroids and fingertips are easily tracked. However, accurately measuring the absolute rotation of hands is more challenging, particularly if the hand is closed or fingers are otherwise occluded. Conversely, IMUs are capable of providing very accurate absolute orientation but cannot accurately track position. By combining the ring's IMU with the HWD's depth camera, we can enable tracking in 6 degrees of freedom (DoF).

In our implementation, this sensor combination allows windows to be freely manipulated in 3D space (Figure 4, Figure 5). After grabbing a window, by closing a hand around its virtual bezel, the user can position the window anywhere within reach.

4.2 Dual-Tier Selection

During an analytic task, users may apply various operations, such as moving windows to improve the layout, selecting data, or manipulating filter controls. Our implementation allows selection operations to be applied to both windows at layout level and to the contents at the container level. We provide two natural selection methods, for near and distant windows, respectively.

The first method uses absolute, direct input; the user simply 'taps' the window's virtual surface with an extended finger (Figure 6a, Figure 8a). The second method uses a virtual ray [12] projected from the user's hand. A tap on the ring enables the ray and a second tap selects a window (Figure 7a, Figure 8b), or disables the ray. With either technique, controls or objects closest to the point of intersection are selected. Once selected, either windows or contents can be used further by combining additional operations (Figure 6b, Figure 7b – see Control and Small Object Selection, below).

4.3 Control and Small Object Selection

Hand position tracking allows operation of virtual controls on a 2D interface; however, extended use of direct manipulation can quickly cause arm fatigue [2, 8]. Our system supports a combination of direct and indirect interaction methods. For instance, after selecting the filter panel, the user can cycle through the vertically-aligned sliders by swiping up or down on the ring pad and move the selected slider by swiping left or right (Figure 6b).

Analytic tasks may also require the selection of data points on dense visualizations. Even assuming that current methods allow sufficient precision, research has shown that input precision suffers without a haptic surface [17]. After a coarse selection, we disable any points outside a defined threshold and allow refinement by cycling through the remaining points (Figure 7b).

4.4 Ray-Grabbing

When 6 DoF manipulation is not ideal, the ring's IMU allows alternate 3D interaction techniques [1, 16]. For example, a user can 'ray-grab' [1] a window with a tap-and-hold gesture on the ring (Figure 9a, Figure 10a). The grabbed window can then be repositioned by mapping ring rotation to window translation in 2 DoF on a body-centric sphere (Figure 9b, Figure 10b). Lifting the thumb from the ring pad releases the window. The window can then be shifted along the third axis (depth) [16] by swiping up or down on the ring pad (Figure 9c, Figure 10c).

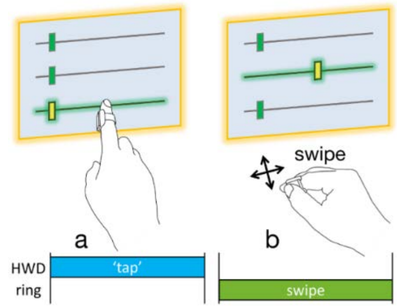

Figure 6: a) Selecting a virtual window with a 'tap' gesture also selects the nearest control within the window. b) Swiping gestures on the ring pad may be used to change the control selection and change the value of the selected control.

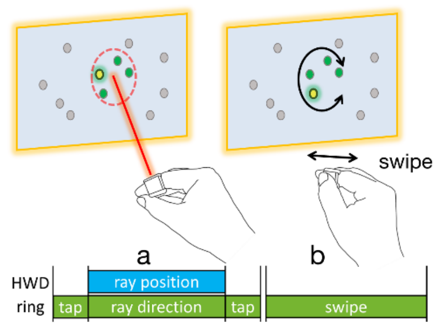

Figure 7. a) Items can also be selected with a virtual ray. Data points beyond a threshold distance from the ray selection point are disabled. b) Swiping the ring pad (horizontally or vertically) cycles through the enabled data points.

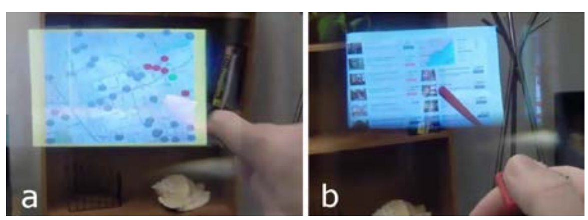

Figure 8. Our implementation uses the combined HWD and ring sensors to provide natural input methods. Users can select a window using a direct 'tap' gesture (a) on the virtual window surface, or by pointing a ray and tapping the ring pad (b).

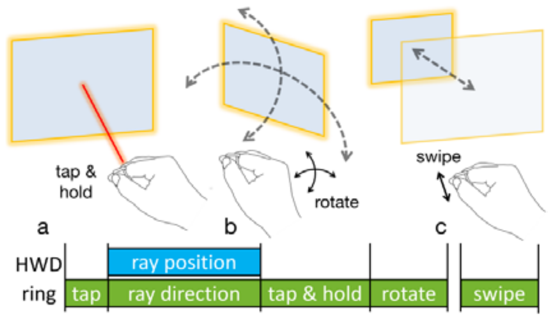

Figure 9. 'Ray-Grabbing' [1] constrains the window movement to two axis. a) The user taps and holds to select a window. b) Wrist rotation moves it along the surface of a body-centric sphere. c) The window's depth is manipulated using up and down swipes on the ring's touch pad.

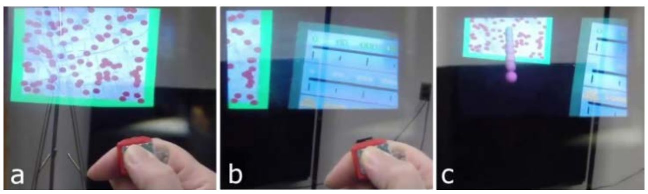

Figure 10. Raycasting is useful for interacting with windows that are out of reach. a) A user grabs a window (the ray is hidden, but stays connected), and places it beside another (b). c) The window is then moved in depth using swipe gestures.

5. CONCLUSION

We implemented a prototype system that combines ring device input with hand-tracking on a head-worn display. With this system, we demonstrate how a variety of interaction techniques can be created using various combinations of input values. These techniques are aimed at supporting natural interaction with SAIs, which require precise selection for everyday analytic tasks. Hand-tracking supports natural interaction using hand gestures. Ring input provides precision and can be used in a relaxed posture with low fatigue. The variety of available sensors allow a wide variety of combinations to develop a rich interaction language.

In future work we plan to develop a refined framework to describe hybrid interactions developed for SAIs, to help inspire new techniques. Further work will test these techniques in user studies, to determine which variations can best support seamless dual-tier control of both the window layout and content levels.

As we continue to develop an arsenal of techniques, we will apply these interactions effectively to analytic tasks. For example, we intend to develop features that help users switch their attention efficiently between coordinated views, and to design new controls for filtering and exploring data. These features must overcome the limitations of HWDs such as view size, and should be designed for in-situ use in mobile contexts. With this work, we hope to improve the usability of wearable systems and raise the bar beyond the productivity level of current mobile devices.

REFERENCES

[1] Bowman, D.A. and Hodges, L.F. 1997. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proc. I3D '97*, 35-ff.

[2] Bowman, D.A., McMahan, R.P. and Ragan, E.D. 2012. Questioning naturalism in 3D user interfaces. *Commun. ACM*. 55, 9 (Sep. 2012), 78–88.

[3] Chan, L., Liang, R.-H., Tsai, M.-C., Cheng, K.-Y., Su, C.-H., Chen, M.Y., Cheng, W.-H. and Chen, B.-Y. 2013. FingerPad: private and subtle interaction using fingertips. *Proc. UIST '13*, 255–260.

[4] Ens, B. and Irani, P. 2016. Spatial Analytic Interfaces: Spatial user interfaces for in-situ visual analytics. *IEEE Computer Graphics and Applications*, PP, 99 (Mar. 2016), 1-1.

[5] Ens, B.M., Finnegan, R. and Irani, P.P. 2014. The Personal Cockpit: A spatial interface for effective task switching on head-worn displays. *Proc. CHI '14*, 3171–3180.

[6] Forlines, C., Vogel, D. and Balakrishnan, R. 2006. HybridPointing: fluid switching between absolute and relative pointing with a direct input device. *Proc. UIST '06*, 211–220.

[7] Ha, T., Feiner, S. and Woo, W. 2014. WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception. *Proc. ISMAR '14*, 219–228.

[8] Hincapié-Ramos, J.D., Guo, X., Moghadasian, P. and Irani, P. 2014. Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. *Proc. CHI '14*, 1063–1072.

[9] Kienzle, W. and Hinckley, K. 2014. LightRing: always-available 2D input on any surface. *Proc. UIST '14*, 157–160.

[10] Lee, M., Billinghurst, M., Baek, W., Green, R. and Woo, W. 2013. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*. 17, 4 (Nov. 2013), 293–305.

[11] McCallum, D.C. and Irani, P. 2009. ARC-Pad: Absolute+relative cursor positioning for large displays with a mobile touchscreen. *Proc. UIST '09*, 153–156.

[12] Mine, M.R. 1995. *Virtual Environment Interaction Techniques*. *UNC Chapel Hill computer science technical report TR95-018 (1995)*, 507248-2.

[13] Nancel, M., Chapuis, O., Pietriga, E., Yang, X.-D., Irani, P.P. and Beaudouin-Lafon, M. 2013. High-precision pointing on large wall displays using small handheld devices. *Proc. CHI '13*, 831–840.

[14] Ogata, M., Sugiura, Y., Osawa, H. and Imai, M. 2012. iRing: intelligent ring using infrared reflection. *Proc. UIST '12*, 131–136.

[15] Piumsomboon, T., Altimira, D., Kim, H., Clark, A., Lee, G. and Billinghurst, M. 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. *Proc. ISMAR '14*, 73–82.

[16] Poupyrev, I., Billinghurst, M., Weghorst, S. and Ichikawa, T. 1996. The Go-go interaction technique: Non-linear mapping for direct manipulation in VR. *Proc. UIST '96*, 79–80.

[17] Szalavári, Z. and Gervautz, M. 1997. The Personal Interaction Panel – A two-handed interface for augmented reality. *Computer Graphics Forum*. 16, 3 (Sep. 1997), C335–C346.

[18] Vanacken, L., Grossman, T. and Coninx, K. 2009. Multimodal selection techniques for dense and occluded 3D virtual environments. *International Journal of Human-Computer Studies*. 67, 3 (Mar. 2009), 237–255.

[19] Vernier, F. and Nigay, L. 2000. A framework for the combination and characterization of output modalities. *International Workshop on Design, Specification, and Verification of Interactive Systems* (2000), 35–50.

[20] Yang, X.-D., Grossman, T., Wigdor, D. and Fitzmaurice, G. 2012. Magic finger: always-available input through finger instrumentation. *Proc. UIST '12*, 147–156.