Wireless User Interface Devices for Connected Intelligent Environments

Andrew D. Wilson
Microsoft Research
One Microsoft Way, Redmond, WA
awilson@microsoft.com

BIOGRAPHY

I am a research scientist at Microsoft Research. My present work involves designing novel user interfaces that exploit various sensing modalities, including wireless sensor packages such as accelerometers, as well as computer vision. One of the design goals is to construct simple input devices which, rather than trying to solve the entire UI problem in isolation, rely on the intelligence of the connected intelligent environment to correctly interpret the user's intention. Here I describe the XWand research prototype system, further described in (Wilson et al, 2003)..

INTRODUCTION

We present the XWand, a hardware device (Figure 1) and associated signal processing algorithms for an interface that may control multiple connected devices in a natural manner. The main idea is that the user should merely point at the device to be controlled, and use simple gestures or speech to control the device. The intelligent environment system interprets the user's manipulation of the wand to determine an appropriate action in context. The ultimate goal of such a natural interface is to provide an interface that is so simple that it requires no particular instruction or special knowledge to use, and instead relies on the intelligence of the environment to figure out what to do.

For example, the user may turn on a light in the room by pointing the wand at the light and pressing the button. Alternatively, the user may point the wand at the light and say "turn on". The user may then point the wand at the stereo amplifier and roll clockwise or counter-clockwise to turn the volume up or down.

HARDWARE DEVICE

We have constructed an early hardware prototype of the XWand, a handheld device which embeds a variety of sensors which in combination support pointing and gesture recognition tasks (Figure 1). The XWand has the following features:

- Analog Devices ADXL202 2-axis MEMS accelerometer. When motionless, this senses the acceleration due to gravity, and so can be used to sense the pitch and roll angle of the device.
- Honeywell HMC1023 3-axis magnetoresistive permalloy magnetometer. This senses the direction of the Earth's magnetic field in 3 dimensions, and can be used to compute the yaw angle of the device.
- Murata ENC-03 1-axis piezoelectric gyroscope. This is an angular rate sensor, and is placed to sense motion about the vertical axis (yaw).
- BIM 418MHz FM transceiver (38kbps). The transceiver is used to send and receive digital information to a matching base station, which then communicates to a host PC via RS-232.
- PIC 16F873 flash-programmable microcontroller running at 20MHz. The microcontroller reads each of the sensor values, formats data communication packets, decodes received packets, controls timing, power management, etc.
- Infra-red (IR) LED. Invisible to the naked eye, this LED can be seen by cameras equipped with an IR pass filter. This is used to support position tracking of the wand.
- Green and red visible LEDs. These can be used to display status information.
- Pushbutton.
- 4 AAA batteries. Quiescent power when awake is approximately 52mA, less than 1mA while asleep.

The output of the accelerometer and magnetometer may be combined to compute the full 3-d orientation of the wand

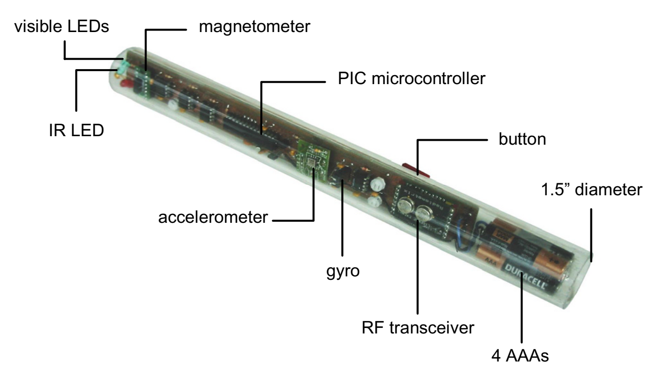

Figure 1: XWand hardware prototype.

with respect to the room. Computer vision techniques are used to find the 3-d position of the wand using the IR LED. We presently use a pair of calibrated cameras equipped with IR pass filters and some simple image processing techniques to triangulate the blinking IR LED

The orientation and position of the wand may be used to compute what the user is pointing at with the wand, given a 3-d model of the room and its contents. Objects may be placed in the 3-d model by using the wand itself, either by waving over the desired location, or by pointing at the object from different locations in the room (see Figure 2).

MULTIMODAL INTERPRETATION

We use a Bayes network to combine the outputs of various modalities and interpretation processes such as pointing targets, wand gestures, and speech, to arrive at a unified interpretation that instructs the system on an appropriate course of action. This framework decomposes the desired action (e.g., "turn up the volume on the amplifier") into a command ("turn up the volume") and referent ("amplifier") pair. Presently, the referent may be determined from the wand pointing target or speech recognition events, while the command may be specified by wand gesture, a button press event, or a speech recognition event. With this command/referent representation, it is possible to effect the same action in multiple ways. For example, all the following actions on the part of the user will result in a light turning on:

- Say "turn on the desk lamp"
- Point at the lamp and say "turn on"
- Point at the lamp and perform the "turn on" gesture
- Say "desk lamp" and perform the "turn on" gesture
- Point somewhere closer to the desk lamp than the floor lamp and say "lamp" and perform the "turn on" gesture
- Point at the lamp and click the button

The wand pointing target is determined by combining the 3-d orientation and position of the wand with a 3-d model

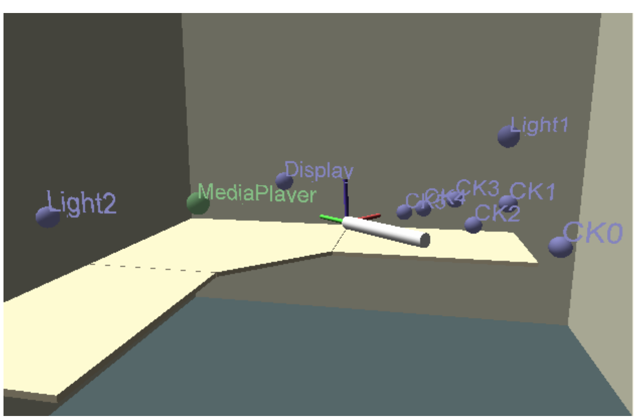

Figure 2: A 3-d graphics visualization of the wand world model with several trained targets in an office space. The wand (foreground) is shown as a white cylinder and coordinate axes.

of the room and the objects within it.

DEVICE CONTROL

We have assembled a demonstration of the wand used to control a variety of devices in a living room-like scenario. The user may control the following with the wand:

- X10 lighting: Multiple lights in the room may be turned on and off by pointing and clicking, or uttering the phrases "turn on" and "turn off". The lights may be dimmed or brightened by gesturing down and up.
- Windows Media Player: Pointing and clicking starts the media player playing or pauses it. Rolling left or right changes the volume, gesturing up and down moves the previous and next tracks in the play list. "Volume up", "volume down", "next" and "previous" utterances are mapped appropriately.
- Cursor control: Pointing and clicking at the computer display gives control of the cursor to the wand, with the wand button taking the function of the left mouse button. Clicking on a special button in the corner of the display exits cursor control mode.
- Color Kinetics lights: Pointing at these special computer controlled arrays of red, green, and blue lights brightens them over time. Rolling left and right changes the red, green and blue combination sent to the selected light, changing the light's color. When the user points away, the color gradually decays.

DISCUSSION

The XWand system is designed around a general purpose input device, with sensors chosen to support pointing and gesture recognition. Only with the centralized interpretation of the XWand sensor values in the context of the geometric model and application model does this general interface become useful.

It is interesting to note that users of the XWand focus their attention on the devices under control rather than the UI device itself. Contrast this behavior with typical use of today's universal remote, which demands the user's attention for all but the simplest interactions. Users have referred to the XWand as the "one-button universal remote."

I believe that the approach of providing a simple interface whose function is based on sophisticated notions of application and user context may solve some of the problems associated with the complexity of tomorrow's ubiquitous computing environments. Already we find collections of devices with interfaces that overlap in frustrating ways. For example, text entry into a cell phone is no easier when the phone is next to a full keyboard. While Bluetooth promises to solve this particular combination of devices, it still does not handle the association of the two devices easily. Put another way, in this case Bluetooth provides the communications link, but not the interaction. A user may be more likely to suffer the difficulty of using the phone for a short text entry rather than go through the trouble of using Bluetooth.

The challenge of the approach of the XWand system and other systems which rely on intelligent environments is in specifying the nature of the infrastructure itself. What capabilities does the intelligent environment require for a given set of interactions? Will these interactions be worth the trouble and expense of installing this infrastructure? How does a user know what abilities a given environment has, when those abilities may vary from environment to environment?

Clearly, the weight of the infrastructure is related to the nature of the interaction. After developing the XWand system, we pursued the question of what abilities, if any, did the system have if we removed the two cameras? The cameras raise various problems, including:

- Two or more cameras must be permanently mounted in the room. Besides the difficulty of installation, such cameras inevitably draw objections related to privacy.
- The cameras must be carefully calibrated to the room geometry upon installation, and recalibrated if they are moved.
- At least two cameras must have clear sight-lines to the wand at all times.
- The three dimensional position of each active device in the room must be known.
- Small errors in the orientation and position information translate to inaccuracy in pointing, possibly disrupting the interaction.

After conducting a simple user which analyzed the impact on pointing performance with and without audio feedback in pointing, and with and without cameras to provide precise position information, we developed the WorldCursor device, which eschews the original cameras, and adds a small tele-operated laser to the intelligent environment (see Figure 3). This laser provides feedback as to where the system believes you are pointing, much in the same way that the mouse cursor and mouse device work together. This system has a number of interesting advantages over the original XWand system, including a higher degree of precision in pointing, which enables seamless integration of the laser cursor with the Windows desktop cursor, for example.

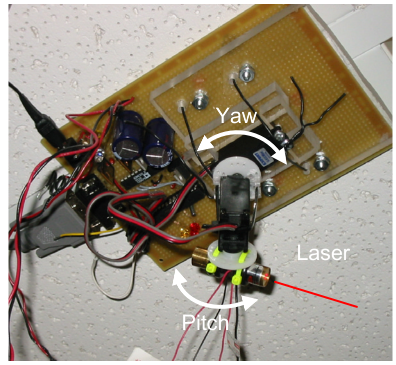

Figure 3: The WorldCursor device.

Our development of the WorldCursor device as a response to the original XWand system highlights the value of thinking seriously about the nature of feedback in intelligent environments, how that in turn impacts the nature of the infrastructure required, and how multiple devices may work together to enable new user interfaces.

REFERENCES

Wilson, A. and S. Shafer (2003), XWand: UI for Intelligent Environments, *in CHI 2003*.

Wilson, A. and H. Pham (2003), Pointing in Intelligent Environments with the WorldCursor, *in Interact 2003*.