Demonstration of the XWand Interface for Intelligent Spaces

Andrew D. Wilson and Steven Shafer
Microsoft Research
One Microsoft Way, Redmond, WA
awilson, stevensh@microsoft.com

ABSTRACT
The XWand is a novel wireless sensor package that enables styles of natural interaction with intelligent environments. For example, a user may point the wand at a device and control it using simple gestures. The XWand system leverages the intelligence of the environment to best determine the user's intention. We demonstrate the wand hardware device and multimodal (wand and speech) interpretation process in variety of device control scenarios, including lighting, media player, and cursor control.

INTRODUCTION
Increasingly our environment is populated with a multitude of intelligent devices, each specialized in function. The modern living room, for example, typically features a television, amplifier, DVD player, lights, and so on. In the near future, we can look forward to these devices becoming more interconnected, more numerous and more specialized as part of an increasingly complex and powerful integrated intelligent environment.

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

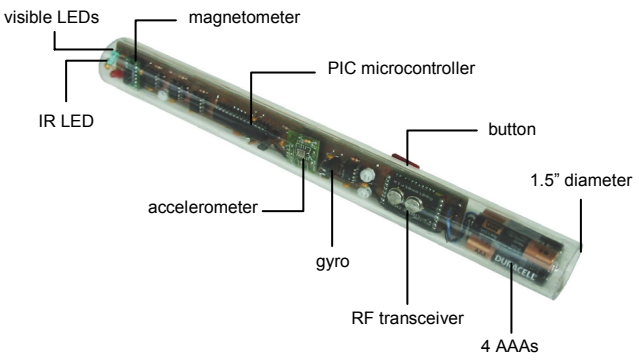

Figure 1: XWand hardware prototype.

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 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.

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 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 (**Figure 2**). 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.

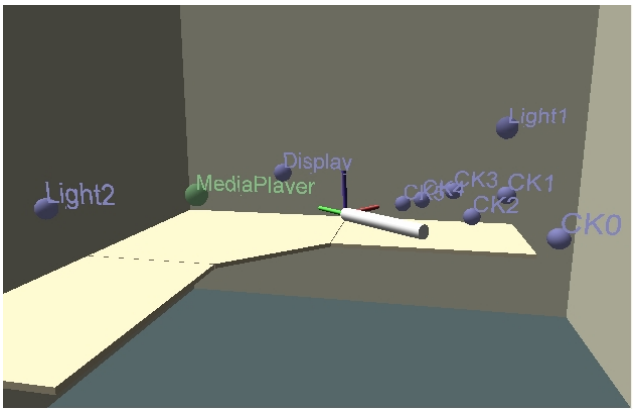

Figure 3: A 3-d graphics visualization of the wand world model with several trained targets in an office space. The wand (foreground) is shown as a white cylinder and coordinate axes.

Figure 2: Controlling the Media Player, with X10 controlled lights and video camera shown.