The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring

Thad Starner, Jake Auxier, and Daniel Ashbrook
College of Computing, GVU Center
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
{thad,jauxier1,anjiro}@cc.gatech.edu

Maribeth Gandy
Interactive Media Technology Center
Georgia Institute of Technology
Atlanta, GA 30332-0130 USA
maribeth.gandy@imtc.gatech.edu

Abstract

In this paper we present a wearable device for control of home automation systems via hand gestures. This solution has many advantages over traditional home automation interfaces in that it can be used by those with loss of vision, motor skills, and mobility. By combining other sources of context with the pendant we can reduce the number and complexity of gestures while maintaining functionality. As users input gestures, the system can also analyze their movements for pathological tremors. This information can then be used for medical diagnosis, therapy, and emergency services.Currently, the Gesture Pendant can recognize control gestures with an accuracy of 95% and user-defined gestures with an accuracy of 97% It can detect tremors above 2HZ within $\mp$*.1 Hz.*

1. Introduction

The Broadband Institute's Residential Laboratory at Georgia Tech [1] has been constructed to allow researchers to develop and evaluate technologies for the homes of the future in a real-life setting. One focus of research in the house is the Aware Home project [8]. The goal of this work is to sense, recognize, and interpret the activity of those living in the house, thus creating a home that is aware of its occupants. This awareness results in a living area that can assist the human residents, can anticipate their needs, and can improve the quality of their lives.

To sense the activity in the house, efforts are being focused on creating an infrastructure of sensors and computing within the house, from floors that can identify those who walk on it [11], to RF transmitters that can provide resident location information [9], to cameras and microphones in the ceilings to recognize and track people in the house [3]. While a ubiquitous computing architecture built into the house is necessary for many applications, we are also interested in placing some of this sensing and computing power on the body. By using wearable computing on the body to interface with the technologies in the environment of the house we increase the functionality, portability, and privacy of services available to the residents. For example, the house contains an infrastructure of cameras, which are and will be used to recognize people, track their movements, and observe their activities. However, problems due to occlusion and lighting can be minimized if a camera and some computing capacity are also placed on the bodies of the residents.

The goal of the Gesture Pendant is to allow the wearer to control elements in the house via hand gestures. Devices such as home entertainment equipment and the room lighting can be controlled with simple movements of the pendant wearer's hand.Building on previous work that used a wearable camera and computer to recognize American Sign Language (ASL) [15], we have created a system that consists of a small camera worn as a part of a necklace or pin. This camera is ringed by IR LEDs and has a IR pass filter over the lens (Figure 1.) The result is a camera that can track the user's hands, even in the dark. This is a design similar to the Toshiba "Motion Processor" project which uses a camera and IR LEDs as an input to a gesture recognition system for interaction with desktop and portable computers [16]. The Gesture Pendant video is analyzed and gestures are recognized. The wearer indicates that he/she wants the system to begin gesture recognition by pressing a button on the pendant. In this way we can avoid erroneous control information that might be generated by normal hand motions. The gestures are then used to trigger various home automation devices. In our current system, the user can control devices via a standard X10 network or a Nirvis Slink-E box that mimics remote controls for various consumer electronics. For example, the wearer can simply raise or lower a

[1] http://www.broadband.gatech.edu/

flattened hand to control the light level and can control the volume of the stereo by raising or lowering a pointed finger. By putting the sensing and computing on the body, this same pendant can be used to control things in the office, in the car, on the sidewalk, or at a friend's house. Privacy is also maintained since the wearer controls the video and the resulting data about their activities. The user does not have to worry about who is viewing the data and what is being done with it.

Figure 1. The Gesture Pendant

2. Motivation

But why do we want to use hand gestures to control home automation? Home automation offers many benefits to the user, especially the elderly or disabled; however, the interfaces to these systems are poor. The most common interface to a system such as X10 is a remote control with small, difficult to push buttons and cryptic text labels that are hard to read even for a person with no loss of vision or motor skills. This interface also relies on the person having the remote control with them at all times. Portable touchscreens are emerging as a popular interface, however they have many of the same problems that remotes have, with the additional difficulty that the interface is now dynamic and harder to learn. Other interfaces include wall panels, which require the user to go to the panel's location to use the system, and phone interfaces, which still require changing location and pressing small buttons. While speech recognition has long been viewed as the ultimate interface for home automation, there are many problems in this domain. First, in a house with more than one person, a speech interface could result in a disturbing amount of noise, as all the residents would be constantly talking to the house. Also, if the resident is listening to music or watching a movie, he/she would have to speak very loudly to avoid being drowned out by the stereo or televison. This ambient noise can also cause errors in the speech recognition systems. Finally, speech is not necessarily a graceful interface. Imagine you are hosting a dinner party and you want to lower the lights in the room. If you were using a speech interface you might have to excuse yourself from the dinner conversation and then loudly state a phrase such as "computer, lower lights to level 2." In this case the speech interface would be disruptive and non-ideal. However, if you had been using the gesture pendant, you could have continued your dinner conversation and simply lowered the light level to your liking by gesturing up and down with your hand. Further support for the validity of this gestural approach to user interface comes from the work of Harrison *et al*, where users made gestures that mimicked real life movements to control virtual objects [4]. For example, making a gesture that resembled turning a page turned the virtual pages on a palm computer. Therefore the user interface metaphor was the real world. In our work we are using this same interface philosophy not to control virtual objects, but to manipulate remote objects in the real world.

The gesture pendant can be used alone or in conjunction with various types of contextual awareness. By using various types of context the number and complexity of gestures can be reduced without reducing the number of functions that can be performed in the house. The following are some of the currently implemented and future configurations of the gesture pendant.

1. The gesture pendant alone. In this case the user defines a number of gestures to control home automation. The user defines a different gesture for each function. For example, if he/she wants to be able to change the volume on the stereo, the level of the thermostat, and the light level, three different gestures are required.

2. The gesture pendant combined with speech recognition: In this configuration speech is used as a modifier for a smaller set of user defined gestures. For example, the user can have one gesture to control the volume, the thermostat and the light. Before performing the gesture the user simply indicates the desired target by speaking its name. This multi-modal approach has already been implemented successfully in the QuickSet system, which enables users to create and position objects on a virtual map using speech and pen-based gestures [12].

3. The gesture pendant combined with pendant orientation: In the current system, the transmitter that sends out the IR codes to control the house devices is placed in the room(s) where the devices reside. However, we have also used the system with the IR transmitter place next to the pendant on the body. By making this transmitter part of the pendant, the user can indicate

the target for a gesture by aiming his/her body at the device. For example, the user may have the same gesture for controlling a stereo's volume, a thermostat's setting, and the room's lighting level by turning towards the appropriate device before gesturing. This results in fewer gestures but also requires the user to move and to be in the general location of the devices. Another problem arises if two or more devices are close together, in which case the user must use separate gestures to control them.

4. The gesture pendant combined with RF location: Another project that is a part of the Aware Home is a RF location system [9]. This system can determine with room level accuracy the location of house residents. In this configuration the location information could serve as the modifier. This would require the user to move to the room with the target device before gesturing. If multiple devices were located in the same room, different gestures would be required for each.

5. The gesture pendant combined with fiducial detection: By putting the IR transmitter on the body we could very crudely determine where the wearer is looking; however, by using fiducials on devices we could have a more precise determination. If the fiducials consisted of IR LEDs, the pendant camera could easily detect them. The use of IR transmitters for communication between the environment and a computer with a camera has already been demonstrated in the research of physical icons [10]. This configuration would still have some of the problems outlined in number 3, but would have more precision.

These various modifier technologies could be combined in any permutation to create a system that uses gesture along with sound, location, and/or orientation to control home automation. To date we have created the subsystems for configurations 1-4 and have begun experimentation in adapting our software to handle fiducials.

3. As an enabling technology

Obviously, all types of people can use the gesture pendant system. However, the focus of the Aware Home project is "Aging in Place" [8]. We feel that those who have the most to gain from the features of the Aware Home are the elderly. The technologies in the Aware Home can allow them to remain independent, living in their own homes longer. The elderly can have their health status monitored, and can have help with day-to-day tasks without losing their dignity or privacy. The features of many home automation interfaces that make them hard to use by healthy adults make them unusable by the elderly or disabled. An elderly person may suffer from Parkinson's disease, stroke, diabetes, arthritis and other ailments that can result in a reduced motor skills, reduced mobility, and/or loss of sight. Also, in an emergency, the resident might need the automation to assist them, but might be unable to speak. The gesture pendant can be used despite such impairments, enabling the resident to perform automation assisted tasks such as locking or opening/closing doors, using appliances, and accessing emergency systems.

The same interface problems are faced by those with disabilities such as cerebral palsy and multiple sclerosis. However, a study has shown that even people with extremely impaired motor skills due to cerebral palsy are able to make between 12 and 27 distinct gestures[14], which could be used as input to the gesture pendant. Therefore, we see the gesture pendant as an interface alternative that could allow people who are unable to use some of the more traditional interfaces to take advantage of the independence that home automation could afford them.

4. Medical Monitoring

As a user makes movements in front of the gesture pendant, the system can not only look for specific gestures, but we can also analyze how the user is moving. Therefore, a second use of the gesture pendant is as a monitoring system rather than as an input device. The parameter of movement that the pendant detects is a tremor of the hand as the user makes a gesture.

As discussed above, the target population for the gesture pendant is the elderly and disabled. Many of the diseases that this population suffers have a pathological tremor as a symptom. A pathological tremor is an involuntary, rhythmic, and roughly sinusoidal movement [2]. These tremors can appear in a patient due to disease, aging, and drug side effects; these tremors can also be a warning sign for emergencies such as insulin shock in a diabetic. Currently, we are interested in recognizing essential tremors (4-12 HZ) and Parkinsonian tremors (3-5 Hz)[2], since determination of the dominant frequency of the tremor can be helpful in early diagnosis and therapy control of such disorders [5].

The medical monitoring of tremors can serve several purposes. This data can simply be logged over days, weeks, or months for use by the doctor as a diagnostic aid. Upon detecting a tremor or a change in the tremor, the user might be reminded to take medication, or the physician or family members could be notified. Tremor sufferers who do not respond to pharmacological treatment can have a deep brain stimulator implanted in their thalamus [7]. This stimulator can help reduce or eliminate the tremors, but the patient must control the device manually. The gesture pendant data could be used to provide automatic control of the stimulator.

Another area in which tremor detection would be helpful is in drug trials. The subjects involved in these studies must be closely watched for side-effects and the pendant could provide day-to-day monitoring.

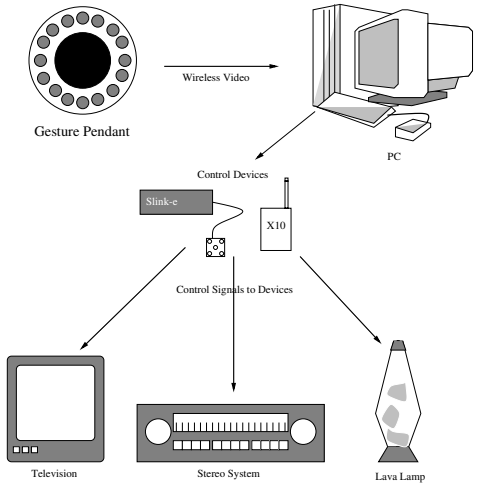

Figure 2. Experimental gesture pendant system design for home automation control.

5 Gesture Pendant Hardware

The motivation behind the Gesture Pendant called for a small, lightweight wearable device. At first we considered a hat mount, but concluded that gestures would be too hard to recognize if made in front of the body, and difficult to perform if made in front of the hat. Due to the off-the-shelf nature of the components (leading to larger size and heavier weight than ideal) we decided that a pendant form was the only reasonable one. Using custom-made parts, the hardware could be shrunk considerably, and other form factors such as a brooch or, assuming sufficient miniaturization, a shirt button or clasp could be possible.

Since the goal of the Gesture Pendant was to detect and analyze gestures quickly and reliably, we decided upon an infrared illumination scheme to make color segmentation less computationally expensive. Since black and white CCD cameras pick up infrared well, we used one with a small form factor (1.3" square) and an infrared-pass filter mounted in front of it (Figure 3.) To provide the illumination, we used 36 near-infrared LEDs in a ring around the camera. The first incarnation had a lens with a roughly 90 degree field of view, but that proved to limit the gesture space too much. A wider angle lens of 160 degrees turned out to work much better, despite the fisheye effect.

The eventual goal is to incorporate all components of the gesture pendant into one wearable device; however, for the sake of rapid prototyping we used a desktop computer to do the bulk of the image processing. This also allowed us to easily centralize the control system, by using standard peripheral home automation devices such as the Slink-E and X10. To send the video to the desktop, we used a 900 MHz video transmitter/receiver pair (Figure 2.) The transmitter is powerful enough that cordless 900 MHz phones do not interfere with it, and the receiver can be tuned to a range of channels to avoid conflicting signals from multiple pendants.

Because all of the components of the Gesture Pendant are off-the-shelf, it is currently power inefficient. The camera itself uses about 300 mW, and the LEDs can be assumed to use about 6 mW each. This, coupled with the camera's requirement for a power input of 12v, led us to use two Sony NP-F330 lithium-ion camcorder batteries in series with a step-down voltage converter. The transmitter requires 9v input, and to keep design changes simple, was just attached to a standard 9v alkaline battery. For future versions of the Pendant, we will use a single battery to power everything.

Almost all technology necessary to integrate the components of the Gesture Pendant system into a single unit exists today. There are now available general-purpose ARM processor-based computers which are smaller than a business card and yet have almost as much processing power as a desktop computer [3]. By creating a specific-purpose computer for this task, the size could be reduced even further. The major obstacle to total integration lies in the power requirements. While the ARM processor is very low power, many other components, such as the camera or video grabber board, have few low power analogs. The power needed for the total setup would require a large battery or frequent battery changes, neither of which is currently conducive to sustained wear.

Since one of the groups that we feel this device can be of the most use to is the elderly, it is important to make it as non-obtrusive as possible. This means it must be inconspicuous, lightweight, and non-complicated. Also, since the Gesture Pendant is a wearable device, and one that is constantly in full view of others, it will be important to make it more attractive. The ring of LEDs makes it appear somewhat jewelry-like, but with a smaller form-factor and some principles of design applied to it, it will become more appealing.

6. Gesture Recognition

The recognition system incorporates two kinds of gestures: control gestures and user-defined gestures. Control gestures provide continuous control of a device. This type of gesture provides continuous output while the gesture is

[3] http://www.brightstareng.com/arm/nanoeng.htm

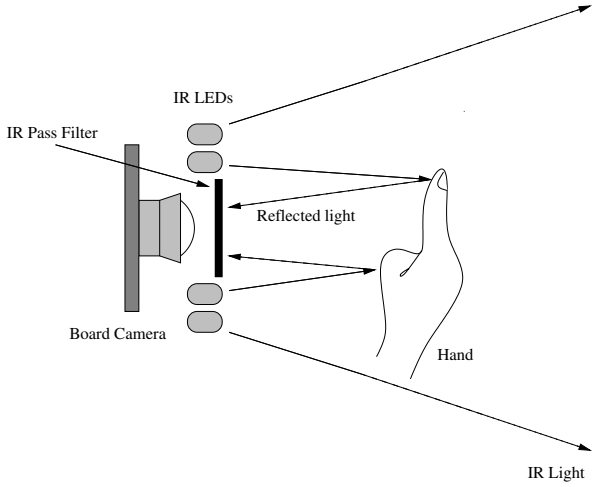

Figure 3. Sideview of pendant with IR reflecting off the user's hand

being performed. User defined gestures, recognized by hidden Markov models (HMM's) provide discrete output for the single gesture.

Data is gathered by scanning an image from the camera line by line. The algorithm used to find a blob in the image looks for a pixel with a pre-determined color. In this case, since a black and white camera is used, and the object is sufficiently illuminated, the color is a saturated white. Given an initial pixel as a seed, the algorithm grows the region by checking if any of its eight neighbors are white. This is similar to the algorithm used in Starner *et al.* [15]. If a region grows above a certain mass, it is considered a blob and certain statistics are computed for it. For this project, eight statistics were gathered from the blob: the eccentricity of the bounding ellipse, the angle between the major axis of this ellipse and horizontal, the length of the ellipse's major and minor axes, the distance between the blob's centroid and the center of the its bounding box, and the angle determined between horizontal and the line drawn between the centroid and the center of the bounding box. The last two features help determine if the fingers are extended on the hand and their rough orientation.

Variations in the skin tone of the user will not effect the recognition rate of the system. Due to our use of IR to illuminate the hand, color is not a significant factor in the detection of the hand, instead the successful segmenting of the hand is a function of the illumination of the IR on its surface.

6.1. Control Gestures

Control gestures should be simple because they need to be interactive and will be used more often. User defined gestures, on the other hand, can be more complicated and powerful since they will be used less frequently.

	vf	hf	hfh	op
vf	112	1	4	0
hf	4	125	0	0
hfh	4	15	114	0
op	0	1	0	125

Figure 4. Confusion matrix for the control gestures: hf = horizontal finger, vf = vertical finger, hfh = horizontal flat hand, op = open palm

Control gestures are those that are needed for continuous output to devices, for example, a volume control on a stereo. These are needed because gestures described by HMM's are discrete and will indicate an action, but will not let the action proceed in increments (at least in our implementation). To get a continuous control effect, the gesture would have to be done repeatedly. With a control gesture, on the other hand, the displacement of the gesture determines the magnitude of the action.

The set of features used for control gestures includes the eccentricity, major and minor axes, the distance between the blob's bounding box's centroid and the blob's centroid, and the angle of the two centroids. Eight gestures were defined for the Gesture Pendant. The gestures are determined by continual recognition of hand poses and the hand movement between frames. These hand poses consist of: "vertical pointed finger" (vf), "horizontal pointed finger" (hf), "horizontal flat hand" (hfh), and "open palm" (op). The gestures were "horizontal pointed finger up", "horizontal pointed finger down", "vertical pointed finger left", "vertical pointed finger right", "horizontal flat hand down", "horizontal flat hand up", "open palm hand up", and "open palm hand down" (Figure 5).

Assuming independence, random chance would result in an accuracy of 25%. The Nearest Neighbor algorithm was used for pattern recognition. The training and test sets were obtained in the same manner, and both sets were taken independently. The training set consisted of 1000 examples per hand pose for a total of 4000 examples. The test set consisted of 117 examples of vf, 129 examples of hf, 134 examples of hfh and 126 examples of oh. One-Nearest Neighbor on the test and training sets resulted in a 95% correct classification of the gestures. Figure 4 shows the confusion matrix of the hand poses.

6.2. User-defined Gestures

The user-defined gestures are intended to be one or two-handed discrete actions through time. Thus, a slightly dif-