Gesture Interpretation for Domestic Appliance Control

Takayuki Hattori, Yorinobu Mine, and Michael Cohen
Spatial Media Group, University of Aizu
Aizu-Wakamatsu, Fukushima-ken 965-8580; Japan
{m5111104,m5111103,mcohen}@u-aizu.ac.jp

Jun Yamadera
Eyes, Japan, Co. Ltd.; 9-15 Higashi-Sakae
Aizu-Wakamatsu, Fukushima-ken 965-0872; Japan
yamadera@nowhere.co.jp

Abstract

We present a gestural interpretation interface that optically senses hand positions and motions. These postures and gestures are recognized and processed, closing the interaction loop by controlling computer media— such as Apple TV *or Google Earth— as well as physical devices such as domestic appliances— including adjusting curtains, changing television channel and volume, turning lights and fans on and off, and piloting a cleaning robot.*
Keywords: ambient information systems, calm technology, computer vision, disappearing computer, image processing, multimodal interaction, remote controls, roomware, smart spaces and aware environments.

1. Introduction

Figure 1. Control of Robot Vacuum via Natural Gestures

Optical gesture recognition systems are not new; they have been explored for at least thirty years, since the pioneering, seminal "Videoplace" work[1] of Myron Kreuger [8] in the 80s. Kreuger coined the phrase "artificial reality" to describe such advanced interfaces that allowed users to, for instance, interpolate a high-wire between the silhouette of one's outstretched fingers, along which a shadow puppet would stride. Devices like the VPL DataGlove allowed interpretation of hand signs but were somewhat cumbersome. In the 90s, systems such as the Vivid Mandala allowed players to watch chromakeyed reflections of themselves composited into video scenes featuring position tracking for "mirror VR" systems, which musical, athletic, and fantasy games were popular in location-based entertainment installations such as science museums. This decade has seen the growing popularity of gesture recognition as a focus of research [4] [3] [10] [12].

The emergence of software like ToySight,[2] CamSpace,[3] and XTR[4] start to exploit the affordability of computer vision systems [7], the proliferation of webcams, and the incorporation of cameras into laptops and mobile phones [13].

1.1. Computer Vision "Vision"

The paradox of digital living is that user interfaces tend to become harder to use. Our group is motivated by the "calm computing" goal [15], embracing the ubicomp vision and recognizing the need to make computer interfaces accessible to ordinary users (perhaps especially "silver society" senior citizens) who aren't necessarily adept at GUI/WIMP idioms familiar to the computer literate. Natural interfaces, the "disappearing computer" [14] without mouse or keyboard [2], are more convenient and intuitive than traditional devices and enable new computing environments.

[1] www.open-video.org/details.php?videoid=8211
[2] www.freeverse.com/games/game/?id=45
[3] www.camtraxtechnologies.com
[4] www.xtr3d.com

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

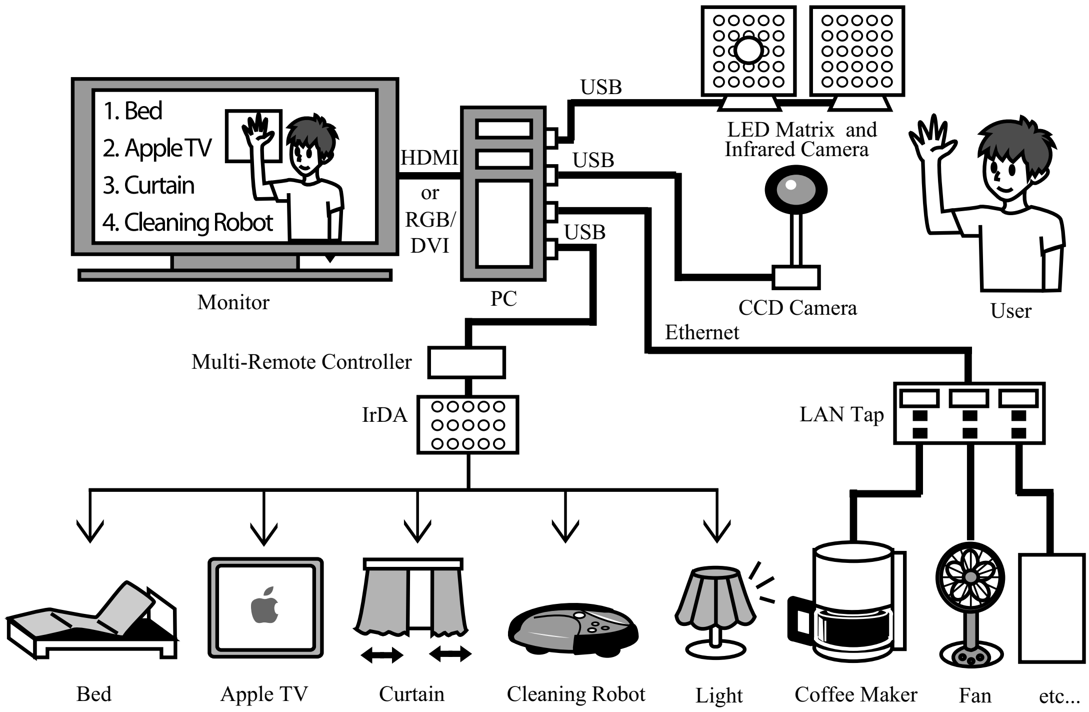

Figure 2. System Architecture

We are accustomed to controlling appliances— TV, air conditioner, lighting, etc.— by remote control. We present here a home automation [1] gestural system featuring operation by natural actions. Computer vision has some advantages compared to marker-based motion capture, including lack of requirement for reflectors or special clothing, and less restriction on location. We introduce a purely optical posture and gesture recognition system featuring a novel combination of visible and infrared cameras, allowing interpretation of gestural commands to control domestic appliances [9] [6].

For crafting an intuitive interface, we focus on gesticulation because gestures come naturally, without training, and many are international and cross-cultural. With our interface, gestures are used to control electronic media as well as physical devices, including a robot vacuum cleaner, as illustrated by Figure 1. The interface features automatic gaze-activated recognition of user gestures, hand contour and displacement, using image processing techniques for interpretation to invoke remote controls.

2. Implementation

The "roomware" digital living system is organized as shown in Figure 2. The hardware configuration for the system server is an ordinary modern PC: Windows XP Professional operating system on an Intel®Core™2 Duo CPU running at 2.10 GHz, with 3.00 GB RAM and an NVIDIA GeForce 8400M GS graphics board with 512 MB. Peripherals include a CCD camera for gesture capture, an infrared camera for eyegaze detection, and an infrared transmitter for wireless appliance control. The CCD camera is used to grab video frames; the infrared camera is used to detect eyegaze; the infrared transmitter is used to send commands to wireless appliances after interpretation of gestures to control game displays, steer the robot, etc.

Computer performance has improved to make practical image processing and computer vision interfaces which had heretofore been too heavy for realtime response. The main processes of our system are skin-color detection and shape recognition. We used OpenCV,[5] Intel's Open Source Computer Vision library of realtime image processing functions, which provide powerful methods for image processing, exposed as C++ and C functions.

2.1. Image Processing Pipeline

The image processing flow of the computer vision system is traced in Figure 3. When the program is launched, the main thread is started and the processing loop begins. The program extracts a background image for background subtraction and calibrates the user's skin-color by having the user hold his or her hand over an outline, from which the system derives a range for flesh tones. A frame grabber captures a new frame and passes it to the image processing

[5] www.intel.com/technology/computing/opencv

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

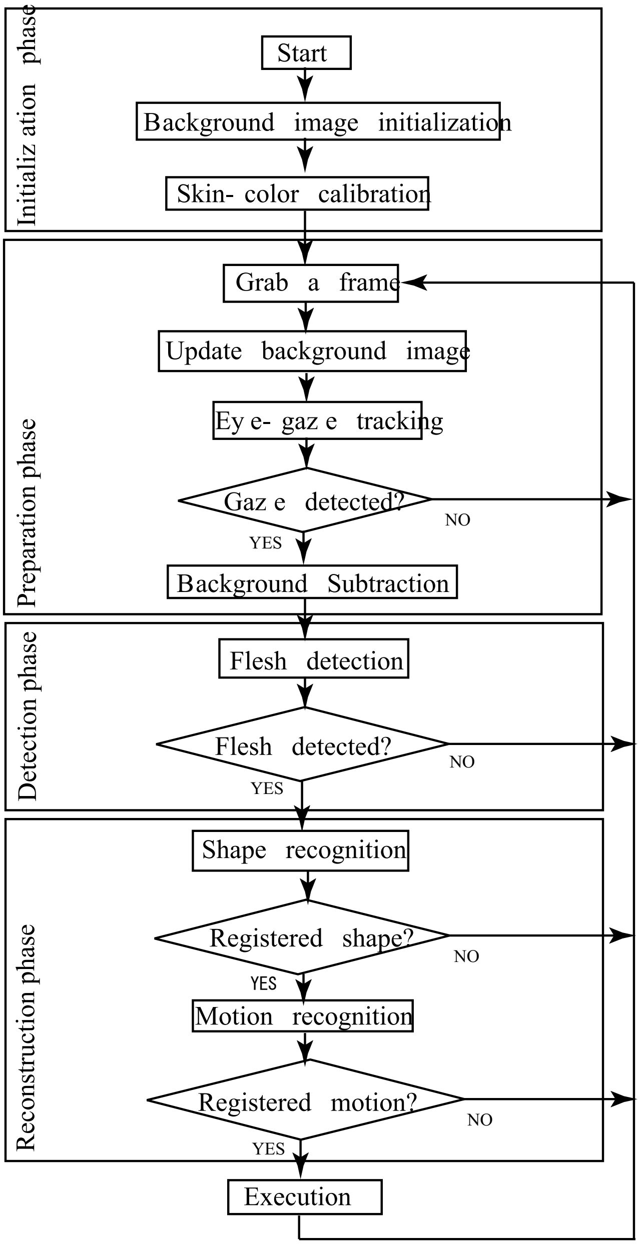

Figure 3. System Data Flow

module. In preparation phase, the system attempts to find users' eyes from each frame captured by the infrared camera. In detection phase, flesh-colored segments extracted from the moving area passed from the preparation phase are used to determine the contours of skin-color area. Finally, in recognition phase, the system judges whether or not a user's gestures represent commands. If a gesture is recognized as imperative, the corresponding command is executed. These phases are elaborated in more detail below.

2.2. Background Subtraction

In image processing, noise rejection is critical for isolating salient signals. The largest noise is background, so we engineered effective background removal. The system captures the background image and separates moving area from background by comparing each captured frame with the background. The background may alter with time as, for example, lighting conditions change, so it is updated incrementally each frame using a history-weighted moving window average: background = $(1 - \alpha) \times$ background + $\alpha \times$ frame, where α is the mixture fraction and frame is the current image. By updating the background image, the system can adapt over time and smooth over noise, like a low-pass filter. We have found empirically that a value of $\alpha = 0.005$ yields satisfactory results.

2.3. Skin-Color Detection

Skin-color detection is critical for extracting areas and contours of users' hands. We performed experiments to find an optimal representation in the system by comparison of four major color spaces: HSV, HSL, YCbCr, and RGB. These tests were performed across various conditions— altering time-of-day, lighting, and users. As a result of such experiments, we selected the YCbCr color space as the most appropriate for our application, at least for Asian and Caucasian skin tones. This is consistent with the general practice of many researchers working with skin-color detection who also use YCbCr color space because it favors the identification of flesh tone intervals.

2.4. Shape Recognition

Extracted skin-color areas are candidates for hand interpretation by the shape recognition process. First, the system extracts contours of hand candidates for determination of how many fingers the user holds up. Finger postures are reserved for top-level selections, which modes determine interpretation of hand gestures. Next, the system traces the contour on frontal (coronal) plane and determines the command vector.

2.5. Gesture Interpretation

The user simply performs natural gestures while glancing towards the monitor, besides which is the infrared camera. The system must recognize hand motion because gesture is interpreted based upon hand shape and displacement, including grasping and moving. Both hands can be detected. A menu-driven modal interface chooses the target appliance at which subsequent gestures are directed, for an effective "noun–verb" syntax [11]. Recognized gestures indicating commands include the following:

Grab Mouse click.

Grab, release, regrab Double click.

Grab, move, and open hand Drag and drop.

Move open hand Displacement to move cursor, steer robot, etc.

Our system can presently recognize only two-dimensional hand-signs because of the difficulty of recognizing three-dimensional gestures with a single camera.

2.6. Command Invocation, including IrDA and LAN-tap Transmission

The interpreted commands are sent to the respective appliance, including computer media such as Apple TV [5] and Google Earth, as well as physical devices. One can control appliances for medical applications or home care (like an assisted living care bed). The gesture recognition server can invoke recognized commands by wireless "spoofing" transmission to respective appliances. A custom-built circuit trained with the patterns of infrared appliance control emulates such protocols by a USB-connected transmitter. As industry standards move away from IrDA to WiFi and Bluetooth, this link can be easily upgraded accordingly.

Other ordinary appliances are also remotely controllable. A network "LAN-tap," an ethernet-connectable power strip, is also connected the recognition and control server via the internet, so simple electric appliances like lights and fans can be switched on and off by gesture.

3. Conclusion and Future Research

Sensor techonologies to monitor activities, biometrics, geospatial location, proximity, and contextual influences are starting to fufill the promise of the ubicomp, calm computing asymptote. We have presented the implementation of a gesture recognition system suitable for interaction with arbitrary appliances. We haven't yet performed rigorous quantitative performance evaluation or validation of the system, but informally it seems fairly robust across various conditions. Accuracy is somewhat degraded if the backlight is strong, if skin-color noise is significant, or if the background lighting or objects (such as people walking behind) changes too quickly. Variations like short-sleeved shirts, which expose arm flesh, sometimes cause tracking problems. Command interpretation is sensitive to scale and distance, since contour extraction is imprecise when the gestures are performed too far from the camera, but unaffected by rotation. We are considering extending the sensing with multiple CCD cameras, to enable to recognition of three-dimensional gestures. Our ultimate goal is to explore ways that simple, minimalist interfaces can be applied to everyday "mundane" activities, multimodal interaction that is nonintrusive and respectful, to gracefully improve social welfare and quality of life.

References

[1] R. A. Brooks. The intelligent room project. In J. P. Marsh, C. L. Nehaniv, and B. Gorayska, editors, *Proc. CT: Int. Conf. on* **C***ognitive* **T***echnology*, pages 271–278, Aizu-Wakamatsu, Japan, Aug. 1997. IEEE. ISBN 0-8186-8084-9. 2

[2] J. R. Cooperstock, S. S. Fels, W. Buxton, and K. C. Smith. Reactive environments: Throwing away your keyboard and mouse. *Communications of the* ACM, 40(9):65–73, Sept. 1997. 1

[3] L. Y. Deng, D.-L. Lee, Y.-J. Liu, and N. C. Tang. Human computer interaction based on hand gesture ontology. In *ICCOMP'07: Proc. 11th WSEAS Int. Conf. on Computers*, pages 26–31, Stevens Point, Wisconsin, USA, 2007. World Scientific and Engineering Academy and Society (WSEAS). 1

[4] J. M. S. Dias, P. Nande, N. Barata, and A. Correia. O.G.R.E.— Open Gestures Recognition Engine, a Platform for Gesture-Based Communication and Interaction. In *Gesture Workshop*, pages 129–132, 2005. 1

[5] W. T. Freeman and C. D. Weissman. Television control by hand gestures. In *ICCOMP'07: Proc. IEEE Intl. Wkshp. on Automatic Face and Gesture*, pages 179–183, June 1994. 4

[6] N. Gershenfeld, R. Krikorian, and D. Cohen. The internet of things. *Scientific American*, pages 76–81, Oct. 2004. 2

[7] B. Kisacanin, V. Pavlovic, and T. S. Huang, editors. *Real-Time Vision for Human-Computer Interaction*. Springer-Verlag, Secaucus, NJ, USA, 2005. 1

[8] M. W. Krueger. *Artificial Reality II*. Addison-Wesley, Reading, MA, 1991. ISBN 0-201-52260-8. 1

[9] T. Lewis. Information appliances: Gadget netopia. *(*IEEE*) Computer*, pages 59–68, Jan. 1998. ISSN 0018-9162. 2

[10] S. Marcel. Gestures for Multi-Modal Interfaces: A Review. Technical Report IDIAP-RR-02-34, LIDIAP: Idiap Research Institute, Ecole Polytechnique Fédérale de Lausanne, 2002. 1

[11] Z. Mo and U. Neumann. Lexical gesture interface. In *ICVS '06: Proc. Fourth IEEE Int. Conf. on Computer Vision Systems*, page 7, Washington, DC, USA, 2006. IEEE Computer Society. 3

[12] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):677–695, July 1997. 1

[13] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proc. Fourth Int. Symp. on Wearable Computers*, pages 87–94, Atlanta, 2000. IEEE Computer Society. 1

[14] N. Streitz, A. Kameas, and I. Mavrommati, editors. *The Disappearing Computer: Interaction Design, System Infrastructures and Applications for Smart Environments*. Springer, 2007. 1

[15] M. Weiser and J. S. Brown. The coming age of calm technology. http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm, 1996. 1

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278