Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 19–29
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Object interaction detection using hand posture cues in an office setting

Brandon Paulson*, Danielle Cummings, Tracy Hammond

TAMU Sketch Recognition Lab, 3112 TAMU, College Station, TX 77843, USA

Received 4 January 2010; received in revised form 21 September 2010; accepted 27 September 2010
Communicated by E. Motta
Available online 14 October 2010

Abstract

Activity recognition plays a key role in providing information for context-aware applications. When attempting to model activities, some researchers have looked towards Activity Theory, which theorizes that activities have objectives and are accomplished through interactions with tools and objects. The goal of this paper is to determine if hand posture can be used as a cue to determine the types of interactions a user has with objects in a desk/office environment. Furthermore, we wish to determine if hand posture is user-independent across all users when interacting with the same objects in a natural manner. Our experiments indicate that (a) hand posture can be used to determine object interaction, with accuracy rates around 97%, and (b) hand posture is dependent upon the individual user when users are allowed to interact with objects as they would naturally.

Keywords: Glove-based interaction; Activity recognition; Haptics; Cyberglove; Context-aware; Hand posture; Hand gesture

1. Introduction

As the future of computing heads toward ubiquity, wearable computing, coupled with context-aware applications, will become more prevalent. Context-aware applications are those which adapt one's computing environment based on "where you are, who you are with, and what resources are nearby" (Schilit et al., 1994). According to Dey and Abowd (2000), the most important types of context are location, identity, activity, and tim. Of these forms of context, activity is one of the hardest to capture and is used less frequently by many context-aware applications (Dey et al., 1999). However, we believe activity-based context can play a significant role in applications, particularly those involving wearable and pervasive computers.

As a simple example, imagine a scenario in which a user is at her place of employment and currently engaged in conversation over her office telephone. Someone else, meanwhile, attempts to call the user on her mobile phone. If the mobile phone has access to the user's activity context (i.e. the user is currently on her office telephone), then the mobile phone could respond by informing the caller that the user is currently engaged in another conversation on her office phone. Another example could be a coffee machine that can determine when to start a new brew based on the number of times a user has picked up his mug to take a drink. Other office applications could involve detecting when a user has been holding a particular position or performing a repetitive task for too long and requires a break to prevent injury.

Hand postures can also be used as metrics to evaluate the use of new products without having an observer present. In an office setting, perhaps a vendor wants to evaluate the use of a new type of phone or mouse to determine if it contributes to safety and productivity. In this scenario, hand postures can provide measurable data relating to various usage patterns. For example, if the vendor wants to determine how effective a track ball mouse is over a regular wired mouse, they could collect data showing the different types of hand postures that are used to interact with the mouse. Using hand posture data, they could possibly develop new product designs tailored to their customers' most common hand posture of choice. The vendor could

*Corresponding author. Fax: +1 979 847 8578.
E-mail addresses: bpaulson@cse.tamu.edu (B. Paulson), dcummings@cse.tamu.edu (D. Cummings), hammond@cse.tamu.edu (T. Hammond).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2010.09.003

also use hand posture metrics to determine whether or not these hand postures are ergonomically correct in order prevent strain or injury to the hands and/or wrists.

Other possible applications outside the office environment include gaming scenarios in which the activity recognition can be translated into similar actions generated by an avatar in a virtual environment. Imagine a gaming environment where certain hand postures can generate unique virtual objects with which the user could interact. This level of interaction could result in a more realistic experience for the gamer.

Activity-based context can also play a significant role in military scenarios. For example, consider the use of activity recognition to deliver team status information. A military team placed in an environment in which visibility of other team members is limited or non-existent can benefit from receiving activity data about the team. Perhaps if a team member is suddenly engaging hand postures that are defensive in nature, this information will alert his/her teammates that they should on their guard or possibly provide assistance.

Hand posture recognition can also be used as a means of communication between team members. Basic hand signals are often used in military situations where verbal communication is unsafe or impossible. Hand posture recognition would allow team members to communicate silently over long distances or in low visibility situations without the use of radios. These scenarios are just a few possibilities that show the potential benefit of knowing activity-related context.

The goal of activity recognition is to aid context-aware applications by providing information to help explain what a user is currently doing (i.e. what activity the user is engaged in). An issue activity recognition researchers face, however, is how to define what an activity is and how to determine when it is taking place. One answer may lie in *Activity Theory* (Kuutti, 1995; Nardi, 1995). According to this theory, activities have objectives and are accomplished via tools and objects. Therefore, one can assume that if we can determine the object that a user is interacting with, then we may be able to imply something regarding the activity that the user is currently engaged in. Some frameworks have been created to model activities in this manner, but were implemented in virtual environments in which interaction with objects is assumed to be given by some form of sensor values (Surie et al., 2007a, b). When applying such frameworks to a real-world domain, we still face the issue of determining when an appropriate interaction is taking place. In order to achieve full contextual-awareness, one must address the category of contextual sensing as it is the lowest, most basic part of context-aware applications (Pascoe, 1998).

The two most common approaches to tracking haptic activity and interaction in a real-world setting are vision-based through the use of cameras or sensor-based through the use of wearable sensors such as RFID tags. Vision-based techniques require cameras that are either placed within a room (Moore et al., 1999), or that are wearable (Mayol and Murray, 2005). Stationary cameras placed in a room have the advantage of being less obtrusive to the user, but makes the context-capturing system static to that one location. Wearable cameras allow for context-capturing systems to become mobile, but still have the problem which most vision-based approaches experience when dealing with the interaction of objects: occlusions (which typically occur because of the object itself).

Our challenge is to find new methods of tracking and interpreting user activity that allow for natural interaction with objects and detection tools. By working toward eliminating limitations placed on the detection environment and user activity, we hope to facilitate the creation of a ubiquitous computing experience while simultaneously delivering accurate context information about the user to nearby applications; this is needed in order to create the scenarios mentioned above.

We also want to present a new method with an activity recognition accuracy rate comparable to those resulting from the use of cameras or sensors. A new activity detection method could be used in conjunction with other commonly used approaches to strengthen accuracy and improve performance. Understanding immediate cues for object interaction and using this information to determine both the object in use and the most likely activity is the another challenge that we explore. We believe that hands are the best focus for understanding these cues.

It can be argued a person's hands are his primary means of interacting with tangible objects. They also can serve as a secondary source of communication through gesturing. Because of this, we focus our attention solely on haptic input experienced via the hands (Paulson and Hammond, 2008b). In this paper we present hand posture recognition as another means tracking haptic activity. Fig. 1 gives an idea of where hand posture may fit as a new approach to activity recognition in the overall scheme of context-aware computing. By ‘‘hand posture’’ we refer to the static positioning and orientation of the hand, which includes the fingers and palm. This differs from dynamic ‘‘hand gestures’’ which refer not only to changes in the positioning of the fingers and palm, but also changes in the orientation of the hand in 3-dimensional Euclidean space. In essence, a hand gesture is a continuous path or sequence of hand postures.

Because our interests lay more with interaction and less with writing vision-based algorithms to handle occlusions, we decided to use glove-based sensor input provided by Immersion's 22-sensor CyberGlove II (Fig. 2). The primary goal of our work to determine if hand posture can be used as a cue to help determine the objects a user interacts with, thus providing some form of activity-related context. To give some real-world practicality to our problem, we chose to perform our experiments in an office domain, a setting we believe could benefit from context-aware applications.

A secondary goal of our work is to determine the variability of hand postures between different users who