The Bar of Soap: A Grasp Recognition System Implemented in a Multi-Functional Handheld Device

Brandon T. Taylor
MIT Media Lab
20 Ames St. E15-346
Cambridge, MA 02139 USA
bttaylor@media.mit.edu

V. Michael Bove, Jr.
MIT Media Lab
20 Ames St. E15-368B
Cambridge, MA 02139 USA
vmb@media.mit.edu

CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 978-1-60558-012-8/08/04.

Abstract
We propose a vision of a grasp-based interaction system where users' intentions are inferred by the way they hold and interact with a device. In this paper we specifically discuss the Bar of Soap, a multi-function handheld prototype that uses grasp-based interactions to switch between modes. This prototype relies on the hypothesis that users share a set of stereotyped grasps associated with common multi-function modes. We show that using common machine learning techniques our device can reliably distinguish five separate classes based on the users' grasps. While this interaction is currently implemented in a multi-function handheld, we anticipate the existence of many scenarios where grasp recognition could provide a more intuitive or useful interface.

Keywords
Grasp, mode, handheld, switching

ACM Classification Keywords
H.1.2: Models and Principles: User/Machine Systems. H.5.2: Information Interfaces and Presentation: User Interfaces. K.8.0: Personal Computing: General. J.7: Computer Applications: Computers in Other Systems.

Introduction

Designing device interfaces is a difficult balancing act. Superficially simple questions such as how many buttons or where to place them can quickly become complicated when considering aspects of intuitiveness, functionality and ergonomics. Even when a device's complete functionality is known in advance, it is difficult to choose the best set of affordances. Devices like game controllers or open source handheld devices, which do not even offer this constraint, increase the problem even more. Given that electronic devices are becoming more and more ubiquitous and, at the same time, their functionalities continue to increase, designing effective interfaces will be an ongoing struggle. While there is likely no complete solution, we argue that creating devices that understand situations and adapt their interfaces accordingly will be a necessary, if not sufficient, approach.

This is not exactly a new concept in and of itself. Cameras that switch between portrait and landscape mode or cell phones that powers off their screens when held up to the users ear are both examples of devices making intelligent use of passive user interactions. We feel, however, that these types of responsive devices are not close to reaching their full potential. Thus, we have been exploring ways to use grasp recognition (a passive measurement of device orientation and user hand placement) to make devices more responsive.

In this paper, we will focus on the Bar of Soap, a prototype handheld device that uses grasp recognition as a means of mode selection, as a sample implementation of this technology. We will provide a detailed description of the Bar of Soap's hardware, briefly overview its pattern recognition techniques, and discuss a preliminary user study. In addition, we will try to relate the grasp recognition concept to areas beyond multi-functional mode switching and discuss related avenues of research.

Related Works

A parallel, independent study performed at the Samsung Advanced Institute of Technology (SAIT) provided further validation for this method of grasp recognition [2,3]. Using a different device geometry and set of device modes, they achieved similar recognition accuracies across a variety of classifiers.

Aside from this study, most efforts related to grasp recognition have come from efforts in robotics and virtual reality to develop whole-hand interfaces. While this has typically been done using glove-based systems or machine vision, a spherical device called the Tango [4] uses an array of pressure sensors to detect finger positions and squeezing. Like other whole-hand interface, though, the Tango is implemented to provide an arbitrary hand model rather than to explore how users are grasping a specific object.

Multi-Function Handhelds

As users, we face an irritating choice with our handheld devices: should we carry a small army of gadgets (a phone, an mp3 player, a camera, a PDA), each of which accomplishes its task elegantly, or should we carry one device that accomplishes each task poorly, and makes it hard to even switch the device from one mode to the other? Clearly, the ideal situation would be the emergence of a single device that can elegantly perform a variety of tasks.

figure 1. Commercially available multi-function handheld devices. The Helio Ocean (A) employs a sliding screen that reveals different button sets. The Apple Iphone (B) uses virtual buttons on a touchscreen. The Palm Treo (C) has a touch screen in addition to a variety of physical buttons

Figure 1 demonstrates the variety of approaches that have been taken with this goal in mind. Solutions range from completely relying on touchscreens and virtual buttons, to using sliding screens to reveal buttons, to loading devices with full keyboards and a touchscreen. We feel that the key to such a device will lie in coupling the adaptability of virtual buttons with the familiar methods of interaction. Toward this end, we have developed the Bar of Soap.

Bar of Soap Feasibility Study

When we first began exploring the idea of a grasp-based sensing system, much of our work focused on finding adequate sensor resolution and pattern recognition techniques. Our first user study, performed with the screen-less first version of the Bar of Soap, assured us that our sensor arrangement could adequately distinguish different grasps.

Bar of Soap V1 Prototype

The prototype, shown in figure 2, is a 4.5x3x1.3 inch rectangular box containing a 3-axis accelerometer and with 48 capacitive sensors covering 5 of the 6 faces (24 on the back, 7 on each longer side, 5 on shorter side, 0 on front). A microcontroller in the device samples these sensors approximately 3 times per second and reports the results to a PC via bluetooth. Each capacitive sensor is treated as a binary button, reporting a '1' when the user's hand covers a sensor or a '0' otherwise. The accelerometer's data is output as an 8-bit integer that corresponds to a range of +/-1.5g.

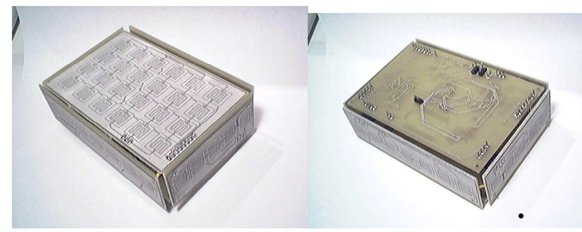

figure 2. The Bar of Soap V1 prototype with (a) back side up and (b) front, sensor-less side up.

Data Collection

Users were seated with the Bar of Soap in front of them on a table. They were told that they would be given a specific mode (e.g. "the device is a phone") and that they should then pick up the device and hold it however they saw fit until instructed to put it back down. Users were also informed that any display they would expect from the given mode would appear on the front, sensor-less face of the device. It is important to note that no suggestion of appropriate grasps was given to the user either before or during the tests.

After giving these instructions to the user, we would begin recording the data stream from the Bar of Soap and then tell the user which mode to treat the device as. Once the user had established a relatively stable pose with the device, we would label and save the data sample and have the user place the device back on the table. We would then repeat this process with each user until we had three data samples from each of the five tested modes: camera, gamepad, PDA, phone, remote control. Typical pose examples for each of the five poses are shown in figure 3.

figure 3. The Bar of Soap held in five poses: remote control, PDA, camera, game controller and phone. Note that other grasps were used by subjects in the tests.

Since in this study we were interested only in the way the user grasps the device and not in the dynamics of how they pick it up, each data sample was trimmed to a single measurement of the 48 capacitive sensors and 3 accelerometer axes. This was done by simply averaging over the final four measurements in each sample to account for any minor changes in the way the user held the device.

Using this data collection process we generated two distinct data sets, each containing 39 grasp samples for each of the five device modes. The first set was obtained from a single user to see how consistently one person held the device. For the second set, many different users held the device to see how grasps varied across a population.

Feasibility Study Results

In order to analyze the data we first had to explore methods of reducing the 51 independent features to a more manageable feature space.   We explored many different techniques including Principal Component Analysis (PCA) [3], Fisher Linear Discriminant Analysis [1], data subsets (such as only using accelerometer data), and grouping capacitive sensors. Of the methods we tried, the one that provided the highest recognition rates across multiple users was by grouping sensors on each face and accounting for symmetries.

With this method the 48 capacitive sensors formed 6 sensor groups (two 5-sensor sides, two 7-sensor sides, and two 12-sensor halves of the back face) with the value of each sensor group being a count of the active sensors in that group. Since the device should not have a preferred orientation (aside from the sensor-less face) we grouped the buttons according to accelerometer orientation rather than their physical location. Thus if the device were rotated by 180 degrees the capacitive sensors on the left side would always be treated as the left side group even though the specific sensors would move. Finally, to give equal weight to both the accelerometer and sensor groups, all the data were normalized on a scale of 0 to 1.

Next, we employed a wide range of classifiers including Templates, a Neural Net, Bayesian Classification, k-Nearest Neighbors, Parzen Windows and Generalized Linear Discriminates [3]. Each classifier was trained, cross-validated on single user data and then tested on both single and multiple-user data. Each of the classification techniques we explored performed well compared to random, with all of them achieving at least 80% accuracy for a single user and 70% for multiple users. Overall, the simple Bayesian Classification provided the best combination of accuracy (95% for single user, 79% for multiple users) and ease of implementation

Current Bar of Soap Prototype

The results of our feasibility study along with the results from the SAIT study convinced us that our sensor layout could provide an adequate grasp-recognition system. Our next step was to build a more advanced prototype that could be used to go beyond demonstrating technical feasibility and explore how coupling grasp recognition and flexible control layouts can enhance handheld interfaces.

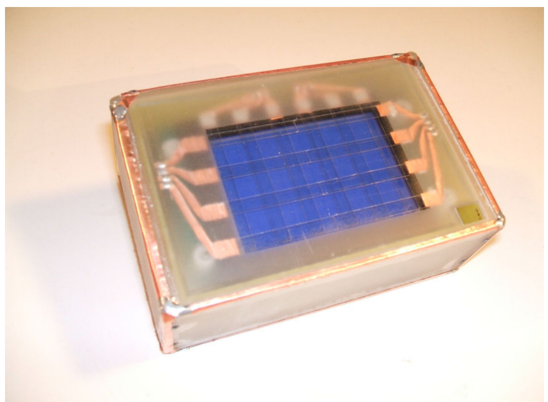

figure 3. The most current Bar of Soap prototype

The current Bar of Soap prototype, shown in figure 3, remains similar to the original in physical size and sensor density. The primary difference with the current versions is the addition of cholesteric LCD screens on both the front and back of the device. Consequently, we developed transparent capacitive sensors to overlay the screens. The end result being a total of 72 individual capacitive sensors covering all six faces of the device.

Current User Study

Our first concern was to verify that the new prototype was able to provide grasp recognition results similar to the original. We did this by essentially repeating the feasibility study and achieved similar results in a more limited sample size.

By equipping the Bar of Soap with LCDs overlaid with transparent capacitive sensors we can explore how virtual button configuration affect user interactions. While the course grid of 24 discrete sensors over the display does not provide the resolution necessary for typical touchscreen interactions, it does provide adequate grasp recognition while avoiding the multi-touch issues inherent with most touchscreens.

With our current studies we are placing a stronger focus on identifying reasons for differing interaction styles. For example, we want to know if our classifier is failing because a user holds their phone atypically or if it just because their hand is larger than average. We are also exploring how design options, such as button layouts, impact user interactions.

Conclusions and Future Work

With the results from our originally feasibility study and the preliminary results from our current study, we are confident that our grasp-recognition system can provide an effective user interface for a multi-function handheld device. By recognizing users' natural methods of interaction, the Bar of Soap provides an intuitive method of mode selection.

The method is not without its limitations of course. First, as with any system that relies on classification techniques, there will undoubtedly be errors. While our research indicates that these can be greatly reduced by training the system to the individual, we have yet to explore the cost of such errors in terms of user experience. Another limitation is the fact that we have used a rather limited set of modes in our study. While we are confident that we could extend the number of modes without significantly increasing classification errors, there is a limit to how many distinct ways a user can hold the Bar of Soap. A better understanding of approximately how many natural grasps can be recognized would be useful.

Despite these drawbacks, we feel that the Bar of Soap is a promising demonstration of the possibilities for grasp-recognition. Given the minimal cost of the sensors and negligible computing power required for the classification, the benefits are significant. Instead of forcing users to dig through additional menus or deal with additional buttons, designers can customize interfaces based on users' natural interactions. For users with limited dexterity or vision, such an interface may provide a way to use an otherwise inaccessible device.

Additional Grasp Recognition Implementations
Beyond the work we have done with the Bar of Soap, we are planning to explore how grasp-recognition can be used as a general interface method. We are currently designing devices of other geometries, which we will equip with similar sensors arrays. For example, we envision a baseball video game where the pitch type is selected not by pushing a button, but by appropriately gripping a sensor equipped ball. Similarly, it is not hard to imagine equipping a stylus with grasp recognition so that switching between pen and paint brush modes in an art program can be done by merely holding the stylus as one or the other. We look forward to developing such systems and exploring their possibilities.

Acknowledgements
We thank Jeevan Kalanthi, Daniel Smalley and Matt Adcock for their assistance with the analysis of classification techniques. Also, thanks to Quinn Smithwick, James Barabas, Ana Santos and all the other user study participants.

References
[1] Duda, R., Hart, P., Stork, D. *Pattern Classification*, John Wiley & Sons, Inc., 2001.

[2] Chang, W., et al., Recognition of Grip-Patterns by Using Capacitive Touch Sensors. *Proc. ISIE 2006*, IEEE Press (2006), 2936-2941.

[3] Kim, K., et al., Hand Grip Pattern Recognition for Mobile User Interfaces. *Proc. 21st AAAI Conf. Artificial Intelligence*, AAAI Press (2006), 1789-1794.

[4] Pai, D., et al., The Tango: A Tangible Tangoreceptive Whole-Hand Human Interface. *Proc. of World Haptics 2005*, IEEE Press (2005), 141-147.