Mobile Interaction Based on Human Gesture Analysis

Gamboa Ricardo[1], Guerreiro Tiago[1], Gamboa Hugo[2], Jorge Joaquim[1]

[1] Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal
[2] Plux, Engenharia de Biosensores

Abstract
Mobile devices' user interfaces have some similarities with the traditional interfaces offered by desktop computers, but those can be highly problematic when used in mobile contexts. Human gesture recognition in mobile interaction appears as an important area to provide suitable on-the-move usability. In our work, we use inertial data to detect both contextual information and explicit gestures within the user's body frame. Those gestures allow the user to naturally interact with mobile devices with no movement limitations.

Introduction

The actual interaction with mobile devices is not easy. Their interfaces continue to be produced mainly based on graphical components, similar to desktop computers. Those graphical interfaces, already limited by the small screen area, are also extremely hard to control when a user needs to do it while moving. As result, user interfaces become cumbersome and slow to use, demanding high levels of users workload. Besides, the number of physical buttons is decreasing in each mobile device, in order to make them smaller and with larger screens. All these emerging problems have been addressed with some solutions that popped in the market, such as voice and key shortcuts. However, those solutions are limited and generally abandoned: voice shortcuts have low recognition rates and low acceptance when used in public, while key shortcuts fail on long-term usage due to low correspondence between actions and buttons [8?].

Gestural interaction is an area of interaction that is still not adapted in commercial mobile devices. Gestures are one of the most important channels of communication between humans, and gestural interaction is already explored in diverse interfaces and applications. Our work is also based on the idea that gestures combined with body parts give even more emphasis and meaning to gestures (i.e., apologising with a hand over the heart or asking for the time with a touch on the wrist are common cultural gestures). Both gestures and body parts can be combined, generate mnemonical shortcuts and, when recognized on mobile devices, revolutionize the way we interact, approaching this communication with the one between humans.

Gestural Mobile Interfaces

Gesture recognition with mobile devices is being studied with the support of a wide variety of sensors. Touch Screens, Radio Frequency Identification (RFID), Cameras, Electromyography (EMG) and Accelerometers were found to be the most used auxiliary sensors. Touch screens are able to recognize 2D finger movements [2]; RFID provide point recognition when used on clothes or personal object, as suggested and prototyped by Headon and Coulouris [1]; Cameras can provide gestural recognition through visual tags but especially using the optical flow to detect movement, rotation and tilting of the phone [3]; EMG is the best choice for a subtle interaction based on muscle contractions (Contanza et al[4]). Finally, accelerometers are distinguished from the others because they are clearly the most used sensor, based on its small size, ability to capture acceleration on multiple axes and deliver it with great accuracy and low power consumption for mobile and small addon devices. Accelerometers are used to detect vibrational [5] and tilt [6] input but also other complex movements such as numbers drawn in the air [7]. While EMG and RFID approaches are dependent on the use of extra hardware on the body, inertial sensing with accelerometers is capable to achieve the same results without the referred limitation.

Mnemonical Shortcuts

In order to study the actual state of interaction with mobile devices and shortcut usage, we advanced with a task analysis. In parallel, we also focused on the validation of the concept of combining both gestures and body for mobile interaction [8]. In the first stage, twenty users were questioned and tested about their habits when using mobile devices. The majority (75%) uses key shortcuts on a daily basis, with an average of 5 programmed shortcuts, while voice shortcuts are not used. However, difficulties on memorizing the shortcuts were reported. When trying to reach the three most important applications, users spent an average of 4 keystrokes, and 5 keystrokes to the three most important contacts. Observation demonstrated that people still rely on menu selection to reach those applications and contacts, even when a key shortcut is available. To test the concept of body-based mnemonical shortcuts, a RFID prototype was developed and tested. This prototype was the combination of RFID to be sticked on clothes and a reader on the mobile device. The evaluation consisted on twenty users asked to chose the 5 most important tasks and associate them with body parts and key shortcuts. In a controlled environment, they separately performed 20 shortcuts with gestures with the RFID prototype and keys, repeating the test one week later. Results demonstrate that are some clusters of meaningful relations (contacts to chest, message with hand, call with mouth, etc.) and the RFID prototype was characterized with a low error rate (0.8%) when compared with key shortcuts (9%). The relation was maintained one week later, with 6% error of mnemonical shortcuts and 22% of regular key shortcuts.

Accelerometer-based Interaction

Task analysis suggests that a new interaction paradigm is important to increase mobile devices' usability and evaluation on the RFID prototype demonstrated that mnemonical gestures are a good candidate solution. However, a RFID-based system has some inconvenient regarding the need of using RFID tags on clothes or personal objects to allow the interaction. Following the line of the major part of the related work on this area, we decided to use accelerometers for a new prototype, mainly because of its precise measure of acceleration and easy integration on mobile devices.

For development, we used a Bioplux4 wireless system and an ADXL330 MEMS tri-axial accelerometer. The three channels of the accelerometer were connected to three of the analog channels of the device that delivers the RAW data of the accelerometer through Bluetooth connection, with a sample rate of 1024 samples per second. The data was captured and processed on a Pocket LOOX 720 using .Net programming (C#). Our efforts on the usage of the accelerometer for recognizing the human interaction went beyond the concept of mnemonical gestures, and were divided in two main areas – Contextual and Explicit interaction. The fusion of these different types of interaction is intended to create a total gesture-based interface. This interface, when accompanied with appropriate audio feedback, may be the beginning of the creation of small devices without visual interface and keypads.

Contextual Interaction

A mobile device with motion sensing capabilities has the ability to capture implicit data about the user movement, providing different functionalities in each situation. Thresholds were used to detect the presence of movement (when the user is walking or running) and a lower threshold to detect if the user is holding the device. When both thresholds are not passed, it indicates that the device is stable. Those characteristics can be useful to decrease energy consumption of the device or chose different user profilers for different motion characteristics. However, the most interesting investigation we made on this area was focused on the fall detection. This kind of contextual data is given much importance because a good recognition rate can provide faster medical care to the elder. Our algorithm is based on two main characteristics when people fall: an unusually large acceleration magnitude and an unexpected final angle of the device. When both characteristics are combined, it is possible to recognize falls and have a low false-positive rate.

Explicit Interaction

Considering explicit interaction, we define two main areas to be studied. In one hand, we find important to use the accelerometer to provide users a different model of navigation on mobile devices. On the other hand, the main objective is still the construction of an algorithm to recognize mnemonical gestures related to body parts. We find these two functionalities as the most important when defining a complete interface for mobile devices.

The navigation on the device software was achieved through the recognition of tilt movements along two axes, considering that there is no dynamic motion on the device and all the alterations are due to the rotation of the device, passing the gravity acceleration between axes. Using this assumption, it is possible to calculate the orientation of the device in the three axis in each instant. The tilt movement is recognizable because it has a specific variation on the angle of the signal, based on the initial position. As it’s illustrated on the graphic x, a threshold of 30 degrees is enough to clear the noise and detect the tilt in both directions.

The main goal, the recognition of mnemonical gestures, was achieved with a totally different approach. The two main characteristics to take in account in the construction of such algorithm were its high recognition rate and also the importance of being lightweight to be executed on mobile devices with low processing capabilities. Instead of using feature-based algorithms, we decided to map the dislocation of the mobile device on a 2D plan, calculating the distance between an initial and fixed point (the chest) and a final point. The distance calculation was based on a double integration of the signal. However, since this integration delivers some error and the mobile device may suffer some unexpected rotation, we also applied a moving average filter and a threshold to isolate de part of the signal where the real movement was present. With this approach, it was possible to rapidly detect the movement on both x and y axis. Mnemonical gestures are recognized in two different ways: one option is to train the system and relate the given value with body parts; the other option is to pre-process data, which permits some default gestures based on the height of the person, taking off the necessity of further training.

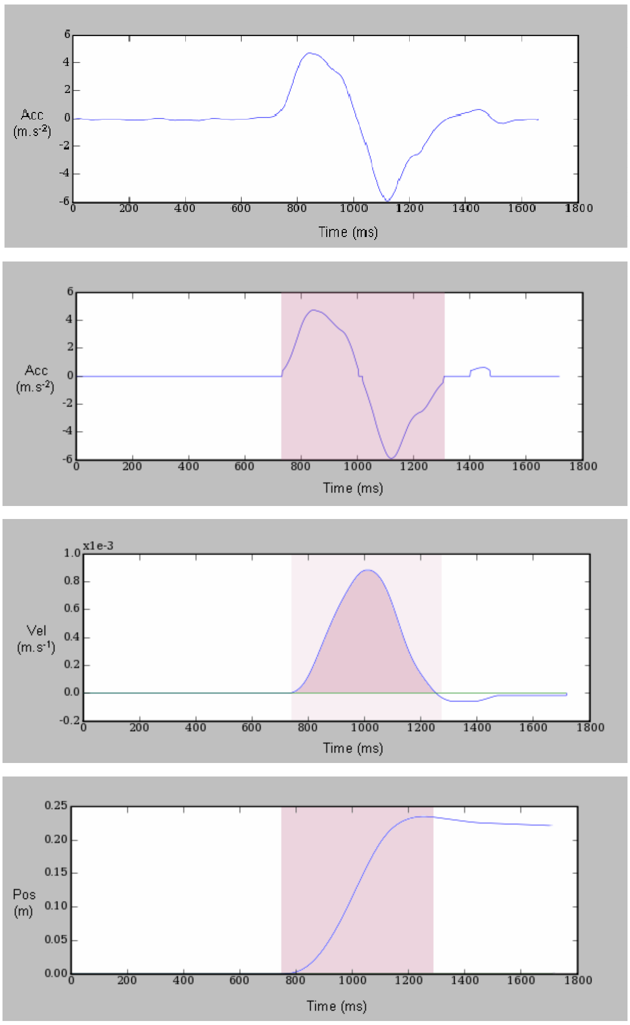

Figure 1. Signal Processing Evolution

a) Raw Signal b) Filtered c) Velocity d) Position

Discussion

The prototypes in both contextual and explicit interaction are in final adjustments, but user testing is needed to find out which are the weaker points on the interface. However, we are aware of some limitations. The fall detection is effective for regular falls but complex falls, when the final rotation of the device is almost vertical, will not be detected. In tilt detection, the major limitation is the inexistence of an automatic adaptation to the rotation of the device because the user has to re-start the movement acquiring when he changes the device initial orientation. The position detection has one major limitation, regarding the impossibility to rotate the device during the mnemonical gesture. The solutions to these problems are being studied and will be discussed evaluated with real users in within real daily life scenarios.

Conclusions and Future Work

We proposed the usage of inertial sensors to capture the human motion and build an interface based on gesture recognition, in an attempt to increase the usability of the actual mobile devices when users are on-the-move. Prototypes were successfully developed in both contextual interaction analysis (motion characteristics and fall detection) and implicit interaction analysis (Tilt-based navigation and body-frame mnemonical gestures). User tests are needed to evaluate the interface and clarify the potentiality of an all-gestural interface for mobile devices.

References

[1] Headon, R., Coulouris, G. Supporting Gestural Input for Users on the Move. Proc IEE Eurowearable '03, pp 107—112 2003.

[2] P. Pirhonen, S. A. Brewster, and C. Holguin, "Gestural and Audio Metaphors as a Means of Control in Mobile Devices," ACM-CHI'2002, pp. 291 - 298, 2002.

[3] M. Rohs, “Real-world interaction with camera-phones”, In: International Symposium on Ubiquitous Computing Systems, 2004.

[4] E. Constanza, S. Inverso, R. Allen. “EMG as a Subtle Input Interface for Mobile Computing”, MobileHCI, pp.426-430,2004

[5] S. Strachan, R. Murray-Smith. *Muscle Tremor as an Input Mechanism*. In Annual ACM Symposium on User Interface Software and Technology 2004.

[6] J. Rekimoto, Tilting operations for small screen interfaces, Proceedings of the 9th annual ACM symposium on User interface software and technology, pp.167-168, 1996.

[7] E. Choi, W. Bang , S. Cho, J. Yang, D. Kim, S. Kim. Beatbox music phone: gesture-based interactive mobile phone using a tri-axis accelerometer, ICIT 2005**.**

[8] R. Gamboa, T. Guerreiro, J. Jorge, *Mnemonical Body Shortcut,* Gesture Workshop 2007.