Binding an Handheld Device with its Owner

Maurício Sousa and Joaquim Jorge

INESC-ID / IST-UTL
antonio.sousa@ist.utl.pt,jaj@inesc-id.pt

Abstract. We present a gesture based method to associate personal handheld devices to their users in a three-dimensional interaction space to support cross-device interactions and collaboration. This work aims to provide a base model to support future collaborative environments with multiple users and a mix of private and public interactive surfaces, where mobile devices play a powerful role in data sharing within the ecosystem. The proposed method uses the correlation between hand acceleration and the handheld device's accelerometer data to find a matching pair.

Keywords: Cross-device Interactions, Handheld Devices, Accelerometer, Pattern Matching.

1 Introduction

Mobile devices have become ubiquitous, are equipped with several sensors to better understand the world surrounding them, and they hold storage capacity for large amounts of data. Also, mobile devices are, almost exclusively, personal and linked to one user. Therefore, this type of handheld devices are proxies of their users and can be used as a personal identifier.

Although these devices were designed for communication between geographically separated people, their attributes can be used by collocated groups of people to meet and share information and improve collaboration in business meetings or in informal gatherings. Cross-device interactions can range from transferring files triggered by proximity to joint creation of documents.

Short-range communication technologies, combined with sensor data, enables interactions with other interactive surfaces such as large scale wall mounted displays, public displays and multitouch tabletops.

In this paper, we propose a method for a computer vision user tracking based interactive environments to combine body movements and handheld device's accelerometer data in order to determine which user and which hand holds the device. Enabling usage of *proxemics* interactions [6], personal identity and content access, to aid side by side cross-device interactions. Our approach takes advantage of built-in motion sensors in mobile devices to calculate a positive match with the user's body motion, easily obtained from depth cameras.

2 Related Work

There is considerable research on collocated handheld device interactions in intelligent environments. For an ecosystem with various devices, whether mobile or stationary [4], the first step is the creation of a virtual link between them. Several binding techniques proposals require user intervention, by simultaneously perform the same action on multiple devices [13], doing a shaking gesture [5], bring devices together to close proximity [7] or pushing a button.

Chong and Gellersen [3] devised a study eliciting spontaneous input actions performed by participants to gather association techniques for binding devices together. The study identified *proximity* and *search* & *select* as two of the top five preferred techniques.

Close proximity can be used as an indicative of a non verbal interaction. Vogel and Balakrishnan [19], developed the concept of *proxemics interaction* by applying it to wall mounted displays where user's information displayed ranges from personal to public according to proximity. Marquardt [1][6] proposes the use of *proxemics* to mediate people, devices and non-digital objects, analysing proximity and orientation to allow applications to implicitly change information displayed on a screen or to implicitly trigger events.

More recently, Marquard et al. [7] discussed the **GroupTogether** system that effectively applies the concept of *F-Formations* to define distinct relationships between people engaged in conversation and between their handheld devices. In **GroupTogether**, *F-Formations* are used as a base sociological model to enable exchange of information between personal devices.

Much research has been done in order to obtain users personal identity. Some works make use of intrusive devices to obtain the identity information [8] [10] [15]. However, various techniques are non-intrusive. Some use biometrics [16], some explore what users wear [11], and others rely on computer vision [12].

Within the work related to computer vision, there is a consensus on the usage of data fusion between body movement signals captured by cameras and data from inertial sensors [9] [2][17]. Teixeira et al. [18] proposes the use of inertial sensors that are within mobile devices in combination with a network of surveillance CCTV cameras to locate personnel, track people inside buildings and locate intruders.

On a similar note, Rafouei et al. exposes a similar problem regarding identity in user interactions on touch-enabled surfaces. In his work, a solution is proposed where the user holds a mobile device in one hand and uses the other to touch the interactive surface. The identity is obtained using the similarity between accelerometer sensor data with the accelerations of all hands tracked by a Kinect device [14].

3 Our Approach

Our approach consists in computing the resulting accelerations from hand movements of all tracked users in order to find the relationship of resemblance with accelerometer data obtained from their mobile devices. Pairing an observed user with his personal device, know his identity and open a channel of communication between other handheld devices and interactive surfaces.

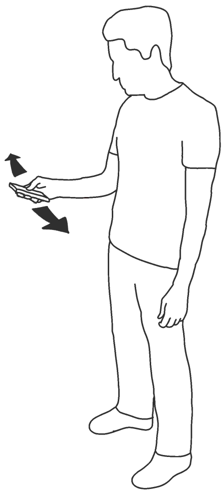

Fig. 1: User, holding an handheld device, performing the horizontal gesture to initiate a session.

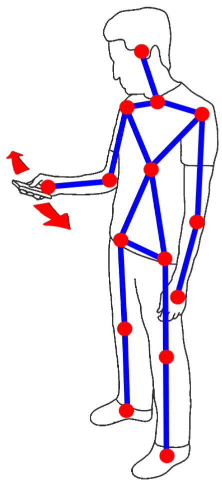

Fig. 2: User Skeleton Model obtained with Microsoft Kinect depth cameras

In order to ensure unambiguous data, we devised a gesture to indicate the intention to bind a user with his phone and to add a that new device to the ecosystem of the interactive environment. The gesture corresponds to an horizontal movement of the hand holding the mobile phone, as depicted in Figure 1, which creates a signal of the accelerometer data within a set time interval.

Simultaneously, we use all users' joints information, from a *Skeleton Model* supplied by one Microsoft Kinect depth camera, to calculate similar acceleration signals from the instantaneous speed given by a change in hands position. Figure 2 shows a representation of the three-dimensional data of the obtained *Skeleton Model*.

Since the values derived from the handheld device accelerometer sensor are calculated using the Earth's gravity, hand tracked values are divided by $1g$ ($g = 9,8m/s^2$) so they can both be at the same scope.

The Figure 3 shows two acceleration signals plotted as waveforms. Where the blue signal corresponds to data collected from the accelerometer sensors, and the

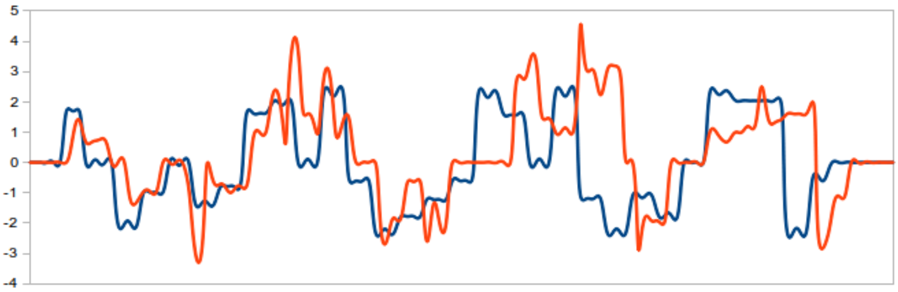

Fig. 3: Plot representing the data set for the computer vision tracked accelerations(in orange) and handheld device accelerometer(in blue) from the same hand, during the horizontal pairing gesture.

color orange corresponds to the calculated accelerations of the users' hand. It is thus apparent that although the signals are similar, there are substantial differences in the amplitude and there is an offset on the time axis. Moreover, the second signal contains some noise due the nature of the hardware chosen to track the users' hands. Every attempt to initiate a binding actions, a list of signals, for each hand of each person observed by the depth camera, is created. Where only one signal will match the data from the mobile device. Thus, to find the corresponding pair we use the *Cross-correlation* algorithm that, given two waveforms, returns a degree of similarity between them, the *correlation coefficient*. In pattern matching, *Cross-correlation* is a standard method of finding the measure to which two signals are correlated. In our case, we use the definition for discrete functions, Equation 1, where f^* is the complex conjugate of f.

$$(f \star g)[n] = \sum_{t_0}^{t_n} f^*[m]g[n+m] \tag{1}$$

For every users' hand signal h, with N being the number of observed hands, and $a_{handheld}$ the signal with accelerometer data, we apply the matching algorithm $(h_N \star a_{handheld})$. The pair that returns the highest *correlation coefficient* is chosen.

Ultimately, the proposed algorithm can easily find the corresponding pair, as the calculated *correlation coefficient* stands out with much higher numerical value than the coefficients for incorrect signals. The *Cross-correlation* method allows the signals to have shifted values in time, different amplitudes, as well as producing positive results with noisy signals.

4 Current Prototype

To create the base model for this experiment, we began with a prototype that follows a Client-Server distributed system, connecting several clients to a *Central*

Server. We developed two different types of clients, each one responsible with the gathering of different sensor data. A Microsoft Kinect depth camera based *User Tracker*, enabling the system to follow users' position and body movements. And a *Mobile Client* to serve as the personal interactive surface and the source of accelerometer data. We conceived this topology to ensure that the system is scalable to the extent that the number of users increases in the same proportions as new instances of the *User Tracker* client are added.

The *Central Server* receives data from multiple clients and implements the domain model to manage the position and identity of users in order to be able to determine the intent of interaction between actors, whether they are people or devices.

Communication between the various components was established through a protocol built over TCP/IP in a wireless local area network.

5 Conclusions and Future Work

This paper explores a gesture based method to match handheld device accelerometer data with the body motion of its user as a way to add new devices to a multiple interactive surface ecosystems. In addition, we proposed the usage of *Cross-correlation* as an signal processing algorithm to find the match between accelerations of the user's hand and the device.

Our current and future work focuses on improving the user experience and, ultimately, eliminate the horizontal gesture to allow users to log their devices into the system, even with the device still in his pocket. Also, further experimental testing, regarding security issues, will be performed to ensure that our method does not return false positives which could eventually lead to a breach of privacy.

Acknowledgements

The work described in this paper was partially supported by the Portuguese Foundation for Science and Technology (FCT) through projects CEDAR PTDC/EIA-EIA/116070/2009 and PEst-OE/EEI/LA0021/2013.

References

1. Applying proxemics to mediate people's interaction with devices in ubiquitous computing ecologies. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 1–1, New York, NY, USA, 2010. ACM.
2. G. Bahle, P. Lukowicz, K. Kunze, and K. Kise. I see you: How to improve wearable activity recognition by leveraging information from environmental cameras. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pages 409–412, 2013.
3. M. K. Chong and H. Gellersen. How users associate wireless devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1909–1918, New York, NY, USA, 2011. ACM.

4. J. Cortez, D. A. Shamma, and L. Cai. Device communication: a multi-modal communication platform for internet connected televisions. In *Proceedings of the 10th European conference on Interactive tv and video*, EuroiTV '12, pages 19–26, New York, NY, USA, 2012. ACM.
5. K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: pen gestures that span multiple displays. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '04, pages 23–31, New York, NY, USA, 2004. ACM.
6. N. Marquardt. Proxemic interactions in ubiquitous computing ecologies. In *PART 1 ———- Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 1033–1036, New York, NY, USA, 2011. ACM.
7. N. Marquardt, K. Hinckley, and S. Greenberg. Cross-device interaction via micro-mobility and f-formations. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 13–22, New York, NY, USA, 2012. ACM.
8. N. Marquardt, J. Kiemer, D. Ledo, S. Boring, and S. Greenberg. Designing user-, hand-, and handpart-aware tabletop interactions with the touchid toolkit. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 21–30, New York, NY, USA, 2011. ACM.
9. W. W. Mayol, A. J. Davison, B. J. Tordoff, and D. W. Murray. Applying active vision and slam to wearables. In *Robotics Research*, pages 325–334. Springer, 2005.
10. T. Meyer and D. Schmidt. Idwristbands: Ir-based user identification on multi-touch surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 277–278, New York, NY, USA, 2010. ACM.
11. A. M. Piper, E. O'Brien, M. R. Morris, and T. Winograd. Sides: a cooperative tabletop computer game for social skills development. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, CSCW '06, pages 1–10, New York, NY, USA, 2006. ACM.
12. R. Ramakers, D. Vanacken, K. Luyten, K. Coninx, and J. Schöning. Carpus: a non-intrusive user identification technique for interactive surfaces. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 35–44, New York, NY, USA, 2012. ACM.
13. J. Rekimoto. Synctap: synchronous user operation for spontaneous network connection. *Personal Ubiquitous Comput.*, 8(2):126–134, May 2004.
14. M. Rofouei, A. Wilson, A. Brush, and S. Tansley. Your phone or mine?: fusing body, touch and device sensing for multi-user device-display interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1915–1918, New York, NY, USA, 2012. ACM.
15. V. Roth, P. Schmidt, and B. Güldenring. The ir ring: authenticating users' touches on a multi-touch display. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 259–262, New York, NY, USA, 2010. ACM.
16. D. Schmidt, M. K. Chong, and H. Gellersen. Idlenses: dynamic personal areas on shared surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 131–134, New York, NY, USA, 2010. ACM.
17. Y. Tao, H. Hu, and H. Zhou. Integration of vision and inertial sensors for 3d arm motion tracking in home-based rehabilitation. *Int. J. Rob. Res.*, 26(6):607–624, June 2007.

18. T. Teixeira, D. Jung, and A. Savvides. Tasking networked cctv cameras and mobile phones to identify and localize multiple people. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 213–222, New York, NY, USA, 2010. ACM.
19. D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM.