

Gestural Interfaces for Mobile and Ubiquitous Applications

Ionut-Alexandru Zaiti and Stefan-Gheorghe Pentiu

Abstract. In the last years there has been a remarkable evolution in acquiring data on human gestures. Increasing efforts are focused on hand gestures as they provide an efficient interaction model for any application. This is more obvious in applications for desktops or gaming systems to be used indoors or relatively stand still activities in which the user does not leave the confines of a room or virtually defined space. However, mobile applications have not benefited the results of recent advances to the same extent. While most of the interaction we have with mobile devices is done using our hands, the model has remained unchanged. We suggest the use of natural hand gestures in both desktop and mobile applications and provide a starting point for achieving an engaging and realistic interaction while still retaining a connection with the surrounding physical space. We go over the work we have done towards this goal and provide some guidelines for designing applications based on hand gestures.

1 Introduction

In the age of ubiquitous computing [24] we are constantly surrounded by multiple computers and smart devices all presenting us with various choices of entertainment or other activities through different ways of integrated interaction. Laptops and desktop computers mainly use the touch pad, mouse and keyboard to allow us to provide input and commands, smartphones and tablets come with touchscreens. To these several other input methods are added as accessories or peripheral devices.

There are joysticks and other game consoles which are specialized to serve a particular purpose, such as those ever popular applications, video games [2].

Ionut-Alexandru Zaiti · Stefan-Gheorghe Pentiu
University Stefan cel Mare of Suceava, 13, Universitatii, Suceava, 720229, Romania
e-mail: ionutzaiti@yahoo.com, pentiu@eed.usv.ro

While these devices serve different purposes and seem to present various and very different ways of interaction they are all used with our hands. Most of the interaction we have with smart devices is done through our hands due to the impressive range of actions that can be performed from the simple moving of the mouse to the calibrated and synchronized text writing through a keyboard.

The newest entries to input devices, for gaming applications or otherwise, broke away from the relatively static model of previous ones. The Nintendo Wii² allowed the user to move around freely and interact with the application using motion gestures to simulate a real life action, such as in the game of tennis where the Wii sensor could be used as a tennis racket. This change contributed towards a more realistic and physically involving experience. The users were no longer bound to their seat and could explore a larger area while holding the game controller. Wireless communication allows the user to move around with a mouse as well but the interaction does not change based on the location of the user. In most cases a larger space is needed to perform various motions required to accomplish a goal.

Similarly, in the case of Microsoft Kinect³, a vision based technique is used to acquire data on the whole body movement. What this means is that the user is presented with the opportunity to execute a wider range of gestures, with an increased accuracy. Both Wii and Kinect create a virtual game space in which the user can move relatively free.

We believe that given the interaction we have with our environment is mostly done through our hands, capable of a varied range of functions with different levels of precision [10], any gestural interface should be at least partially based on hand gestures. Furthermore, hands should not only be used as an actuator but as a sensor as well. We considered that a good opportunity to provide a complete experience in interacting with the virtual space is to use the objects around us as a method of input by reversing the process of hand pose and gesture recognition. In the case of mobile applications, even though most of the interaction is done through our hands, the gestures we are required to use are limited in range. We implemented several applications for smart mobile devices using data not only on the position of the finger on the screen but also on the flexion of the used fingers. Based on our work we provide a discussion and guidelines on using hand gestures in the context of both standstill and mobile applications.

2 Object Based Interaction

We argue that hand postures can inform on the object the user is manipulating. The result is transforming everyday objects into physical interfaces instead of using specialized equipment. While there are multiple choices for acquiring data on hand gestures [3, 7, 8, 18] we used a data glove which allowed a high degree of precision while the user could still execute gestures in any position he saw fit (which is not possible in vision based techniques, depending on the user's position

² Nintendo Wii (<http://www.wii.com/>)

³ Microsoft Kinect (<http://www.xbox.com/en-GB/kinect>)

relative to the camera). We performed a study using the 5DT Data Glove⁴ which includes 14 sensors: two for each finger (the knuckle and first joint) and four sensors that measure proximity between each pair of successive fingers. The data glove captures data at a frequency of 60 Hz. Users had to execute tasks of manipulation for 18 different objects, 6 basic shapes, each in 3 sizes, small, medium and large.

In this context, a gesture can be classified as either simple or complex. For simple gestures or postures the hand and fingers do not move and stand relatively still (such as in simply holding an object). Complex gestures require movement of the hand and fingers to any degree of complexity, such as in closing the fist (Figure 1).



Fig. 1 The gesture of closing the fist represented in four steps from left to right

Hand postures are represented as 14-dimensional vectors and they are classified using the Euclidean distance:

$$p = \{p_1, p_2, \dots, p_{14}\} \in [0, 1]^{14} \tag{1}$$

$$\|p - q\| = \left(\sum_{i=1}^{14} (p_i - q_i)^2 \right)^{\frac{1}{2}} \tag{2}$$

For the first task (object translation) the participants stood in front of a table and they were asked to pick up objects from their right side and move them to the left side. The order of the objects was randomly generated through a software application which instructed participants on their task. Hand posture data was captured during the action of picking up and moving the objects. For the second task (object exploration) each object had a series of digits from 1 to 6 inscribed on several locations of its surface. The participants were required to perform an exploration in order to identify a randomly generated sequence of the digits. We obtained recognition rates over 95% when discriminating between the given objects with data acquired from 13 participants [22]. Our results are in agreement with those reported by existing research [16].

To illustrate our results we applied them for video games, in part due to their popularity [2] and recent adoption of gesture input which allows us to reach a

⁴ <http://www.5dt.com/products/pdataglove14.html>



Fig. 2 Using real objects to interact in a first person shooter: (a) a toy as a gun and (b) a cup as a grenade for a first person shooter [25]

large number of potential users. We also chose games because of their evolution, being one the most dynamic and adaptive type of application. As an example we give Counter Strike, a popular first person shooter, in which we used regular house hold objects to interact with the virtual world [25]. The user can pick up various objects from his surrounding and can use them in the game, thus creating a custom physical interface (Figure 2). We mention previous implementations of object based interfaces [9, 12, 20] where objects were specifically designed for a predetermined action. An advantage to this technique is the increased rate of recognition, each object having its own identity. However, as Sluis et al. [20] reported the users found the objects similar to a TV remote and separate from the environment. The objects had to be redesigned to appear more as decorative and create the impression of belonging to the environment.

2.1 The Holding Posture

To complement the analysis of our previous experiments we provide the results for a third task, the object holding posture, in which the participants were asked to identify the most comfortable holding posture for each object. The experimenter was present during the process and recorded the gesture data when the participant was ready. The task took approximately 5 minutes to perform and an average of 172 postures were gathered for each object, in order to make available the small variations which can appear during holding an object.

2.2 Recognition and Analysis

Following our analysis on the translation and object manipulation tasks we limited our use of classifiers to the nearest neighbour classifier, the k-nearest neighbour classifier and a multilayer perceptron (MLP) in order to recognize object size and

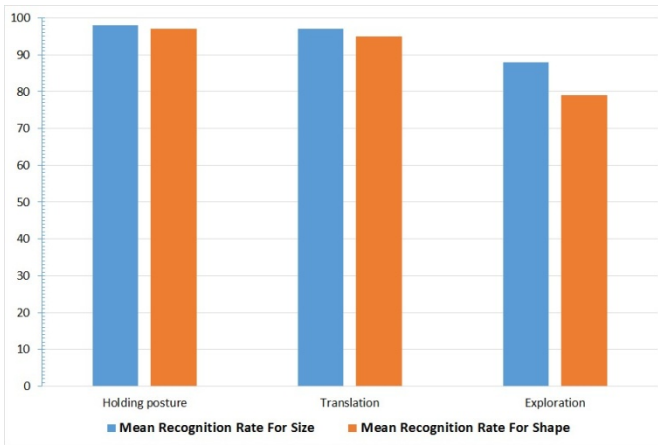


Fig. 3 Mean recognition rates for object shape and size using a nearest neighbour classifier applied on the raw data sets of all three object handling tasks

shape. Both nearest neighbour classifiers used a separating threshold of 0.1 which we determined to provide an acceptable compromise between data precision and recognition accuracy. The technique of calculating recognition rates for the holding posture was similar to the one for the previous tasks. For each object data set we randomly chose a fixed window of w postures which was used for testing while the rest of the data was used as the training set, a process which is repeated 100 times. The recognition rate is given by the formula:

$$\text{Recognition Rate} = \frac{\text{Correct Classifications}}{100 \text{ Trials}} [100\%] \tag{3}$$

We note that we obtained similar and higher recognition rates for both the object size and shape using the nearest neighbour classifier and a window of only 10 postures (Figure 3), which is equivalent to the data collected in a sixth of a second, as opposed to the 30 posture window in our previous experiments. The recognition rates were of 98% for both the object size and shape. In comparison, in the case of the translation task the highest accuracy was obtained using the k-nearest neighbour classifier on the raw data (98% for both size and shape).

The higher accuracy in the case of the most comfortable holding posture can be explained partly through the specifics of the task, which ensure that each data set will provide a rather stable posture as opposed to a range of varied postures in the other tasks. Another factor is the low percentage of shared postures (Figure 4) compared to the other two tasks, 16% in the case of the holding posture as opposed to 22.3% in the translation task and 65.1% in the exploration task. The percentage of shared postures shows how many postures are common to two different object and it is calculated using the techniques established in our previous work [22].

We also wanted to determine how many times the most comfortable holding postures appeared naturally during the two other tasks, the object translation and

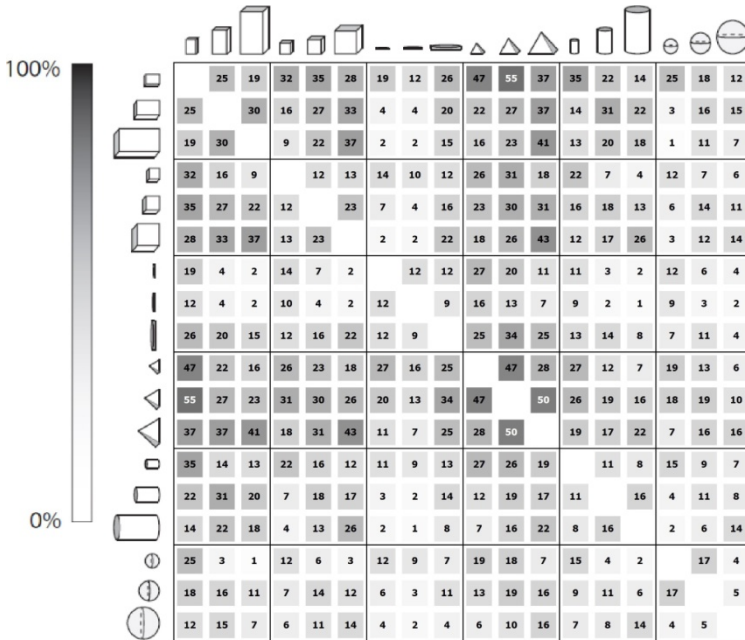


Fig. 4 Shared postures between the objects used in the experiment for the holding posture. A darker colour indicates a higher percentage shared postures between the two corresponding objects.

exploration. In order to calculate the percentage of holding postures used in the other tasks we averaged the holding posture data set for each object of each participant and we compared it to each posture of the corresponding data set from the other tasks. The process was executed for values of the separating threshold between 0.1 and 0.5 with an increasing step of 0.1. The results showed low percentages for the thresholds which provide a better resolution for the interaction postures of an object (Figure 5). Slightly higher percentages were found for the exploration task which provided a more varied range of postures as well as a larger data set, a total of 266595 postures compared to 57606 in the case of the translation task.

Given the size of the window, the used classifier (nearest neighbour) and the low percentage of shared postures between objects, real time recognition of the most comfortable holding posture for an object is possible and recommends it as a potential useful feature in object based interfaces. Considering the low percentage with which it is encountered in manipulating the objects and the high recognition rates, the holding posture offers some opportunities as an identification technique for both the object and the user.

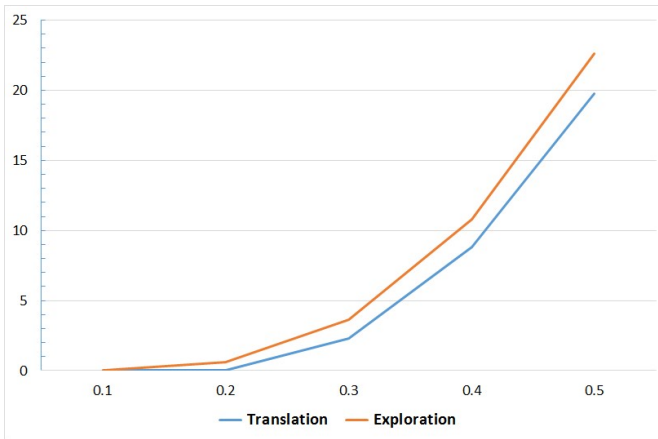


Fig. 5 The encounter percentage of the holding posture for each object in relation to the translation and exploration tasks for various values of the separation threshold

3 Gestural Interfaces for Mobile Applications

In recent years mobile applications have increased in complexity and in the count of their users. The touchscreen remains the main interaction tool for smart mobile devices due to the simple interaction model. We are allowed to directly manipulate displayed entities leading to one of the easiest to use interfaces [6]. However, touchscreens do come with some issues, such as selecting and manipulating a target on the screen, given that the resolution of the human finger is low. This problem is even more troublesome in the case of devices with a small screen estate. Most solutions to this problem originate from Fitts’ law [4] or its variations [19]. Fitts’ law gives us a measure of the time needed to select a target on a device for which we have two dependent parameters *a* and *b* as seen in equation 4. The time to select a target using a cursor is directly proportionate to the distance to the target (*d*) over the width of the target (*w*).

$$T = a + b \log \frac{2d}{w} \tag{4}$$

Various techniques such as offsetting the cursor [1, 13, 17, 23] have been developed to tackle the problem. Improvements have also been brought to the technology behind the touchscreen such as Tactus [21], allowing the touchscreen to become a deformable surface, or TapSense [5], detecting the part of the finger that was used to make contact with the touchscreen (the tip, nail or knuckle).

Even with such improvements the touchscreen still remains limited in the interaction model it allows, more so in relation to the rising complexity of mobile applications. We believe that using the data on hand gestures performed in the interaction with a smart mobile device would greatly benefit both the users and developed applications through the new dimension of supplied input. We implemented several applications using the simplest of hand posture data to show

the potential impact it could have. One of those applications was Paint where the user could draw given a set of simple colours. The addition we made was that the user could associate different colours to different fingers and also had various available actions (Figure 6). An example of such an action is identifying a colour on the screen for which the user could simply touch the screen with his little finger (the colour picker). Rubbing the screen with the side of the hand similar to brushing something off would produce clearing or erasing the screen in the touched area. Similar actions without the use of hand posture data would require either navigating a menu and selecting it or displaying the available actions on the screen and thus losing screen estate.

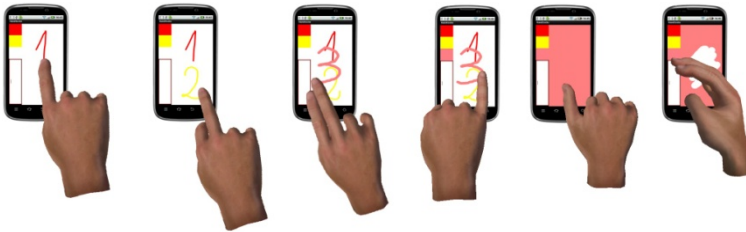


Fig. 6 Paint application using hand postures to differentiate between available actions

We previously mentioned a system in which objects became part of the interface for desktop gaming applications. Object based interfaces have been put forward before in multiple forms [9, 12, 20] but is there an argument for using them in mobile device applications. The main problem to appear with such a system is the handling of both the mobile device and the object at the same time. However, given the continuously evolving interaction between users and mobile devices [15] and the various uses of mobile devices such an interaction is possible and recommended. Current smart mobile devices have the ability to adapt to the user's requirements, being able to control large displays or smart TVs and even transform themselves into desktop systems. A precedent for mobile object based interfaces exists in applications such as Sphero⁵ where the user can either control the Sphero ball with his mobile device as well as the other way around.

4 Creating Gestural Interfaces: Guidelines

Objects from an environment can act as a support for pre-existing gesture commands (as in the Counter-Strike application a cup could be used as a replacement for the grenade which had its own established gesture). However, human gestures are not only a tool for physically interacting with the environment. Supported by our work with object based and gestural interfaces we provide in Table 1 a summary of actions involving hands for such scenarios ranging from simple object manipulation to fine precision gestures.

⁵ Sphero (<http://www.gosphero.com/>)

Table 1 Actions of the hand and fingers and application opportunities

Postures and gestures	Application opportunity
General object manipulation	The interaction we have with common objects all around us (e.g., drinking from a cup)
Precision gestures and complex actions	Using fine finger movements for actions that require precision (e.g., playing the guitar)
Culture-specific gestures	Waving, the 'ok' sign, thumbs-up, etc.
Communication oriented gestures	Conversational hand gestures [11] that assist inter human communication
Involuntary actions	Involuntary hand gestures and finger movements such as finger tapping

While the needed degree of precision and the variation of gestures depend on the application profile we consider that for a ubiquitous application to be complete its gestural interface must use gestures from all those listed in Table 1. The given categories are not to be viewed as independent of one another but overlapping in some contexts. Their goal is not only to cover the range of actions which can be executed by our hands but also to punctuate the varied dimensions of hand gestures and their essential role in our lives.

4.1 Object Manipulation Gestures

An example of such gestures is the Counter Strike application enhanced with hand gestures supported by objects found in the users' surrounding. The user has the option of customizing the interface himself by finding appropriately similar objects to those which are the subject of the given application. The shape and size of objects can be deduced from the gestures performed by the user while manipulating it [16, 22] which would mean that the objects required by the application should be recognizable and easily adjustable to by the user.

We performed a study on students aged 22-23 to provide some insight on how the correlation between physical and virtual environments is made, from the perspective of the available interactive objects. Gaming applications were chosen mainly due to their popularity [2]. We also took into consideration that the first person shooter and role playing game types provide an interaction involving an extensive range of virtual objects such as weapons, tools and household appliances. The participants were asked among others to provide examples of real life objects, with no restriction, that could be related to virtual objects in a game, to represent any entity. As it was expected most correlations were made based on both size and shape, such as using a round plate for a steering wheel or a blanket as a cloaking device in role playing games. However, there were some other results as well regarding perception of scale. Some of the examples did not directly correlate the size of the physical object with the one in the virtual game,

while still allowing a direct manipulation. One example was using a leaf as a boat, another to use a cabinet to represent a building, based on the similar shapes. This provides a potential direction of interaction with virtual entities by change of scale and using gestures defined for shapes as a command interface.

4.2 Precision Gestures and Complex Actions

These gestures refer to those actions which require a high level of accuracy or coordination, such as playing a musical instrument. While such activities may not always be directly required by users they are almost always implied through the requirements of the application itself. Some actions may not be replicated correctly based on the user's hand gestures without the accuracy of the input. However, even if required in order to create a more realistic and involving experience our recommendation is not to ground the application solely on them. From the first study we performed on object exploration and manipulation one of the secondary data we obtained was the reaction of the user. What we noticed was the level of fatigue in the case of finer gestures increased as the needed coordination in such cases requires a greater focus and thus puts a greater strain on the user.

While the first two categories regarded the technical side of gestures and recognition, the next three cover aspects of gestural interfaces that allow the user to feel as a part of the application, instead of just a separate element controlling it.

4.3 Culture Specific Gestures and Communication Oriented Gestures

Culture specific and communication oriented gestures introduce another dimension in gestural interfaces, that of symbolism. Most such gestures represent an idea or a state of mind, thus presenting the application with an input of information, which is much more than a command. The thumbs-up sign, which is executed by holding the fingers closed except for the thumb which is straight, can present multiple meanings depending on the context in which it is executed. It can give a confirmation to a direct question or it can present a state of being. Such gestures have an increased value to applications due to the added information they bring. We see a similar system used in current applications where, based on the current locale, we have present options in the user's language as well as other custom settings. The same concept can be applied to gestures based on the user's own culture, further increasing his integration in the application.

4.4 *Involuntary Actions*

Taking involuntary actions of users into consideration is a step even further from culture specific and communication gestures. These gestures no longer come from the user as a command, they are fully implied data or a raw display of information. Conversational hand gestures along with involuntary actions can be used to deduce the user's emotional state, such as nervousness or anxiousness. AffQuake [14] (based on the first person shooter Quake) or Relax to Win⁶ are examples of pervasive games that use the emotional state of the player to influence various aspects of the game.

5 Conclusion and Future Work

In this paper we presented solutions we implemented for using hand gestures in applications for both desktop systems and smart mobile devices, gestures which were either standalone or supported by objects found in the surrounding environment. This allows users to build custom physical interfaces and have a natural interaction with applications. We complemented our previous study on object manipulation activities (translation, exploration) with an analysis of the holding posture for objects. Following our work and study on gestural interfaces we provided basic guidelines to promote using hand gestures as not only an input but also a provider of information and additional data such as emotional state. We believe our classification of hand gestures by their specific goal and particularities allows a better and a stronger ground for integrating gestures in the development of application interfaces.

Emerging technologies such as LeapMotion⁷ or Myo⁸ provide new ways of acquiring data on hand gestures which could bring benefits and create a momentum in gestural interface development. We believe that an interaction model based on hand gestures either standalone or supported by objects found in the user's environment can be the ground for truly ubiquitous applications. As future work we intend to extend our analysis of the object holding posture, specifically towards user independent recognition. We are also considering the implementation of an object based interaction model for mobile devices to determine the usability of such a model and the study of scale perception and interaction through shape oriented gestures in virtual environments.

Acknowledgements. This paper was supported by the project InteractEdu (Interactive gesture based system for the educational development of school-age children: applications in education, tourism and discovery of patrimony) 588/2012, co-funded by UEFISCDI, Romania and WBI, Belgium”.

⁶ Philips Design - Relax to Win

(http://www.design.philips.com/philips/shared/assets/design_assets/pdf/nvbd/november2009/Getting_emotional1.pdf)

⁷ LeapMotion (<https://www.leapmotion.com/>)

⁸ Myo (<https://www.thalmic.com/en/myo/>)

References

1. Albinsson, P.-A., Zhai, S.: High precision touch screen interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 105–112. ACM, New York (2003), <http://doi.acm.org/10.1145/642611.642631>, doi:10.1145/642611.642631
2. Chatfield, T.: Fun Inc. Why Gaming will Dominate the Twenty-First Century. Pegasus Communications, Inc. (2011)
3. Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X.: Vision-based hand pose estimation: A review. *Comput. Vis. Image Underst.* 108(1-2), 52–73 (2007)
4. Fitts, P.M., Peterson, J.R.: Information capacity of discrete motor responses. *Journal of Experimental Psychology* 67(2), 103–112 (1964)
5. Harrison, C., Schwarz, J., Hudson, S.E.: TapSense: Enhancing Finger Interaction on Touch Surfaces. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST 2011), pp. 627–636. ACM, New York (2011), <http://doi.acm.org/10.1145/2047196.2047279>, doi:10.1145/2047196.2047279
6. Holzinger, A.: Finger instead of mouse: touch screens as a means of enhancing universal access. In: Carbonell, N., Stephanidis, C. (eds.) *User Interfaces for All*. LNCS, vol. 2615, pp. 387–397. Springer, Heidelberg (2003)
7. Huang, Y., Monekosso, D., Wang, H., Augusto, J.C.: A Concept Grounding Approach for Glove-Based Gesture Recognition. In: Proc. of the 7th Int. Conf. on Intelligent Environments (IE 2011), pp. 358–361. IEEE Computer Society, Washington, DC (2011)
8. Huang, Y., Monekosso, D., Wang, H., Augusto, J.C.: A Hybrid Method for Hand Gesture Recognition. In: Proceedings of the 8th International Conference on Intelligent Environments (IE), pp. 297–300 (2012), <http://dx.doi.org/10.1109/IE.2012.30>
9. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1997), pp. 234–241. ACM, New York (1997)
10. Jones, L.A., Lederman, S.J.: *Human Hand Function*. Oxford University Press, Inc., New York (2006)
11. Krauss, R.M., Dushay, R.A., Chen, Y., Rauscher, F.: The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology* 31, 533–552 (1995)
12. van Loenen, E.: On the role of Graspable Objects in the Ambient Intelligence Paradigm. In: Proceedings of the Smart Objects Conference, Grenoble, France, May 15-17 (2003)
13. Lu, H., Li, Y.: Gesture Avatar: A Technique for Operating Mobile User Interfaces Using Gestures. In: CHI 2011 Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 207–216. ACM, New York (2011)
14. Magerkurth, C., Cheok, A.D., Mandryk, R.L., Nilsen, T.: Pervasive games: bringing computer entertainment back to the real world. *Comput. Entertain.* 3(3), 4 (2005), <http://doi.acm.org/10.1145/1077246.1077257>, doi:10.1145/1077246.1077257

15. Nova, N., Miyake, K., Chiu, W., Kwon, N.: Curious Rituals: Gestural Interaction in the Digital Everyday, Near future laboratory (2012), <http://books.google.ro/books?id=XWbelwEACAAJ>
16. Paulson, B., Cummings, D., Hammond, T.: Object Interaction Detection using Hand Posture Cues in an Office Setting. *Int. Journal of Human-Computer Studies* 69(1172), 19–29 (2011)
17. Potter, R.L., Weldon, L.J., Shneiderman, B.: Improving the accuracy of touch screens: an experimental evaluation of three strategies. In: O'Hare, J.J. (ed.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1988)*, pp. 27–32. ACM, New York (1988), <http://doi.acm.org/10.1145/57167.57171>, doi:10.1145/57167.57171
18. Prodan, R.-C., Pentiu, S.-G., Vatavu, R.-D.: An Efficient Solution for Hand Gesture Recognition from Video Sequence. *Advances in Electrical and Computer Engineering* 12(3), 85–88 (2012), doi:10.4316/AECE.2012.03013
19. Sears, A., Shneiderman, B.: High precision touchscreens: design strategies and comparisons with a mouse. *Int. J. Man-Mach. Stud.* 34(4), 593–613 (1991), <http://dx.doi.org/10.1016/0020-7373>, doi:10.1016/00207373(91)90037-8
20. van de Sluis, R., Eggen, J.H., Jansen, J., Kohar, H.: User Interface for an In-Home Environment. In: Hirose, M. (ed.) *Human Computer Interaction, INTERACT 2001*, Tokyo, pp. 383–393 (2001)
21. Tactus Technology, Inc. Taking touch screen interfaces into a new dimension, A tactus technology white paper (2012)
22. Vatavu, R.-D., Zaiti, I.-A.: Automatic recognition of object size and shape via Userdependent measurements of the grasping hand. *International Journal of Human-Computer Studies* (2013), doi:10.1016/j.ijhcs.2013.01.002
23. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: *Proc. CHI 2007*, pp. 657–666. ACM Press (2007)
24. Weiser, M., Brown, J.S.: The Coming Age of Calm Technology. In: *Beyond Calculation*, pp. 75–85. Springer, New York (1997), <http://dx.doi.org/10.1007/978-1-4612-0685-9-6>
25. Zaiji, I.-A., Pentiu, S.-G.: Glove-Based Input for Reusing Everyday Objects as Interfaces in Smart Environments. In: Omatu, S., Neves, J., Rodriguez, J.M.C., Paz Santana, J.F., Gonzalez, S.R. (eds.) *Distrib. Computing & Artificial Intelligence. AISC*, vol. 217, pp. 537–544. Springer, Heidelberg (2013), <http://dx.doi.org/10.1007/978-3-319-00551-5>