Gestural interaction in Virtual Environments: user studies and applications

Fabio Marco Caputo

Department of Computer Science,
Verona

fabiomarco.caputo@univr.it

Abstract. With the current available technology there has been an increased interest in the development of virtual reality (VR) applications, some of those already becoming commercial products. This fact also rose many issues related to their usability. One of the main challenges is the design of interfaces to interact with these virtual environments (VE), in particular for those setups relying on hand tracking as a mean of input. This research project aims to tackle some of the critical aspects of the interaction based on hand tracking in VE by proposing novel interaction technique for object manipulation and gesture based interfaces with intent of addressing relevant issues affecting usability of this kind of VR applications.

1 Background of the research project

Virtual Reality (VR), is a term used for those applications that can simulate physical presence in places of the real world or imagined worlds. These VR applications recreate sensory experiences, which could in theory also include virtual taste, sight, smell, sound, touch, etc. Most current VR environments are mainly composed of visual experiences, through the use of a regular computer screen or with special stereoscopic displays, and some additional sensory information, like sound through headphones or speakers targeted towards users.

The availability of low-cost devices for visualization and interaction gestures are reviving the interest in VR and the technology matured enough to offer many applications of potential interest, not only in games and entertainment, but also for other specific uses such as scientific and medical visualization, augmented reality, etc. What limits the use of immersive 3D environments in these areas, are often the difficulties of interaction that, despite the increasingly sophisticated and relatively cheap tracking devices, make unusable applications that could have significant utility and a large number of users. Therefore it becomes crucial to study and test the different tasks to evaluate which interaction paradigm is the best, before proposing the applications themselves to potential users. Even

though the ability to interact with the virtual environment is not required in order to speak of VR, the most interesting applications of this technology come along with the faculty of performing one or more tasks based on the application purpose. These tasks involve an interaction between both the user and the virtual objects present in the environment (i.e. grabbing a virtual object). Such feature is anything but trivial; depending on the complexity of the task, a large number of problems may arise.

One of the main goal of research in this field is to achieve naturalness of interaction with virtual environments. [4] In order to achieve such goal, research work is currently focusing on two main aspects. The first is the understanding of peoples mental models of interaction with these virtual environments. In fact, due to the interface/devices layer of interaction, users may develop different interaction models from those used in a real physical environment regardless of the kind of task they wish to perform. Understanding this is a key factor in the development of the mid-air interaction interface for a different number of tasks (e.g. virtual assembly, shape modeling, etc.) in order to design an interface perceived as natural by most people. The second one is about the development of actual interaction interfaces possibly based on guidelines derived from the knowledge acquired from the kind of research described above. [8]

The aim of my PhD activity is to investigate on open issues related to object manipulation and gesture-based interfaces in immersive virtual environments both analyzing user preferences and interaction metaphors that implementing and testing practical solutions with low cost hardware.

2 Open issues and research aims

Since the early days of virtual environments, interaction with virtual objects in the scene has been one of the main object of studies. Considering three-dimensional virtual environments, interaction isn't trivial, mainly due to the required mapping between traditional input devices (2D) and the virtual environment (3D). Most common solutions resort to techniques that somehow relate the actions performed in the two-dimensional space of the input device (e.g. mouse cursor or touch) to three-dimensional transformations.

Since it is usual for people to interact with these kind of environments with traditional displays, 3D content is displayed in a 2D rendered image, which hinders content's perception. To overcome both the limitations of the input and the output devices, mainstream solutions for creating and editing 3D virtual content, namely computer-aided design (CAD) tools, resort to different orthogonal views of the environment. This allows a more direct two-dimensional interaction with limited degrees of freedom. Solutions that offer a single perspective view usually either apply the transformation in a plane parallel to the view plane, or resort to widgets that constraint interactions and ease the 2D-3D mapping. [9]

Research has shown that the first approach can sometimes result in unexpected transformations when users are allowed to freely navigate through the virtual environment, and that constrained interactions allow for more accurate manipulations. [10]

Recent technological advances lead to an increased interest in immersive virtual reality settings. Affordable hardware for immersive visualization of virtual environments, such as the Oculus Rift head-mounted display (HMD), ease the perception of three-dimensional content. Moreover, advances in user tracking solutions make possible to know where users' head, limbs and hands are in space. This allows for more direct interactions, mimicking the ones with physical objects. First results of our work showed that mid-air direct interactions with 3D virtual content can reduce tasks' duration and are appealing to users.

Although mid-air interactions show promising results, the accuracy of human spatial interactions is limited. Moreover, the limited dexterity of mid-air hand gestures, aggravated by lack of precision from tracking systems and low-definition of current HMDs, constrain precise manipulations. This precision is of extreme importance when creating or assembling engineering models or architectural mock-ups, for instance.
Even if a large number of methods have been proposed in the literature, most demo systems do not provide easy to use interfaces and the possibility of reaching a high level of precision. This is also due to more intrinsic issues of the tracking task such as the inaccuracy in tracking, occlusions, difficulty in segmentation of different gestural primitives, and other non-trivial problems. There is surely a lot of room for improvements, not necessarily related to a better tracking of body landmark but also to a smart global processing of 3D keypoint trajectories. Several methods have been recently presented for characterizing salient points and for global shape description, with invariance properties and robustness against various kinds of perturbation. Methods with similar characteristics but adapted to the different kind of shape data provided by tracking devices could in principle be applied for the solution of open issues in gestural interaction. An example of smart application of simple geometric processing to realize effective interaction comes from 2D touchscreen interaction, where many gesture recognition applications are not based on complex time series analysis, but on the reduction of the problem to a simple template matching of 2D shapes. The popular 1-dollar recognizer [17] and similar derived methods (also proposed for 3D interaction, e.g. [11] are a clear demonstration of the usefulness of this simplification. This may seem to indicate that in this case the dynamic information may be neglected without losing the meaningful part of the signal. In the following two of the main aspects covered by the thesis are presented along with respective issues that are relevant topic of interest for the HCI community and still need to be further investigated.

Object manipulation in VR environments

There are different ways to interact with virtual object present in a VR scene. The type of interaction is determined by the kind of task the user is allowed to perform and ultimately by the purpose of the application featuring such interactions. One specific kind of interaction is the object manipulation. In this case the user is allowed more or less directly depending on the metaphor proposed by the implemented technique to select an object in the scene e perform one or more different transformations (i.e. translation rather than rotations or scaling operations). This kind of interaction has many limitations, however. First, user's arm has a limited range of interaction due to real world constraints (e.g. it's safe to assume arms are always anchored to user's bodies) and sensors detection volume in case of deviceless setups. This limit can be worked around with the use of a navigation technique. Such technique is required as it also allows different visual perspectives, but it should not be required depending on the application. Manipulation of large objects is also an issue as they tend to obscure the users view during positioning tasks unless additional workarounds are implemented to give the user a way to change the point of a view for a more convenient perspective. [3]

Gestural interaction for VR interfaces

Gestural interaction is particular successful on 2D touchscreens where it is easy to determine the beginning and the end of the gesture by using the contact between fingers and display surface. In 3D deviceless interaction, the beginning of a gesture must be automatically detected from the gesture itself. The use of pattern recognition tools may allow a robust recognition of a gesture learned after the gesture realization, that may be good for a sign language interpreter, for example, but not for a manipulation tool that should give visual feedback within a reasonable time. A possible trick to solve this issue typically applied in interface is to use a second hand or a vocal interface, or the recognition of a coded gesture to tell the system that the gesture is starting or it is finished. Furthermore, in a manipulation interaction gesture start involves also the "grabbing" of the object of interest and gesture end involves also its release. This means that the algorithm should localize with a sufficient accuracy the position of the grab and, more difficult, the desired location of the object release. Especially the last task requires, in our opinion, both a smart geometrical representation of the hand/finger trajectories, possibly invariant to users gestures realization and a smart learning procedure in order to characterize the key actions and the corresponding desired object position. It is a really challenging task, but the ideas that could be applied to find a reasonable solution may be the same applied in classical robust point matching and landmark location. Keypoint trajectories could be simplified and normalized, mappings between trajectories could be encoded as functions, evolution of connected point could be treated as a surface. And, in the same way learning approaches are used to find discriminative keypoints for specialized recognition tasks on 3D meshes [7], it is possible to

think that on geometric encoding of hand trajectories, keypoints able to have a user-independent recognition of gesture limits could be learned through the collection of example data. The use of data collection to learn gestural feature is already applied in HCI, for example in gesture elicitation experiments [14, 1], and, in some sense, with the same approach we could learn how users behave when doing "naturally" simple gestures like grabbing, translating and rotating a virtual object. Registering example gestures and de- coupling 3D trajectories and velocity patterns it would be possible to find specific and invariant keypoints to identify beginning and end of gestures.

Another big problem is related to the accuracy of gestures, that in manipulation tasks is particularly important. The accuracy of object positioning in manipulation is limited by the lack in accuracy of tracking and the possible occlusions of keypoints. However, this problem could be mitigated by the redundancy of the data, and all the research on robust descriptors or partial retrieval can surely help in finding ad hoc solutions for the task. Approaches for partial shape retrieval, like Bag of Words [13, 16] could be, for example, applied to select only the partial information correctly describing the gesture and captured by the device/tracking library used. Furthermore, learning from example, using regression techniques, the relationships between keypoint positions and desired manipulation position could help in improving the accuracy of the gesture localization. Another problem in this particular case is the delay in visual feedback, as the detection of the release gesture requires a backwards analysis and if the grabbed object is moved together with the hand, there is a discrepancy between the expected manipulated object position and the visualized position in the virtual representation.

3 Research work plan

The research project has been summarized in a list of practical goals to achieve by the end of the PhD period, the idea is to cover and provide solutions for interaction with VE based on hand tracking:

- Improve the knowledge of mental models and perception mechanism naturally developed by users, when approaching VR applications, through ad-hoc experiments.
- Derive useful and necessary guidelines to design novel interfaces aimed to achieve a natural interaction with different VR systems.
- Design and implement interactions paradigms and interface modules to be tested with users by creating a set of prototypes to use with appropriate hardware setup for the task.
- Test the validated paradigms and modules in real-world applications (e.g. 3D scene design, virtual museum) with specific user categories.
- Explore different solutions to improve interaction through hand tracking such as machine learning techniques applied to gesture recognition.

With the current state of the research project we already achieved a number of goals while working on the remaining ones in the most recent and future works. The specific methodology applied for each work is described in the next section.

4 Current work and results

A preliminary study has been done to strengthen the knowledge of VR and Human Computer Interaction (HCI) field in scientific research by exploring the most recent literature along with the most relevant works of the past that marked this branch of the field with crucial findings and results. The topics covered by the surveyed works include but aren't limited to those mentioned in Section 1. Subsequent efforts has been put on the development of VR prototypes using novel navigation and interaction methods performing some preliminary studies to evaluate their usability. Both these aspects of the work have been conducted accordingly to the research plan and led to some publications and works in progress.

Classification of manipulation literature The study of literature was a major contribution on the making of a survey on 3D Object Manipulation on which we worked to identify a convenient new taxonomy aimed to classify most of the research works and papers on object manipulation in 3D environments for both immersive VR and not. In our work we examined over 50 works on 3D manipulation techniques and featured over 30 in the survey by classifying them through the new taxonomy. The proposed taxonomy offers two non-exclusive approaches to classify any given work presenting a manipulation technique. The classification of these works can in be in fact based both on their "Environment Properties" (**Figure 1**) and the characteristics of the "Manipulation Metaphor" underlying the technique. (**Figure 2**)

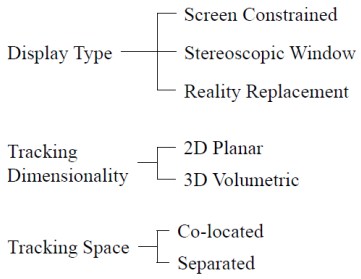

Fig. 1. Taxonomy of Environment Properties

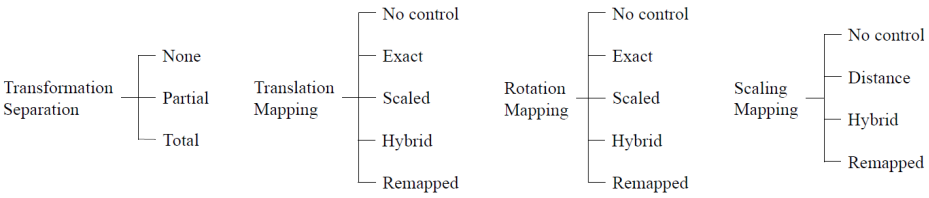

Fig. 2. Taxonomy of Manipulation Metaphors

The literature examined for the survey greatly improved the understanding of the key issues for designing novel interaction paradigms for object manipulation in immersive virtual environment that could possibly bring an interesting contribution to this specific topic when compared to other already known and established paradigms.

Prototyping and evaluation of gestural interface. One of the first works completed in order to derive useful guidelines for future prototypes development was an evaluation of performance and user preferences between 4 different interaction techniques. [6] The interactive environments for our first works have been realized using low-cost setups composed of a Leap Motion controller for hand tracking and a Oculus Rift DK2 for stereoscopic rendering.
In a paper oriented towards Virtual Museum experiences we presented the evaluation results of a series of techniques in VR. These techniques are specifically aimed to cover the basic interaction requirements for a virtual museum experience such as the navigation in the museum space (**Figure 3**) and the and through the information of displayed items. The paper features the details about the interaction design and evaluation test used to validate all the techniques, along with the collected data.

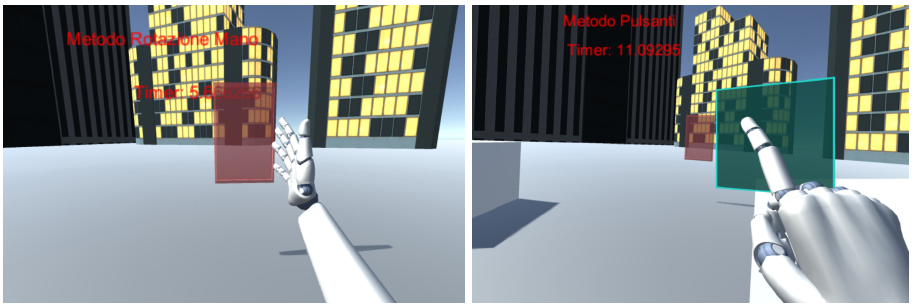

Fig. 3. Examples of two of the four navigation techniques examined [5]

The main contribution of this work was the implementation and evaluation of a number of techniques for information display and the environment navigation:

Information Display techniques:

- Display Buttons
- Swipe
- Object Selection
- Object Picking

Navigation techniques:

- Palm Rotation
- Forward Button
- Mobile Control

Data was collected, for thirty sessions for both tasks. The most interesting results derive from the total time of task completion. These results already show, performance-wise, relevant differences between the proposed solutions. [5]

In more recent works we presented novel solutions for "natural" single-hand manipulation (e.g. picking/translating, rotating and scaling) of object in Virtual Reality environments. The solution is based on the combination of natural gestures with easily recognizable start and end, and no need of explicit or bimanual gesture transitions and smart feedback suggesting potential actions and gesture activation status. The proposed techniques are: the Knob metaphor for performing fast and accurate rotations around selected axis and the Pin method which shows competitive performance results compared to a well known bimanual solution (i.e. the Handle Bar metaphor [2, 15]).

Our goal was to create a set of solutions to allow an intuitive and easy manipulation of objects, given hand and finger information captured by a cheap sensor (e.g. Leapmotion, RealSense), easily adaptable to different contexts and devices. The interaction system acquire the data stream provided by the tracker and processes hand/finger trajectories to determine determine internal state changes, activate manipulation modes and feedback visualization.

Following the hints coming from the literature, we aimed at:

- performing the manipulation control with a single hand
- separating translation from rotation and constrain rotation axes
- being robust against limit of tracking and gesture segmentation accuracy provided by the chosen device/API, especially for object release
- finding an intuitive metaphor for rotation
- avoiding as much as possible the necessity of non-intuitive gestures to to switch modes

The scheme of the designed interaction system for the Knob metaphor is represented in **Figure 4**. The system starts in idle state, when hands are tracked and their position is displayed in the scene, but no interaction with scene objects

is activated. If the hand falls in the region of interaction with a object, specific gestures, however, can trigger the switch to different modes. Furthermore, if the hand is in the right position to start a rotation gesture on the object, a slightly different state is enabled, giving also visual feedback about the enabled rotation axis. In this state a knob rotation gestures starts the rotation as described in the following subsections, locking a rotation direction. Translation and scaling can be activated as well when the hand is in the interaction region, with unambiguous gestures.

In this way basic manipulation modes are completely decoupled, but can be activated easily with a single hand gesture.

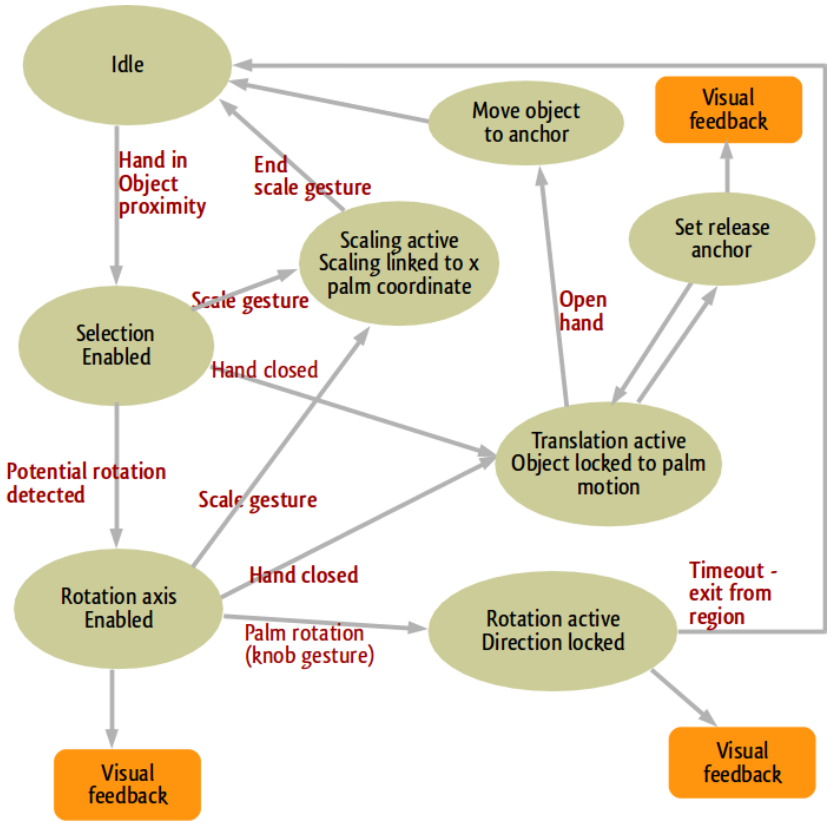

Fig. 4. Knob metaphor FSM representation.

For the Pin method we aim to provide a technique that allows DOF separation between translation, rotation and scaling while keeping the interaction and switching between the manipulation actions as smooth and fluent as possible. In **Figure 5** a snapshot of the prototype currently in development shows the shape of the Pin widget. The two caps (blue and red) work as grabbing points

to enable rotation and scaling mode with a simple tracking of the hand's palm. The translation is enabled by doing the same just with the object itself instead of a pin cap. The idea is to mix advantages of natural interaction for translation with the advantages of a widget like the Pin for the other transformations.

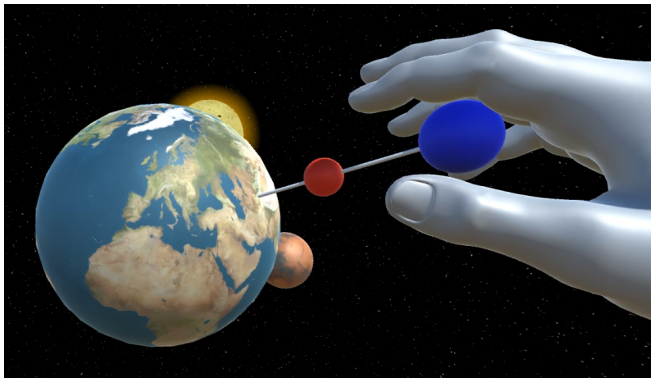

Fig. 5. Sneak peek of the Pin method (WiP).

For both techniques validation has been conducted through an information retrieval task in which the user has to make use of all the possible manipulation actions (translation, rotation, scaling) to extract an "hidden" information (i.e. a small text) on the object rendered in the scene. This particular task was chosen to put more emphasis on learnabilty and interaction fluency rather than high accuracy. Nonetheless the technique also provides a good level of accuracy when considered in the object display applications context.

In our latest work in progress we present a simple 3D gesture recognizer based on trajectory matching in which we provide scores of classification and retrieval of command gestures based on the tracking of single hand 3D trajectories. The work shows good scores and some interesting results such as good performances with trajectories resampled with only few points (from 100 down to 3) or just the initial portion of the original trajectory (down to 20% of the trajectory). Additional results include KNN classification scores for different values of K. The tests have been made on 2 datasets respectively of 26 and 14 gestures. Our method shows how a proper pre-process of data, in particular normalization and rigid transformations, followed by a simple point-to-point distance measure can outperform a previous work presenting a solution for the same task. [12]

5 Short Bio

I am Fabio Marco Caputo, PhD student at the University of Verona. I work on Human-Computer Interaction with focus on Virtual Environments and 3D Manipulation. I'm currently at end of my second year of the PhD programme and my work is supervised by Prof. Andrea Giachetti.

References

1. Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., Zhao, S., Koh, J.: Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Microsoft Research TechReport MSR-TR-2012-111 (2012)
2. Bettio, F., Giachetti, A., Gobbetti, E., Marton, F., Pintore, G.: A practical vision based approach to unencumbered direct spatial manipulation in virtual worlds. In: Eurographics Italian Chapter Conference. (2007) 145–150
3. Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of the 1997 symposium on Interactive 3D graphics, ACM (1997) 35–ff
4. Bowman, D.A., McMahan, R.P., Ragan, E.D.: Questioning naturalism in 3d user interfaces. Communications of the ACM **55**(9) (2012) 78–88
5. Caputo, F.M., Ciortan, I.M., Corsi, D., De Stefani, M., Giachetti, A.: Gestural interaction and navigation techniques for virtual museum experiences. (2016)
6. Caputo, F.M., Giachetti, A.: Evaluation of basic object manipulation modes for low-cost immersive virtual reality. In: Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter, ACM (2015) 74–77
7. Creusot, C., Pears, N., Austin, J.: A machine-learning approach to keypoint detection and landmarking on 3d meshes. International journal of computer vision **102**(1-3) (2013) 146–179
8. Cui, J., Kuijper, A., Fellner, D.W., Sourin, A.: Understanding people's mental models of mid-air interaction for virtual assembly and shape modeling. In: Proceedings of the 29th International Conference on Computer Animation and Social Agents, ACM (2016) 139–146
9. Hand, C.: A survey of 3d interaction techniques. In: Computer graphics forum. Volume 16., Wiley Online Library (1997) 269–281
10. Jankowski, J., Hachet, M.: Advances in interaction with 3d environments. In: Computer Graphics Forum. Volume 34., Wiley Online Library (2015) 152–190
11. Kratz, S., Rohs, M.: A $3 gesture recognizer: simple gesture recognition for devices equipped with 3d acceleration sensors. In: Proceedings of the 15th international conference on Intelligent user interfaces, ACM (2010) 341–344
12. Kratz, S., Rohs, M.: Protractor3d: a closed-form solution to rotation-invariant 3d gestures. In: Proceedings of the 16th international conference on Intelligent user interfaces, ACM (2011) 371–374
13. Lavoué, G.: Combination of bag-of-words descriptors for robust partial shape retrieval. The Visual Computer **28**(9) (2012) 931–942
14. North, C., Dwyer, T., Lee, B., Fisher, D., Isenberg, P., Robertson, G., Inkpen, K.: Understanding multi-touch manipulation for surface computing. In: IFIP Conference on Human-Computer Interaction, Springer (2009) 236–249

15. Song, P., Goh, W.B., Hutama, W., Fu, C.W., Liu, X.: A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems. CHI '12, New York, NY, USA, ACM (2012) 1297–1306
16. Wang, X., Feng, B., Bai, X., Liu, W., Latecki, L.J.: Bag of contour fragments for robust shape classification. Pattern Recognition **47**(6) (2014) 2116–2125
17. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proceedings of the 20th annual ACM symposium on User interface software and technology, ACM (2007) 159–168