Proceedings of the 2nd International Conference on Human-Computer Interaction
Prague, Czech Republic, August 14-15, 2014
Paper No. 124

Evaluating Pointing Accuracy on Kinect V2 Sensor

Hermann Fürntratt, Helmut Neuschmied
Digital – Institute for Information and Communication Technologies
JOANNEUM RESEARCH Graz, Austria
hermann.fuerntratt@joanneum.at; helmut.neuschmied@joanneum.at

Abstract – In this short paper we present an accuracy analysis of the new Kinect V2 depth sensor solely used as a pointing device based on 3D joint positions of the user's arm. Computation of the pointing vector has been done in two ways. First calculated from elbow and wrist position and then from shoulder and wrist position. Based on the evaluation of Fitts' Law we have examined the sensor capabilities with a group of 10 computer affine participants. Results show, that shoulder-wrist based pointing vector calculation performs better (24.3% in average) than elbow-wrist based calculation for clickable targets varying from 39cm to 4.6cm in diameter ordered in distances varying from 28cm to 79.5cm and clicked from a distance of 3.1m. Furthermore we propose that the sensor should be positioned near the ground plane to reduce detection instabilities. Since Kinect V2 preview hardware and API is preliminary and subject to change, we will re-evaluate our findings after official hardware release.

***Keywords*:** 3D pointing device evaluation, Fitts' law, Kinect.

1. Introduction

More than 50 years after Douglas Engelbart has invented the computer mouse (Engelbart 2014), 3D user interfaces are on the rise and research in gesture recognition domain is evolving rapidly (Ibraheem, Khan 2012). Comparisons with traditional input interfaces indicate though, that we are not yet prepared to use all these new natural interfaces in the same way, with the same efficiency as mouse or keyboard (Sambrooks, Wilkinson 2012). Human gestures require a lot of muscular interaction and energy. Figure 1 shows the regions in the primary motor cortex and the corresponding body parts they are in control of. Large areas, like for the hand and fingers indicate, that we can control them with a minimum of energy, not yet physical energy, but also mental energy for eye-hand coordination (Frick et al. 1987).

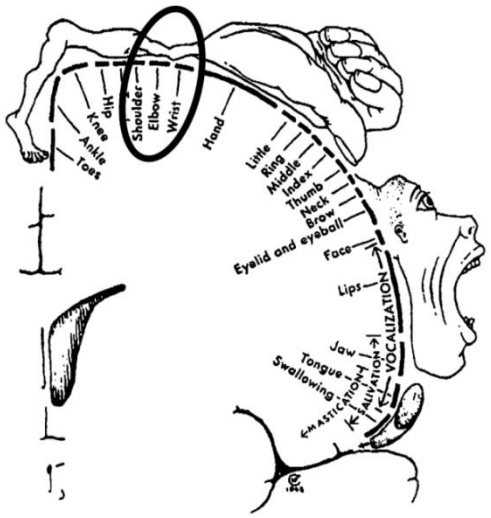

Fig. 1. Mapping of body parts on primary motor cortex.
Source: (Penfield, Rasmussen 1950) with additionally marked parts of interest.

For our test environment we decided to focus on pointing gestures that involve shoulder, elbow and wrist joint positions as base for 3D vector calculation, since hand and finger positions – although more easily to control, as can be concluded from figure 1 – are more difficult to handle due to its smaller effective length, so that a hand rotation of 1.43° in the horizontal plane (yielded by moving

the wrist of a 20cm hand just 5mm aside) results in a pointing displacement of 7.75cm on a screen that is 3.1m away from the user. Pino et al. have evaluated the 2D pointing capabilities of the Kinect V1 (Pino et al. 2013) standing 2m away from a 19'' TFT monitor with 1280x1024 pixel resolution. Computation of the pointing vector did not involve Kinect's 3D joint positions but calculated the cursor location as mean of all hand's pixel x and y coordinates. Depth has not been used for 2D pointing.

2. Fitts' Law

Fitts' Law describes the relation between time to complete an acquisition task e.g. a pointing action, the distance between target and starting point and the size of the acquisition target. It has been developed in 1954 by Paul Fitts and is a fundamental basis for testing the performance of pointing devices (Fitts 1954). Later on it found its way into the ISO 9241-9 standard for evaluating the efficiency and usability of such devices (currently revised by ISO 9241-400 and -410). MacKenzie and Buxton made further improvements and extended it to the 2nd dimension (MacKenzie, Buxton 1992). With Shannon's formulation it denotes to

$$MT = a + b\ log_2\left(\frac{A}{W} + 1\right). \quad (1)$$

The time MT to move a pointing marker into a target area depends on the distance (amplitude A) between start point and target, and the size W of the target area. Term

$$log_2\left(\frac{A}{W} + 1\right) \quad (2)$$

is the so called Index of Difficulty (**ID**). From this formula you can see that a task gets more difficult, the farther away the target is located from the start position and the smaller it is. The time it takes to perform such a task is linearly correlated with its ID (a is start time, b the speed).

3. Experimental Setup

The experiment has been divided into 2 states: a calibration state, in which the participant had to use a laser pointer to point to the top-left and bottom-right corner of the projected test area, and an evaluation state, shown in figure 2, in which Fitts' test has been done in a way, that a participant had to point with one hand (either left or right, as our software detected the pointing hand as the one with the higher wrist) towards targets, and use the other hand to trigger a selection click. This workflow is inspired by results of Schwaller et al. using a Kinect V1 depth sensor for two-handed computer interaction, see (Schwaller et al. 2013) although Kinect V2 offers a new API to trigger events rather easily as described later on.

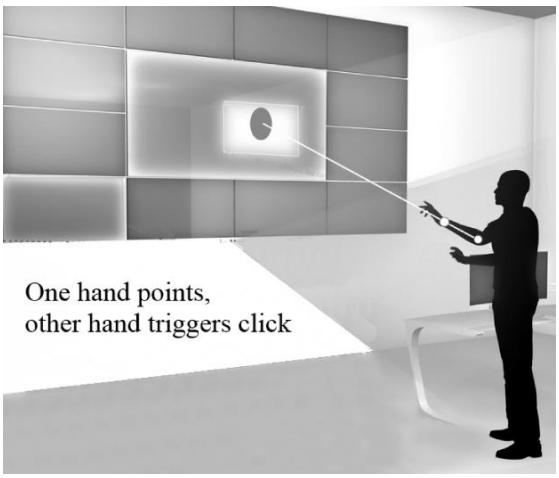

Fig. 2. Initial test setup with Kinect mounted under the canvas and user pointing with his favourite hand.

A single test comprised 5 passes; each pass was defined to click 10 circular targets of a certain radius W, ordered around a circle with diameter A within shortest possible time. After each pass, value A became larger and radius W of the clickable targets smaller, hence increasing the index of difficulty from around 1 to 5.

Table 1. ID values used for evaluation. Effective values (in cm) are taken from beamer canvas.

A [pixel]	W [pixel]	A [cm]	W [cm]	**ID**
70	48	28.0	19.5	**1.28**
100	33	40.5	13.3	**2.00**
135	19	54.5	7.5	**3.05**
160	10	65.0	4.0	**4.11**
196	6	79.5	2.3	**5.15**

The click pattern shown in figure 3 has been chosen similar to the suggested multidirectional tapping task by Soukoreff and MacKenzie with increased ID (Soukoreff, MacKenzie 2004), but with 10 targets per pass and starting position at the right side (East). After running all 5 passes with 50 clicks calculated with shoulder-wrist pointing vector calculation, the test has been repeated with elbow-wrist pointing vector calculation. For evaluation, click and pointer movement data of both test runs have been captured and saved into a local file. In a post processing step, the data has been imported to an adjusted version of the JavaScript based evaluation web page created by S. Wallner, O. Danet, T. Eilersen, and J. Tved (Wallner et al. 2012). Its source code is available under MIT license at github. The results have then been exported to CSV. Test application has been running on a laptop with a 4 core CPU, 4GB of video RAM, 16 GB of RAM, USB 3.0 controller and a 1TB SSD under Windows 8.1. Display resolution of the beamer was full HD (1920 x 1080) although we were operating with depth image resolution of the Kinect V2 (512 x 424). The sensor has been capturing depth data synchronously at 30 fps during the whole experiment at constant artificial light conditions.

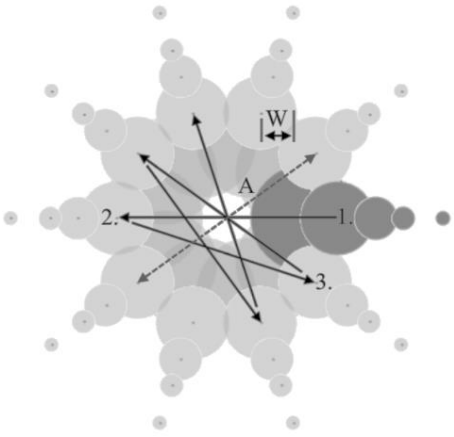

Fig. 3. Click pattern and dimensions of clickable targets with increasing ID.

Since the joint model is prone to instabilities if the user points directly towards the sensor, we have changed the sensor position centred under the beamer canvas to the ground floor at a distance of 2.3m to the user and an inclination of 25° up towards the user's chest.

4. Participants

All 10 test users where male and aged from 27 to 49 (mean 33.5). All of them were highly skilled to use the computer but had no special training on the Kinect sensor. One of them has worn glasses but with proper vision correction. None of them were colour blind. Arm lengths measured from shoulder to wrist varied from 52cm to 59cm (mean 54.6)

Before starting the tests, all users could see themselves on the screen with the skeletal model applied on their body. They have been requested to point to corners and train the clicking behaviour, which has been implemented by evaluating the so called *HandState*, a new property of Kinect V2

API, which works really well. A click is triggered if and only if two *HandState* events occur in the following order: *HandState.Open* (all fingers spread) or *HandState.Lasso* (index finger pointing towards sensor) followed by a *HandState.Closed* (fist). As soon as the user has finished this rather intuitive event sequence, a real mouse click has been emitted by our test software. To calibrate their pointing arms they have been requested to use a laser pointer in the hand pointing in the same direction as their forearm.

5. Evaluation

As result of the test, the time to complete a pointing parcours over 5 IDs with calculation method 1 (pointing vector from elbow-wrist positions) has been compared with test run method 2 (pointing vector computed from shoulder-wrist positions). A scatter plot of time in ms over effective ID for both pointing methods is shown in figure 4. The effective Index of Difficulty ID_e is computed as follows:

$$ID_e = log_2\left(\frac{D_e}{W_e} + 1\right), \quad (3)$$

with D_e as the mean length of the real cursor motion path between start to end points and the effective width W_e defined as

$$W_e = 4.133\boldsymbol{\sigma}, \quad (4)$$

where **σ** is the standard deviation of the hit locations inside the target. **σ** is computed in the target direction and perpendicular to it, and then chosen using the 'smaller-of' heuristic (MacKenzie, Buxton 1992), (Soukoreff, MacKenzie 2004).

Figure 4 – created with adjusted source code of Wallner et al. 2012 – reveals that the average difference between the two pointing vector calculation modes – elbow-wrist and shoulder-wrist – is 24.3%. It also shows that the result is superimposed with noise, originated from instabilities in the joint model. The Dotted line represents results of calculation mode 1 (shoulder – wrist), whereas solid line represents results of calculation mode 2 (elbow - wrist).

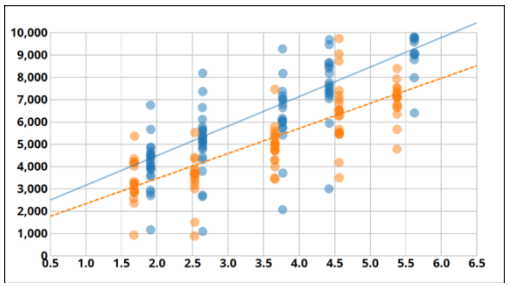

Fig. 4. Movement times MT over the effective ID (ID_e)

6. Conclusion and Future Work

We have presented an experimental study and evaluation of 2 pointing vector calculation modes using the new Kinect V2 depth sensor and its skeletal user model. We have found that quality of pointing calculations heavily depends on the viewpoint of the sensor. Pointing directly towards the sensor results in joint position instabilities that require compensation. Straight forward resolution is to place the sensor on the ground floor directed with a small inclination of 25° towards the user's body.

Since the depth map of the Kinect V2 (preview) sensor is more accurate and the signal to noise ratio higher than for its predecessor, employing the depth values of the arm directly to smoothen joint based instabilities seems to be very promising as this would allow the user to interact without positional restrictions and improved reliability. The influence of human tremor as potential source of jitter will be examined as well.

Acknowledgements

This work has been done under project fusInC which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT). Furthermore, we'd like to thank the authors of the "Visualising Fitts' Law" website – Simon Wallner, Otilia Danet, Trine Eilersen, and Jesper Tved for creating and sharing their work with the HCI community.

References

Engelbart, Douglas (2014): Father of the mouse. Available online at http://www.dougengelbart.org/firsts/mouse.html, checked on 3/3/2014.

Fitts, P. (1954): The information capacity of the human motor system in controlling the amplitude of movement. In *Journal of Experimental Psychology* 47 (6), pp. 381–391.

Frick, H.; Leonhardt, H.; Starck, D. (1987): Spezielle Anatomie II. Eingeweide - Nervensystem - Systematik der Muskeln und Leitungsbahnen: Thieme Verlag.

Ibraheem, N. A.; Khan, R. Z. (2012): Survey on Various Gesture Recognition Technologies and Techniques. In *Computer and Information Science* 5 (3), pp. 110–121.

MacKenzie, S.; Buxton, W. (1992): Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 219–226.

Penfield, W.; Rasmussen, T. (1950): The cerebral cortex of man; a clinical study of localization of function. Oxford, England: Macmillan (XV).

Pino, A.; Tzemis, E.; Ioannou, N.; Kouroupetroglou, G. (Eds.) (2013): Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation. HCI'13 Proceedings of the 15th international conference on Human-Computer Interaction (IV).

Sambrooks, Lawrence; Wilkinson, Brett (2012): Comparison of gestural, touch, and mouse interaction with Fitts' law. In Haifeng Shen, Ross Smith, Jeni Paay, Paul Calder, Theodor Wyeld (Eds.): the 25th Australian Computer-Human Interaction Conference. Adelaide, Australia, pp. 119–122, checked on 2013.

Schwaller, M.; Brunner, S.; Lalanne, D. (Eds.) (2013): Two Handed Mid-Air Gestural HCI: Point + Command. Proceedings of the 15th international conference on Human-Computer Interaction: interaction modalities and techniques - (IV), 388-397.

Soukoreff, William; MacKenzie, Scott (Eds.) (2004): Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. International Journal of Human-Computer Studies: Elsevier Ltd.

Wallner, S.; Danet, O.; Eilersen, T.; Tved, J. (2012): Visualising Fitts' Law. (Source: https://github.com/SimonWallner/uit-fitts-law). Available online at http://www.simonwallner.at/ext/fitts/, updated on 11/17/2012, checked on 2/27/2014.