

Gestairboard: A gesture-based touch typing keyboard using the Kinect camera

Johannes Roith, Tayfur Coskun, Teodora Velikova, Sebastian Klingenbeck, Gudrun Klinker

Fakultät für Informatik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München

{roith, coskun, velikova, klingenb, klinker}@in.tum.de

Abstract: This paper describes a Kinect-based virtual keyboard using finger gestures. Existing systems usually involve expensive range sensors or custom-built hardware equipment. Furthermore, range sensor-based keyboards that are used to detect taps on imaginary keys suffer from shortcomings that were originally reported for touchscreen keyboards: There is a lack of tactile feedback, which has a negative effect on typing performance. The proposed *Gestairboard* extends the *Gestyboard* touchscreen keyboard concept that was suggested in 2011 to overcome exactly that problem, thus adapting it for devices without a touchscreen. It also uses inexpensive off-the-shelf hardware and to the best of our knowledge is the first implementation of a touch typing keyboard using Kinect. The prototype was compared to the *Gestyboard* and to a physical keyboard and evaluated in terms of performance, reliability, learnability and user acceptance.

1 Motivation

The *Gestairboard* originates from a novel gesture-based keyboard concept, known as the *Gestyboard* that was initially developed for use with multitouch input devices. Despite the success of virtual touchscreen keyboards in the mobile market, typing on them has proved to be slower and more error-prone when compared to physical keyboards. In previous work, we mentioned the lack of tactile feedback on conventional touchscreen keyboards as the main reason for their low performance [CAP⁺11] [LP98]. Additionally, any keyboard where users type on flat surfaces is likely to be affected as well. The stated goal of the *Gestyboard* - to overcome this problem while keeping the well-known 10-finger-system - was to be achieved through the use of finger gestures.

When using a conventional English hardware keyboard, users place their left hand fingers on the keys for the characters 'a', 's', 'd' and 'f' and their right hand fingers on the keys 'j', 'k', 'l' and ';', altogether called the *home row*. The thumbs are resting on the spacebar. In the *Gestyboard* concept each home row key defines a finger region. The surrounding keys, which are usually typed by the same finger, are assigned to the respective region. There are

two types of finger gestures: *Tap gestures* are used to activate the home row keys, whereas *sliding gestures* activate all other keys. Since each key can be triggered using a unique gesture, no tactile feedback is needed.

The proposed *Gestairboard* extends the concept of the *Gestyboard* to camera-based virtual keyboards where the user can type directly on a table while exploiting the familiar set of gestures. The Microsoft Kinect motion controller - an input device combining a VGA camera with a depth sensor - enables the detection of finger movements.

2 Related Work

Besides the *Gestyboard*, described above, there is an enormous amount of related literature concerning the use of computer vision in virtual keyboards and the implementation of finger tracking. A good overview of various virtual keyboards, including two computer vision-based touch typing keyboards, is provided in [KT02]. More recently, Huan Du et al. described a system that implements a virtual keyboard by projecting a pattern of colored light onto the users hands. It observes the distortion of the stripes to recover range information from the image captured by a calibrated conventional camera [DC08]. Another very recent virtual keyboard, suggested by Murase et al. [MMS⁺12], additionally uses finger gestures and operates with an ordinary VGA camera, but still requires manual calibration.

The various Kinect software development kits include tracking of body joints except for finger joints, making a custom implementation necessary. The most noteworthy finger tracking projects using Kinect include work by the MIT¹, the open source *Candescent NUI* library, and a method using a detailed 3D model of all finger joints [OKA11]. However, none of these implementations quite met our requirements.

3 Concept and Implementation

The *Gestairboard* was designed to use only a single Kinect camera, to work without a dictionary or word prediction and to enable blind typing at an acceptable speed and error rate. Our prototype is implemented as a C# library that receives a stream of depth images from the Kinect camera, which are subsequently analyzed to detect relevant features. As part of the feature detection module we implemented our own components to detect outlines and orientation of hands, the center of each palm, fingertips (based on [DC08]) and hand movements and to identify and track individual fingers. Layered on top is a module for gesture detection, which, in turn, forms the foundation for a component that maintains the state of the keyboard. The *Gestairboard* shows a simple keyboard on the screen with the current fingertip positions depicted by green circles and additionally displays the computer vision data. Detected keys are sent to the active application to simulate a keystroke.

¹<http://www.ros.org/wiki/mit-ros-pkg>

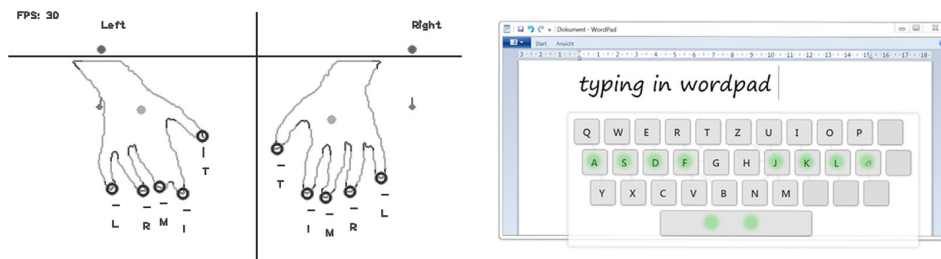


Figure 1: The Gestairboard UI: visualization of detected features (left) and the virtual keyboard overlay (right), which is shown close to the bottom of the display while the keyboard is being used.

4 Evaluation

The prototype was tested with seven participants knowing the 10-finger-system but lacking prior experience with both Gestairboard and Gestyboard. The evaluation included a field-study, an individual interview and a questionnaire. Users had to type 17 English sentences on the Gestairboard, the Gestyboard and a conventional keyboard using the free typing tutor *TIPP10*. Two metrics were used: the error rate (fraction of characters wrongly typed at the first attempt) and the typing speed in characters per minute (*cpm*). The questionnaire included the System Usability Scale (SUS)², the NASA Task Load Index (NASA-TLX)³ and a part with custom questions per tested device. Along with the interviews they enabled us to analyze preferred keys, difficult key sequences, preferred sensitivity of gesture activation, comfort of hand orientation, learnability and mental and physical demand.

5 Results

While the physical keyboard had a significantly lower error rate, we found no significant difference between the error rate of the Gestairboard and the Gestyboard, possibly due to the small number of participants. Tap gestures had a significantly lower error rate than sliding gestures. About 1/3 of all Gestairboard errors were due to unintended spacebar activation. The physical keyboard turned out to be significantly faster than both virtual keyboards (median 184.14*cpm*). Although generally the Gestyboard (median of 27.5*cpm*) appears to be somewhat faster than the Gestairboard (23.99*cpm*), the difference was not statistically significant. This may be due to the sample size. We also found no correlation between typing speed and error rate for either the physical keyboard or the Gestairboard, but could confirm the correlation for the Gestyboard as described in [CAP⁺11].

The Gestairboard obtained a SUS score of 69.9. The NASA-TLX indicated that both tasks were moderately demanding. The Gestairboard has a lower perceived task load with an average score that is just 75% of the Gestyboard score. Temporal demands and

²The SUS is a widely used tool to determine the subjective opinions of the system usability;

³The NASA-TLX is a very popular multi-dimensional scale to determine perceived workload of a task.

frustration were almost twice as high for the Gestyboard prototype. Participants indicated in the Gestairboard questionnaires that they preferred the home row keys. Some keys that required a forward-sliding gesture were disliked, especially if it had to be performed by a ring finger. Some also had problems to lift their ring fingers enough for the system to be able to recognize a tap gesture. Users generally found the Gestairboard pleasant to use and reported little or no frustration. They stressed the importance of better visualization.

Analysis of the recorded computer vision videos (see Figure 1, left) showed that the participants manual dexterity often affected overall performance more than the reliability of the feature detection.

6 Conclusion and Future Work

We have found that the gesture concept works quite well with Kinect, although some users encountered difficulties due to limitations of the human hand anatomy. Most users liked the experience of typing on the Gestairboard. Despite a comparatively high error rate, the evaluation did not indicate that the Gestairboard is significantly less reliable than the Gestyboard. Typing speed is similar to the one of the Gestyboard, although considerably slower than on a physical keyboard. Blind typing is possible as well. It might be worth to reconsider some Gestyboard principles to be able to better adapt the Gestyboard concept to the Kinect hardware. More work is needed to increase reliability, for example by providing better visual feedback. Other future research could include evaluating a modified prototype that features auto-correction or using multiple Kinect cameras to reduce occlusions.

References

- [CAP⁺11] Tayfur Coskun, Eva Artinger, Lorenzo Pirritano, Daniela Korhammer, Amal Benzina, Claudia Grill, Andreas Dippon, and Gudrun Klinker. Gestyboard v1.0: A 10-finger-system and gesture based text input system for multi-touchscreens with no need for tactile feedback. Technical Report 342/22/11 I, 2011.
- [DC08] Huan Du and Edoardo Charbon. A virtual keyboard system based on Multi-Level Feature Matching. *Conference on Human System Interactions, 2008*, 2008.
- [KT02] Mathias Kölsch and Matthew Turk. Keyboards without Keyboards: A Survey of Virtual Keyboards. In *Human Performance*, 2002.
- [LP98] K.T Landauer and P.V. Prabhu. *Handbook of Human-Computer Interaction*. Elsevier Science Inc., New York, New York, USA, 1998.
- [MMS⁺12] Taichi Murase, Atsunori Moteki, Genta Suzuki, Takahiro Nakai, Nobuyuki Hara, and Takahiro Matsuda. Gesture keyboard with a machine learning requiring only one camera. *Proceedings of the 3rd Augmented Human International Conference on - AH '12*, pages 1–2, 2012.
- [OKA11] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis Argyros. Efficient model-based 3D tracking of hand articulations using Kinect. *Proceedings of the British Machine Vision Conference 2011*, pages 101.1–101.11, 2011.