Bartosz POŁOK, Piotr BILSKI
WARSAW UNIVERSITY OF TECHNOLOGY, FACULTY OF ELECTRONICS AND INFORMATION TECHNOLOGY
15/19 Nowowiejska St., 00-665 Warsaw, Poland

A low cost depth sensor-based computer control system for the disabled people

Abstract

The paper presents the system for support interaction of disabled people using a low-cost computer module. The system is based on the Microsoft Kinect device. Its usability strongly depends on the designed software. The architecture of both software and hardware part of the system is discussed. The tests performed on human volunteers are presented as well. The conducted research confirms the usability of the system, showing its disadvantages and limitations.

Keywords: disabled people, human-computer interaction, Kinect, infrared sensors.

1. Introduction

People with physical disabilities are a growing group in the society, encountering difficulties in their daily lives, including the usage of information technology. Computer-controlled devices are designed for the majority of able-bodied users. Therefore people with physical disabilities have limited access to them. In technically advanced countries, such problems are now a prime concern. Devices facilitating interaction with computers include simple solutions such as traditional controllers adapted to the needs of the disabled, modified (usually larger) versions, and specialized equipment with advanced sensors, suitable even for the most severe degrees of disability. Despite the variety of available technologies, most of them are expensive. The most advanced devices cost more than the computer they communicate with. Therefore many people are deprived of the computer, the Internet and their abilities. This calls for the introduction of better and cheaper solutions.

The paper presents the system for the interaction of disabled persons with the computer using the Microsoft Kinect (MK) device. The system allows for the fast and efficient operation of the computer while maintaining a relatively low cost of the equipment used. The structure of the paper is as follows. In the second section, existing solutions for disabled people interaction with the computer are presented. In the third section, the proposed system is briefly introduced. The fourth section presents tests of the system, verifying its usefulness in various environmental conditions. The paper is concluded with the summary and future prospects of the applied solution.

2. Review of existing solutions

Computers are designed without considering the needs of users with disabilities. Therefore, even a slight degree of disability may affect the comfort of using them. Information technologies give the disabled people opportunities of better functioning in the environment, and hence social integration. For this reason, design of devices and software enabling the interaction with the computer is required. The existing solutions are classified into three groups (Fig. 1). They are commercial constructions by transforming standard input/output modules into the specialized versions. Their operation depends on the type of disability to overcome:

- indicator - a metal rod placed in a rack designed to be worn on the head. Using the tip of the rod, one can type on the keyboard or use the touch screen with head movements. The disadvantage of this solution is its low precision. It is supplemented with pads on the keyboard. They have the form of a plate-sized keyboard with holes cut directly above the keys, allowing to prevent the accidental pressing random or multiple keys at the same time. Keyboards with such a cap enable the free position of hands on its surface, which is easier for people with impaired coordination of hands.
- enlarged input devices, such as the keyboard. It is based on the traditional computer keyboard, but larger (making it easier to press a specific button) and heavier, thus making the accidental movement difficult.
- trackball, being the replacement for the mouse. It is available in multiple sizes, selectable according to the disability. The largest ones may be operated using the whole hand or foot.

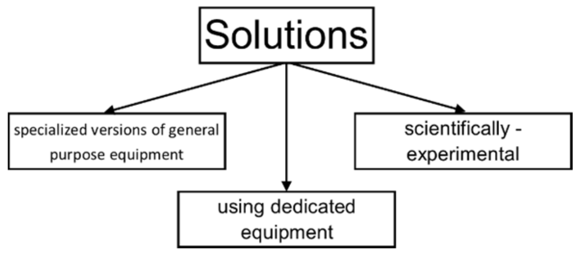

Fig. 1. Classification of solutions supporting integration of disabled people with computers

The specialized general purpose equipment is cheaper than the dedicated hardware. Its prices are higher than for the conventional devices. Enlarged keypad or indicator usually costs about 1,000 PLN ($260), the price for the specialized trackballs is between 500 and 700 PLN ($130 to $180). The efficiency of such equipment is low, because of the slow operating speed.

Commercial solutions using dedicated equipment are based on devices designed specifically for people with disabilities. They are characterized by a better adaptation to the specific disability:

- Intellikeys keyboards use removable boards of many key systems and different size buttons. They allow for wide configuration of parameters, even creating individual tables.
- Magic Wand keyboards. Their usage is implemented by the including stylus and touching the appropriate fields. It does not require the use of force, so the stylus can be held even in the mouth.
- Devices tracking movement of the user's head. The operation involves tracking reflective marker mounted on the head. Its every movement changes the cursor position on the screen. Clicks are implemented by the extra buttons attached to the computer or by stopping the cursor in the desired point for a certain amount of time.
- Eye tracking system is for users with the disabled head. It is placed over the head (the cheaper one), or on the monitor (the more expensive). Both solutions translate eye movements to the movement of the cursor.
- Controllers using movements of the mouth or tongue. The user moves the mouse cursor by manipulating a special mouthpiece, where blowing is interpreted as clicking.

All these solutions have a high precision and speed of operation even when used by severely disabled people. They are also expensive. Prices start from 1500 PLN ($390) to 11000PLN ($2,900) for the first three groups of devices. The most expensive solutions for tracking eye movement cost between 13,000 ($3,400) and 33,000 PLN ($8,550).

The novel solutions for the computer control include scientific and experimental devices, briefly presented below.

Vision systems are based on techniques of analysis and processing of images. They do not need any special equipment. Usually a camera, which is supplied with most of laptops is sufficient. Such systems allow for hands-free computer control. Vision systems use for images of various user body parts their operation. The possible scenario is tracking the head movements. The software keeps track of the selected face fragment (eyes, nose or mouth) and, depending on its position in the frame, moves the mouse cursor. Simulation of pressing mouse buttons is done by stopping the cursor for the specified period of time [1]. Another method of achieving this is to identify the closed mouth. While open, it is used to perform additional actions, such as single or double button clicks [2]. Another option is based on blinking the eyelids. Such solutions typically require a separate interface. They are easy to learn even for completely paralyzed people. the disadvantage is the slow operation [3, 4, 5]. The eye tracking uses the camera and diodes emitting the infrared (IR) light. The latter are installed in the corners of the screen and directed toward the user. The reflection of IR light and the location of the pupil intercepted by the camera allows for determining the part of the screen the user is looking at [6,10]. The solution is extended with an additional camera placed on the user's head directed toward the screen to record the position relative to the IR markers placed in the corners of the screen [7]. A simplified version uses only image captured by the camera. The user's eye is recognized, enabling the cursor control. The camera must be placed close to the user's face [9] or be attached to the glasses, recording directly the eye movement [8]. This approach allows mainly the move of the cursor and click action. It is not suitable for people performing richer computer-based activities.

Solutions with great potential are brain-computer interfaces (BCI). They can be invasive or non-invasive. The former consist of electrodes placed on the surface or in the interior of the cerebral cortex. Electrodes read signals produced even by single neurons. As they are inserted into the user's body, their main application is controlling prosthesis [11]. Non-invasive methods read electrical signals generated by the brain. This is done by electrodes placed on the skull surface and does not require intervention in the tissue. This way only the weak and distorted waveforms are recorded. Three groups of interfaces are used:

- based on responses evoked by visual incentives. Interfaces (called P300 [12]) focus on the sequence of indicated keys. If the key, on which the user's attention is focused is highlighted, the EEG pulse of "brain response" is generated.
- based on processing visual stimuli: the keys flicker continuously, but each flashes at a different frequency. The responses evoked in steady state (SSVEP) are measured here [12].
- using synchronization/desynchronization of the brain motor cortex; when the user imagines an action associated with the movement (like raising the hand), the EEG recording electrodes placed over the appropriate area of motor cortex collect a stronger signal than the ones placed over other areas.

Due to strong interferences, EEG brain-computer interfaces are slow. To detect the user’s intent, multiple P300 signals have to be averaged or long SSVEP responses recorded. Relying on the signals from the motor cortex of the brain requires the user’s training. The information capacity is limited due to the small number of resolvable motor areas on the skull surface [13].

Besides the presented solutions other, less popular approaches, are developed. Such systems are based on different phenomena facilitating reading user's intentions. Such are voice commands and systems for communicating with a single button. Also, electromyography (EMG) or electroculography (EOG) are used. They are at the prototype stage without the commercialization.

3. Details of the designed system

The goal of the presented project was to create the application allowing controlling Windows OS with an MK controller. Two modules of the Kinect SDK were used:

- Face Tracking SDK, which, using data collected from the MK sensor, reads the information about the head's position and face mimicry in the real time. The data is then transformed into information allowing for control of the system.
- Audio API, allowing for defining grammar, according to which spoken words are recognized. It does not require the user to record sound samples. This way commands can be called by every user, who can spell words clearly.

The system was designed to make the whole interaction done using head and face movements with voice commands. This way the Kinect module may become a cheap alternative to the professional equipment and software supporting disabled people, while maintaining the same or better usability. The project consists of two modules described below (Fig. 2).

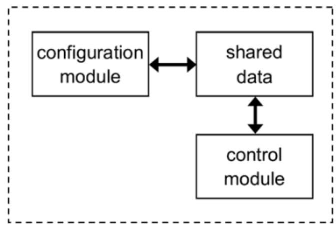

Fig. 2. Control application structure

3.1. Configuration module

It is used to configure the application by adjusting/customizing the range of head movements, triggering the application (Fig. 3). Facial expressions corresponding to the mouse buttons are also defined. It is possible to add voice commands based on the Windows OS "Run" function . The configuration data drives the second module interpreting data collected by the MK and translating them into the cursor movement and actions (like pressing mouse buttons).

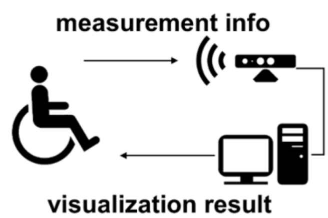

Fig. 3. Remote control system architecture

There is a number of parameters to set up, necessary for the proper operation of the system. The parameters are:

- Deflection of the Kinect
- Movement of the cursor
- Interpretation of left, right and middle mouse buttons
- Voice commands

All parameters are set once for a new user or device location relative to the user. Calibration is done through interaction with new windows of application setup module (Fig. 4). They represent the view of the camera sensor and a graphical representation of the currently read values. The exception is the window of voice commands. Here a box to enter the keyword and a table of user voice commands are present.

Measurement Automation Monitoring, Sep. 2016, no. 09, vol. 62, ISSN 2450-2855

The position of the mouse cursor is based on the user's head. The problem was to adjust the range of movements for various people, by ignoring small, involuntary actions (like convulsions) independent of the monitored person. The range of ignored motion was therefore defined. The selected control scheme operates analogously to the trackpoint (a notebook-like pointer). The head in a natural state does not change the cursor location regardless its position on the screen. The head deflection in one direction starts the cursor movement, while changing the tilt influences its speed. This way the movement in the operating area does not require the excessive force.

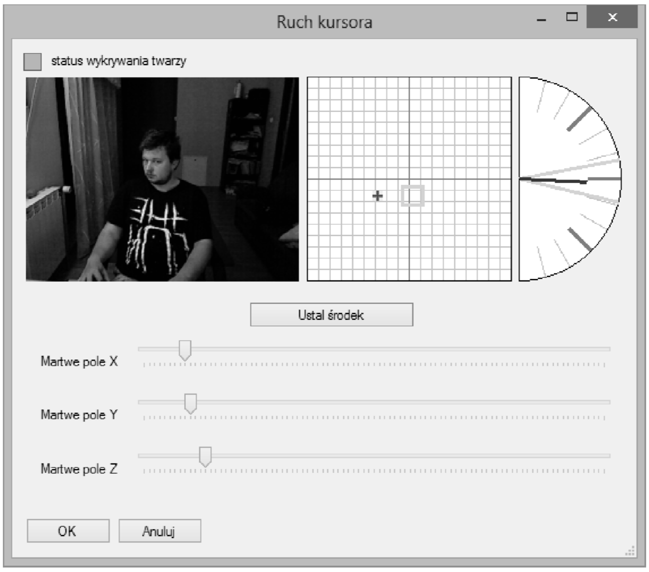

Fig. 4. The configuration screen the range of movement of the head

The next implemented function was scrolling documents. Its implementation uses rotation of the head along the axis perpendicular to the controller. Similarly to the cursor control, increasing the tilt of the head causes acceleration of the scroll. It is also possible to configure the range of ignored motion. Simulation of three mouse buttons was designed so that it is not necessary to use additional accessories. To achieve this, animation units from Face Tracking SDK were implemented. The application allows simulating mouse buttons by presenting the previously configured facial expressions. This way the user can adapt settings to his/her mimicry. Implemented voice commands support control using head movements. This functionality is triggered by the user pronouncing the keyword (here: "kinect"). Afterwards the program starts listening to voice commands. It is active for a specified amount of time; if the command is not recognized, the system enters the stand-by mode, waiting for the keyword again. When the voice command is recognized, the system resets the time to enter the stand-by mode. The following pre-configured commands are interpreted:

Keyboard – allows for entering letters, digits and other characters. The additional command "phonetic" switches regular voice keyboard to the ICAO phonetic alphabet mode. It facilitates entering characters by saying corresponding words("alpha" for "a", "bravo" for "b", etc.)

- Mouse – allows for pressing mouse buttons by issuing commands: "left", "middle", "right" button,
- Start – invokes the "Start" menu of the Windows OS,
- Close – closes an active application,
- Copy, cut, paste – invokes corresponding edition functions.

Besides fixed commands, the user can enter any number of his own voice commands. He defines the command by introducing the word to be detected and a corresponding command (provided the word is correctly recognized).

Commands are defined using the MS Windows "Run" tool. During the setup, the user can verify the entered command by selecting the appropriate line from the list and pressing the "Test Selected" button. If the performance of the action is not possible, the information message appears.

The functions available in Face Tracking SDK allow the configuration module to collect the information about the user's head neutral position, his mimicry and range of the head's motion. This data is presented in the graphical user interface (GUI). This way the user can quickly configure the application according to his needs (Fig. 5).

3.2. Control module

After preparation steps, the control application is executed. It starts minimized without displaying the icon on the taskbar. This way it is not required to minimize any windows at the start. The accidental termination of the system operation is difficult this way. The only visible indication of the system operation is the icon in the system tray. It shows its status: red color means the MK is not properly connected or is unable to detect the user's face. The green color means the face tracking is performed accurately. The violet color is for the listening voice commands mode. Balloons with messages dependent on the application state are also displayed. Pressing the right mouse button or double-clicking on the icon in the system tray invokes the drop-down menu. It allows the user to run a window with a preview of RGB camera or exit the application. Minimizing the GUI is motivated by the practical considerations. The user can place a shortcut to the application in the startup, so it will start automatically with the system and its action will not distract the user in any way.

Thanks to the Face Tracking SDK functions, the controlling module processes the data collected in a real time and converts them into mouse clicks and cursor movement. Based on the Audio API, the system listens to the defined voice commands. In the case of detecting the command, it executes adequate actions.

4. Testing procedure

The system was tested on a group of people in a variety of computer hardware configurations and in different rooms with varying artificial lighting and sunlight. The user's position relative to the equipment was also verified. The application was tested, among others, at a computer desk, on the couch sitting in front of the TV and outdoors during the sunny day. Testers were asked to perform a series of tasks similar to their daily computer activities, as described in subsection 4.2.

Fig. 5. Testing of the control system

4.1. Setup

To control the computer efficiently with MK, it is necessary to have a computer with minimal requirements specified in the device documentation. This includes: Windows 7, dual-core processor 2.66GHz, 2GB RAM, graphic card compatible with DirectX 9.0c). Kinect collects large amounts of data, analyzed and processed in the real time, requiring the sufficiently powerful hardware. Running the application on the system not meeting the requirements causes a significant slowdown of the OS and applications. Another essential step is setting MK according to the manufacturer's guidelines. It must be at least 0.8 m (version for Windows) or 1.2 m (version for Xbox 360) from the user, being in the center of the sensor field and properly illuminated. The incorrect settings cause problems with the detection of head movements and facial expressions. Using the device outside during the sunny day is impossible, as the infrared camera is unable to properly capture the image. Working inside the room may also be problematic, if the user or the computer are directly illuminated by the bright sunlight. Covering the window solves these problems. Before initiating the application, it has to be configured individually for each person. The user has to define his neutral head position, specify blind spot according to the scope of his head's movements and select facial expressions allowing for efficient simulation of mouse clicks. Voice commands are recognized correctly in most cases without any prior configuration. They require clear pronunciation and proper accent, not working properly when problems with speech apparatus are the case. Limitations of Audio API make interpreting commands in English preferable. Usage of the system on-screen keyboard and voice implemented in the application is slower than writing by a person without disabilities on the actual keyboard.

4.2. Experiments

The system has been tested on 12 people of different age. The tests consisted in configuring the system and performing a series of tasks. Tasks included the launch of a web browser, visiting a few websites as well as starting notepad and writing a short text. The tests concluded successfully and each person after acknowledging the control technology was able to perform all the tasks. This proved the system works in practice and experiment outcomes are repeatable. The system was also tested on different hardware configurations. It turned out that it works best on newer generation hardware that exceeds the minimum requirements of the MK. A summary of various configurations and their performance is in Table 1.

Tab. 1. Summary of the hardware used to test the system

Hardware configuration	System startup	Working with the system
Intel Celeron N2840 2.16GHz, 2 GB RAM	yes	impossible
Intel Core 2 Duo P8600 2.4 GHz, 4 GB RAM	yes	possible but slow
Intel Core i7-3610QM 2.3 GHz, 16 GB RAM	yes	comfortable
Intel i7-3630QM 2.4 GHz, 16 GB RAM	yes	comfortable
AMD FX-8320 3.5 GHz, 16 GB RAM	yes	comfortable

The system was launched on each computer, but not in every situation the proper work was possible. The convenience of working with the system was described with the following phrases:

- comfortable – the control was carried out without noticeable delay, system operation does not affect the operation of other software
- possible but slow – the control was carried out with possible delays to approx. 0.5s of reaction to the head movement to move the cursor. Sometimes the operation of the system slows down another software running at the same time.
- impossible – the delay was too big to precisely control the cursor; the launch of the system resulted in maximum CPU load.

5. Summary and conclusions

The application to control the computer by head movements using an MK sensor allows for the efficient computer control under the set of conditions. Due to its modular structure, one can easily make corrections, enhancements of the functionality and adapt the project to the new version of the SDK. The program is a cheap alternative to professional equipment and applications developed for this purpose, while maintaining comparable utility. As a result, it broadens the range of computer users among people with disabilities who cannot afford commercial specialized solutions. It is also an alternative for scientific experiments. After the initial setup, it enables independent and non-contact interaction of a disabled person with a computer, not requiring, in contrast to other solutions, preparation for work each time (e.g., by setting the electrodes to the head of a user or the eye-tracking camera). The user familiar with the application is able to control the mouse cursor as efficiently as the able-bodied person using the trackpoint. Activities such as double-clicking, or clicking and dragging do not cause problems. Entering text is less efficient than using the hardware keyboard. It can be improved by giving up the voice recognizing functionality of the Kinect SDK in favor of Google Speech API. The system is based on the first version/generation of the MK device. It is then limited by all its disadvantages. As the second generation of Kinect has been introduced recently, the improvements are possible. Rewriting the application for a new version of the device and SDK will increase cost of the solution and hardware requirements of the computer (which might not be a problem in the future). The proposed system is therefore economically attractive and flexible to be used in practice.

6. References

[1] Betke M., Gips J., Fleming P.: The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities. IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 10, Issue: 1, pp. 001-010, 2002.

[2] Song Y., Luo Y., Lin J.: Detection of Movements of Head and Mouth to Provide Computer Access for Disabled. Int. Conf. on Technologies and Applications of Artificial Intelligence, Chung-Li, pp. 223-226, 2011.

[3] MacKenzie I. S., Ashtiani B.: BlinkWrite: Efficient text entry using eye blinks. Universal Access in the Information Society, Vol. 10, Issue 1, pp. 069–080, 2011.

[4] Grauman K., Betke M., Lombardi J., Gips J., Bradski G.R.: Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. Universal Access in the Information Society, Vol. 2, Issue 4, pp. 359-373, 2003.

[5] Krolak A., Strumillo P.: Vision-based eye blink monitoring system for human-computer interfacing. Conference on Human System Interactions, Krakow, pp. 994-998, 2008.

[6] Hennessey C. A., Lawrence P. D.: Improving the Accuracy and Reliability of Remote System Calibration-Free Eye-Gaze Tracking. IEEE Transactions on Biomedical Engineering, vol. 56, no. 7, pp. 1891-1900, 2009.

[7] Kocejko T., Bujnowski A., Wtorek J.: Eye mouse for disabled. Conf. on Human System Interactions, Cracov, pp. 199-202, 2008.

[8] Raudonis V., Simutis R., Narvydas G.: Discrete eye tracking for medical applications. 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies, Bratislava, pp. 001-006, 2009

[9] Arai K., Mardiyanto R.: Eye-based HCI with Full Specification of Mouse and Keyboard Using Pupil Knowledge in the Gaze Estimation. Information Technology: New Generations (ITNG), 2011 Eighth International Conference on, Las Vegas, NV, pp. 423-428, 2011.

[10] Oyabu Y., Takano H., Nakamura K.: Development of the eye input device using eye movement obtained by measuring the center position of the pupil. IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, pp. 2948-2952, 2012.

[11] Harrison R. R.: The Design of Integrated Circuits to Observe Brain Activity. Proc. of the IEEE, vol. 96, no. 7, pp. 1203-1216, 2008.

[12] Bin G., Gao X., Wang Y., Hong B., Gao S.: VEP-based brain-computer interfaces: time, frequency, and code modulations [Research Frontier]. IEEE Computational Intelligence Magazine, vol. 4, no. 4, pp. 022-026, 2009.

[13] Muller K. R., Blankertz B.: Toward noninvasive brain-computer interfaces. IEEE Signal Processing Magazine, vol. 23, no. 5, pp. 128-126, 2006.

Received: 16.06.2016 *Paper reviewed* *Accepted: 01.08.2016*

Bartosz POŁOK, MSc, eng.

Bartosz Połok has graduated in computer science from Warsaw University of Life Sciences in 2014. He is currently a PhD student at the Faculty of Electronics and Information Technology of the Warsaw University of Technology. His research interests involve artificial intelligence.

e-mail: bpolok@ire.pw.edu.pl

Piotr BILSKI, PhD, DSc

He graduated from Warsaw University of Technology, Institute of Radioelectronics, obtaining MSc degree in 2001 (with honors), PhD degree in 2006 (with honors) and DSc degree in 2014 in computer science. Currently he is an Associate Professor in the Institute of Radioelectronics and Multimedia Technologies, Warsaw University of Technology. His scientific interests include diagnostics of analog systems, design and analysis of virtual instrumentation, application of artificial intelligence and machine learning methods to the acoustics and environmental sciences.

e-mail: