Software system for contactless text input based on computer vision

Pavel Peresunko[1*], *Alexander* Bochkarev[1], *Egor* Larionov[1], *Evgenia* Peresunko[1], *Victoria* Kosheleva[1], and *Igor* Kovalev[1,2,3,4]

[1]Siberian Federal University, Krasnoyarsk, Russia
[2]Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia
[3]Krasnoyarsk State Agrarian University, Krasnoyarsk, Russia
[4]China Aviation Industry General Aircraft Zhejiang Institute Co., Ltd, China

Abstract. The article is devoted to the problem of contactless human-computer interaction. The task of entering text from an optimized on-screen keyboard using cursor control based on data received from a video camera is considered, including if the user has motor impairments. A software system consisting of two applications is proposed for cursor control and text input based on computer vision algorithms. A diagram of the system operation and interaction of components and basic user settings are presented. As a result, testing was carried out using a focus group, in which participants had to enter free text using the proposed software system. Experiments have shown the functionality of the system.

1 Introduction

Human-computer interaction includes a wide range of actions carried out with the aim of transferring information from the user when solving various problems that are both entertaining and practice-oriented. When using a personal computer, the most popular method of transmitting a signal is using physical input devices - mouse, touchpad, keyboard. However, in the presence of a number of physical, physiological and other restrictions, it is impossible to use devices that require direct contact with the user. An example of such a situation would be control of a computer by a patient with motor impairment.

There is a need to develop a contactless tool to ensure human-computer interaction. Computer vision technologies solve this problem by providing a wide range of algorithms and tools for object recognition that can be used to control the cursor. The control object can be a hand, parts of the face, parts of the body, colored marks, special codes [1-3].

An equally important aspect of interaction between a computer and a person is entering text, either through a physical or on-screen keyboard, or using voice input. For a system of contactless human-computer interaction, we apply an approach using an on-screen keyboard built into the operating system of the user's personal computer and voice input. However, the latter, in turn, has a number of disadvantages associated with the complication of the input process, which increases the number of errors, for example, those associated with speech

* Corresponding author: pvperesunko@yandex.ru

recognition. In addition, in the above conditions of use by a patient with neurological disorders, manifestations of dysarthria are possible, which makes the use of this input method impossible.

For on-screen keyboards in Russian, various layouts can be used. Thus, in the Windows 11 operating system, the Russian layout “YTSUKEN” is used by default, and in a number of distribution kits of the Linux family you can find the “Russian (typescript)” layout, which differs in the location of non-letter characters and the letter E. You can also find the typographic layout of Ilya Birman, which allows type special characters without using character codes. The main disadvantage is that such on-screen keyboards have a layout similar to a physical keyboard for blind ten-finger input, which is not suitable for entering text with a single pointer (mouse cursor) [4].

The text input system can be represented as follows. The user moves the marker to move the computer cursor. The cursor is used to enter characters on a regular on-screen keyboard. A limitation of computer vision cursor-moving systems is the requirement to be able to move the cursor across a rectangle on camera video. Roughly speaking, the position of the relative rectangle marker of the active area is scaled by the position of the cursor on the screen (Figure 1).

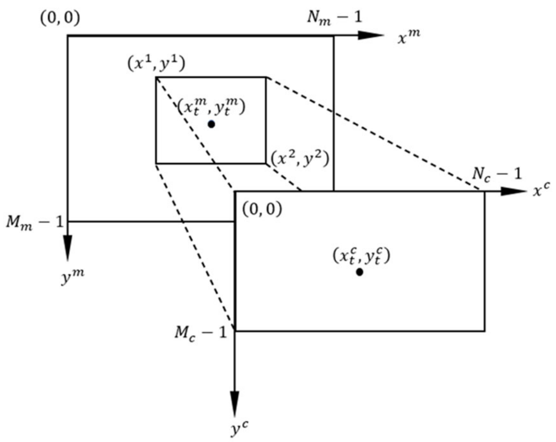

Fig. 1. Projection of the marker position onto the screen when controlling the cursor with absolute positioning.

In addition, standard on-screen keyboard layouts, which are designed for ten-finger input, are not optimal for entering text with a single pointer. The work presents a system that corrects both shortcomings.

2 Methods

Of course, this problem could not go unnoticed, so attempts have already been made to create applications and devices. For example, voice assistants that allow you to control a computer with your voice, special devices that replace standard input devices, or applications for controlling a computer with gestures [5-6].

The system proposed is a complex of two applications, the interaction of which is schematically depicted in Figure 2.

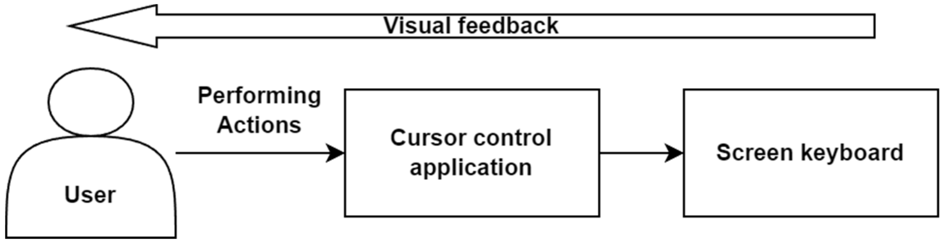

Fig. 2. Interaction of system components.

The first application allows you to control the computer cursor using a video camera. The second is an on-screen keyboard that allows you to enter text with a minimum distance that the cursor needs to travel. Applications can be used both in combination and independently of each other, which allows you to adjust the operation of the system to the needs of a specific user.

A video camera was chosen as the main interface for human-computer interaction, since this device is present in most modern laptops and can be purchased at a low price compared to highly specialized equipment. By processing data from the camera, you can obtain various information: the user's surroundings, the position of certain objects, a person's pose, facial expression, the trajectory of the iris, and so on, which can be used to control the computer.

To develop the system, the Python programming language was used using the PyQt5 library.

2.1 On-screen keyboard application

As stated earlier, the main drawback of typical on-screen keyboards is the shape of the keyboard and the key layout. In this work, the keyboard from [4] was modified. Figure 3 shows the layout.

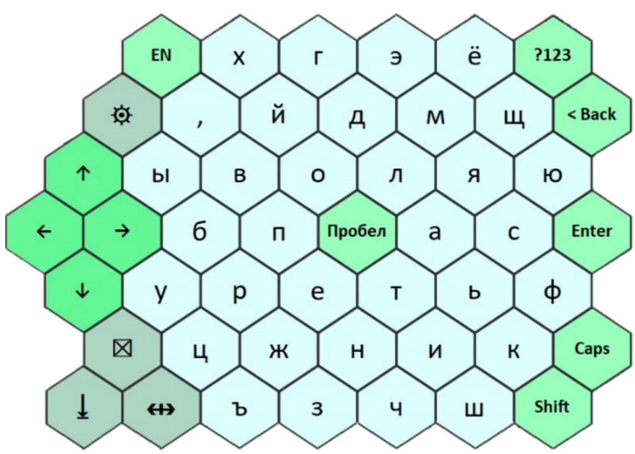

Fig. 3. Screen keyboard

The layout of the alphabet characters was optimized using the particle swarm method, where the function of the weighted pairwise distance between the characters of bigrams of Russian text was used as the optimized function. In addition, auxiliary keys and special characters have been added for control. Keyboard settings allow you to tailor the application to the user, considering key size, color and other characteristics. The click is carried out in two ways - by holding the cursor on the key for a user-specified time and by pressing the left mouse button. In the case of contactless control, a user-defined action can be used as a left mouse button click.

2.2 Cursor control application

An application was developed to control the cursor, which allows you to control the computer using the camera in a non-contact manner. The main task of the application is to analyze user actions and, based on them, move the cursor and click on mouse buttons.

Before the main setup of the application, the user is given the opportunity to select one of the cameras connected to the computer, which will later be used to receive video data. Next, the user needs to select and configure action processing algorithms that will be used to analyze user actions.

One of the possible actions to control the cursor is to move the marker along a user-specified arbitrary path. The trajectory is specified by a curve that is located in the image plane. The position of the marker on this trajectory is the projection of the marker onto the line, calculated as the closest point on the line to the marker. Various objects can act as a marker: colored objects, key points of the body, and so on. The MediaPipe library was used to detect body points. Tapes are supposed to be used as colored objects, since they can be easily attached to any part of the user’s body. To detect colored objects, an algorithm was implemented that allows you to detect the largest object falling within a certain color range. To implement this algorithm, the OpenCV library was used.

Cursor movement can be either relative or absolute. We decided to separate the movement in the spatial dimensions of the screen, that is, moving the cursor along the X axis and the Y axis is performed using different actions.

With absolute movement, the position of the marker on the line corresponds to a certain point on the screen. The implementation of relative cursor movement can be compared to using a joystick. If the marker is at the beginning of the curve, the cursor moves in one direction. If the marker is at the end of the curve, the cursor moves in the other direction. If the marker is in the center of the curve, the cursor does not move. In this case, the speed of cursor movement is directly proportional to the distance from the center of the curve.

Pressing buttons was implemented in a similar way. The curve is divided into a zone of inaction and a zone of action. If the marker is in the first zone, then nothing will happen. If the marker falls into the second zone, a key will be pressed.

Two click processing strategies were implemented. Single click strategy, when the marker hits the first zone, a single click occurs (Fig. 4(a)). A key hold strategy in which the key will be pressed until the marker enters the idle zone (Fig. 4(b)).

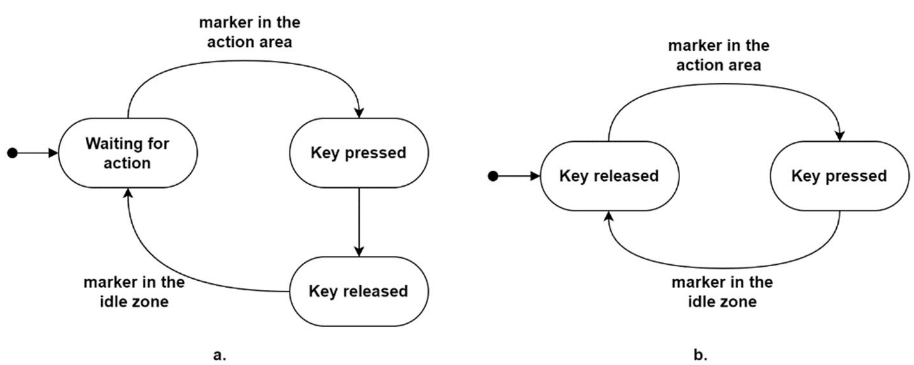

Fig. 4. Key state transition diagram.

2.3 System effectiveness assessment

An action plan was drawn up for testing the system in order to assess the quality of work and the effectiveness of the implemented methods, which includes the following steps: internal testing of the system, testing the system on a focus group of volunteers with free-form feedback, conducting controlled experiments to assess speed and quality printing in comparison with physical devices and contactless analogues. This article describes the results of focus group testing.

The focus group consisted of 25 volunteers with a high level of PC proficiency, aged from 11 to 56 years. The experiment was carried out under the same conditions: the hardware used was a laptop with an Intel Core i7 12gen processor, 16 Gb RAM, 1024 Gb SSD with a built-in webcam with a resolution of 480p and a frame rate of 30 FPS, the Windows 11 operating system, text input software – Notepad++. The users' task was to enter arbitrary text ranging from 10 to 80 characters in length. The evaluation criteria are the usability of the keyboard interface and the clarity of the system settings; reviews are formulated in any form.

The on-screen keyboard was located at the bottom left of the screen. The cursor control application was located at the top right of the screen. It was run in miniature mode, which displayed only the video from the camera, the positions of the detected markers, and the lines along which the markers were moved to control the cursor. This allowed the user to observe the process of analyzing his actions.

Three action handlers were created as settings for the cursor control application:

1) Relative cursor movement along the X axis. The line used is shown in Figure 5(a). The area in which the cursor does not move is highlighted in gray. The marker was a green tape attached to the user's right wrist;

2) Relative movement of the cursor along the Y axis. The line used is shown in Figure 5(b). The markers were the same as in the first processor.

3) Emulation of right mouse button click. The inactivity zone occupied the lower half of the line, and the action zone occupied the upper half of the line. The line used is shown in Figure 5(c). The user's left wrist served as a marker.

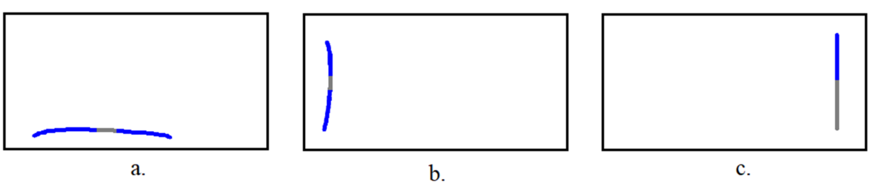

Fig. 5. Lines used to control the cursor.

3 Results

While assessing the effectiveness of the computer control method proposed to users, it was found that the system allows you to perform basic actions when working with a computer without using a physical keyboard and mouse. Analyzing the user's actions, in this case moving the marker relative to a certain trajectory, the application controls the cursor, which allows you to perform some actions when controlling the computer. Adding a keyboard optimized for cursor input can make operating your computer more convenient.

4 Conclusion

The article discusses the development of software for human-computer interaction, designed for non-contact cursor control and typing, including for the purpose of providing computer access to people with disabilities. The proposed system consists of two applications: cursor control software and an on-screen keyboard. The cursor control application provides the ability to customize markers and the paths they will follow, allowing you to customize the system for each individual user.

Further research involves conducting a series of experiments on cursor control and typing with the target group, as well as adapting the system to the task of rehabilitating patients with neurological disorders based on information about the needs of different classes of users.

The use of computer vision methods to develop a system for contactless human-computer interaction makes it possible to expand the scope of user actions used to enter information when working with a personal computer.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No.075-15-2022-1121).

References

1. A. Erdem, E. Erdem, Ya. Yardimci, M. V. Atalay, A. E. Cetin, Computer vision based mouse *IEEE International Conference on Acoustics, Speech, and Signal Processing* Proceedings (USA, Orlando, 2002)
2. D. S. Kulkarni, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering **6(4)** 2716-2723 (2017)
3. A. Erdem, E. Erdem, Ya. Yardimci, M. V. Atalay, A. E. Cetin, Computer vision based mouse *IEEE International Conference on Acoustics, Speech, and Signal Processing*: Proceedings (USA, Orlando, 2002)
4. M. A. Borodynya, V. Doychev, A. Z. Tsyrenzhapov, P. V. Peresunko, Implementation of the simulated annealing algorithm for selecting the optimal key layout on a hex keyboard AIP Conference Proceedings **3021 (**2022) (to be published)
5. H. A. Alhamzawi, Applied Medical Informatics **40(3)** 39-44 (2018)
6. H. Grif, C. C. Farcas, Procedia Technology **22** 657-661 (2016)
7. K. Niyazi, V. Kumar, S. Mahe, S. Vyawahare, International Journal of Information Sciences and Techniques **2(2)** 57-63 (2012)