Proposal of cursor operation method by facial movements for the handicapped

Shunya Mito , Zhang Lifeng

^aFaculty of Engineering Kyusyu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyusyu, Fukuoka, 804-8550, Japan

*<mito.shunya613@mail.kyutech.jp> , <zhang@elcs.kyutech.ac.jp>

Abstract

In recent years, the number of people with physical disabilities has been rising, with about half of them being physically disabled limb. A physically disabled person with a physical disability interferes with daily life, such as missing or malfunctioning fingers. This study focused on the use of computers in daily life. There are three ways for physically challenged people to use a PC: "voice input," "eye input," and "input using a 3D camera. Those problems need to keep the distance between the face and the camera constant, the inability to protect privacy when people are around. Furthermore, 3D cameras are difficult and expensive to install. This study developed a system that allows a handicapped person to operate a computer with only facial movements without using arms or hands. We also evaluated the operability, improving the problems of previous studies.

Keywords: Image processing, machine learning, dlib.

1. Introduction

1.1 Research Background

Currently, there are more than 4 million people with disabilities in Japan. Among them, the number of disabled people aged 65 and older is more than 3 million, accounting for more than 70% of disabled people. The number of people with physical disabilities is expected to continue to increase in the current super-aging society.⁽¹⁾

Table 1 summarizes the number and percentage of people with different types of physical disabilities in 2016.

About half of the disabilities are found to be limbs. Physical disability is a limb, trunk.

Table1. Number and percentage of people by type of disability in Japan

Types of disabilities	Number of people	Percentage (%)
Disability of limbs	1931000	40
Internal Failure	1741000	36
Disabled species unknown	462000	10
Visual impairment	312000	7
Hearing and speech disorders	341000	7
Total	4787000	100

A condition in which a person's ability to perform activities of daily living is impaired by an illness or injury.

1.2 Previous studies

With the number of physical disabilities increasing every year, it is necessary to develop a system to help overcome the disabilities and use computers hands-free.

Several studies and products exist as the three listed below.

(2) (3)

(1) Voice Recognition

Voice recognition allows you to instruct your computer with your voice to launch various application software and enter text.

Advantages of hands-free use and voice input make typing easier than keyboard input.

If there are people around, the disadvantage is that they

will be concerned about their privacy.

(2) Line of sight input

Eye tracking allows the patient user to operate a computer with just eye movements. When mastered, it provides a highly flexible operating environment, just like a mouse. The advantage is that there is a less physical strain on the user. The disadvantage is that the input system is prone to instability and needs to be corrected each time the face and camera position change.

(3) Input with a 3D camera

The "Kinect mouse" is an input interface that uses face and head movements. The 3D camera recognizes facial movements to perform operations such as clicking and double-clicking. The advantage is that it can distinguish more than 50 different gestures and has a wide range of operations. On the other hand, it is expensive because it uses a 3D camera, and the disadvantages are that it reacts to the habitual movements of the face.

1.3 Research Objectives

In this study, we propose a system that enables people with physical disabilities to use computers in their daily lives without using their hands and arms at all, using web cameras and image processing technology to improve the disadvantages of previous studies, such as privacy, need to calibration and high cost, finally, to evaluate the operability of the system.

2. Principle

The system is written in python and the libraries are OpenCV, PyAutoGUI, Numpy and Dlib.

2.1 OpenCV

OpenCV (formally known as the Open Source Computer Vision Library) is a BSD-licensed open source library containing hundreds of computer vision algorithms, developed by Gary Bardsky in 1999 at Intel. Released in 2000, OpenCV now supports a number of algorithms related to computer vision and machine learning, and it is growing every day.

2.2 PyAutoGUI

PyAutoGUI provides cross-platform modules for GUI automation based on human conscious behavior, allowing you to move and click the mouse cursor, drag it, press and hold the keys on the keyboard, and combine them.

2.3 Numpy

NumPy is an extension for efficient numerical computation in the Python programming language. It adds support for typed multidimensional arrays (which can represent vectors and matrices, for example) to Python for efficient numerical computation, and provides a large high-level library of mathematical functions for manipulating them.

2.4 Dlib

Dlib is a general-purpose cross-platform software library written in the C++ language. As of 2016, software components have been developed for processing in a wide range of areas, including n-networks, threading, GUIs, data structures, linear algebra, machine learning, image processing, data mining, XML, and text parsing, mathematical optimization, and Bayesian networks. In recent years, efforts have focused on developing a broad set of statistical machine learning tools. In this system, we used Dlib for face motion discrimination. We explain the discrimination method in the following sections.⁽⁴⁾

2.5 Face motion discrimination method

In this study, face movements were detected, and the coordinate information of the face's feature points was acquired by facial landmark detection in Dlib to calculate the distances between the coordinates and discriminate by the relationship of the distances. As shown in Figure 1, up to 68 coordinate information was obtained for each face, including contour, eyes, nose, and mouth.⁽⁵⁾

In this study, face movements were detected, and the coordinate information of the face's feature points was acquired by facial landmark detection in Dlib to calculate the distances between the coordinates and discriminate by the relationship of the distances. As

shown in Figure 1, up to 68 coordinate information was obtained for each face, including contour, eyes, nose, and mouth.

Fig 1 68-point landmark of the face

The face orientation condition, eye open/close condition, and mouth open/close condition are shown below.

(1) Face Orientation Condition

We estimated the face orientation in terms of the relationship between the distances between the four coordinates, A, B, C, and D, using the six coordinates of the face as shown in the figure2.

Fig 2 The distance between the coordinates used for face orientation determination

First, in comparing the distance of A with the distances of B and C, we tripled the distance of A for the purpose of normalization, and then compared the distances of B, C, and D with that distance (henceforth referred to as the distance of A).

(I) Rightward discrimination condition

When the distance of B is less than one-half the distance of

A and the distance of B is less than one-half the distance of C

(II) Leftward discrimination conditions

When the distance of C is less than one-half the distance of A and the distance of C is less than one-half the distance of B

(III) upward discrimination conditions

When the distance in D is greater than 0.7 times the distance in A

(IV) Downward Discrimination Conditions

When the distance in A is greater than 2.8 times the distance in D

For conditions other than the above four, we had them identified as "front".

(2) Eye opening and closing conditions

The clicking action was implemented to close both eyes for about one second. This is to prevent the click action from being performed when blinking. The Eye Aspect Ratio (EAR) of the eye coordinates was used as a discriminating condition, as shown in the figure 3. The specific calculation method is shown below.

Fig 3 The distance between the coordinates used to determine eye opening and closing

$$EAR = \frac{A + B}{2 * C}$$

If the EAR falls below 0.20 for 15 consecutive frames, the condition was made to judge that the eyes are consciously closed.

(3) Mouth opening and closing conditions

The switch to scrolling mode was implemented to be done when the mouth is consciously opened wide for about one second. Conditions were set to distinguish it from chatting. For opening and closing the mouth, the Mouth Aspect Ratio (MAR) is used as shown in the

figure. The specific calculation method is shown below.

Fig 4 The distance between the coordinates used to determine the opening and closing of the mouth

If the MAR was greater than 0.60 for 15 consecutive frames, it was judged that the mouth was deliberately opened in the shape of an "A".

For 15 consecutive frames, a MAR in the range of 0.10 to 0.20 was taken as a condition for judging the mouth to be consciously opened in the shape of a "yes".

3. System Overview

The proposed cursor manipulation method uses Dlib to determine the orientation of the face, determine the opening and closing of the eyes and mouth, move the cursor, click, and scroll the cursor based on the determination results. This allows physically handicapped users to freely operate the PC and freely browse the Internet without using their hands and arms. Also, when the face is facing the same direction for a certain period, the cursor movement's speed is increased in three levels to reduce the stress of operation. Furthermore, the operation is simplified by setting the click operation condition to close the eyes for more than 1 second, the addition of the scroll mode, and the deletion operation to open the mouth for more than 1 second. In the following, we describe the operation flow of the proposed system.

4. Evaluation

In this experiment, 6 subjects (5 males and 1 female) used this system to evaluate the usability of the proposed system by performing the 6 functions described in Chapter 3,

such as searching the Internet, accessing pages, etc., and then evaluating the functionality (whether the system can be operated without using both hands and whether it has sufficient functions), reliability (whether the system does not force-quit in the middle or malfunction), efficiency (whether the system responds immediately to operations), and usability of the 6 functions on a 5-point scale (1→very bad, 5→very good).

Table 2 Experimental results of the proposed system evaluation

	A	B	C	D	E	F	Average
Functionality	5	4	5	4	4	3	4.2
Reliability	5	3	4	5	5	2	3.8
Efficiency	4	2	3	2	2	3	2.8
Usability	Move the cursor	4	5	4	4	4	4.2
	Left Click	3	5	3	3	4	3.6
	Scroll Mode Switching	3	5	2	5	5	3.8
	Keyboard Display	4	5	3	5	2	3.8
	Page go, back	3	2	4	5	5	3.4

This result shows that efficiency only scored in the two-point range, lower than the others. The reason for this is that the parts of the face differ from person to person and some of them had difficulty in satisfying the set conditions. In order to improve the condition, we need to save the face part coordinates for each person and set the condition where the change is remarkably large.

We have observed that the subjects sometimes make mistakes in operating the system, such as clicking or displaying the keyboard, because their eyes narrow and their mouths widen to the side when they smile. In addition, subjects with glasses sometimes had difficulty in responding to eye movements. To solve these problems, it is necessary to add a condition for judging smiling to prevent incorrect operation and to consider the influence of reflection of glasses lenses.

5. Conclusion

This study solved three problems of previous studies on cursor control for people with disabilities: "Consideration of privacy," "Fixation of the distance between camera and face," and "Difficulty and high cost of installation equipment. However, in evaluating the system, not all the testee feel the system's operability is good, and it took a long time for the system to respond to the user's request. To solve these problems, it is better to set up the conditions corresponding

to the shape of the user's face and the system user's movement. Specifically, we consider that we should store the feature point coordinates and the distance between the coordinates of the user's face in every frame and determine the motion of the face when the difference between the two is significantly large. Although the motions of the faces in this study are all simple, we believe that more facial motions and cursor motions can be coordinated by increasing the number of features and calculating distances at various points.

References

- (1) White Paper on Disabled People 2019 (Entire Version) - Cabinet Office
https://www8.cao.go.jp/shougai/whitepaper/r01hakusho/zenbun/siryo_02.html
- (2) TOKYO UNIVERSITY OF TECHNOLOGY
KATAYANAGI ADVANCED RESEARCH LAB.
Computer-Based Communication Support
Project , Ida Akemi
<https://www.teu.ac.jp/karl/project/mediascience/009956.html>
- (3) Development of ICT equipment utilization evaluation method based on input analysis in special needs education , 2017, Kanamori Katsuhiko
<https://kaken.nii.ac.jp/ja/file/KAKENHI-PROJECT-15K04586/15K04586seika.pdf>
- (4) dlib C++ Library
<http://dlib.net/>
- (5) Get 68 landmarks using Dlib
<https://www.kkaneko.jp/dblab/dlib/landmark.html>
- (6) Find the distance between two points with NumPy
<https://tokibito.hatenablog.com/entry/20121222/1356187172>