A Study on Mouth Mouse

Chan-Myung Han[*], Joon-Ho Park[**], Hwi-Won Kim[***] Young-Woo Yoon[****]

[*] *Graduate School of Computer Engineering, YeungNam University*

[**] *Dept of Computer Science, Kyungwoon University*

[***] *Dept of Computer Science, KyungBuk College*

[****] *Dept. of Computer Engineering, YeungNam University, Korea*

Abstract

Among human body parts, the human face has been studied the most actively for the interface between humans and computers because face has statistic consistency in color, shape and texture. Those characteristics make computers detect and track human faces in images robustly and accurately. The human face consists of eyes, nose, mouth, eyebrows and other features. Detecting and tracking each feature have been researched. The open mouth is the largest in size and the easiest to detect among them. In this study, we present a system which can move mouse pointer using the position and state of the mouth.

1. Introduction

The Mouse has been the most effective tool for computers since Douglas Engelbart invented it in 1963. But in the situation where the keyboard and the mouse are used alternatively and frequently, it gives a lot of difficulty using the computer. People have been thinking over more natural ways of communicating with computers to solve this problem. Unfortunately, anything that can substitute the mouse hasn't been devised yet.

Many alternative mouse devices have been tried for physically challenged people in addition to effort of natural communication with computers. People without hands or people with upper-body paralysis can't use the normal computer with the keyboard and mouse. In this field, alternative mouse devices are being tried constructively.

In this study we deal with a system which can move mouse pointer with the web camera being used broadly. In chapter 2, previous works introduces alternative mouse devices which have been developed until now. In chapter 3, we explain the proposed mouth mouse system and in chapter 4, we discuss conclusions and future works.

2. Previous Works

In an attempt of replacing computer mouse, the trackball, joystick, touchpad, tablet and gyro mouse are invented but they are not a lot different from mouse except but in shape. They need people's hand for operation.

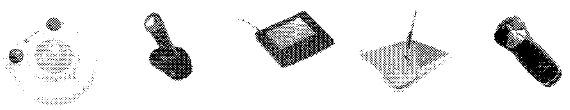

Figure 1. Alternative Mouse Devices

One of alternative mouse devices without requiring hands use electrical signals generated when the pupils move or muscles move[1]. These methods have one disadvantage. Electrodes should be attached to the skin and sweat could be a bother for them in the summer.

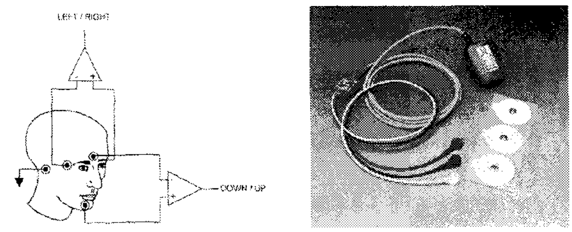

Figure 2. Adhesive Alternative Mouse Devices

To solve it, the web camera method is considered. The face image is captured by the web camera and

[****] Corresponding Author : Young-Woo Yoon

various parameters from it are calculated. Mouse pointer can be moved with these parameters[2]. One of drawbacks for this is that it requires a lot of calculation and it has low speed and low accuracy. For this reason, it didn't become popular.

3. Mouth Mouse Design

3.1. System Outline

The Mouth mouse system is a device which can move mouse pointer with parameters of the position and state of mouth obtained from the face images of the web camera. The hardware configuration of this system simply needs the computer with a web camera. The system move mouse pointer by analyzing the face images captured with the web camera.

Figure 3. Mouth Mouse System

3.2. Image Processing Method

3.2.1. Capture of Face Image

The first step of the system is to obtain video stream from the web camera. The bigger the size of input image is, the more accurate the system is. But 352x288 image is used because processing time get longer if the size becomes bigger. The proper size should be considered to make real-time processing possible.

3.2.2. Extraction of face region

The web camera captures background as well as the human face. The face region should be extracted from the whole image for the mouth is in the face. The face relatively has uniform color information and it is easier to find than the mouth. If the exact face region can be detected, it is very likely to find the mouth region. It helps faster detection of the mouth and reduces the possibility of false mouth detection in the background region.

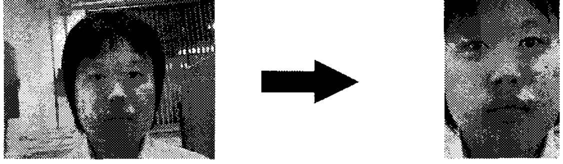

Figure 4. Extraction of face region

Extraction of the face region consists of three stages as figure 5.

Figure 5. Stages of extracting face region

3.2.2.1. Face Threshold Segmentation

Face threshold segmentation is to set pixel values to 1s if they are within the skin color range. Otherwise, pixels are set to 0s. To get skin color range, A-B color model is used to minimize the effect of lighting instead of RGB color model[3]. A-B color model can be described as the below.

$$r = \frac{R}{R+G+B} \qquad g = \frac{G}{R+G+B}$$

$$a = r + \frac{g}{2} \qquad b = \frac{\sqrt{3}}{2}g$$

Skin color range is different from race to race. Skin color range for every race is difficult to define. Many studies obtain the values from statistical distribution of many skin color images. In this study, we also calculated an average and standard deviation of value A and B statistically from experiment images and we define skin color range as the below.

Table 1. Skin Color Range

Skin Color Range
$ave\_a - \delta\_a \le a_{skin} \le ave\_a + \delta\_a$ $ave\_b - \delta\_b \le b_{skin} \le ave\_b + \delta\_b$ ave_a, ave_b : an average of skin color δ _a, δ _b : standard deviation of skin color

Figure 6 shows the result of face threshold segmentation. There are many non-skin color pixels in the face region because eyes, eyebrow, lips and shady parts are not within the skin color range. Nonetheless, it shows the obvious face region.

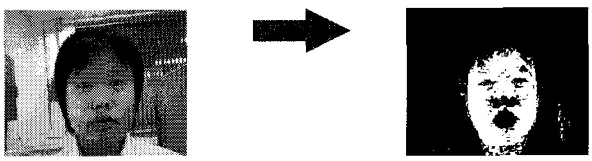

Figure 6. Face Threshold Segmentation

3.2.2.2. Merging

There are many pixels in the face region which are not classified as face although they belong to face. Especially, the eyes, nostrils, facial spots, lips and shady parts are not likely to be classified as the face region. In addition, the border of face has a combination of facial and non-facial pixels. We set up NxN blocks and categorize each block into face region if more than half pixels of one block are recognized as skin color pixels. Figure 7 shows the result of merging.

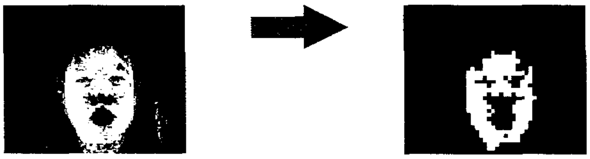

Figure 7. Merging

3.2.2.3. Extraction of Face region

The border of the face is neat after merging and that enables extraction of the face region. Non-facial block within the face region are definitely part of face. Thereby, the face region can be determined considering a rectangular region which includes all face blocks.

3.2.3. Threshold segmentation of open mouth

The open mouth is blackish because it generates a shady region inside the mouth. The open mouth is simply detected with this stark difference of color.

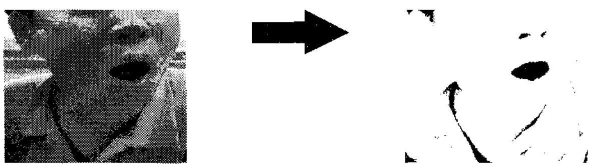

Figure 8. Open Mouth Thresholding

There are dark color pixels inside the open mouth and R, G and B have low values in RGB color model. But R is greater than G and B because it has reddish colors inside the mouth. G and B of pixels are mapped into 0 or 255 with specific threshold values. The face image is transformed as black and white image and transformed black and white image is thresholded with another specific threshold value. Then the thresholded image with black open mouth region can be produced as Figure 8. Threshold values mentioned above can be calculated experimentally.

3.2.4. Region labeling

Hair, pupils, nostrils, facial spots can be detected as black regions as well as open mouth. To group black pixels in the same region, labeling process is required. Labeled regions become candidates for the open mouth region and only these candidates are considered to determine whether they are real mouth region or not. Grassfire algorithm is used as a labeling method. Grassfire algorithm is named because it works as if fire were spreading through dry grass. It labels neighbor pixels around the current pixel until all the pixels are labeled checking neighbor pixels one after another.

3.2.5. Detection of Open Mouth

Dark labeled regions must be more than one. Among them, real open mouth should be found. The Open mouth is bigger than any other elements and it is a round shape. The real mouse is determined with these two criteria. Particularly, A mathematical projection function is used to discern whether a region is round or not.

3.2.5.1. Mathematical Model of Horizontal Projection

To evaluate whether a region is round, both horizontal and vertical projection functions can be used. A horizontal projection function is used for open mouth. To derive the horizontal projection function, we must calculate the distance between point P1 and point p2 which intersect a circle as the below figure.

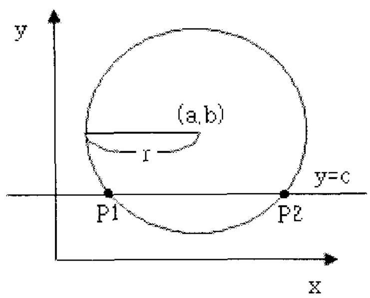

Figure9. Derivation of Horizontal Projection Function

We can get point P1 and P2 by circle equation where y = c as the blow.

$$(x-a)^2 + (y-b)^2 = r^2$$
$$(x-a)^2 + (c-b)^2 = r^2$$
$$x = \pm\sqrt{r^2 - (c-b)^2} + a$$

$$P1 = \left(a + \sqrt{r^2 - (c-b)^2}, c\right)$$
$$P2 = \left(a - \sqrt{r^2 - (c-b)^2}, c\right)$$

The distance between P1 and P2 can be derived as the below.

$$PF = 2\sqrt{r^2 - (c-b)^2} \quad (r - b \le c \le r + b)$$

3.2.5.2. Calculation of Error Rate

Error Rate for each candidate region should be evaluated to find real open mouth. To exclude hair region and noise regions, max area Value and min area value for regions should be set up and regions which are not within this range are excluded before error rate is evaluated. By doing this, precision and speed of processing can be improved. The first step for error rate calculation is to get the radius of the region because radius is needed in PF above. Real horizontal projection values for each candidate region can be evaluated in image. Mathematical values through projection function can also be obtained based on the radius of each region. The difference between real projection values and mathematical values is error.

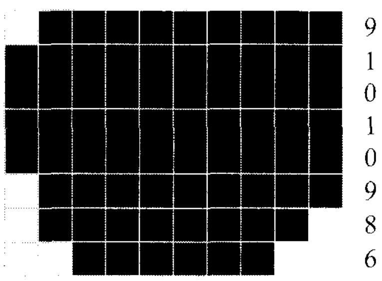

Figure10. Candidate Region for Open mouth

Table2. Calculation of Error Rate

c	PVI(c)	PVF(c)	PVI-PVF
1	9	9.1615	0.16515
	10	9.79795	0.20204
	10	10	0
	9	9.79795	0.79795
	8	9.16515	1.1651
	6	8	2
SUM		55.9262	4.332029
Error Rate			0.077

Error rate is a measurement for how much the candidate Region and the horizontal projection model is similar regardless of their size. The formula for error rate is presented as the blow.

$$\text{Error rate} = \frac{\sum_c |PVI(c) - PVF(c)|}{\sum_c PVI(c)}$$

$PVI(c)$: Projection Values from image

$PVF(c)$: Projection Values from Projection Function

3.2.5.3. Detection of Open Mouth

Error rates for each candidate region are calculated. The Region which has minimum error rate is selected as the real open mouth. After detecting the open mouth, the distance from center of image to open mouth and the state of mouth, whether mouth is open or not are collected on a real time basis. The distance can be utilized as orientation and speed of mouse and the state of mouth is substituted for mouse click information.

3.2.6. Mouse Operation

Mouse and keyboard can be operated with mouth information such as location and openness. There might be various kinds of methods and we don't know which one is better. But usually lattice method where divided rectangular areas have different mouse speed and orientation is useful.

4. Conclusions and Future-work

In this study, we present how to detect the open mouth from the face images captured by web camera and move mouse pointer with mouth information. This method is simple because web camera is only extra device to purchase. And it is very efficient because the open mouth is very distinctive and easy to find. It paved a way for interfacing method with computer especially for physically changed people. But it is very sensitive to illumination. In future, we have to improve this solution so that it can get robust to any illumination situation. In addition, efficient ways to mouse and keyboard operation should be more studied for more convenient and commercialized solution.

5. References

[1] Gregg Norris, MSEE and Eric Wilson, MSEE "The Eye Mouse, An Eye Communication Device." Bioengineering Conference, 1997. Proceedings of the IEEE 1997 23rd Northeast, 4644, 1997

[2] Jilin Tu, Thomas Huang, Hai Tao, "Face as Mouse Through Visual Face Tracking." Proceedings of the Second Canadian Conference on Computer and Robot Vision.

[3] Shinjiro Kawato, Jun Ohya, "Real-time Detection of Nodding and Head-shaking by Directly Detecting and Tracking the Between-Eyes." IEIC Technical Report, **VOL.**99;**NO.**448(PRMU99 101-132); **PAGE.**53-59(1999).