The Head-trace Mouse for Elderly: A Human-Computer Interaction System Based on Detection of Poses of Head and Mouth

You Song, Ye Zhang, Zeqiang Wang, Pengzhi Xie

Beihang University
No.37 Xueyuan Rd. Haidian District
Beijing, 100091, China

songyou@buaa.edu.cn

Abstract

This paper presents the design of a human-computer interaction system to operate a computer by the movements of head and mouth. First, the camera captures the user's head image, and then determines the user's head and mouth poses, including Head Left, Head Right, Head Up, Head Down, Mouth Open and Mouth Close, based on Adaboost detection algorithm. Different combinations of poses of the head and mouth are mapped to different mouse commands to operate a computer. Mouse commands consist of the sets of mouse move, click, drag, etc. This paper describes a map from the poses of head and mouth to mouse operations is designed through the principle of state machine. We also developed a commercial product, which provides a convenient way to operate computers for elderly and disabilities.

Keyword: Face detection, adaboost, state machine, human-computer interaction

I. Introduction

With the proceeding of computer science and technology, the usage of computer has brought about significant facilitation in every aspect of the society. However, the common computer input devices are usually designed for normal capable users, instead of elderly and disabled ones. The use of computers requires a mouse, a touchpad, a keyboard or other external devices. Users with upper limbs disabilities are incapable of controlling the mouse or keyboard easily, which makes it extremely difficult for them to use a computer. For common computer users, the long-term usage of conventional input devices causes chronic sore in hands, shoulders or neck, and greatly increases the risk of getting cervical or vertebral spondylosis.

In order to facilitate the disabled to use computers, extensive work has been carried out and two kinds of solutions are presented. The first solution is to use contact-type auxiliary equipment, e.g. infrared sensors and infrared reflector, to detect the user's movements to control a computer. Takami et al. invented a special kind of eyeglasses with three light-emitting diodes. By sitting in front of a computer with the eyeglasses, the user's image will be captured by camera, and the head movements are judged, so as to operate the computer[1]. Evans et al. used infrared light-emitting diodes and photodetectors as auxiliary equipment to determine the user's head position to operate a computer[2]. Chen et al. produced a mouse and a keyboard which can detect infrared signals. These devices use infrared light to achieve positioning and determining whether it is clicked or not, realizing operations of mouse and keyboard by head[3]. Judging eyes gazing direction by measuring the corneal reflex can also be used to operate the computer[4-6]. Gips et al. used EOG (electrooculographic potential) to detect eyes movements. They designed an EOG-based system that allowed people to control a mouse through moving eyes[6-8]. The results of such research have been applied in the lives of children[9]. The advantage of contact-type solution is that the detection of movement is accurate. However, the solution requires the user to wear special glasses, sensors or other equipments, which brings inconvenience to the user. The cost of these auxiliary devices is very high which greatly limits their wide usage.

The second solution is non-contact-type, which is to determine the user's action by collecting and analyzing the user's facial images with image processing techniques. Masoomeh and Alireza proposed an algorithm to transfer 2D camera motion information into 3D information to realize mouse operations[10]. Margrit et al. realized mouse operations by tracking the facial features and achieved certain effects[11]. Some other systems detected and analyzed the movement of eyes or mouth to operate the mouse[12-16]. Compared with the contact-type method, non-contact-type method is easy to use, free of any contacting devices, and with better experience of users. However, this method requires high-quality camera and image processing equipment. The software that suits this system is yet to be perfected until now.

The paper [17] proposed an algorithm to detect the movements of head and mouth and made a simple demo verification. In this paper, we presented a complete head-trace mouse system via webcam, using image processing technology to recognize the movements of head and mouth. Besides, we made a detailed introduction for the functions of the head-trace mouse.

II. SYSTEM OVERVIEW

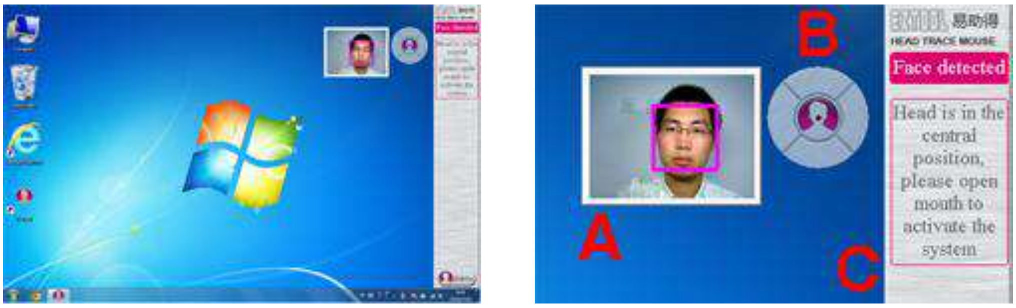
The head-trace mouse system is a human-computer interaction system based on the real-time detection of head and mouth poses. It collects user's head image (this paper refers to the positive region of the face) through camera, and then analyzes the user's head pose (standard, left, right, up, down) and mouth pose (open, close), after that, transfers the poses of head and mouth into the physical operations of a mouse using state machine theory.

The head-trace mouse system can use a camera either built-in or connected through USB interface. The system automatically drives the camera to capture and analyze images of the user in front of a computer, and then transfer the user's action into the shifting of mouse pointer on the screen. As shown in Figure 1, one of the authors is using the head-trace mouse system. The camera fixed on the computer is an ordinary CMOS camera.

To make it convenient for user to determine his own head position, the system displays the real time image captured on the top right of screen, as shown on Figure 2. A user can adjust the camera and his head to make sure the head position is in the middle of the captured images. Part A of Figure 2 (b) shows that the rectangles in the image mark the user's head and mouth detected by the system respectively. Part B, the circular icon, is a mouse pointer specially designed for the system (details in section IV). Part C shows the control panel of the system. When the system detects the user's face and mouth, the system can be activated by opening mouth.

This system realizes real-time detection of input images by recognizing head and mouse poses, with high accuracy and being exempt from any contacting-type devices. It offers significant convenience for the upper-limb disabilities and the seniors. The utilization of the system also notably alleviates the discomforts caused by long-term usage of computer. Furthermore, it can be used to neck exercise and rehabilitation training. It provides a whole new experience for computer users.

Figure 1. Using head-trace mouse



(a) Full-screen screenshot (b) Partial enlarged screenshot

Figure 2. Running screen of the system

III. BASIC HEAD POSES

The system uses Adaboost[18] algorithm to detect position and pose of the head from the input signal, i.e. each frame of captured video stream. The system analyzes information of the detected head and mouth movements, and then maps them into mouse events to operate the computer.

We define six basic head poses in the system, which are shown in Figure 3. These poses are defined as follows:

H0, Standard Head: Pose when face is front and still, mouth is closed.

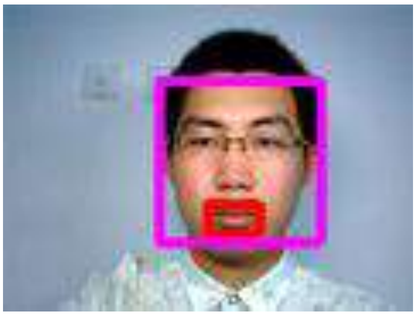
H1, Head Left: Pose when head turns left about 5 to 40 degrees relative to H0.

H2, Head Right: The same as H1, but in the opposite direction.

H3, Head Up: Pose when head turns up about 10 to 40 degrees relative to H0.

H4, Head Down: Pose when head turns down about 10 to 40 degrees relative to H0.

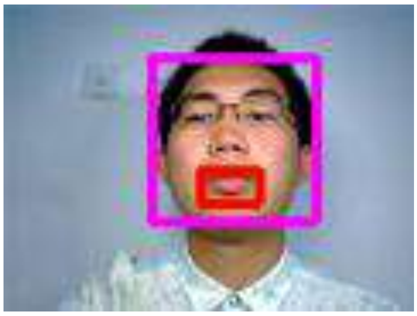
H5, Mouth Open: Pose when open mouth (open mouth as more than 50 percent as one can) from pose H0.



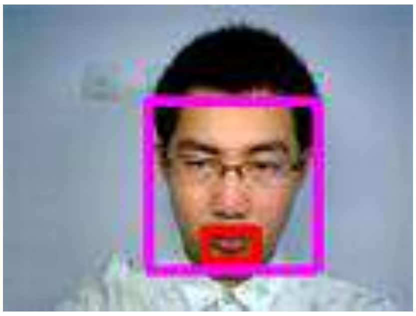
(a) Standard Head (H0)

(b) Head Left (H1)

(c) Head Right (H2)



(d) Head Up (H3)



(e) Head Down (H4)

(f) Mouth Open (H5)

Figure 3. Basic poses of head and mouth

Thus, the set of poses of head and mouth is defined as

$$\Sigma = \{H0, H1, H2, H3, H4, H5\} \quad (1)$$

Users are supposed to combine these basic poses to achieve all kinds of mouse operations.

IV. States and Functions of Head-trace Mouse System

The goal of this system is to operate mouse by inputs of head and mouth movements, which gives outputs of mouse events including still, move, click, drag, etc. For convenient and uniform description, if without special instructions, the mouse states and action are called mouse states for short.

In this system where traditional mouse was not used, we designed a special mouse cursor (Figure 2 (b)) to replace the traditional one in order to help users to operate the system easily. The mouse cursor is a graph including five parts (up, down, left, right and center) and a traditional cursor at upper left corner, which is called five-direction-graph cursor, or five-direction-graph for short, shown in Figure 4. In practice, mouse events are mapped into mouse cursor movements on the screen. Unless noted, the words “mouse” and “mouse cursor” are considered the same for convenient and uniform description. Figure 4 shows six mouse states. Due to space reasons, we haven’t listed all the graphs. Other unlisted mouse states are similar to the presented ones.

According to Figure 4, the mouse states and the running principles of the system are described as follows:

- M0, Mouse Idle: When the system is started and the face is detected, the system can be activated by opening mouth (H5), which goes to state M0. A still five-direction-graph will then be displayed on the screen, shown as Figure 4(a).
- M1, Mouse Moving Left: In the state of M0, if the movement of head left is detected, as shown in Figure 3(b), the five-direction-graph of moving left will be displayed on the screen, shown as Figure 4(b), and the cursor will move left horizontally. If the action of head left is remained, the cursor will keep moving left. When the Standard Head is restored, the cursor

will stop moving and return to the state of M0, as shown in Figure 4(a). Users can adjust moving speed of the cursor by setting corresponding parameters in the menu of the control panel.

(a) Mouse Idle (M0) (b) Mouse Moving Left (M1) (c) Mouse Moving Up (M3)

(d) Mouse Selection Mode 1 (M5) (e) Left Click Mouse (M6) (f) Mouse Selection Mode 2 (M9)

Figure 4. Five-direction-graphs show mouse states, movements or selection modes

- M2, Mouse Moving Right: The same principle as M1. The five-direction-graph is not shown in Figure 4.
- M3, Mouse Moving Up: The same principle as M1. The five-direction-graph is shown as Figure 4(c).
- M4, Mouse Moving Down: The same principle as M1. The five-direction-graph is omitted.
- M5, Mouse Selection Mode 1: In the state of M0, if H5 state (as shown in Figure 3(f)) is detected, the five-direction-graph will be changed to M5, as shown in Figure 4(d). Usually the movement of opening mouth will last about 1 second. Upon seeing the cursor of M5, a user can close mouth and keep in standard head, then the system maintains the state of M5. When the user opens mouth again, the cursor will go back to M0 state. M0 and M5 can be switched through mouth open and close movements.

- M6, Left Click Mouse: In the state of M5, if H1 (as shown in Figure 3(b)) is detected, the five-direction-graph of left click shown as Figure 4(e) will appear, indicating that Left-click command is executed in the computer system. After the command is executed, the user can return to standard head and the cursor will go back to M0.
- M7, Right Click Mouse: The same principle as M6, where the related figure is omitted. If Right-click command is executed, different operation menus will appear according to different applications. Then the Idle state is restored and user can continue to move head or open/close mouse to enter different states and execute different commands.
- M8, Double Click Mouse: The same principle as M6. The five-direction-graph is not shown in Figure 4.
- M9, Mouse Selection Mode 2: In the state of M5, if H4 (as shown in Figure 3(e)) is detected, the five-direction-graph will be changed to M9 (as shown in Figure 4(f)). User can maintain M9 by keeping in standard head.
- M10, Click And Drag Mouse: In the state of M9, if H3 state (as shown in Figure 3(d)) is detected, it is indicated that user executes Left-click and select command (if the window can be dragged, user can drag it or do nothing). User can drag the selected window through head movements. The mouse sub-states are described as follows: M10 (which means Click And Mouse Stay Still), M11 (Click And Drag Mouse Left), M12(Click And Drag Mouse Right), M13 (Click And Drag Mouse Up), M14 (Click And Drag Mouse Down).
- M15, Soft Keyboard: In the state of M9, if H1 (shown as Figure 3(b)) is detected, the system will pop a soft keyboard. If a user goes back to standard head, the cursor will return to M0. Then soft keyboard can be typed through head and mouth movements.
- M16, Reading Assistant (Browse): When web pages or documents are opened and the window is shown on the topmost, a user can scroll up and down by moving head up and down (the same as mouse wheel button). In the state of M9, if H2 is detected, the system will enter M16 state. The principle is same as M10, including the following mouse sub-states: M16

(Mouse Stay Still), M17 (Page Up), M18 (Page Down).

Through the above analyses, the mouse state set is defined as

$$Q = \{M0, M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15, M16, M17, M18\}. \quad (2)$$

It includes 19 kinds of mouse states, which can be grouped into four types:

- standard mouse states (M0, M1, M2, M3, M4);
- mouse select modes (M5, M9);
- mouse events (M6, M7, M8, M10, M11, M12, M13, M14, M16, M17, M18);
- Soft Keyboard (M15).

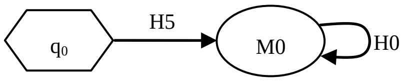
V. Mechanism of State Machine of the System

Mouse states transitions caused by head movements describe the processing logic of the head-trace mouse. The logic is represented through a state machine. This state machine is a 4-tuple defined as $M = (Q, q0, \Sigma, \delta)$, where:

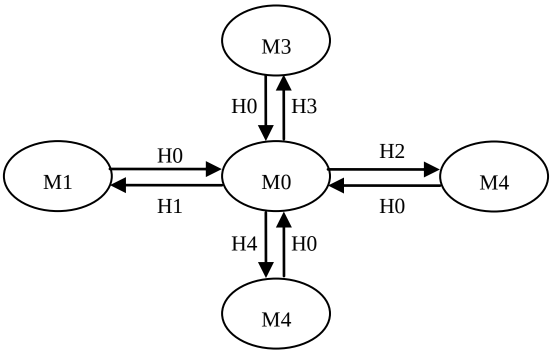
- Q is set of mouse states, defined as equation (2);
- q0 is the initial state (In this system, q0 is mouse still state where no human face is detected, or before the system is activated);
- Σ is set of poses of head and mouth, defined as equation (1);
- δ is a function from $Q \times \Sigma$ to Q, which is called M's transfer function.

Figure 5 shows the state transition diagram of state machine M of the head-trace mouse, where the unfilled oval is mouse state, the unfilled rectangle is mouse event, the shadow-filled oval is mouse select mode, and the shadow-filled rectangle is key represents soft keyboard.

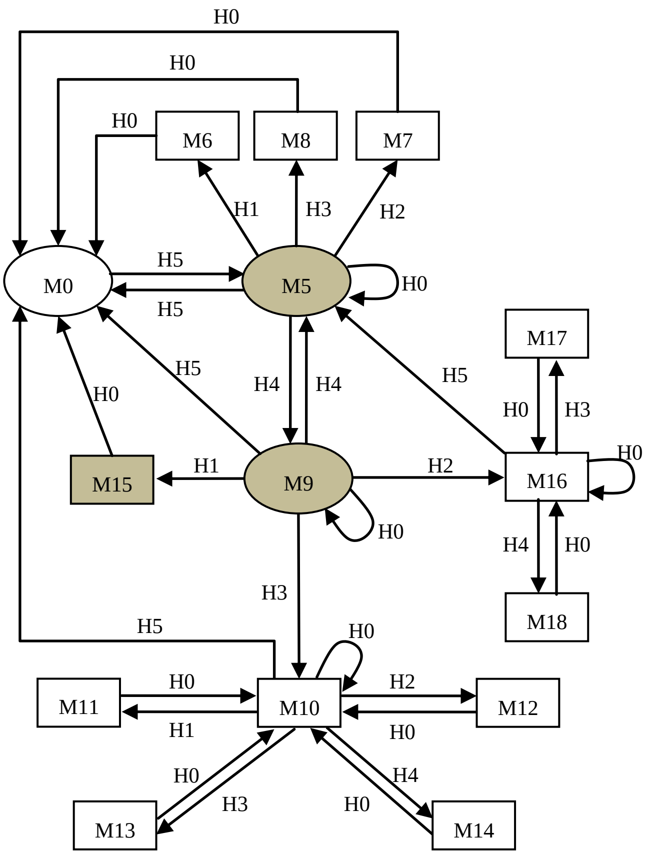
Figure 5 describes the system logic of the head-trace mouse. In Figure 5, the symbol next to the arrow is head movement, which is the input of system. The states connected by arrow illustrate the transition relations between the different states. This system suspends a control panel on the right of computer screen (including system's menu options). User can set system parameters or exit the system in the control panel. If abnormal conditions happen, the system will return to the initial situation to detect and relocate user's head again.



(a) The idle state of mouse



(b) The moving state of mouse



(c) The other operations of mouse's actions

Figure 5. State transition diagram of state machine M of the head-trace mouse

VI. Conclusion

This paper introduced the principles of a computer-human interaction system based on real-time state-detection of head and mouth. And the head-trace mouse system was designed and implemented. It was proved that this system was capable of performing the majority of an ordinary mouse's operations. With this system, users can operate computers by their head and mouth movements in front of web cameras. This system has been tested by an extensive number of persons and has been widely recognized. The commercial products of this system have been produced.

References

[1] M. Betke, J. Gips, P. Fleming, "The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities," Neural Systems and Rehabilitation Engineering, IEEE Transactions on , March 2002, vol.10, no.1, pp.1-10.

[2] R. Barea, L. Boquete, M. Mazo, E. Lopez, "System for assisted mobility using eye movements based on electrooculography," Neural Systems and Rehabilitation Engineering, IEEE Transactions on, Dec. 2002 , vol.10, no.4, pp.209-218.

[3] Y.L. Chen, "Application of tilt sensors in human-computer mouse interface for people with disabilities," Neural Systems and Rehabilitation Engineering, IEEE Transactions, Sept. 2001, vol.9, no.3, pp.289-294.

[4] Q. Ji, Z.W. Zhu, P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," Vehicular Technology, IEEE Transactions, July 2004, vol.53, no.4, pp.1052-1068.

[5] Kenji Yamagishi, J. Hori, M. Miyakawa, "Development of EOG-Based Communication System Controlled by Eight-Directional Eye Movements," Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, Aug. 30 2006-Sept. 3 2006, pp.2574-2577.

[6] C.H Morimoto, D Koons, A Amir, M Flickner, "Pupil detection and tracking using multiple light sources", Image and Vision Computing, 1 March 2000, vol.18, Issue 4, pp.331-335.

[7] S.W. Shih, J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," Systems,

Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions, Feb. 2004, vol.34, no.1, pp.234-245.

[8] A. Sears and M. Young, "Physical disabilities and computing technologies: an analysis of impairments". In The human-computer interaction handbook, Julie A. Jacko and Andrew Sears (Eds.). L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2002, pp. 482-503.

[9] P. DiMattia, F. X. Curran, and J. Gips, "An Eye Control Teaching Device for Students Without Language Expressive Capacity EagleEyes," Lampeter, U.K.: Edwin Mellen, 2001.

[10] M. Nabati, A. Behrad, "Camera Mouse Implementation Using 3D Head PoseEstimation by Monocular Video Camera and 2D to 3D Point and Line Correspondences," 2010 5th International Symposium on Telecommunications (IST'2010), 2010, pp.825-830.

[11] M. Betke, J. Gips, and P. Fleming, "The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People With Severe Disabilities," IEEE Trans. on Neural Systems And Rehabilitation Engineering, March 2002, vol. 10, no. 1, pp. 1-10.

[12] S. Stillittano, and A. Caplier, "Inner lip segmentation by combiningactive contours and parametric models," in Proc. of International Conference on Computer Vision Theory and Applications. 2008, pp. 297-304.

[13] K. E. Yi, and K. S. Kuk, "Eye tracking using neural network and meanshift," LNCS, 2006, vol. 3982, pp. 1200–1209.

[14] H. E. Cetingul, Y. Yemez, E. Erzin, and A. M. Tekalp, "Discriminative analysis of lip motion features for speaker identification and speechreading,"IEEE Trans Image Process, 2006, vol. 15, no. 10, pp. 2879-91.

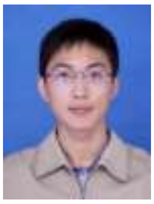
[15] J. J. Magee, M. R. Scott, B. N. Waber, and M. Betke, "EyeKeys: A realtime vision interface based on gaze detection from a Low-grade Video Camera," Computer Science Department, Boston University, 2004.

[16] B. Scassellati, "Eye Finding via Face Detection for a Foveated, Active Vision System," MIT Artificial Intelligence Lab Cambridge, MA, 02139, USA, 1998.

[17] Y. Song, Y.F. Luo, J. Lin, "Detection of Movements of Head and Mouth to Provide Computer Access for Disabled" in Proc. of Conference on Technologies and Applications of Artificial Intelligence, 2011.

[18] Viola, Jones, "Robust Real-time Object Detection," IJCV, 2001.

SONG You received his Ph.D. degree in 2003 from Beihang University. He is currently a professor in the Software School of Beihang University, Beijing, China. His research interests include Software Engineering, Signal and Information Processing, Artificial Intelligence and Human-Computer Interaction.



ZHANG Ye is currently pursuing his master's degree in the Software School of Beihang University. He received his B.Sc. degree in software engineering from Beihang University in 2012. His research interests include Computer Vision, Speech Recognition and Computer Graphics.

WANG Zeqiang is currently pursuing his master's degree in the Software School of Beihang University. He received his B.Sc. degree in software engineering from Beihang University in 2013. His research interests include head movement detection, speech recognition and Multi-sensory systems.

XIE Pengzhi is currently an M.Eng. candidate in the Software School of Beihang University. He received his B.Sc. degree in software engineering from Beihang University in 2013. His research interests include Machine Learning and Human-Computer Interaction.