

COMPUTER VISION BASED UNISTROKE KEYBOARD SYSTEM AND MOUSE FOR THE HANDICAPPED

M. Erkut ERDEM¹, İ. Aykut ERDEM¹, Volkan ATALAY¹, A. Enis CETIN²

¹Dept. of Computer Engineering, Middle East Technical University, Ankara, Turkey
{erkut, aykut, volkan}@ceng.metu.edu.tr

²Dept. of Electrical and Electronic Engineering, Bilkent University, Ankara, Turkey
E-mail: cetin@ee.bilkent.edu.tr

ABSTRACT

In this paper, a unistroke keyboard based on computer vision is described for the handicapped. The keyboard can be made of paper or fabric containing an image of a keyboard, which has an upside down U-shape. It can even be displayed on a computer screen. Each character is represented by a non-overlapping rectangular region on the keyboard image and the user enters a character by illuminating a character region with a laser pointer. The keyboard image is monitored by a camera and illuminated key locations are recognized. During the text entry process the user neither have to turn the laser light off nor raise the laser light from the keyboard. A disabled person who has difficulty using his/her hands may attach the laser pointer to an eyeglass and easily enter text by moving his/her head to point the laser beam on a character location. In addition, a mouse-like device can be developed based on the same principle. The user can move the cursor by moving the laser light on the computer screen which is monitored by a camera.

1. INTRODUCTION

In this paper, a unistroke keyboard and a mouse-like system based on computer vision is described for the handicapped. The system is developed for people with Quadriplegia, Cerebral Palsy, Multiple Sclerosis, Muscular Dystrophy, ALS, Carpal Tunnel Syndrome and any other disability where the user has little or no control of their hands to use a standard mouse or a keyboard. Speech recognition based systems may partially solve the text entry problem in some languages but cannot provide a solution for the mouse. In many agglunative languages including Turkish spoken by more than 200 million people there are no large vocabulary speech recognition systems.

Computer vision may provide alternative, flexible and versatile ways for humans to communicate with computers. In this approach the key idea is to monitor the actions of the user by a camera and interpret them in real-time. For example, character recognition techniques developed in document analysis [4,6,12] can be used to recognize hand-writing or sketching. In [4] we developed a vision based system for recognizing isolated characters drawn by a stylus or a laser pointer on a flat surface or the forearm of a person. The user's actions are captured by a head mounted camera. To achieve very high recognition rates characters are restricted to a single stroke alphabet as in Graffiti in [4]. The beam of the laser pointer is detected in each image of the video and characters are recognized from the trace of the laser beam.

The concept of computer vision based regular QWERTY type keyboards is independently proposed by us [13] and Zhang et.al.[7]. In this system, a character is entered to the computer if its location on the keyboard image is covered by a finger. In this approach the keyboard is a passive device. Therefore it can be made out of paper, fabric or foldable plastic having an image of a QWERTY or other type of regular size keyboard. It can even be displayed on a computer screen. The current version of this keyboard system cannot handle 10-finger typing.

Unistroke keyboards provide a good trade off between 10-finger typing and continuous handwriting recognition. In this paper, we present a computer vision based unistroke keyboard, which is based on a soft keyboard system called Cirrin (the CIRculaR INput device) developed in [14]. The Cirrin was designed for tablet computers and the user draws one stroke per word on a touch-sensitive screen. The key locations are placed circularly and to enter a word the user traces out a path determined by the characters forming the word using a stylus. Whenever the stylus enters a key location the corresponding character is

recognized by the tablet computer. In this paper, we place our key locations on a U-like curve or an upside down U-like curve or instead of a circle as shown in Figure 1. The advantage of this layout design is that the keyboard image can be displayed on the computer screen and whenever the beam of the laser pointer leaves the keyboard area it is treated as a mouse and the cursor is moved towards the direction of the beam. Similarly, the keyboard becomes active when the beam of the laser pointer enters the keyboard area.

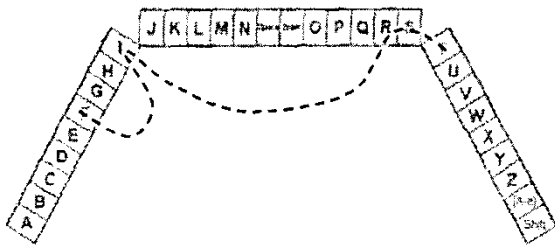


Figure 1. Up side down U shaped keyboard. The word “first” is entered in this example.

This keyboard system is designed for disabled people who have difficulty using their hands and ordinary keyboards and mice. The keyboard image is displayed on the lower part of the screen and a USB web camera is placed in front of the monitor to capture the movement of the laser beam as shown in Figure 2. As pointed out above the user can attach a laser pointer to an eyeglass and control the beam of the laser pointer by moving his or her head. In this way, an inexpensive system is realized for entering text to a computer because there is no need to have a special purpose hardware to realize this keyboard system which is implemented in software. Whereas in commercial head-mouse and keyboard systems, an optical sensor is used to track a small reflective target dot worn on one's forehead or eyeglasses [5]. This special purpose sensor increases the cost of the system. On the other hand the cost of an ordinary red laser pointer and a USB web camera is about 50 US dollars.

In the next section, the video processing algorithm to realize this keyboard and pointing system is described. In Section 3, experimental results and conclusions are presented.



Figure 2. The keyboard system developed for disabled people

2. VIDEO ANALYSIS FOR THE UNISTROKE KEYBOARD SYSTEM

Practical computer vision based human-machine interaction systems can be developed by taking advantage of the advances in computer technology. It is now feasible to process video acquired in real-time from the USB port of a standard PC without requiring any special purpose hardware.

Major functional parts of the image and video analysis algorithm include calibration, extraction of the background image model from the video, and estimation of the foreground image, which should ideally contain the beam of the laser pointer.

Calibration:

In this system, the camera is positioned so that the entire screen can be viewed. As it can be seen in Figure 1 and Figure 2, the keyboard consists of three rectangles, each divided into 10 key regions. When the system is turned on, the keyboard image is displayed on a grayish-white background about a second then it is displayed on a black background. By performing image frame differencing, the boundaries of rectangles containing the key regions are easily determined, because the largest differential change occur on the line segments forming the rectangles.

The next step is the estimation of the center of mass of each key region. This can also be easily carried out because the boundaries of the rectangles forming the keyboard are known.

Estimation of the Background Image:

During the calibration step the locations of the key regions and the keyboard region is determined. It is assumed that the computer monitor is neither moved by the user nor occluded by any object during the text entry process to the computer. Therefore the background image

of the video is stationary most of the time. The keyboard part of images consists of only the background image, and the beam of the laser pointer.

Estimation of the Foreground Image:

The foreground image ideally contains only the beam of the laser pointer in the keyboard area. The beam of the laser pointer is determined by detecting the moving pixels in the current image of the video and from the brightness information. The moving pixels are estimated by taking the image difference of two consecutive image frames. The beam of the laser pointer is brighter than its neighborhood. Therefore, the beam corresponds to a local maximum in the current image. This information is also used to make robust the detection process. By calculating the center of the mass of the bright red pixels among the moving pixels, the position of the beam of the laser pointer is determined.

A character is entered to the computer when the beam of the laser pointer enters the corresponding key region. First, it is checked if the beam is above the lower boundary of the rectangles forming the 1-D keyboard. Then a final decision is reached by calculating the distance of the beam to the center of mass of neighboring character regions. The beam has to stay in a character region at least in two image frames for recognition. In Figure 3, the character "K" is entered to the computer. This image is obtained by overlapping a sequence of images. Small white squares indicate the location of the beam in past images of the video.

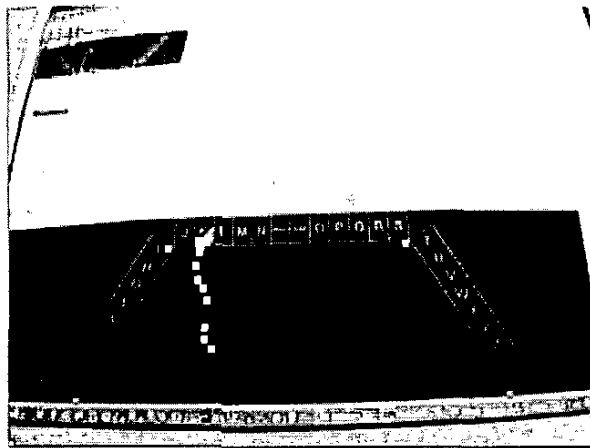


Figure 3. Character "K" is entered. Small white squares indicate the location of the beam in past images.

The vision based mouse system is calibrated in a similar manner. Calibration and registration of the

boundaries of the screen is very important in this case because computer screens are not flat in most cases. A reference calibration image similar to a chess-board is displayed for about a second and locations of reference points on the screen are registered.

The cursor is moved according to the position of the beam of the laser pointer on the current image of the video. Frame differencing and local maxima detection results are fused to determine the location of the beam. In order to smooth out the motion of the beam due to head jiggle a recursive position estimation strategy is implemented as follows

$$Y_k = Y_{k-1} + \alpha(Y_{k-1} - Y_{k-2}) + \beta(X_k - Y_{k-1})$$

where X_k is the actual position of the beam in the k-th image frame, Y_k is the position of the cursor, and the parameters, $\alpha + \beta = 1$.

Since a typical low cost USB camera produces 320 pixel by 240 pixel images the screen of the computer is set to a low resolution screen size e.g., 640 by 480 pixels, and a dark background is chosen to increase the detection accuracy of the beam.

3. EXPERIMENTAL RESULTS AND CONCLUSION

The experimental system setup is shown in Figure 2. An ordinary web camera is placed in front of the monitor and the user enters text to the system by controlling the beam of a red laser pointer. The USB camera produces 320 pixel by 240 pixel color images at about 13 frames per second. All of the processing of the vision based keyboard system is carried out in real time on a Pentium IV- 2Ghz processor with 1 GB memory. The keyboard system is developed using the Microsoft Vision SDK library. It is a free vision library for Microsoft Operating Systems developed by the Vision Technology Research Group in Microsoft Research.

This system has a higher recognition accuracy (96%) than the vision based Graffiti-handwriting recognition system [4]. The text entry speed is about the same in both systems: 10 words per minute (wpm). In order to estimate these values, the system is tested with a set of 20 words which were written at least 20 times.

The current layout of the keyboard is based on alphabetical ordering. With this layout structure, new users can start using the vision based keyboard system with little training. Although it is easy to learn, the current writing speed of 10 wpm is not very high. If this layout is modified then the writing speed can be increased. This can be accomplished by optimizing the keyboard layout by minimizing the median distance of pen travels between letters according to a word dataset. In this way one can reach up to 20 wpm writing speed in expert use after training.

In addition, the intended word can be completed before it is fully written by using a word dataset. This also further increases the writing speed.

4. REFERENCES

- [1] D. Hall, J. Martin, and J.L. Crowley, "Statistical Recognition of Parameter Trajectories for Hand Gestures and Face Expressions", *Computer Vision and Mobile Robotics Workshop*, Santorini, Greece, September 17-18, 1998.
- [2] I. Laptev and T. Lindeberg, "Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features", *Tech. Rep.*, KTH, Sweden, March 2000.
- [3] F. Quek et.al, "Gesture cues for conversational interaction in monocular video", *Proc. of Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Greece, Sept 1999.
- [4] O. F. Ozer, O. Özün, C. O. Tüzel, V. Atalay, A. E. Cetin, "Vision Based Single-Stroke Character Recognition for Wearable Computing", *IEEE Intelligent Systems*, Vol. 16, No. 3, pp. 33-37, May/June 2001
- [5] <http://www.keyalt.com/pointdevices/headmouse.htm>
- [6] O.N. Gerek, A.E. Cetin, A. Tewfik, and V. Atalay, "Subband Domain Coding of Binary Textual Images for Document Archiving", *IEEE Trans. on Image Processing*, Vol.8, No.10, pp.1438-1446, October 1999.
- [7] Ying Wu, Ying Shan, Zhengyou Zhang, Steven Shafer, "Visual Panel: Virtual mouse keyboard and 3-D controller with an ordinary piece of paper", Microsoft Research Tech. Report, 2000.
- [8] D. Goldberg and C. Richardson, "Touch-typing with a stylus", *Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems*, pp.80-87, N.Y, 1993.
- [9] I.S. MacKenzie and S. Zhang, "The immediate usability of Graffiti", *Proc. of Graphics Interface '97*, pp.129-137, 1997.
- [10] A Vardy, J A Robinson, L-T Cheng, "The wristcam as Input Device", *Proc. of the Third Int.Symp. on Wearable Computers*, California, Oct 1999, pp 199-202.
- [11] Starner, Thad, Weaver, Joshua, and Pentland, Alex. "A Wearable Computing Based American Sign Language Recognizer", *Proc. of the First Int. Symp. on Wearable Computers*, Cambridge, MA, IEEE Press, Oct. 13-14, 1997.
- [12] M.E. Munich and P. Perona, "Visual input for pen-based computers", *13th Int. Conf. Pattern Recognition*, pp.33-37, Vienna, 1996.
- [13] Y. Yardimci, A. E. Cetin, "Computer Vision Based Keyboard", Patent Application, Sept. 2000.
- [14] Jennifer Mankoff and Gregory D. Abowd. Cirrin: A word-level unistroke keyboard for pen input. In *Proceedings of UIST '98*. Technical note. pp.213-214.
- [15] Z. Zhang and Y. Shan, "Visual Screen: Transforming an Ordinary Screen into a Touch Screen," *IAPR Workshop on Machine Vision Applications (MVA 2000)*, pages 215-218, Tokyo, Japan, November 2000.