

# HUMAN COMPUTER INTERACTION USING COMPUTER VISION

*Yasemin YARDIMCI<sup>1</sup>, Volkan ATALAY<sup>1</sup>, A. Enis CETIN<sup>2,3</sup>*

<sup>1</sup>Dept. of Computer Engineering, Middle East Technical University, Ankara, Turkey

<sup>2</sup>Dept. of Electrical Engineering, Bilkent University, Ankara, Turkey

<sup>3</sup>Faculty of Engineering, Sabanci University, Istanbul, Turkey

*<http://www.ceng.metu.edu.tr/~vbi>*

## ABSTRACT

In this project we will develop computer vision based weightless keyboards and text entry devices for mobile computing and communication systems, and digital cameras and kiosks. In the keyboard system, the user imitates typing on a plastic or a cloth or even a paper with a keyboard image and his actions are captured by a camera. The characters covered by the fingers are recognized using computer vision techniques. We are also developing a companion system in which the user imitates writing on a surface using a pointing device. In this case the trace of the pointing device is analyzed and the characters are recognized. More information can be found at <http://www.ceng.metu.edu.tr/~vbi>

## 1. INTRODUCTION

There is a high interest for alternative flexible and versatile ways for humans to communicate with computers. In wearable computing flexible and versatile man-machine communication systems other than the ordinary tools of keyboard and mouse are necessary. Examples to the alternative communication systems include touch screens, hand gesture and face expression recognition systems, speech recognition systems, and various key systems [1-5]. Easy data entry to a wearable computer or a mobile communication device is an important problem. Mobile communication and computing devices currently have tiny keyboards with which data entry is difficult. A typical keyboard is too large to carry around. Contrary to the common perception speech recognition based text entry systems are not as efficient as a typical keyboard and they lack privacy [14]. Computer vision based man-machine communication systems can be developed by taking advantage of the character recognition systems developed in document analysis [4,6,7,12]. In addition, human-like capabilities such as perception would be a good feature of systems targeted for man-machine interaction, a specific

gesture or a sign of a hand can be used as a key to a database system.

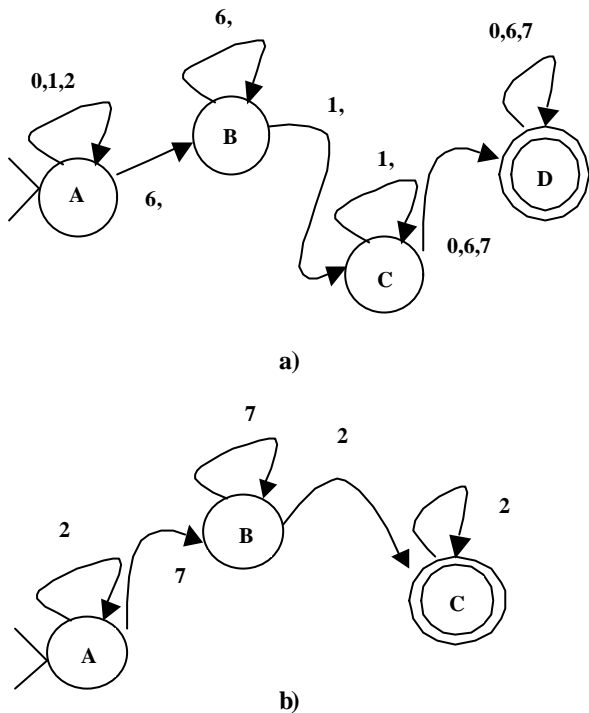
In this project we plan to develop two text entry systems using computer vision. In the computer vision based keyboard system, the user imitates typing on a plastic or a cloth or even a paper with a keyboard image and his actions are captured by a camera. The characters covered by the fingers are recognized using computer vision techniques. In the second system, the user imitates writing on a surface using a pointing device. In this case the trace of the pointing device is analyzed and the characters are recognized. We first describe this system in the next section and in Section 3 we present the vision based keyboard system.

## 2. VISION BASED CHARACTER RECOGNITION

In [4] we developed a system for recognizing isolated characters drawn by a pointer on a flat surface or the forearm of a person. The user's actions are captured by a head mounted camera for wearable computing. We assume that each character is drawn by a single stroke and in an isolated manner as in Graffiti in [4] to achieve very high recognition rates.

Unistroke isolated character recognition systems are successfully used in personal digital assistants in which people feel easier to write rather than type on a small size keyboard [8,9]. In this approach it is assumed that each character is drawn by a single stroke as an isolated character. One of the alphabets that has this property is the Graffiti™. The resulting character recognition system can be also used in mobile communication and computing devices such as mobile phones, laptop computers, handheld computers, and PDAs. The advantages of our computer vision based text entry system compared to other vision based systems [10-12] are the following:





**Figure 2.** Finite state machines for the characters a) “M” and b) “N”.

The use of the above algorithm is illustrated in Figures 1 to 4. Consider the laser beam traces of four characters shown in Figure 3. About 20 consecutive images are merged to obtain the “M” image shown in Figure 1.b and 3.d and the corresponding chain code representation is 32222207777111176666. The FSM for the character “M” is shown in Figure 2.a. When the above chain code is applied as an input to this machine, the first element which is 3 generates an error and the error counter is set to 1. The second element of the chain which is 2 which is a correct value at the starting state of the FSM so the error counter remains at 1 after processing the input 2. The FSM remains in the first state with the other 2s and also with the subsequent 0, as 0,1 and 2 are the inputs of the first state of the machine for M. The input 7 makes the FSM to go to the next state and the subsequent three 7’s let the machine to remain there. Whenever the input becomes 1, the FSM moves to the third state. The machine stays in this state until the single 7 input and this makes FSM go to the final state. The rest of the input data being 6 makes the machine to stay in the final state, and when the input is finished the FSM terminates. The error of the machine for character “M” is 1 for this input sequence. In fact, the above sample chain code is applied to other FSMs corresponding to all of the characters. But, the other machines generate either greater or infinite error values. This can be easily

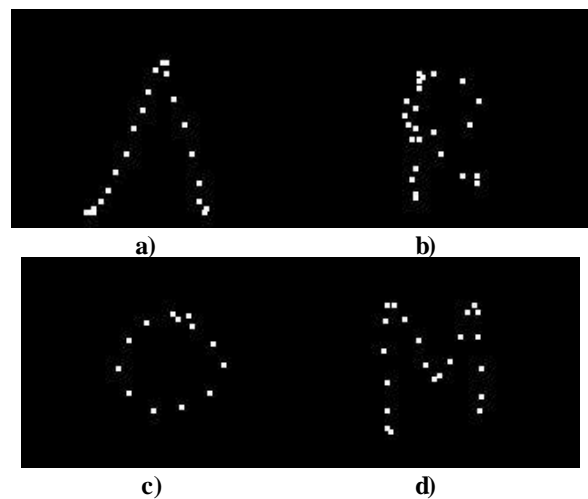
seen on the FSM for the character N which is shown in Figure 2.b. If the above string is given as input to this machine it will never reach to the final state and the error will be set to infinity.

Both the time and space complexity of the recognition algorithm are  $O(n)$ ,  $n$  being the number of elements in the chain code. In order to prevent noisy state changes, look-ahead tokens can be used which acts as a smoothing filter on the chain code.

It is observed that the FSM based recognition algorithm is robust as long as the user does not move his arm or the camera during the writing process of a letter. Characters can be also modeled by Hidden Markov Models which are stochastic FSM’s instead of the deterministic FSM’s to further increase the robustness of the system at the expense of higher computational cost.

## 2.2 VIDEO PROCESSING

The images corresponding to a character are to be processed to extract the marker positions for chain code extraction. If the position of the marker is found in the initial frame, it can be tracked in the consecutive images. In our experiments, we use a red laser pointer to write the characters. The



**Figure 3.** Laser beam traces generated by image sequences corresponding to a) lambda which corresponds to “A” in Graffiti b) R, c) O and d) M.

images are decomposed into red, green and blue components and the red mark can be found by thresholding followed by a connected component analysis

in the red image. If hand gestures are to be used, a skin filter may be necessary. Other pointers such as the tip of a pen can be also extracted and traced in a similar manner. Clearly, a laser pointer is the most robust text entry device to changing lighting and background conditions.

As discussed above, in an image sequence corresponding to a word, characters are separated from each other by discontinuous pointer movements. In the case of a laser pointer, at the end of each character the user turns off the light. This marks the end of each character. Segmentation of the video for each character is based on the jumps of the red mark of the laser pointer. While the user is writing a character, the transition of the pointer positions in consecutive images should be smooth, since only unistroke characters are allowed. The subsequent character will start at a relatively different position since the characters are to be written in an isolated manner. Therefore, a discontinuity is generated between two characters.

There are mainly two problems during the image capture and processing steps: distortion due to perspective projection and occlusion of the marker. Distortion in the characters occurs when the drawing or hand gestures are done in a non-orthographic manner. It is observed that such perspective distortion up to about 45 degrees of difference between the camera and the forearm does not affect the recognition as all of the captured and processed data are in two dimensions. However, if a more robust system is desired, a self-calibration algorithm [13] or video orbits algorithm [14] can be applied. However, these algorithms are too complex for current wearable computational power. Thus, a simplified version based on similar principles should be implemented. For example, the system can be calibrated by initially drawing two hypothetical lines on the forearm. Occlusion is not considered in this system, since the camera is assumed to capture the images in front of the marker.

### 2.3 EXPERIMENTAL RESULTS

The experimental setup is composed of a red laser pointer, a black background fabric and two web cameras one of which is an ordinary Philips PC Camera along with a capturing card, Tekram VideoCap C210 and the second one is a camera with a USB port. The web camera produces 160 pixel by 120 pixel color images at 13.3 frames per second. All of the processing is performed on an Intel Celeron 600 processor with 64MB of memory in real time.

The user draws a Graffiti character using the red pointer on the dark background material. In Graffiti like recognition systems, very high recognition rates are possible [9]. In

our system, in spite of the existence of perspective distortion, it is possible to attain a recognition rate of 97% at about 10 word per minute (wpm) writing speed. It is also observed that the recognition process is writer independent with little training.

In order to estimate the above recognition rate at least 50 samples from each character and a total of 1354 characters are used. An average of 18 image frames per character is required and this can be drawn less than 1.5 seconds which means that more than 40 characters per minute can be entered to the computer on the average. The writing speed can be further improved if the user trains himself or herself to write different characters e.g., the characters I and T can be drawn and recognized with almost 100% accuracy only with 34 frames. On the other hand, the character B needs at least 50 frames (or more than 3.35 seconds) for a reasonable recognition rate accuracy. The overall writing speed of our current system is slightly below the 13 wpm composition rate reported for Graffiti on a PDA. This is due to fact that the frame rate of a wearable camera is much smaller than the sampling rate of a touch screen on a PDA. We believe that we can achieve the same writing speed rates with the advances in digital camera and wearable computer technology.

The perspective distortion plays a minor role in the system since everything is in two dimensions. In our experiments, we have observed that the degradation in recognition is at most 10% around 45 degree difference between the plane on the which writing is performed and the camera.

Several tests are also carried out under different lighting conditions. In day (incandescent) [fluorescent] light the pixel value of the background is about 50 (180) [100] whereas the pixel value of the beam of the laser pointer is about 240 (250) [240]. In all cases the beam of the laser pointer can be easily identified from the dark background. If the user uses his or her finger to write than it is expected that the recognition rate of the current system will be significantly affected.

We have not yet implemented the system on a wearable computer, however the time and space complexity of the employed algorithms are low. The processor on which the experiments are done has similar performance compared to the processors mentioned in current wearable computers. Furthermore, the web camera considered during the experiments has very similar characteristics with the head mounted cameras used in wearable computers or the eyetab.

Although the frame rate of a wearable camera is much smaller than the sampling rate of a touch screen on a PDA, this is compensated by slow writing movements and our recognition algorithms which we believe are more complex and robust compared to the simple recognition algorithms used in PDA's.

The writing speed of our system is lower than the 35 to 40 wpm transcription speeds of septambic keyer developed by Mann [4] and Twiddler [5]. However, regardless of the keyboard the composition writing speed is below 20 wpm for most people. We believe that in a wearable computing environment the composition speed rather than the transcription speed is important. Furthermore, the 20 wpm writing speed with very high accuracy is even possible in our system (or in today's wearable computing technology) if an optimized unistroke alphabet [9] is used instead of Graffiti. In such a case the user has to learn a new alphabet consisting of very simple strokes. The reason that we use the Graffiti alphabet is its almost Latin alphabet like nature.

#### **2.4 FUTURE WORK**

As described above the current system recognizes the characters of a single stroke alphabets, and a laser pointer is used to write the characters. We plan to work on ways to improve the current system:

- (i) The current system understands the beginning and end points of the characters according to the status of the laser light. Whenever the user turns on (off) the laser pointer it means a new character starts (ends). To use an ordinary stylus to enter data into a computer pen up and pen down actions of the user must be recognized which will be determined by a finite state machine within a hidden Markov model framework.
- (ii) Continuous handwriting recognition : The advantage of the current system is that almost 100 % recognition accuracy is feasible as a result of the single stroke nature of the characters. In [12], a continuous handwriting recognition system is developed using computer vision. The main difference between our system [4] and [12] is that a regular pen is used to enter text into a computer in [12]. As pointed above a regular pen is not necessary in our approach and as a result it can be used as a text entry system in wearable computers, mobile communication devices, digital cameras and kiosks. We plan to develop a highly reliable continuous handwriting recognition system recognizing regular characters sketched by a finger or a pointer.

### **3. COMPUTER VISION BASED KEYBOARD**

We plan to develop a weightless keyboard system based on computer vision. In this system, the user imitates typing on a plastic or a cloth or even a paper with a keyboard image and his actions are captured by a camera. The characters covered by the fingers are recognized using computer vision techniques. In this way, computer data (such as messages, e-mails, and long internet addresses etc.) can be easily entered into the mobile device. The proposed keyboard is "wearable" in the sense that the user can fold the fabric keyboard and put it into his or her pocket or bag.

The proposed keyboard can be a full-size, weightless, and wearable keyboard that will eliminate the difficulty of writing messages or e-mails on the small keyboards that are currently available. Future cellular phones will have a built-in camera and video processing capability. Our vision based keyboard product can be implemented in software and incorporated into future cellular phones.

The vision based keyboard can be also used in notebook computers which may have both a removable regular keyboard and the vision based keyboard. The use of vision based keyboard may lead to thinner and lighter notebook computers.

Today, digital image and video cameras have the capability of being connected to a computer. Text entry to a digital camera could also be achieved using a vision based keyboard with. In this way digital cameras can be used as a fax, or a two-way pager or an Internet communicator.

Our current vision based system works, if a user uses only one finger. The user covers the key 'E' with his finger to enter this character as shown in Figure 4. The state of each key is also shown in the same Figure. As a result the character 'E' is recognized.

#### **3.1 FUTURE WORK**

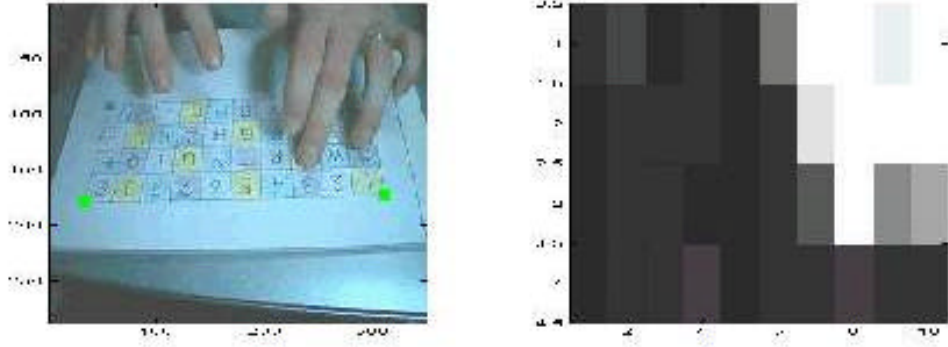
As pointed above the current vision based keyboard system has the ability to track one finger. We plan to handle ten finger typing using a Kalman filter based tracker. The main difficulty to extend the current system to ten finger typing is the occlusion of fingers and keys (in a single finger typing there is almost no occlusion). To overcome the occlusion problem the finger tips will first be determined and the action of each finger tip will be tracked parallelly using Kalman predictors [15]. To determine if a key has been pressed or not a finite state machine will be assigned to each finger. The state of each finger such as

raising, pressing, moving forward, stationary etc. will be estimated within a Hidden Markov Model framework.

In addition we are working on implementing a computer vision based mouse [16] as a companion system to our text entry systems.

## 5. 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", *Technical report CVAP245, ISRN KTH NA/P-00/12—SE*, Department of Numerical Analysis and Computer Science, KTH, Sweden, March 2000.
- [3] F. Quek, D.J. McNeill, R. Ansari, X. Ma, R. Bryll, S. Duncan, K.E. McCullough, C. Kirbas, "Gesture cues for conversational interaction in monocular video", *Proceedings of Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'99)*, Corfu, Greece, September 1999.
- [4] Ömer Faruk Özer, Oguz Özün, C. Öncel Tüzel, Volkan Atalay, A. Enis Çetin, "Vision-Based Single-Stroke Character Recognition for Wearable Computing", *IEEE Intelligent Systems*, Vol. 16, No. 3, pp. 33-37, May/June 2001
- [5] [www.handykey.com](http://www.handykey.com)
- [6] O.N. Gerek, A.E. Cetin, A. Tewfik, and V. Atalay, "Subband Domain Coding of Binary Textual Images for Document Archiving", *IEEE Transactions on Image Processing*, Vol.8, No.10, pp.1438-1446, October 1999.
- [7] E. Oztop, A.Y. Mulayim, V. Atalay, and F. Yarman-Vural, "Repulsive Attractive Network for Baseline Extraction on Document Images", *Signal Processing*, Vol.75, No.1, pp.1-10, 1999.
- [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, New York, 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", *Proceedings of the Third International Symposium on Wearable Computers*, San Francisco, 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 International Symposium on Wearable Computers*, Cambridge, MA, IEEE Computer Society Press, Oct. 13-14, 1997.
- [12] M.E. Munich and P. Perona, "Visual input for pen-based computers", *13<sup>th</sup> Int. Conf. Pattern Recognition*, pp.33-37, Vienna, 1996.
- [13] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.
- [14] Leshner, G.W., Moulton, B.J., & Higginbotham, D.J. (1998). Optimal character arrangements for ambiguous keyboards. *IEEE Transactions on Rehabilitation Engineering*, 6, 415-423.
- [15] A.Murat Bagci, Ismail Yilmaz, Enis Cetin, Mubeccel Demirekler Moving Object Detection and Tracking in Video Based on Higher Order Statistics and Kalman Filtering, IEEE NSIP Conference, Baltimore, 2001.
- [16] A. Erdem, E. Erdem, Yasemin Yardimci, Volkan Atalay, A. Enis Çetin, "Computer Vision Based Mouse", accepted for presentation in IEEE ICASSP 2002, Student Forum Session.



**Figure 4:** Computer Vision Based Keyboard Concept: The key 'E' is pressed (i.e., the region corresponding to the character 'E' is covered by a finger). The output of our system is shown on the right: The character 'E' is recognized. In a single finger typing system the covered key closest to the camera is assumed to be pressed.

## Authors

**A. Enis Cetin**, Project director, (Ph.D. University of Pennsylvania), Professor of Electrical and Electronics Engineering at Bilkent University, Ankara, Turkey, Author of more than 100 scientific papers, Associate Editor of the journal *IEEE Transactions on Image Processing*.  
<http://www.ee.bilkent.edu.tr/~cetin>

**Volkan Atalay** (Ph.D. University of Paris), Associate Professor of Computer Engineering at Middle East Technical University

**Yasemin Yardimci** (Ph.D. Vanderbilt University), Associate Professor, Informatics Institute, Middle East Technical University.