VISION BASED SINGLE STROKE CHARACTER RECOGNITION FOR WEARABLE COMPUTING

Oğuz ÖZÜN[1] , Ö. Faruk ÖZER[2], C. Öncel TÜZEL[1], Volkan ATALAY[1], A. Enis ÇETİN[2,3]
[1]Dept. of Computer Engineering, Middle East Technical University, Ankara, Turkey
[2]Dept. of Electrical Engineering, Bilkent University, Ankara, Turkey
[3]Faculty of Engineering, Sabanci University, Istanbul, Turkey

ABSTRACT

We describe a method for recognizing the regular characters drawn by hand gestures or by a pointer on the forearm of the user 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. Recognition is performed by a bank of finite state machines whose input is the chain code of the hand drawn character.

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 key systems [1-5]. Easy data entry to a wearable computer is a field that requires much attention. One handed chording keyboards such as septambic keyer developed by Mann [4] and Twiddler [5] are interesting new approaches to enter data to wearable computers. Computer vision based man-machine communication systems can be developed by taking advantage of the character recognition systems developed in document analysis [6,7,12]. For example, 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 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.

The purpose of this study is to develop a method for recognizing the characters drawn by hand gestures or by a pointer on the forearm of the user captured by a digital camera which is a new form of data entry into a wearable computer. In this method 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:

- The background is controlled by the forearm of the user. Furthermore, if the user wears a unicolor fabric then the tip of the finger or the beam of the pointer can be detected in each image of the video by a simple image processing operation such as thresholding.
- It is very easy to learn a Graffiti-like alphabet. Only a few characters are different from the regular Latin alphabet. Although it may be easy to learn other text entry systems such as [4] and [5], some people are reluctant to spend a few hours to learn unconventional text entry systems. Furthermore, in addition to the regular characters other single stroke characters can be defined by the user to be used as bookmarks, pointers to databases etc.
- Computationally efficient, low power consuming algorithms exist for the recognition of unistroke characters and they can be implemented in real time with very high recognition accuracy. After a few minutes of studying the Graffiti alphabet, about 86% accuracy is possible. After some practice, accuracy improves to about 97%. Almost 100% accuracy seems to be possible [9].
- Computer vision based text entry systems are almost weightless.

In Section 2 the outline of the recognition system is described. Character recognition is performed by a finite state machine whose input is a chain code. In Section 3,

the FSM structure is describe and in Section 4 simulation examples are presented.

2. RECOGNITION SYSTEM

Consider the situation in which the user draws unistroke, isolated characters by a laser pointer on the forearm captured by a camera mounted on the forehead. Video is first segmented to image sequences corresponding to each character. The image sequence starts with a pointer turn-on action and terminates when the user turns off the marker. In other words, characters are separated from each other by discontinuous marker movements.

In our system unistroke characters are described using a chain code which is a sequence of numbers between 0 and 7 obtained from the quantized angle of the beam of the laser pointer in an equally timed manner. The chain code is extracted from the relative motion of the beam of the laser pointer between consequitive images of the video and it is applied as an input to the recognition system which consists of finite state machines (FSM) corresponding to individual characters. The FSM generating the minimum error indicates the recognized character. However, certain characters such as Q and G may be confused in a feature set consisting of only the chain code. Therefore, in addition, the begining and the end points of strokes are also considered. The weighted sum of the error from a finite state machine and the begining and the end point error determines the final error for a character in the recognition process. The algorithm for the recognition of characters consists of four steps which can be summarized as follows.

Step1. Extraction of chain code.

- The position of the red mark produced by the laser pointer is found in each frame.
- A chain code is generated according to the angle between two consecutive mark positions.
- Beginning and end point coordinates together with the coordinate of a rectangle enclosing the character are determined.

Step 2. Analysis using finite state machines.

- The chain code is applied as input to each state machine.
- State changes are determined in addition an error counter is increased by one if a change is not possible according to the current FSM
- If a chain code does not terminate in the final state, the corresponding character is eliminated.
- Errors in each state are added up to find the final error for each character.

Step 3. Errors due to the beginning and end points.

- Beginning and end points of a stroke are normalized with respect to the enclosing rectangle.
- If the width or the height is larger than a given threshold, it is not considered as a feature.
- An error value is calculated from the comparison of the normalized beginning and end points of the input character and the candidate character stroke.

Step 4. Character determination.

- State machine error and position error are weighted and added up.
- Character with minimum error is the recognized one.

The second feature set used in character recognition is based on position errors due to the beginning and end points of each stroke.

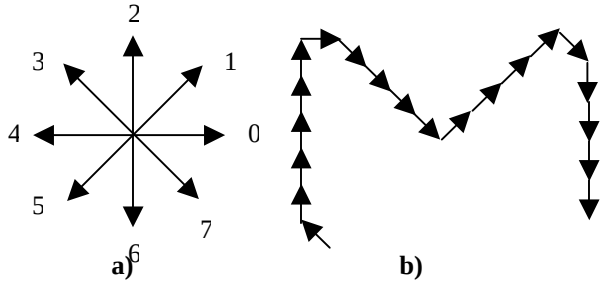

Figure 1. a) Chain code values for the angles, b) a sample chain coded representation of the character "M"=32222207777111176666.

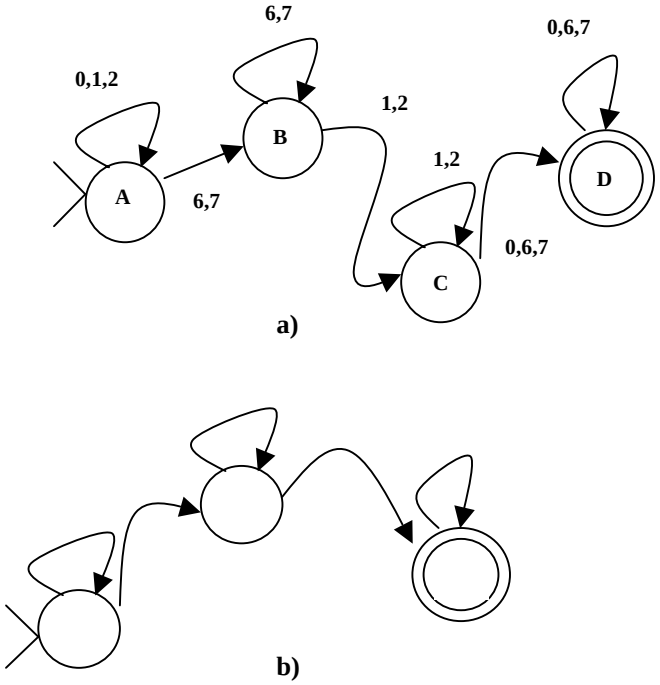

Figure 2. Finite state machines for the characters a) "M" and b) "N".

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.

3. 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

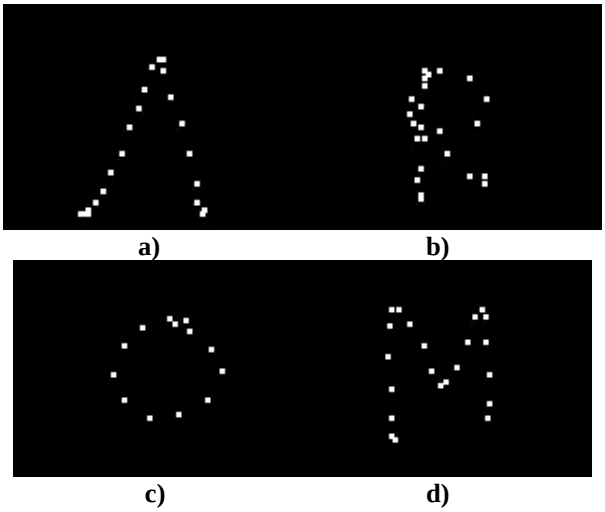

Figure 3. Laser beam traces generated by image sequences corresponding to a) lambda which corresponds to "A" in Graffiti b) R, c) O and d) M.

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 defined by the laser pointer (or regular pointer), the camera and the tangent plane of the forearm does not affect the recognition. The reason that the system fails after 45 degrees is that the chaincode used in the

representation of the characters has a quantization level of 45 degrees. In other words, the unit circle is represented by 8 directions. This problem can be overcome by either increasing the quantization levels and modifying the FSM models accordingly, or by using projective geometry methods developed by Mann [13,14,15] which can provide an efficient solution with the help of a feedback coming from a viewfinder. Occlusion is not considered in this system, since the camera is assumed to capture the images in front of the marker.

4. EXPERIMENTAL RESULTS AND CONCLUSION

The experimental setup is composed of a red laser pointer, a black background fabric and a web camera which is an ordinary Philips PC Camera along with a capturing card, Tekram VideoCap C210. 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.

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 words 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 3-4 frames. On the other hand, the character B needs at lest 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) [fluescerant] 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.

A major application of our system is that it can be used to take notes while watching a presentation. This is only possible, if the system has a viewfinder [16,17,18]. In this way, hand-eye-camera coordination can be carried out in which viewfinder provides the feedback loop so that pointer written characters remain always in the viewing area of the camera and the misrecognized characters can be immediately corrected.

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.

5. ACKNOWLEDGEMENT

We thank to the special issue editor and the anonymous reviewers whose comments and corrections significantly improved the quality of the paper.

6. 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] http://wearcam.org/septambic

[5] 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”, *13th Int. Conf. Pattern Recognition*, pp.33-37, Vienna, 1996.

[13] S. Mann, “Further developments on 'HeadCam': joint estimation of camerarotation+gain group of transformations for wearable bi-foveated cameras”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Volume: 4, pp.2909 –2912, 1997.

[14] S. Mann and R.W.Picard, “Video orbits of the projective group: a simple approach to featureless estimation of parameters”, IEEE Trans. Image Processing, Vol.6, pp. 1281-1295, 1997.

[15] S. Mann, “Humanistic computing: "WearComp" as a new framework and application for intelligent signal processing”, Proceedings of the IEEE, Vol.86, pp.2123 –2151, 1998.

[16] S. Mann, “Smart clothing: the wearable computer and wearcam”, Personal Technologies, Vol.1, 1997.

[17] S. Mann, “'WearCam' (The wearable camera): personal imaging systems for long-term use in wearable tetherless computer-mediated reality and personal photo/videographic memory prosthesis”, Second International Symposium on Wearable Computers, pp.124 -131, 1998.

[18] S. Mann, “Telepointer: Hands-free completely self-contained wearable visual augmented reality without headwear and without any infrastructural reliance”, The Fourth International Symposium on Wearable Computers, pp.177-178, 2000.