Alphabet Recognition in Air Writing Using Depth Information

Robiul Islam, Hasan Mahmud, Md. Kamrul Hasan
Systems and Software Lab (SSL)
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Dhaka, Bangladesh
e-mail: r_islam@live.com, hasan@iut-dhaka.edu, hasank@iut-dhaka.edu

Husne Ara Rubaiyeat
Department of Computer Science
Natural Science Faculty
National University
Bangladesh
e-mail: rubaiyeat@yahoo.com

***Abstract*—We present a data mining approach to recognize air written English Capital Alphabets (ECAs) using depth information. The hand motion while writing the alphabet in the air was captured as depth images by a depth camera. The depth images were then processed as the hand movement time series data, which was matched with standard templates of air writing by Dynamic Time Warping (DTW) algorithm. We collected a dataset of five variations and from them standardized 20 templates for each ECA. The accuracy for ECA is 93-99% with an average of 96.3%.**

Keywords-Air Writing; Gesture Recognition; Depth Information; Time Series; Dynamic Time Warping

I. INTRODUCTION

Human gesture is an important input modality for communication with computers. In hand-based gesture recognition technology, a camera (typical stereo camera) reads the hand movement data, performs the hand tracking and then recognizes a meaningful gesture to control any device or application. For example, a person clapping his hands together in front of a camera can produce the sound of cymbals being crashed together when the gesture is fed through a computer. In hand gesture recognition research, air writing is a prominent and difficult topic. Air writing [2] means gesture based writing in the air through movement of hand fingers by which a computer system can recognize characters and other symbols in natural handwriting. Previously, air writing has been studied by using wearable hand gloves [2], an accelerometer and gyroscope augmented mobile phone [4] and a 3D gesture input device [3]. In our approach, we do not ask for any special wearable device, so that the user can write naturally without any obstruction, shown in Figure 1. Algorithmically, previous approaches studied air writing by converting them into strokes. While writing in the air, users pause unintentionally or bend abnormally and, thus, they create extra strokes into the air written characters. However, recent data mining algorithms enable us to study a gesture signal such as air written character as a time series, as shown in Figure 2, information which can be matched with standard time series character templates. In this paper, we have studied the ECAs as time series information that can be recognized by matching with templates by using a time series data mining algorithm, DTW [1][6].

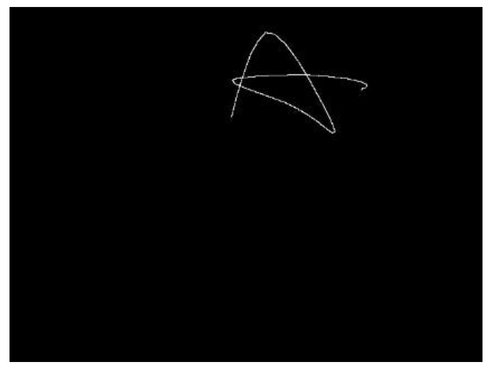


Figure 1: Air writing Character A

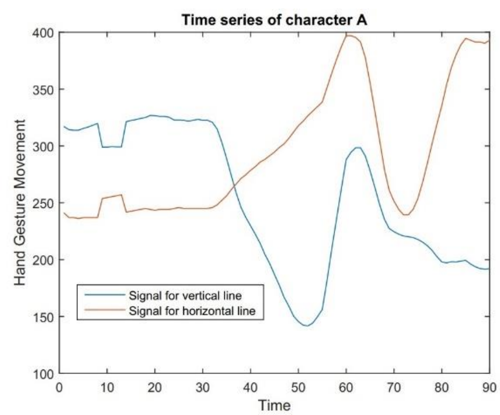


Figure 2 : Time series representation of character A

The rest of the paper has organized as follows: Section II contains the discussion on related work, Section III explains the proposed methodology, Section IV presents the result analysis, and Section V describes the conclusion and future work.

II. RELATED WORK

Agrawal et al. [4] used mobile sensors to recognize ECAs. They used gyroscope bend angles to convert the accelerometer signals into strokes. The concept of detecting strokes in air writing was first introduced in this paper. Amma et al. [2] showed how a wearable device can recognize hand gesture for air writing. A wearable hand motion tracking system captures movement signals using accelerometer and gyroscope. However converting the acceleration signal into distance covered for recognized strokes can be erroneous. Moreover, wearing a special device makes the air writing system cumbersome. Kim et al. [3] showed a way to recognize different people's handwriting on continuous images based on similarity of the different shapes of characters or digits. We studied their research and got inspired to use shape information. Lin et al. [5] showed a way to use Scale Invariance Feature Transform (SIFT) on binary images. In this approach, the motion information is not considered in SIFT descriptors.

We studied the process of generating strokes on ECAs extensively. We found that extra strokes are generated because users put some unnecessary pauses or bend their hands abnormally while writing alphabets. Those extra strokes cannot be merged together to form a regular stroke which demands complex algorithms like HMM [2] or Bayesian networks [3]. So, instead of processing strokes, we prefer to use whole ECAs written in the air as time series signals. Another motivation of using time series signals was to get the hand movement sequence of writing which are important clues to recognize air written ECAs.

III. Proposed System

Our system captures the air writing of a user using Microsoft Kinect depth camera. Each writing of an ECA generates a set of depth images. We segment the hand finger movement using depth values. The hand motion is tracked from image to image, which generates a series of points (x, y). These set of points are actually the time series information of the particular ECA. As the hand movement is noisy, the time series data is smoothed using moving average filter [7]. We did it for all 26 English capital alphabets. For example, we are showing the process of writing "A" in Figure 3.

A. Applying DTW

DTW is time series matching algorithm for which we need to match an ECA time series with standard templates of ECAs. The matching generates a distance score; the smaller the score, the better the matching. So, time series of an unknown ECA will be recognized as the best matching ECA. The calculation of the minimum distance was done using equation 1.

$$RecognizedECA = argminDTW\ (U, TECA) \quad (1)$$

Where U is a time series of unknown ECA. We define the time series U as two dimensional signals x and y; where x is the hand movement along x-axis and y is the hand movement along y-axis. TECA is the template of a particular ECA.

IV. Result Analysis

We have collected data from 5 different variances. Each user was asked to produce ECA gestures in the air five times. From those instances on, we considered as DTW template. So, for each ECA 5 templates were taken from 5 users and the rest of 20 were used to test. The dataset is available in [8]. To measure the performance. we calculated the accuracy and found maximum 100% true positive with average of 52.6% and minimum 93.75% true negative with average of 97.8% shown in Table I for our five variations. In our process, we are depending on true negative rate instance of true positive rate. We applied DTW of signal time series samples to all variation of alphabets and found best matching results for each dataset. Ideally, it gave 20 best matching and 500 are correctly rejected matching. Table I shows the total accuracy, which is calculated using (2).

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Negative}{Total\ population} \quad (2)$$

TABLE I. Result Table

Alphabets	True positive rate	True Negative rate	Accuracy
A	30	99.4	96.73077
B	25	93.75	96.92308
C	25	98	95.19231
D	30	98	95.38462
E	15	100	96.73077
F	55	99.6	97.88462
G	25	99.4	96.53846
H	80	98.4	97.69231
I	85	95.6	95.19231
J	75	98.8	97.88462
K	35	99	96.53846
L	100	97	97.11538
M	55	96.2	94.61538
N	95	97.8	97.69231
O	45	98.6	96.53846
P	70	94.4	93.46154
Q	40	96.6	94.42308
R	60	96.8	95.38462
S	100	99.8	99.80769
T	50	96.8	95
U	65	99	97.69231
V	75	95.8	95
W	60	99.4	97.88462
X	5	98.8	95.19231
Y	15	98.6	95.38462
Z	75	100	99.03846
AVG	52.6	97.822	96.31538

V. CONCLUSION AND FUTURE WORK

We proposed a DTW based time series matching approach to recognize air written English Capital Alphabets. Stroke detection was one of the main challenges in state-of-the-art air writing recognition algorithms, but we have converted the whole image into time series representation to detect in air gestures as ECAs. In the future, we want to collect more datasets for better results.

ACKNOWLEDGMENT

This work is partially supported by ICT Division, Ministry of Posts, Telecommunications and IT, Government of the People’s Republic of Bangladesh.

REFERENCES

[1] P. Doliotis, A. Stefan, C. Mcmurrough, D. Eckhard And V. Athitsos, “Comparing Gesture Recognition Accuracy Using Color And Depth Information,” In Proceedings Of The 4th International Conference On Pervasive Technologies Related To Assistive Environments , New York, 2011.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] C. Amma, M. Georgi and T. Schultz, “Air writing: A wear-able handwriting recognition system,” Personal and Ubiquitous Computing Volume 18, Issue 1, pp 191-203, January-2014, Springer London.

[3] D. H. Kim, H. Il Choi, J. H. Kim, “3D space handwriting recognition with ligature model,” Proceedings of the Third international conference on Ubiquitous Computing Systems, October 11-13, 2006, Seoul, Korea.

[4] Sandip et al., “Using mobile phones to write in air,” Proceedings of the 9th international conference on Mobile systems, applications, and services, June 28-July 01, 2011, Bethesda, Maryland, USA

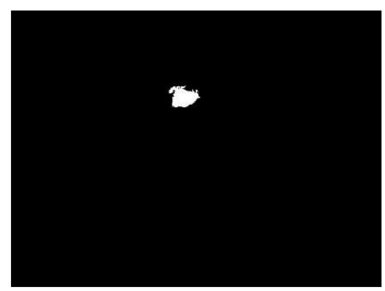
[5] W. Lin, Y. Wu,W. Hung and C. Tang, “A Study of Real-Time Hand Gesture Recognition Using SIFT on Binary Images,” Advances in Intelligent Systems & Applications , SIST 21, pp. 235246.Springer-Verlag Berlin Heidelberg 2013.

[6] S. Salvador and P. Chan, “FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space,” KDD Workshop on Mining Temporal and Sequential Data, pp. 70-80, 2004.

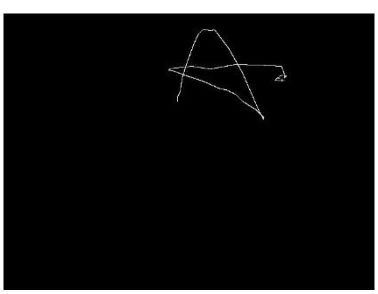
[7] J. F. Kenney and, E. S. Keeping, “ Moving Averages,” Section 14.2 in Mathematics of Statistics, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 221-223, 1962.

[8] SSL ECA dataset 1: http://cse.iutoic-dhaka.edu/group/ssl/SSL_ECE_DATASET_1.rar.

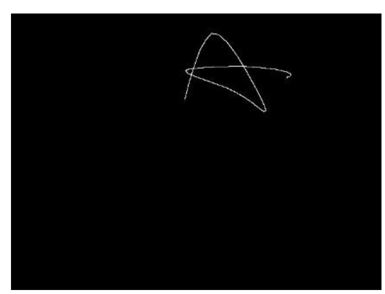
(a)



(b)



(c)



(d)

Figure 3: Preprocessing and generating time series of an ECA: (a) RGB image with depth value (b) Hand segmentation (c) Writing of ‘A’ (d) ‘A’ after smoothing