Use Hand Gesture to Write in Air Recognize with Computer Vision

Najeed Ahmed Khan†
Shariq Mahmood Khan ††
Computer Science & SE department NED University of Engg. & Tech. Karachi, Pakistan.

Maria Abdullah
Computer Systems Engineering Department Dawood University of Engineering & Technology Karachi, Pakistan

{Sana Jamaluddin Kanji, Urooj Iltifat }†††
Computer Science & SE department NED University of Engg. & Tech.Karachi, Pakistan.

Abstract
Human computer interaction has got great demand these days especially if it's serving for the society, consider the best use of technology. Computer vision is also playing a vital role for especial people to interact with the common people of the society. This paper envisages a system that uses Leap Motion device and computer vision techniques to recognize written words in air gestured by human hand. The framework is not only helpful for children to practice writing with fun but it is also useful for the dumb and handy people to communicate with other people. The leap motion device captures hand gestures, track the movements and send the frames to the system. After pre-processing input data, geometric strokes feature are extracted to recognize the written words in air. The recognized text is then display on the computer screen. The test results are out perform for English and Numeric characters with an average accuracy more than 90%.
Keywords
Air Writing; Leap Motion Controller; Hand gesture; strokes.

1. Introduction

Every person starts to talk without learning. How does that happen? When you were a baby, just being around people who were talking was enough to get you start talking. You didn't need to go to talking school or take talking lessons. Human beings' brains are just designed to make talking happen almost automatically.
Children in their young ages, when they are learning their alphabets can be taught by writing in air. As most educationists throughout the world believe and suggest that the first tool to teach children writing shouldn't be markers, crayons, pencils or pens; according to them children should be taught using a multisensory approach. Among which the most common approach used is to make children write letters in air first. This approach basically involves large muscle movements which help children process what they write in air and make it more likely for them to remember. The proposed framework used multisensory approach for teaching small children how to read and write alphabets and numbers. It takes handwritten English characters and numeric as input written in air uses image processing techniques e.g. strokes of the written character, recognize the character and then display on the screen, as shown in figure 1.
The proposed framework is also useful for elderly people, who feel hesitate to learn basics (e.g. writing alphabets, numbers etc.) from other people. The proposed framework can help to start and improve writing skills in hassle free environment and /or with any other dependency.
The people with disabilities or diseased people have difficulty to verbal communicate with normal people. Their communication with others may get easy if other people could understand their gestures. The proposed framework is also providing a user-friendly platform to such people where they will write in air and the text will be displayed on the screen.
This proposed framework can also be used for children learning alphabets and numbers, help patients to do basic physiotherapy for hand or finger and last but not the least can also be used as a source of entertainment as in those who have fun doing things in air.
Similar work has been approached using Kinect [1]. However Kinect is an expensive solution as well as have some limitation, e.g. Mobility [1] .

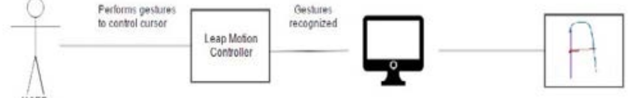


Fig 1. Basic Working of Proposed Framework

This paper has been divided into V sections including this Introduction Section I. Section II covers related literature in the field of computer vision and hand gesture recognition. Section III presented Methodology used for

Manuscript received May 5, 2017
Manuscript revised May 20, 2017

the proposed framework. Section IV describes Implementation and Evaluation Experiments of the framework. Section V discusses Conclusion and Future work.

2. Related work

Many hand gesture applications have been implemented successfully, such as PlayStation Move [2], Xbox Kinect [1] etc. These applications have the ability to identify hand gestures with good accuracy. However, each of the application has been utilized to recognize limited gestures for numeric digits [3] [4] or upper case English alphabets only.

Recognition of upper case English alphabets written in air has been proposed by the authors in [5], used Dynamic Time Warping (DTW) algorithm. They have collected hand movement data using depth cameras. The collected data is than matched with air writing standard template. The average accuracy achieved by their proposed technique is around 96.3%. The limitation of the model is recognition for only uppercase English alphabets.

In [6] the authors has been addressed the similar problem - Air writing recognition of single letter and multiple connected words using Hidden Markov Model. They have divided the recognition process in two different levels. Level one addressed the single letter recognition process whereas level two addressed the multiple connected word recognition. They have generated their own data set which is based in the six-degree of freedom. The error rate of 0.8% and 1.9% has been observed by their proposed system for character and word recognition respectively.

In [7] authors proposed air writing recognition system based on the accelerometers and gyroscopes. For Motion character modelling modified continuous Hidden Markov model was used whereas the character recognition they have used Viterbi Algorithm. Two different databases DB1 and DB2 were used to evaluate their framework.

Unistrokes [8] is a stylus-based interactions framework developed by Xerox PARC. A user is bound to learn a Unistroke alphabet that records each character to a symbol written with single stroke. Once the users get familiar to Unistroke writing, he or she can achieve speeds of 1 to 1:8 characters per second. Graffiti [9] also used single stroke alphabets for handwritten recognition. Graffiti approach looks close to their associated English alphabets. This technique is more user friendly compared to the Unistrokes.

Wii remote is another popular device used by the Nintendo Wii console [10] capable of tracking hand movement. This device uses 3-Axes accelerometer to record backward and forward movements. The optical sensors helped in positioning the device relative to the ground. Similar to these devices the PlayStation Move [2] is equipped with accelerometer and gyroscope sensors used to track the gestures. Unlike these approaches, the proposed framework does not rely on expensive hardware, and does not require training.

Writing English alphabets/words in air which is captured by the Leap Motion Controller [11] is still an open research area. New algorithms are under study to use with the Leap Motion Controller. The main reason is that it is much cost affective compare to the other proposed solution.

3. Methodology

The methodology of the proposed framework is depicted in figure 2. Functionality of the framework begins with user writing a character in air, the Leap Motion Controller gets track the movements of the finger. After pre-processing takes place frames are generated and feature extraction is done. Using computer vision techniques the text written in air is recognized on the basis of strokes. The recognized text is then displayed on the screen. The details of working of the Leap Motion controller and methodology are given below.

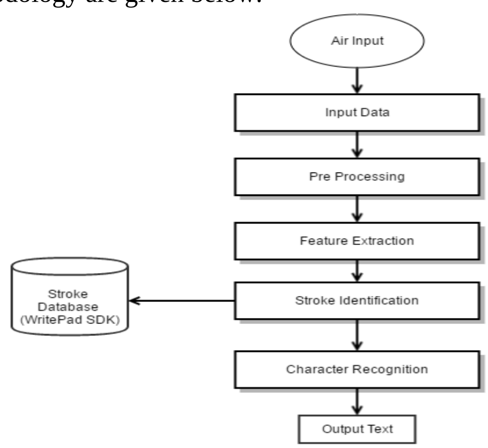


Fig 2. Data Flow Diagram of Proposed Framework

A. Leap Motion Controller

The Leap Motion controller is a small peripheral device shown in figure 3, which is designed to be placed on a physical desktop, facing upward. It can also be mounted onto a virtual reality headset. It consists two monochromatic Infrared (IR) cameras and three infrared LEDs, the device observes a roughly hemispherical area, to a distance of about 1 meter. The LEDs generate pattern-less IR light and the cameras generate almost 200 frames per second of reflected data. This is then sent through a USB cable to the host computer, where it is analyzed by the Leap Motion software using "complex math" in some way synthesizing 3D position data by comparing the 2D frames generated by the two cameras.

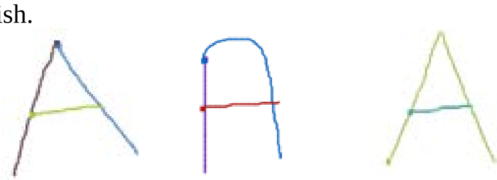
Fig 3. Leap Motion Controller

B. *Data Description*

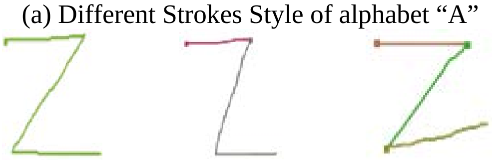
Leap Motion Controller (LMC) [11] is being used as input device capable to record finger gestures perform within the interaction region. The input data is for example, user writing in air with fingers or performing gestures with hands. All these hand movements are being tracked by Leap Motion Controller and are being sent to our application. To calibrate initially the LMC with the computer through hands, we've fixed few gestures. For air writing the finger movements is being tracked and records the strokes. Finger movements for different alphabets in English, numeric, words and some especial gestures for erase and space are recorded as input data. There are 20 people are taken to record the input data to get the variations in writing styles.

Strokes: Stroke is sudden or sharp change in the movement or it is a mark made by drawing a figure in one direction. Mathematically or geometrically it is the maximum curvature in a line curve.

Here are some examples of the strokes recorded against different gestures generated for writing letters or words in English. The alphabet 'A' has three strokes. The system stores the strokes template, which is used for recognition process. Following figure 4 and 5 show the examples of the hand written alphabets, numeric and word ('Casper') in English.



(a) Different Strokes Style of alphabet "A"



Different Strokes Style of alphabet "z"

Fig 4. Strokes Style

After the strokes the corresponding data about all the information of strokes is then taken from inspect element. Following template shows an example of the data stores against the stroke. Due to large size of the template, digits are omitted and few * are inserted instead.

["97,250;101,248;105,244;107,238;113,230;117,224;121,214;125,208;131,196;135,190;143,178;147,172;151,162;155,156;157,152;159,146;165,138;***;257,166;261,166;263,166;265,166;267,166"]

[Example of a Template of the stored data for alphabet 'A']

The recognition of the alphabet or the text is uses the template which is display on the screen accordingly. The recognition process of alphabets and numeric digits is taking place on basis of strokes; an expert system is being used to recognize the letters.

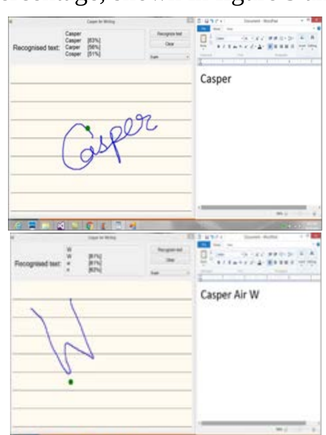
C. *Training*

We had a huge data corresponding to each of alphabet in English in addition to numeric therefore we motivated to use Neural Network and Fuzzy Logic to train the system for the proposed framework.

Neural Network [12] is a field of Artificial Intelligence which finds data structures and algorithms for learning and data classification which is inspired by a normal human brain. Neural Network techniques can be used to learn through examples and create a structure of rules to classify the different kinds of inputs for example the recognition of images.

Fuzzy logic [13] is a computing based approach which is based on 'degrees of truth' rather than the usual 'true and false' of '1 or 0' Boolean logic on which most of the modern computers are based on. In Fuzzy Logic the true value can be anywhere in between 0 and 1 and 0 and 1 are extreme cases of truth.

In the proposed frame work neural network has been used to train the model. The output of the trained model was based on the Fuzzy logic instead of binary classification. The accuracy of the recognition of each test data is measured in percentage, shown in figure 5 and figure 8.



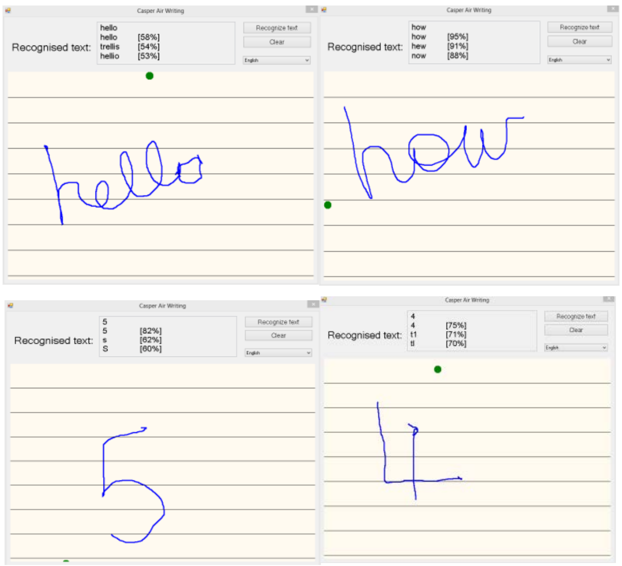


Fig. 5. Output for test writing gestures for single and mutiple connected letter as a complete word.

4. Implementation and Evaluation

The system has been tested by performing number of gestures and wrote single letters, connected letter (words) includes; digits, lower and upper English alphabets, joined words, in air using hand (finger's gestures). Results for each test are evaluated separately and reported in the following tables and figures. Good recognition results are found.

D. Evaluation Methodology / Experiments

For the evaluation of the proposed framework some experiments were created. Those experiments were carried out by some students. Each student was asked to write English alphabets (Upper case or Lower case), numeric digits or a complete word in the air. One by one each student starts writing in the air in front of LMC. The proposed framework starts recognizing the alphabets and numeric digits. The recognition accuracy and the number of attempts were recoded and analysed. The results obtained from these experiments were presented and discussed in the following section.

E. Performance Evaluationvaluation and discussions

The accuracy results obtained after testing is mentioned in Table 1. The result got were good enough for the recognition of the character. The alphabets written in the air would be compared with the features of alphabets stored in the library and the result will then be displayed on the screen giving its accuracy. We unit tested each alphabet separately and for most of the characters (capital alphabets, small alphabets, and numeric values) we got average accuracy greater than 90%.

1.Capital alphabets average accuracy rate

Capital Alphabet	Average Accuracy
A	95%
B	92%
C	96%
D	93%
E	92%
F	92%
G	95%
H	92%
I	96%
J	98%
K	97%
L	98%
M	97%
N	89%
O	90%
P	93%
Q	88%
R	93%
S	98%
T	94%
U	98%
V	97%
W	95%
X	94%
Y	96%
Z	96%

The framework was getting little confused in the recognition of "O", "0" and "Q" for these characters we got accuracy result of 90%,90% and 88% because of the resemblance in their shape. We have recorded the results of each character (capital alphabets, small alphabets and numeric values) in the tables below. The average numbers of attempts for English alphabets and numeric digits were also analysed Table 2 shows the observations of average number of attempts for recognition of numeric digits.

Table 1. average attempts for numeric digits

Numeric Digits	Attempts
0	2
1	1
2	1
3	1
4	1
5	1
6	2
7	1
8	1
9	1

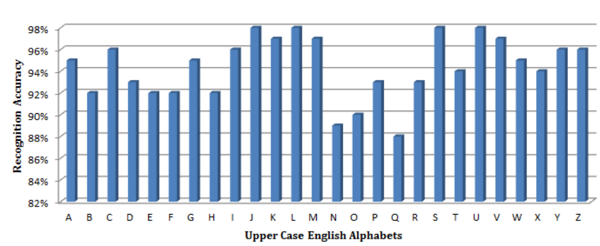


Fig.6. Upper Case Alphabets Recognition Accuracy

Figure 6 shows the average recognition accuracy of all the capital English alphabets. It shows that the system have 98% accuracy rate some alphabets i.e. J, L, S and U. The accuracy rate of the alphabet “Q” is around 88%. The overall average accuracy rate is capital alphabet is more than 90%.

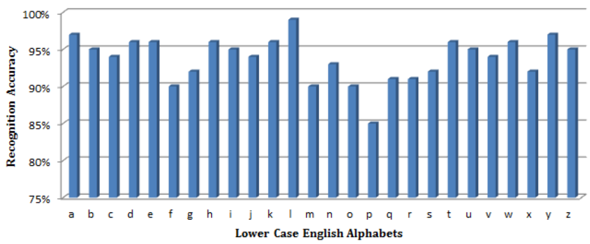


Fig. 7. Small English Alphabets Recognition Accuracy

Figure 7 presented the recognition accuracy of small English alphabet . This graph shows that some of the alphabets got the accuracy rate of 98 % for instance the character “I”. While the accuracy rate of character “p” is around 87%. The overall accuracy rate of small alphabet is more than 90%.

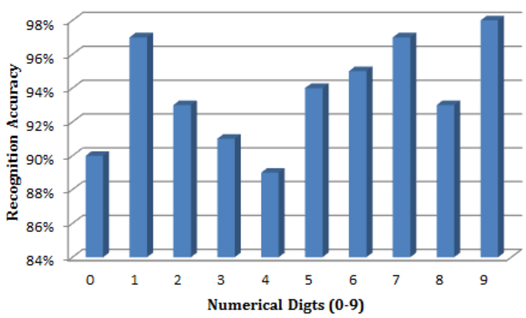


Fig.8 Numeric Value Recognition Accuracy

The average recognition accuracy of numeric values from “0” to “9” is depicted in figure 8 which shows that the average accuracy rate is more that 88%.

Apart from recognizing the single character or numeric digits this framework is also capable of recognizing the complete word. Figure 8 shows the recognition of the complete word “casper”.

5. Conclusion and Future work

This paper proposed a framework which enables its user to literally write in the air through his/her figure and hand movement. The framework records strokes of the gestures and apply computer vision approach to recognize written characters. These characters could be single English alphabets, numeric or a complete connected word e.g. Hello. Experimental results demonstrated high accuracy in recognition. This system will help children and elderly people who want to learn alphabets and number whenever he/she wants to during one day and he doesn’t have to check the availability of any physical person. This framework will also help to do basic physiotherapy for hand or finger which is proposed future direction of this framework. The evaluation of each alphabet separately and for most of the characters (upper and lower case alphabets and numeric digits) the system shows the average recognition rate is greater than 90%. In future this system can be transformed to new system which can enable dyslexic write alphabets naturally in air and to make their learning experience fun.

References.

[1] Zhang, X., et al., A new writing experience: Finger writing in the air using a kinect sensor. IEEE MultiMedia, 2013. 20(4): p. 85-93.

[2] PlayStation Move Available from: https://www.playstation.com/en-us/explore/ps3/. Acces on December 2016.

[3] Zhang, S., C. Yuan, and Y. Zhang. Handwritten character recognition using orientation quantization based on 3D accelerometer. in Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services. 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[4] Dong, Z., et al. Real-time written-character recognition using MEMS motion sensors: Calibration and experimental results. in Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on. 2009.

[5] Islam, R., et al., Alphabet Recognition in Air Writing Using Depth Information. The Ninth International Conference on Advances in Computer-Human Interactions, At Venice, Italy, 2016.

[6] Chen, M., G. AlRegib, and B.-H. Juang, Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions. IEEE Transactions on Human-Machine Systems, 2016. 46(3): p. 403-413.

[7] Xu, S. and Y. Xue. Air-writing characters modelling and recognition on modified CHMM. in Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on. 2016. IEEE.

[8] Goldberg, D. and C. Richardson. Touch-typing with a stylus. in Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems. 1993. ACM.

[9] Költringer, T. and T. Grechenig. Comparing the immediate usability of graffiti 2 and virtual keyboard. in CHI'04 Extended Abstracts on Human Factors in Computing Systems. 2004. ACM.

[10] Nintendo. Wii console. Available from: http://www.nintendo.com/wii. Acces on December 2016.

[11] Motion, L., Leap motion controller. URl: https://www. leapmotion. com, 2015. Acces on October 2016.

[12] Karayiannis, N. and A.N. Venetsanopoulos, Artificial neural networks: learning algorithms, performance evaluation, and applications. Vol. 209. 2013: Springer Science & Business Media.

[13] Gerla, G., Fuzzy logic: mathematical tools for approximate reasoning. Vol. 11. 2013: Springer Science & Business Media.