International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 08, August-2019

Air-Writing Recognition on GPU accelerated Neural Network Architecture

Trusha Patel[1], Dr. Zankhana H. Shah[2]

[1]*Information Technology Department, B V M Engineering College,*
[2]*Information Technology Department, B V M Engineering College,*

Abstract— Air-writing refers to the idea of writing the characters in free-space by the movement of hand-palm or fingers, or any writing device like a pen. In this paper, we use a generic web-camera to recognize isolated characters scripted in the air with a neural network approach. Air-Scripting is performed using a circular-like object that acts as a marker of a fixed color in front of a camera. Color-based segmentation is applied to identify the center and trace its trajectory that forms a character. The recognition is carried out by Multilayer Perceptron and Convolution Neural Network approaches. The performance depends on various illumination conditions owing to color-based segmentation. In a less fluctuating lightning condition, the system is able to recognize isolated upper-case letters with a recognition rate of 91.53% and 96.66% and lower-case letters with a recognition rate of 87.42% and 89.98% for MLP and CNN respectively.

Keywords— CNN,MLP, Air-Writing, Gesture Recognition, character recognition

I. INTRODUCTION

Gesture recognition is an intuitively computing user interface for allowing computers to capture and recognize human gestures as commands. Gesture can be any physical movement, like pointing of a finger, a nod of the head to a pinch or wave of the hand. In some cases, it may also include voice or verbal commands. Movements in air or sounds generally replace the tapping on a touch screen or key-pressing on a keyboard as the data input.

However, motions themselves are not communicative enough to input texts for the motion-based control systems. The idea of using fingers or hand movements to sketch letters, words or numbers in free space is stated as 'Air-writing'. It is useful for interfaces that do not need the user to type on a keyboard or write on a trackpad/touchscreen, or for text input for smart system control, among many applications [1][2].

The inconsistency for the motion text that represents the text is high in air-stroke writing compared to paper as there is no concrete anchors or locations. The individual who performs air-writing uses imaginary co-ordinate in space as reference points [1].

There are several ways of realizing air-writing, as described below and shown in figure 1 & 2.

- Isolated: Writing of isolated letters in an imaginary box
- Connected: Writing sequence of letters in space from left to right.
- Overlapped: Writing multiple letters stack one above others.

Figure 1: Isolated characters [1]

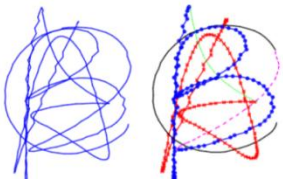

Figure 2: Overlapped characters [1]

Isolated character recognition system has fewer issues and complexities than a 'word' recognition system as the segmentation of each letter in overlapping character case and detection of the end of each letter in word for the case of the connected letter is a challenge. Thus, the suggested framework tries to recognize isolated letters crafted in free space. Nevertheless, the method used here does not involve the use of any special hardware for the implementation and is cost-effective as it employs the use of a simple camera embedded on the monitor screens.

The development of such a system can solve many real-world problems like:

1) It could provide a paperless method to convey messages. Thus, reduce the paper load and maintenance.
2) Many times the paper writing becomes in-feasible like a doctor wants to convey a message while in surgery. In such a case, such systems can be easily approached.
3) It provides an efficient method of communication for deaf and dumb. Patients suffering from pure alexia relies on finger movements to read the characters and communicate which otherwise they are unable to read or write. If speech generation is incorporated with the proposed system, a great help can be provided to the blind as the machine would speak out the recognized characters.
4) The use of keyboard for any character input for computing technology has its limitations in size due to the size of our hands, precisely our fingers. In such cases the writing in free-space can be highly beneficial.

5) The act of air-writing the letters and words creates a big cognitive impression and helps reinforce the word in the child's memory. The exercise also gives the child some valuable practice in writing that will be useful later on in their education.

The organization of the paper is as follows: The following section briefly discusses some other works in this field of study, Section III explains the adopted methodology of this work, Section IV provides the statistics of results in form of recognition rate of the English Characters on each of the two network models and Section V gives the conclusion.

II. RELATED WORK

Many systems have been developed to address the given problem. The use of depth sensors like Kinect [7], leap-motion[8], and motion tracking sensors like accelerometers and gyroscopes embedded in mobile phones are used[5]. The wearable motion control and gestures recognition systems like Myo-Armband [9] are used. Such systems are effective and accurate but they contribute greatly to the high costs. Additionally, the maintenance and use of such a system is a concern.

Chen, et al. [1][2], used leap motion for motion tracking while the HMM model was adopted for recognition of letters and words. The error rate of the word-based system was 0.8% while it was 1.9% for the letter-based system. [6] Proposed the use of neural networks to train the model. The output was based on the Fuzzy logic which described the 'degree of truth' for classification. An average accuracy of 90% was obtained for English characters and numerals.

Dash, et al. [3] used a Kinect Sensor for capturing the depth information to detect the writing activity. The hand finger is segmented and time-series information for each character is generated. Dynamic Time Wrapping algorithm is used for time-series matching of the English alphabet characters. The system gave 100% true positive with an average of 52.6% and a minimum 93.75% true negative with an average of 97.8%.

Amma, et al. [4], used Myo-arm band for sensing motion and novel model which fuse CNN (Convolutional Neural Network) and GRU (Gated Recurrent Unit) was used. The Myo-arm band produced raw IMU and EMG signals that are converted into realistic visualization of digits written in the air using a 2DifViz algorithm and recognized using three classifiers namely, CNN, GRU-NET-1 and GRU-NET-2 that are finally fed to Fusion model to improve the accuracy. It achieved the accuracy of 96.7% on person dependent evaluation and 91.7% for person independent evaluation which was better than classic models like SVM, KNN and HMM.

Khan, et al. [5], used Gyroscopes and accelerometers which are attached to the backside of the hand. The SVM was used to spot the data segments containing handwriting and HMM (Hidden Markov Model) is used to generate text representation of motion data. The error rate of 11% is achieved for a person independent environment and 3% for a person dependent. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. METHODOLOGY

This approach uses a generic web camera embedded in the laptop or any other monitor device thus avoiding the use of additional sensors to reduce the cost. The absence of depth information that is availed with special sensors makes it a bit difficult to process and represent the text data. To address this problem the image processing approach of segmenting the writing object like finger or marker tip is carried out which is followed by tracking its trajectory to project the 3D character stoke data to a 2D plane.

3.1 Writing object segmentation.

Color based segmentation of hand or finger is difficult due to variable skin tones. Hence a simple approach of segmenting an object of uniformly distributed color, here a blue colored object, is applied to act as the writing tip.

For this, the input frame is converted to HSV (Hue, Saturation, and Value) frame. Upper blue and lower blue boundary values are decided to provide masking and thresholding. If R is the range of pixel values that accepted as blue in HSV space than pixel value p at position (x,y) in the original frame is converted to pixel value q as following:

$$q(x,y) = \begin{cases} 1, & if\ p(x,y) \in R \\ 0, & otherwise \end{cases}$$

Erosion, dilation, and morphological operation were performed after thresholding to remove the noise from the segmented image, result of the same is shown in figure 3.

(a) (b) (c)

Figure 1:(a) Image Frame, (b)Segmented image, (c) Noise-removed image

3.2 Writing-Object center identification

After segmentation, the binary image is obtained for which the contours are calculated. A maximum contour area is obtained and it is labelled as a writer. A circle (yellow color) is drawn surrounding the marker in the input frame. The center of the circle marks the initialization of the writing trajectory.

Figure 4: Writing center identification

3.3 Trajectory tracking and Projection on a 2D plane.

Air-writing is continuous path motion unlike discontinuous breakpoints of conventional writing. The start of the trajectory is the first point obtained as the center of a circle surrounded the marker. As the marker moves the circle enclosing the marker area also moves along with the marker that continuously changes the coordinates of the centre. The trajectory is approximated by linearly connecting these central co-ordinates until the marker moves. The end of writing activity is identified when the writing marker moves out of the user input frame and its enclosing circle is no more available. The static image is generated at the end of the trajectory. Otsu's thresholding is performed to obtain the binary form of an image that contains only the character.

Figure 5: User frame writing path tracking

3.4 Recognizing the Character.

The static image generated after the end of the trajectory gives 2D plane representation to 3D character formed by use in the air. Before feeding it to the neural network, Gaussian blurring is performed for smoothing the image by removing edges and noise that results in a pre-processed image. The image is resized to 28x28.

Figure 6: Pre-processed image

3.5 Network Architectures.

CNN: The first layers are feature extraction layers consisting of two convolution layers followed by one max pool layer with pooling size 2x2. Each convolution layer has a kernel of size 3 and 32 and 64 filters respectively. The activation used is ReLU. The feature extraction layer is followed by a classification layer which contains two dense layers, one with 128 neurons with ReLU activation and last with 26 neuron units with softmax activation. Adadelta optimizer is adopted for optimization.

MLP: the Multilayer Perceptron network has 4 fully connected layers each with 784, 512, 512 and 26 neurons respectively. Dropouts of 0.2 were added between two layers to optimize the performance of the network architecture. ReLU activation for the first two layers and softmax activation for last layers is applied. RMSprop algorithm is used for optimization.

IV. RESULTS AND ANALYSIS

4.1 Experimental Setup:

Two neural network models were created and trained with available handwritten character dataset as no air-written character dataset is available as of now. The EMNIST Letter dataset with 145600 image samples of 26 classes corresponding to uppercase and lower case letters (balanced) was used but we reduced the sample size to 124800 images. The images are gray-scaled each of size 28x28. The images were divided into two categories: training, testing in 75% and 25% ratio. Training images were used to train the network and testing images were used to evaluate the network. The training images were further classified into two categories: Training and Validation. Training images were 70% of the total training set and validation consists of 30% of the total training set.

The models have been trained on high-end NVIDIA's GeForce GTX 1050 Ti Graphical Processing Unit with TensorFlow backend in 16 GB RAM computer. The use of GPU accelerates the training process due to simple cores that accounts for the parallel computing through thousands of threads. The training time for MLP was 3.2 minutes and CNN was 4.3 minutes with 10 epochs each.

4.2 Error analysis:

The recognition rates were calculated on the basis of 15 strokes of each alphabetic character.

$$Recognition\ Rate = \frac{Number\ of\ strokes\ with\ True\ Output}{Total\ Number\ of\ strokes}\ 100$$

Alphabets	Recognition Rate
A, B, C	100%
D	93.3%
E, F, G, H, I, J, K	100%
L	0%
M	100%
N	86.6%
O	66.67%
P, Q	100%
R	93.3%
S, T	100%
U	66.67%
V	86.6%
W, Y	93.3%
X, Z	100%

Table 1: MLP Recognition Rate for upper-case letters

Input	Output	Recognition Rate (*(Number of Strokes with Wrong output /Total number of Strokes x100)*
D	O	6.7%
M	N	6.7%
L	I, K	100%
N	M	13.4%
O	D, Q	33.33%
R	K	6.7%
U	V, W	33.33%
V	Y, U	**13.4%**
W	U	6.7%
Y	V	6.7%

Table 1: Wrongly recognized upper-case letters' Recognition Rate for MLP

Alphabets	Recognition Rates
A, B, C, D, E, F	100%
G, H, I, J, K, L	100%
M	93.3%
N	73.33%
O	86.6%
P, Q, R, S, T	100%
U	66.67%
V	93.3%
W, X, Y, Z	100%

Table III: CNN Recognition Rate of upper-case letters

Input	output	Recognition rate *(Number of Strokes with Wrong output /Total number of Strokes x100)*
M	N	6.7%
N	M, W	26.67%
O	D, Q	13.4%
U	V, W	33.33%
V	W	6.7%

Table IV: Wrongly recognized upper-case letters' Recognition Rate for CNN

Alphabets	Recognition Rate
a, g	86.6%
b, c, d, e, f	100%
h	93.3%
i, l	0%
j	100%
k	100%
m	100%
n, o, p, s, v, w, x, y, z	100%
q	86.6%
r, t, u	73.33%

Table V: Recognition Rate of MLP for lower-case letters

Input	Output	Recognition rate *(Number of Strokes with Wrong output /Total number of Strokes x100)*
a	q	13.4%
g	y	13.4%
h	n	6.7%
i	q, k, f, l, j, g	100%
l	k, f	100%
q	a, g	13.4%
r	v, n,y	26.67%
t	k	26.67%
u	v, w	26.67%

Table VI: Wrongly recognized lower-case letters' Recognition Rate for MLP

Input	Recognition Rate
a, n, q	93.3%
b, c, d, e, f, g, h, j, k	100%
i, l	0%
m, o, p	100%
n, q	93.3%
r, t, u	86.6%
s, v, w, x, y, z	100%

Table VII: Recognition Rate of CNN for lower-case letters

Input	output	Recognition rate *(Number of Strokes with Wrong output /Total number of Strokes x100)*
a	q	6.7%
i	i, f, j, k, q, g	100%
l	j, i, f, k, q	100%
n	h	6.7%
q	g	6.7%
r	n, v, y	13.4%
t	k	13.4%
u	v, w	13.4%

Table VIII: Wrongly recognized lower-case letters' Recognition Rate for CNN

The recognition rate of lower-case letters with MLP and CNN are 87.42% and 89.98% respectively.

IV. CONCLUSION & FUTURE SCOPE

We suggested a method that enables robust air-writing recognition in real-time. The strokes in free-space can be either uppercase or lowercase letters of the English Language. The given system provides an efficient communication interface to deaf or dumb people. It also helps in reducing the paper load by providing an approachable method to convey quick and trivial messages. The system evaluates the recognition rate of each alphabet on two neural network architectures. The MLP gave the recognition rate of 91.53% for uppercase letters and 87.42% for lowercase letters while the CNN gave a

higher recognition rate of 96.6% for uppercase and 89.98% for lowercase letters than MLP. This provides with the intuition that CNN outperforms MLP for air-written characters recognition. In the future this system can be combined with a speech interface that can speak the recognized letter to aid the blind and also can be extended to recognize the words.

V. REFERENCES

[1] Chen, M., AlRegib, G., and Juang, B.-H., 2015, Air-writing recognition, part 1: Modeling and recognition of characters, words and connecting motions, IEEE Trans. Human Mach. Syst
[2] M. Chen, G. AlRegib, and B.-H. Juang, Air-writing recognition, part 2: Detection and recognition of writing activity in continuous stream of motion data, IEEE Trans. Human Mach. Syst, Islam, R., Mahmud, H., Hasan, K., 2016, Alphabet Recognition in Air Writing Using Depth Information, The Ninth International Conference on Advances in Computer -Human Interactions
[3] Dash, et al., 2017, AirScript - Creating Documents in Air, 14th International Conference on Document Analysis and Recognition
[4] Amma, C., et al., 2014 Airwriting: a wearable handwriting recognition system, Personal and Ubiquitous Computing, Volume 18, pp-191
[5] Khan, N.A., et al., 2017, Use Hand Gesture to Write in Air Recognize with Computer Vision, IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.5
[6] Microsoft Corporation. Kinect. 2010. URL: https://developer.microsoft.com/enus/windows/kinect/.
[7] Leap Motion Inc. LEAP Motion. 2010. URL: https: / /www.leapmotion.com/.
[8] Thalmic Labs Inc. Myo. 2013. URL: https://www.myo.com/
[9] Y. LeCun, C. Cortes, and C. J. C. Burges. The MNIST Database of Handwritten Digits. 1998.