

Air-Written Digits Recognition using CNN for Device Control

Easily predict the hand written digits drawn in air using Convolution Neural Network to control an IR Device

Akshay J R

Department of Computer Science
and Engineering
Cambridge Institute of Technology
Bengaluru, India

Varalatchoumy M

Associate Professor
Department of Computer Science
and Engineering
Cambridge Institute of Technology
Bengaluru, India

Hitesh C

Department of Computer Science
and Engineering
Cambridge Institute of Technology
Bengaluru, India

Chethan V

Department of Computer Science and Engineering
Cambridge Institute of Technology
Bengaluru, India

Jagadish C

Department of Computer Science and Engineering
Cambridge Institute of Technology
Bengaluru, India

Abstract - Air-Written digits give a complimentary methodology to general human-pc communication. They are intended to be simple so that the users can effectively retain and perform them. Air writing framework alludes to composing a semantic character or digits in free space by moving a marker, or some other handheld gadget. It is generally pertinent where conventional pen-up and pen-down composing frameworks are problematic because of its basic composing style, it has an extraordinary bit of leeway over the motion-based framework. However, it is a difficult task because of the non-uniform characters and diverse composing styles. In this project, we developed an air-writing digit recognition system using a camera that tracks a marker to recognize the digits later these digits are sent to an IR module which can be used to control any device that supports IR signals. For better feature selection, we processed the data by using Erosion and Morphological pre-processing techniques and employed convolutional neural network (CNN) a deep learning algorithm for image recognition. The model was tested and verified by a self-generated dataset. To evaluate the robustness of our system, we tested the recognition model and the IR module with 700 test data with over 5 IR devices respectively. Hence, it verifies that the proposed model is invariant for digits and any IR gadgets.

Keywords - Air-written digits, Camera, CNN, IR-Module.

I. INTRODUCTION

Humans interact with electronic devices like TV using their remote controller but the communication with such devices can be made easy by a unique system of Air-Written digit recognition. Air-writing digit recognition empowers users to communicate with the machine and interact naturally with no mechanical gadgets where air-writing is a procedure of composing something in a 3D space by utilizing characters, signals or trajectory data. It permits clients to write in a touchless framework. Particularly, it is valuable when conventional

writing is troublesome, for example, gesture-based communication, augmented reality (AR), virtual reality (VR), and so forth. In, gesture based writing the number of gestures is constrained by human stance, however it is conceivable to expand the number of gestures by consolidating them. In any case, recalling these is difficult for new users. On the other hand, in air writing, users can write in the same order as traditional method like 0-9. By utilizing this idea of air written digit recognition, it is possible to point a finger towards the camera to perform the right task. So, in this project, a camera was utilized to gather the information of air written digit and once this data is collected erroneous, temporal and spatial noise is removed by applying some filters and normalization techniques like for example Erosion and Morphological techniques. However, zigzag effects may persist due to the special writing style. Users write digits in the 3D space inside an imaginary box, which is not fixed; no delimiter distinguishes the boundary of the region of interest. Thus, the digits are unstable whereas unwanted sequences are drawn that make the framework more challenging. Thus, an efficient deep learning-based algorithm, named the convolutional neural network (CNN) was deployed to predict the image drawn in air. In this we mainly used CNN because it works well with the image dataset. The main contributions of this project are as follows. (1) Designed a proficient deep learning models allowing permitting air-writing digit recognition with an accuracy rate of 93%. (2) Successfully control and IR device by using the thus predicted digits.

II. RELATED WORK

Past attempts to handle air-writing issue were depended upon multi-camera arrangement to judge depth data, depth sensors, for example, LEAP Motion [1], Kinect [2], or motion control gadget, for example, Myo [3] and wearable gesture control gadgets. While these methodologies represent simpler tracking and better precision, they suffer from cost-effective general-

purpose usage because of elementary reliance on the external instruments.

Schick et al. [4] proposed a sensor free marker-less system utilizing numerous cameras for 3D hand tracking followed by recognition with HMM. This strategy is accounted to attain 86.15% recognition rate for characters and 97.54% recognition rate for words recognition. Chen et al. [5], [6] utilized LEAP Motion for tracking and a Hidden Markov Model (HMM) for recognition and reported an error rate of 0.8% for recognition of words and error rate of 1.9% for recognition of letters. Dash et al. [7] used Myo armband sensor on a completely unique Fusion model architecture by consolidating one Convolutional Neural Network (CNN) and two Gated Recurrent Units (GRU). The Fusion model is accounted to surpass other generally utilized models, for example, support vector machine, k-nearest neighbors and accomplished a precision of 91.7% in an individual free assessment and 96.7% in a person dependent assessment.

III. PROPOSED METHODOLOGIES

All the above existing systems have either made use of some special purpose sensors or a camera setup hence these kind of setup increases the overall cost of the system to which it has to implemented and thus the mainstream adoption of these systems is restricted. Therefore, to overcome these restrictions, a generic video camera was used to implement an air-writing system which can be commonly used in any devices (such as TV etc.) consisting of an in-built video camera.

In order to achieve the above aim, this project mainly concentrates on constructing a machine which consists of a camera and an IR circuit, interconnected with each other by using an Arduino and a Raspberry Pi.

The Fig. 1 represents the architecture of the above discussed system/machine consisting of a Raspberry PI, camera, Arduino, IR transmitter and receiver.

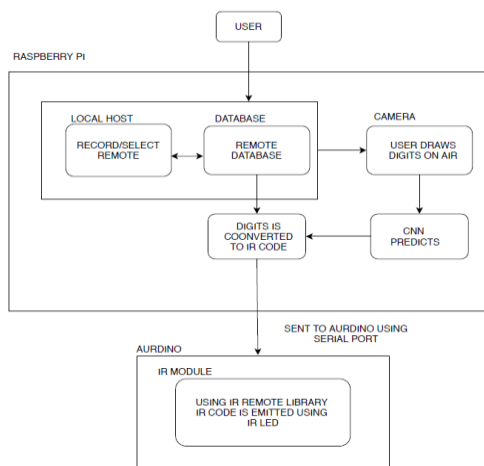


Fig. 1. Representation of System Architecture

The first layer of the architecture is the User layer. User layer will comprise of the people who interacts with the framework for the required results. The next layer in system architecture is the Raspberry PI which consists of Local Host, Data Base, Camera and CNN model for digit recognition. The local host consists of web pages which is developed using HTML, CSS, and PYTHON on DJANGO SERVER. These webpages are accessed by the user using a browser to record new remote, select or delete an existing remote. The Raspberry Pi also consists of a data base which stores the information of all the remotes which are recorded by the user. The information stated here, is the IR codes of the buttons in the remote, it is stored in a '.txt' file named with the name of the remote itself. After selecting an appropriate remote the user draws the digits on air, the camera captures this air written digit and sends it to the CNN module for digit prediction. The CNN module consists of a dedicated CNN model for hand written digit recognition, this CNN model predicts/classifies the digit drawn on air. The digits thus classified/predicted is then converted to the respective IR code, using the IR database. This IR code is then sent to Arduino through a serial port, the Arduino consists of IR module which consists of IR library namely "irremote.h", using this library the IR code is then transmitted to the respective IR device with the help of an IR LED.

A. Marker Segmentation

The major problem we face was the high variability of human skin color, which gave a difficulty in differentiating the segment of the hand from the background using color-based segmentation technique. To overcome this difficulty, a simplistic approach was implemented in the present work. The feasibility of the air-writing system can be demonstrated by ignoring finger detection and hand segmentation, instead made use of marker of a fixed color for writing the digits in the air which is a three-dimensional space.

B. Digits Recognition

The marker tip trajectory for air-written digit are projected from three-dimensional space onto a two-dimensional image plane and in order to predict the written character from the fetched projected image a pre-trained convolutional neural network (CNN) was used. At the time of this work, a dataset for air-written digits was created. Therefore, the CNN model was trained with the created dataset and later the trained model was used to predict the air-written digit.

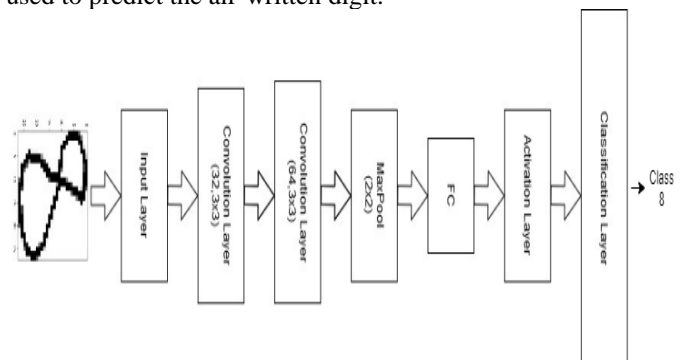


Fig. 2. Representation of CNN model

The architecture of the CNN model is shown in the Fig. 2. As discussed above in this project the CNN model is used for

mapping the images to its corresponding labels. CNN model consists of three-layer namely input layer, hidden layer and output layer. The detailed description of the layer is as follows

- Input Layer

The input layer is where the image is read. The image given to the input layer is of size 28x28. The image is preprocessed in order to remove any noise present in the background and in the edges of the digits. These images are saved in a csv file where each row consist of the digit label and 784 pixels (28x28) information.

- Hidden Layer

This layer is the backbone of the CNN architecture. The feature extraction is performed here with a series of convolution layer, pooling and activation function.

- Convolution Layer

The Convolution layer is used for extracting the features from the images. A kernel of size 3x3 is used in the convolution layer, which is used to scan through the image and find the unique feature. In kernel the weights are selected randomly which is then multiplied to the scanned image from left to right.

- Pooling Layer

The main purpose of this layer is to reduce the spatial dimensionality. The output of the pooling layer is a pooled featured map. The pooling layer is generally placed between two convolution layers. The pooling layer is used for controlling overfitting and feature selection. The commonly used pooling layer is mac=x pooling. Consider a 4x4 convolved is to be converted to 2x2. This is done by taking a 2x2 kernel size and selecting a maximum value present in that kernel. Therefore, the resulting output will a 2x2 convolved layer.

- Activation Layer

This layer is used to introduce non linearity in the system. There main activation function which are SoftMax, Relu etc. The activation function which is used in the present work is nonlinear rectified liner unit. The Relu function is used for normalizing the value present in the layer. The negative values present in the input is replaced with "0" as output and the positive values are kept as it is without any changes.

- Output Layer

This is also known as the classification layer. This is a fully connected feed forward network used as classifiers. In this fully connected layer, the neurons are connected to all the pervious layers neurons. This layer calculates predicted classes based on the combined result of the features learned in the pervious layers. The number of output classes depends on the number of labels in the target dataset. SoftMax function is used in the present work for classifying the input images based on the features generated during the previous layers to various classes based on the trained data.

C. Signal Transmission

Once convolutional neural network (CNN) predicts the character, the respective character is then converted to IR raw code using the IR database stored in Raspberry Pi. The IR raw code consists of an approximately 64 numbers which represents the Address bits and Command bits of an IR signal. Once the IR raw code is fetched from the database it is then sent to the Arduino board which is configured as an 38Khz frequency IR remote. In this project we have particularly configured the Arduino as an 38KHz frequency IR remote because majority of the IR remotes work around this particular frequency. But with further involvement in the present work, it was recognized that there is no fixed representation for zeros and ones with data transmission over different IR methods. There are various different encoding techniques for generating these codes. The significance of 38 KHz is that the frequency at which signal oscillates when logically high that is this is the carrier frequency of the signal. The Fig. 2 shows an example IR signal in NEC Protocol.

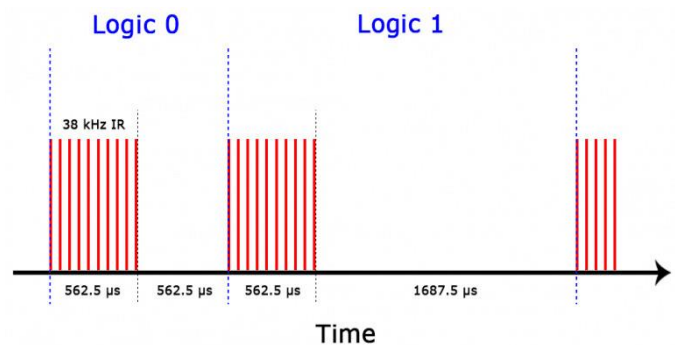


Fig. 3. Representation of an IR signal in NEC protocol

There are many different protocols for IR data transmission and receiving like NEC, RFC etc. among this NEC protocol was used preferably by different device brands, so let take an example to demonstrate how the receiver converts the modulated IR signal to a binary one and zeros.

The NEC protocol follows these rules for transmission and receiving or IR data that is Logical '1' starts with a 562.5 μs long HIGH pulse of 38 kHz IR followed by a 1,687.5 μs long LOW pulse. Logical '0' is transmitted with a 562.5 μs long HIGH pulse followed by a 562.5 μs long LOW pulse.

Therefore, for transmitting a digit through Arduino using the NEC protocol, the message transmission should be carried out in order, first a 9ms leading pulse burst (16 times the pulse burst length used for a logical data bit) and followed be a 4.5ms space, then a 8-bit address for the receiving device the 8-bit and logical inverse of the address, followed by a 8-bit command and the 8-bit logical inverse of the command. Finally, a 562.5μs pulse burst to signify the end of message transmission. The address and command are also sent in the inverse for error reduction and validation. Figure 1 illustrates the format of an NEC IR transmission frame, for an address of 00h (00000000b) and a command of B5h (10101101b).

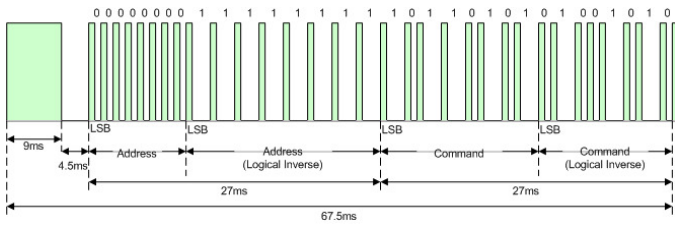


Fig. 4. Representation of IR signal in Analog waveform

From the Fig. 3, we can notice that a 16-bits device address (address + logical inverse) is transmitted in 27ms and the other 27ms are taken to transmit a 16 bits command (command + logical inverse) and 67.5ms to transmit the message frame completely. The inverse of the command and address are sent for reliability and verification purpose.

IV. EXPERIMENTAL RESULTS AND FUTURE ENHANCEMENT

In this project a robust CNN model is proposed for air-written digit recognition in order to control an IR device. To keep away from the troubles of human skin division a marker of uniform shading is utilized. Recognition is performed through tweaking of air-writing information and by a pre-trained CNN model. In this experiment the proposed system accomplished 93% recognition rate over English numerals and also supported all the IR devices. This strategy utilizes a camera and is completely independent on any depth or motion sensor for example Kinect, LEAP Motion, Myo and so an., which is the main advantage of this system over existing methodologies as it decreases the cost drastically.



Fig. 5. Drawing the Numerical

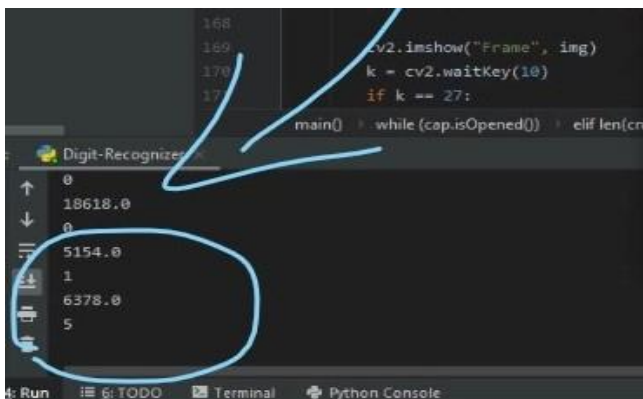


Fig. 6. Prediction from the CNN Model

From the Fig. 5 and Fig. 6 we can note that the CNN model is able to predict the numerical drawn in air.

This work can additionally be upgraded by receiving this system for utilizing straightforwardly on hands with no marker for better unrestricted generalization of the proposed method, the cost of this machine can be further reduced by designing an circuit board which consists of only with the resources /components which are necessary for this framework, enable voice recognition system which helps the user when air-written recognition fails, enabling OTA updates for the framework in order to increase the efficiency of the of the CNN model, dynamically learn different kinds of handwritings, fonts of different digits with the help of user inputs so that the efficiency of the CNN model automatically increases with respect to the user, enable backup of the OS, training data, and CNN model for easier recovery of the system.

REFERENCES

- [1] Leap Motion Inc. LEAP Motion. 2010. URL: <https://www.leapmotion.com/>.
- [2] Microsoft Corporation. Kinect. 2010. URL: <https://developer.microsoft.com/en-us/windows/kinect/>.
- [3] Thalmic Labs Inc. Myo. 2013. URL: <https://www.myo.com>.
- [4] A. Schick, D. Morlock, C. Amma, et al. "Visionbased Handwriting Recognition for Unrestricted Text Input in Mid-air". In: *Proceedings of the 14th ACM International Conference on Multimodal Interaction. ICMI '12*. ACM, 2012, pp. 217–220.
- [5] M. Chen, G. AlRegib, and B. H. Juang. "Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions". In: *IEEE Transactions on Human-Machine Systems* 46.3 (June 2016), pp. 403–413. ISSN: 2168-2291. DOI: 10.1109/THMS.2015.2492598.
- [6] M. Chen, G. AlRegib, and B. H. Juang. "Air-Writing Recognition—Part II: Detection and Recognition of Writing Activity in Continuous Stream of Motion Data". In: *IEEE Transactions on Human-Machine Systems* 46.3 (June 2016), pp. 436–444. ISSN: 2168-2291. DOI: 10.1109/THMS.2015.2492599.
- [7] A. Dash, A. Sahu, R. Shringi, et al. "AirScript—Creating Documents in Air". In: *14th International Conference on Document Analysis and Recognition*. 2017, pp. 908–913.
- [8] P. O. Kristensson, T. Nicholson, and A. Quigley. "Continuous recognition of one-handed and two-handed gestures using 3D full-body motion tracking sensors". In: *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. ACM. 2012, pp. 89–92.
- [9] Xuan Yang, et al., [2016], "Multi - Digit using Convolution Neural Network" in Stanford University.
- [10] T Siva Ajay, [JULY – 2017], "Hand Written Digit Recognition Using Convolution Neural Network", in *International Research Journal of Engineering and Technology (IRJET)*.
- [11] Wan Zhu, [2018], "Classification of MNIST Hand Written Digit Database Using Neural Network".
- [12] Vijayalaxmi R Rudraswaminmath, et al., [June – 2019], "Hand Written digit recognition" in *International Journal of Innovative Science and Technology*.
- [13] Prasun Roy, Subhankar Ghosh, Umapada Pal. "A CNN Based Framework for Unistroke Numeral Recognition in Air-Writing", 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018.
- [14] www.mdpi.com
- [15] Mingyu Chen, Ghassan AlRegib, Biing-Hwang Juang. "Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions", *IEEE Transactions on Human-Machine Systems*, 2016.
- [16] techdocs.altium.com
- [17] www.circuitbasics.com