

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354142408>

AIR CANVAS APPLICATION USING OPENCV AND NUMPY IN PYTHON

Article in International Journal of Research in Engineering and Technology · August 2021

CITATIONS

10

READS

18,737

5 authors, including:



[Saurabh uday Saoji](#)

Bharati Vidyapeeth Deemed University

10 PUBLICATIONS 17 CITATIONS

SEE PROFILE

AIR CANVAS APPLICATION USING OPENCV AND NUMPY IN PYTHON

Prof. S.U. Saoji

*Department of Computer Engineering
Bharati Vidyapeeth
(Deemed to be University)
College of Engineering, Pune*

Nishtha Dua

*Department of Computer Engineering
Bharati Vidyapeeth
(Deemed to be University)
College of Engineering, Pune*

Akash Kumar Choudhary

*Department of Computer Engineering
Bharati Vidyapeeth
(Deemed to be University)
College of Engineering, Pune*

Bharat Phogat

*Department of Computer Engineering
Bharati Vidyapeeth
(Deemed to be University)
College of Engineering, Pune*

Abstract – Writing in air has been one of the most fascinating and challenging research areas in field of image processing and pattern recognition in the recent years. It contributes immensely to the advancement of an automation process and can improve the interface between man and machine in numerous applications. Several research works have been focusing on new techniques and methods that would reduce the processing time while providing higher recognition accuracy.

Object tracking is considered as an important task within the field of Computer Vision. The invention of faster computers, availability of inexpensive and good quality video cameras and demands of automated video analysis has given popularity to object tracking techniques. Generally, video analysis procedure has three major steps: firstly, detecting of the object, secondly tracking its movement from frame to frame and lastly analysing the behaviour of that object. For object tracking, four different issues are taken into account; selection of suitable object representation, feature selection for tracking, object detection and object tracking. In real world, Object tracking algorithms are the primarily part of different applications such as: automatic surveillance, video indexing and vehicle navigation etc.

The project takes advantage of this gap and focuses on developing a motion-to-text converter that can potentially serve as software for intelligent wearable devices for writing from the air. This project is a reporter of occasional gestures. It will use computer vision to trace the path of the finger. The generated text can also be used for various purposes, such as sending messages, emails, etc. It will be a powerful means of communication for the deaf. It is an effective communication method that reduces mobile and laptop usage by eliminating the need to write.

Keywords - Air Writing, Character Recognition, Object Detection, Real-Time Gesture Control System, Smart Wearables, Computer Vision.

I. INTRODUCTION

In the era of digital world, traditional art of writing is being replaced by digital art. Digital art refers to forms of expression and transmission of art form with digital form. Relying on modern science and technology is the distinctive characteristics of the digital manifestation. Traditional art refers to the art form which is created before the digital art. From the recipient to analyse, it can simply be divided into visual art, audio art, audio-visual art and audio-visual imaginary art, which includes literature, painting, sculpture, architecture, music, dance, drama and other works of art. Digital art and traditional art are interrelated and interdependent. Social development is not a people's will, but the needs of human life are the main driving force anyway. The same situation happens in art. In the present circumstances, digital art and traditional art are inclusive of the symbiotic state, so we need to systematically understand the basic knowledge of the form between digital art and traditional art. The traditional way includes pen and paper, chalk and board method of writing. The essential aim of digital art is of building hand gesture recognition system to write digitally. Digital art includes many ways of writing like by using keyboard, touch-screen surface, digital pen, stylus, using electronic hand gloves, etc. But in this system, we are using hand gesture recognition with the use of machine learning algorithm by using python programming, which creates natural interaction between man and machine. With the advancement in technology, the need of development of natural 'human – computer interaction (HCI)' [10] systems to replace traditional systems is increasing rapidly.

This paper's remainder is categorized as follows: Section 2 presents the other pieces of literature that we referred to before working on this project. Section 3 describes the challenges we faced while making this system. In Section 4, we define the problem statement we were solving. Section 5 provides the system methodology and workflow that we followed. The subsections of section 5 include - Fingertip Recognition Dataset Creation and Fingertip Recognition Model Training. Section 6 algorithm of workflow.

II. LITERATURE REVIEW

A. Robust Hand Recognition with Kinect Sensor

In [3], the system proposed used the depth and colour information from the Kinect sensor to detect the hand shape. As for gesture recognition, even with the Kinect sensor. It is still a very challenging problem. The resolution of this Kinect sensor is only 640×480. It works well to track a large object, e.g., the human body. But following a tiny thing like a finger is complex.

B. LED fitted finger movements

Authors in [4] suggested a method in which an LED is mounted on the user's finger, and the web camera is used to track the finger. The character drawn is compared with that present in the database. It returns the alphabet that matches the pattern drawn. It requires a red-coloured LED pointed light source is attached to the finger. Also, it is assumed that there is no red-coloured object other than the LED light within the web camera's focus.

C. Augmented Desk Interface

In [5] Augmented segmented desk interface approach for interaction was proposed. This system makes use and a video projector and charge-coupled device (CCD) camera so that using the fingertip; users can operate desktop applications. In this system, each hand performs different tasks. The left hand is used to select radial menus, whereas the right hand is used for selecting objects to be manipulated. It achieves this by using an infrared camera. Determining the fingertip is computationally expensive, so this system defines search windows for fingertips.

III. CHALLENGES IDENTIFIED

A. Fingertip detection

The existing system only works with your fingers, and there are no highlighters, paints, or relatives. Identifying and characterizing an object such as a finger from an RGB image without a depth sensor is a great challenge.

B. Lack of pen up and pen down motion

The system uses a single RGB camera to write from above. Since depth sensing is not possible, up and down pen movements cannot be followed.

Therefore, the fingertip's entire trajectory is traced, and the resulting image would be absurd and not recognized by the model. The difference between hand written and air written 'G' is shown in Figure 1.



Figure 1: Actual Character vs Trajectory

C. Controlling the real-time system

Using real-time hand gestures to change the system from one state to another requires a lot of code care. Also, the user must know many movements to control his plan adequately.

IV. PROBLEM DEFINITION

The project focuses on solving some major societal problems –

1. People hearing impairment: Although we take hearing and listening for granted, they communicate using sign languages. Most of the world can't understand their feeling, their emotions without a translator in between.
2. Overuse of Smartphones: They cause accidents, depression, distractions, and other illnesses that we humans can still discover. Although its portability, ease to use is profoundly admired, the negatives include life-threatening events.
3. Paper wastage is not scarce news. We waste a lot of paper in scribbling, writing, drawing, etc.... Some basic facts include - 5 liters of water on average are required to

make one A4 size paper, 93% of writing is from trees, 50% of business waste is paper, 25% landfill is paper, and the list goes on. Paper wastage is harming the environment by using water and trees and creates tons of garbage.

Air Writing can quickly solve these issues. It will act as a communication tool for people with hearing impairment. Their air-written text can be presented using AR or converted to speech. One can quickly write in the air and continue with your work without much distraction. Additionally, writing in the air does not require paper. Everything is stored electronically.

V. SYSTEM METHODOLOGY

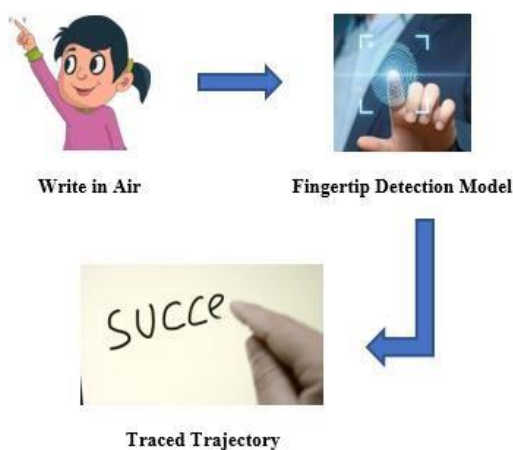


Figure 2: Workflow of the system

This system needs a dataset for the Fingertip Detection Model. The Fingertip Model's primary purpose is used to record the motion, i.e., the air character.

A. Fingertip Detection Model:

Air writing can be merely achieved using a stylus or air-pens that have a unique colour [2]. The system, though, makes use of fingertip. We believe people should be able to write in the air without the pain of carrying a stylus. We have used Deep Learning algorithms to detect fingertip in every frame, generating a list of coordinates.

B. Techniques of Fingertip Recognition Dataset Creation:

a. Video to Images: In this approach, two-second videos of a person's hand motion were captured in different environments. These videos were then broken into 30 separate images, as shown in Figure 3. We collected 2000 images in total. This dataset was labeled manually using

LabelImg[13]. The best model trained on this dataset yielded an accuracy of 99%. However, since the generated 30 images were from the same video and the same environment, the dataset was monotonous. Hence, the model didn't work well for discrete backgrounds from the ones in the dataset.

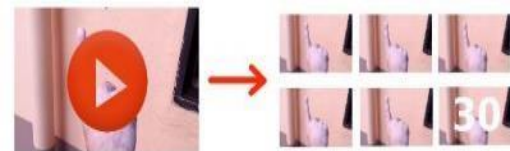


Figure 3: Video to Images

b. Take Pictures in Distinct Backgrounds: To overcome the drawback caused by the lack of diversity in the previous method, we created a new dataset. This time, we were aware that we needed some gestures to control the system. So, we collected the four distinct hand poses, shown in Figure 4.

The idea was to make the model capable of efficiently recognizing the fingertips of all four fingers. This would allow the user to control the system using the number of fingers he shows. He or she could now - promptly write by showing one index finger, convert this writing motion to e-text by offering two fingers, add space by showing three fingers, hit backspace by showing five fingers, inter prediction mode by showing four fingers, and then the show 1,2,3 fingers to select the 1st, 2nd or 3rd prediction respectively. To get out of prediction mode, show five fingers. This dataset consisted of 1800 images. Using a script, the previously trained model was made to auto-label this dataset. Then we corrected the mislabelled images and introduced another model. A 94% accuracy was achieved. Contrary to the former one, this model worked well in different backgrounds.



Figure 4: Taking pictures in different backgrounds

C. Fingertip Recognition Model Training:

Once the dataset was ready and labeled, it is divided into train and dev sets (85%-15%). We used Single Shot Detector (SSD) and Faster RCNN pre-trained models to train our dataset. Faster RCNN was much better in terms of accuracy as compared to SSD. Please refer to the

Results Section for more information. SSDs combine two standard object detection modules – one which proposes regions and the other which classifies them. This speeds up the performance as objects are detected in a single shot. It is commonly used for real-time object detections. Faster RCNN uses an output feature map from Fast RCNN to compute region proposals. They are evaluated by a Region Proposal Network and passed to a Region of Interest pooling layer. The result is finally given to two fully connected layers for classification and bounding box regression [15]. We tuned the last fully connected layer of Faster RCNN to recognize the fingertip in the image.



Figure 5: Word written in air traced on a black image

VI. ALGORITHM OF WORKFLOW

This is the most exciting part of our system. Writing involves a lot of functionalities. So, the number of gestures used for controlling the system is equal to these number of actions involved. The basic functionalities we included in our system are

1. Writing Mode - In this state, the system will trace the fingertip coordinates and stores them.
2. Colour Mode – The user can change the colour of the text among the various available colours.
3. Backspace - Say if the user goes wrong, we need a gesture to add a quick backspace.

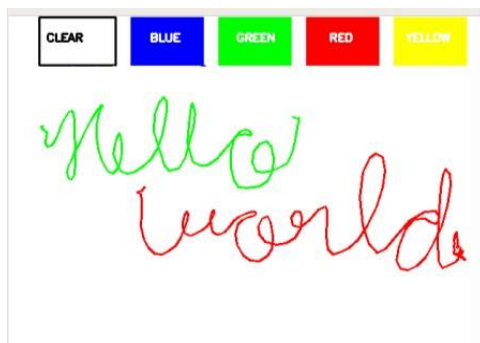


Figure 6: Front End of System

VII. CONCLUSION

The system has the potential to challenge traditional writing methods. It eradicates the need to carry a mobile phone in hand to jot down notes, providing a simple on-the-go way to do the same. It will also serve a great purpose in helping especially abled people communicate easily. Even senior citizens or people who find it difficult to use keyboards will be able to use the system effortlessly. Extending the functionality, the system can also be used to control IoT devices shortly. Drawing in the air can also be made possible. The system will be an excellent software for smart wearables using which people could better interact with the digital world. Augmented Reality can make text come alive. There are some limitations of the system which can be improved in the future. Firstly, using a handwriting recognizer in place of a character recognizer will allow the user to write word by word, making writing faster. Secondly, hand-gestures with a pause can be used to control the real-time system as done by [1] instead of using the number of fingertips. Thirdly, our system sometimes recognizes fingertips in the background and changes their state. Air-writing systems should only obey their master's control gestures and should not be misled by people around. Also, we used the

EMNIST dataset, which is not a proper air-character dataset. Upcoming object detection algorithms such as YOLO v3 can improve fingertip recognition accuracy and speed. In the future, advances in Artificial Intelligence will enhance the efficiency of air-writing.

VIII. REFERENCES

- [1] Y. Huang, X. Liu, X. Zhang, and L. Jin, "A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach, and Application," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, pp. 370-377, 2016.
- [2] P. Ramasamy, G. Prabhu, and R. Srinivasan, "An economical air writing system is converting finger movements to text using a web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, pp. 1-6, 2016.
- [3] Saira Beg, M. Fahad Khan and Faisal Baig, "Text Writing in Air," Journal of Information Display Volume 14, Issue 4, 2013
- [4] Alper Yilmaz, Omar Javed, Mubarak Shah, "Object Tracking: A Survey", ACM Computer Survey. Vol. 38, Issue. 4, Article 13, Pp. 1-45, 2006

- [5] Yuan-Hsiang Chang, Chen-Ming Chang, "Automatic Hand-Pose Trajectory Tracking System Using Video Sequences", INTECH, pp. 132- 152, Croatia, 2010
- [6] Erik B. Sudderth, Michael I. Mandel, William T. Freeman, Alan S. Willsky, "Visual Hand Tracking Using Nonparametric Belief Propagation", MIT Laboratory For Information & Decision Systems Technical Report P-2603, Presented at IEEE CVPR Workshop On Generative Model-Based Vision, Pp. 1-9, 2004
- [7] T. Grossman, R. Balakrishnan, G. Kurtenbach, G. Fitzmaurice, A. Khan, and B. Buxton, "Creating Principal 3D Curves with Digital Tape Drawing," Proc. Conf. Human Factors Computing Systems (CHI' 02), pp. 121-128, 2002.
- [8] T. A. C. Bragatto, G. I. S. Ruas, M. V. Lamar, "Real-time Video-Based Finger Spelling Recognition System Using Low Computational Complexity Artificial Neural Networks", IEEE ITS, pp. 393-397, 2006
- [9] Yusuke Araga, Makoto Shirabayashi, Keishi Kaida, Hiroomi Hikawa, "Real Time Gesture Recognition System Using Posture Classifier and Jordan Recurrent Neural Network", IEEE World Congress on Computational Intelligence, Brisbane, Australia, 2012
- [10] Ruiduo Yang, Sudeep Sarkar, "Coupled grouping and matching for sign and gesture recognition", Computer Vision and Image Understanding, Elsevier, 2008
- [11] R. Wang, S. Paris, and J. Popovic, "6D hands: markerless hand-tracking for computer-aided design," in Proc. 24th Ann. ACM Symp. User Interface Softw. Technol., 2011, pp. 549–558.
- [12] Maryam Khosravi Nahouji, "2D Finger Motion Tracking, Implementation For Android Based Smartphones", Master's Thesis, CHALMERS Applied Information Technology, 2012, pp 1-48
- [13] EshedOhn-Bar, Mohan ManubhaiTrivedi, "Hand Gesture Recognition In Real-Time For Automotive Interfaces," IEEE Transactions On Intelligent Transportation Systems, VOL. 15, NO. 6, December 2014, pp 2368-2377
- [14] P. Ramasamy, G. Prabhu, and R. Srinivasan, "An economical air writing system is converting finger movements to text using a web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1-6.
- [15] Kenji Oka, Yoichi Sato, and Hideki Koike, "Real-Time Fingertip Tracking and Gesture Recognition," IEEE Computer Graphics and Applications, 2002, pp.64-71.
- [16] H.M. Cooper, "Sign Language Recognition: Generalising to More Complex Corpora", Ph.D. Thesis, Centre for Vision, Speech and Signal Processing Faculty of Engineering and Physical Sciences, University of Surrey, UK, 2012
- [17] Y. Huang, X. Liu, X. Zhang, and L. Jin, "A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach, and Application," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, pp. 370-377, 2016
- [18] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 19, NO. 7, JULY 1997, pp.677-695
- [19] Guo-Zhen Wang, Yi-Pai Huang, Tian-Sheeran Chang, and Tsu-Han Chen, "Bare Finger 3D Air-Touch System Using an Embedded Optical Sensor Array for Mobile Displays", Journal Of Display Technology, VOL. 10, NO. 1, JANUARY 2014, pp.13-18
- [20] Napa Sae-Bae, Kowsar Ahmed, Katherine Isbister, NasirMemon, "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," Proc. SIGCHI Conference on Human Factors in Computing System, 2005, pp.977-986
- [21] W. Makela, "Working 3D Meshes and Particles with Finger Tips, towards an Immersive Artists' Interface," Proc. IEEE Virtual Reality Workshop, pp. 77-80, 2005.
- [22] A.D. Gregory, S.A. Ehmann, and M.C. Lin, "inTouch: Interactive Multiresolution Modeling and 3D Painting with a Haptic Interface," Proc. IEEE Virtual Reality (VR' 02), pp. 45-52, 2000.
- [23] W. C. Westerman, H. Lamiroux, and M. E. Dreisbach, "Swipe gestures for touch screen keyboards," Nov. 15 2011, US Patent 8,059,101
- [24] S. Vikram, L. Li, and S. Russell, "Handwriting and gestures in the air, recognizing on the fly," in Proceedings of the CHI, vol. 13, 2013, pp. 1179–1184.
- [25] X. Liu, Y. Huang, X. Zhang, and L. Jin. "Fingertip in the eye: A cascaded CNN pipeline for the real-time fingertip detection in egocentric videos," CoRR, abs/1511.02282, 2015.