Deep Learning Technique Based Character and Speech-to-Text Recognition for Interactive Online Writing with Virtual Pen and Eraser

Jayadharshini.P[a], Dr. Thangarajan.R[b] and Dr. Abirami.T [c,1]

[a]M.Tech(IT) Student, Kongu Engineering College, Perundurai
[b]Professor, Department of IT, Kongu Engineering College, Perundurai
[c]Associate Professor, Department of IT, Kongu Engineering College, Perundurai

Abstract ZeroTouch is the new emergence of sensing technologies as a touch-free interaction of air writing. The main objective of this paper is to use the pen with motion mode of interaction for the efficient online gesture. The virtual pen is detected in webcam with different contour detection is made along with eraser functionalities. But it is made much easier with OpenCV technologies. To display the text from speech in clear format a whiteboard function is added using Tkgui toolkit. A faster region-based convolution neural network technique is proposed for image text detection and then the detected image is converted to text and saved in different image formats.

Keywords: Air writing, Virtual Pen, Virtual Eraser, OpenCV, Faster RCNN, Speech-to-Text

1. Introduction

Gestures are very much important for sharing and communicating information among a huge number of people. (Moeller, Lupfer, Hamilton, Lin, & Kerne, 2011)The technologies like Human-Computer Interaction (HCI) are trending in the developing world by giving an automatic running command to the computers. Humans can communicate with machines by means of hand, facial expressions, with any external objects, voice and touch. This is also called an In-air gesture made by virtual pen and virtual eraser. (U. H. Bohari, Alli, Garcia, & Krishnamurthy, 2021)Free-air writing based on gesture is the hottest topic in development because these allow low-cost recognition of the object. It aims to use a combination of computer vision and deep learning for writing recognition to create a system that acts as a virtual board.

---

[1] Dr. T. Abirami, Department of Information Technology, KonguEngineering College, Perundurai: Email: abiananthmca@gmail.com

It uses a virtual pen and eraser for writing in the air and able to write by facing towards the web camera. (X. A. Chen, Schwarz, Harrison, Mankoff, & Hudson, 2014)The virtual pen is a computer vision program to track object movement and draw the line according to the movement. With some trial and error using OpenCV, successfully With some trial and error using OpenCV, successfully captured an image, Gaussian blurred it and applied binary masking of the targeted pen. It can be thought of as something like closed color curves on compromises that have the equivalent color, it's like an indistinct shape. (Bradski & Kaehler, 2008).

OpenCV and Computer Vision is an open-source machine learning python library used to solve many computer vision problems and improves utilization of machine learning algorithms in real image paramount importance. (Mistry, Maes, & Chang, 2009)The computer vision methodology includes process of reading the real image pixel values as well as resizing and rotating the image. This method has a high image processing power with result of more accuracy and also reduces the implementation cost.

2.   Literature Survey

Previous studies have shown that the hand-based lecture by detecting hand motion using skin tone information from the 'camshaft' tracker. This also states that the human-computer interface completely made with a hand gesture. [2] It states two different hand gesture like palm hand gesture and finger recognition using OpenCV and almost it covers complete hand-based gesture. [3] This paper proposed an efficient way of working with an in-air gesture which relaxes the touch experiences by the in-air sensing method. There are two types for selecting or drawing in a particular region that is drag or text and 'pigtail' in-air.

[4] It was reported in literature that the upper body skeleton tracking for better gesture experience and detection with a better intent prediction model. [5] have also suggested that the best result by using the intangible canvas method which provides the better gesture session by completely monitoring the visual sight, emotional monitoring and also the actions performed by the fingers by direct line of sight information. [6] It proposed the multi-touch force field using OpenCV for freely drawing during the online lecture to give efficient understanding and detection of actions performed by hand.

3.   Proposed Methodology

To initiate the exploration of Pen + Motion communication, an air-based gesture and object recognition when the user was involved in lecture-based interactive tasks. (Oudah, Al-Naji, & Chahl, 2020)A different colour of pen that is made to recognize in-camera for executing air-based gesture session. Initially, the online sessions are videotaped that are held using different colours of object and analysed the movement of the object in the space immediately on the screen. From the above-analysed set of features, the factors suitable for gestural and identified object as an input and also for avoiding collisions with natural pen and finger movements. Further, these factors can be involved in the design of pen-based gestures and how touches can be classified between natural delimiters to segment the different actions.

Participants of online gesture exhibited a different range of in-air: above-screen behaviors of pen. (Ullah, Al-Turjman, Mostarda, & Gagliardi, 2020)While interacting the contents with the screen, lecturers use their pen to point out the content or to highlight the content as well as to spoke similar to that point. In general, there are three main categories for detecting behavior of pen, 1) Path – The trajectory of Pen movement, 2) Position – Pointer of the object should point straight towards the screen, and 3) Repetition – how users repeat certain pen movement for writing and highlighting.

3.1 Pen with motion gestures

(Bose et al., 2016)The Pen + Motion gesture reveals the new way of writing and drawing interaction made between lecturer and the student especially in place of making them understand by writing. (Moeller et al., 2011)This method replaces the usage of the digital board and makes it more user-friendliness for the gestures.
This pen with motion gesture involves mainly 5 steps, such as:

1. Finding the target pen's color range:

•   The initial factor here is to find the appropriate range of colour for the target object (Pen). For this purpose, the colour detection method is followed which converts RGB formatted image into HSV formatted image because the HSV formatted image is easier to detect and manipulate colours on the screen.
•   (Gurav & Kadbe, 2015)Thus, the HSV trackbars are made available in the gesture screen which allows adjusting the lower and upper bounds from HSV channels.
•   Fig.1. By altering the range till everything becomes black leaves only the targeted object to be visible, if not acquired, then it may end up with noises, besides leads to false tracking and detection of the object.
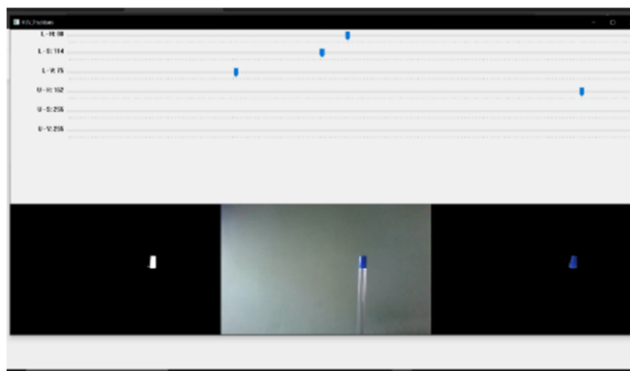

Figure 1. Identifying the target object (Pen)

2. To get rid of noise by maximizing the mask detection:

•   Fig.1. has some white spots on the screen or some noise distractions may occur. These intrusions can get reduced by doing morphological operations such as erosion and dilation.
•   Here the fixed size for kernel is 5×5. But here the kernel size varies for each target object.

- In this place, it is better to tune the value of kernel size to get an efficient result.
- By implementing erosion, small white spots can be get ridden off and by dilating them to enlarge the targeted object.
- When there occurs any intrusion other than the kernel size this can be detected in the further steps.

3. Tracking and writing with pen:

- (Chaudhary, 2017)After decent masking, it detects the pen by applying contour detection. Fig.2. a bounding box is detected around the object to ensure it's being identified all across the screen.
- After setting the above factors the target object is getting ready for tracking and made to virtually draw on the screen.
- At this stage, the x-y location is returned from the previous frame and correlate with the x and y coordinates of targeted object to see the real-time drawing virtual pen.
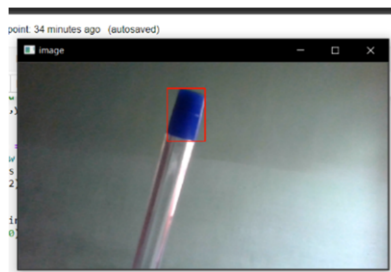

Figure 2. Pen Tracking

4. Adding automated wiping screen:

- To clear or wipe the entire screen, the letter 'C' is pressed on the keyboard or else it can be done by moving the targeted pen towards the camera, which automatically clears the screen.
- The automatic wiping is achieved by the contour i.e., Whenever the object gets close to the camera, the contour increases and if moves away then it starts with the new screen.
- Also, an alert system is proposed to alert the user that the screen will be erased/cleared within few seconds, it helps the user to take the picture or record the doings for their further use.

5. Adding eraser to the screen:

- After including the wiper function the eraser function is included with the pen.
- When the eraser function is clicked, instead of writing pen becomes an eraser that is it performs the functionality of the eraser.
- It erases the part of pen drawings and writings. Fig.3. whenever the user moves their hand on the left top corner of the virtual screen, the switching operation occurs by background subtraction.
- It observes the area for an interruption and makes the virtual button as a pen or as an eraser and restores it to the original state.
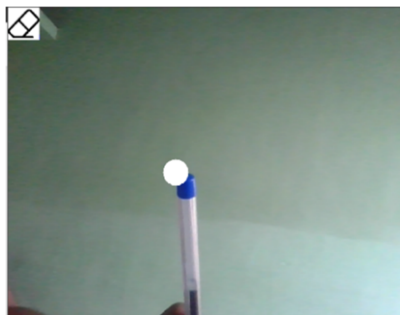
Figure 3. Identification of Eraser

3.2 Speech to text conversion

(Wang et al., 2020)Here the speech should be converted to an electrical signal with the help of a microphone, which is voice as an input from the microphone and performs converting process like audio to signal. And where deep learning applied to convert the speech into text API. (Ravanelli et al., 2020)Subsequently, API's text is understood by the deep learning automatically called the autoML NLP. It finally converts the signal into the text composition and it is displayed in the AI_board automatically. The screen can be saved as any image format into the local disks and even the voice can also be recorded for further use.
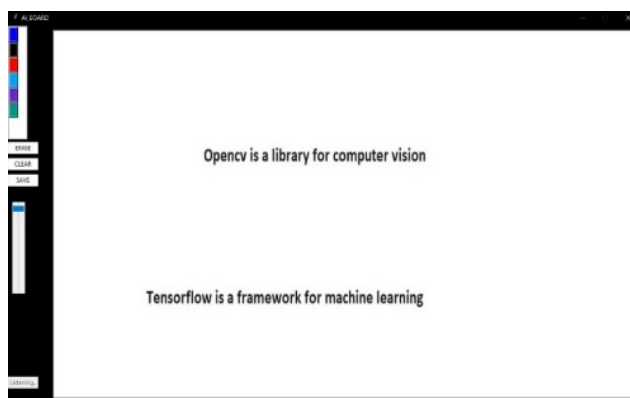

Figure 4. Speech to Text Recognition

3.3 AI – Whiteboard with Speech Recognition

A Canvas application that can recognise speech. This is an application that has all the features that a whiteboard needs. (Meng et al., 2021)Speech Recognition and including the text in the virtual screen.
Steps to use Speech Recognition:
The system requires an effective internet connection for speech recognition to run.
• At first click upon the 'listen' button on the virtual screen, next click wherever on the virtual screen.

- Now it shows 'listening'. Continuously the button seems so, it implies that it is listening.
- Speaking on the microphone, the python function called recorder automatically detects silence.
- When it detects the silence, it sends the text for recognition, and prints the text where clicked.Fig.4.
- This printed text is editable by clicking on it Fig.5. (It cannot yet be edited by voice, but by keyboard).
- If the sent voice is unrecognized, it prints an 'E' onto the canvas (This E can also be deleted with a simple backspace).
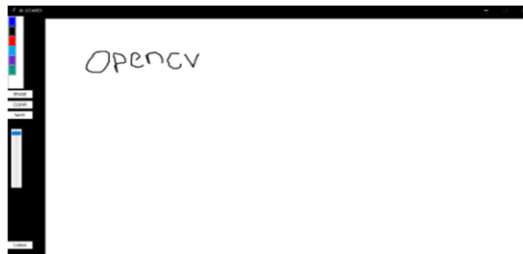

Figure 5. Whiteboard written text recognition

4.   Results And Discussions

In this project, the line segmentation process is executed in webcam for recognizing text in full paragraph mode.For this path planning algorithm is used for separating paragraph into lines for better classification of each and every text. It also reduces background noise for proper real time image detection. Decoding method is used to improve the accuracy. The accuracy is achieved up to~99% by using Dense Net model vanished gradient problem to train the dataset model quickly. The accuracy of result can be increased by Dynamic Feature Propagation by tuning the model. The blocking and adding of layers can be used to augmenting the images.

5.   Conclusion

This project demonstrates about the online gesture using pen and motion tracking method. this method has target object as a pen and the writing operation gets executed with the help of pen. thecolor range of pen is detected for writing and drawing in webcam. the eraser function is also included along with pen and wipes the letters drawn on the webcam by using python in open cv to give more accuracy level. this feature could be more interesting for the users to work in it. it also attracts the listeners with these functionalities which makesstudents to listen the class effectively.

Reference

[1]   Ahlawat, s., choudhary, a., nayyar, a., singh, s., &yoon, b. (2020). Improved handwritten digit recognition using convolutional neural networks (cnn). Sensors, 20(12), 3344.

[2]   Ahmed, b., &gulliver, t. A. (2020). Image splicing detection using mask-rcnn. Signal, image and video processing, 1-8.

[3]   Bohari, u., &chen, t.-j. (2018). To draw or not to draw: recognizing stroke-hover intent in non-instrumented gesture-free mid-air sketching. Paper presented at the 23rd international conference on intelligent user interfaces.

[4]   Bohari, u. H., alli, r., garcia, a., &krishnamurthy, v. R. (2021). Stroke-hover intent recognition for mid-air curve drawing using multi-point skeletal trajectories. Journal of computing and information science in engineering, 21(1).

[5]   Bose, v., roy, r., nadirsha, m., raj, b., ajesh, m., &jacob, p. M. (2016). Gesture based painting system. Paper presented at the 2016 international conference on information science (icis).

[6]   Bradski, g., &kaehler, a. (2008). Learning opencv: computer vision with the opencv library: " o'reilly media, inc.".

[7]   Chaudhary, a. (2017). Finger-stylus for non touch-enable systems. Journal of king saud university-engineering sciences, 29(1), 35-39.

[8]   Chen, x. A., schwarz, j., harrison, c., mankoff, j., &hudson, s. E. (2014). Air+ touch: interweaving touch & in-air gestures. Paper presented at the proceedings of the 27th annual acm symposium on user interface software and technology.

[9]   Chen, z.-h., kim, j.-t., liang, j., zhang, j., & yuan, y.-b. (2014). Real-time hand gesture recognition using finger segmentation. The scientific world journal, 2014.

[10]   Gurav, r. M., &kadbe, p. K. (2015). Real time finger tracking and contour detection for gesture recognition using opencv. Paper presented at the 2015 international conference on industrial instrumentation and control (icic).

[11]   Han, w., zhang, z., zhang, y., yu, j., chiu, c.-c., qin, j., . . . Wu, y. (2020). Contextnet: improving convolutional neural networks for automatic speech recognition with global context. Arxiv preprint arxiv:2005.03191.

[12]   Jha, g., &cecotti, h. (2020). Data augmentation for handwritten digit recognition using generative adversarial networks. Multimedia tools and applications, 79(47), 35055-35068.

[13]   Meng, z., parthasarathy, s., sun, e., gaur, y., kanda, n., lu, l., . . . Gong, y. (2021). Internal language model estimation for domain-adaptive end-to-end speech recognition. Paper presented at the 2021 ieee spoken language technology workshop (slt).

[14]   Mistry, p., maes, p., &chang, l. (2009). Wuw-wear ur world: a wearable gestural interface chi'09 extended abstracts on human factors in computing systems (pp. 4111-4116).

[15]   Moeller, j., lupfer, n., hamilton, b., lin, h., &kerne, a. (2011). Intangiblecanvas: free-air finger painting on a projected canvas chi'11 extended abstracts on human factors in computing systems (pp. 1615-1620).

[16]   Mounika, p., veeranath, p., gopal, r. D., &rao, d. (2016). Digital art drawing in the air through gesture.

[17]   Oudah, m., al-naji, a., &chahl, j. (2020). Hand gesture recognition based on computer vision: a review of techniques. Journal of imaging, 6(8), 73.

[18]   Selvakumar K, Sairamesh, L. User query-based automatic text summarization of web documents using ontology. In: International Conference on Communication, Computing and Electronics Systems; 2021. Singapore: Springer, p. 593–600

[19]   Ullah, z., al-turjman, f., mostarda, l., &gagliardi, r. (2020). Applications of artificial intelligence and machine learning in smart cities. Computer communications, 154, 313-323. Doi:https://doi.org/10.1016/j.comcom.2020.02.069