CURSOR MANIPULATION WITH HAND RECOGNITION USING COMPUTER VISION

Mandar Salvi[1], Shravan Kegade[2], Aniket Shinde[3], Prof. Bhanu Tekwani[4]
[1]Student, Vidyalankar Institute of Technology, Mumbai, India, mandar.salvi@vit.edu.in
[2]Student, Vidyalankar Institute of Technology, Mumbai, India, shravan.kegade@vit.edu.in
[3]Student, Vidyalankar Institute of Technology, Mumbai, India, aniket.shinde@vit.edu.in
[4]Professor, Vidyalankar Institute of Technology, Mumbai, India, bhanu.tekwani@vit.edu.in

ABSTRACT

This paper aims to make a software program which will Track/Monitor your hand movement in front of the screen through a webcam and will move the cursor of the computing system with respect to your hand movement and can do certain fixed tasks like Right Click, Left Click, Scroll, Drag, Switch Between Programs, Go back, Forward, etc. This program will work in background and use convolutional Neural Networks Model (SSD) to convolve each and every video frame coming from input and at the end will classify the image into classes after further processing of the predicted class it will do necessary operations on Mouse/ Trackpad driver to perform desired operations.

Key words: Computer Vision, Deep Learning, Hand Recognition, SSD Neural Networks.

1. INTRODUCTION

Interaction with computer's hand gesture is one of best media between human being and the computer system nowadays. To design a smart as well as well-organized human-computer interface is of utmost importance. Gesture data fetching is remarkable approach in gesture recognition through hand. The method used is categorized as: glove based and vision based. Glove based system mainly need the glove with sensors which can gather information of the movements and it can be more accurate than vision based but these devices are quite expensive for common people. Vision based.

gestures require only cameras to detect movement. This project is vision-based. The primary goal of this project is to make a software program which will run in background on Windows based operating system that will track your hand and fingertip and move your cursor according to your hand movement, with some pretrained gestures you can perform some operations which you generally perform with mouse and trackpad of your computer/laptop. In addition to that some gestures control operations which you can't do normally with mouse. The program will be consisting of an Elegant UI made with Electron JS which will consists of SSD – a Convolution Neural Network that will be pre-trained. It will predict the gestures from input video through the webcam of the computer and it will perform necessary operation on cursor function of the operating system.

2. LITERATURE SURVEY

[1] In the paper proposed by Womg, Shafiee, Li, & Chwyl, it was observed that real time object detection is done using Tiny SSD rather than using a full SSD. As per study it was observed that the Tiny SSD is specifically designed for real time object detection. Also, this paper provides tested examples results that show Tiny SSD can be used in our project scenario.

[2] In the paper proposed by Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi it was observed that YOLO algorithm is fast and precise object detector tool. When YOLO algorithm is attached to a webcam it was concluded that real-time performance is maintained, also the period to capture images from the camera and revealing the detections both work hand-in-hand. YOLO algorithm functions like a tracking system while processing the images and recognizing objects as they move around.

[3] In the paper proposed by Shaoqing Ren Kaiming, we observed that Region Proposal Networks (RPNs) is used to detect the accurate and efficient region. Detection network with some convolutional features helped us study that a object detection system based on deep learning can run at 5- 17 fps which turned out to be centralized. The trained RPN helps to improve proposal quality of the region and the accuracy of object detection.

3. METHODOLOGY

The algorithm used is Single Shot Multi-Box Detector (SSD) which is an architecture of Convolutional Neural Network with some libraries of computer vision like OpenCV and a Machine learning framework PyTorch. There are many architectures that we can use similar to SSD like YOLO, FPN, R-CNN but we are choosing SSD because of its accuracy as it is good as compared to others and most importantly its high-speed testing that processes 46 frames per second which is far better than others.
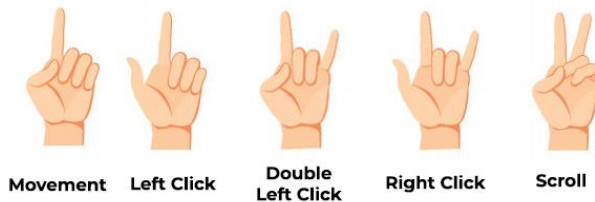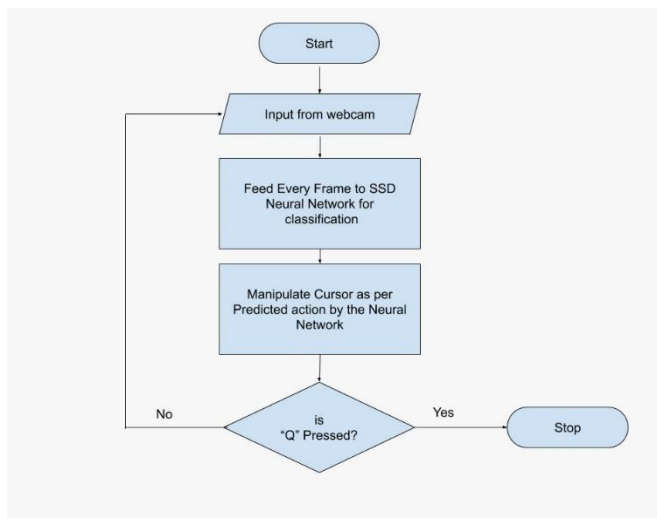

Figure 1: Gestures

Figure 2: Flowchart

Figure 3: Movement Region

For Movement of Cursor, we get bounding box location from SSD with labels, as shown in figure 1 there are 5 labels for respective gestures, if the gesture is Movement then we calculate Center of the box and map it on the regions as illustrated in figure 3, and then call the Respective function to move the cursor to that direction. The speed of movement is slow at start and increases gradually. This is done only in case of Movement Class. Other classes trigger simple clicking functions. There are 2 threads Working simultaneously. First thread Separates the video frames and feeds every frame to the SSD, which then outputs predicted class and Location of the bounding box. Then second thread performs cursor manipulation based on predicted class and mapped location on screen. These two threads communicate with each other via Shared Global Variables. PyAutoGUI library is used to perform cursor manipulation tasks after successful object detection task. Python scripts is allowed by PyAutoGUI library to handle the mouse and keyboard in order to perform automated relations with present applications.

REFERENCES

1. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C. 2016. "**SSD: Single shot multibox detector**," in European conference on computer vision (Vol. 9905 LNCS), Springer Cham, pp. 21–37 (doi: 10.1007/978-3-319-46448-0_2).

2. J. Triesch , C. von der Malsburg, "**A system for person-independent hand posture recognition against complex backgrounds**," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 12, pp. 1449–1453, 2001.

3. Chin Huan Tan J. Triesch , C. von der Malsburg, "**A system for person- independent hand posture recognition against complex backgrounds**," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 12, pp. 1449– 1453, 2001.

4. Hand Gesture Recognition Based on Single-Shot Multibox Detector Deep Learning by Peng Liu, Xiangxiang Li, Haiting Cui, Shanshan Li and Yafei Yuan. (doi: 10.1155/2019/3410348)

5. Erol, A., Bebis, G., Nicolescu, M., Boyle, R. D., and Twombly, X. 2007. "**Vision-based hand pose estimation**," Computer Vision and Image Understanding (doi: 10.1016/j.cviu.2006.10.012).