International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

MOUSE SIMULATION USING NON MAXIMUM SUPPRESSION

Mr.Ravisankar S[1], Mr.Sanjay Nanthan S[2], Ms.Shivani S[3], Mr.Sivasampath B[4] , Ms.Varsha K[5]

[1] *Assistant Professor, Department of CSE, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India*
[2,3,4,5] *Department of CSE, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India*

Abstract - *The main aim of this project is to achieve the various functions of a mouse virtually. In this, the position of the cursor can be controlled without using any electronic devices as an input. For instance, cursor may perform operations like dragging, capturing an image, zooming in and out can be performed with different hand gestures. Hand gestures is captured by using webcam and it is considered as an input device. With the help of this, we can identify the color of the hand and decide the position of the cursor accordingly. Since the environment may contain noises, lighting issues, and background merge of different objects. Therefore, it becomes imperative that the color determining works accurately. Initially, the image is captured by using the webcam and the human hand is extracted amidst the noises in the image. The position of the human hand is stored in the model using the coordinate system. The fingertip location is mapped to RGB images to control the mouse cursor based on a virtual screen. To achieve this, Single Shot Multi box Detection algorithm (SSD) along with the combination of Non- Maximum Suppression (NMS) algorithm is deployed. The motive of this work is to make the machines interact with the human environment and to verify the adaptability to the growing AI-dependent world.*

Key Words: **Virtual mouse, Single shot multibox detection, non-maximum suppression, Open CV, Media Pipe**

1. INTRODUCTION

People want compact electronic devices that enable human – computer interactions. Human computer interactions (HCI) began in the early 1980's as a field of study and practice. One of the simplest and most significant ways of human communication is through hand gestures that people tend to make even unknowingly. The main objective of this project is to setup a system that would reduce the need for major hardware components since most of it face the threat of durability and propose a system that would control the functionalities of a mouse using just hand gestures. The system is designed and implemented to perform the functions of a traditional mouse for which image or object detection plays a major role. To achieve this, Media Pipe is used which uses an algorithm called Non – Maximum Suppression algorithm which aids in detecting the hand gestures accurately to perform functionalities. This is deployed after Single Shot Multi box Detection is used to identify anchor boxes or bounding boxes for the given input images. Using Open CV, web camera is accessed, and video is recorded and converted into number of frames. Computations regarding functionalities for the gestures, after being taken as input, are done within the system itself. The main aim is to create a cost-free hand recognition software for laptops and PCs with external webcams.

1.1 CNN

Deep Learning algorithms such as Convolutional Neural Network (ConvNet/CNN) learn to assign weights and biases to various aspects/objects in an image and determine the importance of each. The pre-processing required in a ConvNet is much lower as compared to other classification. With ConvNets, images are reduced into a form that is easier to process without losing the essential features that are crucial for a good prediction. There are three layers in convolutional neural networks,

1. Convolutional layer
2. Pooling layer
3. Fully connected layer

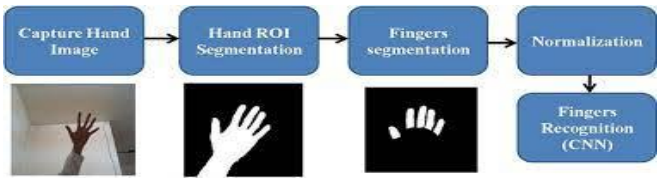

Fig. 1.1 flow chart for CNN

2. Non-Maximum Suppression (NMS)

CNN uses bounding boxes to identify the objects in any images. Bounding box is used to separate the needed object apart from the background. Now, these bounding boxes will be given for every object, in our case for every finger. When there are many bounding boxes, the program will get confused to identify the position of the finger and the gesture that is given as input. This is where NMS comes in to play. What this algorithm does is that it compares the probability of one bounding box to other so it can eliminate the least value. This process goes on until there is one bounding box left. The last standing bounding box will be for

the whole object instead of a part of the object. In our case it will for whole palm instead of each finger. This feature is applied above CNN so that the program can easily identify the hand landmarks and it helps the recognition in live image to be faster.

2.1 Single shot Multi box Detector

The Single-Shot Multibox Detector (SSD) deep algorithm is proposed to apply to the hand gesture recognition. The convolutional neural network is used as a recognition model with learning and training the selected characters end-to-end. The system test results show that the hand gesture recognition system based on the SSD model performs efficiently, reliably, quickly, and accurately.

SSD has two components: a **backbone** model and **SSD head**. *Backbone* model usually is a pre-trained image classification network as a feature extractor. Thus, a deep neural network is left that is able to extract semantic meaning from the input image while preserving the spatial structure of the image albeit at a lower resolution.

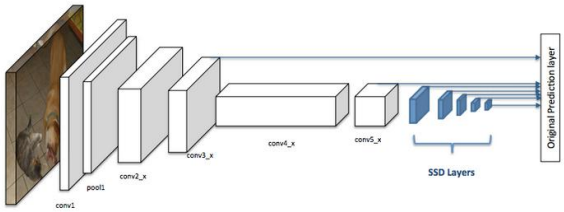

Fig. 2.1 Architecture CNN with an SSD

2.2 Grid Cell

Instead of using sliding window, SSD divides the image using a grid and have each grid cell be responsible for detecting objects in that region of the image. Detection objects simply means predicting the class and location of an object within that region. If no object is present, it is consider as the background class and the location is ignored. For instance, a 4x4 grid can be used in the example below. Each grid cell is able to output the position and shape of the object it contains.

2.3 Anchor Box

Each grid cell in SSD can be assigned with multiple anchor/prior boxes. These anchor boxes are pre-defined and each one is responsible for a size and shape within a grid cell. For example, the swimming pool in the image below corresponds to the taller anchor box while the building corresponds to the wider box.

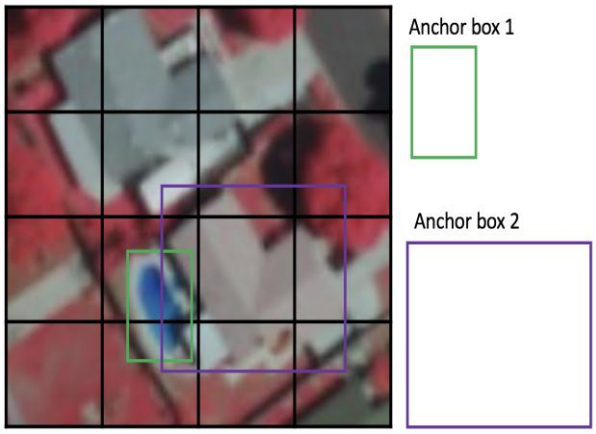

Fig. 2.2 Anchor boxes

3. SYSTEM IMPLEMENTATION

Hand Gestures are an aspect of body language that can be conveyed through finger position and shape constructed through palm. First, the hand region is detected from the original images from the input devices. Then, some kinds of features are extracted to describe hand gestures. Last, the recognition of hand gestures is accomplished by measuring the similarity of the feature data.

The modules used in the system are:

1. Hand recognition
2. Hand gesture recognition
3. Linking hand gestures with mouse operations

Hand recognition

First, instead of training a hand detector, train a palm detector since estimating bounding boxes of inflexible objects like palms and fists is much easier than recognizing hands with articulated fingers. The palm detector operates on full images and outputs an oriented bounding box. They employ a single-shot detector model is deployed. With the help of Palm detection, hand landmark coordinates are identified and thus the hand is detected.

Hand gesture recognition

Following palm detection over the real-time video capture, our next hand landmark model uses regression to accomplish exact key-point localization of 21 3D hand-knuckle coordinates inside the identified hand areas, i.e., direct coordinate prediction. Even with partially visible hands and self-occlusions, the model develops a consistent internal hand posture representation.

Linking hand gestures with mouse operations

When the gestures are done, the system should be able to match the hand landmarks with the directives for the coordinate places given in the algorithm. The code itself includes a hand landmark instruction.

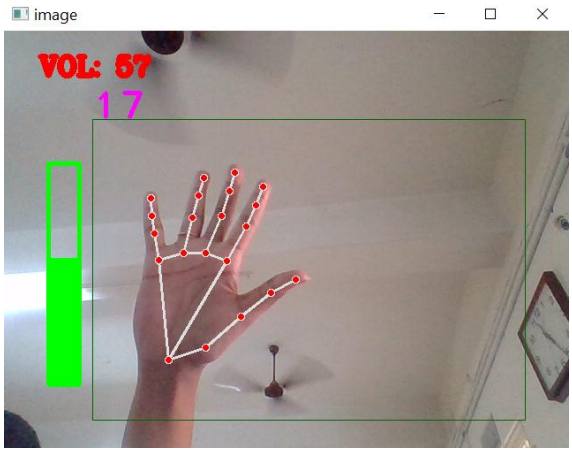

Fig. 3.1 Volume adjustment functionality

4. CONCLUSION AND SCOPE

The main purpose of this project is to reduce the hardware equipment and enable human-computer interaction. As we step into a digital world, our project would adhere to the technological advancements and have immense scope in the field of HCI.

It is used in fields like augmented reality, computer graphics, computer gaming, prosthetics and biomedical instrumentations. Digital Canvas is an extension of this project for creating 2D & 3D images using Virtual Hand Brush tools. In gaming technology, it has a major impact in Human-Computer Interaction. The major extension to work can be done to make system able to work at much complex background and compactible with different light conditions. It also implements Multi-Functional system which can perform a myriad of mouse operations using minimal resources.

REFERENCES

[1] Virtual Mouse using Hand Gestures by Riza Sande, Neha Bhegade, Akanksha Lugade, Prof. Jogdand, Volume 8, Issue 4 of International Journal of Advanced Research in Science and Technology.

[2] Virtual Mouse using Hand Gestures by Gajendra Moroliya, Sahil Patvekar, Prof. Gopnarayan, Volume 5, Issue 5 of JETIR.

[3]Mouse Control using a Web-Camera based on Color Detection by Abhik Banerjee, Abhirup Ghosh, Prof. Hemantha Saikia,Volume9, Issue 1 of International Journal of Computer Trends and Technology (IJCTT).

[4] Virtual Mouse Control using Hand Class Gestures by Vijay Kumar Sharma, Vimal Kumar, Md.Iqbal, Prof.Vishal Jayaswal, GIS Science Journal.

[5]Virtual Mouse using Hand Gestures by Abilash S, Liso Thomas, Naveen Wilson, Prof. Chaitanya, Volume 5, Issue 4 of International Research Journal of Engineering and Technology (IRJET)