IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

A Survey on Hand Gesture Recognition Techniques

Prof.Nidhi Jain[1], Nikhil Jadhav[2], Vrushabh Kakkad[3], Vaibhav Patil[4], Sanket Vetal[5]

Professor, Department of Information Technology, JSPM'S Rajarshi Shahu College of Engineering, Pune, Maharashtra, India[1]

B.E Students, Department of Information Technology, JSPM'S Rajarshi Shahu College of Engineering, Pune, Maharashtra, India [2-5]

ABSTRACT: Hand Gesture Recognition, which is a part of language technology,involves evaluating Human Hand Gestures through Machine Learning Algorithms (Supervised Learning Algorithm).Computer vision algorithms or cameras are used to convert these sign languages. With the rapid development of computer vision, the demand for interaction between human and machine is becoming more and more extensive.Recognition of human gestures comes within the more general framework of pattern recognition. This makes it one of the important applications of Graphic User Interface (GUI). Hand gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction.

KEY WORDS: Machine Learning algorithm, Supervised Learning Algorithm Computer vision, sign language, Graphic User Interface (GUI), Human Computer Interaction.

I. INTRODUCTION

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 dataHand Gesture Recognition serves as a motivating force for research in modeling and analysis. The applications include sign language recognition, augmented reality (virtual reality), sign language interpreters for the disabled,teleconferencing, surveillance and robot control.Gesture recognition, although has been exploring for many years, is still a challenging problem. Complex background, camera angles and illumination conditions make the problem more difficult.

Gesture recognition is a type of perceptual computing user interface that allows computers to capture and interpret human gestures as commands. Most consumers are familiar with the concept through Wii Fit, X-box and PlayStation games such as "Just Dance" and "Kinect Sports."

Image classification is one of the many exciting applications of convolutional neural networks. Aside from simple image classification, there are plenty of fascinating problems in computer vision, with object detection being one of the most interesting. It is commonly associated with self-driving cars where systems blend computer vision, LIDAR and other technologies to generate a multidimensional representation of the road with all its participants. Object detection is also commonly used in video surveillance, especially in crowd monitoring to prevent terrorist attacks, count people for general statistics or analyze customer experience with walking paths within shopping centers.

II. LITERATURE SURVEY

2.1 PAPER 1

AN IMPROVED HAND GESTURE RECOGNITION WITH TWO-STAGE CONVOLUTION NEURAL NETWORKS USING A HAND COLOR IMAGE AND ITS PSEUDO-DEPTH IMAGE.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

2.1.1 Abstract

Robust hand gesture recognition has been playing a significant role in the field of human-computer interaction for a long time, but it is still full of challenges due to many accept such as cluttered backgrounds and hand self-occlusion. With the help of depth information, depth-based methods have better performance, but the depth cameras are not as widely used and affordable as color cameras. Therefore, in this paper, we propose a two-stage deep convolutional neural network (CNN) architecture for accurate color-based hand gesture recognition. The first stage performs generation of pseudo depth hand images from color images and the second stage recognizes hand gesture classes using both the color image and its pseudo-depth hand image. The generation stage architecture is based on an image-to-image translation network. In the recognition stage, a two-stream CNN architecture with color image and its pseudo depth image is proposed to improve the color image-based recognition performance. We also propose two strategies in two-stream fusion: feature fusion and committee fusion. To validate our approach, we construct a new dataset called MaHG-RGBD dataset. Experiments demonstrate that our approach significantly improves the performance in RGB-only recognition for hand gestures.

2.1.2 In this paper they have Proposed

They proposed an RGB hand gesture recognition model by using a two-stage CNN architecture. In order to improve the color image recognition performance, we focused on generated pseudo depth hand as an effective step in our methods. We also designed a two-stream CNN for the second stage of the network which extracts the useful feature from the raw color and generated pseudo-depth image to obtain high classification accuracy. We evaluated the proposed method on MaHG-RGBD database, and the experimental results show that the pseudo depth image, as an important part, can improve the accuracy of gesture recognition. In the future work, we intend to find better ways to exploit advantages of pseudo-depth generation by adding a gesture preserved loss.

2.1.3 From this paper we have referred

By referring this paper, we came to know that CNN was used, we replaced it with YOLOv5 thus, reducing no. of stages. Also, instead of using Depth Camera, we used normal camera.

2.2 Paper 2

HAND GESTURE RECOGNITION BASED ON COMPUTER VISION: A REVIEW OF TECHNIQUES.

2.2.1 Abstract

Hand gestures are a form of nonverbal communication that can be used in several fields such as communication between deaf-mute people, robot control, human–computer interaction (HCI), home automation and medical applications. Research papers based on hand gestures have adopted many different techniques, including those based on instrumented sensor technology and computer vision. In other words, the hand sign can be classified under many headings, such as posture and gesture, as well as dynamic and static, or a hybrid of the two. This paper focuses on a review of the literature on hand gesture techniques and introduces their merits and limitations under different circumstances. In addition, it tabulates the performance of these methods, focusing on computer vision techniques that deal with the similarity and difference points, technique of hand segmentation used, classification algorithms and drawbacks, number and types of gestures, dataset used, detection range (distance) and type of camera used. This paper is a thorough general overview of hand gesture methods with a brief discussion of some possible applications.

2.2.2. In this paper they have Proposed

Hand gesture recognition addresses a fault in interaction systems. Controlling things by hand is more natural, easier, more flexible and cheaper, and there is no need to fix problems caused by hardware devices, since none is required. From previous sections, it was clear to need to put much effort into developing reliable and robust algorithms with the help of using a camera sensor has a certain characteristic to encounter common issues and achieve a reliable result. Each technique mentioned above, however, has its advantages and disadvantages and may perform well in some challenges while being inferior in others.

2.2.3 From this paper we have referred

They have proposed the use of sensor. Our proposal avoids use of sensor as YOLO is capable of training camera well.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

III. ALGORITHM

3.1 YOLO

3.1.1 YOLOv1

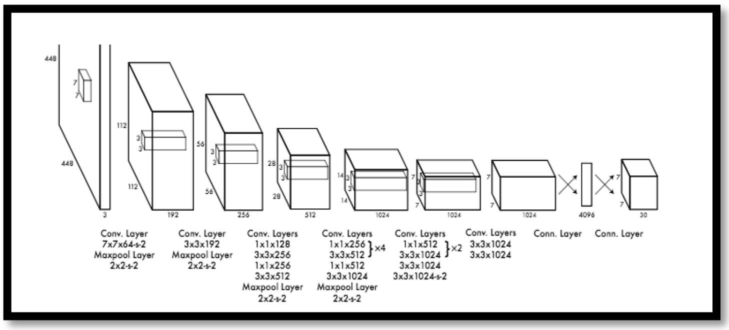

Fig.1:YOLOv1 Architecture

YOLOv1 is different from traditional method as it treats the problem of image detection as a regression problem rather than a classification problem and supports a single convolutional neural network to perform all the above-mentioned tasks. The unification of all the independent tasks into one network has the following benefits:

1. **SPEED:** YOLO is extremely fast compared to its predecessors as it uses a single convolution network to detect objects. The convolution is performed on the entire input image only once to yield the predictions.

2. **LESS BACKGROUND MISTAKES:** YOLO performs the convolution on the whole image rather than sections of it due to which it encodes contextual information about the classes and their appearances. It makes less mistakes in predicting background patches as objects as it views the entire image and reasons globally rather than locally.

3. **HIGHLYGENERALIZABLE:** YOLOlearns generalizable representations of objects due to which it can be applied to new domains and unexpected inputs without breaking.

3.1.2 YOLOv2

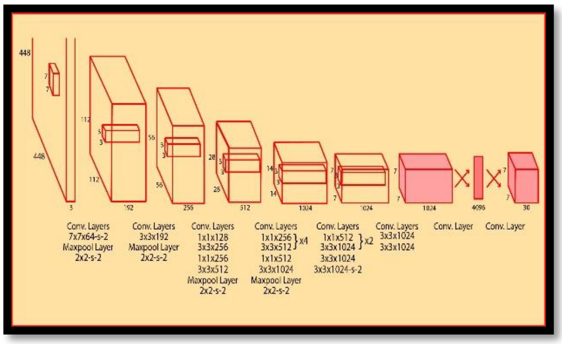

Fig. 2: YOLOv2 Architecture

The previous YOLO architecture has a lot of problems when compared to the state-of-the-art method like Fast R-CNN. It made a lot of localization errors and has a low recall.
YOLOv2 not only to improve these shortcomings of YOLO but also to maintain the speed of the architecture. There are some incremental improvements that are made in basic YOLO. Let's discuss these changes below:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

- **Batch-Normalization:**
  By adding batch normalization to the architecture the convergence of the model can be increased that leads us for faster training. This also eliminates the need for applying other types of normalization such as Dropout without overfitting. It is also observed that adding batch normalization alone can cause an increase in mAP by *2%* as compared to basic YOLO.
- **High-Resolution-Classifier:**
  The previous version of YOLO uses *224 *224* as input size during training but at the time of detection, it takes an image up to size 448*448. This causes the model to adjust to a new resolution that in turn causes a decrease in-map.
  The YOLOv2 version trains on higher resolution *(448 * 448)* for *10* epochs on ImageNet data. This gives network time to adjust the filters for higher resolution. By training on 448*448 images size the mAP increased by by *4%*.
- **Use Anchor Boxes for Bounding Boxes:**
  YOLO uses fully connected layers to predict bounding boxes instead of predicting coordinates directly from the convolution network like in Fast R-CNN, Faster R-CNN.
  In this version, removal of fully connected layer is done and instead add the anchor boxes to predict the bounding boxes.
  The fully connected layer responsible for predicting bounding boxes is removed and replacedwith anchor boxes prediction.
- **Dimensionality-clusters:**
  The number of anchors (priors) generated need to be identified so that they provide the best results. Let's take as K for now. Our task is to identify the top-K bounding boxes for images that have maximum accuracy. The K-means clustering algorithm is used for that purpose. But don't need to minimize the Euclidean distance instead, maximize the IOU as the target of this algorithm. YOLO v2 uses K=5 for the better trade-off of the algorithm.

3.1.3 YOLOv3

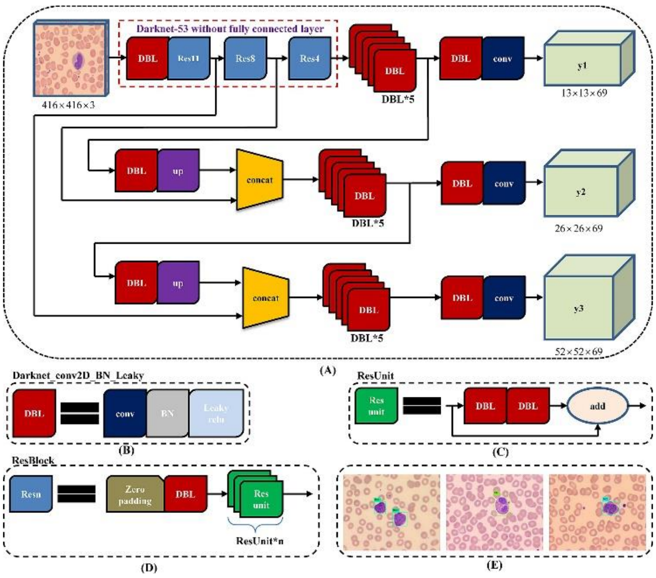

Fig.3: YOLOv3 Architecture

- For its time YOLO 9000 was the fastest, and also one of the most accurate algorithms. However, a couple of years down the line and it's no longer the most accurate with algorithms like Retina Net, and SSD outperforming it in terms of accuracy. It still, however, was one of the fastest.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

- But that speed has been traded off for boosts in accuracy in YOLO v3. While the earlier variant ran on 45 FPS on a Titan X, the current version clocks about 30 FPS. This has to do with the increase in **complexity of underlying architecture called Darknet.**

- YOLO v2 used a custom deep architecture darknet-19, an originally 19-layer network supplemented with 11 more layers for object detection. With a 30-layer architecture, YOLO v2 often struggled with small object detections. This was attributed to loss of fine-grained features as the layers down sampled the input. To remedy this, YOLO v2 used an identity mapping, concatenating feature maps from a previous layer to capture low level features.

- However, YOLO v2's architecture was still lacking some of the most important elements that are now staple in most of state-of-the art algorithms. No residual blocks, no skip connections and no up sampling. YOLO v3 incorporates all of these.

- First, YOLO v3 uses a variant of Darknet, which originally has 53-layer network trained on ImageNet. For the task of detection, 53 more layers are stacked onto it, giving us a **106 layer fully convolutional underlying architecture for YOLO v3**. This is the reason behind the slowness of YOLO v3 compared to YOLO v2. Here is how the architecture of YOLO now looks like.

3.1.4 YOLOv4

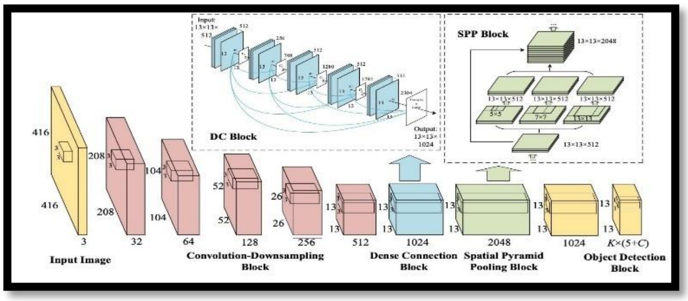

Fig.4: YOLOv4 Architecture

- YoloV4 is an important improvement of YoloV3, the implementation of a new architecture in the **Backbone** and the modifications in the **Neck** have improved the **mAP** (mean Average Precision) by **10%** and the number of **FPS** (Frame per Second) by **12%**. In addition, it has become easier to train this neural network on a single GPU.

- The Realtime object detection space remains hot and moves ever forward with the publication of YOLO v4. Relative to inference speed, YOLOv4 outperforms other object detection models by a significant margin.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

3.1.5 YOLOv5

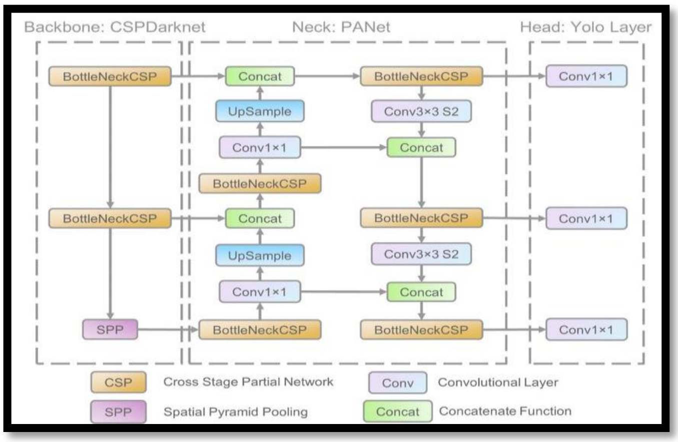

Fig. 5: YOLOv5 Architecture

- **YOLOv5** was released by a company called Ultralytics in 2020. It was published in a GitHub repository by Glenn Jocher, Founder & CEO at Ultralytics, and quickly gained traction soon after its publishing.
- The YOLO network consists of three main pieces.

1) **Backbone** - A convolutional neural network that aggregates and forms image features at different granularities.

2) **Neck** - A series of layers to mix and combine image features to pass them forward to prediction.

3) **Head** - Consumes features from the neck and takes box and class prediction steps.

4. OPENCV - LIBRARY

- OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

- OpenCV Library Modules
  - Core Functionality
  - Image Processing
  - Video
  - Video I/O
  - calib3d
  - features2d
  - Objdetect
  - Highgui

IV. ADVANTAGES

1) Functioning:

1.1) The output of the sign language will be displayed in the text form in real time.
1.2) The communication with hearing and speechimpaired people become easier.
1.3) The images captured from a continuous video stream of a webcam are passed in a yolo model and the output is text. Thus, this feature of the system makes communication very simple and delay free.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

2) No external hardware required:

2.1) As traditional approach is using sensors or gloves, the implemented system suggests the direct use of images from a web camera for recognition which reduces the cost.

2.2) The only requirement here is a webcam, a computer (for dedicated application) or a server (for a web application which is to be deployed on a server).

3) Portable:

3.1) If the Entire project is implemented on a Raspberry Pi computer, the entire system becomes portable and can be taken anywhere.

3.2) Deploying the project on a web application will increase its access without having to setup manually.

3.3) Deploying this project as a dedicated Desktop application will enable the model to use a dedicated GPU on it which in turn will increase its performance.

4) Does not get damaged through use:

4.1) As no specific hardware is required here, and the project is just a software, there is no degrade in performance with respect to time.

4.2) The only variable factors for the proposed system are that it depends on how powerful the computer it is deployed on is.

V. DISADVANTAGES

1. The proposed system uses only cameras for input purpose, so illumination of the room shadows, background can hinder the output.
2. Palm must be facing towards the camera for accurate results.
3. The performance of the system also depends on the type of hardware used. Higher end systems can yield higher accuracy as well as high frames per second.

VI. CONCLUSIONS

The traditional approach seems to be quite expensive and a complex hardware system is required. This also demands high maintenance. On the other hand, the proposed system focuses on recognizing gestures through the camera, thus reducing hardware requirement, complexity and is cost effective. Both techniques mentioned above, however, have their advantages and disadvantages and may perform well in some challenges while being inferior in others.

VII. ACKNOWLEDGEMENT

The authors would like to thank the project guide Prof. Nidhi Ma'am for this opportunity, valuable support and encouragement.

REFERENCES

1. https://en.wikipedia.org/wiki/Gesture_recognition
2. Gesture Recognition Developments- Vrushabh Jayesh Kakkad, Volume:07 Issue: 12, Dec 2020, e-ISSN: 2395-0056, p-ISSN: 2395-0072.
3. https://www.irjet.net/archives/V7/i12/IRJET-V7I12365.pdf
4. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques Munir Oudah 1, Ali Al-Naji 1,2, * and Javaan Chahl 2, published: 23 July 2020
5. https://doi.org/10.3390/jimaging6080073
6. https://www.ijser.org/paper/Sign-Language-Recognition-System.html#:~:text=The%20output%20of%20the%20sign,displayed%20at%20the%20same%20time.
7. https://github.com/ultralytics/yolov5
8. https://www.hindawi.com/journals/tswj/2014/267872/

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905137|

9. Research on Hand Gesture Recognition based on Deep Learning.
Jing-Hao Sun; Ting-Ting Ji; Shu-Bin Zhang, Guang-Rong Ji
https://ieeexplore.ieee.org/document/8634348
10. https://medium.com/adventures-with-deep-learning/yolo-v1-part-1-cfb47135f81f
YOLOv1: Part 1, Divakar Kapil
11. https://www.geeksforgeeks.org/yolo-v2-object-detection/
12. https://towardsdatascience.com/yolo-v3-objectdetection53fb7d3bfe6b#:~:text=For%20an%20input%20image%20of,boxes%20at%203%20different%20scales
13. https://s3-eu-west-1.amazonaws.com/ppreviews-plos-725668748/15596051/preview.jpg
14. https://www.google.com/url?sa=i&url=https%3A%2F%2Fmedium.com%2Fvisionwizard%2Fyolov4-version-3 proposedworkflowe4fa175b902&psig=AOvVaw2yuPrMkeO7F5stOlWTTThy&ust=1621854307399000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCPiZqqHU3_ACFQAAAAAdAAAAABAD
15. https://becominghuman.ai/explaining-yolov4-a-one-stage-detector-cdac0826cbd7
16. https://www.researchgate.net/publication/349299852/figure/fig1/AS:991096905678849@1613307324973/The-network-architecture-of-Yolov5-It-consists-of-three-parts-1-Backbone-CSPDarknet.jpg
17. https://viso.ai/deep-learning/yolov5-controversy/#:~:text=YOLOv5%20was%20released%20by%20a,traction%20soon%20after%20it's%20publishing.
18. https://blog.roboflow.com/yolov5-improvements-and-evaluation/
19. https://www.tutorialspoint.com/opencv/opencv_overview.htm#:~:text=OpenCV%20is%20a%20cross%2Dplatform,face%20detection%20and%20object%20detection.
20. Hand Gesture Recognition using YOLOv5
Nikhil Jadhav, Vrushabh Kakkad, Vaibhav Patil, Sanket Vetal https://www.irjet.net/archives/V8/i5/IRJET-V8I5464.pdf
21. An Improved Hand Gesture Recognition with Two-Stage Convolution Neural Networks using a Hand Colour Image and Its Pseudo-Depth Image.
Jiaqing Liu; Kotaro Furusawa; Tomoko Tateyama; Yutaro Iwamoto; Yen-Wei Chen
https://ieeexplore.ieee.org/document/8802970
22. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques.
Munir Oudah,Ali Al-Naji, Javaan Chahl
https://www.mdpi.com/2313-433X/6/8/73
23. Hand Gesture Recognition: A Literature Review
Rafiqul Zaman Khan, Noor Ibraheem
https://www.researchgate.net/publication/284626785_Hand_Gesture_Recognition_A_Literature_Review
24. Hand Gesture Recognition with Skin Detection and Deep Learning Method: Hanwen Huang, Yanwen Chong, CongchongNie, Shaoming Pan
https://iopscience.iop.org/article/10.1088/1742-6596/1213/2/022001/pdf
25. Hand posture recognition with the fuzzy glove T. Allevard, L. Foulloy
https://www.sciencedirect.com/topics/computer-science/gesture-recognition
26. https://appsilon.com/object-detection-yolo-algorithm/
27. Recent methods in vision-based hand gesture recognition, Haitham Badihttps://link.springer.com/article/10.1007/s41060-016-0008-z
28. https://www.marxentlabs.com/what-is-gesture-recognitiondefined/#:~:text=Gesture%20recognition%20is%20a%20type,commands%20based%20on%20those%20gestures.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.488

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH
IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com