International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072

TOUCHLESS OBJECT INTERACTION SYSTEM USING GAZE DETECTION AND GESTURE RECOGNITION

Suriya·K[1], Anusha.R[2], Devatha.N[3], Hareeni.D[4]

[1] *Assistant Professor, Department of Electronics and Communication Engineering, Easwari Engineering College, Ramapuram. Chennai – 600 089, Tamil Nadu, India.*
[2,3,4] *UG Scholar, Department of Electronics and Communication Engineering, Easwari Engineering College, Ramapuram. Chennai – 600 089, Tamil Nadu, India.*

Abstract: Internet of Things (IoT) as known is the system of collecting and exchanging data over a wide area of networks. A novel interaction system called "Touchless Object Interaction System using Gaze Detection and Gesture Recognition" is introduced, where the user especially the visually impaired will be able to utilize and control a target device by looking at the target device and operating the same through simple hand gestures. The proposed system consists of modules that involve 1) Detection of the object, 2) Recognition and estimation of direction of the head,3) Recognition of hand gestures, 4) Voiced intimation of the target device or object. The target device is identified by various techniques which majorly involves machine learning algorithm. The direction of the head is detected focusing on the methods of Roll, pitch, and YAW. Recognition of the hand gestures pertains to controlling the target device which is then fed to the IoT platform. Here, the accuracy of the modules depends on the user interaction and the corresponding module results.

Keywords: CNN module, IoT, Spyder IDE, Jupyter notebook, Deep-learning, Gesture sensor, proximity-based sensors.

I. Introduction

Touchless interaction plays an important role in the automation industry. The first developed system was the home automation with speech recognition with IoT enabled devices. The next was developed with future advancements in it. The gesture recognition also made its entry to the world of IoT and touchless interaction. Speech recognition is a technique where the person says the command that needs to be done, like switching on or off the device. This command is sent to the target device using wireless modules like Bluetooth [2], Zigbee [3], WIFI [4], etc. Finally, the device performs the operations that the user says. This can be used to control the target device and also plays a major role in various industries. Apart from controlling objects, speech to text helps in mitigating risks, saves time to write, and helps in doing the work efficiently. Gesture recognition is another technique where people use hand gestures to control any device. The basic idea behind this is, the gestures are captured using some sort of cameras or controllers and are processed using 3D algorithms. After decoding the type of gesture, the device does its operation based on it.

Y. Zou et al [1] proposed a novel device-free gesture recognition system based on phase information output by COTS RFID devices. It was proved more accurate than most other similar systems. N. Alioua [5] published a paper on appearance-based discrete head pose estimation aiming to determine the driver's attention level from a visible spectrum image. Similarly, M. Patacchiola [6] proposed a convolutional neural network-based head posed estimation with an adaptive gradient method. Here head movement direction is detected and is used along with gesture control to activate and control the target device. CNN is a technique where the input image passes through several convolution layers and is finally integrated to classify an object. Since this uses pixels and filters, all the minute details can be extracted. This can also be used by people with speaking disabilities since only gestures are needed to control. This entire system completes only in the presence of the Internet of Things. It is a system were exchanges, collection, and control of data information is done over a wide area of a network.

To have an active interaction between the target and the user, the IoT platform must consist of 1) sensors2) data connectivity 3) data processing 4) user-interface. The working phenomenon of an IoT device can also be divided as 1) collect data 2) transfer data 3) analyze data and action. Even though IoT is the main platform used to communicate between devices, another key technique used here is deep learning. With CNN and R-CNN all the data is obtained, processed, and finally communicated through the internet with the other device.

II. Existing and Proposed System

- **Existing System**

1) Detection of the object, 2) Recognition and estimation of direction of the head,3) Recognition of hand gestures, 4) Voiced intimation of the target device or object are used individually for different purposes. Gaze estimation is based on psychology research, majorly for people with motor disabilities. The core purpose is to identify a picture or a photo or an object following the movement of the eye. In the same way, object detection is focused on identifying the object based on the movement and direction of the user's head. This concept makes use of the concept of CNN. Hand gesture recognition is also a touchless feature that involves human-computer interaction using the background subtraction method using proximity-based sensors. All these features are successfully implemented individually with low accuracy. The future system requires accuracy and more feasibility of usage which involves not just one but many usages which will be beneficial at a time. The existing system involves the usage of both YOLO and CNN algorithms. The proposed system is modified to use only the CNN algorithm with better accuracy as an added advantage.

- **Proposed System**

The proposed prototype is based on CNN where all the features which are Detection of the object, Recognition, and estimation of direction of the head, Recognition of hand gestures, Voiced intimation of the target device or object are implemented under one roof which is more beneficial in the case of touchless operation and to people with motor disabilities. The voice output from the system helps even blind people to control the objects from wherever they are. Python is replaced by JAVA for efficient integration, ease of working on different platforms, uncomplicated syntax, and lesser lines of programming.

III. Block Diagram

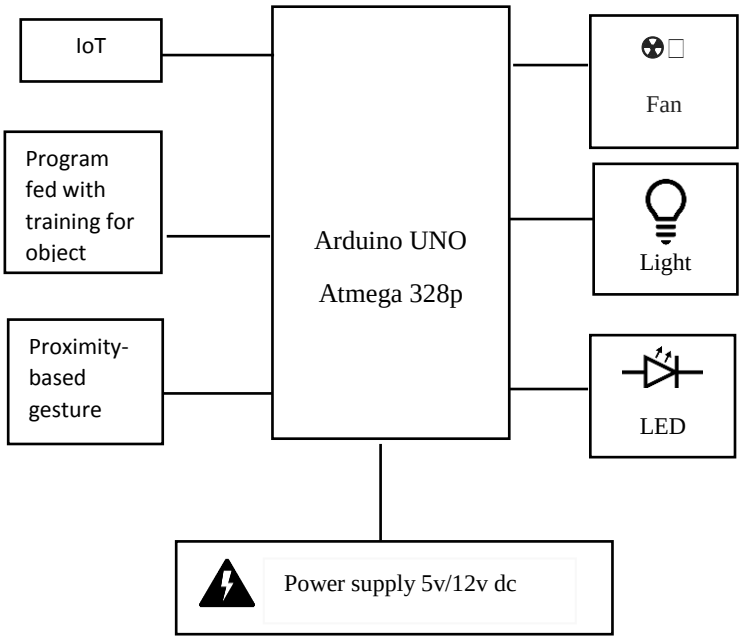


Fig 1: Implementation Overview of Touchless Object Interaction System using Gaze Detection and Gesture Recognition

Implementation Overview of Touchless Object Interaction System using Gaze Detection and Gesture Recognition shows that the objects are first recognized by a camera in the PC and the target object is identified by changing the person's head positions which activates the device through the IoT module connected to the Arduino. The identified object is priorly trained to attain its advantage of identification with ease. The object being targeted is displayed and this is particularly useful for the visually impaired as an intimation of what object is identified is conveyed to the user through a voice. The gesture sensor which is a proximity-based

sensor sense the user's hand gestures and controls the object through which the target object can be operated. In this case, a fan, LED, and light are used for experimenting purposes.

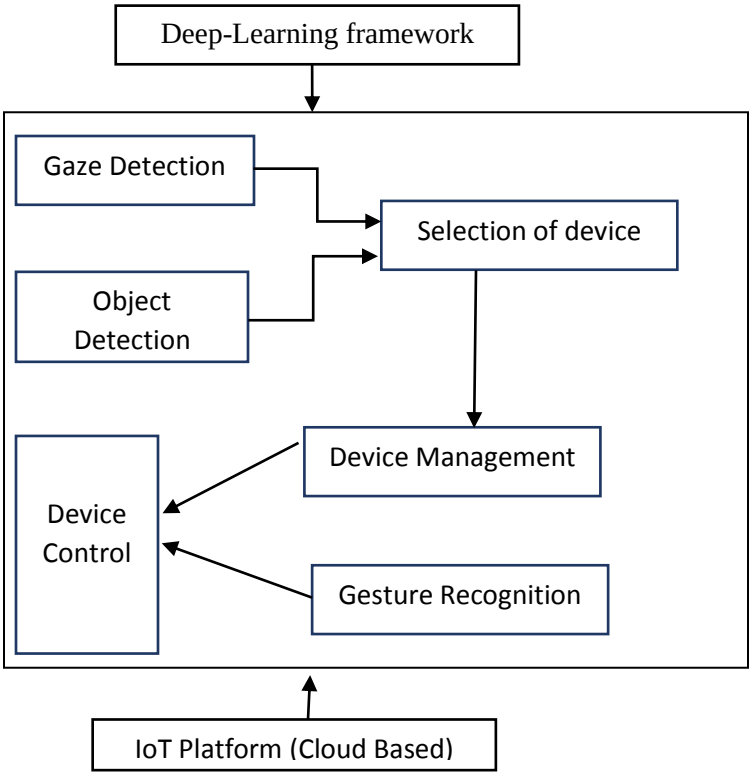
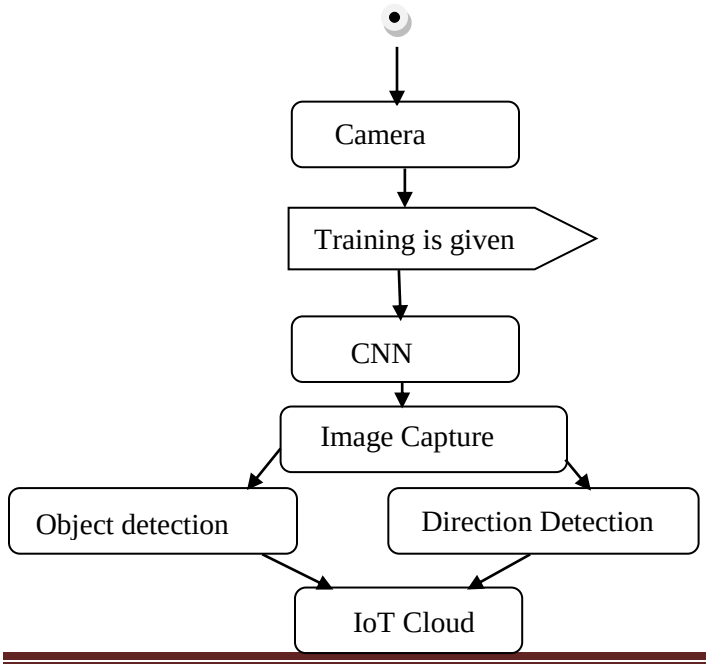


Fig 2: Architecture Diagram



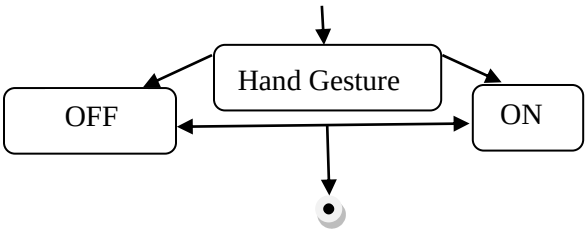


Fig 3: Activity Flow Chart

The above flow diagram represents the activity flow chart which involves the entire process of the three modules incoherence. The camera is set and the modules which are trained are uploaded to the program. Further, the input image is captured and identified. The direction of the head is detected and the object towards which the head is directed is identified and intimated to the user through a voice. This object is operated using hand gestures.

IV. Overview of Modules:

At first, the object is activated based on the head movement and that object is operated based on the hand gesture. The object is identified using features extracted priorly from the prefeed input images. The head movement detection makes use of Roll, Pitch, and YAW. This proposed system majorly involves the usage of the YAW method. The direction of the head determines the location of the object to be controlled and this object is operated using the movement of the hand at particular directions. All these features are fed under one roof using the Internet of Things. IoT architecture has different layers, in which the network layer is responsible for transmitting data between different applications using heterogeneous and communication network protocols.

Module Description

1. Detection of the object

Fig 4: Identification of Input Image

The first step focuses on checking out the varied objects present around the person. Here, the machine is trained to classify those objects by feeding them pictures of these objects as input and feeding an equivalent with certain features and labels so that it can predict correctly with what it's learned. As a result, the machine learning model is trained and most of the objects are identified using this module. In the above case given as an example, the object identified may be a person. The equivalent is often implemented with various other objects like mobile, fan, light, table, and so on. The results were found to be 99% accurate.

2. Recognition and Estimation of the direction of the head

Once the objects are identified, the next step involves selecting the objects that require control. The gaze of the user is analyzed using deep learning algorithms: CNN and RCNN. The direction where the head points to is identified (here it is left, right, top, bottom). The frame is divided into grid forms so that the angle at which the head is tilted is detected accurately. From the process, the object to be controlled can be focused on. With the assistance of principal axes called yaw, pitch and roll the head pose is identified. In the example, the person is looking towards the right, which implies that the object located there is a table fan and intimating that the fan is activated. Similarly, various objects are targeted.

3. Recognition of Hand gestures

Device	Gesture	Meaning
Fan	Forward Backward Left Right	Turn on Turn off Speed up Speed down
Light	Forward Backward	Turn on Turn off

Table 1: Example, of device types and their corresponding gestures

Once the object is identified with the direction of the head, it has to be controlled to be operated. With simple hand gestures, mentioned as examples in the above table are used to turn on or turn off the identified object. The data collected is transmitted to the device through an IoT and the response is obtained with 99% accuracy.

Components Used

1) Arduino UNO: Atmega328P based microcontroller
2) Proximity-based gesture sensor
3) IoT Module (ESP288)
4) Fan
5) Light
6) LED
7) Laptop with a webcam (Processor I3 and above)
8) The relay circuit board and connecting wires

Software Used

- Coding language: Python and Anaconda
- Operating System of PC: Windows (64 bit)
- Anaconda is installed with Spyder IDE which is an advanced development tool with inbuilt packages for developing python applications.
- Jupiter Notebook.

V. Results and Discussion

The surrounding objects are all recognized and displayed in the PC using deep learning methods. All three target objects are activated by positioning the head towards it and voice output is generated to intimate the object identified and this target object is controlled using different gestures. The implementation resulted in touchless interaction between humans and the object with better efficiency. Since the activation is only through head pose calculations through the R- CNN algorithm, there is no requirement of a Smartphone, Smartwatch, or any iris trackers which will involve the addition of further complications.

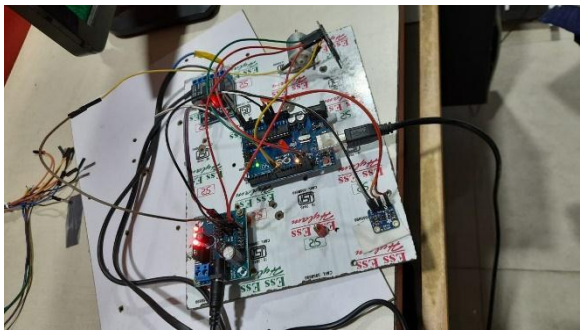


Fig 5: Hardware Output Connections

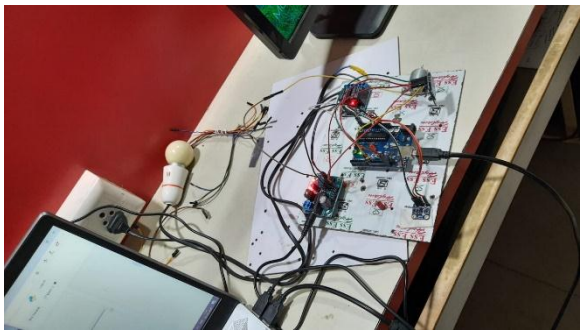


Fig 6: Hardware Output Overview

Future Advancements:

- Multiple cameras can be used to capture the movements of the user and deeper object detection techniques like Single shot multibox algorithm or Fast CNN can be implemented.

- The future for touchless operation techniques is in high demand. The head detection module plays a major role in the advancements of touchless automation in aviation, hotels, and many other related areas in various parts of the World.

VI. Conclusion

The system helps all people especially the visually impaired to locate, identify, and operate the particular target object without touching it. Amidst the current situation, a touchless operating system is of high use and also plays a major role in preventing the spread of the virus.

References

[1] Y. Zou et al., "GRfid: A device-free RFID-based gesture recognition system," IEEE Trans. Mobile Comput., vol. 16, no. 2, pp. 381–393, 2017.

[2] K. Wu et al., "HJam: Attachment transmission in WLANs," in Proc. IEEE INFOCOM, Orlando, FL, USA, 2012, pp. 1449–1457.

[3] W. Kluge et al., "A fully integrated 2.4-GHz IEEE 802.15.4-compliant transceiver for Zigbee applications," IEEE J. Solid-State Circuits, vol. 41, no. 12, pp. 2767–2775, Dec. 2006

[4] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: Recognizing human behavior with commodity

[5] N. Alioua, A. Amine, A. Rogozin, A. Bensrhair, and M. Rziza, "Driver's head pose estimation using efficient descriptor fusion," EURASIP J. Image Video Process., vol. 2016, no. 1, Dec. 2016, Art. no 2

[6] M. Patacchiola and A. Cangelosi, "Head pose estimation in the wild using convolution neural networks and adaptive gradient methods, "Pattern Recognit., vol. 71, pp. 132–143, Nov. 2017.

[7]. Erden and A. E. Çetin, "Hand gesture-based remote-control system using infrared sensors and a camera," IEEE Trans. Consum. Electron., vol. 60, no. 4, pp. 675–680, Nov. 2014.

[8][11] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[9]J. Liu, K. Furusawa, T. Tateyama, Y. Iwamoto and Y. Chen, "An Improved Hand Gesture Recognition with Two-Stage Convolution Neural Networks Using a Hand Color Image and its Pseudo-Depth Image," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 375-379, DOI: 10.1109/ICIP.2019.8802970.

[10] D. Lian et al., "Multiview Multitask Gaze Estimation with Deep Convolutional Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 10, pp. 3010-3023, Oct. 2019, DOI: 10.1109/TNNLS.2018.2865525.