Target Identification by Various Deep Learning-based Gaze Estimation and Object Detection

Ponramya.J[1], Nivedha.J[2], Jackulin Sam Gini.A[3]

[1,2]UG Student, Dept. of Information Technology, Jeppiaar SRR Engineering College, Tamil Nadu, India
[3]Assitant professor, Dept. of Information Technology, Jeppiaar SRR Engineering College, Tamil Nadu, India

---

Abstract The Internet of Things (IOT) attempts to help people access internet-connected devices, applications, and services anytime and anywhere. However, how providing an efficient and intuitive method of interaction between people and IOT devices is still an open challenge. In this work, we propose a novel interaction system where users can control an IOT device by gazing at it and doing simple gestures. The proposed system mainly consists of four categories such as object detection module, gaze estimation module. Hand The Internet of Things (IOT) attempts to help people access internet-connected devices, applications, and services anytime and anywhere. However, how providing an efficient and intuitive method of interaction between people and IOT devices is still an gesture recognition module and IOT controller module. The target device is identified by Region Based Convolution Neural Network & CNN Algorithm based gaze estimation and object detection techniques. Afterwards, hand gesture recognition is applied to generate an IOT device control command which is transmitted to the IOT platform. The experimental results and case studies demonstrate the feasibility of the proposed system and imply the future research directions

Key Words: IOT, IOT devices, object detection, target device, CNN algorithm, RCNN algorithm,

1. INTRODUCTION

Open challenge. In this work, we propose a novel interaction system where users can control an IOT device by gazing at it and doing simple gestures. The proposed system mainly consists of four categories such as object detection module, gaze estimation module. Hand gesture recognition module and IOT controller module. The target device is identified by Region Based Convolution Neural Network & CNN Algorithm based gaze estimation and object detection techniques. Afterwards, hand gesture recognition is applied to generate an IOT device control command which is transmitted to the IOT platform. The experimental results and case studies demonstrate the feasibility of the proposed system and imply the future research directions.

2. RELATED WORK

Charith Perera, Arkady Zaslavsky [1] Context Aware Computing for the Internet of Things: A Survey. As we are moving towards the Internet of Things (IOT), the number of sensors deployed around the world is growing at a rapid pace. Market research has shown a significant growth of sensor deployments over the past decade and has predicted a significant increment of the growth rate in the future. These sensors continuously generate enormous amounts of data. However, in order to add value to raw sensor data we need to understand it. Collection, modelling, reasoning, and distribution of context in relation to sensor data plays critical role in this challenge.

Gururaj Kulkarni, S. H Manoor [2] Enabling Technologies, Protocols, and Applications: A Detailed Survey on IOT. The Internet of things is used as a parasol catch word for combining and covering the major aspects related to the extension of the Internet and Web into the phenomenon, by means of vast positioning of spatially distributed devices that contains embedded identification, sensing and/or actuation capabilities. Internet of Things (IOT) consists of a large number of connected objects that are communicating with each other.

Ala Al-Fuqaha, Mohsen Guizani [3] Internet of Things: A Survey on Enabling Technologies, Protocols and Applications. This paper provides an overview of the Internet of Things (IOT) with emphasis on enabling technologies, protocols and application issues. The IOT is enabled by the latest developments in RFID, smart sensors, communication technologies and Internet protocols. The basic premise is to have smart sensors collaborate directly without human involvement to deliver a new class of applications. The current revolution in Internet, mobile and machine-to-machine (M2M) technologies can be seen as the first phase of the IOT. In the coming years, the IOT is expected to bridge diverse technologies to enable new applications by connecting physical objects together in support of intelligent decision making.

Luigi Atzori, Antonio Iera [4] The Internet of Things: A survey. This paper addresses the Internet of Things. Main enabling factor of this promising paradigm is the integration of several technologies and communications solutions. Identification and tracking technologies, wired and wireless sensor and actuator networks, enhanced communication protocols (shared with the Next Generation Internet), and distributed intelligence for smart objects are just the most relevant. As one can easily imagine, any serious contribution to the advance of the Internet of Things must necessarily be the result of synergetic activities conducted in different fields of knowledge, such as telecommunications, informatics, electronics and social science.

3. PROPOSED SYSTEM

The ultimate goal of the proposed system is to develop an efficient interaction method between users and IOT devices so that any user including patients, children, and the elderly can control the devices intuitively. The object detection module records the opposite side of the user to detect and recognize the types of IOT devices installed in the room. Then, the Gaze estimation module detects the user's head region and then computes a fine-grained head pose information (i.e., pitch, yaw, and roll) to estimate the user's gaze position. With this information, the proposed system can identify the target device. Later, Hand gesture module perform Gesture recognition using a commercial gesture sensor module. A combination of hand gesture information and the type of selected IOT device is then translated into an IOT command and transmitted to IOT platforms. The target device is identified by Region Based Convolution Neural Network & CNN Algorithms which are used for gaze estimation and object detection techniques. This algorithm gives more accurate real time output .It produces output in the form of voice command. Blind people uses voice command to operate the target object.

4. METHODOLOGY

The System architecture is designed with three modules such as object detection module, gaze estimation module. Hand gesture recognition module and IOT controller module. The target device is identified by Region Based Convolution Neural Network & CNN Algorithm based gaze estimation and object detection techniques. Afterwards, hand gesture recognition is applied to generate an IOT device control command which is transmitted to the IOT platform.

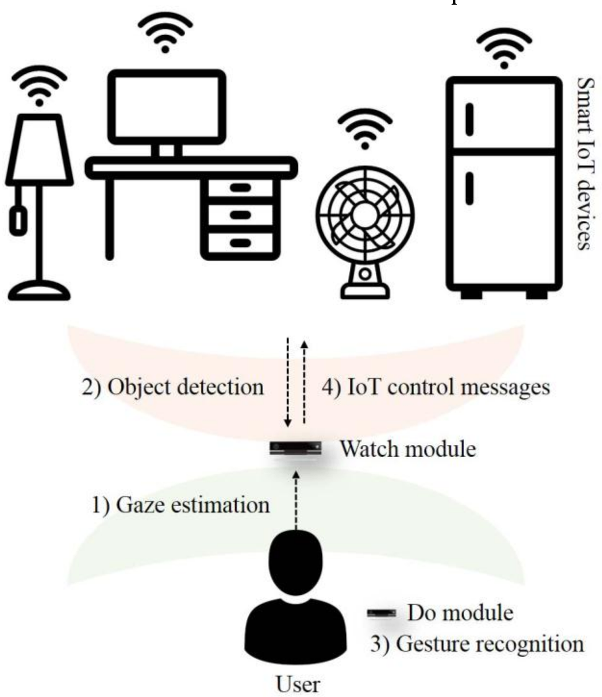


Fig-1: Working architecture

4.1 Working of Object detection module

The object detection module also records the opposite side of the user to detect the IOT devices installed in the room and recognizes their types. For object detection and recognition, recent deep learning approaches have shown excellent performance in terms of accuracy. The Deep Learning framework divides each input image into S x S grids and each grid predicts N bounding boxes and their confidence scores which indicate whether the bounding box contains an object or not

4.2 Working Gaze estimation module

In the Gaze estimation module, we use the orientation of the face (i.e., pitch, yaw, and roll) and the distance between the Watch module and the user as head pose information for training a gaze estimator. A deep learning-based approach called Deep Gaze was employed to compute the value of face orientation. In this gaze estimation the Deep Gaze uses a convolutional neural network (CNN) to output the values of face orientation from the given face image.

Fig-2: examples images for training the gaze estimator

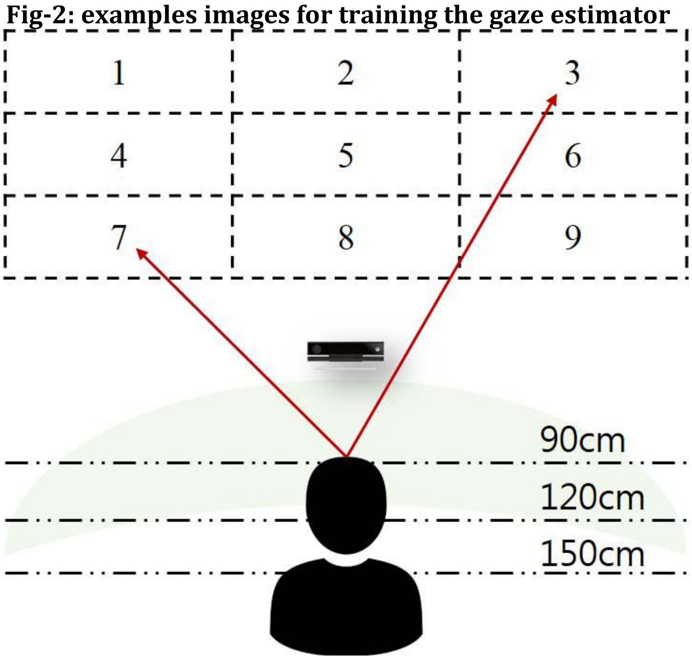


Fig-3: Overview of indoor gaze estimation by head pose estimation

4.3 Working of Hand gesture recognition module

Gesture recognition is performed by a commercial gesture sensor module. This sensor supports the following nine hand gestures: "Up", "Down", "Left", "Right", "Forward", "Backward ","Clock wise " , Counter clockwise", and "Wave". User's hand gestures are associated to different actions according to the type of the selected IOT device. User's hand gestures are associated to different actions according to the type of the selected IOT device.

Supported hand gestures. Each image represents "right", "left", "backward", and "forward" gesture from left to right, respectively.

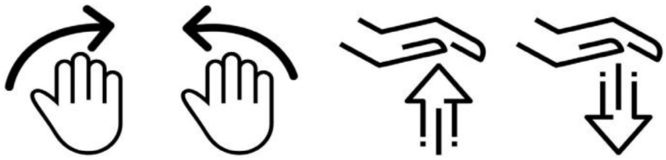


Fig-4: Hand gesture controls diagram

4.3 RCNN Algorithm

R-CNN was first introduced by Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik in 2014. Region based Convolutional Network(R-CNN) combines two ideas: (1) one can apply high-capacity Convolutional Networks (CNN'S) to bottom-up region proposals in order to localize and segment objects and (2) When labeling data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific-fine-tuning, boosts performance significantly.

In this chapter we will talk about the Region based methods for object detection. In particular, R-CNN (Regional CNN), the original application of CNNs to this problem, along with its descendants Fast R-CNN, and Faster R-CNN. In classification, there's generally an image with a single object as the focus and the task is to say what that image is. But when we look at the world around us, we carry out far more complex tasks. We see complicated sights with multiple overlapping objects, and different backgrounds and we not only classify these different objects but also identify their boundaries, differences, and relations to one another. The goal of a Region based CNN is to take an image input to detected and localize the object in the image. All three above mentioned architectures were used for object detection and localization for a driving scenario to detect cars, pedestrians and traffic signs etc. and performed a time study to conclude which of the three architectures can be used as a detector for a real-world driving scenario for detection of objects present on road. We will also discuss their architecture and describe how they perform the task of object detection and localization given the input image.

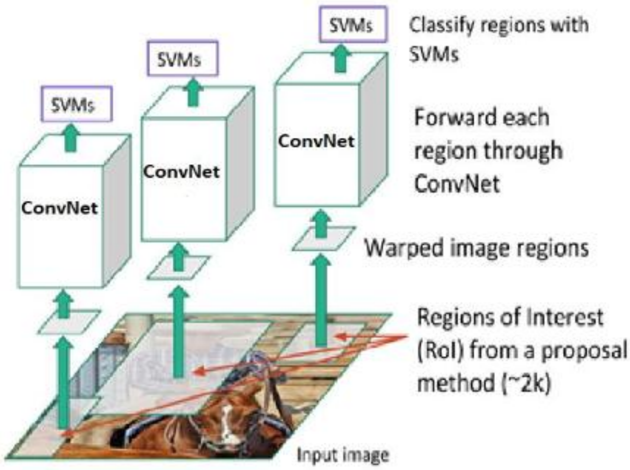
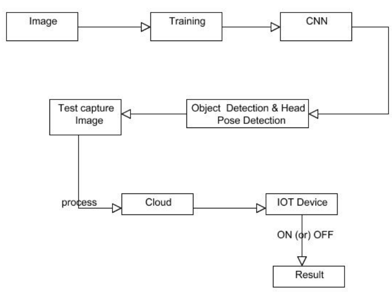


Fig-5: RCNN Architecture

4.4 CONVOLUTION NEURAL NETWORK

Convolutional neural networks (CNN) sounds like a weird combination of biology and math with a little CS sprinkled in, but these networks have been some of the most influential innovations in the field of computer vision. 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's Image Net competition dropping the classification error record from 26% to 15%, an astounding improvement at the time. Ever since then, a host of companies have been using deep learning at the core of their services. Facebook uses neural nets for their automatic tagging algorithms, Google for their photo search, Amazon for their product recommendations, Pinterest for their home feed personalization, and Instagram for their search infrastructure. However, the classic, and arguably most popular, use case of these networks is for image processing. Within image processing, let's take a look at how to use these CNNs for image classification.

Fig-6: diagrammatic representation of CNN



5. EXPERIMENTAL OUTPUT

The experimental results and case studies demonstrate the feasibility of the proposed system. We conducted quantitative experiments with a prototype. The prototypes of the object detection module and gaze estimation module were configured using commercial webcams and small single-board computers. The RCNN and CNN algorithms are used for gaze estimation and object detection were trained using a desktop PC by training process called back propagation it's the way the computer is able to adjust its filter values (or weights). Mainly the experiment runs by the idea of being given an image and a label is the training process that CNNs go through the IOT platform used in our experiment is a prototype implementation from the authors' previous work.

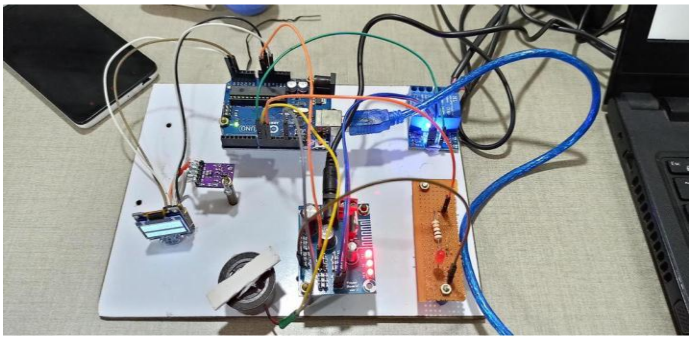


Fig-7: Experimental setup of a hand gesture censor connected to a prototype of fan, light, television and a lED display

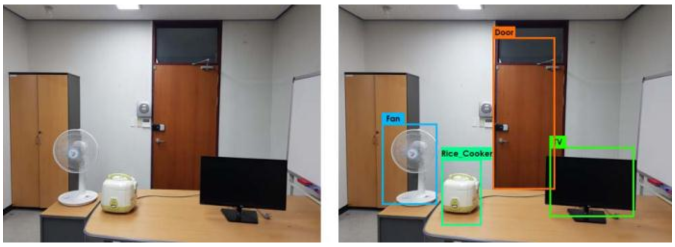


Fig-8: Output of the object detection module detecting IOT devices in the room

5. CONCLUSION

In this work, we proposed a novel smart IOT interaction System. The proposed method helps users easily control the IOT devices by gazing at a device and performing simple hand gestures. The main contributions of this work are as follows. First, we proposed a novel approach to utilize head pose estimation and object detection technologies for IOT interaction added to that we perform voice command output of the identified target object which helps the blind people to identify the objects. The previous works on gaze estimation focused on tracking the gaze position on a display. However, these approaches cannot be applied to the IOT environment that requires indoor gaze estimation. In this work, we solve this problem by a combination of object detection and implicit gaze estimation from the fine-grained head pose information. Second, without any additional devices, such as a headband or watch, the users can specify the target device. The previous works used wearable devices to detect user's head pose or gaze direction, which can be uncomfortable for many users. With this work, the users can interact with the IOT devices intuitively by installing simple modules

6. REFERENCES

[1] O. B. Sezer, E. Dogdu, and A. M. Ozbayoglu, "Context-aware com- puting, learning, and big data in Internet of Things: A survey," IEEE Internet Things J., vol. 5, no. 1, pp. 1–27, Feb. 2018.

[2] J. B. B. Neto, T. H. Silva, R. M. Assunç ao, R. A. F. Mini, and A. A. F. Loureiro, "Sensing in the collaborative Internet of Things," IEEE Sensors J., vol. 15, no. 3, pp. 6607–6632, Mar. 2015.

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

[4] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi," IEEE Commun. Mag., vol. 55, no. 10, pp. 105–111, Oct. 2017.

[5] 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.

[6] H. B. Pandya and T. A. Champaneria, "Internet of Things: Survey and case studies," in Proc. EESCO, Visakhapatnam, India, 2015, pp. 1–6.