

FACE RECOGNITION AND GESTURE BASED SECURED LOCKER CONTROL AND HOME APPLIANCE CONTROL SYSTEM FOR DISABLED PEOPLE

Submitted in partial fulfillment of the requirements for the award of Bachelor
of Engineering degree in Computer Science and Engineering

by

HANSINI P (Reg. No. 37110263)

HARSHITAA A (Reg. No. 37110272)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND
TECHNOLOGY(DEEMED TO BE
UNIVERSITY)**

**Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by
AICTE**

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH - 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC | 12B Status by UGC | Approved by
AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Ms. **HANSINI P**
(Reg. No. 37110263) and **HARSHITAA A** (Reg. No. 37110272) who carried out the project
entitled "**FACE RECOGNITION AND GESTURE BASED SECURED LOCKER CONTROL**
AND HOME APPLIANCE CONTROL SYSTEM FOR DISABLED PEOPLE"
under my supervision from November 2020 to March 2021.

Internal Guide
Dr. P. Asha, M.E., PH.D.,

Head of the Department
Dr.L.Lakshmanan M.E.,Ph.D.,
Dr.S.Vigneshwari M.E.,Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **HANSINI P (Reg. No. 37110263)** and **HARSHITAA A (Reg. No. 37110272)** hereby declare that the Project Report entitled " **FACE RECOGNITION AND GESTURE BASED SECURED LOCKER CONTROL AND HOME APPLIANCE CONTROL SYSTEM FOR DISABLED PEOPLE**" done by me under the guidance of **Dr. ASHA. P, M.E., PH.D.**, Is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E.,Ph.D.**, Dean, School of Computing and **Dr.L.Lakshmanan M.E.,Ph.D.**, and **Dr.S.Vigneshwari M.E.,Ph.D.**, Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. Asha.P M.E.,Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Automation is the essential need for the present world. There are various types of automation like building automation, industrial automation, Home automation, Artificial Intelligence, etc. can be a potential application which provides support to all the persons. Home automation is the use and control of home appliances remotely or automatically. Day by day the gap between machines and humans is being reduced. Nowadays hand gesture-based home automation is getting more importance. Gesture recognition refers to recognizing the motion of the human parts like hand, face, etc. Most of the electronic components manufactures focuses on the hand gesture basis. In this proposed work we will be using artificial intelligence which recognizes human gestures without the use of any kind of sensors with the help of a camera and also an additional security is provided using face recognition and if the registered human is successfully detected then the locker is unlocked along with the gesture recognition. If an unauthorized person attempts to unlock the locker an alert is sent to the user's mobile application. In case of general home appliances only the human gestures are recognized to effectively operate the. Thus, our project aims at discovering an automation system without the use of sensors or a mobile app with the help of artificial intelligence.

TABLE OF CONTENTS

	ABSTRACT	v
	LIST OF FIGURES	viii
	LIST OF ABBREVIATIONS	ix
	LIST OF TABLES	x
CHAPTER NO.	TITLE	PAGE NO.
1.	INTRODUCTION	1
	1.1 OUTLINE OF PROJECT	1
	1.2 TECHNOLOGIES USED	3
	1.3 REACT NATIVE APP DEVELOPMENT	9
2.	LITERATURE SURVEY	11
	2.1 LITERATURE REVIEW	11
3.	AIM AND SCOPE OF THE PRESENT INVESTIGATION	23
	3.1 AIM OF THE PROJECT	23
	3.2 PROBLEM STATEMENT	23
	3.3 OBJECTIVES	23
	3.4 EXISTING SYSTEM	23
	3.5 PROPOSED SYSTEM	27
4.	EXPERIMENTAL OR MATERIALS AND METHODS;	29
	ALGORITHMS USED	
	4.1 ARCHITECTURAL DIAGRAM	29
	4.2 MODULE DESCRIPTION	31
	4.3 SOFTWARE REQUIREMENTS	39
	4.4 HARDWARE REQUIREMENTS	39
5.	RESULTS AND DISCUSSION, PERFORMANCE	40
	ANALYSIS	
6	SUMMARY AND CONCLUSIONS	53

REFERENCES	54
APPENDIX	
A. SOURCE CODE	56
B. PLAGIARISM REPORT	63

LIST OF FIGURES

FIGURE NO.	NAME OF THE FIGURE	PAGE NO.
1.1	Deep Learning Layers	4
1.2	Deep Learning Process	5
1.3	Types of Deep Learning Networks	7
1.4	CNN	8
4.1	Face detection architecture	29
4.2	Gesture recognition	30
4.3	Ultralight face detector flow diagram	34
4.4	MobileFaceNet architecture	36
4.5	Hand Gesturenets diagram	38
4.6	Mobile App Development	39
5.1	Hardware setup	40
5.2	Login page	41
5.3	Home page	42
5.4	Hand gesture three	43
5.5	Hand gesture four	43
5.6	Running command in terminal	44
5.7	Accessing device 1 in hardware setup	44
5.8	Switching ON device 1 in hardware setup	45
5.9	Device 1 Switched ON	45
5.10	Accessing device 2 in hardware setup	46
5.11	Switching ON device 2 in hardware setup	46
5.12	Device 2 Switched ON	47
5.13	Accessing the locker with gesture 3	47
5.14	Authorized face recognition	48
5.15	Theft detection	49
5.16	Alert to the mobile application when unauthorized person	50
5.17	The locker is opened	51
5.18	Accuracy generation after training process	51

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
RNN	Recurrent Neural Networks
HTS	Hand tracking and segmentation

LIST OF TABLES

TABLE NO.	TABLE NAME	PAGE NO.
4.1	Mobilefacenet architecture	36

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Smart homes utilize top innovations such as advanced lighting elements and security systems to bring comfort and convenience to your living experience. As smart houses become more popular and the technological advances are more widely used, owning a smart home has its advantages. Nevertheless, many factors still keep homeowners hesitant about purchasing a smart home.

Installing state-of-the-art features inside a home results in a higher price tag for the property. The cost of an intelligent home that makes our lives convenient is high because some of the technology is relatively new. In new smart homes, wireless cameras and light sensors, a central touchscreen and automated systems are just a few advances that are standard in a new home. The cost-of-living expenses such as utilities, maintenance and repair of the technology can be expensive as well.

Owning a smart home means having to learn how to use your home. Unlike traditional homes, smart home technology requires you to adapt to the innovations within your living area such as security systems, air units and a remote that controls your entire house. For the technology-savvy family, the smart home will help achieve convenience faster, but for others, it will take reading manuals and learning how-to before the benefits of convenience pay off.

Video surveillance can be a wonderful tool in heightening security and deterring crime, but when the technology falls into the wrong hands, issues of privacy can occur. Security sensors within the doors and walls of a smart home use wireless technology to transfer signals to a central control unit that notifies emergency officials of any foreign activity. Video surveillance feeds are also transferred wirelessly to a part of the home that can be monitored. If the security feeds from the video and sensor fall into the wrong hands, your smart home could be monitored by those who have managed to hack it.

Pros

1. Energy Savings

Home automation systems have definitely proven themselves in the arena of energy efficiency. Automated thermostats allow you to pre-program

temperatures based on the time of day and the day of the week. And some even adjust to your behaviors, learning and adapting to your temperature preferences without your ever inputting a pre-selected schedule. Traditional or behavior-based automation can also be applied to virtually every gadget that can be remotely controlled – from sprinkler systems to coffee makers.

Actual energy savings ultimately depend on the type of device you select and its automation capabilities. But on average, product manufacturers estimate the systems can help consumers save anywhere from 10 to 15 percent off of heating and cooling bills.

2. Convenience

In today's fast-paced society, the less you have to worry about, the better. Right? Convenience is another primary selling point of home automation device, which virtually eliminate small hassles such as turning the lights off before you go to bed or adjusting the thermostat when you wake up in the morning.

Many systems come with remote dashboard capabilities, so forgetting to turn off that coffee pot before you leave no longer requires a trip back to the house. Simply pull up the dashboard on a smart device or computer, and turn the coffee pot off in a matter of seconds.

3. Security

Remote monitoring can put your mind at ease while you're away from the house. With remote dashboards, lights and lamps can be turned on and off, and automated blinds can be raised and lowered. These capabilities – combined with automated security systems – can help you mitigate the risks of intrusions: you will be alerted immediately if something uncharacteristic happens.

Cons

1. Installation

Depending on the complexity of the system, installing a home automation device can be a significant burden on the homeowner. It can either cost you money if you hire an outside contractor or cost you time if you venture to do it yourself.

2. Complex Technology

Automating everything in life may sound extremely appealing, but sometimes a good old-fashioned flip of the switch is a lot easier than reaching for your smart phone to turn lights on and off. Before you decide which system is right for you, think about how far you really want to take home automation in your household.

3. System Compatibility

Controlling all aspects of home automation from one centralized platform is important, but not all systems are compatible with one another. Your security system, for example, may require you to log in to one location to manage settings, while your smart thermostat may require you to log in to another platform to turn the air conditioner on and off. To truly leverage the convenience of home automation, you may need to invest in centralized platform technology to control all systems and devices from one location.

4. Cost

Even though the price of home automation systems has become much more affordable in recent years, the cost to purchase and install a device can still add up. Consumer Reports offers a wide range of information and insights – including costs – on the best home automation systems on the market and In each and every time the WiFi is not connected so, this is main disadvantage of home automation.

1.2 TECHNOLOGIES USED

Deep learning is a computer software that mimics the network of neurons in a brain. It is a subset of machine learning and is called deep learning because it makes use of deep neural networks. Deep learning algorithms are constructed with connected layers.

- The first layer is called the Input Layer
- The last layer is called the Output Layer

- All layers in between are called Hidden Layers. The word deep means the network join neurons in more than two layers.

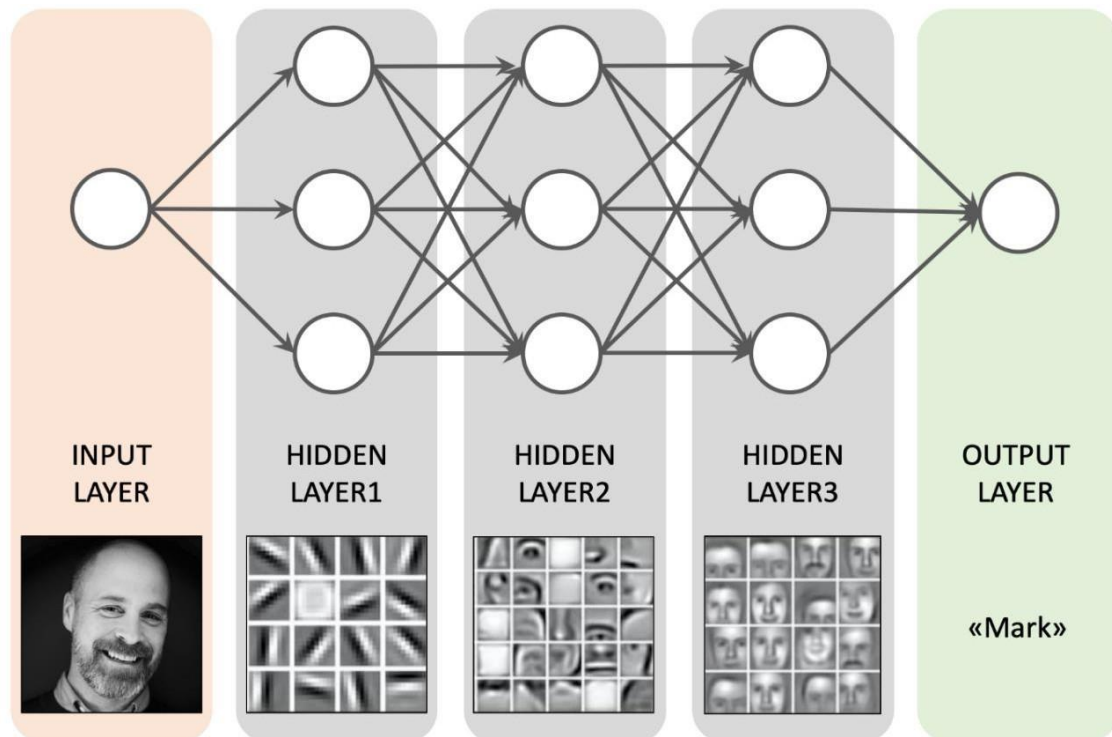


Figure 1.1 Deep Learning Layers

Each Hidden layer is composed of neurons. The neurons are connected to each other. The neuron will process and then propagate the input signal it receives the layer above it. The strength of the signal given the neuron in the next layer depends on the weight, bias and activation function.

The network consumes large amounts of input data and operates them through multiple layers; the network can learn increasingly complex features of the data at each layer.

1.2.1 Importance of Deep Learning:

Deep learning is a powerful tool to make prediction an actionable result. Deep learning excels in pattern discovery (unsupervised learning) and knowledge-based prediction. Big data is the fuel for deep learning. When both are combined, an organization can reap unprecedented results in term of productivity, sales,

management, and innovation.

Deep learning can outperform traditional method. For instance, deep learning algorithms are 41% more accurate than machine learning algorithm in image classification, 27 % more accurate in facial recognition and 25% in voice recognition.

1.2.2 Deep Learning Process:

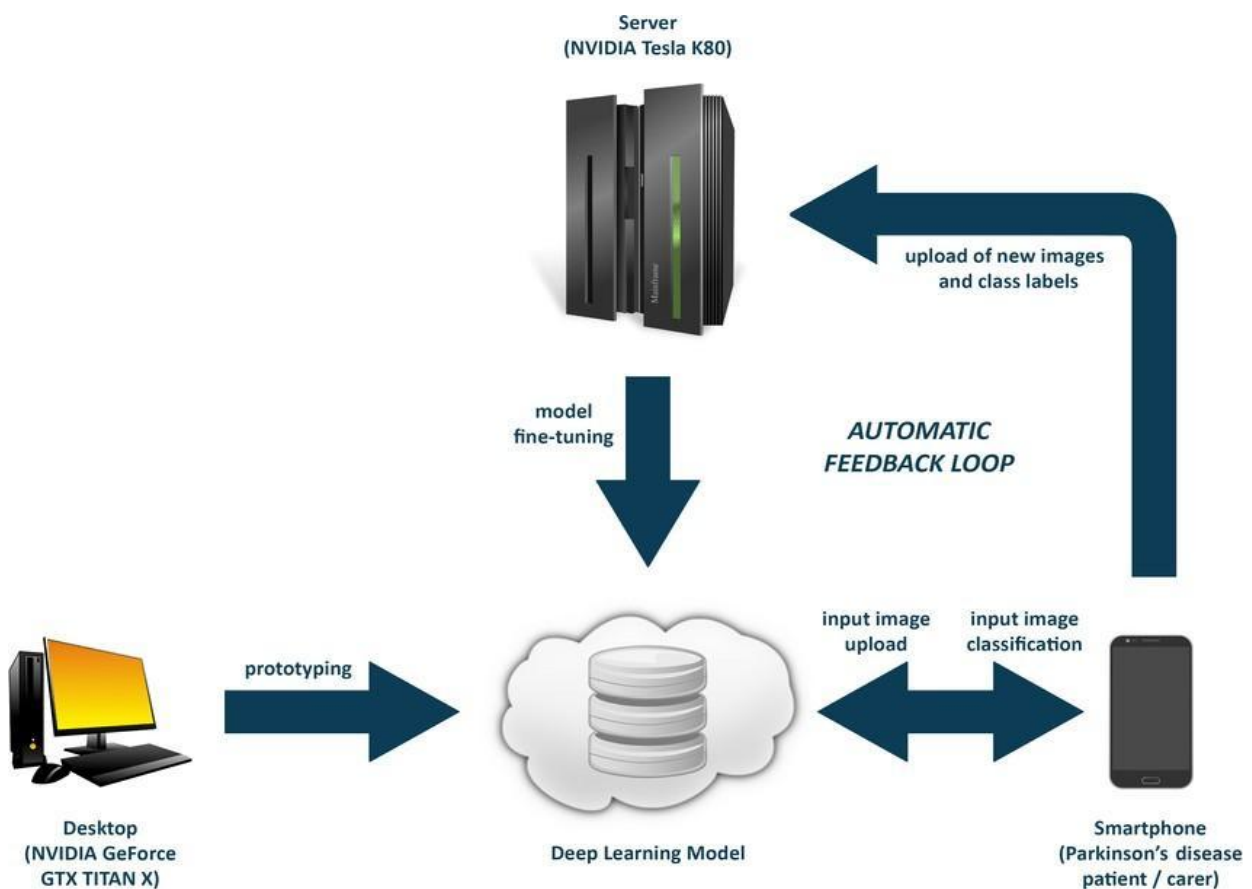


Figure 1.2 Deep Learning Process

A deep neural network provides state-of-the-art accuracy in many tasks, from object detection to speech recognition. They can learn automatically, without predefined knowledge explicitly coded by the programmers.

To grasp the idea of deep learning, imagine a family, with an infant and parents. The toddler points objects with his little finger and always says the word 'cat.' As its parents are concerned about his education, they keep telling him 'Yes, that is a cat' or 'No, that is not a cat.' The infant persists in pointing objects but becomes more accurate with 'cats.' The little kid, deep down, does not know why he can say it is a cat or not. He has just learned how to hierarchies complex features coming up with a cat by looking at the pet overall and continue to focus on details such as the tails or the nose before to make up his mind.

A neural network works quite the same. Each layer represents a deeper level of knowledge, i.e., the hierarchy of knowledge. A neural network with four layers will learn more complex feature than with that with two layers.

The learning occurs in two phases.

The first phase consists of applying a nonlinear transformation of the input and create a statistical model as output.

The second phase aims at improving the model with a mathematical method known as derivative.

The neural network repeats these two phases, hundreds to thousands of time until it has reached a tolerable level of accuracy. The repeat of this two-phase is called an iteration.

1.2.3 Classification of Neural Networks:

Shallow neural network: The Shallow neural network has only one hidden layer between the input and output.

Deep neural network: Deep neural networks have more than one layer. For instance, Google LeNet model for image recognition counts 22 layers.

Nowadays, deep learning is used in many ways like a driverless car, mobile phone, Google Search Engine, Fraud detection, TV, and so on.

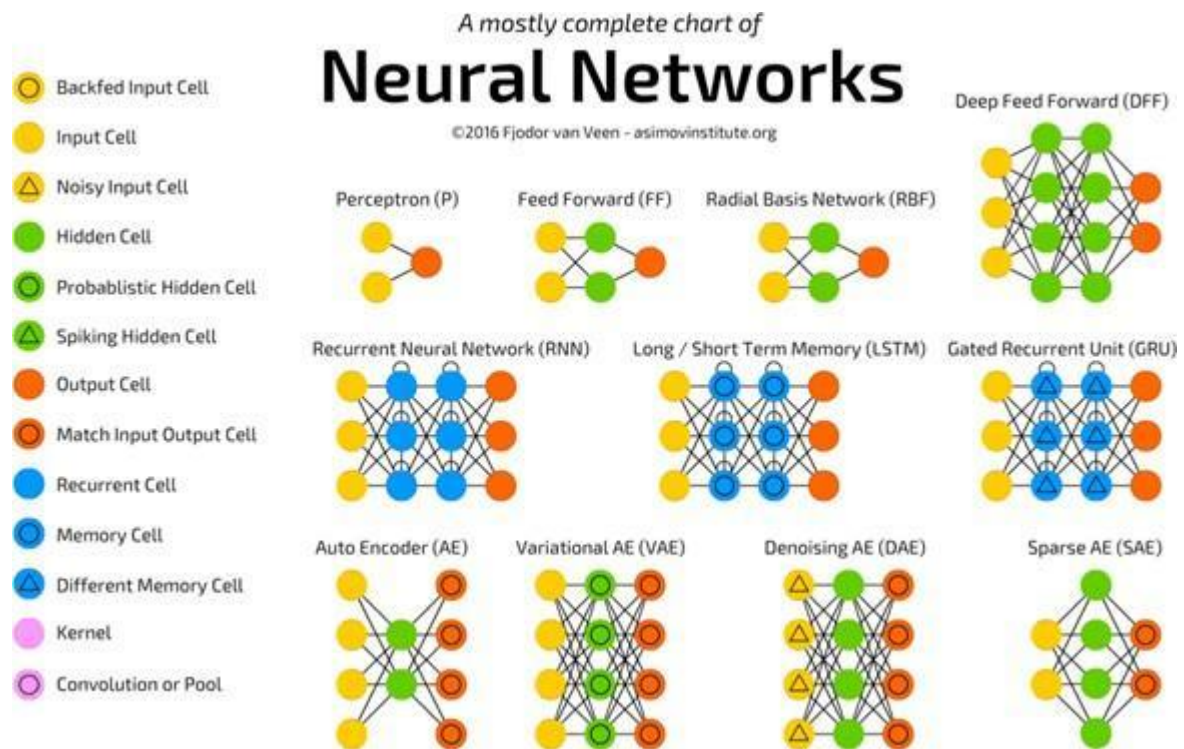


Figure 1.3 Types of Deep Learning Networks

1.2.4 Feed-Forward Neural Networks:

The simplest type of artificial neural network. With this type of architecture, information flows in only one direction, forward. It means, the information's flows starts at the input layer, goes to the "hidden" layers, and end at the output layer. The network does not have a loop. Information stops at the output layers.

1.2.5 Recurrent Neural Networks (Rnns):

RNN is a multi-layered neural network that can store information in context nodes, allowing it to learn data sequences and output a number or another sequence. In simple words it an Artificial neural network whose connections between neurons include loops. RNNs are well suited for processing sequences of inputs.

- The RNN neurons will receive a signal that point to the start of the sentence.
- The network receives the word "Do" as an input and produces a vector of the

number. This vector is fed back to the neuron to provide a memory to the

network. This stage helps the network to remember it received "Do" and it received it in the first position.

- The network will similarly proceed to the next words. It takes the word "you" and "want." The state of the neurons is updated upon receiving each word.
- The final stage occurs after receiving the word "a." The neural network will provide a probability for each English word that can be used to complete the sentence. A well-trained RNN probably assigns a high probability to "café," "drink," "burger," etc.

1.2.6 Common uses of RNN:

- Help securities traders to generate analytic reports
- Detect abnormalities in the contract of financial statement
- Detect fraudulent credit-card transaction
- Provide a caption for images
- Power chatbots
- The standard uses of RNN occur when the practitioners are working with time-series data or sequences (e.g., audio recordings or text).

1.2.7 Convolutional Neural Networks (Cnn):

CNN is a multi-layered neural network with a unique architecture designed to extract increasingly complex features of the data at each layer to determine the output. CNN's are well suited for perceptual tasks.

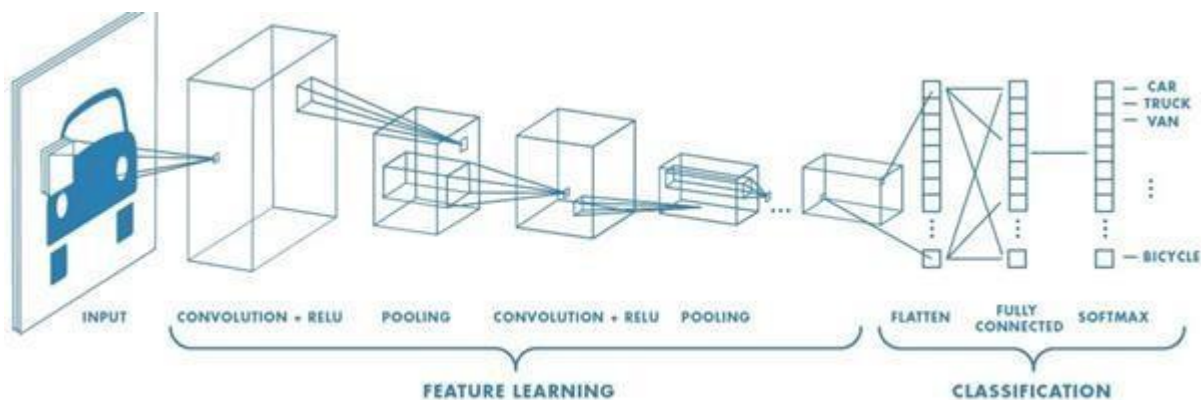


Figure 1.4 CNN

CNN is mostly used when there is an unstructured data set (e.g., images) and the practitioners need to extract information from it

For instance, if the task is to predict an image caption:

- The CNN receives an image of let's say a cat, this image, in computer term, is a collection of the pixel. Generally, one layer for the greyscale picture and three layers for a color picture.
- During the feature learning (i.e., hidden layers), the network will identify unique features, for instance, the tail of the cat, the ear, etc.
- When the network thoroughly learned how to recognize a picture, it can provide a probability for each image it knows. The label with the highest probability will become the prediction of the network.

A Convolutional Neural Network (CNN, or ConvNet) are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal preprocessing. The ImageNet project is a large visual database designed for use in visual object recognition software research. The ImageNet project runs an annual software contest, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where software programs compete to correctly classify and detect objects and scenes. Here I will talk about CNN architectures of ILSVRC top competitors.

1.3 React native app development:

Cross-platform development has become a great alternative to fully native mobile app development. Following the native mobile development approach, you create separate apps for Android and iOS. The cross-platform development allows you to cut expenses and save time by using the same code across both platforms. The React Native framework is a rising mobile solution and is considered the future of cross-platform mobile app development.

React Native is an open-source framework that allows to build a mobile app

with only JavaScript. It was introduced by Jordan Walke, a Facebook software engineer, as a new technology for simpler development and a better user

experience. The main distinctive of this framework is that React Native apps function just like native apps.

They don't differ from apps built on Java, Objective-C or Swift and they use the same UI building blocks as native iOS or Android apps. But with React Native, building a mobile app is much faster and less expensive.

Community matters

React Native is an open-source platform. That means all documentation related to this technology is open for everyone and is available for free to everyone in React Native community. There's a great advantage to using a community-driven technology. For example, if you face any issue related to React Native development, anyone can get help from community experts or find information online.

Reusable code & pre-built components

This is probably the most important advantage of React Native. With the convenience of this framework, developers don't need to build a separate mobile app for each platform, as almost 90 percent of React Native code can be reused between iOS and Android.

Simplified UI

React Native is all about the mobile UI. Comparing this framework to AngularJS or MeteorJS, it looks more like a JavaScript library than a framework.

It's important to create a sequence of actions when building a mobile app, and React Native makes an implementing order just perfect. In addition, user interfaces designed in React Native are more responsive, decrease load time, and provide a smoother feel.

Third-party plugin support

The React Native framework is still advancing, so it might lack some components in the core framework. To fill this gap, React Native provides two types of third-party plugins: native modules and JavaScript modules.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE REVIEW

2.2.1 *Andrea Bandini, Member, IEEE, and Jos´e Zariffa, Senior Member, IEEE, “Analysis of the hands in egocentric vision: A survey” [2020]*

Egocentric vision (a.k.a. first-person vision – FPV) applications have thrived over the past few years, thanks to the availability of affordable wearable cameras and large annotated datasets. The position of the wearable camera (usually mounted on the head) allows recording exactly what the camera wearers have in front of them, in particular hands and manipulated objects. This intrinsic advantage enables the study of the hands from multiple perspectives: localizing hands and their parts within the images; understanding what actions and activities the hands are involved in; and developing human-computer interfaces that rely on hand gestures. In this survey, we review the literature that focuses on the hands using egocentric vision, categorizing the existing approaches into: localization (where are the hands or parts of them?); interpretation (what are the hands doing?); and application (e.g., systems that used egocentric hand cues for solving a specific problem). Moreover, a list of the most prominent datasets with hand-based annotations is provided.

Pros:

A list of the most prominent datasets with hand-based annotations is provided.

Cons:

The position of the wearable camera (usually mounted on the head) allows recording exactly what the camera wearers have in front of them.

2.2.2 Qiang Nie, Jiangliu Wang, Xin Wang, Yunhui Liu, Fellow, IEEE," View-invariant Human Action Recognition Based on A 3D Bio-constrained Skeleton Model", IEEE Transactions on Image Processing, Vol. 14, No. 8, August 2019.

Skeleton-based human action recognition has been a hot topic in recent years. Most existing researches are based on the skeleton data obtained from Kinect which is noisy and unstable, in particular in the case of occlusions. To cope with the noisy skeleton data and variation of viewpoints, this paper presents a view-invariant method for human action recognition by recovering the corrupted skeletons based on a 3D bio- constrained skeleton model and visualizing those body-level motion features obtained during the recovery process with images. The bio-constrained skeleton model is defined with two types of constraints: 1) constant bone lengths and 2) motion limits of joints. Based on the bio-constrained model, an effective method is proposed for skeleton recovery. Two types of new motion features, the Euclidean Distance Matrix between joints (JEDM) which contains the global structure information of the body, and the local dynamic variation of the joint Euler angles (JEAs), are used in describing human action. These two types of features are encoded into different motion images which are fed into a two-stream convolutional neural network for learning different action patterns. A method for view-invariant skeleton-based action recognition based on a proposed 3D bio-constrained skeleton model. In the 3D bio-constrained skeleton model, lengths of bones are constant and joint motion limits recommended by the American Association of Orthopedic Surgeons (AAOS) are utilized. Skeleton recovery with the defined bio-constrained skeleton model is introduced to deal with corrupted skeletons or error joints.

Pros:

Skeleton recovery with the defined bio-constrained skeleton model is introduced to deal with corrupted skeletons or error joints.

Cons:

The project does not focus on the hand gestures.

2.2.3 An-An Liu*, Ning Xu*, Wei-Zhi Nie*, Yu-Ting Su, and Yong-Dong Zhang,” Multi-Domain & Multi-Task Learning for Human Action Recognition”, IEEE Transactions on Image Processing, VOL.1057-7149,2018.

Domain-invariant (view-invariant & modality-invariant) feature representation is essential for human action recognition. Moreover, given a discriminative visual representation, it is critical to discover the latent correlations among multiple actions in order to facilitate action modeling. To address these problems, we propose a multi-domain & multi-task learning (MDMTL) method to (1) extract domain-invariant information for multi-view and multi-modal action representation and (2) explore the relatedness among multiple action categories. Specifically, we present a sparse transfer learning-based method to co-embed multi-domain (multi-view & multi-modality) data into a single common space for discriminative feature learning. Additionally, visual feature learning is incorporated into the multi-task learning framework, with the Frobenius-norm regularization term and the sparse constraint term, for joint task modeling and task relatedness-induced feature learning. To the best of our knowledge, MDMTL is the first supervised framework to jointly realize domain-invariant feature learning and task modeling for multi-domain action recognition. We aim to solve the multi-domain action recognition problem by exploring and fusing view-invariant and modality-invariant information. We proposed the multi-domain & multi-task learning (MDMTL) framework, which jointly addresses the problems of domain-invariant feature learning and multi-task modeling.

Pros:

Jointly addresses the problems of domain-invariant feature learning and multi-task modeling.

Cons:

The results are not reliable compared to any other projects.

2.2.4 Ping Wei, Member, IEEE, Hongbin Sun, Member, IEEE, and Nanning Zheng, Fellow, IEEE,” Learning Composite Latent Structures for 3D Human Action Representation and Recognition”, VOL.1520-9210,2018.

3D human action representation and recognition are important issues in many multimedia applications. While latent state approaches have been widely used for action modeling, previous works assume the latent states of actions are single- attribute. This assumption is inaccurate for representing structures of complex actions. In this paper, we propose that latent states have composite attributes and introduce a novel composite latent structure (CLS) model to represent and recognize 3D human actions with skeleton sequences. A human action is modeled with a hierarchical graph, which represents the action sequence as sequential atomic actions. An atomic action is represented as a composite latent state which is composed of a latent semantic attribute and a latent geometric attribute. A discriminative EM-like algorithm is proposed to learn the model parameters and the composite latent structures of human actions. Given a 3D skeleton sequence, a composite attribute iterative programming algorithm is proposed to recognize the action and infer the action's latent temporal structure. We evaluate the proposed method on three challenging 3D action datasets - MSR 3D Action Dataset, Multiview 3D Event Dataset, and UTKinect-Action3D Dataset.

Pros:

Evaluate the proposed method on three challenging 3D action datasets - MSR 3D Action Dataset, Multiview 3D Event Dataset, and UTKinect-Action3D Dataset.

Cons:

The results of the project are not very accurate, hence is not applicable for real time usage.

2.2.5 Zhigang Tu, Member, IEEE, Hongyan Li, Dejun Zhang, Member, IEEE, Justin Dauwels, Senior Member, IEEE Baoxin Li, Senior Member, IEEE and Junsong Yuan, Senior Member, IEEE, “Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition”, IEEE Transactions on Image Processing, VOL.1057-7149,2018.

Despite outstanding performance in image recognition, convolutional neural networks (CNNs) do not yet achieve the same impressive results on action recognition in videos. This is partially due to the inability of CNN for modeling long-range temporal structures especially those involving individual action stages that are critical to human action recognition. In this paper, we propose a novel action-stage (ActionS) emphasized spatio-temporal Vector of Locally Aggregated Descriptors (ActionS-ST-VLAD) method to aggregate informative deep features across the entire video according to adaptive video feature segmentation and adaptive segment feature sampling (AVFS-ASFS). In our ActionS-ST-VLAD encoding approach, by using AVFS-ASFS, the key frame features are chosen and the corresponding deep features are automatically split into segments with the features in each segment belonging to a temporally coherent ActionS. Then, based on the extracted key frame feature in each segment, a flow-guided warping technique is introduced to detect and discard redundant feature maps, while the informative ones are aggregated by using our exploited similarity weight. Furthermore, we exploit an RGBF modality to capture motion salient regions in the RGB images corresponding to action activity.

Pros:

Exploit an RGBF modality to capture motion salient regions in the RGB images corresponding to action activity.

Cons:

The results are not up to the mark, hence is not applicable for long term purposes.

2.2.6 Nour El Din Elmadany, Student Member, IEEE, Yifeng He, Member, IEEE, and Ling Guan, Fellow, IEEE,"Information Fusion for Human Action Recognition via Biset/Multiset Globality Locality Preserving Canonical Correlation Analysis",IEEE Transactions on Image Processing,VOL.1057-7149,2018.

The problem of human action recognition, in which each action is captured by multiple sensors and represented by multisets. We propose two novel information fusion techniques for fusing the information from multisets. The first technique is biset globality locality preserving canonical correlation analysis (bglpcca), which aims to learn the common feature subspace between two sets. The second technique is multiset globality locality preserving canonical correlation analysis (mglpcca), which aims to deal with three or more sets. The proposed bglpcca and mglpcca are able to learn a low-dimensional common subspace that preserves the local and global structures of data samples. Moreover, two novel descriptors are presented for both depth and skeleton. We then propose a new human action recognition framework employing the proposed bglpcca or mglpcca to learn the shared subspace from multiple sets of features including skeleton, depth, and optical flow. Biset globality locality preserving canonical correlation analysis (bglpcca) for learning the common subspace from two sets, and multiset globality locality preserving canonical correlation analysis (mglpcca) for learning the common subspace from three or more sets. In feature extractions, two new descriptors were introduced. Boa was proposed for skeleton and hp-dmm-cnn for depth.

Pros:

The project gives solutions to two major challenges for robotic applications.

Cons:

The project is fundamentally not satisfying the problem of the hand gesture.

2.2.7 Sid Ahmed Walid Talha¹, Mounir Hammouche^{1,2}, Enjie Ghorbel^{1,3}, Anthony Fleury^{*1}, Member, IEEE, Sébastien Ambellouiss,” Features and Classification Schemes for View-Invariant and Real-Time Human Action Recognition”,IEEE Transactions on Cognitive and Developmental Systems,VOL.2379-8920,2018.

Human Action recognition (HAR) is largely used in the field of Ambient Assisted Living (AAL) to create an interaction between humans and computers. In these applications, it cannot be asked to people to act non-naturally. The algorithm has to adapt and the interaction has to be as quick as possible to make this interaction fluent. To improve the existing algorithms with regards to that points, we propose a novel method based on skeleton information provided by RGB-D cameras. This approach is able to carry out early action recognition and is more robust to viewpoint variability. To reach this goal, a new descriptor called Body Directional Velocity is proposed and a real-time classification is performed. a novel approach to perform human action recognition using RGB-D sensors. We have focused our work on two major challenges for robotic applications, the robustness to viewpoint changes and the early recognition property.

Pros:

The project focuses primarily on two major challenges for robotic applications, the robustness to viewpoint changes and the early recognition property.

Cons:

The project uses the RGB-D cameras that are very poor in quality.

2.2.8 Aouaidjia Kamel, Bin Sheng, Po Yang, Ping Li, Ruimin Shen, and David Dagan Feng, Fellow, IEEE, "Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures", IEEE transactions on systems, man, and cybernetics: Systems, 2018.

A method (action-fusion) for human action recognition from depth maps and posture data using convolutional neural networks (cnns). Two input descriptors are used for action representation. The first input is a depth motion image that accumulates consecutive depth maps of a human action, whilst the second input is a proposed moving joints descriptor which represents the motion of body joints over time. In order to maximize feature extraction for accurate action classification, three cnn channels are trained with different inputs. The first channel is trained with depth motion images (dmis), the second channel is trained with both dmis and moving joint descriptors together, and the third channel is trained with moving joint descriptors only. The action predictions generated from the three cnn channels are fused together for the final action classification. We propose several fusion score operations to maximize the score of the right action. The experiments show that the results of fusing the output of three channels are better than using one channel or fusing two channels only. Our proposed method was evaluated on three public datasets: 1) microsoft action 3-d dataset (msraction3d); 2) university of texas at dallas-multimodal human action dataset; and 3) multimodal action dataset (mad) dataset.

Pros:

The results of fusing the output of three channels are better than using one channel or fusing two channels only.

Cons:

The project cannot be trusted, since this project lacks human face detection.

2.2.9 Hongsong Wang, and Liang Wang, Senior Member, IEEE," Beyond Joints: Learning Representations from Primitive Geometries for Skeleton-based Action Recognition and Detection", IEEE Transactions on Image Processing, 2018.

Skeleton-based action recognition becomes popular owing to the development of cost-effective depth sensors and fast pose estimation algorithms. Traditional methods based on pose descriptors often fail on large-scale datasets due to the limited representation of engineered features. Recent recurrent neural networks (RNN) based approaches mostly focus on the temporal evolution of body joints and neglect the geometric relations. In this paper, we aim to leverage the geometric relations among joints for action recognition. We introduce three primitive geometries: joints, edges and surfaces. Accordingly, a generic end-to-end RNN based network is designed to accommodate the three inputs. For action recognition, a novel viewpoint transformation layer and temporal dropout layers are utilized in the RNN based network to learn robust representations. And for action detection, we first perform frame-wise action classification, then exploit a novel multi-scale sliding window algorithm. we aim to learn representations from primitive geometries of human skeletons, i.e., joints, edges and surfaces. For action recognition, we propose a novel RNN based architecture to accommodate the three inputs. For action detection, we first perform frame-wise action classification, then design a multi-scale sliding window search algorithm to generate detection results.

Pros:

First performs frame-wise action classification, then design a multi-scale sliding window search algorithm to generate detection results.

Cons:

There are some accuracy issues because it focuses primarily on the recognition part and not the security part.

2.2.10 Hossein Rahmani, Ajmal Mian and Mubarak Shah, “Learning a Deep Model for Human Action Recognition from Novel Viewpoints”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,VOL.0162-8828,2016.

Recognizing human actions from unknown and unseen (novel) views is a challenging problem. We propose a Robust Non-Linear Knowledge Transfer Model (R-NKTM) for human action recognition from novel views. The proposed R-NKTM is a deep fully-connected neural network that transfers knowledge of human actions from any unknown view to a shared high-level virtual view by finding a set of non-linear transformations that connects the views. The R-NKTM is learned from 2D projections of dense trajectories of synthetic 3D human models fitted to real motion capture data and generalizes to real videos of human actions. The strength of our technique is that we learn a single R-NKTM for all actions and all viewpoints for knowledge transfer of any real human action video without the need for re-training or fine-tuning the model. Thus, R-NKTM can efficiently scale to incorporate new action classes. R-NKTM is learned with dummy labels and does not require knowledge of the camera viewpoint at any stage. an algorithm for unsupervised learning of a Robust Non-linear Knowledge Transfer Model (R-NKTM) for cross-view action recognition.

Pros:

Robust Non-linear Knowledge Transfer Model (R-NKTM) for cross-view action recognition is best when comes to recognition.

Cons:

R-NKTM is learned with dummy labels and does not require knowledge of the camera, making it not very user-friendly.

2.2.11 Jian Cheng, Haijun Liu, Feng Wang, Hongsheng Li, and Ce Zhu, Senior Member, IEEE, "Silhouette Analysis for Human Action Recognition Based on Supervised Temporal t-SNE and Incremental Learning", IEEE transactions on image processing, Vol. 24, No. 10, October 2016.

A human action recognition method for human silhouette sequences based on supervised temporal t-stochastic neighbor embedding (st-tsne) and incremental learning. Inspired by the some and its variants, is proposed to learn the underlying relationship between action frames in a manifold, where the class label information and temporal information are introduced to well represent those frames from the same action class. As to the incremental learning, an important step for action recognition, we introduce three methods to perform the low-dimensional embedding of new data. Two of them are motivated by local methods, locally linear embedding and locality preserving projection. Those two techniques are proposed to learn explicit linear representations following the local neighbor relationship, and their effectiveness is investigated for preserving the intrinsic action structure. The rest one is based on manifold-oriented stochastic neighbor projection to find a linear projection from high-dimensional to low-dimensional space capturing the underlying pattern manifold.

Pros:

Learns explicit linear representations following the local neighbor relationship, and their effectiveness is investigated for preserving the intrinsic action structure.

Cons:

The proposed project is based on the neighboring embeddings making it very inefficient.

2.2.12 Sungpill Choi, Seongwook Park, and Hoi-Jun Yoo ,” A Memory-efficient Hand Segmentation Architecture for Hand Gesture Recognition in Low-power Mobile Devices”

Hand gesture recognition is regarded as new Human Computer Interaction (HCI) technologies for the next generation of mobile devices. Previous hand gesture implementation requires a large memory and computation power for hand segmentation, which fails to give real-time interaction with mobile devices to users. Therefore, in this paper, we presents a low latency and memory-efficient hand segmentation architecture for natural hand gesture recognition. To obtain both high memory-efficiency and low latency, we propose a streaming hand contour tracing unit and a fast contour filling unit. As a result, it achieves 7.14 ms latency with only 34.8 KB on-chip memory, which are 1.65 times less latency and 1.68 times less on-chip memory, respectively, compare to the best-in-class.

Pros:

It achieves 7.14 ms latency with only 34.8 KB on-chip memory, which are 1.65 times less latency and 1.68 times less on-chip memory, respectively, compare to the best-in-class.

Cons:

It is very poor in terms of optimization, hence making the whole system very inefficient.

2.2.13 Renliang Weng, Jiwen Lu, Senior Member, IEEE, and Yap-Peng Tan, Senior Member, IEEE, “Robust Point Set Matching for Partial Face Recognition”[Vol 1057-7149,2015]

Over the past three decades, a number of face recognition methods have been proposed in computer vision, and most of them use holistic face images for person identification. In many real-world scenarios especially some unconstrained environments, human faces might be occluded by other objects and it is difficult to obtain fully holistic face images for recognition. To address this, we propose a new

partial face recognition approach to recognize persons of interest from their partial faces. Given a pair of gallery image and probe face patch, we first detect key

points and extract their local textural features. Then, we propose a robust point set matching (RPSM) method to discriminatively match these two extracted local feature sets, where both the textural and geometrical information of local features are explicitly used for matching simultaneously. Lastly, the similarity of two faces is converted as the distance between these two aligned feature sets. Experimental results on four public face datasets show the effectiveness of the proposed approach.

Pros:

Experimental results on four public face datasets show the effectiveness of the proposed approach.

Cons:

It is very poor in terms of optimization, hence, making the whole system very inefficient.

2.2.14 Abhijith Punnappurath, Ambasamudram Narayanan Rajagopalan, Senior Member, IEEE, Sima Taheri, Student Member, IEEE, Rama Chellappa, Fellow, IEEE, and Guna Seetharaman, Fellow, IEEE, "Face Recognition Across Non-Uniform Motion Blur, Illumination, and Pose", VOL. 24, NO. 7, JULY 2015

Existing methods for performing face recognition in the presence of blur are based on the convolution model and cannot handle non-uniform blurring situations that frequently arise from tilts and rotations in hand-held cameras. In this paper, we propose a methodology for face recognition in the presence of space-varying motion blur comprising of arbitrarily-shaped kernels. We model the blurred face as a convex combination of geometrically transformed instances of the focused gallery face, and show that the set of all images obtained by non-uniformly blurring a given image forms a convex set. We first propose a nonuniform blur-robust algorithm by making use of the assumption

of a sparse camera trajectory in the camera motion space to build an energy

function with l_1 -norm constraint on the camera motion. The framework is then extended to handle illumination variations by exploiting the fact that the set of all images obtained from a face image by non-uniform blurring and changing the illumination forms a bi-convex set. Finally, we propose an elegant extension to also account for variations in pose.

Pros: Cons:

The project an elegant extension to also account for variations in pose.

The project is very insecure since anyone can perform the gesture to unlock

the actions.

2.2.15 Jiwen Lu, Member, IEEE, Venice Erin Liong, Xiuzhuang Zhou, Member, IEEE, and Jie Zhou, Senior Member, IEEE," Learning Compact Binary Face Descriptor for Face Recognition", [Vol 0162-8828,2015]

Binary feature descriptors such as local binary patterns (LBP) and its variations have been widely used in many face recognition systems due to their excellent robustness and strong discriminative power. However, most existing binary face descriptors are hand-crafted, which require strong prior knowledge to engineer them by hand. In this paper, we propose a compact binary face descriptor (CBFD) feature learning method for face representation and recognition. Given each face image, we first extract pixel difference vectors (PDVs) in local patches by computing the difference between each pixel and its neighboring pixels. Then, we learn a feature mapping to project these pixels difference vectors (PDVs) into low-dimensional binary vectors in an unsupervised manner, where 1) the variance of all binary codes in the training set is maximized, 2) the loss between the original real-valued codes and the learned binary codes is minimized, and 3) binary codes evenly distribute at each learned bin, so that the redundancy information in PDVs is removed and compact binary codes are obtained. Lastly, we cluster and pool these binary codes into a histogram feature as the final representation for each

face image. Moreover, we propose a coupled CBFD (C-CBFD) method by reducing the modality gap of heterogeneous faces at the feature level to make our

method applicable to heterogeneous face recognition. Extensive experimental results on five widely used face datasets show that our methods outperform state-of-the-art face descriptors.

Pros:

Results on five widely used face datasets show that our methods outperform state-of-the-art face descriptors.

Cons:

This system is hand crafted making it a very tiresome procedure to follow, hence is not recommended.

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 PROBLEM STATEMENT

Semi paralyzed or physically challenged people face difficulty in operating the appliances. Locker insecurity due to lack of any technological solution. Difficulty in controlling the appliances manually in work place or in crowded areas.

3.2 AIM OF PROJECT

Smart security system has become indispensable in modern daily life. The proposed security system has been developed to prevent robbery in highly secure areas like home environment with lesser power consumption and more reliable standalone security device for both Intruder detection and for door security. The door access control is implemented by using face recognition technology, which grants access to only authorized people to enter.

3.3 OBJECTIVES

The main objective of the project is to determine the gestures of the human and provide a solution to improve locker security. Its automatically switch on the particular device based on the gesture of the person and makes the life of the physically challenged people easy. Save the time of the people. And also accomplish automation with the help of AI without use of sensors.

3.4 EXISTING SYSTEM

Egocentric vision (a.k.a. first-person vision – FPV) applications have thrived over the past few years, thanks to the availability of affordable wearable cameras and large annotated datasets. The position of the wearable camera (usually mounted on the head) allows recording exactly what the camera wearers have in front of them, in particular hands and manipulated objects. This intrinsic advantage enables the study of the hands from multiple perspectives: localizing hands and

their parts within the images; understanding what actions and activities the hands are involved in; and developing human-computer interfaces that rely on hand

gestures. In this survey, we review the literature that focuses on the hands using egocentric vision, categorizing the existing approaches into: localization (where are the hands or parts of them?); interpretation (what are the hands doing?); and application (e.g., systems that used egocentric hand cues for solving a specific problem). Moreover, a list of the most prominent datasets with hand-based annotations is provided.

Disadvantages of existing system

The existing system for automation has only Hand gesture recognition.

Not reliable if no internet connection is available.

Normal people cannot afford.

Failure of sensor modules will lead to inefficient automation.

3.5 PROPOSED SYSTEM

Automation is the essential need for the present world. There are various types of automation like building automation, industrial automation, Home automation, Artificial Intelligence, etc. can be a potential application which provides support to all the persons. Home automation is the use and control of home appliances remotely or automatically.

Day by day the gap between machines and humans is being reduced. Nowadays hand gesture-based home automation is getting more importance. Gesture recognition refers to recognizing the motion of the human parts like hand, face, etc. Most of the electronic components manufactures focuses on the hand gesture basis. In this proposed work we will be using artificial intelligence which recognizes human gestures without the use of any kind of sensors with the help of a camera and also an additional security is provided using face recognition and if the registered human is successfully detected then the locker is unlocked along with the gesture recognition. If an unauthorized person attempts to unlock the locker an alert is sent to the user's mobile application. In case of general home appliances only the human gestures are recognized to effectively operate the.

Thus, our project aims at discovering an automation system without the use of sensors or a mobile app with the help of artificial intelligence.

Advantages of proposed system

Face recognition-based locker control

Hand gesture auto appliance control

More reliable

No sensors required

No internet connectivity needed

Suitable for physically challenged people

Applications

Smart homes

Lockers

Banks

CHAPTER 4

EXPERIMENTAL OR MATERIALS AND METHODS; ALGORITHMS USED

4.1 ARCHITECTURAL DIAGRAM

The architectural diagram is shown in

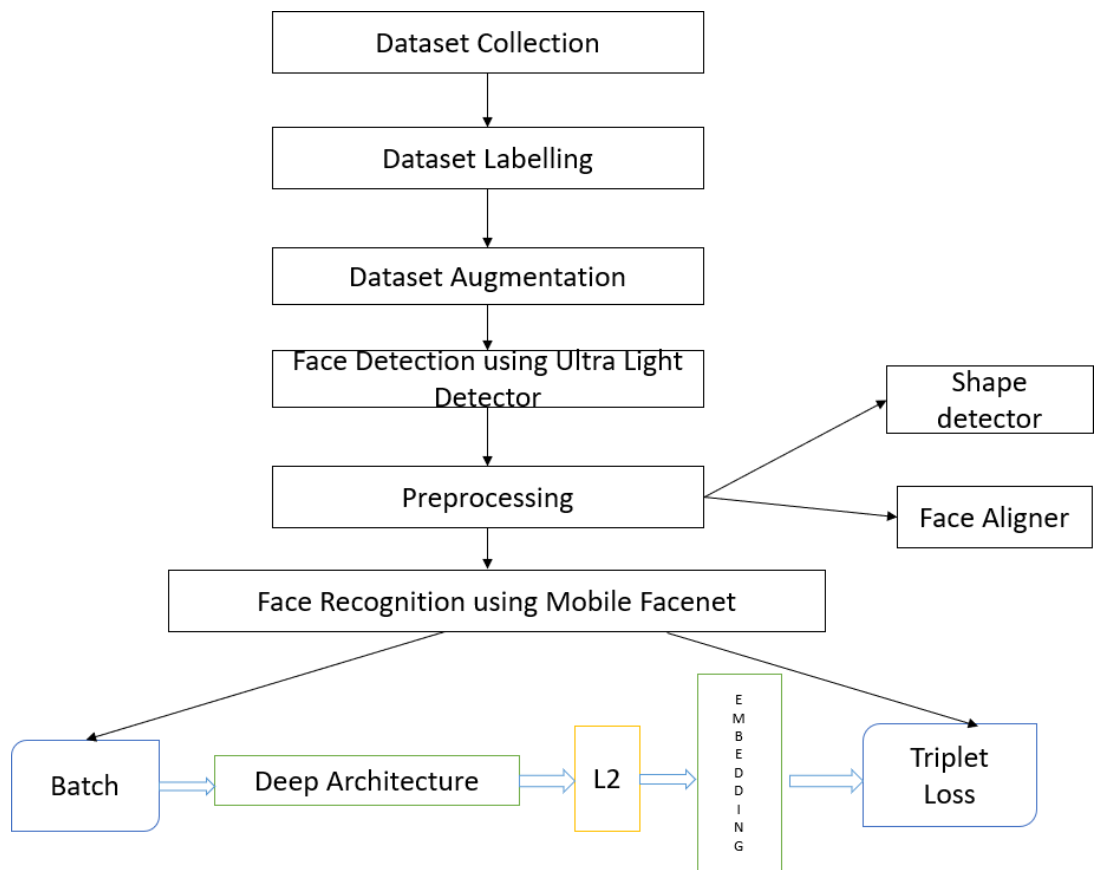


Figure 4.1 Face detection architecture

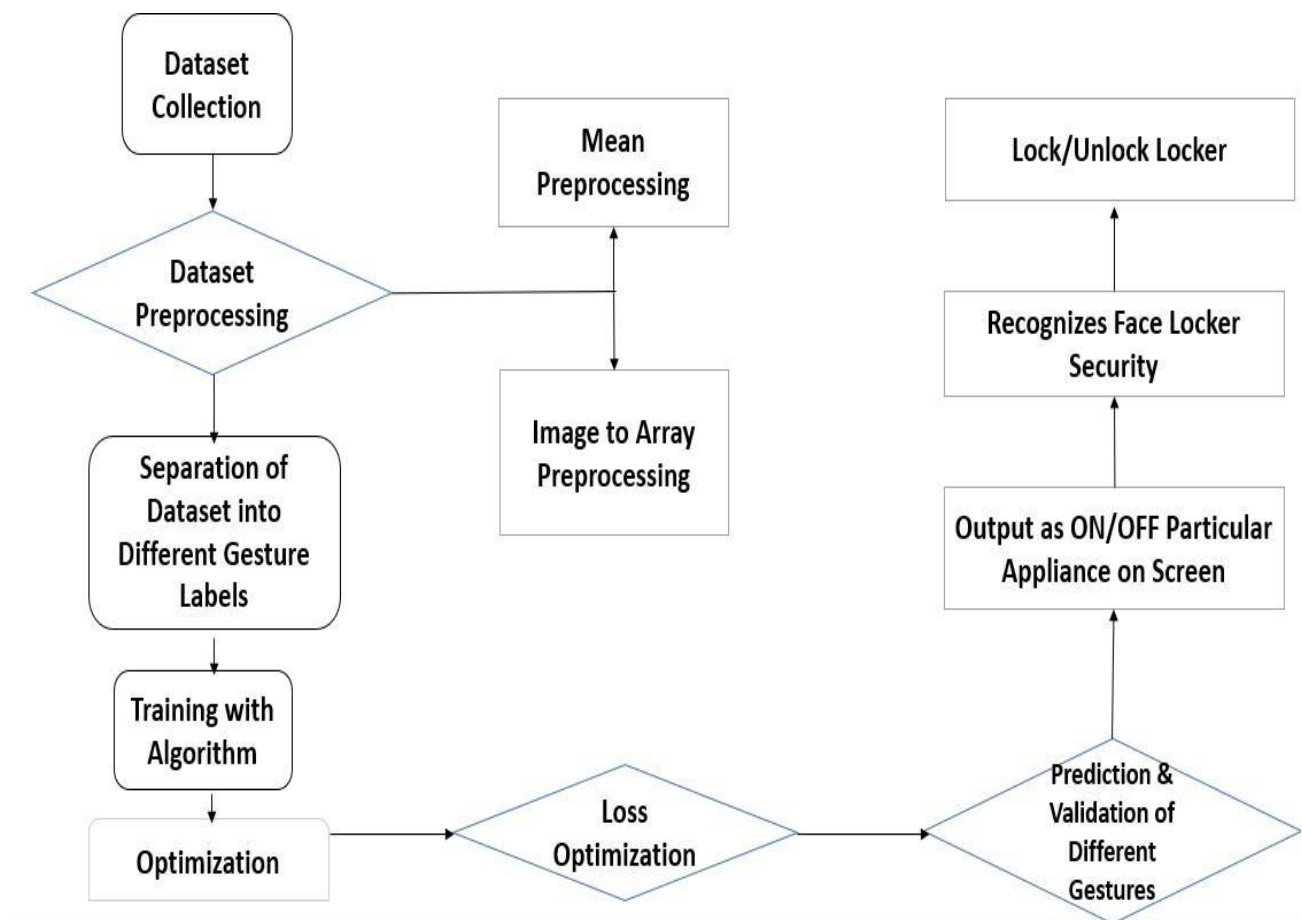


Figure 4.2 Gesture recognition

WORKING

In this project, we will be designing a hand gesture recognition-based home automation system. The initial step would be to collect the datasets of the faces of all registered people living in the house. Where only when the face of the person is recognized the hand gestures of the person are the accepted. will. The next step is to collect the dataset of different gestures by taking pictures of hands and collecting from various resources from internets. so, both these datasets are them trained by raspberry pi or the pc to identify the faces and particular gestures in order to switch on and off the appliances. so, initially raspberry pi will be configured with raspbian OS, it will be given with OpenCV libraries which will be used to trained the model in order to recognize the gestures. A camera module will be

interfaced with raspberry pi/ PC system which will lively stream the gestures shown

in front of it. These gestures are captured and given to the raspberry pi which will analyses it to switch on and off the appliances. From the dataset collected initially we will train the model and preprocess the gestures in order to get the approximate output. A driver circuit with a relay is also connected to the raspberry pi which will switch on and off high voltage appliances. Whenever the camera gives a face input, and then accepts the hand gesture input only if the face is recognized and then compares as to which appliance should be switched on based on that particular gesture. From this model we will be able to control appliances without any internet connection using artificial intelligence.

4.2 MODULES DESCRIPTION:

- Dataset Collection Module
- Face detection module
- Face recognition module
- Hand Segmentation
- Training with Gesturennet
- Compare the gestures and Switching the devices on off
- Mobile application development

4.2.1 Data Collection Module:

A data set is a collection of data. Deep Learning has become the go-to method for solving many challenging real-world problems. It's definitely by far the best performing method for computer vision tasks. The image above showcases the power of deep learning for computer vision. With enough training, a deep network can segment and identify the "key points" of every person in the image. These deep learning machines that have been working so well need fuel lots of fuel; that fuel is data. The more labelled data we have, the better our model performs. The idea of more data leading to better performance has even been explored at a large-scale by Google with a dataset of 300 Million images! When deploying a Deep Learning model in a real-world application, data must be constantly **fed** to continue improving its performance. And, in the deep learning era, data is very well

arguably your most valuable resource. There are three steps of collecting data

4.2.1.1 Scraping from the Web

Manually finding and downloading images takes a long time simply due to the amount of human work involved. The task probably has some kinds of common objects are to be detected. And so that becomes the keyword for web-scraping. It also becomes the class name for that object. From the sounds of it this is of course very easy for a task such as image classification where the images annotations are quite coarse. But to do something like instance segmentation? We need labels for *every single pixel in the image*. To get those, it's best to use some really great image annotation tools that are already out there. The paper shows how to create a model that, given a rough set of polygon points around an object, can generate the pixel labels for segmentation. Deep extreme cut is also quite similar except they use only the four extreme points around the object. This will then give some nice bounding box and segmentation labels. Another option is to use an existing image annotation GUIs. Label someone very popular where one can draw both bounding boxes and set polygon points for segmentation maps. Amazon Mechanical Turk is also a cheap option.

4.2.1.2 Third-party:

Since data has become such a valuable commodity in the deep learning era, many start-ups have started to offer their own image annotation services they'll gather and label the data. Given a description of what kind of data and annotations needed. Mighty is one that has been doing self-driving car image annotation and has become pretty big in the space were at CVPR 2018 too. Payment AI are less specialized than Mighty AI, offering image annotation for any domain. They also offer a couple more tools such as video and landmark annotations.

4.2.2 Face Detection Module

The "Ultra-Light-Fast-Generic-Face-Detector" is designed for general-purpose face detection applications in low-power computing devices and is applicable to both Android and iOS phones as well as PCs (CPU and GPU). The model is a real-time ultra-lightweight universal face detection model designed for

edge computing devices or low-power devices. It can be used in low-power computing devices such as ARM for real-time common scene faces.

Facial recognition technology is widely applied in security monitoring,

surveillance, human-computer interaction, entertainment, etc. Detecting human faces in digital images is the first step in facial recognition, and an ideal face detection model can be evaluated by how quickly and accurately it performs.

The Face-Detector-1MB stands out in terms of speed — the model's default FP32 precision (.pth) file size is 1.1MB, and the inference frame int8 is quantized to a size of 300KB. In terms of model calculation, the input resolution of 320×240 is only about 90 to 109 MFlops. The Face-Detector-1MB training process used a VOC dataset generated by the WIDER FACE dataset, a face detection benchmark. WIDER FACE was released in 2015 and consists of 32,203 images and 393,703 face bounding boxes with a high degree of variability in scale, pose, expression, occlusion and illumination.

The 1MB lightweight model comes in a version-slim with slightly faster simplification, and a version-RFB with a modified RFB module for higher precision. The model was tested on Ubuntu16.04, Windows 10, Python3.6, Pytorch1.2, CUDA10.0, etc.

Features

In terms of model size, the default FP32 precision (pth) file size is 1.04~1.1MB, and the inference frame int8 is about 300KB.

In the calculation of the model, the input resolution of 320x240 is about 90~109 M-Flops.

There are two versions of the model, version-slim (slightly faster simplification), version-RFB (with the modified RFB module, higher precision).

Provides pre-training models using wider face training at 320x240 and 640x480 different input resolutions to better work in different application scenarios.

Support for onnx export, easy to transplant.

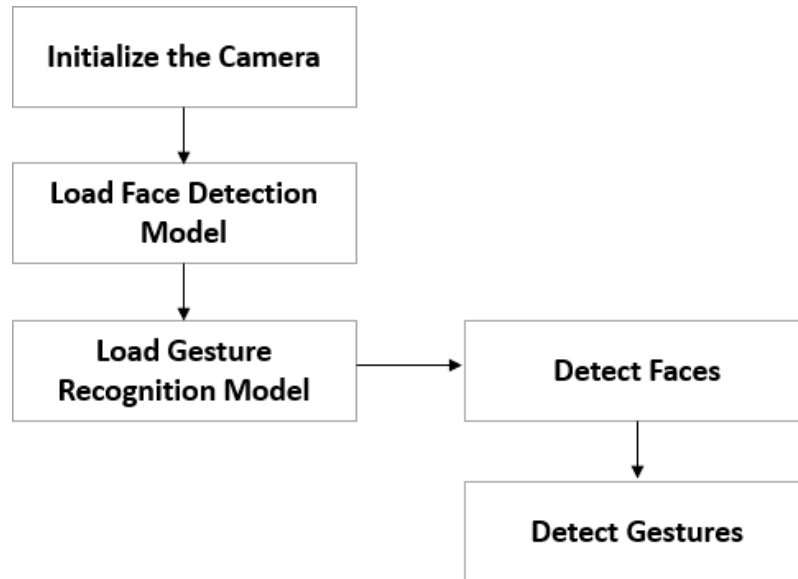


Figure 4.3 Ultralight face detector flow diagram

4.2.3 Face Recognition Module

For recognizing the face we use mobile facenet which is a more accurate in classifying face. MobileFaceNet is a neural network and obtains accuracy upto 99.28 percent on labelled faces in the wild (LFW) dataset, and a 93.05 percent accuracy on recognising faces in the AgeDB dataset. The network used around a million parameters taking only 24 milliseconds to run and produce results on a Qualcomm Snapdragon processor. We can compare this performance to accuracies of 98.70 percent and 89.27 percent for ShuffleNet, which has many more parameters and takes a little longer to execute on the CPU.

The researchers have made it easy to replace the global average pooling layer in the CNN with a depth wise convolution layer, which improves performance on facial recognition. This development is really important as the artificial intelligence world searches for efficient models that run on small compute powers.

Another approach for obtaining lightweight facial verification models is by compressing pretrained networks by knowledge distillation. Such approaches have achieved 97.32 percent facial verification accuracy on LFW with 4.0 MB model size. The remarkable achievement is that MobileFaceNets achieves comparable accuracy with very small budget.

MobileFaceNet architecture is partly inspired by the MobileNetV2 architecture. The residual bottlenecks proposed in MobileNetV2 are used as our main building blocks. The researchers use PReLU as the non-linearity, which is better suited for facial verification than using ReLU. The researchers also use a fast downsampling strategy at the beginning of the network, and a linear 1×1 convolution layer following a linear global depthwise convolution layer as the feature output layer. The detailed architecture is mentioned in the table below:

Input	Operator	t	c	n	s
$112^2 \times 3$	conv3x3	-	64	1	2
$56^2 \times 64$	depthwise conv3x3	-	64	1	1
$56^2 \times 64$	bottleneck	2	64	5	2
$28^2 \times 64$	bottleneck	4	128	1	2
$14^2 \times 128$	bottleneck	2	128	6	1
$14^2 \times 128$	bottleneck	4	128	1	2
$7^2 \times 128$	bottleneck	2	128	2	1
$7^2 \times 128$	conv1x1	-	512	1	1
$7^2 \times 512$	linear GDConv7x7	-	512	1	1
$1^2 \times 512$	linear conv1x1	-	128	1	1

Table 4.1 MobileFaceNet architecture

The primary MobileFaceNet network uses 0.99 million parameters. To reduce computational cost, the researchers decided to change input resolution from 112×112 to 112×96 or 96×96 . The the linear 1×1 convolution layer after the linear GDConv layer was also removed from MobileFaceNet. This gives a resulting network called MobileFaceNet.

The researchers have used MobileNetV1, ShuffleNet, and MobileNetV2 as the baseline models. All MobileFaceNet models and baseline models are trained on CASIA-Webface dataset from scratch by ArcFace loss, for a fair performance comparison among them. The training is finished at 60K iterations.

To pursue further excellent performance, MobileFaceNet, MobileFaceNet (112×96), and MobileFaceNet (96×96) are also trained on the cleaned training set of MSCeleb-1M database with 3.8 million images from 85,000 subjects. The accuracy of our primary MobileFaceNet is boosted to 99.55 percent and 96.07 percent on LFW and AgeDB-30, respectively.

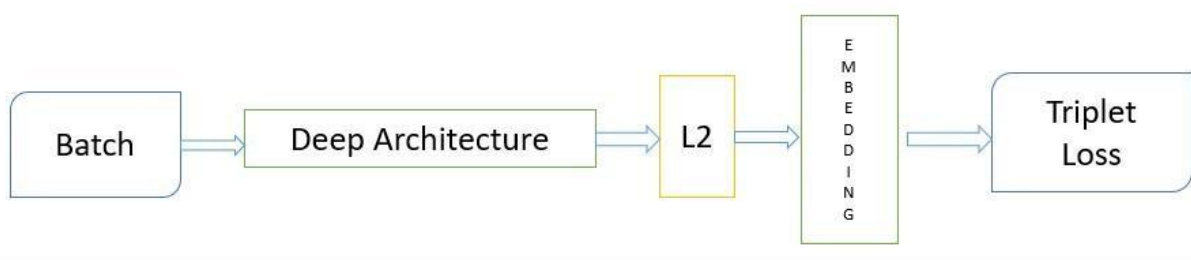


Figure 4.4 MobileFaceNet architercture

4.2.3 Hand Segmentation:

Since, efficient hand tracking and segmentation is the key of success towards any gesture recognition, due to challenges of vision based methods, such as varying lighting condition, complex background and skin color detection; variation in human skin color complexion required the robust development of algorithm for natural interface. Color is very powerful descriptor for object detection. So for the segmentation purpose color information was used, which is invariant to rotation and geometric variation of the hand. Human perceives characteristics of color component such as brightness, saturation and hue component than the percentage of primary color red, green, and blue. Color models are useful for to specify a particular color in standard way. It is space-coordinated system within which any specified color represented by single point. Here, three techniques were introduced using different color spaces for robust hand detection and segmentation. Hand tracking and segmentation (HTS)

technique using HSV color space is identified for the preprocessing of HGR

system.

4.2.5 Training with GestureNet:

The original images used for hand gesture recognition in the work are demonstrated. These images are captured with a normal camera. These hand images are taken under the same condition. The background of these images is identical. So, it is easy and effective to detect the hand region from the original image using the background subtraction method. However, in some cases, there are other moving objects included in the result of background subtraction. The skin color can be used to discriminate the hand region from the other moving objects. The color of the skin is measured with the HSV model. The HSV (hue, saturation, and value) value of the skin color is 315, 94, and 37, respectively. The image of the detected hand is resized to make the gesture recognition invariant to image scale. The output of the hand detection is a binary image in which the white pixels are the members of the hand region, while the black pixels belong to the background. An example of the hand detection. Then, the following procedure is implemented on the binary hand image to segment the fingers and palm. The palm point is defined as the center point of the palm. It is found by the method of distance transform. Distance transform also called distance map is a representation of an image. In the distance transform image, each pixel records the distance of it and the nearest boundary pixel. An example of distance transform is demonstrated is a binary image and is the distance transform image. The block city distance is used to measure the distances between the pixels and the nearest boundary pixels. As is shown in the figure, the center point of the binary image is with the largest distance. Thus, in the distance transform image of the binary hand image, the pixel with largest distance is chosen as the palm point. The found palm point is marked with the point of the green color. In this, we introduce a new algorithm.

It's named as GestureNet. Using this algorithm with different gestures are trained with high accuracy whenever, the input is given videos streaming that will be recognize the model automatically.



Figure 4.5 Hand Gesturenent diagram

4.2.6 Compare the gestures and switch the devices ON and OFF:

Combination of gestures to control the fan & light ON & OFF. Without internet connection the gestures can control the fan & light ON & OFF. Incoming gestures is always monitored and type of gestures which will be used to switch the appliances ON & OFF are compared from the incoming gestures. With the gestures compared, the fan/light are switched ON or OFF. The aged people and the physically challenged people can use this efficiently. Thus this project helps in automatic control of appliances using artificial intelligence.

4.2.7 Mobile App Development:

React Native is a framework that builds a hierarchy of UI components to build the JavaScript code. It has a set of components for both iOS and Android platforms to build a mobile application with a native look and feel. Mobile development has witnessed unprecedented growth. According to statistica, mobile applications will generate an estimated 188 billion U.S. dollars in revenue via app stores, advertising and in-app purchases by the year 2020. Single and business users require high-standard apps with flawless performance, multiple screens, easy navigation and good design. On the other side, high-performing, good quality native apps are very time-consuming to develop compared to cross-platform apps that provide faster development but compromise on performance and support. React Native seems to be a viable solution for building high-quality apps in a short time with the same performance and user-experience standards that native apps provide.

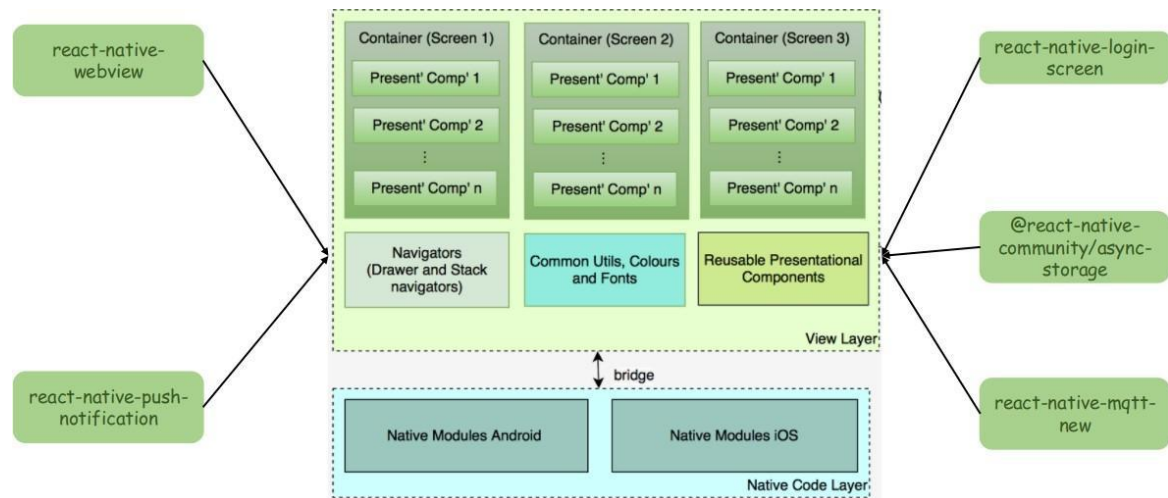


Figure 4.6 Mobile App Development

The architecture of React Native helps us in structuring a project for a multi-platform mobile application, keeping the logic of the business in a reusable and maintained sub-module. In some projects, we can see that the containers could be important to be in the core module but that depends on the applications. React Native uses different mechanisms to create an efficient, consistent and reusable visual identity for the applications.

4.3 SOFTWARE REQUIREMENTS

Python
Visual Studio

4.4 HARDWARE REQUIREMENTS

PC
Web Camera
Power Supply – 2A 5V DC adapter
12V DC electromechanical Relay
32GB Sandisk SD Card

230V Light bulb

CHAPTER 5

RESULTS AND DISCUSSIONS OF PERFORMANCE ANALYSIS

We have developed hand gesture recognition-based home automation system. The below image shows the hardware setup of the project:

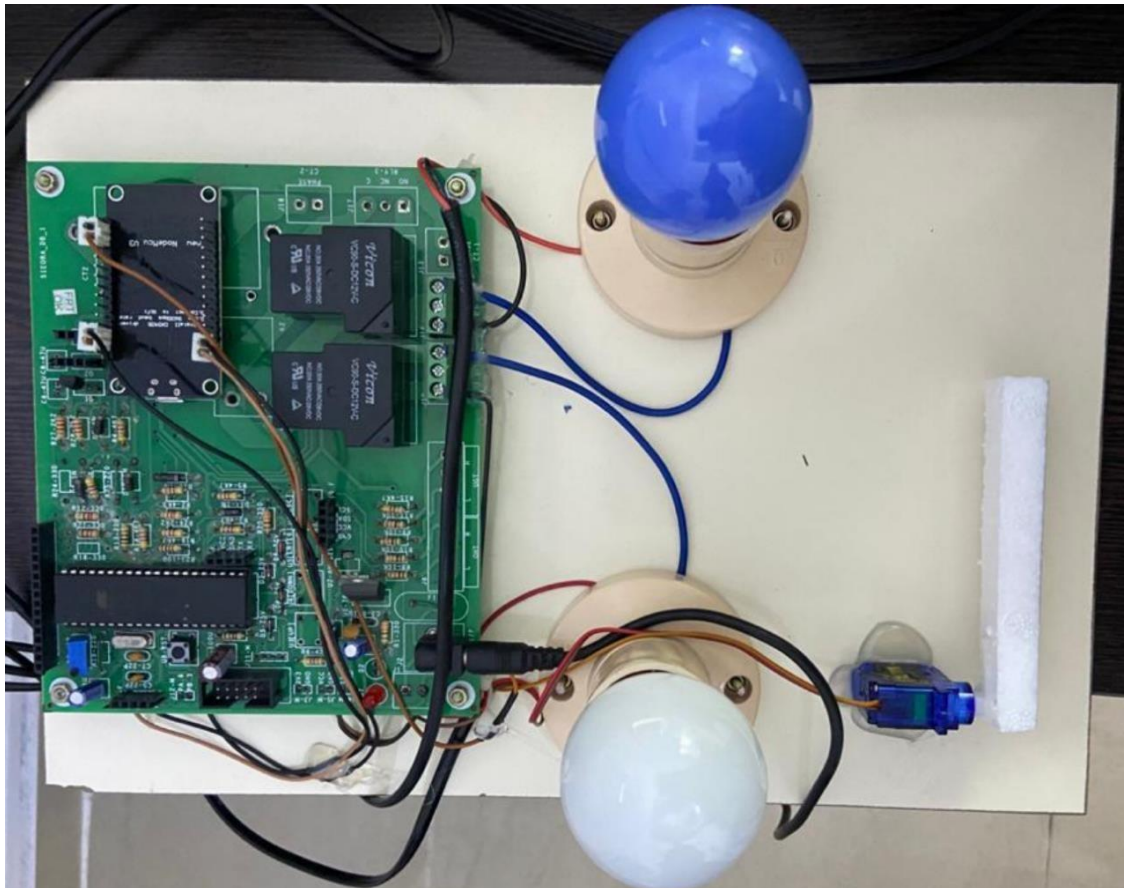


Figure 5.1 Hardware setup

A mobile application using react native is developed as an interface to the user notifying him/her when an unauthorized person tries to operate the locker. The below image shows the login page to the mobile application:

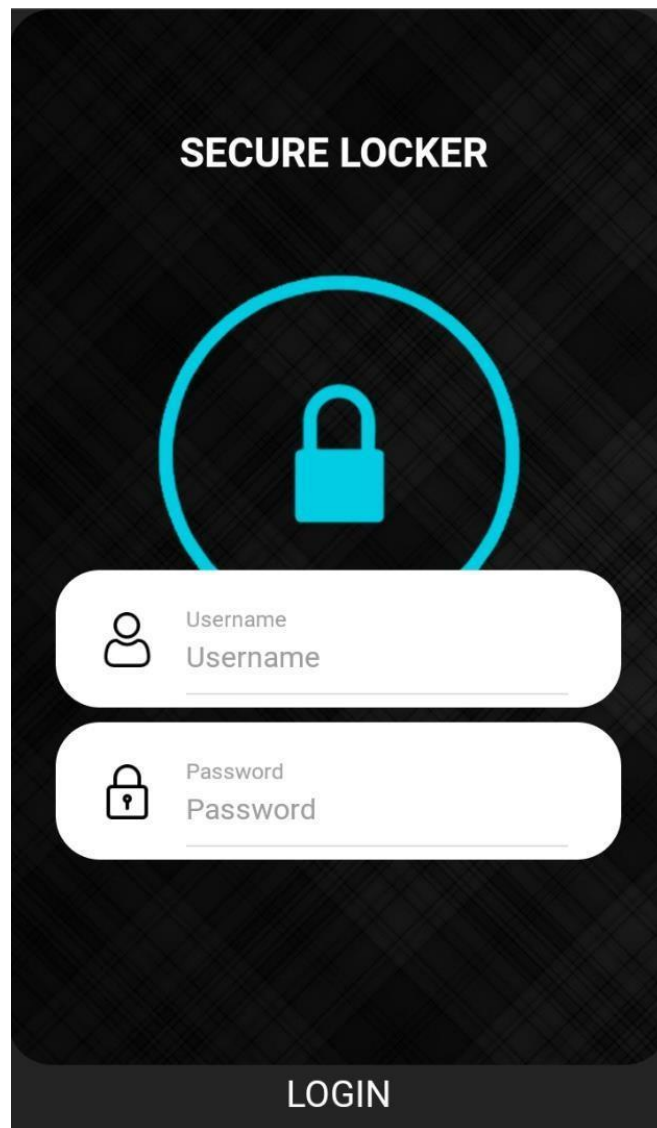


Figure 5.2 Login page

On successful login to the mobile application, the below image shows the screenshot of the mobile application home page:

Home Locker Status



DOOR IS CLOSE AND SAFE

Figure 5.3 Home page

To develop the system for hand gesture recognition, the first step will be dataset collection which can be collected using real time video streaming. that can be seen in the following figure.

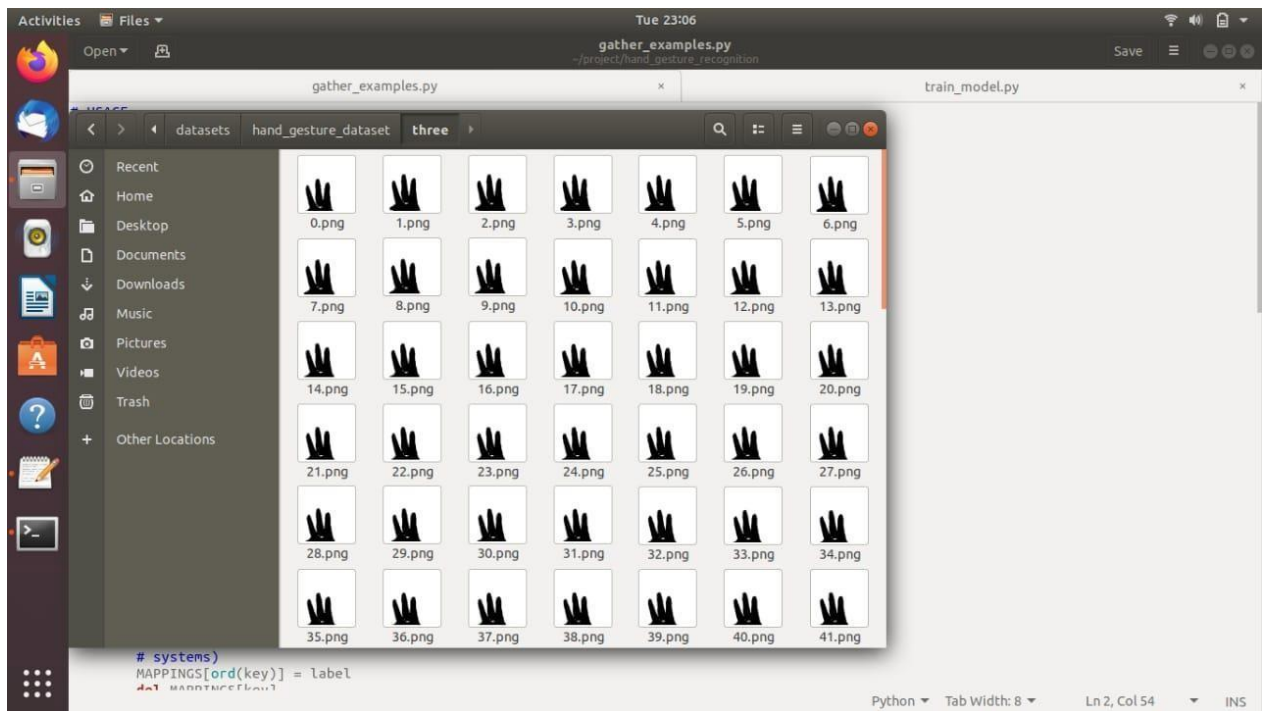
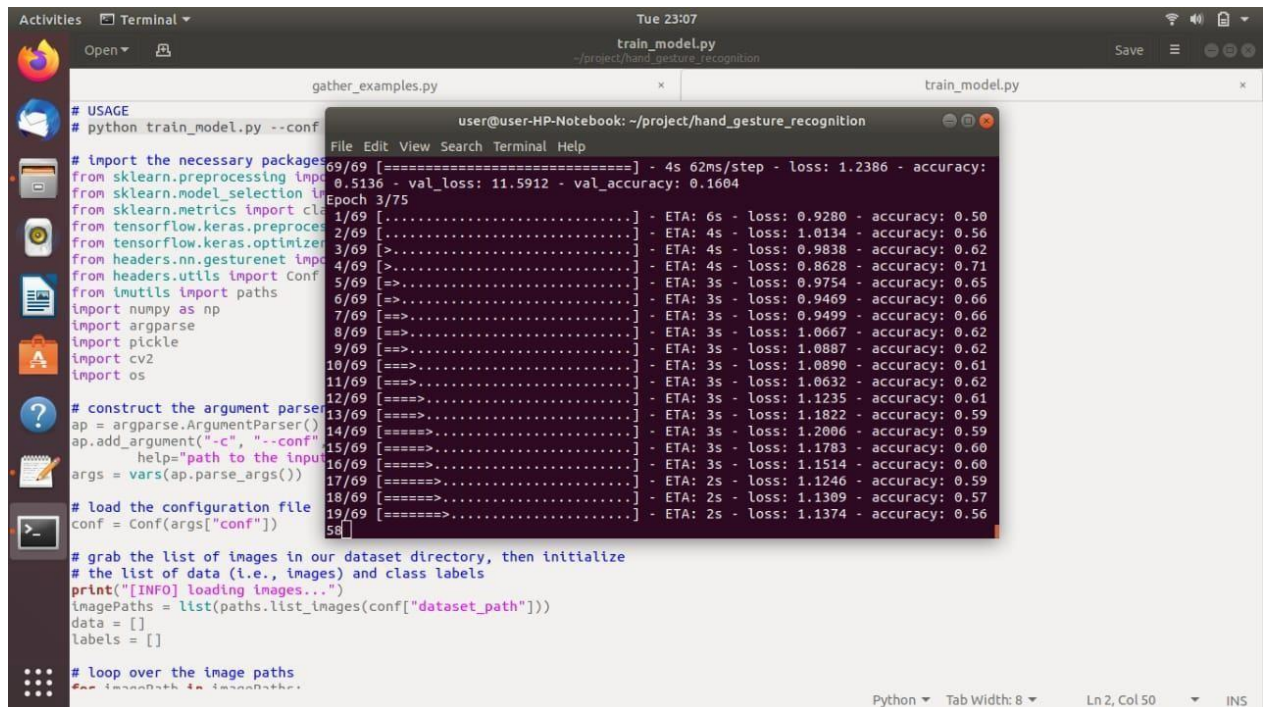


Figure 5.4 Hand gesture three



Figure 5.5 Hand gesture four

As the dataset collection is over, these datasets are trained for gesture recognition. The command for training is run in the terminal that can be seen in the following figure.



```
# USAGE
python train_model.py --conf

# import the necessary packages
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.optimizers import Adam
from headers.nn.gesture_net import Net
from headers.utils import Conf
from imutils import paths
import numpy as np
import argparse
import pickle
import cv2
import os

# construct the argument parser
ap = argparse.ArgumentParser()
ap.add_argument("-c", "--conf", help="path to the input configuration file")
args = vars(ap.parse_args())

# load the configuration file
conf = Conf(args["conf"])

# grab the list of images in our dataset directory, then initialize
# the list of data (i.e., images) and class labels
print("[INFO] loading images...")
imagePaths = list(paths.list_images(conf["dataset_path"]))
data = []
labels = []

# loop over the image paths
for imagePath in imagePaths:
```

```
69/69 [=====] - 4s 62ms/step - loss: 1.2386 - accuracy: 0.5136 - val_loss: 11.5912 - val_accuracy: 0.1604
Epoch 3/75
1/69 [.....] - ETA: 6s - loss: 0.9280 - accuracy: 0.50
2/69 [.....] - ETA: 4s - loss: 1.0134 - accuracy: 0.56
3/69 [.....] - ETA: 4s - loss: 0.9838 - accuracy: 0.62
4/69 [.....] - ETA: 4s - loss: 0.8628 - accuracy: 0.71
5/69 [====>] - ETA: 3s - loss: 0.9754 - accuracy: 0.65
6/69 [====>] - ETA: 3s - loss: 0.9469 - accuracy: 0.66
7/69 [====>] - ETA: 3s - loss: 0.9499 - accuracy: 0.66
8/69 [====>] - ETA: 3s - loss: 1.0667 - accuracy: 0.62
9/69 [====>] - ETA: 3s - loss: 1.0887 - accuracy: 0.62
10/69 [====>] - ETA: 3s - loss: 1.0890 - accuracy: 0.61
11/69 [====>] - ETA: 3s - loss: 1.0632 - accuracy: 0.62
12/69 [====>] - ETA: 3s - loss: 1.1235 - accuracy: 0.61
13/69 [====>] - ETA: 3s - loss: 1.1822 - accuracy: 0.59
14/69 [====>] - ETA: 3s - loss: 1.2006 - accuracy: 0.59
15/69 [====>] - ETA: 3s - loss: 1.1783 - accuracy: 0.60
16/69 [====>] - ETA: 3s - loss: 1.1514 - accuracy: 0.60
17/69 [====>] - ETA: 2s - loss: 1.1246 - accuracy: 0.59
18/69 [====>] - ETA: 2s - loss: 1.1309 - accuracy: 0.57
19/69 [====>] - ETA: 2s - loss: 1.1374 - accuracy: 0.56
58
```

Figure 5.6 Running command in terminal

On initiating the system camera in order to access the device 1 in the hardware setup hand gesture 1 should be given as input. The below image shows the hand gesture 1 to access the device 1 in the hardware setup:



Figure 5.7 Accessing device 1 in hardware setup

On accessing the device 1 hand gesture three should be given as input to switch ON the device 1 and hand gesture zero is used to switch OFF the device 1. The below image shows the hand gesture 3 to switch on the device 1



Figure 5.8 Switching ON device 1 in hardware setup

The below image shows the device 1 being switched on:

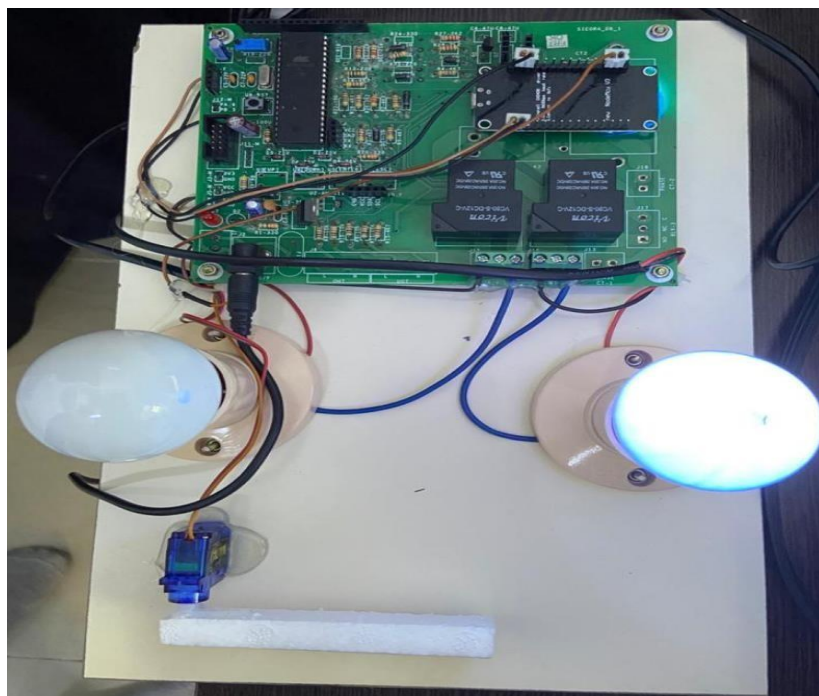


Figure 5.9 Device 1 switched ON

In order to access the device 2 in the hardware setup hand gesture zero should be given as input. The below image shows the hand gesture zero to access the device 2 in the hardware setup:



Figure 5.10 Accessing device 2 in hardware setup

Similar to the device 1 the hand gesture three is used to switch ON the device 2 and zero is used to switch OFF the device 2. The below image shows the hand gesture three to switch ON the device 2:

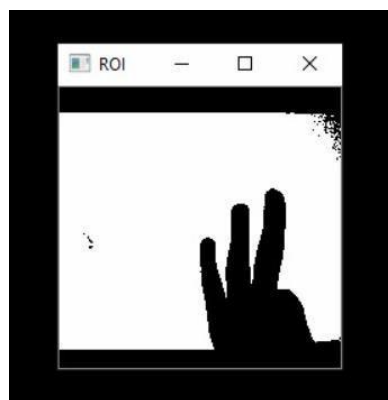


Figure 5.11 Switching ON device 2 in hardware setup

The below image shows the device 2 being switched on:

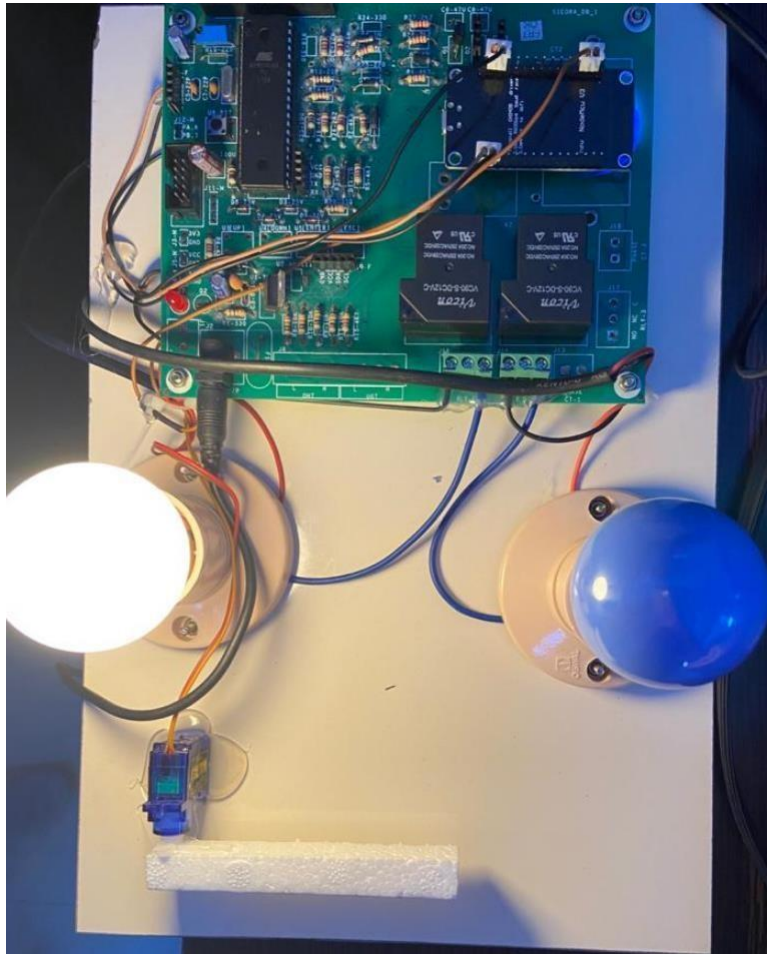


Figure 5.12 Device 2 switched ON

In order to access the locker hand gesture three is required to be given as input. The below image shows the hand gesture 3 given for accessing the locker:

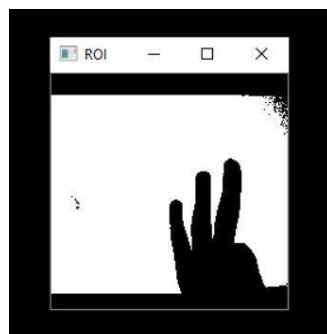


Figure 5.13 Accessing the locker with hand gesture 3

On accessing the locker face recognition is enabled to confirm that the user is an authorized person. The below image shows the face recognition initiation:



Figure 5.14 Authorized face recognition

If an unauthorized face is detected than theft is detected. The below image shows the unauthorized face detection:



Figure 4.15 Theft detection

If the person trying to access the locker is an unauthorized person then an alert is sent to the mobile application of the user. The below image shows the alert received at the mobile application of the user when the unauthorized person tries accessing the locker:



Figure 5.16 Alert to the mobile application when unauthorized person tries to access the locker

If the person trying to access the locker is recognized to be an authorized user then when given an input of hand gesture zero the locker is opened and hand gesture three is given as input to close the locker, the below image shows the locker opened:



Figure 5.17 The locker is opened

Thus, from the above results it is proved that we achieved home automation with the help of artificial intelligence where there will be no use of internet to achieve automation. By this project, the scope of the project is achieved.

```

1 # USAGE
2 # python train_model.py --conf config/config.json
3
4 # import the necessary packages
5 from sklearn.preprocessing import LabelBinarizer
6 from sklearn.model_selection import train_test_split
7 from sklearn.metrics import classification_report
8 from tensorflow.keras.preprocessing.image import ImageDataGenerator
9 from tensorflow.keras.optimizers import Adam
10 from headers.nn.gestureNet import GestureNet
11 from headers.utils import Conf
12 from imutils import paths
13 import numpy as np
14 import argparse
15 import pickle
16 import cv2
17 import os
18
19 # construct the argument parser and parse the arguments
20 ap = argparse.ArgumentParser()
21 ap.add_argument("-c", "--conf", required=True,
22                 help="path to the input configuration file")
23 args = vars(ap.parse_args())
24
25 # load the configuration file
26 conf = Conf(args["conf"])
27
28 # grab the list of images in our dataset directory, then initialize
29 # the list of data (i.e., images) and class labels
30 print("[INFO] loading images...")
31 imagePaths = list(paths.list_images(conf["dataset_path"]))
32 data = []
33 labels = []

```

```

C:\Windows\System32\cmd.exe
ined and being set to 0.0 in labels with no predicted samples. Use "zero_division" para
meter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))
precision    recall    f1-score   support
four         0.33      1.00      0.50      25
ignore       1.00      1.00      1.00      26
one          1.00      0.44      0.61      25
three        0.00      0.00      0.00      25
two          0.07      0.04      0.05      26
accuracy          0.48      0.50      0.50      127
macro avg         0.48      0.50      0.43      127
weighted avg      0.48      0.50      0.43      127

[INFO] saving model...
[INFO] serializing label encoder...
C:\Users\Harshita\Desktop\gesture_reg>

```

Figure 5.18 Accuracy generation after training process

MobileFaceNet is utilized here for recognizing face as it is more exact in characterizing face. MobileFaceNet acquires exactness up to 93.05 percent exactness on perceiving faces in the AgeDB dataset and about 99.28 percent on labelled faces in the wild (LFW) dataset as it is a neural network. More than a million parameters are utilized by the network taking just 24 milliseconds to execute and deliver outcomes on a Qualcomm Snapdragon processor. We can contrast this presentation with correctness of 98.70 percent and 89.27 percent for ShuffleNet, which has numerous more parameters and takes somewhat more time to execute on the CPU. MobileNetV2 architecture partly motivated or contributed for our by MobileFaceNet architecture that is, our fundamental building blocks is obtained from the residual obstracles proposed in MobileNetV2. The PReLU is used as the non-linearity, which is more qualified for facial confirmation than the ReLU. The MobileNetV1, ShuffleNet, and MobileNetV2 are used as the baseline models. To get a reasonable outcome when compared among the listed that is, all MobileFaceNet models and baseline models are trained 34 on CASIA-Webface dataset without any preparation .

CHAPTER 6

SUMMARY AND CONCLUSIONS

SUMMARY

In this project we will be using artificial intelligence which recognizes human gestures without the use of any kind of sensors with the help of a camera and acts accordingly to control any kind of home appliance. By this project physically challenged people can easily switch off the lights and fan or any home appliance. So, we reduce the time required for manual work.

CONCLUSION

We would conclude the paper, stating that the system proposed helps the disabled and elderly people to control the electronic home appliances and providing more secured locker control system than those which are existing with the help of face and hand gesture recognition, without any internet connectivity. We avoided the usage of sensors which are expensive and even need proper maintenance and replace them with the help of deep learning algorithms and artificial techniques. As the internet connectivity is not needed the proposed system can reach out to much more people where internet connectivity is not possible to establish or not available, for example rural areas. The gestures made by the person from a fixed position is captured by camera which is available at the room and gets switched on or off the required device respectively.

REFERENCES

- [1] Abhijith Punnappurath, Ambasamudram Narayanan Rajagopalan, Senior Member, IEEE, Sima Taheri, Student Member, IEEE, Rama Chellappa, Fellow, IEEE, and Guna Seetharaman, Fellow, IEEE, “ Face Recognition Across Non-Uniform Motion Blur, Illumination, and Pose” VOL. 24, NO. 7, JULY 2015
- [2] Andrea Bandini, Member, IEEE, and Jos’e Zariffa, Senior Member, IEEE, “Analysis of the hands in egocentric vision:A survey” [Vol 0162-8828,2020]
- [3] An-An Liu*, Ning Xu*, Wei-Zhi Nie*, Yu-Ting Su, and Yong-Dong Zhang, “Multi-Domain & Multi-Task Learning for Human Action Recognition”, IEEE Transactions on Image Processing, VOL.1057-7149, 2018.
- [4] Aouaidjia Kamel, Bin Sheng , Po Yang, Ping Li , Ruimin Shen, and David Dagan Feng , Fellow, IEEE, “Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures”, IEEE transactions on systems, man, and cybernetics: Systems, VOL.2168-2216, 2018.
- [5] Hongsong Wang, and Liang Wang, Senior Member, IEEE, “Beyond Joints: Learning Representations from Primitive Geometries for Skeleton-based Action Recognition and Detection”, IEEE Transactions on Image Processing, VOL.1057-7149, 2018.
- [6] Hossein Rahmani, Ajmal Mian and Mubarak Shah, “Learning a Deep Model for Human Action Recognition from Novel Viewpoints”, IEEE Transactions on Pattern Analysis and Machine Intelligence , VOL.0162-8828, 2018.
- [7] Jian Cheng, Haijun Liu, Feng Wang, Hongsheng Li, and Ce Zhu, Senior Member, IEEE, “Silhouette Analysis for Human Action Recognition Based on Supervised Temporal t-SNE and Incremental Learning”, IEEE transactions on image processing, Vol. 24, No. 10, October 2016.

- [8] Jiwen Lu, Member, IEEE, Venice Erin Liong, Xiuzhuang Zhou, Member, IEEE, and Jie Zhou, Senior Member, IEEE, "Learning Compact Binary Face Descriptor for Face Recognition"[Vol 0162-8828,2015]
- [9] Nour El Din Elmadany, Student Member, IEEE, Yifeng He, Member, IEEE, and Ling Guan, Fellow, IEEE, "Information Fusion for Human Action Recognition via Biset/Multiset Globality Locality Preserving Canonical Correlation Analysis", IEEE Transactions on Image Processing, VOL.1057-7149,2018.
- [10] Ping Wei, Member, IEEE, Hongbin Sun, Member, IEEE, and Nanning Zheng, Fellow, IEEE, "Learning Composite Latent Structures for 3D Human Action Representation and Recognition", VOL.1520-9210,2018.
- [11] Qiang Nie, Jiangliu Wang, Xin Wang, Yunhui Liu, Fellow, IEEE, "View-invariant Human Action Recognition Based on A 3D Bio-constrained Skeleton Model", IEEE Transactions on Image Processing, Vol. 14, No. 8, August 2019.
- [12] Renliang Weng, Jiwen Lu, Senior Member, IEEE, and Yap-Peng Tan, Senior Member, IEEE, "Robust Point Set Matching for Partial Face Recognition" [Vol 1057-7149,2016]
- [13] Sid Ahmed Walid Talha¹, Mounir Hammouche^{1,2}, Enjie Ghorbel^{1,3}, Anthony Fleury^{*1}, Member, IEEE,, Sébastien Ambellouiss, "Features and Classification Schemes for View-Invariant and Real-Time Human Action Recognition", IEEE Transactions on Cognitive and Developmental Systems, VOL.2379-8920,2018.
- [14] Sungpill Choi, Seongwook Park, and Hoi-Jun Yoo, "A Memory-efficient Hand Segmentation Architecture for Hand Gesture Recognition in Low-power Mobile Devices" [Vol 1057-7149, 2017]

- [15] Zhigang Tu, Member, IEEE, Hongyan Li, Dejun Zhang, Member, IEEE, Justin Dauwels, Senior Member, IEEE Baoxin Li, Senior Member, IEEE and Junsong Yuan, Senior Member, IEEE ,“Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition”, IEEE Transactions on Image Processing,VOL.1057-7149,2018.

APPENDIX

A. SOURCE CODE

```
# USAGE
# python train_model.py --conf config/config.json
# construct the argument parser and parse the arguments
ap = argparse.ArgumentParser()
ap.add_argument("-c", "--conf", required=True,
                help="path to the input configuration file")
args = vars(ap.parse_args())

# load the configuration file
conf = Conf(args["conf"])

for imagePath in imagePaths:
    # extract the class label from the filename
    label = imagePath.split(os.path.sep)[-2]

    # load the image, convert it to grayscale, and resize it to be a
    # fixed 64x64 pixels, ignoring aspect ratio
    image = cv2.imread(imagePath)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    image = cv2.resize(image, (64, 64))

    # update the data and labels lists, respectively
    data.append(image)
    labels.append(label)

# convert the data into a NumPy array, then preprocess it by scaling
# all pixel intensities to the range [0, 1]
data = np.array(data, dtype="float") / 255.0
```

```

# reshape the data matrix so that it explicitly includes a channel
# dimension
# the data for training and the remaining 25% for testing
(trainX, testX, trainY, testY) = train_test_split(data, labels,
    test_size=0.25, stratify=labels, random_state=42)

# construct the training image generator for data augmentation
aug = ImageDataGenerator(rotation_range=20,
    zoom_range=0.15,
    width_shift_range=0.2, height_shift_range=0.2, shear_range=0.15,
    horizontal_flip=True, fill_mode="nearest")

# initialize our gesture recognition CNN and compile it
model = GestureNet.build(64, 64, 1, len(lb.classes_))
opt = Adam(lr=conf["init_lr"],
    decay=conf["init_lr"] / conf["num_epochs"])
model.compile(loss="categorical_crossentropy", optimizer=opt,
    metrics=["accuracy"])

# train the network
H = model.fit_generator(
    aug.flow(trainX, trainY, batch_size=conf["bs"]),
    validation_data=(testX, testY),
    steps_per_epoch=len(trainX) // conf["bs"],
    epochs=conf["num_epochs"])

# evaluate the network
print("[INFO] evaluating
network...")
predictions = model.predict(testX, batch_size=conf["bs"])
print(classification_report(testY.argmax(axis=1),
    predictions.argmax(axis=1), target_names=lb.classes_))
print("[INFO] serializing label encoder...")

```

```
f = open(str(conf["lb_path"]), "wb")  
f.write(pickle.dumps(lb))  
f.close()
```

```
# USAGE
```

```
# python recognize.py --conf config/config.json
```

```
# import the necessary packages
```

```
from pyimagesearch.utils import Conf
```

```
#from pyimagesearch.notifications import TwilioNotifier
```

```
from imutils.video import VideoStream
```

```
from imutils import paths
```

```
from threading import Thread
```

```
from tensorflow.keras.preprocessing.image import img_to_array
```

```
from tensorflow.keras.models import load_model
```

```
from datetime import datetime
```

```
from datetime import date
```

```
import RPi.GPIO as GPIO
```

```
import numpy as np
```

```
import argparse
```

```
import imutils
```

```
import pickle
```

```
import time
```

```
import cv2
```

```
import os
```

```
# set the Pi Traffic Light GPIO pins
```

```
red = 9
```

```
yellow = 10
```

```
green = 11
```

```
# setup the Pi Traffic Light
```

```
GPIO.setmode(GPIO.BCM)
```

```
GPIO.setup(red, GPIO.OUT)
```

```
GPIO.setup(yellow, GPIO.OUT)
```

```
GPIO.setup(green, GPIO.OUT)
```

```
def correct_passcode(p):
```

```
# actuate a lock or (in our case) turn on the green light
```

```
GPIO.output(green, True)
```

```
# print status and play the correct sound file
```

```
print("[INFO] everything is okay :-)")
```

```
play_sound(p)
```

```
def incorrect_passcode(p, tn):
```

```
    # turn on the red light
```

```
    GPIO.output(red, True)
```

```
    # print status and play the incorrect sound file
```

```
    print("[INFO] security breach!")
```

```
    play_sound(p)
```

```
    # alert the homeowner
```

```
    hhmmss = (datetime.now()).strftime("%I:%M%p")
```

```
    today = date.today().strftime("%A, %B %d %Y")
```

```
    msg = "An incorrect passcode was entered at " \
```

```
        "{} on {} at {}".format(conf["address_id"], today, hhmmss)
```

```
    tn.send(msg)
```

```
def reset_lights():
```

```
    # turn off the lights
```

```
    GPIO.output(red, False)
```

```
    GPIO.output(yellow, False)
```

```
    GPIO.output(green, False)
```

```
def play_sound(p):
```

```
    # construct the command to play a sound, then execute the command
```

```
    command = "aplay {}".format(p)
```

```
os.system(command)
```

```
print(command)
```

```

# construct the argument parser and parse the arguments
ap = argparse.ArgumentParser()
ap.add_argument("-c", "--conf", required=True,
    help="path to the input configuration file")
args = vars(ap.parse_args())

# load the configuration file and initialize the Twilio notifier
conf = Conf(args["conf"])
#tn = TwilioNotifier(conf)

# grab the paths to gesture icon images and then initialize the icons
# dictionary where the key is the gesture name (derived from the image
# path) and the key is the actual icon image
print("[INFO] loading icons...")
imagePaths = paths.list_images(conf["assets_path"])
icons = {}

# loop over the image paths
for imagePath in imagePaths:
    # extract the gesture name (label) the icon represents from
    the # filename, load the icon, and then update the icons
    dictionary label = imagePath.split(os.path.sep)[-1].split(".")[0]
    icon = cv2.imread(imagePath)
    icons[label] = icon

# grab the top-left and and bottom-right (x, y)-coordinates for the
# gesture capture area
TOP_LEFT = tuple(conf["top_left"])
BOT_RIGHT = tuple(conf["bot_right"])

```

```
# load the trained gesture recognizer model and the label binarizer  
print("[INFO] loading model...")
```

```
model = load_model(str(conf["model_path"]))
lb = pickle.loads(open(str(conf["lb_path"]), "rb").read())

# start the video stream thread
print("[INFO] starting video stream thread...")
vs = VideoStream(src=0).start()
#vs = VideoStream(usePiCamera=True).start()
time.sleep(2.0)

# initialize the current gesture, a bookkeeping variable used to keep
# track of the number of consecutive frames a given gesture has been
# classified as
currentGesture = [None, 0]

# initialize the list of input gestures recognized from the user
# along with the timestamp of when all four gestures were entered
gestures = []
enteredTS = None

# initialize two booleans used to indicate (1) whether or not the
# alarm has been raised and (2) if the correct pass code was entered
alarm = False
correct = False

# loop over frames from the video stream
while True:
    # grab the frame from the threaded video file stream and grab the
    # current timestamp
    frame = vs.read()
    timestamp =
    datetime.now()
```

```
# resize the frame and then flip it horizontally  
frame = imutils.resize(frame, width=300)  
frame = cv2.flip(frame, 1)
```

```

# clone the original frame and then draw the gesture capture area
clone = frame.copy()

cv2.rectangle(clone, TOP_LEFT, BOT_RIGHT, (0, 0, 255), 2)

# only perform hand gesture classification if the current gestures
# list is not already full
# if len(gestures) < 4:
#     # extract the hand gesture capture ROI from the frame, convert
#     # the ROI to grayscale, and then threshold it to reveal a
#     # binary mask of the hand
roi = frame[TOP_LEFT[1]:BOT_RIGHT[1], TOP_LEFT[0]:BOT_RIGHT[0]]
roi = cv2.cvtColor(roi, cv2.COLOR_BGR2GRAY)
roi = cv2.threshold(roi, 75, 255, cv2.THRESH_BINARY)[1]
visROI = roi.copy()

# now that we have the hand region we need to resize it to be
# the same dimensions as what our model was trained on, scale
# it to the range [0, 1], and prepare it for classification
roi = cv2.resize(roi, (64, 64))
roi = roi.astype("float") / 255.0
roi = img_to_array(roi)
roi = np.expand_dims(roi, axis=0)

# do a bit of cleanup
cv2.destroyAllWindows()
vs.stop()

```

B. PUBLICATION WITH PLAGIARISM REPORT

Gesture Based Home Appliance Control System For Disabled People

Harshita A
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
harshitaakavi@gmail.com

Hansini P
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
hansinip04@gmail.com

Dr.Asha P
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
asha.cse@sathyabama.ac.in

Abstract—Home appliances can be controlled using gesture recognition for the disabled people. Gestures can be the movements or motions of human body parts. The gesture recognition can be used for many applications such as automation sign language recognition, human interactions and much more. Mostly hand gestures are used for controlling home appliances for better functionality. The proposed system shows that we shall use hand gestures to control the electronic home appliances with ease. Here concepts like deep learning, artificial intelligence, automation is used to achieve the target. These techniques are used to control the home appliances via the camera without the usage of sensors. Thus, the result of our project is to help the disabled people to operate their home appliances from a given fixed position with ease via the camera with the help of deep learning algorithms, artificial intelligence without the use of sensors and internet connectivity.

Keywords—Deep Learning, Gesture recognition, Home Appliances, Disabled People, GesturNet Algorithm.

I. INTRODUCTION

Deep learning is a major subset of the machine learning, which is a software that behaves like the neurons in the brain. There are many deep learning architectures like Convolutional neural network (CNN), Recurrent neural network (RNN), Artificial neural network (ANN), Deep neural network (DNN) which help us in the fields like computer vision, gesture recognition, natural language processing, audio recognition and more. The first layer is called the input layer and the last layer is called the output layer, in between these layers there maybe one or more layers called the hidden layers which contains set of many neurons. These neurons are connected with each other in the following layers. The Deep learning algorithms are all very effective in various fields, for example these are 41% more accurate than the ML algorithms in image classification, 27% more in facial recognition and 25% more in voice recognition[8].

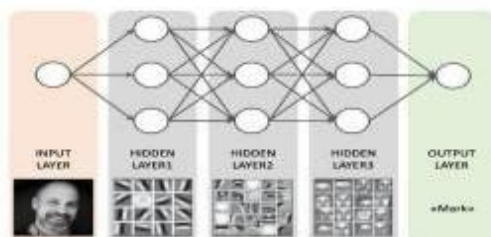


Fig 1.Types of layers in Deep Learning

The neural networks can be mainly classified into two types, shallow neural network, which has only one hidden layer between the input and output layers, the next is deep neural network, which consists of more than one hidden layer between the input and output layers respectively. The Feed forward neural network is a type of artificial neural network, where the information can flow only in one direction and that direction is forward, it flows through input layer followed by many of the hidden layers and then the output layer, it has no loops. The Recurrent neural network (RNN) is also a artificial neural network and is a multilayered neural network as it has multiple layers, it also contains loop. The Convolutional neural network (CNN) is also a multilayered neural network and is mostly used on unstructured dataset so as to extract complex features at every single level till the output layer to get the desired output.

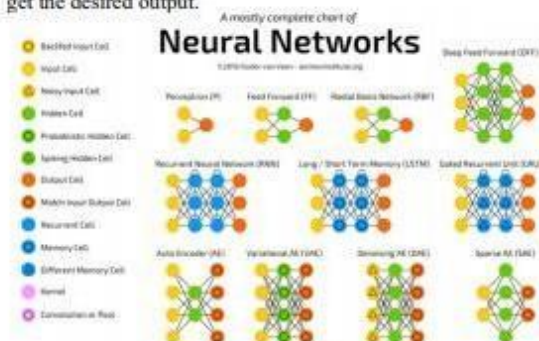


Fig 2.Types of Neural Networks

Artificial intelligence is a branch of computer science, which is used for building smart and intelligent machines to perform some specified tasks which is done by human. The major contribution of artificial intelligence is to develop computer functioning like the human intelligence as the name itself says such as learning, problem analysis and reasoning. It is used in various fields like gaming, speech recognition, vision systems, robots etc. There are mainly two different types of Artificial intelligence, weak artificial intelligence which is used to perform one job at one time and strong artificial intelligence which is used to execute more complex features.

Elderly or physically challenged people face difficulty in operating the electronic home appliances as they face problem to move from one place to another for operating the appliances. Keeping this in mind we have proposed this

system to help the disabled and elderly people to control their home appliances from a fixed position in a room via camera with ease without any usage of sensors and also works in the absence of internet connectivity.

II. RELATED WORKS

According to Sungpill Choi et al.[15], the next generation of mobile devices will be Hand gesture recognition as it is regarded as new Human Computer Interaction (HCI) technology. Before more computational power and huge memory was required to implement the hand segmentations for obtaining hand, which resulted in failure for the real-time interaction with mobile devices along with users. Therefore, in this paper memory efficient and low latency hand segmentation architecture is used to obtain natural hand gesture recognition. A streaming hand contour tracing system and fast contour filling system is used to get high memory and less latency as desired.

According to An-An Liu* et al.[1], the paper deals with the essential part of human recognition such as Domain-invariant, multiple actions correlation is discovered for a discriminative. Therefore to solve the above problem multi-domain learning and multi-tasking learning methods is proposed (MDMTL). Focusing on this paper, the co-embed multi-domain (multi-modality & multi-view) data converting it to a single general space for the learning of discriminative feature using sparse transfer learning-based method is presented. Moreover, multi-task learning framework consists of visual feature learning incorporated into it, with the regularization term of Frobenius-norm and the sparse constraint term, for task relatedness-induced feature learning and joint task modeling. The first supervised framework is MDMTL which realized multi-domain action recognition from both task modelling and domain-invariant feature learning.

According to Sid Ahmed Walid Talhal et al.[14] Ambient Assisted Living (AAL) is the field where Human Action recognition (HAR) is used largely which creates an interaction/communication between computers and humans. In these systems, people are asked to act naturally. To make the interaction fluent the following algorithm needs to adapt and interact as quick as required. In this paper, proposed is an ideal method provided by RGB-D cameras based on information which are skeleton, to improvise the already existing systems or algorithms. By applying the given above concept it is able to detect early action recognition and is very robust to the variability of viewpoint. A new descriptor named Body Directional Velocity is offered and a classification based on real-time is executed to reach the desired goal, which is an apt method to perform various human action recognition with the help of RGB-D sensors. The two major challenges for robotic applications are focused here they are, the early recognition action property and the viewpoint changes due to robustness.

According to Aouaidia Kamel et al.[3] By using data of postures and depth maps using convolutional neural networks (CNN) an approach has been developed for human action recognition which is an action-fusion. For action representation, two input descriptors are being used. The first input which accumulates subsequent human action depth maps is called as the depth motion image, while the next input, which represents the body joints motions with time is a proposed moving joints descriptor. In accurate action classification the feature extraction can be maximized when different inputs are being trained with three channels namely CNN. Depth motion images (DMIS) trains the first channel, both DMIS and moving joint descriptors together train the second channel, and lastly the moving joint descriptors alone used to train the third channel. In the final action classification all the predictions of action obtained by all the three CNN channels are combined together. To maximize the score of the action which is right, have proposed various score operations that include fusion. The inference that we get from the experiments is that the result of the output is better when the three channels are fused rather than the result of only one channel or only two channels being fused. The method proposed was evaluated based on three main public datasets: 1) microsoft action 3D dataset (msraction3d); 2) multimodal human action dataset from university of Texas at dallas; and 3) multimodal action dataset (mad) dataset.

According to Andrea Bandini et al.[2], the paper is based on the egocentric vision application. The datasets are collected using wearable cameras and focus only the study of hands is focused in various perspectives like their hands and other parts being localized within the image and recognize the actions involved in our hands. In this paper, the literature tells about egocentric vision used for hand gestures, the existing methods can be categorized into: localization; interpretation; and application. Adding to this, hand-based annotations with most prominent datasets lists are provided.

III. INFERENCE FROM RELATED WORK

The inference we get from the above literature survey is that those existing systems had only hand recognition without any application involved in it. Failures of sensors led to inefficient automation and even if there were home appliance controlling system they couldn't be used if there was no internet connectivity. Hence making the systems insecure. Due to the misplacement and low quality of camera which results in low accuracy.

IV. PROPOSED SYSTEM

In the proposed system, we make use of deep learning algorithms and artificial intelligence techniques to avoid the usage of sensors which are expensive and need more maintenance. Here the home appliance can be controlled even without the internet connectivity, so it is beneficial in areas where there is internet connectivity such as rural areas. Hence the elderly and disabled people can use their home appliance with ease under our proposed system.

A. Modules Implemented

There are mainly four modules to be implemented. They are:

- Dataset collection
- Hand segmentation
- Training with Gesturennet
- Compare the gestures to switch on and off the devices

In the dataset collection module, the deep learning algorithms help in training mostly in the computer vision tasks and for training we will need various set of different data, these data which are labeled provide better performance for the model. Dataset collection is a never ending process that is the data should be fetched constantly from time to time to improve the performance of the model even after the creation of the model into real-time application. Here data is the most precious component for deep learning techniques. Here we use mainly two steps to collect data they are, scraping from the web and third party.

In the hand segmentation module, efficient tracking and segmentation of the hand naturally improves the performance of gesture recognition otherwise the following challenges may cause issues such as skin color variation and detection, complex background adjustment, adjusting various lighting effects and more. The most powerful descriptor for the gesture recognition of a object is the color. In the segmentation process, we use the information about color irrespective to rotation of the hand. The characteristics of color can be seen through brightness, hue, saturation than the percentage of the primary red, green, blue colors. Thus there are mainly three techniques for various color spaces to robust the segmentation and detection of hand.



Fig.3 Hand Segmentation

In the training with Gesturennet module, the images are captured by a normal or wearable camera such that the background of the images are identical because it is effective and easy in the analysis of hand region from the original image by background subtraction method although in some cases, moving objects are also included as the outcome of background subtraction. The skin color is used for the differentiation of hand region; hence it is measured by HSV model. The value of skin color in HSV model is 315, 94 and 37. The output of the hand will be taken in as binary image with white pixels which belongs to hand and black pixels which belongs to the background. Addition to it, the distance transform is used for palm point. The palm point is marked in green color for that we use a new deep learning algorithm called the Gesturennet. By using this algorithm, we can train different gestures to yield high accuracy. The input is given from the video stream and recognizes the model.

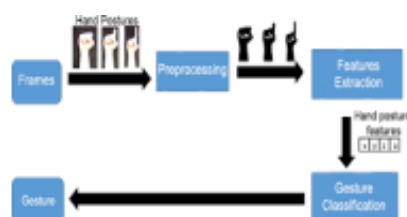


Fig.4 Gesturennet Algorithm

In the Compare the gestures to switch on and off the devices module, we are using appliances such as light and fan to on and off without the use of internet they can be made on and off. Here the incoming gestures are monitored and used to operate the switch to on and off. Hence will be useful for the physically challenged and aged people with the help of automatic control of home appliance using artificial intelligence.

B. System Architecture

In this project, the initial step is to collect the datasets of gestures by taking different pictures of hand, and then these datasets are trained by raspberry pi to identify a face for switching ON/OFF the appliances. So, raspberry pi will be configured by raspian OS, and OpenCV libraries are used to train the dataset models to recognize gestures. A camera will be connected by PC which lively stream the gestures, which is shown in the front of the camera. These gestures are captured and analyzed to switch ON/OFF appliance. From the dataset collection we train the models to get the accurate output. A driver circuit relay is connected to raspberry pi which is control high voltage to switch ON/OFF appliances.



Fig. 5 Gesture Recognition

When the face input is given in camera, the face is recognized first, then compares with gestures to switch based ON/OFF appliances. So, by this algorithm we can control home appliances without using sensors with Artificial Intelligence.

V. CONCLUSION

We would conclude the paper, stating that the system proposed helps the disabled and elderly people to control the electronic home appliances with ease without any internet connectivity. We avoided the usage of sensors which are expensive and even need proper maintenance and replace them with the help of deep learning algorithms and artificial techniques. As the internet connectivity is not needed the proposed system can reach out to much more people where internet connectivity is not possible to establish or not available, for example rural areas. The gestures made by the person from a fixed position is captured by camera which is available at the room and gets switched on or off the required device respectively.

REFERENCES

- [1] An-An Liu*, Multi-Domain & Multi-Task Learning for Human Action Recognition, IEEE Transactions on Image Processing, VOL.1057-7149, 2018.
- [2] Andrea Bandini, Analysis of the hands in egocentric vision: A survey, IEEE, 2020.
- [3] Aouaidjia Kamel, Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures ,IEEE transactions on systems, man, and cybernetics: Systems, 2018.
- [4] Hongsong Wang, Beyond Joints: Learning Representations from Primitive Geometries for Skeleton-based Action Recognition and Detection, IEEE Transactions on Image Processing, 2018.
- [5] Hossein Rahmani, Learning a Deep Model for Human Action Recognition from Novel Viewpoints, IEEE Transactions on Pattern Analysis and Machine Intelligence ,VOL.0162-8828,2016.
- [6] <https://www.alamy.com/stock-photo/number-1-hand-gesture.html>
- [7] <https://www.bouvet.no/bouvet-deler/an-introduction-to-deep-learning>
- [8] <https://www.guru99.com/deep-learning-tutorial.html>
- [9] <https://jenstirrup.com/2017/08/01/how-do-you-evaluate-the-performance-of-a-neural-network-focus-on-azureml/>
- [10] Jian Cheng, Silhouette Analysis for Human Action Recognition Based on Supervised Temporal t-SNE and Incremental Learning ,IEEE transactions on image processing, Vol. 24, No.10October2016.
- [11] Nour El Din Elmadany, Information Fusion for Human Action Recognition via Biset/Multiset Globality Locality Preserving Canonical Correlation Analysis ,IEEE Transactions on Image Processing, VOL.1057-7149, 2018.
- [12] Ping Wei, Learning Composite Latent Structures for 3D Human Action Representation and Recognition, IEEE, VOL.1520-9210, 2018.
- [13] Qiang Nie, View-invariant Human Action Recognition Based on A 3D Bio-constrained Skeleton Model ,IEEE Transactions on Image Processing, Vol. 14, No. 8, August 2019.
- [14] Sid Ahmed Walid Talha1, Features and Classification Schemes for View-Invariant and Real-Time Human Action Recognition ,IEEE Transactions on Cognitive and Developmental Systems, VOL.2379-8920, 2018.
- [15] Sungpill Choi, Seongwook Park, and Hoi-Jun Yoo , A Memory-efficient Hand Segmentation Architecture for Hand Gesture Recognition in Low-power Mobile Devices",IEEE, 2017.
- [16] Zhigang Tu, Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition ,IEEE Transactions on Image Processing, VOL.1057-7149, 2018.

PLAGIARISM REPORT

Home_appliances_paperwork

ORIGINALITY REPORT

14%

SIMILARITY INDEX

11%

INTERNET SOURCES

10%

PUBLICATIONS

5%

STUDENT PAPERS

Gesture And Face Recognition Based Locker Control System

Hansini P
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
hansinip04@gmail.com

Harshita A
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
harshitaakavi@gmail.com

Dr.Asha P
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai, India
asha.cse@sathyabama.ac.in

Abstract—Locker control system with the help of gesture recognition and additional security is given with the use of face recognition. Gestures are nothing but the motion of the human body parts. The face detection or recognition is used for various applications such as, in biometrics, photography, marketing, emotional inference etc. The gesture recognition can also be used for various applications such as automation sign language recognition, human interactions and much more. Our proposed system shows the use of gesture and face recognition to unlock the locker when the registered face is detected along with the required gesture recognition. If there occurs any unauthorized attempt to open the locker, then the owner of the locker would receive an alert through the mobile application. Thus, the result of our project is to develop a secured locker control system simply using face recognition along with gesture recognition through artificial intelligence techniques and without the usage of any sensors.

Keywords—Deep Learning, Gesture recognition, Mobile application, Face recognition, GesturNet Algorithm, MobileFaceNet Algorithm, Artificial Intelligence.

1. INTRODUCTION

Deep learning is a significant subset of the Artificial Intelligence, which is a product that carries on like the neurons in the cerebrum. There are numerous deep learning designs like Convolutional neural organization (CNN), Recurrent neural organization (RNN), Artificial neural organization (ANN), Deep neural organization (DNN) which help us in the fields like PC vision, signal acknowledgment, regular language handling, sound acknowledgment and that's just the beginning. The main layer is known as the information or the input layer and the last layer is known as the output or the yield layer, in the middle of these layers there perhaps are at least one or more layers called the hidden layers which contains set of numerous neurons. These neurons are associated with one another in the accompanying layers. The Deep learning calculations are largely exceptionally powerful in different fields, for instance these are 41% more precise than the ML calculations in picture arrangement, 27% more in facial acknowledgment and 25% more in voice acknowledgment [8].



Fig 1. Classification of Neural Networks

The neural networks can be basically characterized into two kinds, shallow neural networks, which has just one hidden layer between the information or input and yield or output layers, the following is deep neural networks, which comprises of more than one hidden layer between the information and yield layers separately. The Feed forward neural network is a kind of artificial neural networks, where the data can stream just in one way and that bearing is forward, it courses through input layer followed by numerous individuals of the hidden layers and afterward the yield layer, it has no loops. The Recurrent neural network (RNN) is likewise an artificial neural network and is a diverse neural network as it has multiple layers, it additionally contains loops. The Convolutional neural network (CNN) is also a multilayered neural network and is for the most part utilized on unstructured dataset in order to separate complex highlights at each and every level till the yield layer to get the ideal yield.

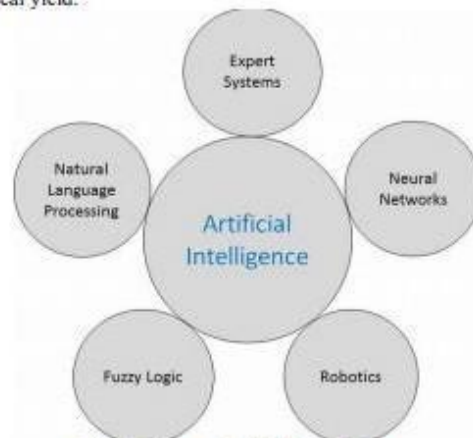


Fig 2. Usage of Artificial Intelligence

Artificial Intelligence is a part of computer science, which is utilized for building intelligent and smart machines to play out some predetermined tasks which are performed by human. The significant commitment of artificial intelligence consciousness is to create computer working like the human insight as the actual name says like learning, issue examination and thinking. It is utilized in different fields like gaming, discourse acknowledgment, vision frameworks, robots and so forth. There are mostly two distinct sorts of Artificial intelligence, weak artificial intelligence which is utilized to perform one occupation at one time and strong

artificial intelligence which is utilized to execute more complex features.

The React Native system is an upcoming mobile arrangement and is viewed as what's to come of cross-stage mobile application improvement. React Native is an open-source system which implies all documentation identified with this innovation is open for everybody and is accessible for nothing to everybody in React Native group, that permits to assemble mobile application with just JavaScript. It was presented by Jordan Walke, a Facebook software engineer, as another innovation for less complex turn of events and a superior client experience. The fundamental particular of this framework is that React Native applications work very much like native applications. They don't contrast from applications based on Java, Objective-C or then again Swift and they utilize a similar UI building blocks as native Android or iOS applications. Yet, with React Native, assembling a mobile application is a lot quicker and more affordable. The main benefit of React Native is reusability of code and pre-built parts. The client interfaces planned in React Native are more responsive, decline load time, and give a smoother charge. In addition to these listed features, the React Native structure is as yet progressing, so it may not have a few components in the central structure. Hence, React Native gives two sorts of third-party plugins such as native modules and JavaScript modules.

Most of the locker control systems nowadays are passcode oriented, which can be easily hacked. Hence in our proposed system, we have developed a secure locker control system through face and gesture recognition without any usage of sensors and the owner of the locker is notified whenever there is an unauthorized access to the locker through mobile app.

II. RELATED WORKS

According to Renliang Weng et al.[1], in the course of recent many years, various face acknowledgment strategies have been proposed in computer vision, and a large portion of them utilize comprehensive face pictures for individual identification. In some true situations particularly some unconstrained conditions, human faces may be blocked by different articles and it is hard to get completely encompassing face pictures for acknowledgment. To address this, they proposed another incomplete face acknowledgment way to deal with perceive people of interest from their halfway faces. Given a couple of display picture and test face fix, we initially identify key focuses and concentrate their neighborhood textural highlights. At that point, propose a vigorous point set coordinating (RPSM) technique to discriminatively coordinate with these two removed neighborhood include sets, where both the textural and geometrical data of nearby highlights sets are unequivocally utilized for coordinating at the same time. In conclusion, the similitude of two faces is changed over as the distance between these two adjusted capabilities. Test results on four public face datasets show the adequacy of the proposed approach.

According to Abhijith Punnappurath et al.[2], existing strategies for performing face acknowledgment within the sight of obscure depend on the convolution model and can't

deal with non-uniform obscuring circumstances that every now and again emerge from slants and turns close by held cameras. In this paper, they have proposed a strategy for face acknowledgment within the sight of room shifting movement obscure containing self-assertively molded pieces. We model the obscured face as a raised blend of geometrically changed examples of the engaged display face and show that the arrangement of all pictures got by non-consistently obscuring a given picture frames a curved set. They have initially proposed a non-uniform haze vigorous algorithm by utilizing the suspicion of an inadequate camera direction in the camera movement space to assemble an energy work with $l1$ -norm imperative on the camera movement. The system is then reached out to deal with illuminating varieties by misusing the way that the arrangement of all pictures got from a face picture by non-uniform obscuring and changing the enlightenment frames a bi-convex set. At last, they have proposed a rich expansion to likewise represent varieties in present.

According to Jiwen Lu et al.[3], binary component descriptors like local binary patterns (LBP) and its varieties have been broadly utilized in many face acknowledgment frameworks because of their fantastic strength and solid discriminative force. Nonetheless, most existing binary face descriptors are hand-created, which require solid earlier information to design them by hand. In this paper, they have proposed a reduced compact binary face descriptor (CBFD) include learning strategy for face portrayal and acknowledgment. Given each face picture, we first concentrate pixel difference vectors (PDVs) in local fixes by registering the contrast between every pixel and its adjoining pixels. At that point, we become familiar with an element planning to project these pixel contrast vectors (PDVs) into low-dimensional double vectors in an unsupervised way, where 1) the change of all binary codes in the preparation set is amplified, 2) the misfortune between the first genuine esteemed codes and the learned binary codes is limited, and 3) binary codes equally circulate at each scholarly canister, so the excess duplicate data in PDVs is taken out and smaller binary codes are acquired. In conclusion, we group and pool these binary codes into a histogram include as the last portrayal for each face picture. In addition, they have proposed a coupled CBFD (C-CBFD) strategy by lessening the methodology hole of heterogeneous appearances at the element level to make our technique appropriate to heterogeneous face acknowledgment. Broad trial results on five generally utilized face datasets show that our strategies beat cutting edge face descriptors.

According to Sungpill Choi et al.[4], the up and coming age of cell phones will be Hand motion acknowledgment as it is viewed as new Human Computer Interaction (HCI) innovation. Before more computational force and immense memory was needed to execute the hand divisions for getting hand, which brought about disappointment for the constant association with cell phones alongside clients. Along these lines, in this paper memory effective and low dormancy hand division engineering is utilized to get common hand motion acknowledgment. A streaming hand contour tracing system

and fast contour filling system is utilized to get high memory and less inactivity or less latency as wanted.

According to An-An Liu* et al.[5], the paper manages the fundamental piece of human acknowledgment, for example, Domain-invariant, various activities connection is found for a discriminative. Consequently, to tackle the above issue multi-domain learning and performing multi-tasking learning techniques is proposed (MDMTL). Focusing on this paper, the co-implant multi-domain (multi-modality and multi-view) information changing it over to a solitary general space for the learning of discriminative component utilizing scanty exchange learning-based technique is introduced. In addition, perform multiple tasks learning structure comprises of visual component learning fused into it, with the regularization term of Frobenius-norm and the inadequate imperative term, for task relatedness-initiated include learning and joint assignment demonstrating. The originally administered structure is MDMTL which acknowledged multi-domain activity acknowledgment from both task modelling and domain-invariant component learning.

III. INFERENCE FROM RELATED WORK

The inference we get from the above literature survey is that those existing systems either had face recognition or gesture recognition for their application but there was no combination of both gesture and face recognition. The system is inefficient due to poor optimization. There are security issues as anyone can make the gestures actions involved to unlock. The procedures are difficult to follow as it is hand crafted and not recommended.

IV. PROPOSED SYSTEM

In the proposed system, we have developed a locker control system with the combination of both face and hand gesture recognition which is more secured than the passcode locker control system which can be easily hacked. Here we use the deep learning algorithms and artificial intelligence techniques instead of using sensors which are expensive and need more maintenance.

A. Modules Implemented

There are five main modules to be implemented. They are:

- Dataset collection
- Face detection module
- Face recognition module
- Hand segmentation
- Mobile application development

In the dataset collection module, the deep learning algorithms help in preparing for the most part in the computer vision errands and for preparing we will require different arrangement of various data, these data which are marked give better execution to the model. Dataset collection is an endless interaction that is the information ought to be gotten

continually from time to time to adverse the exhibition of the model even after the formation of the model into real-time application. Here data is the most valuable part for deep learning strategies. Here we utilize basically two stages to gather data they are, scraping from the web and the third party.

There are many face detection applications in which the "Ultra-Light-Fast-Generic-Face-Detector" is intended for general purpose which can be operated in low-power computing gadgets and is material to both Android and iOS software as well as PCs (CPU and GPU). This model is designed for low-power devices or edge computing gadgets, hence is known as real-time ultra-lightweight universal face detection model. The initial phase in facial recognition is to detect human faces in digital images, and an ideal face recognition model can be assessed by how rapidly and precisely it performs. The Face-Detector-1MB preparing measure utilized a VOC dataset produced by the WIDER FACE dataset, a face detection benchmark. WIDER FACE was delivered in 2015 and comprises of 32,203 pictures and 393,703 face bounding boxes with a serious level of fluctuation in scale, pose, appearance, impediment and light. The 1MB lightweight model comprises of two versions one is version-slim which has marginally quicker improvement, and the other one is the version-RFB which is an altered RFB module for higher accuracy.



Fig.3 Flow diagram for ultra-light face detector

MobileFaceNet is utilized here for recognizing face as it is more exact in characterizing face. MobileFaceNet acquires exactness up to 93.05 percent exactness on perceiving faces in the AgeDB dataset and about 99.28 percent on labelled faces in the wild (LFW) dataset as it is a neural network. More than a million parameters are utilized by the network taking just 24 milliseconds to execute and deliver outcomes on a Qualcomm Snapdragon processor. We can contrast this presentation with correctness of 98.70 percent and 89.27 percent for ShuffleNet, which has numerous more parameters and takes somewhat more time to execute on the CPU. MobileNetV2 architecture partly motivated or contributed for our MobileFaceNet architecture that is, our fundamental building blocks is obtained from the residual obstacles proposed in

MobileNetV2. The PReLU is used as the non-linearity, which is more qualified for facial confirmation than the ReLU. The MobileNetV1, ShuffleNet, and MobileNetV2 are used as the baseline models. To get a reasonable outcome when compared among the listed that is, all MobileFaceNet models and baseline models are trained 34 on CASIA-Webface dataset without any preparation by ArcFace loss.

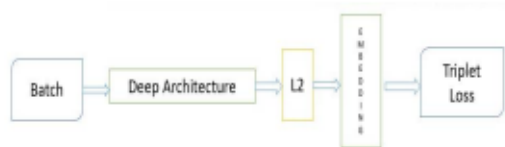


Fig.4 MobileFaceNet Architecture

In the hand segmentation module, productive tracking and division of the hand normally improves the exhibition of gesture acknowledgment in any case the accompanying difficulties may cause issues, for example, skin color variation and detection, complex background adjustment, changing different lighting impacts and more. The most remarkable descriptor for the gesture recognition of an item is the tone. In the segmentation interaction, we utilize the data about shading independent to pivot of the hand. The attributes of color can be seen through splendor, hue, immersion than the level of the essential red, green, blue tones. In this way there are for the most part three methods for different color spaces to strong the segmentation and detection of hand.

The JavaScript code is assembled with the help of framework called the react native which forms a chain of command for the UI components. It has a bunch of segments for the two iOS and Android stages to construct a versatile mobile application with a native look and feel. There has been a phenomenal developmental growth for the mobile development. React Native is by all accounts a suitable answer for building top notch applications in a brief timeframe with a similar presentation and client experience norms that native applications facilitate. The architecture of React Native causes us in organizing a venture for a multiplatform mobile application, keeping the rational of the business in a reusable and kept up sub-module. Respond Local uses various components to make a productive, reliable and reusable visual character for the applications.

B. System Architecture

In our project, we have planned a hand gesture and face recognition secured locker control system. The collection of datasets of all the faces of the registered people living as a family is the initial advancement to be done and only if both the face of the person and hand gesture which was registered is recognized will the next step be executed. The subsequent stage is to gather the dataset of various gestures by taking pictures of hands and gathering from different assets from the web. In this way, both these datasets are then trained by raspberry pi or the pc to distinguish the appearances and specific gestures in request to turn open and close the locker. At whatever point the camera gives a face information, and afterward acknowledges the hand gesture information just if the face is perceived and at that point analyzes regarding whether to open or close the locker or report as theft due to unauthorized face accessing the locker and sends a mobile notification to the owner's mobile using the mobile application created.

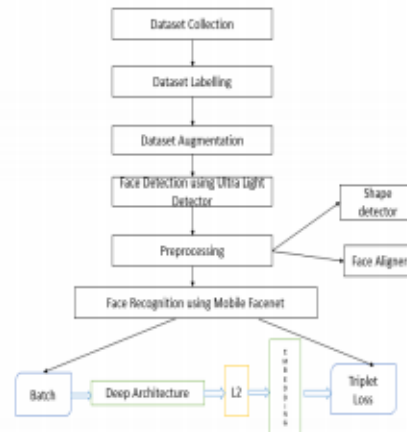


Fig. 5 Face Recognition architecture Diagram

V. CONCLUSION

We would conclude the paper, stating that the system proposed helps in providing more secured locker control system than those which are existing with the help of face and hand gesture recognition. We have avoided the use of sensors as they are expensive and even need proper maintenance, so we have replaced them with the help of deep learning algorithms and artificial techniques.

REFERENCES

- [1] An-An Liu*, Multi-Domain & Multi-Task Learning for Human Action Recognition, IEEE Transactions on Image Processing, VOL.1057-7149, 2018.
- [2]Abhijith Punnappurath, Ambasamudram Narayanan Rajagopalan, Senior Member, IEEE, Sima Taheri, Student Member, IEEE, Rama Chellappa, Fellow, IEEE, and Guna Seetharaman, Fellow, IEEE, "Face Recognition Across Non-Uniform Motion Blur, Illumination, and Pose", VOL. 24, NO. 7, JULY 2015
- [3]<https://www.educba.com/classification-of-neural-network/>
- [4] <https://www.guru99.com/deep-learning-tutorial.html>
- [5]https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_research_areas.htm
- [6] Jiwen Lu, Member, IEEE, Venice Erin Liong, Xiuzhuang Zhou, Member, IEEE, and Jie Zhou, Senior Member, IEEE, "Learning Compact Binary Face Descriptor for Face Recognition", [Vol 0162-8828,2015]

[7]Renliang Weng, Jiwen Lu, Senior Member,IEEE, and Yap-Peng Tan, Senior Member,IEEE, "Robust Point Set Matching for Partial Face Recognition",IEEE[Vol 1057-7149,2015]

[8]Sunpill Choi, Seongwook Park, and Hoi-Jun Yoo , A Memory-efficient Hand Segmentation Architecture for Hand Gesture Recognition in Low-power Mobile Devices",IEEE, 2017.

PLAGIARISM REPORT

Face_locker_paperwork_updated

ORIGINALITY REPORT

16%	11%	10%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS