2
3
4
5
6
7
8

VIRTUAL MOUSE

A project phase I report submitted in partial fulfilment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science & Engineering

Submitted by

ASWIN SIVADAS
ASHUAL JESON
ADHIL MUHAMMAD K
ASWIN C

Federal Institute of Science And Technology (FISAT)®
Angamaly, Ernakulam

Affiliated to

APJ Abdul Kalam Technological University
CET Campus, Thiruvananthapuram
December, 2022

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY
(FISAT)
Mookkannoor(P.O), Angamaly-683577

CERTIFICATE

This is to certify that the project phase I report for the project entitled "**VIRTUAL MOUSE**" is a bonafide report of the project presented during VIIth semester (CSD415 - Project Phase I) by **ASWIN SIVADAS(FIT19CS036)**, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology (B.Tech) in Computer Science & Engineering during the academic year 2022-23.

	Anuranj P	**Dr. Jyothish K John**
Staff in Charge	Project Guide	Head of the Department

ABSTRACT

The PC mouse is one of the wondrous developments of people in the field of Human-Computer Interaction (HCI) innovation. In the new age of innovation, a remote mouse or a contactless mouse actually utilizes gadgets and isn't liberated from gadgets completely, since it utilizes power from the gadget or might be from outside power sources like battery and gain space and electric power, likewise during COVID pandemic it is encouraged to make social separating and keep away from to contact things which gave by various people groups

This research introduces a novel method for controlling mouse movement with a real-time camera. Adding more buttons or repositioning the mouse´s tracking ball are two common ways. Instead, we recommend that the hardware be redesigned. Our idea is to employ a camera and computer vision technologies to manage mouse tasks (clicking and scrolling), and we demonstrate how it can do all that existing mouse devices can. This project demonstrates how to construct a mouse control system.

Contribution by Author

I examined each document that was related to the base paper that was chosen for the project's adaptation and improvement. I read through a variety of reference sources to get a better understanding of how we may approach the provided issue statement. In order to help with the implementation of the same and support my team during the early stages of development, I researched opencv and how it can be utilised to construct the gesture-based input system. In addition to the foundational work, research was done. I conducted research on the reference paper in addition to the basis work, from which I learnt how to address any implementation-related concerns

Aswin Sivadas

Contribution by Author

I analyzed a few papers related to the chosen base paper to which we came to a consensus about. I went through numerous reference papers as well as to get an even clearer idea of how we can approach this problem statement at hand. To help with the implementation of the same, I learned the basics of React and open cv so that I could help my team with the first phase of development of things.

Aswin.c

Contribution by Author

I studied several papers related to the main paper that we agreed upon as a team, and also looked at additional reference papers to gain a better understanding of how to approach the problem we are trying to solve. In order to assist with implementing the solution, I learned the basics of React so that I could support my team during the initial development phase. It's important to collaborate and get feedback from others to make sure everyone is on the same page and that the approach is effective

Ashual Jeson

Contribution by Author

I researched documents related to the base paper chosen for adapting and improving the project. To gain a better understanding of how we might approach the given issue statement. In order to aid in the implementation of the gesture-based input system and assist my team with the early stages of development, I researched opencv and how it can be used to implement it. In addition to the foundational work, research was carried out. In addition to the foundational work, research on the reference paper was carried out.

Adhil Muhammad K

ACKNOWLEDGMENT

Behind every achievement lies an unfathomable sea of gratitude to the almighty, without whom it would ever have come into existence I can barely find words to express all the wisdom, love, and support from the almighty.

I like to express my utmost gratitude to **Dr. Manoj George**, Principal, Federal Institute of Science and Technology, Angamaly. I am fortunate to be blessed with the guidance and encouragement of **Dr. Jyothish K** John, Head of Department, Computer Science and Engineering, FISAT, Angamaly.

It gives me immense pleasure to express my sincere and wholehearted sense of gratitude to **Mr.Anuranj.p** for their valuable and untiring guidance and supervision throughout the tenure of my work. To derive benefits from their in-numerous experiences is a matter of great privilege for me. I also take this opportunity to express my sincere thanks to all the staff in the computer science department, who extended their wholehearted cooperation, and moral support, and rendered ungrudging assistance whenever and wherever the need has arisen. I am very much thankful to them.

Aswin Sivadas
Aswin C
Ashual Jeson
Adhil Muhammad K

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

As computer technology continues to develop, people have smaller and smaller electronic devices. Increasingly we are recognizing the importance of human computing interaction (HCI) and in particular vision-based gesture and object recognition. In our project, we propose a novel approach that uses a video device to control the mouse system(Mouse tasks).

We employ several image-processing algorithms to implement this. A virtual

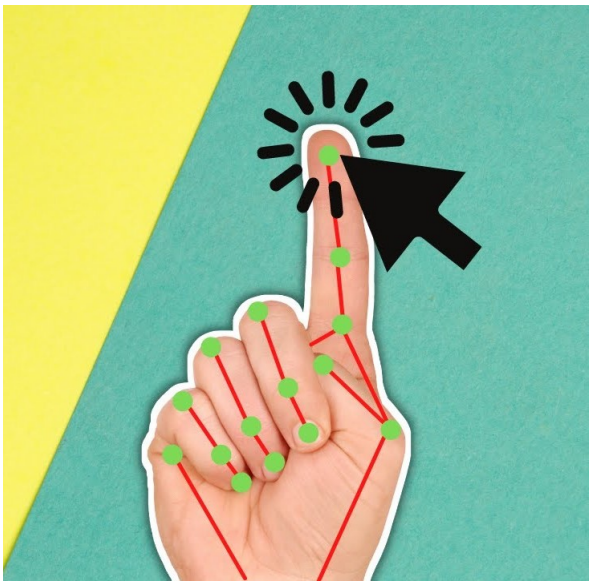


Figure 1.1: Virtual Mouse

mouse is a device that allows a user to control a computer, laptop, or smart pad without physically touching it. Instead, the user simply makes gestures in front of an infrared (IR) camera, which is modeled as a pen and connected to the device through a Human Interface Device (HID) and an I2C interface. This allows the device to act as a mouse and a marker, allowing it to be used in classrooms and conferences as a replacement for traditional wired mice, chalk boards, and markers. The device is also able to track the position of the user's gestures and convert them into X and Y coordinates, which are then used to move the mouse pointer to the desired location. This technology can be implemented using a high-end microcontroller and an IR camera, and can be configured to use a WII remote as a virtual marker with an IR source. However, this method may have some delays due to the use of software to process and obtain coordinates, and it can only function as a marker.

The virtual mouse device is a useful tool for those who want to interact with a computer or other system in a more intuitive and natural way. It can be particularly helpful in educational settings, where it allows teachers to demonstrate concepts and interact with students without the need for physical boards and markers. It can also be useful in conference settings, where it allows attendees to collaborate and share ideas without the need for traditional mouse and keyboard input devices. Additionally, the virtual mouse can be a convenient option for those who prefer to use gestures instead of physical buttons and levers to control their devices. Overall, the virtual mouse is a versatile and innovative tool that offers many benefits for users in a variety of settings.

1.2 Problem Statement

The virtual mouse using hand signal structure could in like manner be familiar with beat issues inside the spot like things where there isn't any space to use a genuine mouse and set up for individuals who have issues in their grip and don't appear, to be prepared to manage a real mouse. Moreover, in the COVID circumstance, it isn't safeguarded to include the devices reaching them as an eventual outcome of its intention to achieve what is happening of spread out of the disease by reaching the contraptions, that the projected AI virtual mouse could in like manner be adjusted vanquished these issues since hand sign and hand Tip disclosure is used to manage the device mouse limits by using a camera or a characteristic camera like a webcam.

While using a remote or a Bluetooth mouse, a couple of devices especially like the mouse, the contraption to connect with the pc, and besides, battery to drive the mouse to control a used, So all through this, the client uses his/her natural camera or visual camera and usages his/her hand movements to manage the PC mouse action.

1.3 Objective

This project urges substitute to the common and obsolete mouse construction to perform and the board as far as possible, and this could be accomplished with the assistance of an inside net camera that gets the hand developments and hand tip then, at that point, processes these lodgings to play out the specific mouse performs like left click, right snap, and investigating perform.

The goal is to manage computers and other devices with gestures rather than pointing and clicking a mouse or touching a display directly. It is believed that the approach can make it not only easier to carry out many existing chores but also take on trickier tasks such as creating 3-D models, browsing medical imagery during surgery without touching anything.

Reduce cost of the hardware.

Chapter 2

Literature Review

2.1 Paper1

On-Air Hand-Drawn Doodles for IoT Devices Authentication During COVID-19

This research paper proposes a new method for authenticating human interaction with Internet of Things (IoT) devices using hand gestures in the air to create virtual hand-drawn passwords. This approach is particularly relevant during the COVID-19 pandemic as it allows for secure authentication without physical contact. The proposed method utilizes a computer vision technique with a single camera, two lightweight deep CNN models, and a Kalman filter for signal processing to correct the path of the drawn line in the air.

The proposed method for authenticating human interaction with IoT devices using hand gestures in the air is intended to be a simple and secure alternative to traditional authentication methods. By using hand gestures to create a virtual hand-drawn password, users are able to interact with the device in a natural and intuitive way. The method utilizes a computer vision technique with a single camera and two lightweight deep CNN models to accurately detect and interpret the hand gestures.

One potential benefit of the proposed method for authenticating human interaction with IoT devices using hand gestures in the air is its potential to improve accessibility for users with disabilities. Traditional authentication methods, such as passwords or fingerprint scanners, may be difficult or impossible for some users to access. By using hand gestures as an alternative method, more users may be able to successfully authenticate and interact with the device.

In addition, the proposed method may be more convenient for users in certain situations. For example, if a user is wearing gloves or has dirty hands, they may be unable to use a traditional touch-based authentication method. In these cases, the ability to authenticate using hand gestures in the air could be a useful alternative.

Overall, the proposed method for authenticating human interaction with IoT devices using hand gestures in the air offers a simple, secure, and potentially more accessible option for protecting against unauthorized access. Its combination of computer vision techniques, deep CNN models, and signal processing with a Kalman filter allows it to be more accurate and efficient than existing approaches, and it has been shown to be acceptable to users in terms of usability and satisfaction.

2.1.1 Architecture

The proposed system consists of four parts:

Computer vision: A comproposes technique was used for hand detection and the creation of a virtual pen. Thisbecausel pen is automatically picked up by the hand when the hand is in front of the camera.

First lightweight deep model : A lightweight deep CNN for dynamic hand gesture recognition was used to classify three hand gestures. An open-index finger gesture was used for the drawing. An open-hand gesture was used for the erasing. A closed-hand gesture was used to save.

Kalman Filter: The Kalman filter is a simple and lightweight algorithm.This algorithm was used to smooth hand-drawn symbols on the air.

Second lightweight deep model: The login authentication and verification consist of three stages. First, the authentication key symbols are drawn to the login. Second, keys were extracted. Third, a lightweight deep CNN was used to verify authentication keys. The following subsections provide an in-depth explanation for each part.

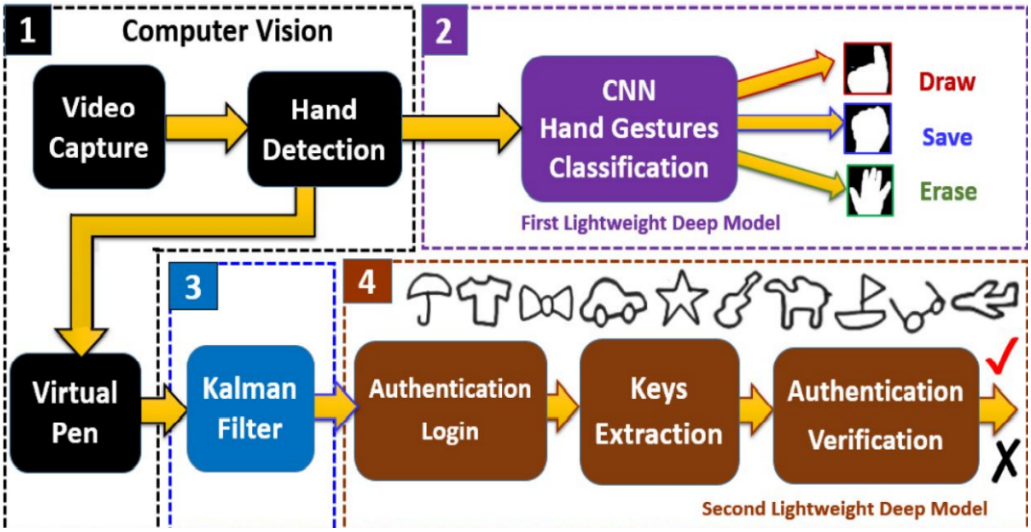


Figure 2.1: Architecture of proposed model

2.1.2 Hand Detection And Catch Virtual Pen

The goal of this stage was to identify three specific features of a hand using a camera: the presence of skin, the center of the hand, and the topmost point representing the fingertip. To achieve this, an image of the hand was captured and processed through a series of filters. The following sections provide more detailed information about each part of this process.

Color Filter: The preliminary processing stage is the most important step in the proposed system because any failure at this stage will affect the subsequent stages. This stage involves detecting the hand in the image, which can be challenging due to variations in skin color among individuals or differences in lighting conditions.

To address this issue, the system performs calibration by having the user place their hand in front of the camera for five seconds before starting the drawing process. This allows the camera to capture the correct skin tone for each user based on the lighting conditions and ensures more reliable hand detection.

Mask Technique (Background subtraction) : The mask technique involves image processing to subtract the background. The proposed system uses a mask to obtain only the hand object and ignores the rest of the background of the image.

Binary Image Converter (Bitwise AND Filter) : A binary image converter or bitwise AND filter is used to convert the image into white and black images. In this filter, a particular pixel is turned off (black pixel) or turned on (white pixel). If the pixel value is zero, it is turned off (black pixels). If the pixel value is greater than zero, it is turned on (white pixels).The proposed system applies this filter to the resulting images from the mask filter. The result of the filter is a binary image with noise.

Opening Filter : An openconsister was used to remove small noise around the hand object. The opening filter was an erosion filter, followed by a dilation filter.

Dilation Filter The proposed system uses another dilation filter to highlight the feature off the hand object because the dilation filter joins the broken parts of the hand object in the image, which increases the area of the hand object.

Hand Contour Detection : Hand contour This stage aimedding the largest contour in the entire image, which is the hand object The proposed system determines the center of the largest contour in the entire image, which is the center of the hand.

2.1.3 Extraction of the image

The process of authenticating a password symbol drawn by the user in real-time involves extracting and processing the authentication symbol image, then inputting it to the model for verification. The symbol extraction stage includes several filters to isolate the symbol from the entire image. The image is first converted to grayscale, then to a binary image using a threshold filter. The dilation filter is used to increase the size of the symbol in the image, and the largest contour in the image is extracted as the symbol. In the preprocessing stage, the size and background color of the symbol image are adjusted to match the size and background color of the trained image dataset. This ensures that the model can accurately compare the extracted symbol to the stored passwords.

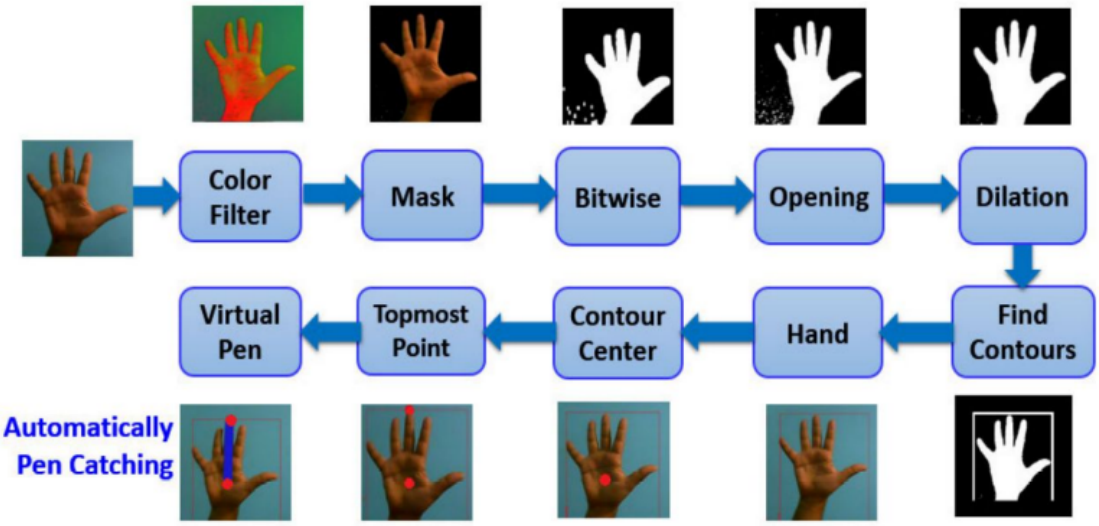


Figure 2.2: Image extraction

2.1.4 Result

Users	1	2	3	3	5	6	7	8	9	10
User_1	44	38	41	48	45	24	42	28	26	40
User_2	48	28	24	43	32	36	55	51	27	38
User_3	35	43	29	41	37	55	56	41	40	28
User_4	25	51	45	54	35	26	36	52	48	45
User_5	28	39	33	41	45	28	33	43	36	53
User_6	48	44	43	44	33	39	60	28	32	38
User_8	51	43	28	36	42	30	37	27	48	37
User_9	44	41	27	51	29	37	50	48	32	52
User_10	42	37	41	36	43	30	48	41	57	41
User_11	49	50	60	41	36	47	44	45	27	46
User_12	32	37	38	38	45	38	35	42	31	26
User_13	35	33	46	25	41	36	44	53	39	35
User_14	37	45	45	37	33	42	31	38	49	40
User_15	26	50	32	46	31	29	35	54	38	50
User_16	32	55	53	38	53	25	47	33	27	45
User_17	42	42	24	42	46	37	31	42	55	37
User_18	58	25	38	43	47	35	48	41	60	37
User_19	36	35	37	48	44	45	37	34	36	34
User_20	31	37	30	44	42	57	42	49	30	34

Figure 2.3: Image extraction

The table includes cells that are red, indicating failed attempts, and cells that are not red, indicating successful attempts. The values in the cells represent the time in seconds it took to complete the task. The paragraph also mentions that some statistical information about the time spent on this task is available.

- Average time for successful attempts (40 s).
- Median time for successful attempts (39 s)
- Mode time for successful attempts (37 s).
- Maximum time for successful attempts (60 s).
- Minimum time for successful attempts (24 s)

2.1.5 Advantages

The convolutional neural network (CNN) proposed in this study has been trained to recognize cars through various methods of drawing their shapes. This allows users to choose a drawing method that is easy and comfortable for them, and input the resulting hand-drawn symbol to the CNN model for identification.

The hand-drawn symbols can be drawn in the air without any specific orientation, size, or grid restrictions, and are easy to remember and associate with specific devices in the Internet of Things (IoT) environment, where there may be many smart devices requiring individual passwords. This makes it easier for users to remember a large number of passwords and link each symbol with a private device.

2.2 Paper2

Video Hand Gestures Recognition Using Depth Camera and Lightweight CNN

This research paper by David González León, Jade Gröl, and Sreenivasa Reddy Yeduri proposes a method for recognizing hand gestures in video using a depth camera and a lightweight convolutional neural network (CNN) model. The recognition of human actions or gestures has been a topic of significant research in recent years, and the use of artificial intelligence and sensor technology has become increasingly popular in order to improve the autonomy of individuals. Hand gestures recognition has a range of potential applications, including sign language detection, smart home technology, autonomous vehicles, healthcare, augmented and virtual reality, driver monitoring in autonomous cars, and automatic surgical tasks..

•The researchers created a dataset of video-based hand gestures using RGB-D camera.

• They proposed lightweight deep CNN model for the hand gesture classification from video sequences.

• When compared to RGB camera-based hand gestures, the proposed depth camera-based hand gestures are more reliable and robust.

• Furthermore, when compared to image-based hand gesture recognition, the proposed video-based hand gestures have a number of practical applications.

• It is also simple to incorporate the additional video gestures without requiring major changes to the proposed model.

• They also included a detailed analysis on the reduced number of frames in a video gesture, which is extremely useful in other domains as video-based

• Through extensive experiments, we evaluate the performance of the proposed method in terms of classification accuracy and inference time.

• Further, they have deployed the model on the edge computing system to show the capability of the model for the real-time applications

2.2.1 System Description

The dataset for this study was created using an Intel RealSense Depth Camera D435, which is a RGB-D camera with a maximum range of 3 meters and two channels: one for the RGB stream and one for the depth stream. The camera has a depth field of view (FOV) of 87○×58○ and an RGB sensor FOV of 69○ × 42○. The setup for extracting both RGB and depth versions of the images is shown in Figure 2.5, which depicts the camera mounted on a tripod stand (as shown in Figure 2.4)

Figure 2.4: The camera module

Figure 2.5: The complete setup for creating the dataset.

All the recorded sequences for this study were taken in the same room with the same framing, which included the area from the waist to the top of the head and had a field of view of 1.5 meters.

2.2.2 Dataset Details

The dataset consists of 762 sequences, each of which includes 40 frames and has both a RGB and depth version generated by the two channels of the camera. The images are resized to 25% of their original size before being saved.

2.2.3 Gesture Details

Scroll Up and Scroll Down,Scrol left and Scroll right,Zoom in and Zoom out

For the scroll up gesture, the movement starts from below the waist and goes

to the top of the head, while for the scroll down gesture, the movement goes in the opposite direction. Figure 2.6 shows the RGB version of the starting and ending positions of the hand in the video for the scroll up gesture, and Figure 2.7 shows the depth version of the same positions for the scroll up gesture.

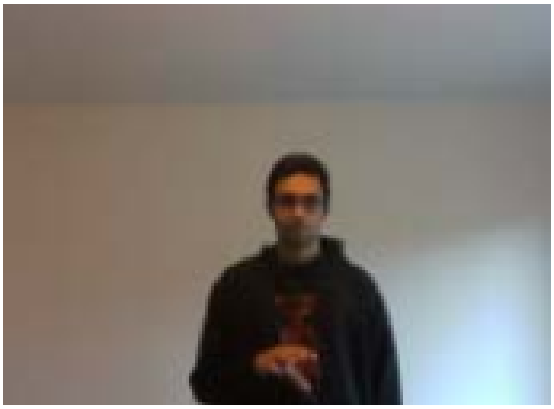
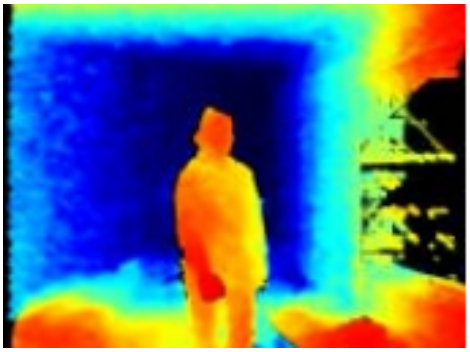


Figure 2.6: A complete set of RGB version of gestures.



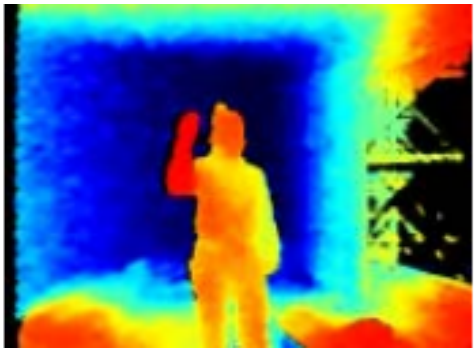


Figure 2.7: A complete set of depth version of gestures.

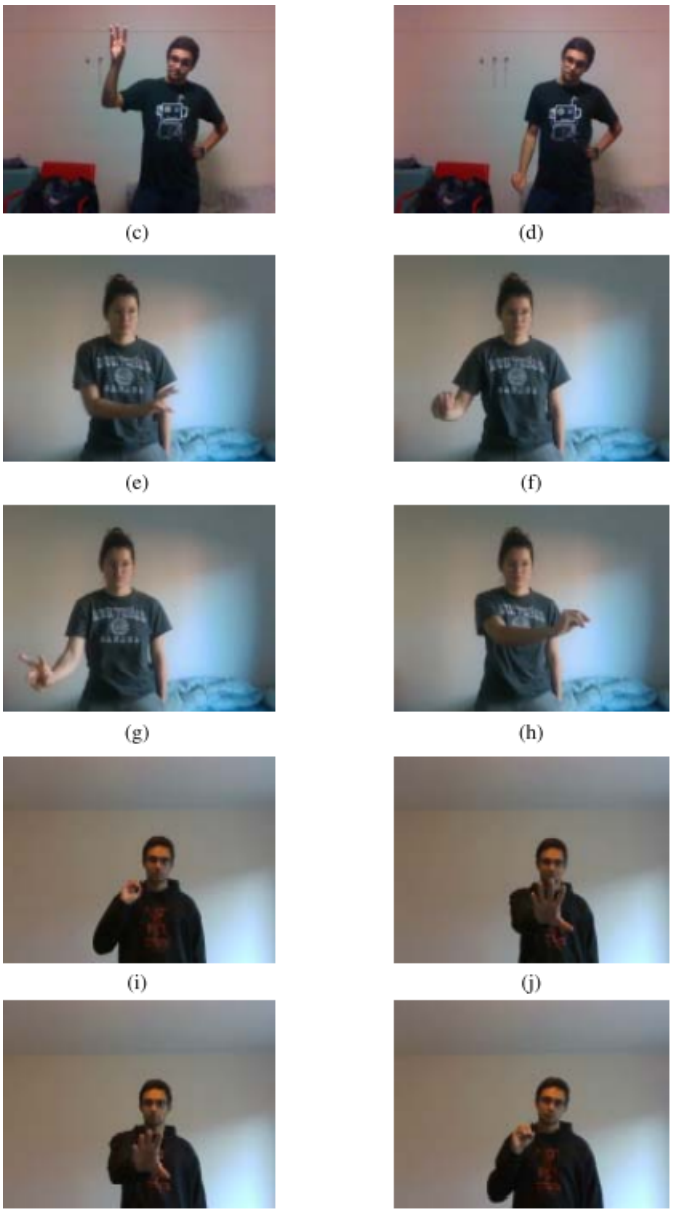


Figure 2.8: A complete set of RGB version of gestures.

Figure 2.8 shows the RGB version of the starting and ending positions of the hand in the video for the scroll down, scroll right, scroll left, zoom in, and zoom out gestures.

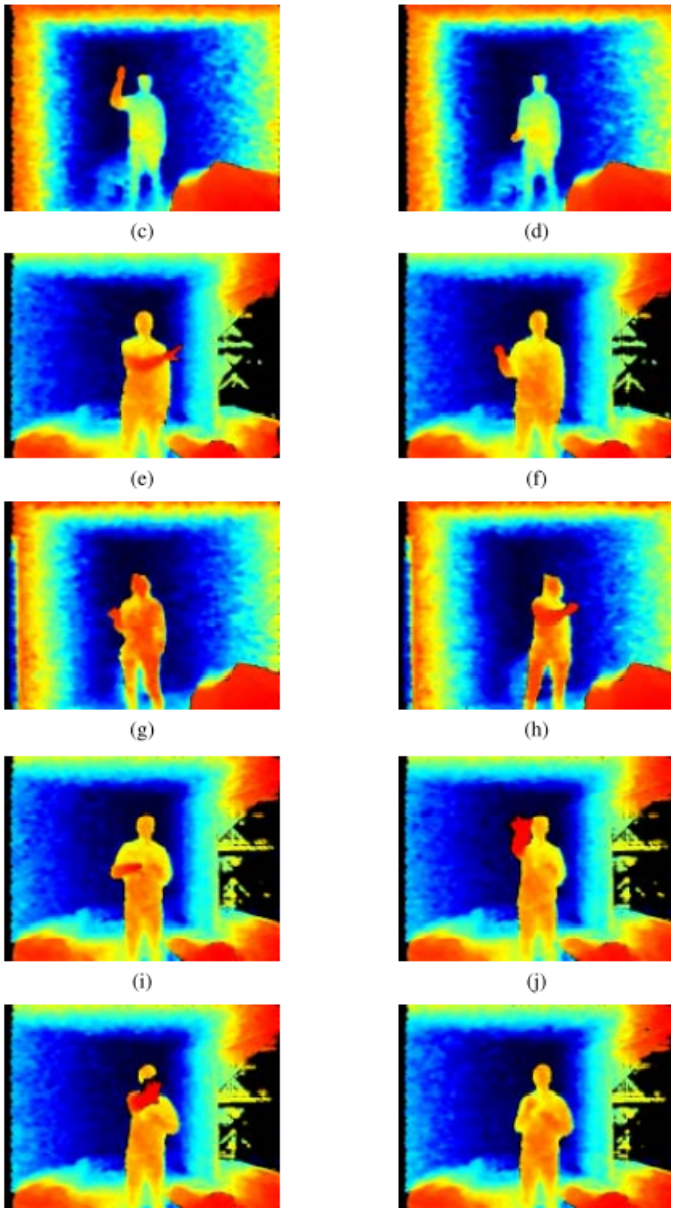


Figure 2.9: A complete set of depth version of gestures.

Figure 2.9 shows the depth version of the starting and ending positions of the hand in the video for the scroll down, scroll left, scroll right, zoom in, and zoom out gestures.

2.2.4 Light Weight CNN For Video Based Hand-Gestures Classification

The convolution 3D layer: applies cuboidal convolution filters to 3D input using a convolution kernel, resulting in a tensor of outputs. It works based on the principle of convolving the layer input with the convolution kernel..

Max Pooling 3D layer: This is used to reduce the dimentionality of the input which then helps the CNN to train faster.

Batch Normalization layer: It allows each network layer to perform learning

independently. It also helps tobnormalize the output layer from the corresponding input layer, which reduces the network initialization sensitivity and speeds up the training.

Flatten layer: Flatten layer takes matrix as input and creates a vector. Then, it will be connected to the fully-connected layer for classification.

Dropout layer: This layer helps in preventing the over-fitting by setting randomly selected input units to 0 with a frequency of r at each step during training time.

Dense layer: Dense layer is basically used for changing the dimensions of the vector. It generates an 'm' dimensional vector with matrix-vector multiplication.

2.2.5 Results

S.No	Model/Dataset size	Test Accuracy rgb	Train Accuracy rgb	Test Accuracy depth	Train Accuracy depth	Overall Accuracy rgb	Overall Accuracy depth
1	Proposed Model/full	97.25% (+-0.55)	99.60% (+-0.23)	98.86% (+-0.73)	99.48% (+-0.38)	99.23%	99.18%
2	Proposed Model/1 of 2	97.50% (+-1.03)	99.51% (+-0.33)	98.53% (+-0.33)	98.94% (+-0.39)	98.20%	99.04%
3	Proposed Model/1 of 4	95.91% (+-2.62)	99.05% (+-0.22)	98.47% (+-0.50)	99.43% (+-0.31)	98.99%	99.16%

Figure 2.10: Performance Comparison Of Proposed Model On Different Sizes Of The Dataset

A dataset containing two video sequences each with 40 frames has been created for both RGB and depth version. The proposed model has been compared to the state-of-theart models in terms of classification accuracy. The proposed model has achieved an accuracy of 99.23percent and 99.18percent on RGB and depth version of test datasets.

2.2.6 Advantages

Provides greater flexibility than the existing system.

Less prone to physical damage due to absence of a fixed physical device.

2.3 Paper3

Real-Time Gesture Detection Based on Machine Learning Classification of Continuous Wave Radar Signals

This paper aims to explore various methods for efficiently classifying simple human gestures recorded with a low-cost radar system, and compares classical Machine learning (CML) methodologies with novel Machine learning (nML) approaches such as neural networks (NNs). It investigates the mutual performance of different solutions such as threshold detection (THD) algorithms, and aims to determine the additional functions required by each approach (THD, cML, nML) to work effectively, as well as the accuracy that can be expected and achieved with each approach. The study also examines the feasibility of running the designed networks on low-cost standalone computers and Microcontroller Units (MCUs).

2.3.1 Architecture

Gesture recognition allows for communication between humans and other beings or machines. This can include the detection of motion-based gestures, such as a hand wave, through various sensors and methods such as cameras, lasers, and radar. Hybrid systems that combine these approaches may also be used. This technology allows for the detection of gestures through optical means and through the use of radar echoes.

This study looks at basic human gestures such as hand swipes or foot kicks, and evaluates the performance of different algorithms based on their true positive rate, false-positive rate, real-time capability, computational power, and implementability on low-cost hardware. The goal is to compare the effectiveness of these algorithms in detecting and interpreting these gestures.

A gesture recognition door sensor only activates the door if it detects a specific, valid gesture. Invalid signals, such as those made by a passerby, are ignored. In order for the system to be user-friendly and interact effectively with the low-cost radar system, the valid gestures must be of a certain size and cannot be small micro gestures like "touching fingertips." In this case, factors like computational power, real-time capability, and low cost are not as important. Real-time capability refers to the ability of the algorithm to classify the signals and provide a result within 250 milliseconds, so that the user does not have to wait for the door to open. The algorithm begins to classify the signals as soon as the user starts to execute the gesture, and is able to present classification probabilities while the gesture is still being performed. This allows for seamless and convenient use of the system

In the context of radar hardware, it is more efficient to use continuous-wave signals rather than frequency-modulated continuous wave signals. Continuous-wave signals can only provide information about the Doppler frequency shift and intensity of the radar signal, while frequency-modulated continuous wave signals can also provide information about velocity and distance. To avoid an overly complex and expensive solution, it is important to define the gestures that need to be detected beforehand. In this study, the focus is on the detection of natural gestures made

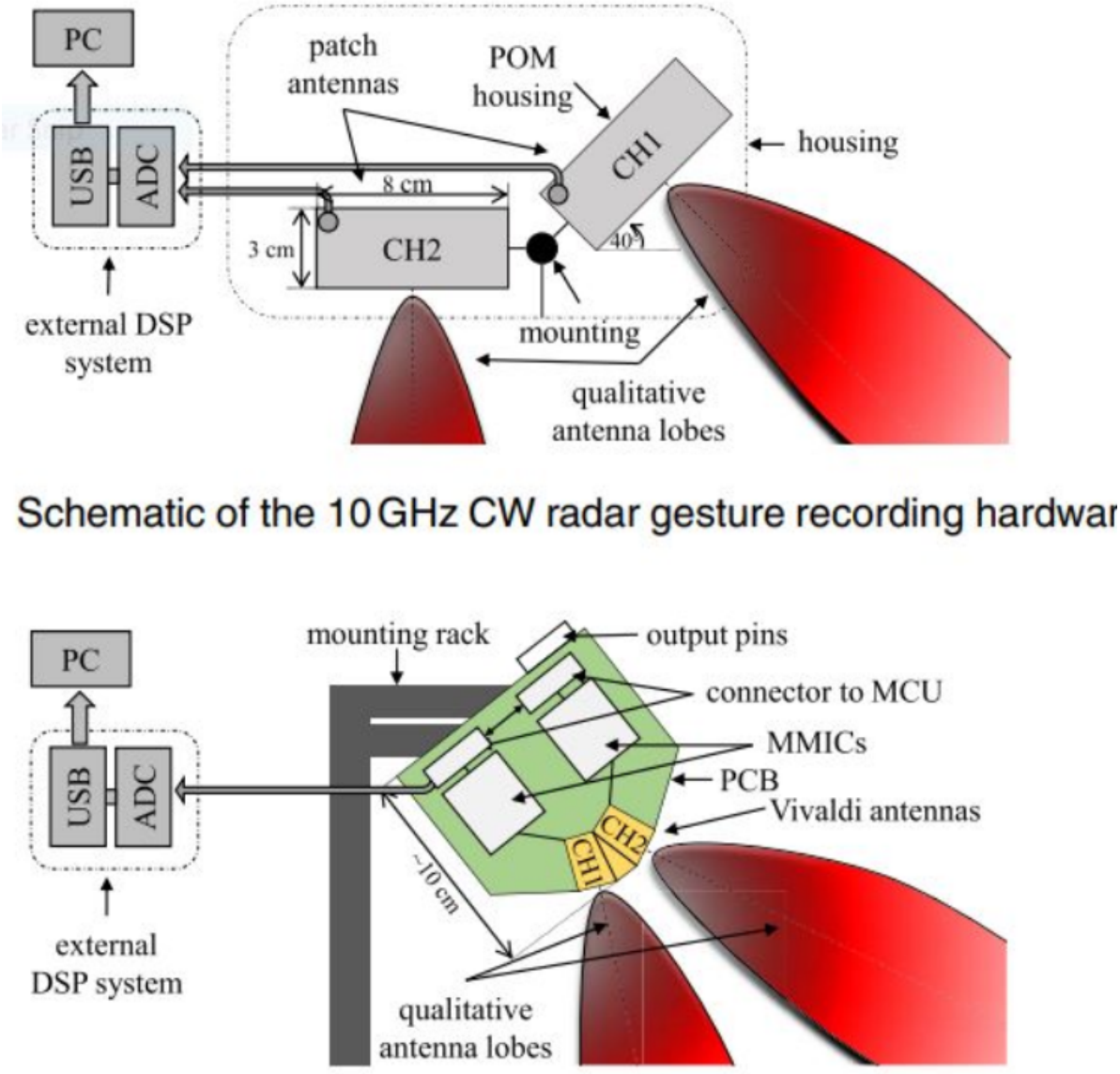


Figure 2.11: Architecture of the proposed model

with human limbs that can be used to interact with everyday applications, such as a hand wave motion to open a door. This system only recognizes valid gestures and ignores invalid signals, such as those made by a passerby. The valid gestures must be of a certain size and cannot be small micro gestures like "touching fingertips" in order to be detectable by the low-cost radar system.

This study uses two different radar systems for its investigations: one operating at 10 GHz in the X-band, and one operating at 24 GHz in the K-band. The X-band system is built with discrete components and was designed as a low-cost feasibility study, while the K-band system is based on a more industrializable monolithic microwave integrated circuit.

10 GHz System: The first employed radar system operates in the X-band at 10 GHz, . The system works with CW and a single mixer, hence, only the in-phase part of the signal is available The effective coverage range of the 10 GHz radar system is approximately one meter. The Doppler signal in baseband is sent to an external analogue-to-digital converter (ADC), the ADC sends the data to an MCU or personal computer (PC) either via universal serial bus (USB), or via I2C. The HW for this radar system costs approximately 10 EUR without manufacturing costs.

24 GHz System: The second radar system operates in the industrial, scientific, and medical band at 24 GHz and consists of a single printed circuit board with all the necessary electronic components, including power supply and RF monolithic microwave integrated circuits from Infineon's BGT24 series. These circuits are set to continuous-wave mode with two different frequencies to prevent interference, and have a stable output frequency thanks to a phase locked loop. While the circuits could also be set to frequency-modulated continuous-wave mode to obtain distance information, this is not necessary for effective gesture recognition. Similar to the 10 GHz system, the 24 GHz system has two monostatic channels with 40 degree angular separation and uses Vivaldi antennas.

2.3.2 Gesture Recognition Methodoligies:

1.Naïve Gesture Recognition Approach: The first focus for naive gesture recognition is on time-harmonic decomposition (THD) algorithms. These algorithms use fixed, equally distributed threshold levels for all input data rather than adaptive ones. After a scattered signal is received by the radar system from a moving object, person, or animal, it is mixed down, filtered, and downsampled. The second focus is on Classical Machine learning (cML) methodologies, which offer a range of high-performance algorithms that can be trained with a dataset. To compare the effectiveness of these approaches, various CML methods will be trained and evaluated.

- Naïve Bayes (NB)
- Support vector machine (SVM)
- Stochastic gradient descent (SGD)
- Decision tree (DT) .

Datasets Two different datasets are used to train the machine learning algorithms, both of which were recorded using low-cost radar hardware. The focus of this study is on designing scalable neural networks (NNs) that can be run on low-cost Microcontroller units (MCUs) for gesture recognition. To investigate classical machine learning methods, the 10 GHz continuous-wave radar system is used. Nine different gestures were defined, four of which are labeled as valid and five as invalid. The four valid gestures are different hand movements: straight kick, sideways kick, kick from left to right and vice versa, and a swiping movement. The five invalid gestures are a passerby, a person walking towards the sensor, a person lingering in front of the sensor, a person bending down in front of the sensor, and a scenario where the sensor is washed by a hand (this is the most challenging invalid event). The dataset consists of 3600 gestures recorded individually, each 5 seconds long and containing one valid or invalid gesture.

2.3.3 Result

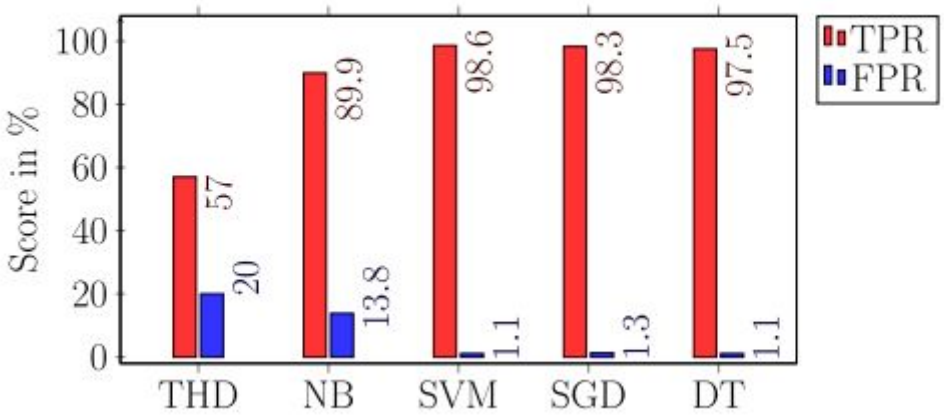


Figure 2.12: Architecture of the proposed model

The time-harmonic decomposition (THD) classifier was unable to effectively classify the gestures, so the focus shifted to classical machine learning (CML) methodologies. These methodologies offer a range of high-performance algorithms that can be trained with a dataset, and the results of the training processes are shown here. All classifiers were optimized to improve classification performance, while the THD algorithm remained in its unoptimized and unmodified state. The blue bars represent the true positive rate (TPR, the higher the better) and the red bars represent the false positive rate (FPR, the lower the better). The results show that the THD approach is not effective for classifying the gestures, and the naive Bayes (NB) classifier performs better but is still unsatisfactory. However, the more complex cML approaches, with their extensive optimization potential and tunable parameters, are able to achieve excellent classification results. Among these, the support vector machine (SVM) has the best TPR and generalizes well to new data.

2.3.4 Advantage

Radar systems have an advantage over optical systems because they are less computationally expensive and less affected by weather conditions. This is because camera-based gesture recognition requires more computational power, and cameras can be disrupted by poor weather conditions such as darkness or fog.

2.4 Paper4

Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation

The aim of sign language recognition is to create systems that can recognize and interpret the hand gestures and movements used in sign languages for deaf communities. These systems can be used to facilitate communication between deaf and hearing individuals, or to translate sign languages into spoken or written languages. However, sign language recognition is a difficult task due to the complexity and variety of sign languages, as well as the variability in the way that different users produce signs.

This paper presents a method for sign language recognition that uses deep learning, a type of machine learning that trains artificial neural networks on large datasets. Deep learning models can learn intricate patterns and relationships in the data and are effective for various tasks such as image and video recognition, natural language processing, and speech recognition.

In the context of sign language recognition, deep learning models can be trained to identify and classify different hand gestures using various hand gesture representation techniques. These techniques may include depth maps, which are images that show the distance of objects from the camera; skeletons, which are simplified versions of hand and finger bones; or 2D images of the hand. The type of representation method used can impact the model's performance, as well as its complexity and the amount of data needed for training.

2.4.1 Architecture

The proposed system consists of 3 phases:

Input Preprocessing: The input videos are converted into sequences of RGB frames of varying lengths, and linear sampling is used to normalize the temporal dimension by selecting only 16 frames from each video sequence.

Feature Learning: Feature learning is the process of automatically extracting useful and relevant features from raw input data to perform a specific task. In the context of sign language gesture recognition, feature learning involves finding patterns and characteristics in hand gestures that indicate the specific sign being made. These features may include the hand's shape and movement, the position and orientation of the fingers, and the relative distances between different parts of the hand.

Feature Fusion and Classification Feature fusion is the process of combining multiple features or representations of the input data to improve the performance of a machine learning model. In the context of sign language gesture recognition, feature fusion may involve combining different hand gesture representations, such as depth maps and skeletons, or combining features extracted from multiple types of data, such as images and videos.

In cases where a single representation or feature set may not contain enough information to accurately classify hand gestures, feature fusion can be helpful. By

combining multiple sources of information, the model may be able to make more precise predictions. However, feature fusion can also make the model more complex and may require more data for training.
Classification is the process of assigning a label or class to an input based on its characteristics. In the context of sign language gesture recognition, classification involves identifying a specific sign or word for a given hand gesture.
Possible classification methods for sign language gesture recognition include decision trees, k-nearest neighbors, and support vector machines. The choice of classification method may depend on the characteristics of the data and the needs of the task.

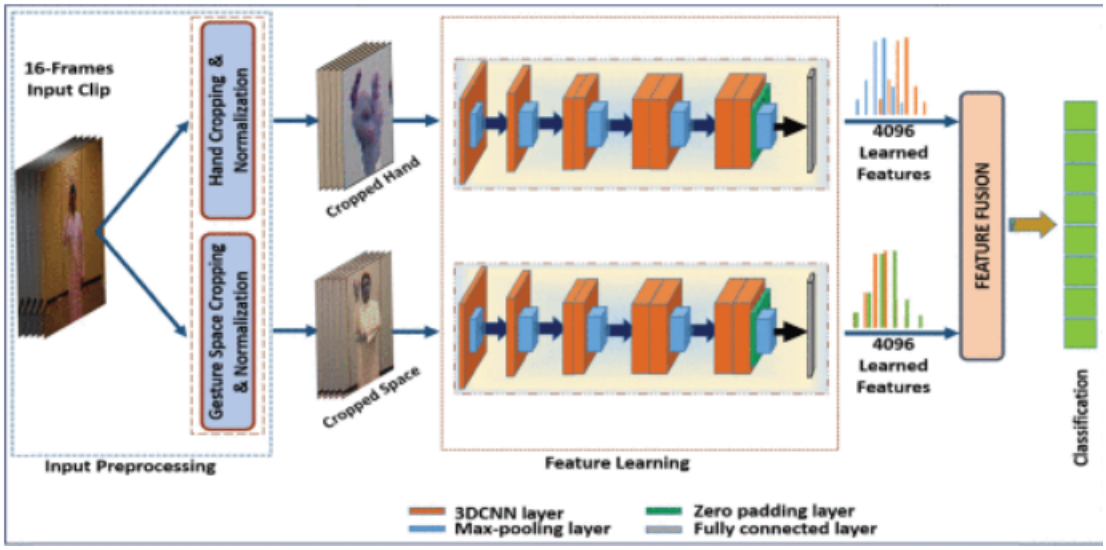


Figure 2.13: Architure of proposed model

2.4.2 Experimental Results

To evaluate the proposed system, the experiment conducted in 2 scenarios:

Signer-dependent mode: In this situation, the samples were randomly shuffled and divided into two subsets for training and evaluation. This means that the samples for each signer were randomly split into a training and evaluation set.
Signer-independent mode: In this case, the signers were split into two sets. The samples from the first set of signers were used for training, and the samples from the second set were used for testing.

The results of the experiment in terms of evaluation loss and recognition accuracy are shown below.

The highest performance, 80.94%, was obtained by fine-tuning all layers except the first one.

2.4.3 Advantages

There are several advantages to using a deep learning-based approach for sign language gesture recognition:

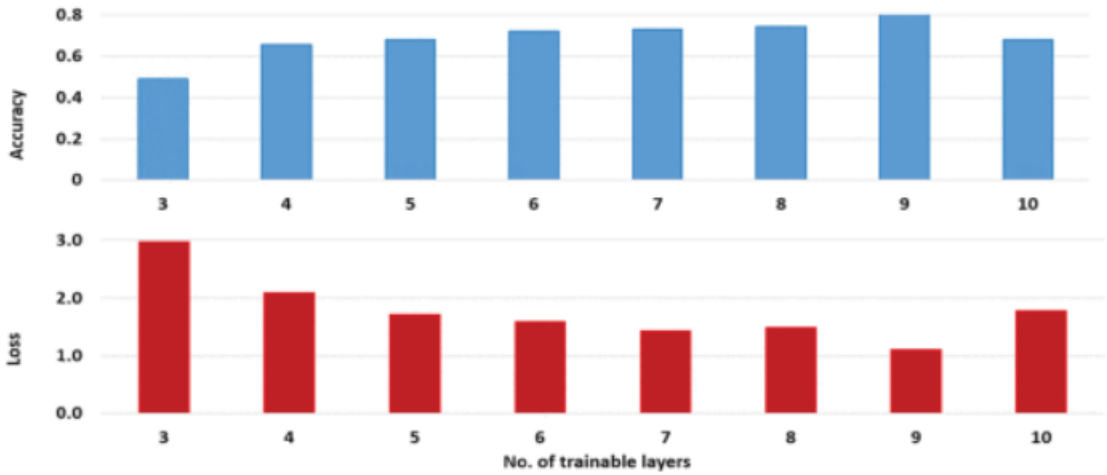


Figure 2.14: Architure of proposed model

High accuracy:: Deep learning models are able to learn intricate patterns and relationships in the data and can achieve a high level of accuracy for various tasks, including sign language gesture recognition.

Generalization: Deep learning models are able to generalize well to new, unseen data, making them well-suited for tasks where the data may be highly variable or diverse.

Efficiency: Deep learning models can be designed to be computationally efficient, enabling them to run in real-time on various devices such as smartphones and smartwatches.

Scalability: Deep learning models can be trained on large datasets, which enables them to handle a wide range of classes and variations in the data. Additionally, these models can learn automatically from the data, without the need for manual feature engineering, making them a useful tool for tasks where the underlying features are not well understood.

Overall, deep learning techniques can be used to develop highly accurate and efficient systems for sign language gesture recognition, which can enable the translation of sign languages into spoken or written languages and facilitate communication between deaf and hearing individuals.

2.5 Paper5

A Hand Gesture Based Interactive Presentation System Utilizing Heterogeneous Cameras

A hand gesture based interactive presentation system that uses a network of cameras with different capabilities or types to detect hand gestures for controlling presentations can be utilized in various settings, such as conference rooms, classrooms, or other places where presentations are given.
This hand gesture based interactive presentation system can be useful for presenters who need to be mobile because it allows them to control the presentation without using a physical device, such as a remote control. This can enable the presenter to move around the room while giving the presentation.
The use of multiple cameras with different capabilities in this hand gesture based interactive presentation system can help improve the accuracy and reliability of detecting hand gestures. This is because the different cameras may be able to capture hand gestures in a variety of situations, such as in low light conditions or with high resolution. By using a network of cameras with these different capabilities, the system may be better able to detect hand gestures in a variety of conditions.
A system that allows users to .control a presentation using hand gestures, detected by a network of cameras with varying capabilities, could improve the accuracy and reliability of detecting the gestures. This type of system could be useful for presenters who need to be mobile and free from physical devices, and could potentially be used in various settings such as conference rooms or classrooms.

2.5.1 Architecture

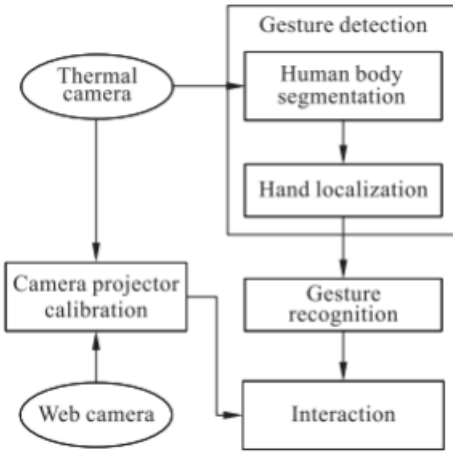


Figure 2.15: Architure of proposed model

The architecture of a hand gesture based interactive presentation system utilizing heterogeneous cameras could vary depending on the specific design and requirements of the system. However, some common elements that might be included in the architecture are:

Cameras: The system would also need a method for processing the images captured by the cameras in order to detect and recognize hand gestures. This may involve using machine learning algorithms or other techniques to analyze the images and determine the gestures being made. The system would also need a way to interpret the detected gestures and map them to specific actions or commands, such as advancing to the next slide in a presentation. Finally, the system would need a way to communicate the commands to the presentation software or other system being controlled.

Image processing and gesture recognition: The system would require some type of image processing and gesture recognition component to analyze the images captured by the cameras and detect the hand gestures. This could be done using machine learning algorithms or other techniques. The system would also need a way to translate the detected hand gestures into actions, such as advancing to the next slide in a presentation. This could be done through the use of a control interface or some other type of mechanism.

Communication and control:The system would require a mechanism for communicating the detected hand gestures to the presentation software and using them to control the presentation. This could be done through a wired or wireless connection.

Presentation software:The system would need to include a network of cameras with different capabilities for capturing images of hand gestures, an image processing and gesture recognition component for analyzing the images and detecting the hand gestures, and a means of communication with the presentation software to control it based on the detected hand gestures. The system should also be compatible with a variety of presentation software and have the ability to connect wirelessly or through a wired connection.

User interface: The system would need to be able to detect and interpret hand gestures made by the presenter, communicate with the presentation software, and be compatible with a variety of presentation software and potentially include a user interface for configuration.

To summarize, a hand gesture based interactive presentation system utilizing heterogeneous cameras is a system that allows users to control a presentation using hand gestures, which are detected by a network of cameras with different types or capabilities. The system would include components for capturing images, processing and recognizing hand gestures, communicating with the presentation software, and providing a user interface for the presenter. This system could be useful for individuals giving presentations, as it would allow them to control the presentation in a more intuitive and natural way, without the need for a physical device. The use of heterogeneous cameras in the system could improve the accuracy and reliability of detecting hand gestures in a variety of situations.

2.5.2 Results

.

Table 2 Gesture recognition results

Gesture	Number of gestures	Number recognized	Recognition rate (%)	Overall rate (%)
Lining	84	81	96.4	
Circling	52	50	96.2	96.7
Pointing	43	42	97.7	

Figure 2.16: Results

2.5.3 Advantages

- Increased flexibility and mobility for the presenter: By allowing the presenter to control the presentation using hand gestures, the system could allow them to move freely around the room without the need for a physical control device.
- Enhanced audience engagement: The use of hand gestures as a means of controlling the presentation could make the experience more interactive and engaging for the audience.
- Improved reliability and accuracy: The use of a network of cameras with different capabilities could allow for better detection of hand gestures in a variety of situations, resulting in a more reliable and accurate system.

2.6 Paper 6

Robust Hand Gesture Recognition Based on RGB-D Data for Natural Human–Computer Interaction.
paper presents a method for recognizing static and dynamic hand gestures using RGB-D data. The method involves extracting the hand gesture contour and identifying the palm center using the Distance Transform algorithm. The fingertips are located using the K-Curvature-Convex Defects Detection algorithm, and the distances between the pixels on the hand gesture contour and the palm center and the angles between the fingertips are used as auxiliary features to create a multimodal feature vector. A recognition algorithm is then applied to robustly identify the static hand gestures.

This paper presents a method for recognizing both static and dynamic hand gestures using RGB-D data. The static hand gesture recognition process involves extracting the hand gesture contour and identifying the palm center and fingertips. A multimodal feature vector is then constructed using auxiliary features such as the distances of pixels on the contour to the palm center and the angles between the fingertips. A recognition algorithm is then used to robustly recognize the static hand gestures. For dynamic hand gesture recognition, a unifying feature descriptor is generated by combining the Euclidean distance between hand joints and the shoulder center joint with the modulus ratios of skeleton features. An improved dynamic time warping algorithm is then used to recognize the dynamic hand gestures. Finally, the static and dynamic hand gesture recognition algorithm is tested and verified through extensive experiments and used to create a low-cost, real-time application for natural interaction with a virtual environment through hand gestures.

2.6.1 Architecture

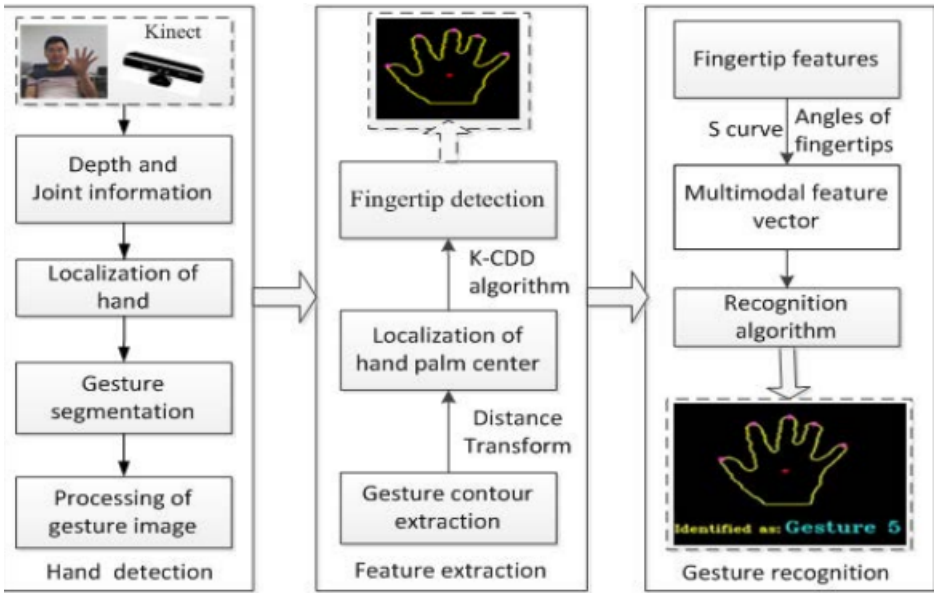


Figure 2.17: Architecture of proposed model

- Data collection: The recognition of static hand gestures involves identifying the contour of the hand, locating the palm center and fingertips, and constructing a multimodal feature vector based on the distances and angles between these points. An algorithm was then proposed to recognize the static hand gestures using this feature vector. For dynamic hand gestures, a unifying feature descriptor was generated based on the distance between hand joints and the ratios of skeleton features. An improved dynamic time warping algorithm was then used to recognize the dynamic hand gestures. The performance of both the static and dynamic hand gesture recognition algorithms was tested and verified through extensive experiments, and the results were used to create a low-cost, real-time system for natural interaction with a virtual environment using hand gestures.

- Preprocessing: The raw RGB-D data is also used to extract hand gesture contours, identify the palm center, and locate the fingertips. These features are then used to construct a multimodal feature vector for static hand gesture recognition.

- Hand detection: Once the hand is detected, the system extracts the contour of the hand from the RGB-D data. The palm center is identified using a distance transform algorithm, and the fingertips are localized using a curvature and convex defects detection algorithm. These features are used to construct a multimodal feature vector that represents the static hand gesture. A recognition algorithm is then applied to this feature vector to classify the static hand gesture.

- Hand tracking: The movement of the hand is analyzed to recognize static or dynamic hand gestures. This may involve extracting features such as the distances between hand joints and the modulus ratios of skeleton features. These features are used to create a feature vector, which is then compared to a set of known gestures to identify a match. If a match is found, the gesture is recognized and the appropriate action is taken, such as advancing to the next slide in a presentation. The improved dynamic time warping (IDTW) algorithm is used to recognize dynamic hand gestures. This involves comparing the feature vector for the current gesture to a set of known gestures using a dynamic time warping method to determine the best match. Finally, the recognition results are used to control a virtual environment or perform other tasks.

- Gesture recognition: The recognition of static hand gestures involves extracting the contour of the hand and identifying the palm center and fingertips using the DT and K-CCD algorithms, respectively. The distances between the pixels on the hand contour and the palm center, as well as the angles between the fingertips, are used as auxiliary features to construct a multimodal feature vector. A recognition algorithm is then applied to this feature vector to robustly recognize the static hand gestures.

- Output: The hand gesture recognition system described in this paper involves collecting and preprocessing data from a depth camera to detect and track the movement of a hand. Machine learning algorithms are then used to recognize the gestures being made by the hand and interpret them to control a device or machine. This allows for natural interaction with virtual environments or other devices using hand gestures.

2.6.2 Results

No.	Descriptions	Functions	No.	Descriptions	Function
01	Both hands keep down	Initialization	09	Right hand swipes left and left hand keeps straight forward	Visual angle turns left
02	Left hand swipes up and right hand down	Virtual miner walks	10	Right hand swipes right and left hand keeps straight forward	Visual angle turns right
03	Right hand swipes left and left arm keeps down	Virtual miner turns left	11	Left hand swipes up and right hand keeps straight forward	Camera and visual angle move up
04	Right hand swipes right and left arm keeps down	Virtual miner turns right	12	Left hand swipes down and right hand keeps straight forward	Camera and visual angle move down
05	Zero	Push virtual mine car	13	Left hand swipes left and right hand keeps straight forward	Camera and visual angle move left
06	One	Press virtual devices bottom	14	Left hand swipes right and right hand keeps straight forward	Camera and visual angle move right
07	Right hand swipes up and left arm keeps straight forward	Visual angle turns up	15	Both hands straight forward, then swipe to left and right	Camera moves forward and view zooms in
08	Right hand swipes down and left hand keeps straight forward	Visual angle turns down	16	Both hands straight to left and right, then swipe forward	Camera moves backward and view zooms out

Figure 2.18: Results

2.6.3 Advantages

Improved accuracy: The use of RGB-D data and advanced machine learning techniques enables the system to accurately and efficiently recognize static and dynamic hand gestures, enabling natural interaction with virtual environments through hand gestures.

Robustness to lighting changes: By using both color and depth information, the system can more accurately detect and recognize hand gestures, even in challenging conditions such as low light or when the hand is partially occluded. This can improve the reliability and usability of the system.

Enhanced user experience:This research presents a hand gesture recognition system that uses depth camera data to accurately detect and recognize static and dynamic hand gestures. The system is able to detect and track the hand in the scene, extract features from the RGB-D data, and use machine learning algorithms to recognize the gestures. The system is designed to be robust to variations in lighting and can be used to control a device or machine through a user interface. The use of depth information in addition to color information can improve the accuracy and robustness of the gesture recognition. Overall, this system enables a more natural and intuitive human-computer interaction experience.

Improved flexibility: Using RGB-D data to track hand gestures allows for more accurate and reliable recognition of hand movements, even in challenging lighting conditions or when the hand is partially occluded. This makes it possible to use the system in a variety of situations and settings. The system can be configured to recognize a wide range of static and dynamic gestures, allowing for a high degree of flexibility in how it is used.

2.7 Paper 7

Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks,2021
The paper introduces a deep learning model for recognizing hand gestures in videos. The model combines multiple techniques to identify both short and long-term features in the video input while minimizing computational requirements.

The model first divides the video into groups of frames, then selects one frame from each group to create an RGB image and an optical flow snapshot. These images are combined and passed through a Convolutional Neural Network (ConvNet) to extract features.

The output from all of the ConvNets is then fed into a Long Short-Term Memory (LSTM) network, which produces the final classification prediction for the hand gesture

2.7.1 Architecture

Convolutional neural networks (CNNs):Convolutional Neural Networks (CNNs) are a type of neural network that excel at tasks involving image classification and processing. They are effective at analyzing sequential data like video frames, and are able to identify features such as edges, corners, and textures within the input data.

Recurrent neural networks (RNNs):Recurrent Neural Networks (RNNs) are a type of neural network that are designed to handle sequential data. They have the ability to retain information from previous time steps through feedback connections, which enables them to use past context to inform their current predictions. This makes RNNs effective for tasks that involve processing sequential data, such as language translation and speech recognition, where the meaning of the input is dependent on the context provided by previous inputs.

Combination of CNNs and RNNs:It is possible to use both Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to create a dynamic hand gesture recognition system. In this type of system, a CNN could be used to extract features from the input data, while an RNN could be used to analyze the feature maps and make predictions based on the most recent data points.

2.7.2 Results

Change of Loss During training, the network parameters are adjusted using the gradient descent algorithm to minimize cross-entropy loss. The loss of the model decreases over time, indicating that the training process is effective.

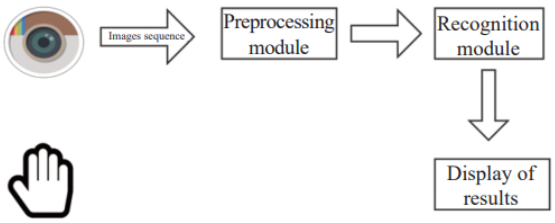


Figure 2.19: Architecture of proposed model

The model achieves the best performance with an accuracy of 95.73% when every input video is divided into 9 groups. The best result on the "zoomedout" Jester dataset is 95.39%.

2.7.3 Advantages

Real-time processing:Short-term sampling neural networks are able to process data in real-time, making them effective for tasks like hand gesture recognition where the input data is constantly changing.

Continuous updating: Short-term sampling neural networks continuously process a stream of data, allowing them to continuously update their predictions as new data becomes available. This is especially useful for tasks like hand gesture recognition, where the user's gestures may change rapidly over time.

2.8 Paper 8

Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frames,2022

In this paper, a method is presented for continuously identifying and recognizing finger gestures based on the similarities between the start and end frames of the gesture. A convolutional neural network (CNN) trained on a large dataset of finger gestures is used to detect and classify gestures in real-time. The experiments show that the proposed approach performs well and achieves high accuracy in various lighting and background conditions. The proposed method has the potential to be used in a wide range of areas, including human-computer interaction, sign language recognition, and virtual reality. Finger gestures are a natural and intuitive way for humans to communicate and interact with each other and with computer systems

2.8.1 Architecture

Input: The system may use a camera or other type of sensor to capture images or video of the fingers and hand performing the gesture.

Preprocessing: The raw input data may undergo processing to remove noise, improve contrast, and extract features that are relevant to the gesture. This may involve techniques such as filtering, edge detection, and image segmentation.

Feature extraction: The processed data is then analyzed to extract features that describe the start and end frames of the gesture. These features may include characteristics such as the shape, size, and orientation of the hand and fingers, as well as the motion and trajectory of the gesture.

Classification: The extracted features are then fed into a classifier, such as a convolutional neural network (CNN), that has been trained to recognize the gestures. The classifier outputs a label indicating the type of gesture that has been performed.

Output: The system may output the recognized gesture in various formats, such as displaying it on a screen, providing visual or auditory feedback, or triggering some other action.

2.8.2 Results

The system should be able to accurately identify and recognize a diverse range of finger gestures with a high level of reliability. The system's performance may be evaluated using various metrics, including accuracy, precision, recall, and F1 score.

The real-world performance of the system should also be taken into consideration, including factors such as speed, robustness to noise and variations in the input data, and the ability to handle a large number of gestures.

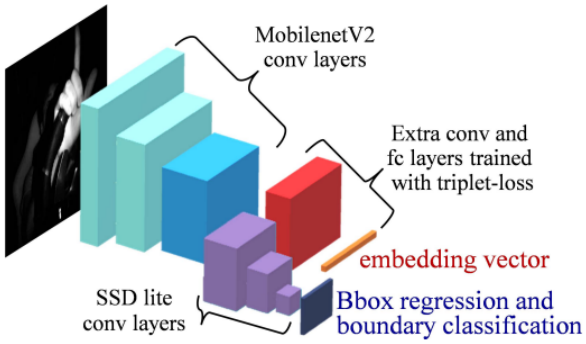


Figure 2.20: Architecture of proposed model

2.8.3 Advantages

Intuitive: Finger gestures are a natural and intuitive way for humans to communicate and interact with each other and with computer systems. This makes them easy for people to learn and use.

Real-time: The system is able to recognize gestures in real-time, allowing for continuous and seamless interaction.

Robust and Versatile

2.9 Paper 9

Gesture-Based Human Machine Interaction Using RCNNs in Limited Computation Power Devices ,2021
Manual hand gestures can be used in real-time to allow humans to communicate with and control computers or other electronic devices using hand movements. This is made possible through the use of motion tracking, computer vision, and machine learning technologies. Hand gestures can be used to carry out a range of tasks, such as navigating menus, controlling media playback, and interacting with virtual or augmented reality environments. While hand gesture-based interaction has the potential to enhance the usability and accessibility of computing devices, it also presents challenges in terms of gesture design and technology accuracy and reliability.

2.9.1 Architecture

Input devices: These are used to capture the hand gestures made by the user. Examples include depth cameras, infrared sensors, and motion capture systems.

Gesture recognition software: This refers to the software responsible for interpreting the hand gestures captured by the input devices and translating them into commands or actions. This software may use machine learning algorithms to identify patterns in the hand gesture data and map them to specific actions.

Output devices: These are used to display the results of the hand gestures to the user. Examples include displays, speakers, and haptic feedback devices.

User interface: This is the interface through which the user interacts with the system. It may include menus, buttons, and other elements that allow the user to perform tasks or access information.

Networking and communication: If the system is connected to other devices or the internet, it may include networking and communication hardware and software to enable communication and data transfer.

Hardware and software platform: The hardware and software platform of the system refers to the physical hardware and operating system that the system runs on. This can include a computer, mobile device, or specialized hardware designed for gesture recognition.

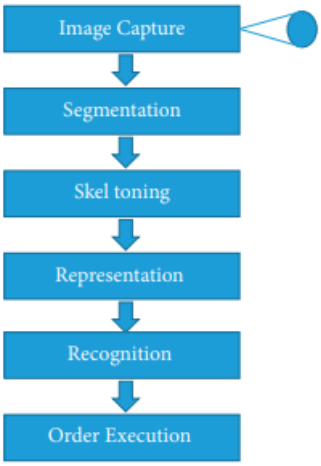


Figure 2.21: Block diagram of the system

2.9.2 Result

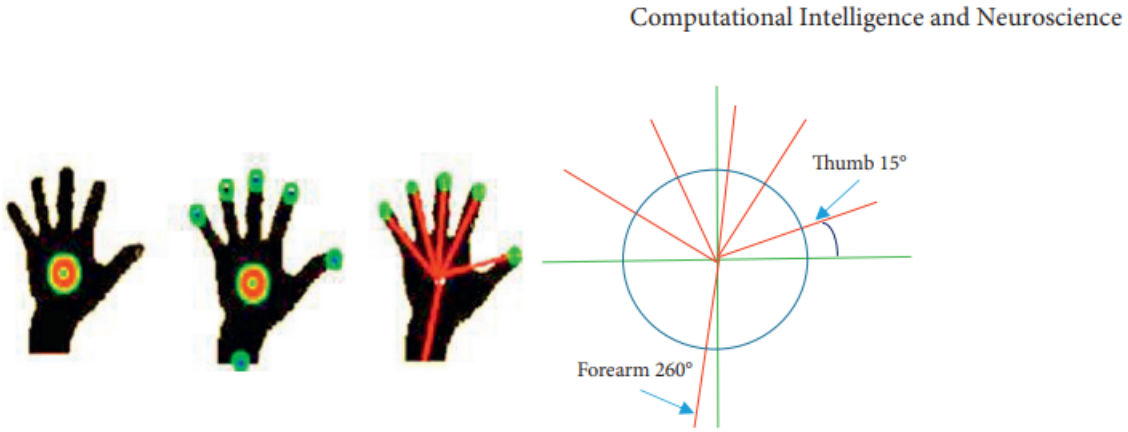


Figure 2.22: Angles relative to the forearm

- True hits 79.27%
- True rejections 99.50%
- False hits 0.27%
- False rejections 38.39%

2.9.3 Advantages

Increased usability:Hand gestures can provide a more natural and intuitive way to interact with computers and other digital devices, making them easier and more enjoyable to use.

Enhanced user experience: Manual hand gestures can enhance the overall user experience by allowing users to interact with devices in a more immersive and interactive way.

Increased efficiency: Hand gestures can often be quicker and more efficient than other forms of input, like typing on a keyboard or using a mouse, allowing users to complete tasks faster.

Greater accessibility:Hand gestures can provide a more accessible means of interacting with computers and other devices for users with disabilities or mobility issues.

2.10 Paper 10

A hand Gesture Based Interactive Presentation System Utilizing Heterogeneous cameras.

Hand gestures are detected by a network of cameras with different types or capabilities. This system could be used in various settings, such as conference rooms, classrooms, or other locations where presentations are given.

One potential use of this type of system is to enable the presenter to control the presentation without the need for a physical remote control or other device. This could be especially beneficial for individuals who are giving a presentation and need to be mobile, as it would allow them to move around the room without being tied to a device.

The use of a network of cameras with different capabilities, or heterogeneous cameras, in the system could lead to improved accuracy and reliability in detecting hand gestures. Different cameras may have unique features, like higher resolution or the ability to capture hand gestures in low light conditions. By using a variety of cameras, the system may be able to more effectively detect hand gestures in various situations.

In summary, a hand gesture-based interactive presentation system that utilizes heterogeneous cameras could be a useful tool for individuals giving presentations, allowing them to control their presentations in a more intuitive and natural way.

2.10.1 Architecture

The architecture of a hand gesture-based interactive presentation system utilizing heterogeneous cameras may vary depending on the specific design and requirements of the system. However, some common elements that might be included in the architecture are:

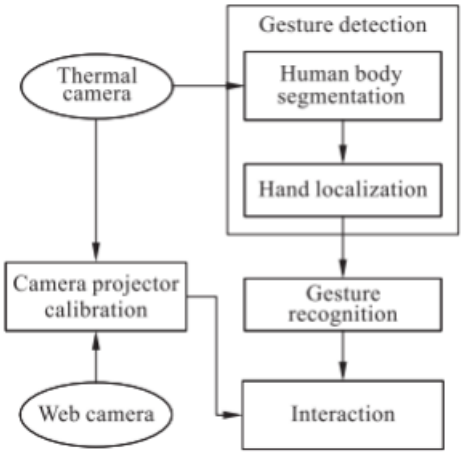


Figure 2.23: Architecture of proposed model

Cameras: The system would require a network of cameras with different types or capabilities to capture images of the presenter's hand gestures.

Image processing and gesture recognition:The system would need an image processing and gesture recognition component to analyze the images captured by the cameras and detect the hand gestures. This could be accomplished using machine learning algorithms or other techniques.

Communication and control: The system would need to have a means of communicating with the presentation software and controlling it based on the detected hand gestures. This could be done using a wired or wireless connection.

Presentation software: The system would need to be compatible with a variety of presentation software, such as PowerPoint or Google Slides.

User interface: The system may include a user interface, like a touch screen or physical control panel, to enable the presenter to configure the system and adjust its settings.

In general, the architecture of a hand gesture-based interactive presentation system utilizing heterogeneous cameras would need to include components for capturing images, processing and recognizing hand gestures, communicating with and controlling the presentation software, and providing a user interface for the presenter.

2.10.2 Results

Gesture	Number of gestures	Number recognized	Recognition rate (%)	Overall rate (%)
Lining	84	81	96.4	
Circling	52	50	96.2	96.7
Pointing	43	42	97.7	

Figure 2.24: Result of proposed model

2.10.3 Advantages

Improved presenter mobility: The system could allow the presenter to move freely around the room without the need for a physical control device by enabling them to control the presentation using hand gestures.

The use of hand gestures to control the presentation could make the experience more interactive and engaging for the audience.

Increased flexibility: The use of heterogeneous cameras could improve the system's ability to detect hand gestures in various situations, making it more flexible and adaptable to different environments.

Greater convenience:A hand gesture-based interactive presentation system could be more convenient for the presenter, as they would not need to bring a separate device to control the presentation.

2.11 Comparison Table

Title	Dataset	Methodology	Disadvantages
Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks,2021	Jester dataset and Nvidia.	STSNN model	The current optical flow algorithm is still costly..
Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frames ,2022	VIVA and Nvidia &Deep network approach,Hand crafted based approach	This proposed model have robustness to scale and orientation change	Gesture spotting is not perfect
Gesture-Based Human Machine Interaction Using RCNNs in Limited Computation Power Devices,2021	EMG Dataset	Region based CNN approach	High computation capacities are needed for the majority of the problems that use CNN
Human-Computer Interaction Using Manual Hand Gestures in Real Time,20	Nvidia	PPI together to DMA	When implemeting a recursive funtion,a space limitation problem was faced since the ADSP BF533 processor must save the context in each iteration
A hand Gesture Based Interactive Presentation System Utilizing Heterogeneous Cameras	Nvidia	ML and Computer Vision	Currently one gesture interaction is limited to one hand ,web camera is only used for caliberation——

Table 2.1: Comparison Table of papers

A Hand Gesture Recognition Sensor Using Reflected Impulse	Jester dataset	(1-d)CNN,IR sensors	Low accuracy for some letters.
On-Air Hand-Drawn Doodles for IoT Devices Authentication During COVID-19	Nvidia	CNN and Kalman filter.	Cant use in the dark room Causes zigzag line due to unstable .
Video Hand Gestures Recognition Using Depth Camera and Lightweight CNN	VIVA	Depth camera and light weight CNN	Camera Hardware's are expensive Need separate hardware's.
Real-Time Gesture Detection Based on Machine Learning Classification of Continuous Wave Radar Signals	Nvidia and jester	Threshold detection (THD) classical machine learning (CML) Support vector machine (SVM) Stochastic gradient descent (SGD).	Need separate radar system. .
Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation	Nvidia	3D CNN	3DCNN requires more parameters than 2DCNN which is one of its disadvantages.

Table 2.2: Comparison Table of papers

Chapter 3

Methodology and Design

3.1 Proposed System

This paper presents a proposed AI virtual mouse system that uses hand gestures and hand tip detection to perform mouse functions on a computer using computer vision. The goal of the proposed system is to enable computer mouse cursor functions and scroll functions to be performed using a web camera or built-in camera on the computer, instead of using a traditional mouse device. Hand gestures and hand tip detection are used as a human-computer interface (HCI) with the computer. The AI virtual mouse system allows for the tracking of the fingertip of the hand gesture using a built-in camera or web camera, and enables the performance of mouse cursor operations and scrolling functions, as well as the movement of the cursor.

In this system, the user can control the mouse using hand gestures and hand tip gestures captured by a built-in camera or webcam. The captured frames are processed, and the recognized gestures are used to perform specific mouse functions. This method of mouse control is an alternative to using a wireless or Bluetooth mouse, which requires the use of additional devices such as a mouse, dongle, and battery.

The AI virtual mouse system was created using Python and OpenCV for computer vision. It used MediaPipe to track the hands and the tips of the hands. Pynput, Autopy, and PyAutoGUI were used to move around the window screen and perform actions like left click, right click, and scrolling. The model was able to work accurately and effectively in a real-world environment without requiring a GPU, only a CPU.

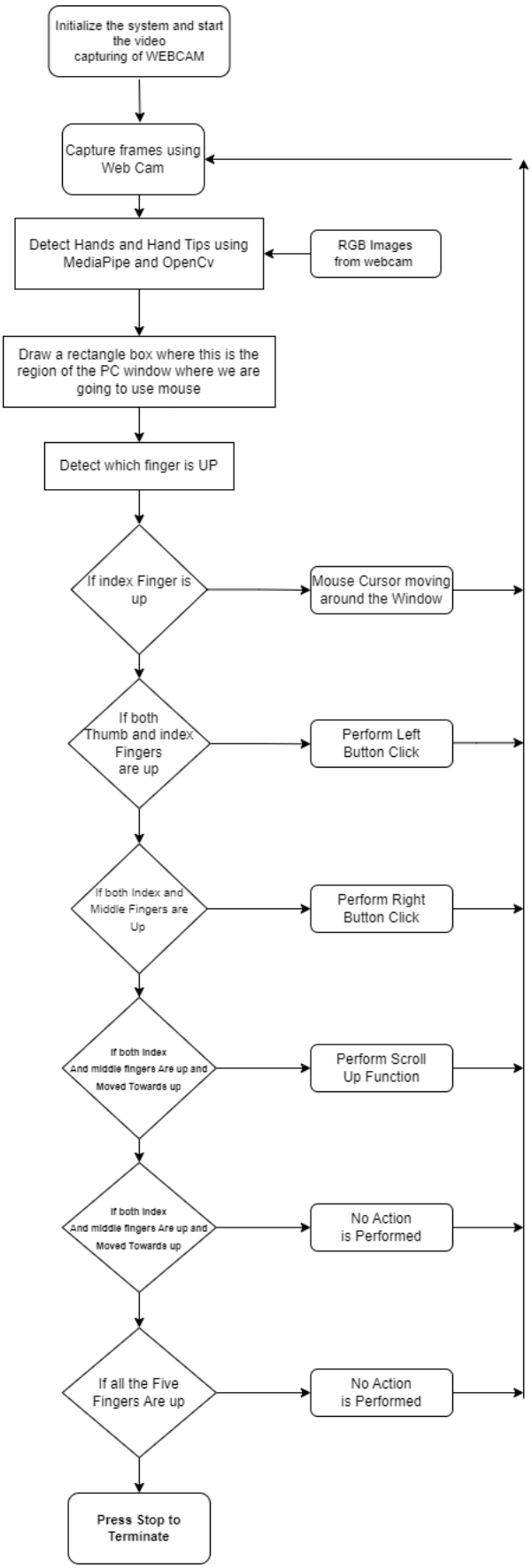
3.1.1 Problem Description and Overview

The AI virtual mouse is meant to be an alternative to traditional physical mice for people who may have difficulty using them due to physical limitations or limited space. For example, individuals with limited hand or wrist mobility may find it hard to use a physical mouse, which requires precise and controlled movements to work properly. The AI virtual mouse allows these users to control their computer's mouse functions through hand gestures and hand tip detection, which may be easier for them to do. The COVID-19 pandemic has also raised concerns about the transmission of the virus through physical contact with shared or public surfaces, including computer peripherals like mice. The AI virtual mouse allows users to control their computer's mouse functions without physical touch, reducing the risk of virus transmission through shared devices.

3.1.2 Objective

The AI virtual mouse is designed to be a replacement for the standard physical mouse that is typically used to control a computer's cursor and perform various functions. It uses a webcam or built-in camera to capture hand gestures and hand tip movements made by the user. These captured frames are then processed by the system to perform specific mouse functions, such as left clicking, right clicking, and scrolling. The use of hand gestures and hand tip detection allows the user to control the mouse functions without physical contact with a device, making it a more convenient and potentially safer alternative to traditional mouse systems.

3.2 Architectural Diagrams



widthheightfigure[h!] Conceptual Diagram

Chapter 4

Work Plan

4.1 Phase 1 plan

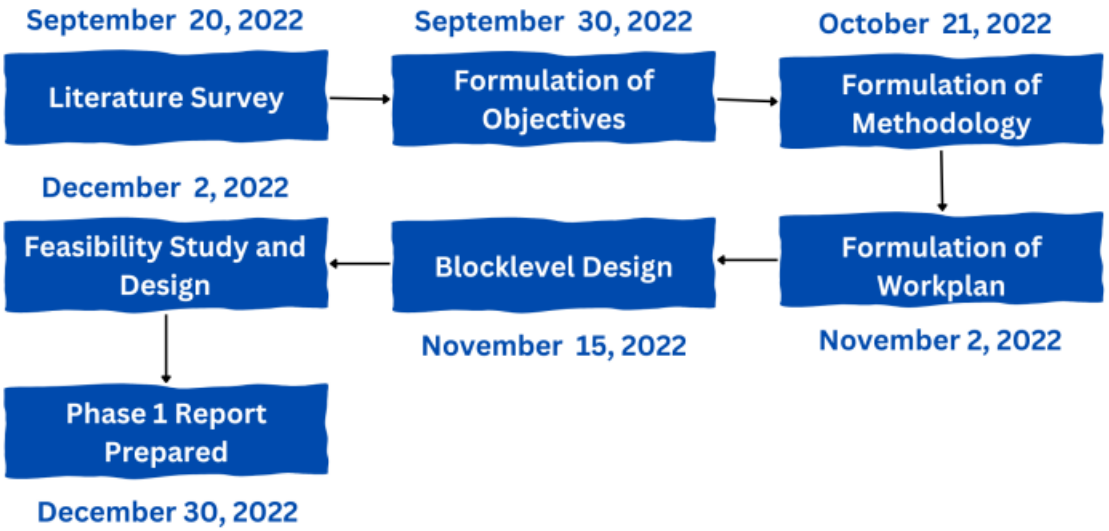


Figure 4.1: Phase1 Plan

4.2 Phase 2 plan

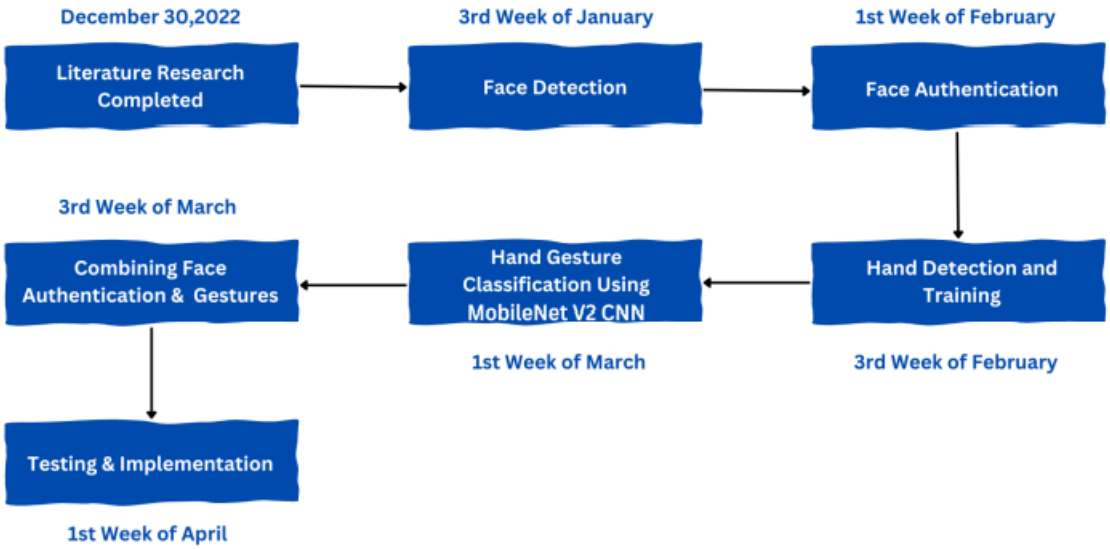


Figure 4.2: Phase2 Plan

Chapter 5

Conclusion

The AI virtual mouse system aims to allow users to control their computer's mouse cursor and perform various functions using hand gestures instead of a physical mouse. This is done by using a webcam or built-in camera to capture and detect hand gestures and hand tip movements made by the user. The system processes these captured frames to perform specific mouse functions, such as left clicking, right clicking, and scrolling. The AI virtual mouse can be used in a variety of real-world situations, including those where there is limited space to use a physical mouse or where individuals may have difficulty using one due to hand mobility issues. It can also be useful during the COVID-19 pandemic as it allows users to control their computer's mouse functions without the need for physical touch, reducing the risk of virus transmission through shared devices.

Bibliography

[1] Gibran Benitez-Garcia , Muhammad Haris, Yoshiyuki Tsuda , and Norimichi Ukita, , "Continuous Finger Gesture Spotting and Recognition Based on Similarities Between Start and End Frame" ,VOL. 23, NO. 1, JANUARY 2022

[2] Wenjin Zhang, Jiacun Wang, Senior Member,"Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks",VOL. 8, NO. 1, JANUARY 2021

[3] David Gonz´alez Le´on , Jade Gr¨oli , Sreenivasa Reddy Yeduri , Daniel Rossier , Romuald Mosqueron , Om Jee Pandey , and Linga Reddy Cenkeramaddi , "Video Hand Gestures Recognition Using Depth Camera and Lightweight CNN", VOL. 22, NO. 14, 15 JULY 202

[4] J. P. Giuffrida, A. Lerner, R. Steiner, and J. Daly, "Upper-extremity stroke therapy task discrimination using motion sensors and electromyography," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 16, no. 1, pp. 82–90, Feb. 2008.

[5] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," J. Neuroeng. Rehabil., vol. 9, no. 1, p. 21, 2012

[6]] A. Hollinger and M. M. Wanderley, "Evaluation of commercial forcesensing resistors," in Proc. Int. Conf. New Interfaces Musical Expression, Jun. 2006, pp. 1–4.

[7] H.-B. Xie, Y.-P. Zheng, and J.-Y. Guo, "Classification of the mechanomyogram signal using a wavelet packet transform and singular value decomposition for multifunction prosthesis control," Physiolog. Meas., vol. 30, no. 5, pp. 441–457, 2009

[8] N. Li, D. Yang, L. Jiang, H. Liu, and H. Cai, "Combined use of FSR sensor array and SVM classifier for finger motion recognition based on pressure distribution map," J. Bionic Eng., vol. 9, no. 1, pp. 39–47, 2012

[9]] R. Xu, S. Zhou, and W. Li, "MEMS accelerometer based nonspecificuser hand gesture recognition," IEEE Sensors J., vol. 12, no. 5, pp. 1166–1173, May 2012

[10] Y. Tenzer, L. P. Jentoft, and R. D. Howe, "The feel of MEMS barometers: Inexpensive and easily customized tactile array sensors," IEEE Robot. Autom. Mag., vol. 21, no. 3, pp. 89–95, Sep. 2014

[11] P. Pławiak, T. So´snicki, M. Nied´zwiecki, Z. Tabor, and K. Rzecki, "Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms," IEEE Trans. Ind. Informat., vol. 12, no. 3, pp. 1104–1113, Jun. 2016.

[12] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance," Scand. J. Rehabil. Med., vol. 7, no. 1, pp. 13–31, 1975

[13]] A. D. Roche, H. Rehbaum, D. Farina, and O. C. Aszmann, "Prosthetic myoelectric control strategies: A clinical perspective," Current Surgery Rep., vol. 2, no. 3, p. 44, 2014.

[14] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967

[15] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," Artif. Intell. Rev., vol. 43, no. 1, pp. 1–54, Jan. 2015.