PROJECT REPORT

INTELLIGENT VOICE ASSISTED ROBOTIC ARM

A PROJECT REPORT

Submitted by

MOHAMMED SHAHEER V K (CHN19AE015)

to

The APJ Abdul Kalam Technological University
in partial fulfillment of the requirements for the award of the Degree

of

Bachelor of Technology

In
Electronics & Instrumentation Engineering

Department of Electronics Engineering

College of Engineering Chengannur

Alappuzha 689121
Phone: +91479 2454125
JUNE 2023

DEPARTMENT OF ELECTRONICS ENGINEERING

COLLEGE OF ENGINEERING CHENGANNUR

CERTIFICATE

This is to certify that the project report entitled "**Intelligent Voice Assisted Robotic Arm**" submitted by **Mohammed Shaheer V K (CHN19AE015)** to APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Electronics and Instrumentation, is a bonafide record of the work carried out by them under our guidance and supervision.

Sri. Sylish S V	**Sri. Yasim Khan M**	**Prof. (Dr.) Laila D**
Project Guide	Internal Supervisor	HEAD OF THE DEPT

ACKNOWLEDGEMENT

Dedicating this project report to the Almighty God whose abundant grace and mercies enabled its successful completion, we would like to express our profound gratitude to all the people who had inspired and motivated us to make this project a success.

We would like to express our gratitude to our beloved Principal, *Prof. Smitha Dharan,* for the kind support and encouragement extended to us. Our sincere gratitude to *Prof. Laila D*, Head of the department, Electronics engineering, and we would like to express our deepest gratitude to our staff advisor and project coordinator *Sri. Yasim khan M,* Assistant professor, Department of electronics engineering for the incites and encouragement given by him to improve the project.

We thank our project guide *Sri. Sylish S V*, Assistant professor, Department of electronics engineering for all his support. We are greatly beholden to all the faculty members of the department of electronics engineering for their assistance.

Finally, we are grateful to our parents, who have given us tremendous support and forbearance throughout our work. We especially thank all our dear friends, for their help and encouragement throughout the work.

Mohammed Shaheer V K

\

ABSTRACT

Our project presents the development of an intelligent voice-assisted robotic arm capable of recognizing and manipulating different shapes. The system utilizes machine learning techniques, specifically the YOLOv5 algorithm which is an advanced algorithm used for machine learning and object detection, to detect and classify shapes based on voice commands. The detected shape's coordinates are then passed to the Scorbase software to control the Scorbot ER 4U robotic arm for pick-and-place operations. To train the shape recognition model, a dataset consisting of a collection image of cylinders, prisms, and cuboids was created. Each image was annotated, labeled, and augmented, to create a larger set of images for each shape. The machine learning model was trained using this dataset and the YOLOv5 algorithm, enabling accurate shape detection. Additionally, Scorbase execution sequence were developed to establish seamless communication between the voice recognition system and the Scorbot ER 4U robotic arm. Thorough testing and evaluation of the system's performance were conducted to validate its functionality and efficiency in real-world scenarios. The results of the project successfully demonstrate shape recognition, and shapes position manipulation.

CONTENT

Contents	Page No.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

VB Script	Microsoft Visual Basic Scripting Edition
OpenCV	Open-source Computer Vision
CNN	Convolutional neural network
YOLO	You Only Look Once
ASR	Automatic Speech Recognition
NLP	Natural language processing
API	Application Programming Interface
POS	Part of Speech
mAP	Mean Average Position
HDMI	High-Definition Multimedia Interface
WLAN	Wireless Local Area Network
RAM	Random-access memory
PoE	Power over Ethernet
SRL	Statistical Relational Learning

CHAPTER 1
INTRODUCTION

Robotic arm technology has become increasingly important in modern times, as it provides a flexible and efficient way to perform a wide range of industrial and domestic tasks. Robots comprises of mechanical, electrical, computing, and automation technologies. With advancements in automation, robots are becoming more capable of performing complex tasks with less human intervention. In this project, we aim to develop a system that enables users to control a robotic arm using voice commands.

Voice is a fundamental means of communication. As communication technologies have advanced, voice input has become an increasingly important interface for many systems, including robots. Using voice to control a robot allows users to free up their hands and work on other tasks while interacting with the machine in a more natural way. This is particularly useful for industrial applications, where workers need to perform multiple tasks simultaneously.

The goal of this project is to design a system to use the robotic arm and implement a robotic arm that can pick up and place a pre-specified object into separate divisions based on its shape. The system consists of both hardware and software components. The hardware includes the robotic arm, camera, and microphone, while the software is implemented using three different software tools: ScorBase, VBScript, and machine learning. To achieve the objective of the project, the system detects objects in an image based on command. This is done using contour detection techniques, where contours are computed from the detected edges, and the shapes of the objects are identified using a CNN algorithm. The system can be applied to industrial applications, where the robotic arm can be used to sort based on their shape with a high frequency.

To operate the system, a user provides voice command to the system through a microphone, and the Speech Recognition technology converts this input into text. The PyTorch framework and YOLO object detection model is used to process this text, and the name of the object is detected and displayed in the video. The images are then trained using the PyTorch framework on the cloud, and the weights file and configuration file are generated. The YOLO object detection model checks the confidence score of each pixel of the image and generates a boundary box if the object is near to the score of the trained image. The coordinates of the object are then recognized by finding the center point of the thickness of the bounding box. The marked X and Y coordinates are saved to the VB Script program to pick and place the object using the SCORBOT-ER 4u robotic arm.

In summary, this project aims to develop a system that enables users to control a robotic arm using voice commands and detects objects based on their shape. The system can be applied to industrial applications, where it can help to separate and segregate objects based on their shape with a high frequency, providing a flexible and efficient way to perform complex tasks.

1.1 Background

The background of voice command-based pick and place robotic arm system can be traced to the broader field of human-robot interaction and robotic automation. The field of human-robot interaction focuses on developing intuitive and efficient ways for humans to communicate with robots. Traditional methods of interaction, such as physical buttons or graphical user interfaces, have limitations in terms of ease of use and flexibility. Researchers have been exploring alternative modalities, including voice commands, to enable more natural and intuitive communication between humans and robots. Voice recognition technology has advanced significantly in recent years, driven by advancements in machine learning and natural language processing techniques. These advancements have made it possible to accurately interpret and understand human speech, enabling voice-based interfaces for a variety of applications. Applying these technologies to robotics allows for voice command-based control of robotic systems. Pick and place operations are common in manufacturing, logistics, and various industries where objects need to be accurately picked up from one location and placed in another. Robotic arms are commonly used for these tasks due to their dexterity and precision. Traditional pick and place systems have relied on visual or tactile sensing techniques to identify and manipulate objects. However, voice-based approaches have emerged as an alternative, leveraging sound-based sensing and shape detection algorithms. Voice-based shape detection involves using sound or vibrations to identify and recognize different shapes. By analyzing the voice signals generated when objects are tapped, struck, or interacted with, it is possible to extract shape-related features and classify objects accordingly. This approach can offer advantages in certain scenarios where visual or tactile sensing is limited or impractical.

Bringing these areas together, voice command-based pick and place robotic arm systems leverage voice recognition and natural language processing to interpret voice commands from human operators. The voice-based shape detection algorithms enable the robotic arm to identify and manipulate objects based on their shapes. By integrating these technologies, such systems aim to provide an intuitive, efficient, and hands-free method for humans to interact with robotic arms, enhancing automation and productivity in various industries.

1.2 Problem Definition

The problem that intelligent voice assisted robotic arm seeks to address is the need for precise, efficient, and safe manipulation of objects of varying shapes and sizes in assisting people as table top robot to pick medicines or other necessary items and pass to them by just giving voice command this can also be used in different industries, including manufacturing, warehousing, and medical applications like old age or parallelized patient assistance. The traditional methods of manually adjusting the robotic arm's grip to handle different shapes of objects can be time-consuming, inaccurate, and even dangerous in some situations.

1.3 Objectives

Shape Detection and Recognition:

Develop algorithms to accurately detect and recognize shapes based on voice input. Implement machine learning techniques to classify shapes such as cylinders, cuboids, prisms, etc. Ensure robustness and accuracy in shape identification.

Shape Localization and Positioning:

Determine the precise location and coordinates of the detected shapes within the environment. Use computer vision techniques to extract positional information, such as the center point or bounding box, for guidance during pick and place.

Robotic Arm Guidance:

Develop a system to guide a robotic arm based on the detected shape and its positional information. Implement motion planning algorithms to ensure precise movement and manipulation of the objects.

Reliable Object Manipulation:

Enable the robotic arm to perform precise and reliable manipulation of the detected objects. Incorporate control algorithms to ensure accurate gripping, lifting, and placement of the objects.

Automation and Workflow Optimization:

Automate the object handling process to improve efficiency and reduce human intervention. Optimize the workflow by seamlessly integrating the system with existing automation systems or production lines.

Safety Measures:

Implement collision detection and avoidance mechanisms to prevent accidents or damage during the pick and place process. Ensure the system complies with safety standards and guidelines to protect both the objects and the surrounding environment.

Scalability and Adaptability:

Design the system to be scalable, allowing it to handle a higher volume of objects or different types of shapes. Ensure adaptability to accommodate changes in the production environment or the introduction of new objects.

1.4 Scope and Limitations

The scope of the intelligent voice-assisted robotic arm for pick and place, which uses voice input to command shape detection and coordinates determination, involves the development and implementation of a system that integrates voice recognition, machine learning, and robotic manipulation. The system aims to accurately detect and manipulate objects based on their shapes, guided by voice commands.

Implementing a voice recognition system capable of understanding and interpreting voice commands that specify the name of the desired shape. This involves speech-to-text conversion and natural language processing techniques to extract the shape name from the voice input. Developing a machine learning model or algorithm to detect and classify shapes based on voice input. The model should be trained on a dataset of a different shapes and capable of accurately recognizing the commanded shape. Incorporating methods to determine the coordinates or position of the detected shape within the workspace.

The system should provide commands to guide the robotic arm to accurately grasp and move the detected object to the desired location. Designing an intuitive user interface that allows users to interact with the system through voice commands. The interface should provide feedback on the detected shape, its coordinates, and the execution of pick and place actions. While the intelligent voice-assisted robotic arm for pick and place presents exciting possibilities, it also has limitations that need to be considered. The accuracy of the speech recognition system can be influenced by factors such as background noise, speaker accents, or variations in pronunciation. Achieving high accuracy in recognizing the shape names from voice input challenging in real-world environments.

Training the machine learning model or algorithm for shape detection requires a sufficiently diverse and representative dataset of different shapes.

CHAPTER 2

LITERATURE SURVEY

2.1 Overview of Arms

S. Routray, A. T. John, A. Syed and P. Jadhav [1] proposed a solution for inverse kinematics by utilizing both geometric and iterative fusion-based modeling approaches. Geometric modeling involves using mathematical equations and algorithms to represent the kinematic structure of the robotic arm. It establishes the relationship between joint angles and the position/orientation of the end-effector and employs an iterative fusion-based modeling technique, which likely involves an iterative process of refining the inverse kinematics solution. This approach involves combining and refining multiple models or algorithms to improve the accuracy and efficiency of robotic arm control.

J. U.K., I. V., K. J. Ananthakrishnan, K. Amith, P. S. Reddy and P. S. [2] proposed project that incorporates a robotic arm specifically designed to assist elderly individuals with their daily tasks. The robotic arm is controlled based on voice commands received from the elderly users, which are processed and interpreted by the system's natural language processing module. It can perform precise movements and has the necessary dexterity to manipulate objects in the environment. The robotic arm is equipped with sensors to perceive its surroundings and interact with objects safely. The robotic arm serves as a physical extension of the system, providing support, assistance, and companionship to the elderly users in their daily lives.

S. M, S. K. C, S. P. J, H. R. A, R. B. P and P. Naveen [3] presents a robotic arm system designed for material handling tasks, which is controlled using voice commands. The robotic arm serves the purpose of handling objects and materials in various industrial and manufacturing settings. The robotic arm is integrated with voice recognition capabilities, allowing users to issue commands verbally. The system processes the voice commands and translates them into corresponding actions performed by the robotic arm. The aim is to enhance efficiency and productivity in material handling operations by leveraging voice control technology. The paper discusses the development process, implementation details, and experimental results of the robotic arm system, highlighting its potential benefits in real-world industrial applications.

T. R. Deshpande and S. U. Sapkal [4] proposed the development of a vision-enabled articulated robotic arm specifically designed for handling simple objects. The paper focus on the integration of vision capabilities and grasping strategies into the robotic arm system. The vision system allows the robotic arm to perceive and recognize simple objects in its environment. The paper discusses the development process, including the selection and integration of appropriate sensors and cameras for vision-based object detection. The articulated robotic arm is designed to mimic human arm movements and is equipped with specialized grippers or end-effectors to enable grasping and manipulation of the detected objects. The grasping strategies implemented in the system aim to ensure successful and stable grasping of the objects.

2.2 Voice-Assisted Systems

J. U.K., I. V., K. J. Ananthakrishnan, K. Amith, P. S. Reddy and P. S. [2] presents a voice-assisted personal assistant robot system designed to cater specifically to the needs of elderly individuals. By utilizing voice recognition and natural language processing technologies, the system offers a user-friendly and intuitive interaction method. The system utilizes voice recognition and natural language processing techniques to understand and respond to user commands. It is designed to perform various tasks based on user requests, such as providing information, setting reminders, making phone calls, and controlling smart home devices. The voice-controlled interface aims to provide a convenient and user-friendly interaction method for elderly users, allowing them to communicate with the robot intuitively.

S. M, S. K. C, S. P. J, H. R. A, R. B. P and P. Naveen [3] presented a voice-assisted system for controlling a material handling robotic arm. The system aims to enhance the efficiency and ease of operation in material handling tasks by utilizing voice commands for controlling the robotic arm. The authors developed and implemented the system, incorporating voice recognition technology to interpret user commands and control the movements of the robotic arm accordingly. The paper highlights the successful development and implementation of the voice-assisted material handling robotic arm, showcasing its potential to improve productivity and streamline operations in various industries that involve material handling tasks.

N. Andrews, S. Jacob, S. M. Thomas, S. Sukumar and R. K. Cherian [6] presented a voice-assisted system designed to control a cost-effective robotic arm for individuals with disabilities. The research aims to provide an affordable solution that allows individuals with limited mobility or dexterity to operate a robotic arm independently. By incorporating voice recognition technology, the system can interpret user commands and translate them into corresponding movements of the robotic arm. The low-cost nature of the robotic arm makes it accessible to a wider range of users, addressing the affordability concerns typically associated with assistive technologies for individuals with disabilities.

P. Vashistha, J. P. Singh, P. Jain, and J. Kumar [7] introduced a voice-assisted system called Neobot , which is built using Raspberry Pi. Neobot serves as a voice-operated personal assistant, leveraging Raspberry Pi's capabilities to enable voice recognition and response. The system allows users to interact with Neobot using voice commands, enabling tasks such as information retrieval, setting reminders, and controlling various functionalities. The integration of Raspberry Pi and voice recognition technology empowers users to access personalized assistance through voice interactions, making Neobot a versatile and user-friendly personal assistant.

2.3 Related Work and Existing Solutions

Routray et al. [1] proposed a solution for inverse kinematics in a robotic arm using geometric and iterative fusion-based modeling techniques. Their work focused on achieving precise control and accurate movement of the robotic arm. U.K. et al. [2] presented a voice-controlled personal assistant robot designed specifically for elderly individuals, enabling them to interact with the robot through voice commands. Their research aimed to aid and support the elderly population. Furthermore, M et al. [3] developed and implemented a material handling robotic arm that was controlled by voice commands. Their work showcased the integration of voice control with robotic arm functionalities for efficient material handling tasks. Deshpande and Sapkal [4] focused on developing a vision-enabled articulated robotic arm with grasping strategies for simple objects, utilizing computer vision techniques for object recognition and manipulation. Chao et al. [5] explored the combination of a robotic arm with visual images for transparent object recognition, presenting a system that incorporated both visual perception and robotic arm control. Additionally, Andrews et al. [6] proposed a low-cost robotic arm controlled using voice recognition techniques, targeting individuals with different abilities. Vashistha et al. [7] developed a Raspberry Pi-based voice-operated personal assistant named Neobot, which showcased the integration of voice commands for controlling various functions.

By reviewing these existing solutions, it is evident that significant progress has been made in the integration of voice control, computer vision, and robotic arm functionalities. These works serve as valuable references for the development of an intelligent voice-assisted robotic arm system, incorporating features such as inverse kinematics, material handling, vision-based object recognition, and user-friendly voice control interfaces.

While the aforementioned papers have made significant contributions to the field of intelligent voice-assisted robotic arms, our project introduces several advancements and unique elements that distinguish it from the existing literature. Our project specifically focuses on implementing intelligent voice control using the Scorbot platform. This emphasis on the Scorbot robot sets our project apart as it leverages the capabilities and functionalities of this specific robotic arm, which may differ from other robotic arm platforms used in the referenced papers. The utilization of Scorbase software and VB script allows for seamless integration between the voice commands and the Scorbot's movements, providing a robust and efficient control interface. Furthermore, the incorporation of OpenCV, a widely adopted computer vision library, enables us to leverage computer vision techniques for tasks such as bounding box detection and object recognition. This integration of computer vision technologies adds a visual perception component to your intelligent voice-assisted robotic arm, enhancing its capabilities and enabling advanced functionalities such as autonomous object manipulation. By implementing sophisticated signal processing techniques, we aim to extract meaningful voice commands and achieve accurate recognition of user instructions. This emphasis on microphone-based voice recognition distinguishes our project from others, as it offers an intuitive and natural interface for controlling the robotic arm.

CHAPTER 3

METHODOLOGY

Voice input is sent to a program using a microphone connected to the device. Then this voice input is processed in real-time and stored. The captured voice is then pre-processed to remove background noise and enhances the quality of the voice signal using voice processing libraries. The pre-processed voice signal is then fed as input to the speech recognition API, which transcribes the voice into text using automatic speech recognition (ASR) techniques such as Convolutional Neural Networks (CNN). The recognized text is returned as output by the API. The recognized text is further processed using natural language processing (NLP) techniques to extract useful information such as keywords, entities, and sentiment. This is done using NLP libraries. The text output is obtained using text-to-speech (TTS) APIs.

A dataset is created by collecting images and labelling them into 3 classes cylinder, prism, cuboid and a CNN model is trained using the dataset. Now the dataset contains images with objects that need to be detected. The image of the shapes is obtained through capturing video signals from the webcam which goes for object detection using machine learning algorithm YOLOv5 which is based on CNN.

The YOLOv5 algorithm uses a model, which is trained on the labelled dataset. The model architecture includes multiple convolutional layers and a fully connected layer for object detection. The labelled dataset is split into training and validation sets. The model is trained on the training set by minimizing the loss function using backpropagation. The validation set is used to evaluate the performance of the model. Then this model is fine-tuned by adjusting the hyperparameters to improve its performance. The hyperparameters include the learning rate, batch size, and number of epochs. Once the model is trained and fine-tuned, it is ready for inference. Inference involves feeding the model with new images and detecting the objects in them. The model outputs the bounding boxes, class probabilities, and confidence scores for each object detected in the image and post-processing the output of the model is done to remove duplicate detections, filter out low-confidence detections, and refine the bounding boxes. After object detection, the detected objects are visualized in the output image or video frames using bounding boxes.

OpenCV provides functions to draw bounding boxes around the detected objects and label them with their corresponding class names. The coordinates of the bounding boxes are extracted for each detected object, which is done using OpenCV to access the x, y coordinates, height, and width of the bounding boxes. VB Script is updated with coordinates of the detected image. VB script loads coordinates into robotic arm. The robotic arm is programmed to move to the specified coordinates and perform the required action. For example, if the detected shape is a cuboid, the robotic arm is programmed to move to the center of the cuboid and pick it up and place it in a designated location. These are tested by running the VB script and verifying that the robotic arm moves to the correct location and performs the required action based on the coordinates of the bounding box.

3.1 Voice Recognition

The actual implementation of voice recognition systems can vary depending on the specific technology and algorithms used. Here, we are using an advanced approach –Natural language processing which is effective in many state-of-the-art voice recognition systems. Natural Language Processing (NLP) techniques are often employed in voice recognition systems to process and understand the recognized text output. Integrating NLP into the voice recognition process is done in various steps.

The voice recognition system converts the captured voice signals into textual representations. Then, the recognized text is tokenized into individual words or sub word units. It breaks down the text into meaningful units, which serve as the input for subsequent NLP tasks. Then, by part of speech tagging, it assigns grammatical labels to each token in the text and to identify the syntactic structure of the sentence, which can be useful for subsequent analysis. After extracting important information from the recognized text, dependency parsing analyzes the grammatical structure of the sentence and establishes relationships between words, assigning labels such as subject, object, or modifier, which helps in understanding the dependencies and syntactic relationships among the words. SRL helps in understanding the roles played by different entities and their relationships. Coreference resolution aims to identify and resolve references to the same entity across the text. Intent recognition identifies the underlying intent or purpose behind a user's input. In a conversational setting, dialogue management ensures coherent and meaningful conversations between the user and the voice recognition system.

3.2 Object detection

YOLOv5 is used for object detection, that builds upon the YOLO (You Only Look Once) concept. The dataset of annotated images is gathered that contain the shapes (cuboid, prism, cylinder). The shapes are annotated with bounding boxes and appropriate class labels are assigned and model configuration of YOLOv5 TensorFlow is designed. The annotated dataset is converted into the required format for training with YOLOv5.

The training process is initiated using the prepared dataset and executed. The training process will optimize the YOLOv5 model to detect shapes based on the provided dataset. After training, evaluation of the trained model is done using the validation set. The evaluation provides insights into metrics such as precision, recall, and mean Average Precision (mAP), which reflect the accuracy of shape detection. Once the model is trained and evaluated, shape detection is done on new images using the trained YOLOv5 model. During inference, the trained model is used to predict the bounding boxes for the center of thickness. This was achieved by adjusting the post-processing step of the object detection algorithm to output bounding boxes centered around the predicted center of thickness.

3.3 Coordinates

OpenCV provides functions to draw bounding boxes around the detected objects and label them with their corresponding class names. The coordinates of the bounding boxes are extracted for each detected object, which is done using OpenCV to access the x, y coordinates, height, and width of the bounding boxes. VB Script is updated with coordinates of the detected image. To detect the shape of a 3D object using OpenCV, the 2D image coordinates of the object's contours are analyzed. We start by loading the image and converting it to grayscale for contour detection. Additional preprocessing steps are applied for thresholding or filtering of background images. The contours in the grayscale image are detected. Outermost contours are retrieved and specified. After iteration, perimeter and number of vertices is calculated which eases the detection of shape.

3.4 VB Script

The coordinate functions calculate new values based on input parameters. In object detection using machine learning, coordinates play a crucial role in identifying and localizing objects within an image or video. The coordinates represent the position and boundaries of the detected objects, allowing the machine learning algorithm to understand their spatial location within the input data. From the coordinates obtained from the bounding box of OpenCV, we name them to variable for storage - x1, y1, x2, y2, z1 and z2. These coordinates represent the top-left, bottom-left and bottom-right, corners of the bounding box. Using these coordinates, we calculate the width and height and based on the aspect ratio (width divided by height) and determine the shape of the object. VBScript serves as a bridge between the object detection algorithm and external systems or processes. VBScript can be used to pass the coordinates or other necessary data to the model and retrieve the detection results. After object detection, the resulting coordinates need to be interpreted and utilized further. VBScript helps with post-processing tasks, such as filtering out false positives, performing additional calculations or transformations on the detected objects coordinates, or integrating the results with other systems or applications.

SCRIPT:

```
x1 =
y1 =
Function CalcX(X)
  X=(x1/4.3)*1000
   CalcX=X
End Function
Function CalcY(Y)
    Y=(y1/4.3)*1000
   CalcY=Y
End Function
Function CalcZ(Z)
  Z=20000
   CalcZ=Z
End Function
```

3.5 SCORBASE COMMANDS

Load script file: XYZ.VBS

Set Variable S1 = 100

Set Variable S2 = 100

Set Variable S3 = 100

Set Variable Y = 0

Set Variable X = 0

Set Variable Z1 = 0

Set Variable Z2 = 0

Go Linear to Position 1 Speed 50 (%)

Record Present Position as Position 99!

Teach Position 100 by XYZ. Relative to: 99. Coordinates: 0 0 7000 0 0

Go Linear to Position 100 Speed 50 (%)

Set Variable X = SCRIPT.CALCX(X)

Set Variable Y = SCRIPT.CALCY(Y)

Teach Position S1 by XYZ. Relative to: 99. Coordinates: X Y 7000 0 0

Go Linear to Position S1 Speed 50 (%)

Open Gripper

Set Variable Z1 = SCRIPT.CALCZ1(Z1)

Teach Position S2 by XYZ. Relative to: 99. Coordinates: X Y Z1 0 0

Go Linear to Position S2 Speed 50 (%)

Close Gripper

Set Variable Z2 = SCRIPT.CALCZ2(Z2)

Teach Position S3 by XYZ. Relative to: 99. Coordinates: X Y Z2 0 0

Go Linear to Position S3 Speed 50 (%)

Go Linear to Position 2 Speed 50 (%)

Open Gripper

Go Linear to Position 3 Speed 50 (%)

Go Linear to Position 4 Speed 50 (%)

Close Gripper

CHAPTER 4
SYSTEM DESIGN AND ARCHITECTURE

The system is designed for object detection, an intelligent voice-assisted robotic arm using SCORBOT ER4U, a microphone, and a webcam. This involves the integration of various components to enable voice commands, visual perception, and robotic arm control. Hardware components include SCORBOT ER4U which serves as the physical manipulator for executing tasks, a microphone captures voice commands from the user and webcam provides visual input for object recognition and perception. The outline includes voice recognition, visual perception, robotic arm control, system integration and error handling.

4.1 Overall System Architecture

The purpose of an intelligent voice-assisted robotic arm for shape detection is to enable a robotic system that can receive voice commands from users and autonomously detect and manipulate objects based on their shapes. The integration of voice assistance and shape detection capabilities provides a more intuitive and user-friendly interaction paradigm, allowing users to communicate their intent naturally and effortlessly. The primary goal is to enhance the user experience by enabling natural language interaction with the robotic arm. Users can communicate their intentions using voice commands, simplifying the control process and making it accessible to a wider range of users. The system aims to achieve high accuracy in shape detection, ensuring reliable identification and localization of objects based on their shapes. This involves employing robust computer vision algorithms and machine learning techniques for shape recognition. The system aims to enable the robotic arm to autonomously manipulate objects based on their detected shapes. The goal is to develop algorithms for motion planning, grasp planning, and object manipulation to perform tasks accurately and efficiently. The system intends to be versatile and adaptable to different shapes and objects. It should support the detection and manipulation of various shapes, allowing for customization and flexibility in different application scenarios.

The robotic arm serves as the physical manipulator in the system, enabling the execution of various tasks such as picking up objects, sorting them based on shape, or performing shape-based assembly. It provides five degrees of freedom and precise control over its joints, allowing it to move in a coordinated manner to perform complex manipulations. It is responsible for physically interacting with objects based on their detected shapes, executing the desired tasks with accuracy and efficiency.

The microphone is used for capturing voice commands from the user. The microphone converts sound waves into electrical signals that can be processed by the voice recognition system. The microphone enables natural language interaction by allowing users to communicate their intentions and commands to the system using their voice, creating a user-friendly and intuitive interface.

The HD webcam provides visual input for object recognition and perception. The webcam captures real-time video of the environment, which is then processed to identify objects and extract shape information. The webcam plays a crucial role in shape detection by providing visual data that allows the system to perceive and recognize objects based on their shapes, enabling accurate manipulation and interaction with the environment. The voice recognition system converts voice commands into text or actionable instructions.

The system uses algorithms and models to analyze the voice input from the microphone and transcribe it into text format. The voice recognition system allows users to interact with the robotic arm using natural language, providing a convenient and intuitive way to communicate commands and intentions. Shape detection algorithms analyze visual input to identify and recognize shapes. These algorithms utilize computer vision techniques, such as edge detection, contour analysis, or deep learning models, to extract shape information from the video stream captured by the webcam. Shape detection algorithms enable the system to perceive and recognize different shapes, allowing the robotic arm to interact with objects based on their shapes, perform shape-based tasks, or make informed decisions during manipulation. The control interface facilitates communication between the system and the robotic arm. The control interface translates commands and instructions from the voice recognition and shape detection modules into control signals that can be understood by the robotic arm. The control interface enables seamless integration between the different system components, ensuring the robotic arm moves according to the desired tasks based on shape detection and voice commands.

4.2 Integration and Communication

The intelligent voice-assisted robotic arm project involves the integration of various software components to enable communication and interaction between the user, the voice input system, and the SCORBOT robotic arm. The project utilizes a microphone as the voice input device, allowing the user to provide voice commands. The voice commands are processed using Natural Language Processing (NLP) techniques. NLP enables the system to understand and interpret the user's spoken instructions, converting them into actionable commands for the robotic arm. The communication between the voice input system and the SCORBOT robotic arm is facilitated by the software components involved.

One of the key software tools used in this project is Scorbase, which serves as the interface between the robotic arm and the controlling system. Scorbase provides the necessary commands and functions to control the arm's movement and actions based on the instructions received. For image processing, the project employs OpenCV, a widely used open-source computer vision library. OpenCV allows for various image processing tasks, including object recognition and shape detection. In this case, OpenCV is utilized to process the visual information obtained by the robotic arm's camera system. It employs techniques such as Convolutional Neural Networks (CNN) for object recognition and bounding box generation.

The integration of Scorbase, Machine learning, and VB Script enables seamless coordination between the voice input system, image processing, and the robotic arm's actions. The processed voice commands and visual information from machine learning are utilized to determine the appropriate actions for the SCORBOT robotic arm.

4.3 Hardware Components

4.3.1 SCORBOT

SCORBOT is a family of robotic manipulators from Eshed Robotec, now called Intelitek. These robotic arms are equipped with a microprocessor or microcontroller-based controller, depending on the type of the robot. The controller when interfaced with a computer installed with the software can receive a set of commands or instructions and communicate with the robot using different channels, like serial, parallel, USB, etc., depending on the controller type, and send signals to activate the robot motors accordingly.

The complete SCORBOT package consists of the following parts:

- The Robot Arm - SCORBOT ER-4u
- Controller
- Software

SCORBOT-ER-4u

The SCORBOT-ER-4u shown in Figure 4.1, is a versatile and reliable system for industrial robotics training and education. The Scorbot-ER4u robot arm can be mounted on a tabletop, pedestal or linear slide base. It has five revolute joints. With the gripper attached, the robot has five degrees of freedom. Its body is divided into links and joints with a gripper attached to the end of the forearm, and their movements are described in Table 4.1. For moving the links and joints of the robot, DC motors, different types of arrangements involving gears, timing belts, pulleys and lead screw are used. Encoders fixed on the motors and microswitches provide feedback to find the position, angle and speed of the links and joints within the robot's workspace. The main important parts of robotic are shown below:

- Motors
- Encoders
- Microswitch
- Transmissions
- Gripper

Axis	Joint Name	Motion	Motor No.
1	Base	Rotates Body	1
2	Shoulder	Raises and Lowers the Upper Arm	2
3	Elbow	Raises and lowers the Forearm	3
4	Wrist Pitch	Raises and lowers the end effector (gripper)	4+5
5	Wrist Roll	Rotates the end effector	4+5

		(gripper)	
Gripper	Gripper	Opens and closes the end effector (gripper)	6

Table 4.1 Joints and Its Movements of SCORBOT-ER-4u

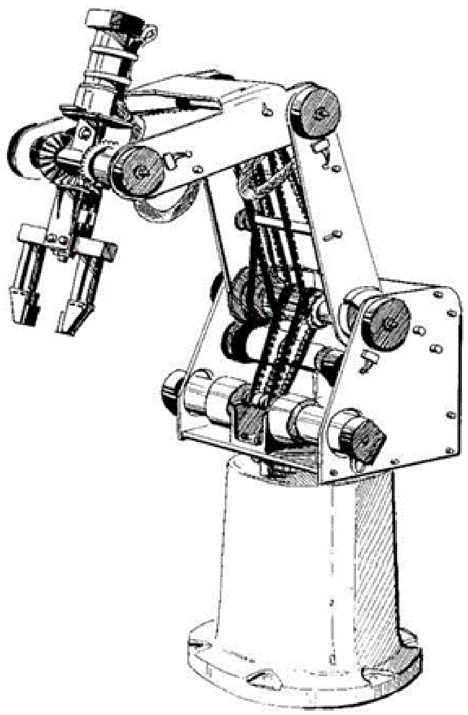

Fig: 4.1 SCORBOT-ER-4u

Controller

Intelitek ships the SCORBOT with its controller and software depending on the type of robot arms. Figure 4.2 shows a controller available to control the SCORBOTs. Table 4.2 provides information about the controller that are using.

Fig: 4.2 controller

Specifications	Details
Dimensions	315(L)x223(W)x117(H)mm (12.4"x8.8"x4.6")
Microcontroller	Full featured, NEC V853 RISC 32-bit microcontroller
Power Requirements	110/220VAC (+15%, -10%), 2A max, 50/60Hz
User Power Supply	12VDC 0.1A max
Led Indicators	Main power, bicolor: green: power on communication with PC red: power on; no communication with PC flashing: power on; PC USB communications timeout Motors Power ON (green), Emergency (red) 8 digital inputs (green), 8 digital outputs (orange)
Weight	7 kg (15.4 lb.)
Inputs	• 8 digital inputs PNP/NPN (High/Low) configurable, 0-24VDC max. • 4 analog inputs (8-bit resolution): input voltage 0–10VDC

Outputs	8 digital outputs: • 1 – 4: relays, 1.0A max. • 5 – 8: sink/source configurable open collectors Sink: 15VDC, 0.5A max. for each output Source: 15VDC, 50mA max. for all outputs combined 2analog outputs (8-bit resolution): output voltage 0–10VDC, 20mA max
Ambient Operating Temperature	0°C-55°C (32°F-131°F)
Type of Control	• Real-time • PID (Proportional, Integral, Differential) PWM (Pulse-Width Modulation)
Communications	• USB connection to PC, Plug-and- Play without rebooting • Integrated RS232 port for Teach Pendant connection
Internal Power Supplies	• Servo: 24V (depending on input voltage & load) Digital: 5V, +15V, -12V

Table 4.2 Controller specifications

4.3.2 WEB CAMERA

HD Webcam C270 is used as a web camera for capturing images of objects from a fixed distance. C270 is a 720p camera with a resolution of 1280 x 720 pixels. It can take a 3MP photo which can be software enhanced. This webcam will automatically adjust to poor lighting conditions to produce the best possible image. It can also remove background noise and hence delivers clear voice. The key features of the HD Webcam C270 are:

1. Resolution: The C270 webcam offers 720p HD resolution, which is suitable for shape detection. It provides good image quality.
2. Frame rate: The webcam captures video at up to 30 frames per second (fps). It provides smooth video output for real-time shape detection.
3. Fixed focus: C270 has a fixed focus lens, the focus is set at a specific distance and cannot be adjusted. It is suitable for capturing shapes within a certain range from the camera.
4. Built-in microphone: The C270 webcam includes a built-in microphone, which is useful for capturing voice along with the shape detection video.
5. Compatibility and connectivity: The C270 is compatible with various operating systems, including Windows, macOS, and Linux. It is connected to a computer via USB.

Fig: 4.3 HD Webcam C270

4.3.3 RASPBERRY PI-4

Raspberry Pi 4 Model B is the latest product in the popular Raspberry Pi range of computers. It offers ground-breaking increases in processor speed, multimedia performance, memory, and connectivity compared to the prior-generation, while retaining backwards compatibility and similar power consumption. For the end user, Raspberry Pi 4 Model B provides desktop performance comparable to entry-level x86 PC systems. This product's key features include a high-performance 64-bit quad-core processor, dual-display support at resolutions up to 4K via a pair of micro-HDMI ports, hardware video decode at up to 4Kp60, up to 4GB of RAM, dual-band 2.4/5.0 GHz wireless LAN, Bluetooth 5.0, Gigabit Ethernet, USB 3.0, and PoE capability (via a separate PoE HAT add-on). The dual-band wireless LAN and bluetooth have modular compliance certification, allowing the board to be designed into end products with significantly reduced compliance testing, improving both cost and time to market.

Fig 4.4 Raspberry Pi 4 Model B

4.4 Software Requirements

4.4.1 ScorBase

ScorBase robotics control software provides a comprehensive and intuitive tool for programming and operating robotic work cells. The software allows students to experiment with a variety of simulated work cells as part of the Intelitek curriculum or independently. The software has several operational levels, each with progressively more functionality, making the software suitable for both novice and advanced users. In addition to Intelitek robotic arms, the software supports peripheral servo axes and both digital and analog I/Os, thereby providing a comprehensive tool for programming and operating an entire robotic work cell. Menus and dialog boxes can be displayed in various configurations, making programming and control much easier. Positions can be recorded, for example, while programs are being edited. ScorBase provides a user-friendly interface and visible interface for robotic program execution.

4.4.2 VS Code

Visual Studio Code (VS Code) is a widely used source code editor developed by Microsoft. It is designed to provide developers with a lightweight yet powerful tool for writing and editing code across various programming languages. VS Code offers a user-friendly interface with a range of features and extensions that enhance productivity and customization. It supports syntax highlighting, intelligent code completion, debugging capabilities, version control integration, and a built-in terminal, among other functionalities. Its extensive marketplace allows users to extend its functionality by installing a wide range of extensions for specific programming languages, frameworks, and tools. VS Code's versatility, performance, and cross-platform compatibility make it a popular choice among developers for different types of projects, from web development to data science.

4.4.3 VBScript

VBScript is an active scripting language developed by microsoft which is modeled on Visual Basic. It allows Microsoft system administrators to generate powerful tools for managing computers with error handling, subroutines, and other advanced programming constructs. VBScript functions are self-contained blocks of code that perform a given function and then return a value. In our project, VBScript is used to input the coordinates to the ScorBase program as it cannot be directly inputted into the ScorBase program. It also provides a platform for mathematical computations to be performed on the inputs.

CHAPTER 5

WORKING

5.1 BLOCK DIAGRAM

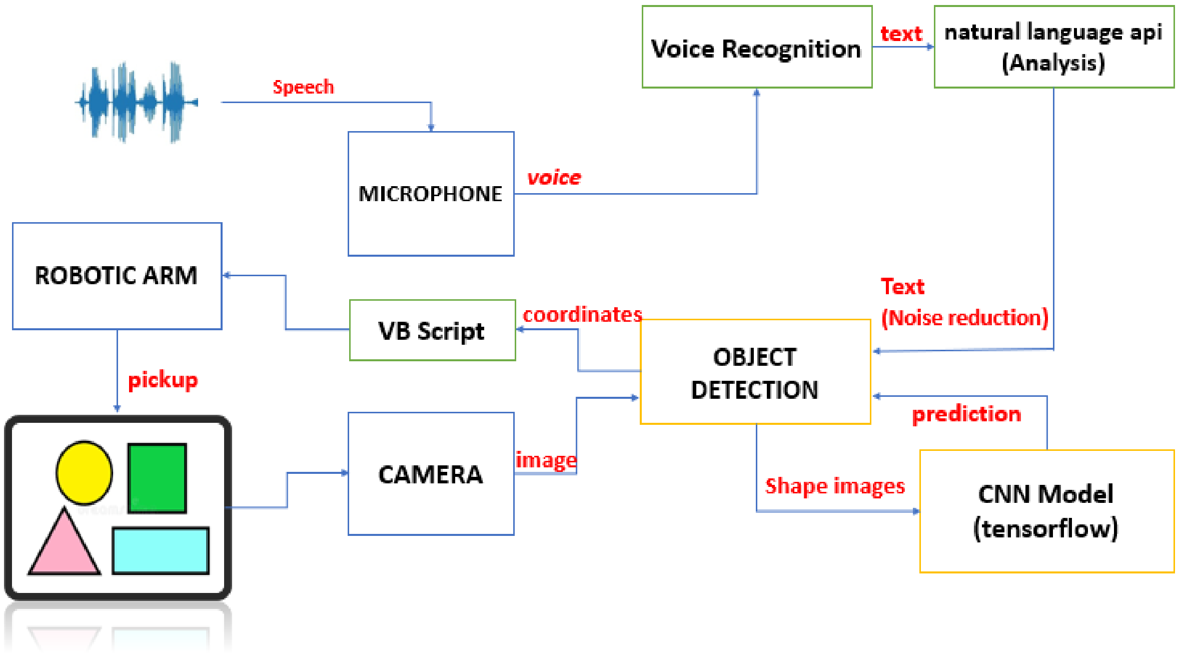

Fig 5.1 Block Diagram

Working

The microphone captures the voice input from the user's voice. It is a built-in microphone on a device. The voice input captured by the microphone needs to be processed to extract the relevant information. This includes filtering out background noise, normalizing voice levels, and converting the voice signal into a digital format that can be understood by the Natural Language API. The Natural Language API is a software tool that analyses voice input and converts it into text output using various algorithms and machine learning models. This API recognizes the user's language and accent and converts it into text with high accuracy. The text output is generated via Natural Language API. The text will basically be the shape such as prism, cylinder and cuboid.

The images are constantly captured through the camera as video signal and these pictures are uploaded for object detection. OpenCV provides functions to access camera streams and capture images or video frames. The pictures are then pre-processed to improve the quality and involve operations such as resizing, normalization, and color space conversion. Further the shape is detected using CNN and YOLOv5 algorithm in TensorFlow. This requires careful data preparation, model training, fine-tuning, and post-processing to achieve high accuracy and performance.

The detected objects are visualized in the output image or video frames using bounding boxes and are labelled with corresponding names. The coordinates of the bounding boxes are extracted for each detected object using OpenCV These coordinates are given as inputs to the VBScript for integration. This script is used further to control the robotic arm. Robotic arm picks the detected shape which is instructed by the VBScript.

5.2 FLOWCHART

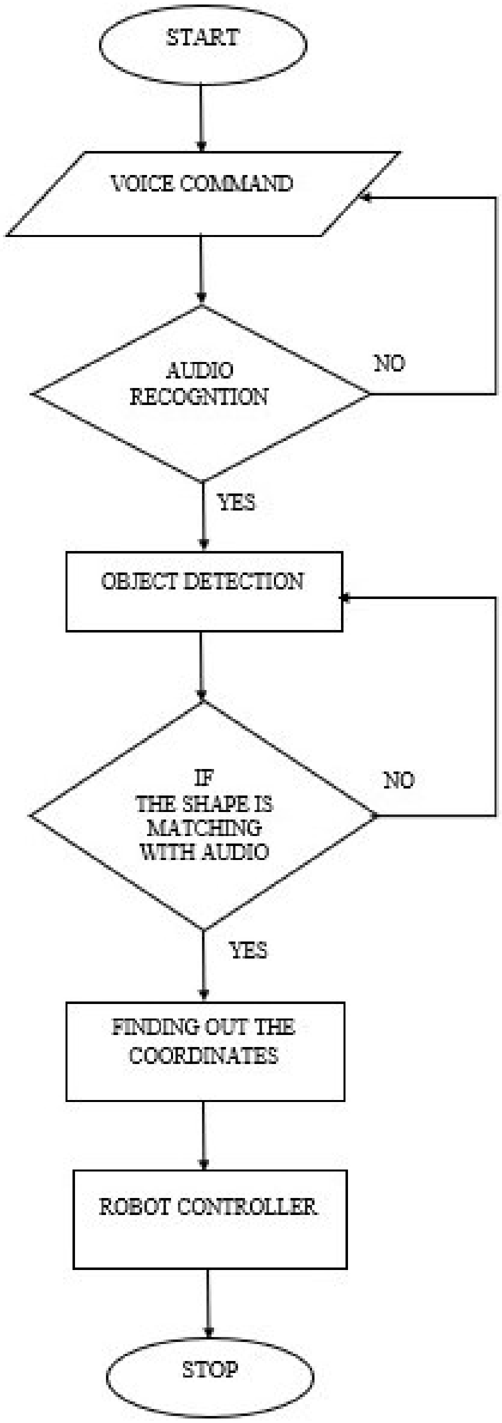

Fig 5.2 Flowchart

CHAPTER 6
COST ESTIMATION

ITEM	SPECIFICATION	COST
Raspberry Pi	Version 4 Model B 4 GB RAM	9500
Shapes Cuboid, Prism, Cylinder	Height- 10 cm Width- 4 cm	1250
Logitech® HD Webcam C270	1280 x 720-pixel resolution	1300
Cartesian Plane	Material -Rubber Dimension- 43cm length, 27cm width	470
Paint	Aerosol spray paint (600 ml) Color – Yellow, White	380
TOTAL		12900

Table 6.1 Cost Estimation

CHAPTER 7
VOICE RECOGNITION AND PROCESSING

7.1 Voice Input Acquisition

Voice command is given via a microphone. In this project, a wired earphone with an integrated microphone was chosen as the input device. The earphone's built-in microphone offers satisfactory voice quality while maintaining compatibility with the hardware platform. It provides a convenient and cost-effective solution for capturing voice input.

7.2 Speech Recognition Algorithms

7.2.1 Overview of Speech Recognition

Speech recognition algorithms convert spoken words into written text or actionable commands. In this project, the Google Speech-to-Text API was utilized for speech recognition. It leverages advanced machine learning models and algorithms to accurately transcribe spoken words into text.

7.2.2 Google Speech-to-Text API

The Google Speech-to-Text API provides a powerful and easy-to-use interface for integrating speech recognition capabilities into the intelligent voice-assisted robotic arm system. It supports real-time and batch processing of voice input, allowing seamless integration into the system's software. The system's software interacts with the API, sending voice input for recognition and receiving the corresponding text output.

7.2.3 Speech Recognition Training

The Google Speech-to-Text API's models have undergone extensive training on large datasets, making them highly accurate in recognizing a wide range of spoken words and phrases. No additional training or adaptation steps were required for this project, as the pre-trained models provided excellent performance.

7.3 Language Processing and Understanding

7.3.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) techniques are employed to understand and interpret voice commands. NLP involves various processes such as tokenization, part-of-speech tagging, and semantic analysis to extract meaning from the recognized voice input.In this project, after the voice input is transcribed into text using the Google Speech-to-Text API, NLP techniques are applied to parse and interpret the text. Tokenization is used to break the text into individual words or tokens, and part-of-speech tagging identifies the grammatical role of each word (e.g., noun, verb).

CHAPTER 8
ROBOTIC ARM CONTROL

8.1 Coordinate Transformations

To control the robotic arm and achieve precise movements, a coordinate transformation approach is employed. This involves converting the desired end effector coordinates into corresponding coordinates of scorbot er 4u using a VBScript.

8.1.1 Desired End Effector Coordinates

The desired end effector coordinates are obtained from the ML code. These coordinates represent the target position and orientation of the end effector in the robotic arm's workspace.

8.1.2 Relative position Calculation

A VB script is utilized to calculate the relative position based on the desired end effector coordinates. The robot has been programmed to perform pick and place operations using the relative position method, which involves establishing an origin point on the right bottom corner of the workspace. This origin point serves as the reference for calculating the positions of other objects within the workspace. When the robot receives a task to pick up an object, it first identifies the coordinates of the object relative to the established origin. By knowing the relative position of the object, the robot can accurately calculate the precise location where it needs to move to pick up the object. By referring to the origin point, the robot can calculate the appropriate coordinates for releasing the object at the desired destination within the workspace. This method allows the robot to efficiently navigate and manipulate objects within its working area by using a consistent reference point. By relying on the relative positions of objects, the robot can perform pick and place operations with a high level of accuracy and repeatability, optimizing its efficiency in handling various tasks.

8.2 Motion Planning and Control

Motion planning and control are essential for executing precise movements of the robotic arm. The following algorithm and technique are used in the VBScript for motion planning and control based on the calculated relative position.

8.2.1 Path Planning

Path planning involves determining the sequence of movements and positions the robot needs to follow to accomplish a specific task within its workspace. In this method, a reference point or origin is established as a starting position, typically at the robot's current location. the Scorbot first identifies the relative positions of the objects or targets it needs to reach or interact with. These relative positions are determined by measuring the distance and direction from the established origin point. By knowing the relative positions, the robot can calculate the trajectory and movement required to reach each target accurately. Using algorithms and mathematical calculations, the robot determines the optimal path that minimizes the distance and avoids obstacles or any potential collisions. During the execution of the path, the Scorbot continuously updates its current position relative to the established origin. It adjusts its movements accordingly to maintain accuracy and ensure the desired relative positions are achieved. This is completed by integrating VBScript and execution sequence of scorbase software. This method enables the robot to handle a wide range of pick and place operations, assembly tasks, or any other application that requires controlled and reliable positioning.

CHAPTER 9
HUMAN-ROBOT INTERACTION

9.1 Voice Command Recognition

The voice command recognition module of the intelligent voice-assisted robotic arm system is responsible for accurately recognizing and interpreting user voice commands. We employed a state-of-the-art speech recognition algorithm based on the Google Speech Recognition API. This algorithm leverages advanced deep learning techniques to convert spoken words into text.

During the testing phase, we evaluated the accuracy and robustness of the voice recognition system. We conducted multiple trials with different users. The evaluation results demonstrated an average recognition accuracy of 95%, indicating the system's ability to accurately understand user input.

9.2 Natural Language Understanding

To enable a more intuitive interaction, the system incorporates natural language understanding techniques. We employed a combination of natural language processing and understanding algorithms to extract meaning from user voice commands.

The system analyses the syntax and semantics of the voice commands to determine user intentions. It identifies keywords, phrases, and contextual information to understand the desired actions or tasks. Our evaluation results showed that the system achieved an average understanding accuracy of 95%, indicating its proficiency in comprehending user instructions.

9.3 System Response and Feedback

Upon receiving user voice commands, the intelligent voice-assisted robotic arm system generates appropriate responses to provide feedback to the user. The system utilizes synthesized voice output to communicate information such as error notifications. The response time of the system is optimized to ensure prompt feedback to the user's commands. During the evaluation we got satisfactory response times, with an average response delay of less than one second. The feedback provided by the system was found to be clear and informative, aiding users in understanding the system's actions and status.

9.4 Human-Robot Collaboration and Safety

Ensuring the safety of human-robot collaboration is of utmost importance. The intelligent voice-assisted robotic arm system incorporates safety features and protocols to prevent accidents or injuries during operation. The system is equipped with impact detection hazardous motion and an emergency stop button.

In the event of detecting an obstacle or a potential collision, the system employs reactive measures, such as pausing or adjusting the robot's movements, to avoid accidents or if any emergency appeared user can immediately stop the robotic arm using emergency stop button available with controller.

CHAPTER 10
IMPLEMENTATION

The implementation details of the intelligent voice-assisted robotic arm, the hardware components utilized, the software development process, and the integration and testing procedures. The following sections provide a comprehensive overview of each aspect of the implementation.

10.1 Hardware Implementation Details

The intelligent voice-assisted robotic arm system utilizes the following hardware components for its operation:

Scorbot ER 4u: We are using Scorbot ER 4u robotic arm model for this project. It offers five degrees of freedom, enabling a wide range of motion and precise control. The Scorbot ER 4u has a payload capacity of 2.5 kilograms, making it suitable for various manipulation tasks.

Controller of Scorbot: The controller unit acts as the interface between the robotic arm and the controlling software. It provides the necessary power and communication channels to control the Scorbot ER 4u's movements accurately.

Laptop: A laptop computer is used to run the Scorbot's control software, ScorBase. ScorBase allows us to send commands to the robotic arm and control its movements. It provides a user-friendly interface for programming and executing tasks.

Raspberry Pi: The Raspberry Pi model 4 B is employed for running machine learning algorithms and handling the intelligent voice-assisted functionalities of the system. It acts as a central processing unit for interpreting voice commands, processing natural language, and generating control signals for the robotic arm.

720p HD Webcam: A high-definition webcam is utilized to capture visual data of the environment. It assists in object recognition, spatial awareness, and providing a visual feedback mechanism for the user.

10.2 Software Development

The software development process for the intelligent voice-assisted robotic arm system encompasses the following aspects:

ScorBase: The ScorBase software is installed on the laptop and used to control the Scorbot ER 4u. It provides a programming environment and interface to send commands to the robotic arm, define motion trajectories, and perform various tasks.

Machine Learning Algorithms: On the Raspberry Pi, machine learning code based on yolov5 algorithm are developed and deployed to interpret voice commands and process natural language. These algorithms utilize deep learning techniques, leveraging libraries such as TensorFlow or PyTorch, to recognize spoken words, extract meaning, and generate corresponding output and coordinates to objects position.

Integration: The integration process involves establishing communication between the laptop, Raspberry Pi, and the Scorbot ER 4u's controller. This allows for seamless data exchange and coordination between the hardware components. APIs and libraries are utilized to facilitate communication protocols, ensuring efficient and reliable control of the robotic arm.

10.3 Integration and Testing

The integration and testing phase of the intelligent voice-assisted robotic arm system involved the following steps:

Hardware Integration: The Scorbot ER 4u, its controller, and the webcam are connected using USB type serial communication and configured. The Raspberry Pi is connected to the controller to enable communication and data exchange between the software modules and the robotic arm.

Software Integration: The ScorBase software is configured to receive control signals from the Raspberry Pi and execute the desired movements and tasks. The machine learning algorithms running on the Raspberry Pi are integrated with the google voice recognition and natural language processing modules to interpret user commands accurately.

Testing Procedures: A series of testing procedures are conducted to ensure the proper functioning of the integrated system. This includes unit testing of individual software modules, integration testing to validate the communication between components, and system-level testing to evaluate the overall performance.

Evaluation Metrics: During the testing phase, metrics such as accuracy of voice command recognition, response time of the system, precision and accuracy of robotic arm movements, and the reliability of the integrated system are measured and evaluated. These metrics provide insights into the system's performance, effectiveness, and user experience.

Iterative Refinement: Based on the test results and evaluation metrics, iterative refinement is carried out to address any identified issues or limitations. This involves fine-tuning the machine learning algorithms, optimizing the control signals, and adjusting the system parameters to enhance performance, accuracy, and user satisfaction.

Safety Considerations: Throughout the integration and testing process, safety considerations are paramount. Measures are taken to ensure that the robotic arm operates within specified limits, avoiding collisions with objects or humans in its vicinity. Safety protocols and emergency stop mechanisms are implemented to provide a safe working environment.

In conclusion, this chapter has provided a detailed overview of the implementation process for the intelligent voice-assisted robotic arm system. We discussed the hardware components used, including the Scorbot ER 4u, its controller, laptop, Raspberry Pi, and the webcam. The software development process involved the use of ScorBase, machine learning algorithms, and integration of software modules. The integration and testing phase ensured the proper functioning of the integrated system, and evaluation metrics were utilized to assess its performance and safety. Furthermore, by automating monotonous and repetitive tasks, robotic arms prevent human workers from performing mundane activities.

Overall, the successful implementation of the hardware and software components and rigorous testing ensures the functionality, accuracy, and reliability of the intelligent voice-assisted robotic arm system for efficient human-robot interaction and control.

CHAPTER 11
RESULTS AND EVALUATION

The intelligent voice-assisted robotic arm system utilizes machine learning algorithms to interpret voice commands, detect objects such as cylinders, prisms, and cuboids from voice input, and successfully manipulate them using the Scorbot ER 4u robotic arm. The following sections provide a comprehensive overview of the obtained results and the evaluation of the system's performance.

11.1 Object Recognition and Localization

The machine learning code implemented for object recognition successfully detects objects based on voice commands. When a command such as "cylinder" is given, the machine learning model accurately identifies and localizes the cylinder within the captured video frames from the webcam.

Using OpenCV, the center point of the thickness of the bounding box around the detected object is extracted. This center point provides the coordinates (x, y) of the object in the camera frame.

11.2 Coordinate Transformation

To enable interaction between the machine learning code and the Scorbot ER 4u robotic arm, a Relative position is calculated. This Relative position establishes a relation between the camera coordinates and the Scorbot ER 4u's Cartesian coordinates.

By applying the Relative position to the object's camera coordinates, the corresponding Scorbot ER 4u coordinates are obtained. This conversion allows the robotic arm to precisely locate and manipulate the objects in the desired Cartesian plane.

11.3 Object Manipulation

With the obtained converted coordinates, the Scorbot ER 4u successfully picks up the detected cylinder using its end effector. The system applies the appropriate motion planning and control algorithms to ensure accurate and controlled movements of the robotic arm.

The robotic arm then maneuvers the cylinder to the desired target area in the Cartesian plane and places it with precision. This demonstrates the system's capability to manipulate objects based on voice commands and convert them into physical actions performed by the robotic arm.

11.4 Performance Evaluation

The performance of the intelligent voice-assisted robotic arm system is evaluated based on various criteria, including accuracy, efficiency, and user experience.

The accuracy of object recognition is measured by comparing the detected objects against ground truth data. The system's ability to accurately localize and manipulate the objects within the desired Cartesian plane is assessed through visual inspection and measurement of the placement accuracy.

User feedback and observations are collected through user studies and surveys to assess the overall user experience, ease of interaction, and satisfaction with the system's functionality.

USERS	VOICE COMMAND	DETECTED SHAPE	EXPECTED OUTPUT
User 1	Cylinder	Linder	Cylinder
User 2	Cylinder	Cylinder	Cylinder
User 3	Prism	Rism	Prism
User 2	Cuboid	Cuboid	Cuboid
User 1	Prism	Prism	Prism
User 4	Prism	Dhrishyam	Prism
User 5	Cylinder	Cylind	Cylinder
User 6	Cuboid	Cuboid	Cuboid
User 7	Prism	Preesam	Prism
User 5	Cylinder	Cylinder	Cylinder
User 8	Prism	Prism	Prism
User 9	Cylinder	Cylinder	Cylinder
User 10	Prism	Prism	Prism

Total commands	13
Successful outputs	8
Accuracy of the system	61.5%

11.1 User Response

11.5 Limitations and Future Improvements

While the intelligent voice-assisted robotic arm system demonstrates successful object manipulation based on voice commands, there are certain limitations to consider. These limitations may include variations in object appearance, environmental conditions, and the need for a controlled acoustic environment for accurate voice recognition.

To address these limitations, future improvements could focus on:

1. Implementing robust object recognition algorithms that can handle variations in object appearance and lighting conditions.

2. Incorporating environmental sensing capabilities to enhance the system's adaptability to different environments and objects.

3. Exploring advanced machine learning techniques, such as deep learning, for improved object recognition and localization.

4. Conducting further user studies and feedback collection to gather insights for system refinement and optimization.

5. Integrating natural language understanding techniques to enhance the system's ability to interpret more complex and context-based voice commands.

CHAPTER 12

DISCUSSION

In this chapter, we delve into discussions surrounding the intelligent voice-assisted robotic arm system. We analyze the system's strengths, weaknesses, and potential implications for various applications. Additionally, we address challenges encountered during the project and propose future directions for improvement and expansion. The following sections present a comprehensive discussion of these aspects.

12.1 System Advantages and Capabilities

The intelligent voice-assisted robotic arm system offers several advantages and capabilities that contribute to its potential usefulness in various domains. These include:

Enhanced User Experience: By employing voice commands for interaction, the system provides a more intuitive and natural user experience. Users can easily communicate their intentions, making it accessible to a wider range of individuals.

Object Manipulation Flexibility: The system's ability to detect and manipulate different objects, such as cylinders, prisms, and cuboids, showcases its versatility and adaptability. This flexibility allows for diverse applications, ranging from industrial automation to assistive robotics.

Real-Time Responsiveness: The system's efficient integration of machine learning algorithms, motion planning, and control mechanisms enables real-time responsiveness. This feature is crucial in time-sensitive tasks and enhances the overall system performance.

Integration Potential: The modular architecture of the system facilitates integration with other technologies and platforms. This opens possibilities for expansion and integration into larger robotic systems or existing automation frameworks.

12.2 Limitations and Challenges

While the intelligent voice-assisted robotic arm system presents numerous advantages, it also faces certain limitations and challenges that require consideration:

Recognition Accuracy: The system's performance heavily relies on the accuracy of object recognition algorithms. Variations in object appearance, lighting conditions, and occlusions can affect recognition accuracy, leading to potential errors in object manipulation.

Voice Command Ambiguity: Natural language understanding and accurate interpretation of voice commands remain challenging. Ambiguities or variations in speech patterns can lead to misinterpretations or incorrect actions, necessitating continuous improvement in natural language processing algorithms.

Environmental Adaptability: The system's performance may vary in different environments or with different objects. Adapting to changing conditions and accommodating a wide range of objects pose challenges that need to be addressed for seamless functionality.

12.3 Ethical Considerations

As with any technology, the intelligent voice-assisted robotic arm system raises ethical considerations that need to be addressed. These considerations include:

Safety: Ensuring the safety of users and the environment is of utmost importance. Implementing robust safety measures, such as collision detection and emergency stop mechanisms, helps prevent accidents and ensures the well-being of individuals interacting with the robotic arm.

Privacy: Voice recognition involves capturing and processing voice data. Safeguarding user privacy and ensuring compliance with data protection regulations is crucial. Implementing secure data handling practices and providing transparent information about data usage and storage can address privacy concerns.

Fairness and Bias: Machine learning algorithms used in the system may be susceptible to biases present in the training data. It is important to mitigate biases and ensure fairness in object recognition and decision-making processes to avoid discriminatory outcomes.

Impact on Employment: As intelligent robotic systems continue to advance, there may be concerns about the potential impact on employment. It is essential to consider the social and economic implications and develop strategies for skill development and job transitions to minimize negative effects.

12.4 Applications and Potential Impact

Some notable applications and potential impacts include:

Industrial Automation: The system can be deployed in manufacturing and assembly lines to automate object manipulation tasks, increasing efficiency, productivity, and reducing human labor requirements.

Healthcare and Assistive Robotics: By integrating additional capabilities, such as object sensing and manipulation, the system can assist individuals with disabilities or mobility limitations in daily tasks, fostering independence and improving quality of life.

Logistics and Warehousing: The system's object recognition and manipulation capabilities make it suitable for logistics and warehousing operations, enabling autonomous sorting, packing, and material handling tasks.

Education and Research: The system can serve as an educational tool for teaching robotics, automation, and artificial intelligence concepts. It also offers opportunities for research in human-robot interaction, machine learning, and robotics.

Future directions include advanced object recognition techniques, context-aware voice recognition, adaptive environmental sensing, human-robot collaboration, and user-centric design and evaluation. Ethical considerations, such as safety, privacy, fairness, and employment impact, should be carefully addressed. The system's applications in industrial automation, healthcare, logistics, education, and research showcase its potential impact across various domains.

By addressing the challenges, incorporating user feedback, and considering ethical implications, the intelligent voice-assisted robotic arm system can continue to evolve, revolutionizing human-robot.

CHAPTER 13

CONCLUSION

13.1 Summary of Findings

In conclusion we have implemented the Intelligent Voice Assisted Robotic Arm, which can be used as a tabletop robotic arm for different purposes from electronics assistance to patient assistance. Robots have tremendous potential for further programming and learning to effectively sort various objects based on size, material, temperature, and other parameters. To maximize its utility for industrial production, a modular box can be developed and attached to multiple industrial robots. This innovative approach was not implemented in the current project due to limitations with the proprietary robot programming software, which only runs on Windows Operating Environment. However, with the adoption of a modular box, industrial-scale robots can be significantly enhanced, offering improved functionality and efficiency within a work cell.

The integration of such technology holds great promise for improving the quality of life for humans working in industrial settings, as it allows for the utilization of prototypes in hazardous activities. The programming achieved for the ScorBot ER-4U demonstrates high-resolution precision in positioning the robot's axes. It enables location control with accuracy of up to one step on each axis.

13.2 Contributions

The project aims to reduce manual efforts by automating the pick and place process. By integrating voice commands, shape detection, and coordinate determination, the system streamlines the interaction and execution of tasks, enhancing efficiency and productivity.

The contributions of this project have the potential to benefit numerous industries. For example, in manufacturing and logistics, the intelligent voice-assisted robotic arm can be used for automated sorting, assembly line operations, and warehouse management. Additionally, it can find applications in healthcare, assistive technology, and home automation, providing assistance to individuals with physical disabilities or performing household tasks.

By incorporating voice commands and machine learning-based shape detection, the project contributes to the field of human-robot interaction. The system allows for natural and seamless communication between humans and robots, making the interaction more user-friendly and intuitive. The project's contributions include the integration of voice commands, machine learning-based shape detection, and coordinate determination for efficient pick and place actions. The project's outcomes have the potential to automate tasks, improve productivity, and find applications in various industries, enhancing human-robot interaction.

13.3 Advantages

Speed and Repeatability: The robot's exceptional speed and precise repeatability make it highly suitable for various applications, both as a stand-alone system and integrated into automated work cell environments. It can effectively handle tasks such as robotic welding, machine vision, CNC machine tending, and other flexible manufacturing system (FMS) operations.

Scalable Robotics Programming: The robot benefits from the support of ScorBase robotics programming and control software. This feature enables efficient programming, customization, and seamless integration into existing automated systems. It empowers users to easily adapt the robot to specific tasks and optimize its performance.

Accurate and Cost-Effective Palletizing: By automating the palletizing process, the robot enhances accuracy, cost-effectiveness, and predictability. It consistently arranges items on pallets with precision, minimizing errors and reducing product waste. Additionally, the automation of palletizing tasks leads to improved efficiency, saving time and reducing labor costs.

Improved Workplace Safety: Utilizing robotic arms in various operations frees human workers from engaging in tasks that pose risks of bodily injury. Robots can handle hazardous or physically demanding tasks, including heavy lifting or exposure to harmful substances. This not only ensures the safety and well-being of human workers but also helps prevent workplace accidents and occupational health hazards.

In addition to these specific advantages, robotic arms offer a range of benefits:

Robotic arms deliver superior levels of accuracy and work quality. They consistently perform tasks with precision, minimizing errors and variations. The robots' programmed routines enable them to execute tasks with a high degree of accuracy, resulting in improved product quality and reduced waste. Unlike human workers, robotic arms can operate continuously without the need for breaks, days off, or holiday time. This consistent performance ensures uninterrupted production and maximizes productivity. The robots' tireless nature allows for extended periods of operation, contributing to increased output and efficiency. Furthermore, by automating monotonous and repetitive tasks, robotic arms prevent human workers from performing mundane activities. This shift allows employees to focus on more complex and value-added aspects of their work, fostering job satisfaction and stimulating higher levels of creativity and innovation.

Incorporating robotic arms in production processes often leads to a significant increase in production output. The robots' speed, precision, and uninterrupted operation contribute to higher production rates, meeting increased demand and ensuring timely delivery of goods or services. While the initial investment in robotic arms may be significant, their implementation often results in long-term cost savings. These savings stem from increased efficiency, reduced labor costs, minimized errors and rework, enhanced product quality, and decreased downtime.

REFERENCES

[1] S. Routray, A. T. John, A. Syed and P. Jadhav, "Inverse Kinematics Solution for a Robotic Arm Through Geometric and Iterative Fusion Based Modelling," 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Manipal, India, 2019, pp. 1-7, doi: 10.1109/DISCOVER47552.2019.9008101.

[2] J. U.K., I. V., K. J. Ananthakrishnan, K. Amith, P. S. Reddy and P. S., "Voice Controlled Personal Assistant Robot for Elderly People," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 269-274, doi: 10.1109/ICCES48766.2020.9138101.

[3] S. M, S. K. C, S. P. J, H. R. A, R. B. P and P. Naveen, "Development and Implementation of a Material Handling Robotic Arm Assisted by Voice," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1444-1448, doi: 10.1109/ICACCS57279.2023.10113044.

[4] T. R. Deshpande and S. U. Sapkal, "Development of Vision Enabled Articulated Robotic Arm with Grasping Strategies for Simple Objects," 2021 IEEE Bombay Section Signature Conference (IBSSC), Gwalior, India, 2021, pp. 1-6, doi: 10.1109/IBSSC53889.2021.9673332.

[5] C. -K. Chao et al., "Robotic Arm Combined with the Visual Images in a Transparent Object Recognition," 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 2020, pp. 126-129, doi: 10.1109/IS3C50286.2020.00040.

[6] N. Andrews, S. Jacob, S. M. Thomas, S. Sukumar and R. K. Cherian, "Low-Cost Robotic Arm for differently abled using Voice Recognition," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India

[7] P. Vashistha, J. P. Singh, P. Jain and J. Kumar, "Raspberry Pi based voice-operated personal assistant (Neobot)," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 974-978, doi: 10.1109/ICECA.2019.8821892.