Innovative Software and Hardware Design for a User-Friendly Tour Guide Robot

Prof. Dr. M.M.Bhoomkar
Mechanical engineering(HOD)
PVG COET
Pune, India

Prof. Shubhangi Deshpande
Information Technology
PVG COET
Pune, India

Rijvi Kulkarni
Information Technology
PVG COET
Pune, India

Aanchal Sarda
Information Technology
PVG COET
Pune, India

Rohan Chitale
Information Technology
PVG COET
Pune, India

Sudhanshu Sali
Information Technology
PVG COET
Pune, India

Kshitij Ghanekar
Mechanical
PVG COET
Pune, India

Atharva Potnis
Mechanical
PVG COET
Pune, India

Abstract - Museum curators are constantly exploring new ways to integrate technology into their displays. They may use interactive touchscreens or virtual reality technology to create a more immersive experience however, the cost of this technology is very high. We are proposing an autonomous, interactive tour-guide robot 'Theano-The Chaperone', who will offer a distributed, modular software and hardware architecture that combines modules for user interface, interaction, localization and collision avoidance. The software strategy depends on artifact detection, voice-enabled information delivery, voice assistant capable of answering user questions, reinforcement learning and path planning algorithms to give a complete museum tour experience. Our strategy involves the utilization of an indoor tour-guide robot that operates on an embedded system, the Raspberry Pi 4.

Index Terms - Image Dataset, You Only Look Once, Voice assistance, Text To Speech, Mechanical Design, Navigation, Proportional-Integral-Derivative, Kinematics

I. INTRODUCTION

Museums are institutions that collect, preserve and display objects and artifacts of cultural, historical, artistic or scientific significance for public viewing and education. Wonders can be found in museums. The designers of museums continuously strive to enhance the visitor experience. The relationship between humans and robots has been a significant area of interest for researchers and engineers. A specific area of development in this field is the tour guide robot. Our approach focuses on designing an autonomous indoor tour guide robot capable of assisting visitors by giving them a tour of the Museum. This approach will provide visitors with information and assistance in identifying the artifacts, it will also be able to respond to a few fundamental inquiries regarding the artifact. The design of a robot is mainly focused on Reinforcement Learning, which dynamically adapts its path depending on the presence of obstacles. In addition to aiming for low costs, this robot will be designed to be as effective and dependable as possible. With the help of this system, visitors will be able to gain information without having to read the lengthy notes, visually impaired can benefit through the voice enabled information and the need for human guides will be reduced. The rest of the paper is organized in the following way- In Section 2, we review previous papers related to the domain. Section 3 describes software modules, hardware architecture, and navigation, which are listed in Section 4. Advantages, future scope, and conclusions are mentioned in the last section.

II. LITERATURE SURVEY

A successful tour guide robot is judged based on how well it localizes itself around a certain place and how well it interacts with the humans. [1] Several types of tour guide robots have been introduced in the past, each with a unique navigation technique. Researches [2] Yelamarthi et al proposed a tour. guide robot equipped with an RFID reader for localization and sonar and IR sensors for obstacle detection and avoidance. However, passive RFID readers tend to have a limited operating range which makes them less reliable as the robot has high chances of missing a tag. Furthermore, RFID readers are quite costly. Instead we propose a line sensor array, which will be used to detect the contrast between the robot's surrounding surface and the line to be followed.

[3]The system employs a speak-and-retreat interaction model that approaches visitors who display an interest and gives them the required information about the artifact. The distances between the person and the artifact, as well as the angles, are used to estimate the attention. Cameras are attached to the ceiling, and people's head and body orientations are recorded. RFID tags are also used to uniquely identify each person.

Our proposed system will make use of an intelligent voice bot that will not only deliver the information associated with the artifact but also answer the fundamental questions asked by the users using a chatbot. In order to identify individuals and other obstacles, our system will make use of object detection algorithms like Yolo, Fast-RCNN etc.

The integration of hardware and software components in tour-guide robots requires a multidisciplinary approach involving expertise in mechanical design, electronics, artificial intelligence, and human-computer interaction. The design should prioritize safety, reliability, and user-friendliness to provide an engaging and informative experience for visitors.

III. PROPOSED METHODOLOGY

A. *Software Specifications:*

[redacted]***Image Detection:*** Since the robot will detect the visitors and other background elements, Object Detection will be used which can combine localization and classification to find objects in an image or video. The algorithms produce thousands of candidate regions which might contain objects. Various feature extraction techniques can be employed to extract features from these regions, and a classifier can be trained on top of these features. Faster Region-based Convolutional Neural Network (Fast R-CNN) is an improvement over R-CNN both in terms of accuracy and speed. In R-CNN, various candidate proposals which are overlapping are fed into a CNN independently, which results in a lot of repeated computation. They use a CNN trained on the Image Classification task as a base network, by excluding the final Soft max layer. The feature maps are then passed to a Region Proposal Network(RPN) which generates a fixed number of candidate regions, the same as the selective search algorithm in R-CNN, but much faster. A Roi Pooling layer is introduced, which produces fixed-length features corresponding to objects in the region proposals. The output layer is made of 2 nodes:- a Soft max to generate class probabilities and an FC layer to generate bounding box coordinates**.**

You Only Look Once (YOLO) is a single-shot object detection algorithm which generates class probabilities and bounding box coordinates in a single pass, unlike the F-RCNN, which uses 2 passes. The entire architecture is made of convolutional layers where the first few are pre-trained on the ImageNet dataset and the last layer outputs a (m x m) prediction. The input image is divided into multiple grids each of dimension (n x n) and whichever grid contains the center of the object should ideally detect that object. Each grid generates a fixed number of bounding boxes, confidence scores which indicate how certain the model is that the bounding box contains the object and the class of object. During training, the bounding box which has the highest IOU with the ground truth is selected. If there are multiple objects in a single grid, the algorithm fails to detect all the objects, which is the major downside of the algorithm.

1.1 Training Phase: A well-thought data collection methodology is made which aims at capturing images from different angles, under different lighting conditions, and different occlusion events. The data is stored in a MongoDB database which handles unstructured data efficiently and it can scale horizontally if augmentation is used which will increase the number of images. In data preprocessing, Tf-records will be generated which stores data in binary format, taking much less space on disk. This format is now easily compatible with not just TensorFlow but also other frameworks like PyTorch. The dataset will be partitioned into training and testing sets with an 80:20 ratio in such a way that the ratio of number of images in the train set to that in the test set is the same for each class. The weights of state-of-the-art models mentioned above and their different versions are downloaded from their GitHub repositories. Training will take place on a machine with an I7 processor, 16 GB RAM, and RTX 3050 GPU with 4 GB RAM. Once all the models start training without any errors and system issues, hyper-parameter tuning will be conducted to select the best set of hyper-parameters. Tensor board-dev will be used to log different hyper-parameter tuning experiments, where for each experiment a tensor board is used to show classification loss and evaluation metrics.

1.2 Prediction phase: Once the image is processed, the model is initialized with the trained weights and the image is passed to get the predictions. The information associated with the object is retrieved and passed to the Voicebot module, which will actually play it.

2. Voice enabled Information: A voice-enabled system, also known as a text-to-speech (TTS) system, works by converting written text into spoken words. The system uses a combination of natural language processing (NLP) and speech synthesis technologies to generate an audio output that closely mimics the sound and intonation of human speech. The process begins with the system analyzing the text input to determine the appropriate pronunciation, tone, and pacing for each word and sentence. This involves breaking down the text into its constituent parts, such as syllables and phonemes, and using rules and algorithms to generate the appropriate sounds.Once the audio output has been generated, the system uses a speech synthesizer to convert the digital audio signal into sound waves that can be heard through speakers. The quality and clarity of the audio output can vary depending on the quality of the synthesizer and the complexity of the text being read. Our proposed system will consist of a camera that detects the exhibits and retrieves information about it from the database. To achieve this, the system will employ image recognition algorithms(YOLO) that identify the object in the exhibit. Once the object is identified, the system will search the database (MongoDB) for relevant information related to the exhibit.The retrieved information is in textual form and will be converted to speech using a text-to-speech engine. The system will use a speaker to read out the information to the visitor. The text-to-speech engine converts the retrieved textual information into audio output that the visitor can hear, while the speaker enables audio output. Overall, the system will provide a user-friendly and interactive experience for visitors at the museum.

3. VoiceBot*:* Additionally, the system will have voice- enabled,question-answer interactions for visitors to learn more about the exhibit.

Algorithm for VoiceBot:

1. User speaks to the robot.
2. The robot will use Voice Activity Detection (VAD) to detect the user's voice and distinguish it from background noise
3. The robot then uses Automatic Speech Recognition (ASR) technology to transcribe the user's speech into written text using the Speech-to-Text (STT) engine.

4. The natural language Understanding (NLU) engine analyzes the text to extract the user's intent and essential information from the text.
5. The voicebot selects an appropriate response based on the user's intent
6. The response is synthesized into speech using the Text-to-Speech (TTS) engine
7. The robot finally delivers the synthesized response to the user.

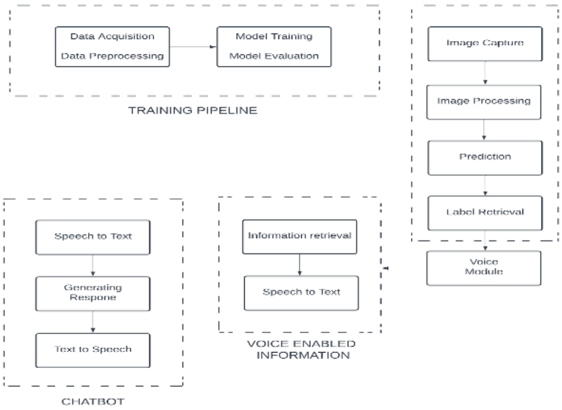

Fig. 1. (Software Block Diagram)

B. Hardware Specifications:

1. Planned Components: Our autonomous tour-guide robot's hardware components include a Raspberry PI 4B 8GB module as the central processing unit and a battery pack for power supply. It also has a 5MP camera module for capturing images, a microphone for recording audio input, and a proximity sensor for obstacle detection. Output modules include an 18 cm LCD touch screen display and speakers for audio output. The system utilizes Nema 17 stepper motors for movement, along with TB6600 stepper motor drivers for precise control. The selected hardware components ensure the smooth functioning and high-quality performance of the system.

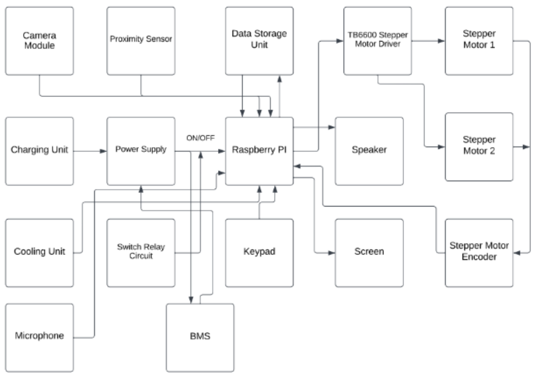

Fig. 2 (Hardware Block diagram)

2. Mechanical design of system:The autonomous tour-guide robot's mechanical design will be optimized for optimal functionality, aesthetics, and load-bearing capabilities. The proposed CAD model is highlighted in Fig.3.a, while structural simulations in Fig.3.b were performed to ensure the system operates safely and reliably under various boundary conditions.

Fig. 3.a (Rendered CAD model)

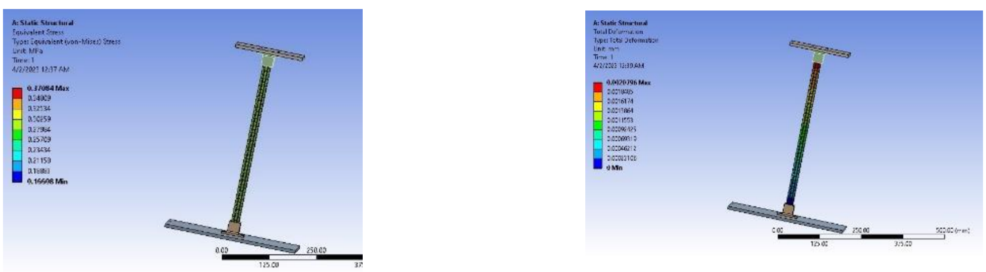

Fig. 3.b (Static Structural Analysis Results)

3. **Navigation**: A line following wheeled robot employs a sensor-based technique to achieve its desired motion. The robot will be equipped with a set of sensors, typically a line sensor array, which is used to detect the contrast between the robot's surrounding surface and the line to be followed.

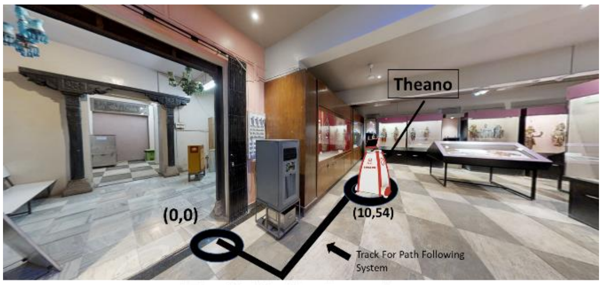

Fig.4 (Navigation visualization in actual environment)

The robot's control system will utilize the data from the sensors to make adjustments to the robot's movement in order to maintain its position relative to the line. The control system analyzes the output of the sensor array to determine the position of the line relative to the robot's current position.The line following technique will be implemented using proportional-integral-derivative (PID) control. This algorithm uses the sensor data to compute the necessary adjustments to the robot's motion, and it will be tuned to optimize the robot's performance based on the specific requirements of the application.

The equations governing the motion of such a robot can be derived from the **kinematic model** of the system. The kinematic model we used for the differential drive consists of the following equations:

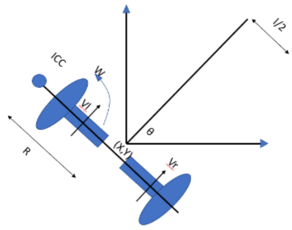

Fig.5 (Nomenclature of differential drive)[6]

Velocity of the left wheel: $V_L = \frac{r*(2v - wL)}{2L}$ Velocity of the right wheel: $V_R = \frac{r*(2v + wL)}{2L}$

Velocity of the robot's center of mass: $V = \frac{V_L + V_R}{2}$ Angular velocity of the robot: $w = \frac{V_R - V_L}{L}$

In these equations, V_L and V_R are the velocities of the left and right wheels, respectively, r is the radius of the wheels, L is the distance between the two wheels(also known as the wheelbase), V is the velocity of the robot's center of mass, and w is the angular velocity of the robot.[5]

There are changes in velocity of the two wheels of a differential drive robot for different motion scenarios:

CASE 1- While the robot is moving straight:
When the robot moves straight, both wheels need to rotate at the same speed to maintain the straight-line motion. Therefore, the velocities of the left and right wheels (V_L and V_R) are equal, and the angular velocity of the robot (w) is zero. This means that the robot is not turning, and its motion is purely translational.[4]

CASE 2- While the robot takes a left turn:
When the robot takes a left turn, the left wheel needs to rotate slower than the right wheel to create a differential velocity that turns the robot to the left. The magnitude of the velocity difference depends on the desired turn radius, which can be controlled by adjusting the angular velocity of the robot (w). The equation for the velocity of the left wheel (V_L) is given by

$$V_L = r*(2v - wL)2L,$$

where v is the desired linear velocity of the robot, r is the radius of the wheels, and L is the distance between the two wheels. To turn left, we must be positive, which means that V_L is smaller than V_R. The equation for the velocity of the right wheel (V_R) is

$$V_R = r*(2v + wL)2L$$

where v is the same as before, and the positive value of w causes V_R to be larger than V_L[4]

CASE 3- While the robot takes a right turn:
When the robot takes a right turn, the right wheel needs to rotate slower than the left wheel to create a differential velocity that turns the robot to the right. The magnitude of the velocity difference depends on the desired turn radius, which can be controlled by adjusting the angular velocity of the robot (w). The equation for the velocity of the left wheel (V_L) is given by

$$V_L = r*(2v - wL)2L$$

where v is the desired linear velocity of the robot, r is the radius of the wheels, and L is the distance between the two wheels. To turn right, we must be negative, which means that V_L is larger than V_R. The equation for the velocity of the right wheel (V_R) is

$$V_R = r*(2v + wL)2L,$$

where v is the same as before, and the negative value of w causes V_R to be smaller than V_L.

IV. Advantages of Proposed System

The user-friendly interface is easy to use and understand without having to consult a manual or receive external trainingThe proposed system is beneficial for the visually impaired as it allows them to experience the museum through auditory means, enabling them to fully enjoy the information presented without relying on reading.
Autonomous navigation reduces human dependency.

V. CONCLUSION

Our proposed system will utilize a line sensor array to accurately detect the contrast between the robot's surrounding surface and the desired line to be followed, providing a more reliable and cost-effective solution compared to RFID tags.Furthermore, object detection algorithms are used to identify individuals as well as obstacles.By utilizing an intelligent voice bot, visitors not only receive information about artifacts but also have their fundamental questions answered by the chatbot.

VI. Future Scope

With further improvements, Theano will soon be capable of comprehending and responding in numerous languages, thereby enabling individuals of all linguistic backgrounds to enjoy the museumImplementation of the 3D Laser mapping system instead of the proposed coordinate system for mapping will increase the overall mapping efficiency of the model.The chatbot will be fine-tuned with precise information about the exhibit to increase its accuracy.Theano will be standardized to make it applicable for all types of museums; the museum will only need to supply the dataset and choose the model; after that, the robot will automatically train itself and be prepared for use in that specific museum.

VII. References

[1] Diallo, Alpha & Gobee, Suresh & Durairajah, Vickneswari. (2015). Autonomous Tour Guide Robot using Embedded System Control. Procedia Computer Science. 76. 126-1339.

[2] Yelamarthi, K.; Sherbrook, S.; Beckwith, J.; Williams, M.; Lefief, R. (2012) An RFID based autonomous indoor tour guide robot. Circuits and Systems (MWSCAS), 2012 IEEE 55th International Midwest Symposium on , vol., no., pp.562,565.

[3] Iio, T., Satake, S., Kanda, T. et al. Human-Like Guide Robot that Proactively Explains Exhibits. Int J of Soc Robotics 12, 549–566 (2020)

[4] Ben-Ari, M., Mondada, F. (2018). Robotic Motion and Odometry. In: Elements of Robotics. Springer,Cham.

[5] Lu - Han Ma, Yong-Bo Zhong, Gong-Dong Wang, Nan Li, "The Kinematic and Dynamic Modeling and Numerical Calculation of Robots with Complex Mechanisms Based on Lie Group Theory", Mathematical Problems in Engineering, vol. 2021, Article ID 6014256, 34 pages, 2021.

[6] Dudek, G., & Jenkin, M. (2010). Computational Principles of Mobile Robotics (2nd ed.). Cambridge: Cambridge University Press.