ECE 445 Project Proposal

Project 30: Search and Identify

Name	Email
Shitian Yang	shitian.20@intl.zju.edu.cn
Yitao Cai	yitao.20@intl.zju.edu.cn
Ruidi Zhou	ruidi.20@intl.zju.edu.cn
Yilai Liang	yilai.20@intl.zju.edu.cn

Supervisor: Prof. Howard Yang & Gaoang Wang

ZJU-UIUC Institute
Zhejiang University, Haining, China
02/27/2024

Contents

1 Introduction

1.1 Problem

In contemporary households, managing daily tasks and locating various objects can become a time-consuming and sometimes frustrating endeavor. This challenge is mag-nified for individuals with mobility or visual impairments, for whom navigating through cluttered or unfamiliar environments to find objects can be particularly daunting. Fur-thermore, the increasing complexity of modern homes, filled with a myriad of gadgets and personal items, necessitates a smarter, more efficient approach to object localization and interaction. Despite advancements in smart home technologies, there remains a gap in intuitive, interactive solutions that bridge the human-robot communication divide, enabling seamless integration of artificial intelligence in daily life.

With the emergence of artificial intelligence, including ChatGPT [1], multimodal models [2–4], and innovative attention mechanism models [3, 5], we now have better and more extensive tools to tackle everyday tasks, while the corresponding application domains remain to be explored. This development offers promising avenues for enhancing the way we interact with our environment, making it more accessible and user-friendly, especially for those with disabilities. By leveraging these advanced technologies, we can bridge the gap in human-robot communication, facilitating a seamless integration of artificial intelligence into daily life, and opening up new possibilities for smart home solutions that are intuitive and effective.

1.2 Solution

Our final product is a voice-seeking robot for home use. The robot will read the user's regular-speed-and-tone voice commands and scan its surroundings for a specified item that best matches the description. It will then stop rotating with a laser pointer lit to shone on the desired object. Our solution consists of both software and hardware components:

- **Software Component:**

  1. **Speech Recognition Module:** Acquires the user's voice commands of common speed and tone in English, converts them into logical questions and feeds the recognized desired object name to the image recognition module.
  2. **Image Recognition Module:** Recognizes the captured environment images under different lighting conditions and matches the target objects by using innovative attention mechanism models.

- **Hardware Component:**

  1. **Image Acquisition Module:** Acquires images of the environment through camera with different positions and angles controlled by the steering motor, can also perform appropriate zooming and stitching and transmit the visual data to software for analysis.

2. **Control module:** Adopts digital circuit and logic chips, transforming the condition signals into digital signals in controlling the powering module. For example, this control module will change its output when the software component judges that it suddenly found the desired object.
3. **Drivetrain and Powering Module:** This module is the core of the hardware component, where the micro-controller STM32F103c8t6 chip is adopted in controlling and providing the steering engine with impulse signal. The steering engine will be continuously rotating uniformly, until input signal from the control module shifts. A 12V power source is adopted in charging the micro-controller.

1.3 Physical Design

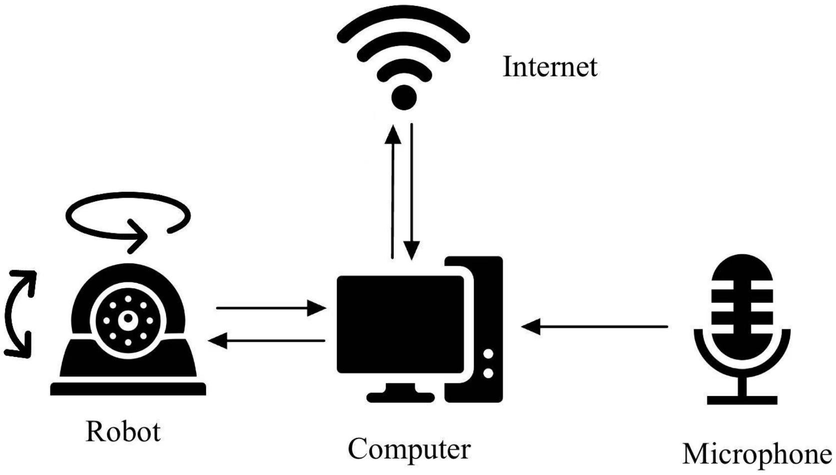

Figure 1: Physical Design System Diagram

1.4 High-level Requirements list

1. **Accurate functionality of steering motor system:** Our hardware system mainly centers around the steering system, it should possess the capability to ac-curately indicate the object based on a specific given coordinate. Once turned on and without target, the steering engine will be continuously rotating uniformly. The indicator will only come to a stop when target is found and must precisely display a direction within an angular deviation of no more than 7° from the target's horizontal and vertical rotation.

2. **Accuracy of Speech Recognition:** The voice recognition system must accurately activate when the user calls it, and correctly extract a clear and concise description of the desired object from the user's commands, achieving a minimum accuracy rate of 95%. This involves discerning the specifics of the desired item from the user's commands under various household noise conditions.

3. **Accuracy of Vision Module:** The vision module should identify and localize the target object within the camera's view with a minimum accuracy of 85%. This requires the generation of an approximate bounding box around the object and precise calculation of its center relative to the robot's position. Performance must be consistent across varying lighting conditions, object orientations, and backgrounds.

2 Design and Requirements

2.1 Block Diagram

The block diagram of our project is divided into components of software and hardware.

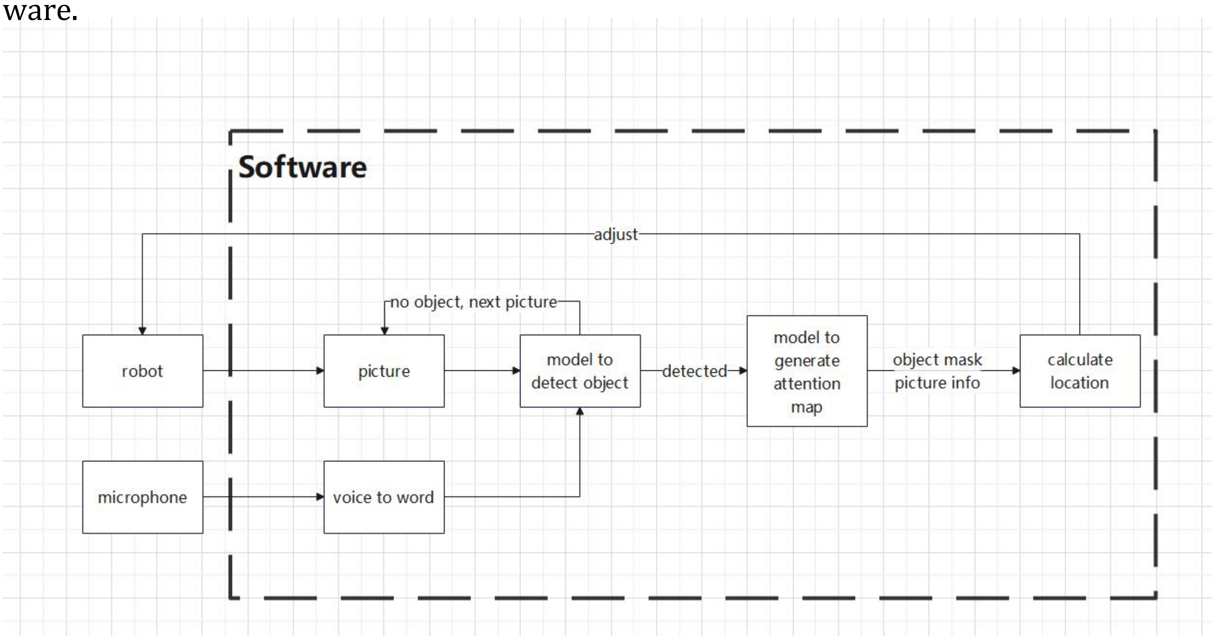

Figure 2: Block Diagram for the Software Component

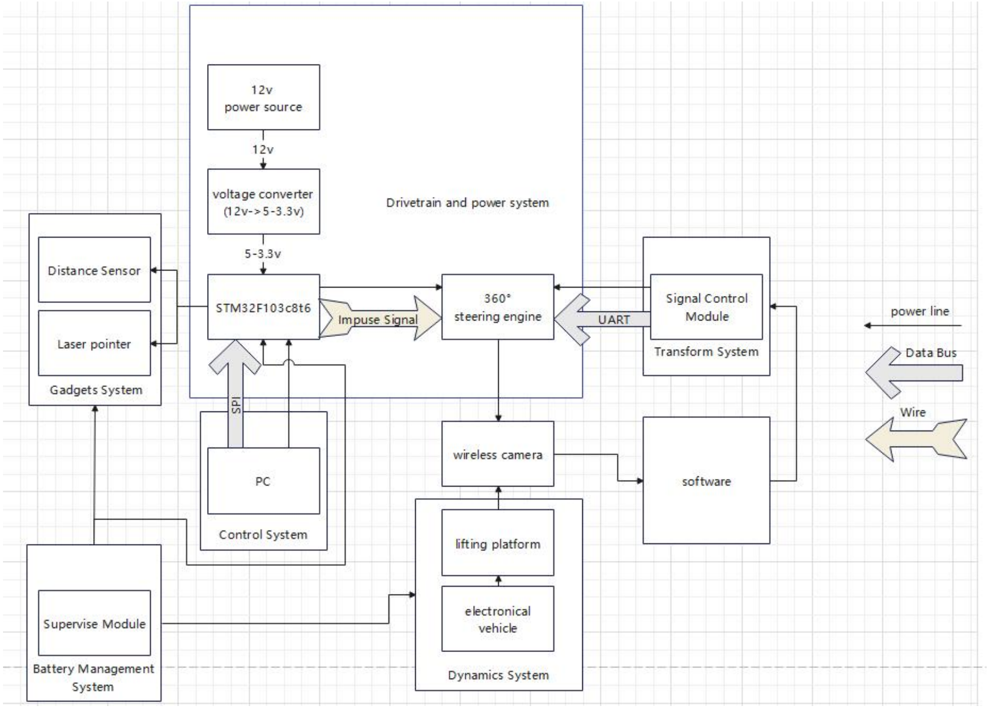

Figure 3: Block Diagram for the Hardware Component

2.2 Block Subsystem Functions & Requirements

2.2.1 Microphone Voice-to-Text Conversion System:

This mature model is responsible for converting audio captured by the microphone into corresponding text.

Requirements: The initial audio input will undergo noise reduction to eliminate background disturbances, ensuring the conversion output is precise, with an emphasis on critical adjectives and nouns.

2.2.2 Object Detection System:

This module detects the presence of relevant objects in images captured by the camera based on the text converted from audio.

Requirements: Object detection must be fast and accurate, ideally pinpointing the approximate location of objects and performing a secondary detection based on adjectives in the commands. It should quickly ascertain if such objects are present in the image, employing both broad and detailed screening to determine compliance with requirements and location. Multiple models may be used for cross-validation to prevent errors and omissions, if time permits .

2.2.3 Object Mask Generation System:

Utilizing images from the object detection module and corresponding text requirements and instance segmentation methods, this module generates masks for identified objects.

Requirements: These masks should possess precise contours rather than being mere indicative boxes.

2.2.4 Calibration and Computation System:

Based on the generated object masks, this module calculates the centroid of the object and determines its actual 3D angles in spherical terms using the image information.

Requirements: The module must accurately match the information from the masks, processed through various models, to prevent loss of positional information due to image distortion caused by limitations within the models.

2.2.5 Drivetrain and Power System:

1. **12V Power Source:** The 12v power supply provides a stable 12v voltage and 10A current, and inputs into the voltage converter, providing power for the entire drivetrain and power system.

   Requirements: The power source must supply stable current and voltage to ensure the normal operation of the system.

2. **Voltage Converter:** The voltage converter connects the power supply with the micro-controller, ensuring the voltage and current do not exceed the limits required by the micro-controller.

   Requirements: Convert 12v to 5-3.3v, the current is 3A for 5v and 1A for 3.3v.

3. **STM32F103c8t6 Micro-controller:** Through hardware programming, we en-sure the steering engine can rotate according to the set target. After rotating a specific angle, it stays for a predetermined amount of time, then automatically ro-tates the same specific angle again, until it completes a full rotation.

   Requirements: STM32F103c8t6 micro-controller’s nominal voltage is 3.3v.

4. **360° Steering Engine:** The steering engine can automatically rotate according to the set angle, then stay at the set angle, thus controlling the angle. The steering engine connects with the wireless camera to take the picture of the surrounding environment.

   Requirements: The steering engine can accept the 3.3v signal voltage or 5-8.5v power voltage.The steering engine can carry up to 30kg items.

2.2.6 Transform System:

Based on the results of the software part, determine whether to output a 0/1 signal (low level or high-level signal) to the steering engine. The steering engine will stop running after accepting 0 signal and keep running after accepting 1 signal.

Requirements: Signal Control Module should output the 3.3v signal voltage to the steering engine.

2.2.7 Control System

PC writes the program into the micro-controller and gives the angles we want.

Requirements: the steering engine can only rotate in 0-360°, so the input angles should be in the range 0-360°.

2.2.8 Dynamics System:

1. **Lifting Platform:** Raise the camera to allow the camera to cover a larger field of view, achieving three division rotation.

   Requirements: The power source in the lifting platform should supply the platform enough energy.

2. **Electronical Vehicle:** Carry the lifting platform and the wireless camera to search and identify the specific items in a broader area.

   Requirements: Battery attached to the vehicle should power the vehicle when needed.

2.2.9 Gadgets System:

1. **Laser pointer:** When detecting the goal items, laser pointer will be activated and point at the items.

   Requirements: Micro-controller should give the signals to the laser pointer after finding the goal items.

2. **Acoustic distance sensors:** When detecting the goal items, the distance sensor should cal-culate the distance between the goal items and the distance sensor so that the software can calculate the specific coordinates.

   Requirements: Micro-controller should give the signals to the distance sensor after finding the goal items.

2.2.10 Battery Manage System

This system consists of the supervising module which mainly handles two tasks. First of all, this system will keep track of State of Charge (SOC) of the batteries powering the camera, micro-controller and the dynamic cart, and at the same time, monitor their bat-tery health. The system is also capable alerting users of unexpected circumstances such as overheating or voltage spikes from the safety perspective.

Requirements: The supervising module should be able to show real-time data relate to the electricity for all of our electrical components. Warnings should be noticeable when physical values detected exceeds our set threshold.

2.3 Tolerance (Risk) Analysis

A big difficulty in this project is that the angle control program of the micro-controller cannot receive external data, thus the rotation angle of the servo motor cannot change according to the results of the software part. Through exact theoretical calcu-lations, we converted the results of the software part into a low- and high-level signal output through logic gate, to control whether the entire hardware part is running.

A simple logical equation is:

$$Y = (A \cdot B)' = (A') + (B')$$

A and B represents the results from software and hardware.

Also, different hardware items require different nominal voltage and current. If it exceeds the required range of voltage or current, there are safety hazards, and it will cause damage to the hardware. We ensure that the voltage and current passing through each piece of hardware meet the requirements through accurate theoretical calculations and voltage conversion. Furthermore, the connection of the line between different hardware items is prone to loosening, so the servo motor cannot run normally. We can ensure the stability of the connection of the route by welding.

In terms of software, the AP50 of YOLO-World-Small is 62.3, and AP75 is 49.9. Calculated with the maximum error for AP50, the center of the bounding box can deviate from the object's center by r (where r is the object's average radius). For some everyday objects, such as a cup (r=12cm) and glasses (r=6cm), the generated error is not as significant as mechanical error. Moreover, after processing with segment-anything, in a clean environment, the object contours are extracted. By comparing centroid weights, it's possible to pinpoint the mask error within r/2, which for general small objects is within 6cm, an acceptable range.

On the mechanical side, the minimum adjustable precision is 1 degree. With the camera's applicable range for small objects (detection up to 5m), the relative error is about 5*tan(1 degree)*sqrt(2)/2 = approximately 6.1cm. Thus, the cumulative error is around 12cm for small objects detected at 5m. However, after further visual adjustments, the error can be reduced to about 6cm, close to the average radius of general small objects, which is acceptable.

3 Ethics and Safety

There are indeed several concerns on safety and ethics with our project. First of all, a laser pointer is considered to be mounted to the camera for target directing. Though the laser pointer is only designated to be turned on when the desired object is found, it can pose serious risks to human health, including eye injuries and skin burns. To avoid inappropriate pointing, we will fix a baffle and protector to limit the laser pointer in certain angles with a low level, and always keep the power off the during testing to comply with relevant safety regulations and to minimize the risk of harm to users or bystanders [6]. While our design also adopts electric power sources and motor, special care will be paid on robust testing and validation procedures to ensure the reliability of the system and to prioritize user safety. This complies with the ACM Code of Ethics, Section 2.9, that "Design and Implement Systems That Are Robustly Secure [6]." As a project involving visual and vocal data utilization, it's crucial that such data is handled securely and with respect for user privacy. With the scope of the course ECE 445, our team members will mainly be the operators and users, and that we will take on responsibility not to use or spread others' data without formal and proper permission [7]. Other users of our project will have total autonomy over whether to or to what degrees would they like to engage with our robot.

Another concern rises in transparency and explainability. Even if users permit our usage of their vocal data, it's our unshirkable duty to provide clear explanations of its decision-making processes, especially regarding object recognition and task execution, to ensure users understand and trust the system's behavior [6].

Last but not least, to ensure the users' safety and convenience, we adopts a battery management system which will monitor the charge on all our electric components since they are working wireless. This design not only provides the users information about charging condition of our robot but can also warn ahead if something unexpected or unsafe is about to occur, which also complies with Section 2.9 of the ACM Code of Ethics [6].

All of our group members carefully affirm that we will strictly follow the IEEE and ACM Code of Ethics.

References

[1] OpenAI. Chatgpt can now see, hear, and speak. OpenAI Blog, 2024. Accessed: 2024-01-10.

[2] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In Proceedings of the IEEE International Conference on Computer Vision, 2015.

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers, 2020.

[4] Baifeng Shi, Trevor Darrell, and Xin Wang. Top-down visual attention from analysis by synthesis, 2023. Version 2, submitted to CVPR2023. Project page: [URL].

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[6] ACM Ethics. Acm code of ethics and professional conduct. ACM Ethics - the Official Site of the Association for Computing Machinery's Committee on Professional Ethics, jan 2022.

[7] IEEE Code of Ethics. https://www.ieee.org/about/corporate/governance/p7-8.html.