上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

THESIS OF BACHELOR

Project Title Advance Development on State-of-Art Robots

Name1 Sidong Wang	ID1 515370910170	Major	ECE
Name2 Jiazhen Ji	ID2 515370910173	Major	ECE
Name3 Tailun Liu	ID3 515370910169	Major	ME
Name4 Ziyou Wu	ID4 515370910134	Major	ME
Name5 Xinyu Lin	ID5 515370910153	Major	ME

Supervisor Chengbin Ma

School UM-SJTU Joint Institute

Semester 2018-2019-Summer

上海交通大学

毕业设计（论文）学术诚信声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：李家枝 韩俐俐 刘泰伦 吴坤樑 王思琛

日期：2019 年 8 月 5 日

上海交通大学
毕业设计（论文）版权使用授权书

本毕业设计（论文）作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业设计（论文）。

保密□，在___年解密后适用本授权书。

本论文属于

不保密☑。

（请在以上方框内打"√"）

作者签名：李宇梭 林靖涌 沈博伦 吴啸宇 王思桥　　指导教师签名：郭云龙 代签

日期：2019 年 8月 5日　　日期：2019 年 8月 5日

Abstract

A software solution with a robot for some specific scenarios in our daily life is better than a human solution. Our problem is to develop a software solution on an open source robot platform for a specific scenario decide by ourselves. We decide to develop a photography service robot that aims at taking group photos in meetings or exhibitions. Group photos can be taken by person now or can be taken by setting a timer. However, it is not a perfect solution since you cannot always have someone waiting behind the camera or you would not know if the photo is good until it is shot. Therefore, our project aims at solving this problem. According to the customer requirements and engineering specifications, four sub-systems should be developed according to the needs. They are: a face recognition system, a sound recognition system, a user interface and a movement control system. Our final design is to detect sound instructions using Xun Fei platform, detect face using OpenCV from pictures taken by a Logitech C920 camera, develop a user interface in Linux, use LiDar to detect objects around and control the car using Arduino. The total cost is 1388, which is within the budget.Several improvements can be made if we have more time. The motion system is currently controlled by Arduino. To make it achieve a more reliable performance, ROS can be used to integrate the encoder message, and give speed or position target input. Also, sound recognition now has a 1-2s delay depending on Internet speed. We may fix this out. Finally, the user interface and the robot itself can be further decorated. Our project is a photography service robot aiming ta taking a good group photo. We think our final will satisfy all the customer requirements and fits well in the scenario.

Keywords: Service robot, ROS, LiDAR, face detection, voice recognition

摘要

使用机器人的软件解决方案在应对一些生活场景时要好于使用人力资源的解决方案。我们的课题是开发一个给予开源机器人平台的软件解决方案来应对一些生活中的场景。我们决定设计一台拍摄集体照的服务型机器人。现在的集体照或可以用人来拍摄或可以定时。但他们要么需要一个人在摄像机前要么就无法知道照片的好坏。我们的项目旨在解决这个问题。根据客户需求，四个子系统应被设立，他们是：语音识别系统，人像识别系统，用户界面以及机器人运动控制。我们的最终设计是一使用讯飞平台进行语音识别，使用OpenCV进行人像识别罗技C920摄像头拍摄的画面，使用雷达来探测障碍物，使用Arduino来控制机器人运动，在Linux环境下开发用户界面。我们的全部花费在1388元。这在我们的预算里。如果时间足够我们可以在如下内容作出提升。运动控制现在由Arduino完成。我们可以使用带有编码器的马达来提升运动控制准确度。声音控制有1-2秒延迟，这也可以做出一些提升。最后，用户界面和机器人本身都可以被更好的装饰。我们的项目是一个旨在拍摄集体照的服务型机器人。我们认为我们的最后结果可以满足客户需求与工程细节。我们的项目可以很好地解决选定场景中的问题。

关键词： 服务型机器人，机器人操作系统，激光雷达，人脸识别，语音识别

Executive Summary

A software solution with a robot for some specific scenarios in our daily life is better than a human solution. Our problem is to develop a software solution on an open source robot platform for a specific scenario decide by ourselves. We decide to develop a photography service robot that aims at taking group photos in meetings or exhibitions. Group photos can be taken by person now or can be taken by setting a timer. However, it is not a perfect solution since you cannot always have someone waiting behind the camera or you would not know if the photo is good until it is shot. Therefore, our project aims at solving this problem.

Customer requirements are generated from the scenario, and engineering specifications are generated according to the customer requirements. They are 1) Semantic recognition accuracy$\geq$97%. 2) Sound recognition accuracy$\geq$98%. 3) LiDar accuracy$\geq$0.1m. 4) LiDar max detection distance$\geq$2m. 5) Frames collected per second$\geq$30. 6) CPU frequency$\geq$2.0GHz. 7) Camera resolution$\geq$1200w pixels. 8) Face position within center 1/3 of picture. 9) Battery capacity>5Ah.

According to the customer requirements and engineering specifications, four sub-systems should be developed according to the needs. They are: a face recognition system, a sound recognition system, a user interface and a movement control system. We use OpenCV to detect face since it is a cheap way with high efficiency and accuracy. A Logitech C920 camera is used since it fits other sub-systems and provides a high image quality. We use Xun Fei platform to do the sound detect, since it fits for both Chinese and English, cheap and accurate enough to satisfy our engineering specifications. We develop the user interface in Linux since it best fits our other systems and easy to operate on. We use a LiDar to detect object since it is easy to program and has a good enough accuracy. The movement control is done in Arduino since it is easy to program and fits well with other systems. Our final design is to detect sound instructions using Xun Fei platform, detect face using OpenCV from pictures taken by a Logitech C920 camera, develop a user interface in Linux, use LiDar to detect objects around and control the car using Arduino. The total cost is 1388, which is within the budget.

The test results are: CPU frequency=2.4GHz, speech recognition accuracy=98%, semantic recognition accuracy=98%, fps=30, object detection range=12m, distance detection precision =0.002m and center face is within 1/3 of photo.

Several improvements can be made if we have more time. The motion system is currently controlled by Arduino. To make it achieve a more reliable performance, ROS can be used to integrate the encoder message, and give speed or position target input. Also, sound recognition now has a 1-2s delay depending on Internet speed. We may fix this out. Finally, the user interface and the robot itself can be further decorated.

Our project is a photography service robot aiming ta taking a good group photo. We think our final will satisfy all the customer requirements and fits well in the scenario.

Keywords: Service robot, ROS, LiDAR, face detection, voice recognition

Contents

1 Introduction

1.1 Background

Some daily scenario such as tour guide or caring for the elderly lack a complete software solution under robot platforms. They are now done by human and has many problems. As human resource gets more and more expensive, human solution for some specific scenario is no longer available. Also, human solution for some scenario is not as perfect as a robot solution, since human will not be able to focus and work for a long time. Therefore, a robotic solution for such scenario is in need. We are required to develop robot applications on open source robot platforms such as NAO or Cruzr to meet the requirements from a specific scenario, and providing a complete traceable technical document. The scenario it faces can be decided by our own.

1.2 Significance

Imagine when you are having a meeting or a party with some important people. After the meeting you want to take a photo with them. The expected photo is like this: In order to have a good group

Figure 1: Expected group photo[1]

photo, you need to find someone to photo for you, but it is not always possible to find a photographer. Besides, you can photo with your own cell phone. However, if the number of people is too large, holding your phone and have a selfie is not a good option. You can also set a timer and let the camera photo itself, but you would not know if the photo is good until you back to see the photo. In these cases, what you need is a robot that can take a good group photo for you.

1.3 About our project

We decide to develop a robot aiming at taking group photos. Our project aims at solving the kind of problem described in the significance section. The service robot we develop should meet the requirements under the chosen scenario. It is based on an open source robot platform that is provided by our sponsor Wecan Startup Accelerator. A little adjustment to the hardware can be accepted. Besides, ROS (Robot Operating System) should be used in developing the software. The final outcome should be a robot that can take a group photo in a professional manner.

1.4 Customer needs

According to the scenario we choose, several requirements are generated. They are listed below.

1. The robot should be able to adjust itself to the group to be photographed to ensure that all people being photographed is in the figure and they are close to the center of the picture.
2. The robot should always let the customer see what the current figure looks like.
3. The robot should be able to be controlled by sound to move its position and adjust its camera.
4. The robot should avoid collision when moving.
5. The robot should be able to send you the photo taken

We think that those requirements will include all real life needs of customers under such scenario. They will be explained in detail below.

- The robot should be able to adjust itself to the group to be photographed to ensure that all people being photographed is in the figure and they are close to the center of the picture.

When you are taking a group photo using this service robot, you would like it to adjust the position by itself first. This will save your time on adjusting the position of the picture. Of course, the robot will ensure that all people is in the picture and they are close to the center of the picture. In this way, the picture out should be in a professional manner.

- The robot should always let the customer see what the current figure looks like.

One problem of setting a timer on a normal camera is that when you go back to your position in the picture, you would not know if the picture now is good or not. Maybe actually you are in a strange post. Only when the picture is taken and you back to see the picture, you would know if it is a good one. If it is not, then you have to do all the things again. Therefore, our robot would let you always see your current picture. As long as you think the photo is good enough, it will take the photo.

- The robot should be able to be controlled by sound to move its position and adjust its camera.

Auto-adjust is not always reliable. In case you think the robot position is not good, you can adjust it. This should be done when you are in position, so a sound control is used to control the movement of the service robot after it has adjusted itself.

- The robot should avoid collision when moving.

This requirement does not need much explanation. Since the robot can move, it should always avoid to collide with a person.

- The robot should be able to send you the photo taken

After the photo is taken, you would like it to be send to your phone. Therefore, the robot should have a means to communicate with your phone, either using WiFi or Bluetooth.

1.5 Background research

Service robot is not new. We have searched photography robots facing similar scenario as ours. A typical one is show in Figure 2

Figure 2: Eva service robot[2]

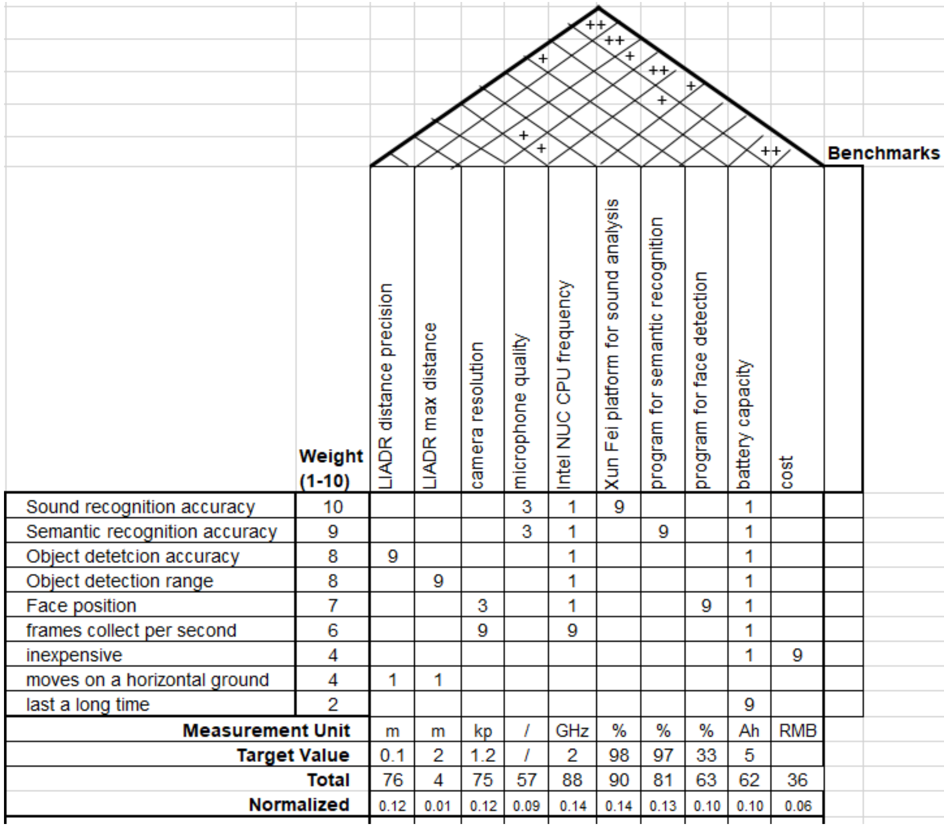
This is an Eva photography robot. It is aimed at taking photos for one or more people. A big screen is set on its front to let people see the current photo. It will navigate itself around a room and use face detection to catch people and ask if it can take a picture. After the picture is taken, it will be able to send the figure to your social media. The scenario is show in Figure 3

Figure 3: Eva service robot in use[2]

The good thing for this product is that it can navigate around a close area and find people to be photographed by itself. The problem, on the other hand, is obvious. It provides you with the current figure on its front screen, but it does not provide you an option to adjust its position on your instruction. You have to move yourself to make figure looks good. Our project should be able to do this as well.

2 Engineering Specifications

The engineering specifications are shown in the QFD chart in Figure 1.



	Weight (1-10)	LIADR distance precision	LIADR max distance	camera resolution	microphone quality	Intel NUC CPU frequency	Xun Fei platform for sound analysis	program for semantic recognition	program for face detection	battery capacity	cost	Benchmarks
Sound recognition accuracy	10				3	1	9			1		
Semantic recognition accuracy	9				3	1		9		1		
Object detetcion accuracy	8	9				1				1		
Object detection range	8		9			1				1		
Face position	7			3		1			9	1		
frames collect per second	6			9		9				1		
inexpensive	4									1	9	
moves on a horizontal ground	4	1	1									
last a long time	2									9		
Measurement Unit		m	m	kp	/	GHz	%	%	%	Ah	RMB	
Target Value		0.1	2	1.2	/	2	98	97	33	5		
Total		76	4	75	57	88	90	81	63	62	36	
Normalized		0.12	0.01	0.12	0.09	0.14	0.14	0.13	0.10	0.10	0.06	

Figure 4: QFD chart

The sponsor does not have any requirement on our project, as long as we use the open source platform he provides. We figured out the engineering specifications on our own according to the scenario we chose.

Since the users should be able to control the robot remotely with voice, we want the sound recognition accuracy and semantic recognition accuracy to be high. Therefore, the following to requirements have a high weight and high specification.

1. Semantic recognition accuracy $\geq$97 %
2. Sound recognition accuracy $\geq$98 %

Collision is very dangerous in real life usage. Therefore, collision avoidance is the next important thing we need to ensure, which is realized by a LiDAR. The accuracy and maximum range of the LiDAR is required as follows.

1. Accuracy $\geq$0.1
2. Maximum distance $\geq$2m

We want our customer see the current picture from the screen. When the robot moves, the customer should see the figure changes. We want the figures to be refreshed quickly. Therefore, the fps should be regulated. Also, the CPU of the robot should be powerful enough to support such kind of calculation. The specifications are given below.

1. CPU frequency $\geq$2.0GHz

2. Frames collected per second ≥30

We want the final picture taken to be good and professional, and therefore we want all faces are in the picture and all people are close to the center of the picture. The given specification are as follows.

1. Camera resolution ≥1200w pixels
2. Face position within center 1/3 of picture

The most important specifications, from the QFD chart, is the CPU frequency and Xun Fei platform used for sound recognition, since they are the base of our project. The specifications on the two should be met first.

3 Concept Generation

The sub-systems of our project is generated with brainstorms according to the requirements of our selected scenario. To satisfy the needs, four sub-systems are generated. The morphological chart is shown in Figure 2.

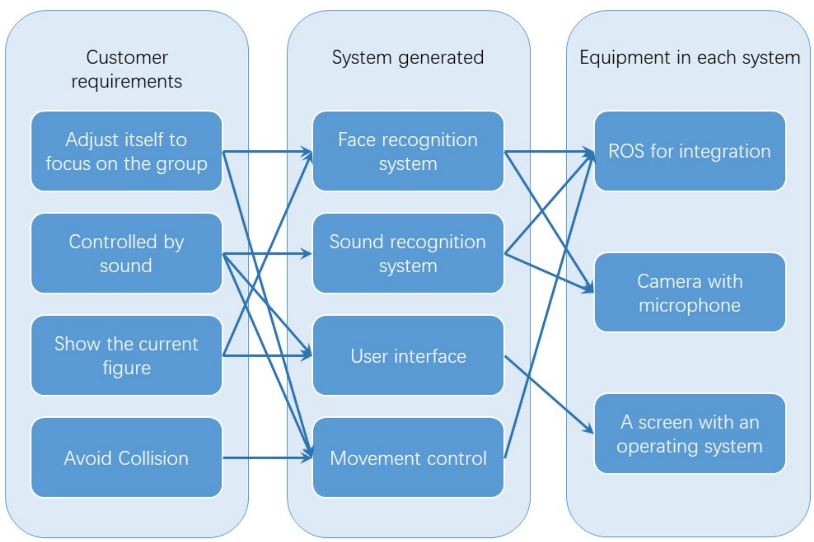


Figure 5: Morphological chart

We can see that we begin from the main customer requirements to generate the systems we need. Then from the system we generated, we figure out what equipment should be used in each system and how they should work with each other.

To summarize, each system and what they are used for are described below in detail.

1. In order to let the robot move, a movement control system is needed. Also, the movement control system will have to ensure that the robot will not collide in any object or human during moving.
2. In order to let the robot adjust itself according to the group being photographed, the robot has to detect all the faces within the picture and locate them close to the center of the picture. This

will require a face detection system. According to the faces detected, the robot will communicate with the movement control to let the robot move left, right, forward or backward and move the camera upwards or downwards to make all faces close to the center of the figure.

3. To let the customer control the robot with sound, the robot has to recognize the sound taken by the microphone. A sound recognition system is needed to recognize the voice instructions from the customer and translate them to commands that can be used by the movement control system.

4. A user interface should be developed to show the current figure to the customer. Also, any instruction came from the customer such as send the picture to phone can be showed in the user interface.

4 Concept Selection

4.1 Chassis Control System

The chassis consists of two DC motors with high-precision encoders and two corresponding wheels, which are provided by our sponsor. The main function of the chassis control system is to detect obstacles and avoid collisions, while the function of moving has already been built in ROS. To implement this function, several engineering specifications are found as:

1. The measurement of distance should be precise as 0.01 m.
2. The hardware should be able to measure surroundings at any direction.
3. The detection range should be greater than 0.5 m, which is the safe range we define.

With these specifications, we find three commonly-used device as:

1. LiDAR
2. Depth Camera
3. Ultrasonic sensor

The corresponding scoring matrix is shown as in Fig 6.

We can see from Fig 6 that LiDAR gets the highest scores compared to the other two. For all design criterion, LiDAR shows itself as the best solution to our requirements.

Hence, LiDAR is chosen to be implemented with motors and wheels to handle detection.

4.2 Sound Recognition System

Sound Recognition System is used to manually adjust the position of robot and camera to make the photo better. So, we hope the voice recognition system can distinguish our order quickly and accurately. To implement this function, several engineering specifications are found as:

1. This system should support speech wake-up function. Also, this function should be naturally.
2. This system should have a high accuracy of speech recognition.
3. This system should have high semantic resolution accuracy.
4. This system shouldn't cost too much.
5. This system should support Linux operation system.

Design criterion	Weight factor	Unit	LiDAR			Depth Camera			Ultrasonic Sensor		
			Value	Score	Rating	Value	Score	Rating	Value	Score	Rating
Detection range	0.30	m	4	8	2.4	1.5	4	1.2	5	9	2.7
Distance precision	0.30	mm	0.1	9	2.7	0.5	6	1.8	1.5	4	1.2
Cost	0.05	RMB	500	8	0.4	1200	4	0.2	650	6	0.3
Response time	0.20	ms	0.2	10	2	16.7	6	1.2	20	4	0.8
Durability	0.15	/	High	8	1.2	High	8	1.2	Fair	6	0.9
					8.7			6.4			5.9

Figure 6: Chassis control system scoring matrix

With these specifications, we find three commonly-used open platform as:

1. Xun Fei Open Platform
2. Baidu Open Platform
3. Tencent Cloud

The corresponding scoring matrix is shown as in Fig 7.

Design criterion	Weight factor	Unit	Xun Fei			Baidu			Tencent		
			Value	Score	Rating	Value	Score	Rating	Value	Score	Rating
Speech Wake-up Naturality	0.30	%	98	9	2.7	93	8	2.4	0	0	0
Speech Recognition Accuracy	0.30	%	98	9	2.7	97	8	2.4	94	6	1.8
Semantic Resolution Accuracy	0.30	%	97.4	9	2.7	96	8	2.4	92	5	1.5
Cost	0.05	RMB	200	6	0.3	0	10	0.5	0	10	0.5
Support Linux	0.05	/	Yes	10	0.5	Yes	10	0.5	Low	6	0.9
					8.9			8.2			4.7

Figure 7: Sound recognition system scoring matrix

We can see from Fig 7 that Xun Fei Open Platform gets the highest scores compared to the other two. For all design criterion, Xun Fei shows itself as the best solution to our requirements.

Hence, Xun Fei Open Platform is chosen.

4.3 User Interface

The concept selection process of the subsystem of user interface mainly concerns on the following three aspect:

1. The CPU capability of the device. Our device also serves as the server for our sound recognition algorithm and face detection algorithm. These process can be a burden for the device. We require a device with CPU at least capable of containing all the processes at the same time.
2. The system running on the device. Since the device serves as the terminal for robot-base-control, camera, microphone, speaker and LiDAR. We need a system on which python, ROS and all these hardware can all work well.
3. The User Interface application workload. At last we need to design a user interface application running on the device interacting with the user interface. Due to the limited time provided for the project, we need a system easy for development ensuring the best user experience.

Other affecting facts include the performance of Bluetooth which we use to send our photo to our user. A solid touchscreen where users interact with the system. We also should take the cost into consideration.

With all these facts and requirements in head, we have come up with the following top five concepts to select from, they are all named by the device name:

1. Intel NUC with monitor (Linux)
2. Microsoft Surface (Windows 10)
3. Android pad (Android)
4. Apple IPad (IOS)
5. Raspberry Pi with touchscreen (Linux raspbian)

To compare between the five concepts, we make the following scoring matrix.

Design criterion	Weight factor	UNIT	Intel NUC with Monitor			Android Pad			iPad		
			Value	Score	Rating	Value	Score	Rating	Value	Score	Rating
Hardware and Ros Compatibility	0.30	/	mature	10	3	mature	10	3	not mature	0	0
UI workload	0.20	week	4	10	2	8	5	1	8	5	1
CPU capability	0.20	GHz	2.4	8	1.6	1.8	6	1.2	2.49	8	1.6
Bluetooth performance	0.10	/	Solid	10	1	Solid	10	1	Solid	10	1
WIFI performance	0.10	/	Solid	10	1	Solid	10	1	Solid	10	1
Cost	0.10	RMB	4000	8	0.8	3500	9	0.9	4000	8	0.8
Total					9.4			8.1			5.4

Figure 8: UI scoring matrix (part1)

Design criterion	Weight factor	UNIT	Raspberry Pi with Touchscreen			Microsoft Surface		
			Value	Score	Rating	Value	Score	Rating
Hardware and Ros Compatibility	0.30	/	not mature	0	0	mature	10	3
UI workload	0.20	week	4	10	2	8	5	1
CPU capability	0.20	GHz	1.5	5	1	2.6	9	1.8
Bluetooth performance	0.10	/	Solid	10	1	Solid	10	1
WIFI performance	0.10	/	Solid	10	1	Solid	10	1
Cost	0.10	RMB	1000	10	1	6400	5	0.5
Total					6			8.3

Figure 9: UI scoring matrix (part2)

From the matrix we can find that Intel NUC with monitor and Linux system get the highest rating. From the weight factor of different design criterion, we can find that the compatibility is the deciding factor in this concept selection. Concepts with immature compatibility are eliminated even if they may have the advantages in other category. Besides, our selected concept, which is the Intel NUC mini computer, has lighter UI workload and reasonable CPU capability. These advantages help it defeat the other two candidate.

Besides, the Bluetooth performance and the WIFI performance are also important for function realization. However, since all of the top five concepts to choose from have solid performance. Their weight is also reduce to 10 percent.

As a conclusion, we choose NUC mini computer from Intel as the device running UI.

5 Final Design

5.1 Chassis Control System

5.1.1 Engineering Design Analysis

As mentioned in Section 4.1, the engineering specifications for the chassis control system are detection precision, scanning angle and maximal detection distance. LiDAR fits all these specifications as it scans the surrounding for 360 degrees with the precision of 0.1 mm and maximal distance of 4 m. Since the robot kit is provided by our sponsor, which has ROS built in within, we choose to use simultaneous localization and mapping (SLAM). This is a common package in ROS regarding robots' localization and navigation. Combined with other packages in ROS, we can write our own program to control the movement of our robot.

LiDAR is then installed on our robot in the middle layer. The chassis of our robot is provided by our sponsor, so are LiDAR and the depth camera. Thus, we can only choose one of them. LiDAR is chosen due to the highest score of the socring matrix shown in Fig 6.

5.1.2 Design Description

The software part is described by the flow chart in Fig 10. Since the hardware, such as LiDAR and motors, are provided by our sponsor, which they purchased from Intel, a detailed layout drawing with dimensions cannot be made.

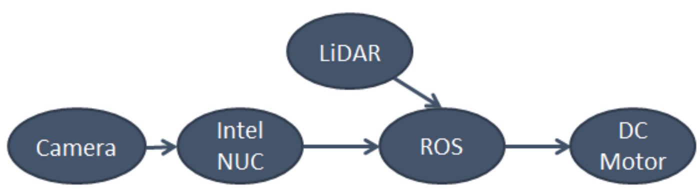


Figure 10: Chassis control system flow chart.

The algorithm behind our robot is first to gather image and data from the camera, which has a microphone. Then the host, Intel NUC, can calculate the best position for a group photo, if there are no voice commands; otherwise, Intel NUC can separate the key word from the voice and convert it to the command. In either case, Intel NUC sends the moving commands to ROS, which is connected to LiDAR. While ROS controls the motors according to the commands, it monitors the surrounds. If a person, or something, is detected by LiDAR to be too close, ROS will immediately stop the robot until the path is clear. The codes which implement functions on our robot are listed in Appendix.

5.2 Sound Recognition System

5.2.1 Engineering Design Analysis

As we mentioned before, the sound recognition system is used to transfer our speech to the final order. Here, Speech Wake-up Naturality, Speech Recognition Accuracy and Semantic Resolution

Accuracy are all important to the effect of our project. The naturality of speech wake-up decideds whether we can wake the system up quickly at the time we want. Besides, it also ensures that the system won't waste resources when we don't need it. The high accuracy of speech recognition provides us with the assurance that he can accurately identify our instructions. Also, since sometimes we need to accept instructions in Chinese, we need our system to support semantic resolution. These three parameters are the most important for the performance of the sound recognition system. What's more, the support of Linux and the cost of it are also crucial. Using scoring matrix, Xun Fei is finally choosed. It has an extremely high accuracy rate of 98, 98 and 97.4 percents in these three aspects. It also support the Linux system and has a low cost.

5.2.2 Design Description

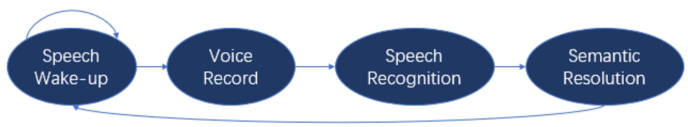


Figure 11: Sound recognition system flow chart.

As shown in flow chat in Fig 11. The sound recognition system will keep detecting for key words to wake-up. Once it detects the key words we set, it will start to record the voice using the microphone and save it as a wav file. After that, the wav file will be uploaded to the open platform and be transferred to a sentence there. Then, we will use word segmentation technology to extract the instructions. Then the system will sleep ant wait for the next wake-up instruction. The corresponding codes are listed in Appendix as *record.py* and *voice.py*

5.3 User Interface

5.3.1 Engineering Design Analysis

The engineering fundamentals support this concept is the ability to check the compatibility between hardware and the terminal, the communication between different programming environment as well as a concept of MVC frame for the UI application. The compatibility between hardware determine whether the camera, microphone, LiDAR, base-element and speaker can function as they are expected on our robot. The communication between languages allow data and signals to be received and emitted in the terminal. The MVC frame concept enables us to establish a frame on which we build our UI.

The determination of specific parameter are mainly based on a positive-negative mechanism. Most of the compatibility test would give out a result of yes or no which directly tells us whether the device meet the requirements. Three of our parameters on determined by the standard level of technology. The pixar, the time for Bluetooth to send the photo and the FPS are of the average level of the products we can acquire from the market. Also the parameter of CPU is determined the calculation amount of the algorithm.

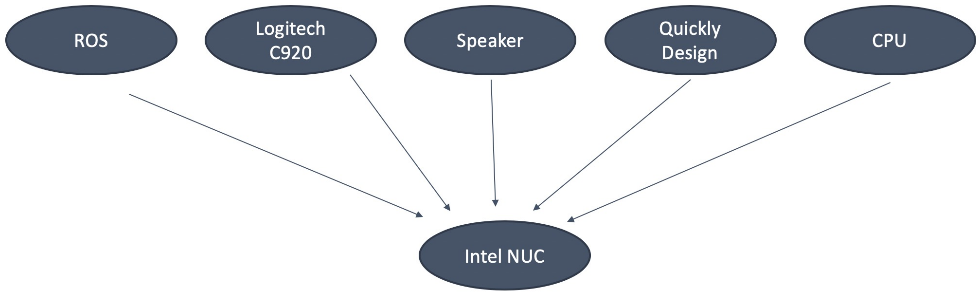


Figure 12: Rationale Behind Decision of UI

The rationale behind the decision can be split into several parts.

1. ROS. ROS is an operation platform dedicated to robot development based on the Linux system. With the platform, huge amount of open source codes and scripts can be found for base-control, camera detection and communications between different parts. It also have it's unique file system with Linux to enable the synchronization of multi-process. These property of ROS leads us to a Linux operation system.

2. Logitech C920. USB-camera is now widely supported on Linux system and can be detected by the node "usb-cam" embedded in ROS. Also there are drivers out there for the PID and VID of USB camera product from logitech. Due to it's high photo quality, reasonable size and weight, we resemble it with our Linux terminal.

3. Quickly Design. Quickly Design is a tool enabling us to quickly setup a UI framework on which we can write our UI program. The tool provides graphic interface for UI design and established python files to link functions with buttons. The tool lead us to our Linux system because it can significantly lower the workload of UI.

5.3.2 Design Description

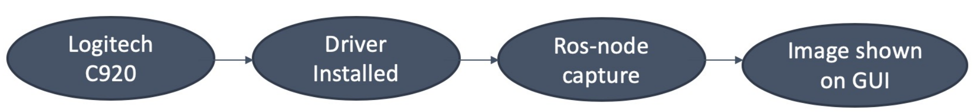


Figure 13: UI Flow chat 1

After sufficient research is done to prove that Logitech C920 can work well with ROS and Linux. We purchase the camera from online website. Than we install the open source drive written for this camera model on Linux. Then we run the ROS and the command "Roslaunch usb-cam" in the shell. In the last step we capture the stream from the camera from a GUI tool named rqt-image-view. The hardware is successfully installed into the system

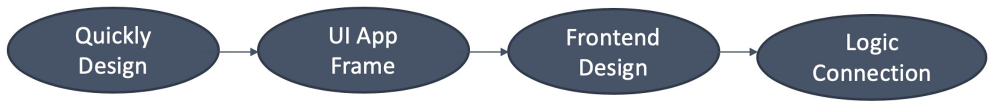


Figure 14: UI Flow chat 2

First we need to install the tool named Quickly. With this tool installed, we can use it to create an App project on ubuntu with basic MVC frame already written inside. Than we use a UI design tool called "glade" to design the graphic part of the UI. In the end, we connect different actions and functions through the code in python files the project has already built up. The UI is now functioning as we expected.

6 Implementation Plan

Our sponsor WECAN company provided Ros Kit robot to us, and we integrated a lead screw with Logitech C920 Camera onto the robot. The prototype is shown in Figure 15. The hardware component needs a corresponding software implementation, which is listed on the right.

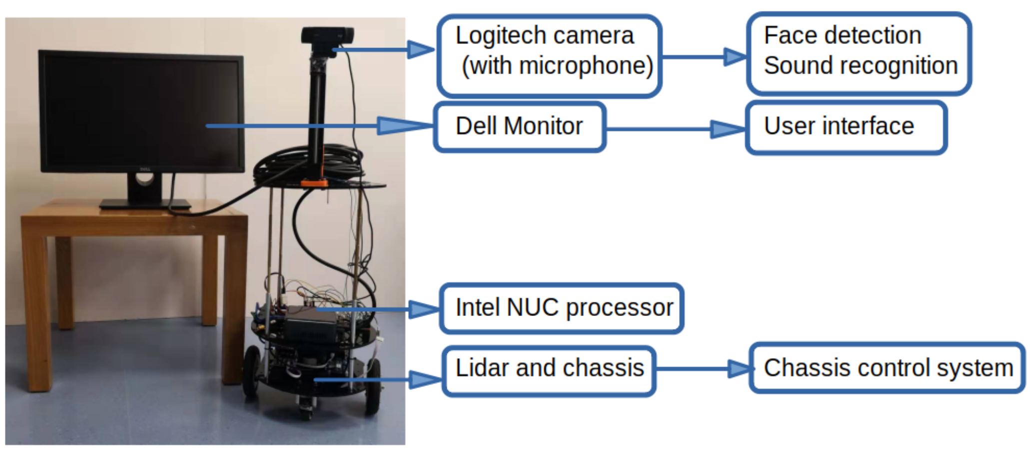


Figure 15: Photograph robot prototype

All software implementation uses Python. The face recognition system uses opencv library to detect the position of faces in the photo. The sound recognition system uses Xunfei online platform to get JSON string from audio input. The chassis movement control is implemented on ROS and Arduino platform. We use Arduino to control the DC motors, and use the navigation library in ROS to avoid collisions based on LiDAR messages. The user interface uses ROS to capture the camera photograph, and Quickly to integrate all the functions inside a graphical tool. The detailed material list is in section 12.

The schedule of our project is summarized in the gantt chart attached in Figure 16. The labels behind each task are the assigned teammates' first character of his/her first name. We spent around 40 hours in service robot idea generation, market investigation and platform decision. We spent 35 hours on robot sensor and camera selection and purchasing, 35 hours on software structure design. Then, we spent 30 hours each, on movement control, UI, sound recognition and face detection implementation. Finally, we spent 40 hours on testing and refine. In total, our team spent 270 hours on this project.

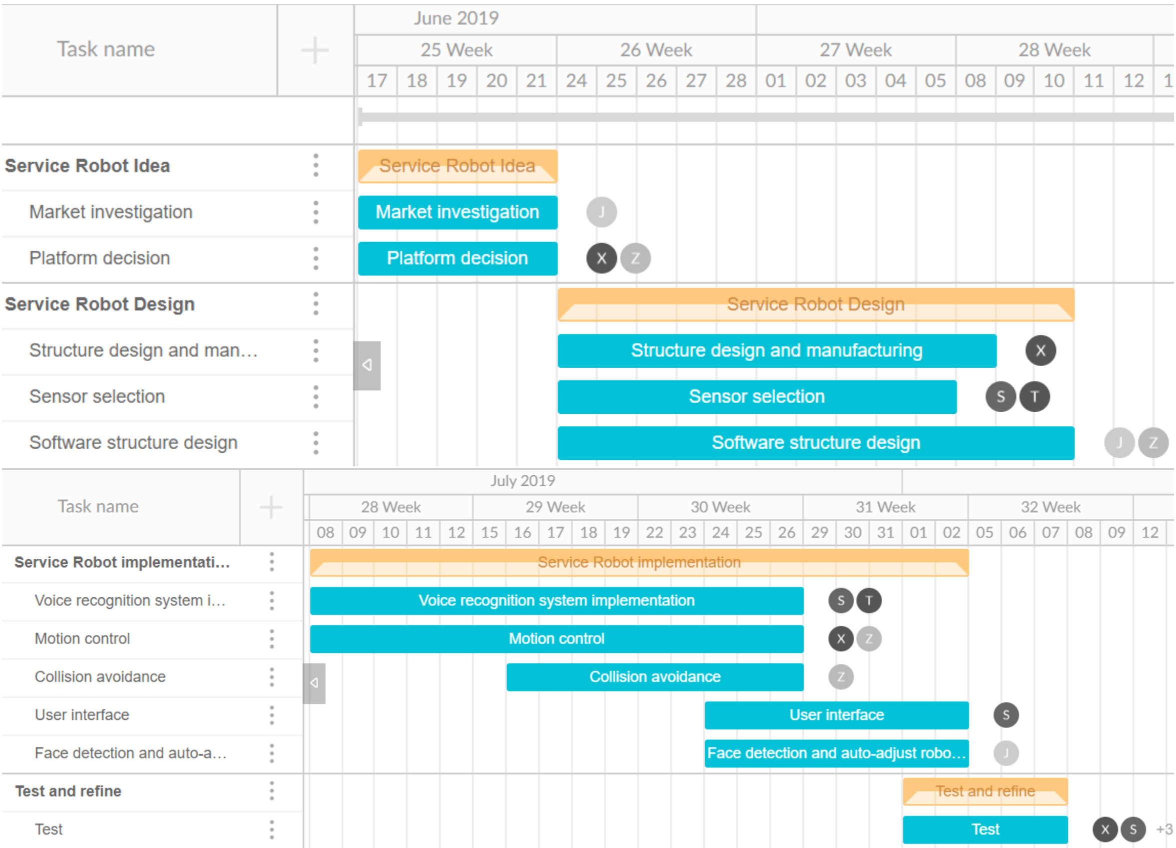


Figure 16: Gantt Chart

7 Test Results

To ensure the customer requirements are met with our prototype, we will test our robot in the following scenario. The robot first automatically adjust itself to ensure the users' faces are around the center of the photo. Then, the user can let it adjust more according to the voice instructions. When it detects an object and stops.

In our test we use a ruler to measure the distance between the edge of the robot and the object. The accuracy of sound recognition and semantic recognition can be assured by the platform we used. After we take a picture, the place of the face in center of a group relative to the whole picture can be measured using a ruler. Some other specifications can also be read from the program.

The test results are:

1. CPU frequency = 2.4GHz
2. Speech recognition accuracy = 98%
3. Semantic recognition accuracy = 98%
4. frames collected per second = 30
5. Object detection range = 12m
6. Distance measured precision = 0.002m
7. Center face place: in middle 1/3 of the photo

8 Engineering Changes Notice (ECN) - Changes in Design since Design Review #3

The four subsystem components design hasn't been changed since design review #3. There are some minor changes on movement control and UI interface screen selection. We planned to use ROS on DC motor control, but finally, we used Arduino on that part. Also, to ensure our users to see the photos clearly at realtime, we changed our platfrom from a 11 inch touchscreen on the robot to a 24 inch display screen on a desk. There is a long HDMI wire connecting the robot and the screen. The wired connection will assure the realtime photo transmission, comparing with wireless. The ECN is in the appendix.

9 Discussion

Our design functions well in satisfying our customer requirements, as discussed in the test results. There are also several parts that can be improved in the future to enhance customers experience.

1. **Robot platform selection:** Our design is based on the Ros Kit, provided by our sponsor. This robot size limits the screen size that can be integrated onto the robot. That screen will not enable the users to view there photo clearly. Thus, we use along HDMI wire to connect the robot with the display screen put on a desk. However, if the robot size can be larger we can integrated screen onto the robot. This will enhance the mobility freedom of the robot.
2. **Movement system:** The motion system is currently controlled by Arduino through voltage input. To make it achieve a more reliable performance, ROS can be used to integrate the encoder message, and give speed or position target input.
3. **Sound recognition speed:** We use Xunfei online platform to perform sound recognition. This will cause 1-2 seconds delay, based on current Internet speed.
4. **User interface:** Our user interface have all the functionality that required by the system. However, it can be further decorated to enhance users' experiences.

10 Recommendations

In our project, we implemented sound recognition, face detection, motion control and graphical user interface subsystems. We carefully designed these four subsystems before start implementation. However, the major difficulty for this project becomes how to integrate all the subsystems together on to the robot afterwards. We highly recommend future teams to clearly design not only the subsystem structure before implementation, but also think of how the connection and communications works between subsystems.

On the implementation part, the robot platform need to be larger and the movement system can be improved by using ROS. The sound recognition delay can be improve by using offline recognition or using more reliable network, and the user interface can be further decorated.

11 Conclusions

To design, build and implement a indoor photograph robot, we divided our tasks into four parts: sound recognition, face detection, motion control and graphical user interface. With these four function, our robot can move around and adjust its position automatically according to where the faces are in the photo. Also, people can further adjust its position using voice command, and it will avoid collisions with people passing-by. After taking a photo, people can get the image through our user interface via Bluetooth.

After investigating and compare several platforms fro implementation, we make the following decision. Motion control system is implemented on Robot operating system (ROS) and Arduino. The collision avoidance uses LiDAR to measure the distance. Sound recognition system is implemented by XunFei online platform. Face detection system is implemented using opencv. User interface is implemented on Intel NUC with a display screen out of the robot.

All of the implementations are using Python. After integrating all the parts onto the robot and finish the implementation, the validation result shows these three systems satisfies our engineering specification. The speech recognition accuracy is greater or equal to 98%. The camera frames captured is greater than 30 per second. The LiDAR detection range is greater or equal to 2m, precision is greater or equal to 0.1m. The automatically adjust can ensure the faces are in the middle 1/3 of the photo.

12 Information Sources and Reference List

1. UM-SJTU JI ‖ http://umji.sjtu.edu.cn/news/2019-0408-1700/. [Retrieved at Aug 2, 2019].
2. Warrington Robotics ‖ https://www.servicerobots.com/photography-robot/.[Retrieved at Aug 2,2019].
3. ROS.org ‖ Powering the world's robots. http://www.ros.org/. [Retrieved at July 23, 2019].
4. iflyrec.com ‖ https://www.iflyrec.com/. [Retrieved at July 23, 2019].
5. gnome.org ‖ A user interface designer. https://glade.gnome.org/ [Retrieved at July 23, 2019].

Appendix

Bill of Materials

The parts we used in this project are listed in table 1. The total cost is within the expected budget.

Part Description	Quantity	Purchase From	Part Number	Price(RMB)
Ros Kit	1	WECAN(Leboshi)		0
HDMI Cable	1	JD	1233220	199
Dell Monitor	1	UMJI Lab		0
Logitech C920	1	Tmall	37176770174	499
Metal Connecting rod	1	Taobao	565472938246	60
Arduino Uno Board	2	Tmall	592121292238	240
Electric Motor with encoders	2	Taobo	597532302078	90
Electric Lead Crew with driver	1	Tamll	553656246409	300
Total				**1388**

Table 1: Budget

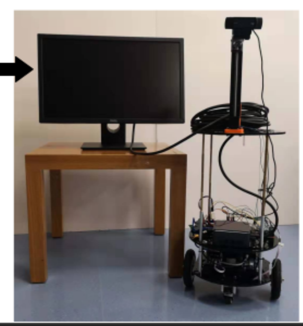
Description of Engineering Changes since DR#3

WAS:

11inch touchscreen

IS:

24inch monitor



Notes:
Need to replace 11inch screen with a larger screen. A HDMI wire is also used to show realtime photo without any delay.

WECAN	
project: Photograph Robot	
Ref Drawing: Display Setup	
Engineer: Tailun Liu	07/29/19
Proj. Mgr: Sidong Wang	07/29/19
Mgmt: Chengbin Ma	07/30/19