# Developing an Identity Recognition Low-cost Home Service Robot Based on Turtlebot and ROS

Chengguang Xu[1], Wenyu Li[1], Jeffrey Too Chuan TAN[2], Zengqiang Chen[1], Han Zhang[1], *Feng Duan[1]

1. College of Computer and Control Engineering, Nankai University, Tianjin, 300350, China
*Corresponding author, E-mail: duanf@nankai.edu.cn

2. Institute of Industrial Science, University of Tokyo, Tokyo, 171-0033, Japan
E-mail: jeffrey@iis.u-tokyo.ac.jp

**Abstract:** Aging population and disabled assistance have become serious social problems these years. However, it is impossible to make sure that every old people has a caregiver. Therefore, the home service robot is a good method to mitigate these issues. Although current robotic technology is mature, service robots with advanced functions are very expensive. In this paper, we developed an identity recognition low-cost home service robot based on Turtlebot and Robot Operating System (ROS). In order to verify the performance of the service robot, we designed a series of experiments in a simulated home environment. The average successful rates of the designed experiments were all above 90%. The results confirmed that the low-cost service robot satisfied our design target.

**Key Words:** Aging population, Disabled assistance, Home service robot, Low-cost

## 1 INTRODUCTION

The aging population problem has become more and more serious these years. According to the 2015 Social Service Development Statistical Communique [1], the population of the aged grew to 222 million in China and took up 16.1% of the population. However, because of the shortage of the caregivers and the expensive labor cost, it is impossible to provide all the aged with personal care. With the development of robotics, home service robot technology seems to be a good method to deal with the issue.

Numerous studies have been done in developing the home service robot. KangWoo Lee presented a human-robot interaction framework [2]. This framework consisted of multimodal, cognitive and emotional interaction modules. He aimed to make the interaction more convenient for human. Ha Manh Do designed an open framework [3] based on active auditory learning. In this framework, sound event could be recognized automatically. In order to prompted the home-service-related language recognition capability, Tzuu-Hseng S. Li created a speech recognition system [4] based on the hidden Markov model (HMM). Some real robots were also developed. Tapio Taipalus developed a robot named MARY [5]. It could fetch things for the aged in a home environment. David Fischinger and his team created Hobbit [6], a care robot supporting the independent living at home. These robots were all fully autonomous. Meanwhile, semi-autonomous robot like Care-Bot-3 [7] was also developed to serve the elder people.

The works described above are excellent and impressive. However, Sandra Bedaf did a research on the relationship between the service robots and the living independent [8]. According to his investigation results, we find that the home service robot developed above have the following weaknesses:

- **Deprived**: The service robots described above are highly intelligent and can finish some daily tasks for the old. However, doing some simple housework is good for the physical health of the old people. Highly intelligent robots may somehow deprive some simple daily abilities from the aged people.
- **Expensive**: The cost of developing service robots above is expensive. Because many precise sensors and equipment are used on the robots. Thus, it is impossible for an ordinary people buying a service robot.
- **Poor Scalability**: The robots are developed on different platforms and systems. Therefore, it is really hard to update the robot without changing the architecture.

In this work, we develop a home service robot based on Turtlebot and ROS. Turtlebot is a cheap and reliable mobile platform compared to the PR2, Pepper, Nao and so on. In order to confirm that the low-cost robot has the designed performance, we optimize the software structure and perception algorithms. We design a four level robot control system based on ROS. All advanced robot functions are managed as ROS packages and stored in the third level of the system. Therefore, it will be convenient to update the robot functions in the future. Finally, we realize the identity recognition system, navigation system and speech interaction system on the low-cost service robot.

This work is supported in part by the Natural Science Foundation of China (No. 61673224), the Research Fellowship for International Young Scientists (No. 61550110251, 61650110511, 61650110510), the Tianjin Research Program of Application Foundation and Advanced Technology (No. 14JCYBJC18300), and the Key Technologies R & D Program of Tianjin (No. 14ZCZDSY00008).

This work is constructed as follows. Section 2 introduces overall construction of the robot. Section 3 describes the realization of the three subsystems in details. In the section 4, a simulated household environment is established and a series of experiments are designed to test the robot performance. The conclusion and the future work of this article are concluded in Section 5.

## 2 ROBOT OVERALL DESIGN

The design of the home service robot consists of the hardware entity construction and the software system construction. Figure 1 shows the robot entity.

### 2.1 Robot Entity Construction

The navigation system, identity recognition system and the speech interaction system are the three main systems of the home service robot. In order to keep the balance of the developing cost and the performance of each system, the hardware components for each system are selected and described in Table1.

Table1. Hardware Components of the Service Robot

| System | Components | Features |
|---|---|---|
| Navigation System | Kuboki Base | 1: Wheel differential mobile system.<br>2: Multi-scalable interfaces. |
| | Kinect Camera | 1: RGB-D depth camera.<br>2: Effective detection distance 0.8m-2.0m.<br>3: Depth data based on Point Cloud Library. |
| Identity Recognition System | Primesense 1.09 Camera | 1: RGB-D depth camera.<br>2: Effective detection distance 0.34m-1.4m.<br>3: Depth data as Kinect camera. |
| Speech Interaction System | Microphone | 1: Noise Suppression.<br>2: Small size. |
| | Speakers | 1: Light weight.<br>2: USB power. |

According to Table1. The reasons for selecting these hardware components are the followings.

(1) **Navigation system**: The space of a home environment is limited. Therefore, mobile platform with big turning radius is not suitable for a home environment. Thus, we choose the wheel differential Kuboki base. Because it is small size and having a small turning radius. In the indoor environment, it is good enough to collect the environment data with an infrared sensor. Therefore, we choose the Kinect depth camera instead of a Lidar.

(2) **Identity Recognition system**: The Primesense camera is a short distance RGB-D camera. This camera is smaller and lighter than the Kinect camera. Thus, it is easy to be fixed on the top of the robot. The most important thing is that the effective detection distance of this camera is 0.34m to 1.4m. Therefore, it is suitable to do a close-up object detection.

(3) **Speech Interaction system**: The noise in the environment will interferes the speech system and may sometimes lead to bad recognition results. In order to reduce the voice noise maximally, we choose a professional microphone used in the news interview. Because the microphone can suppress the noise in the home environment. The speakers are common USB speakers, which is selected just for reducing the weight of the robot.


Fig 1. Robot Hardware Entity

### 2.2 Software System Construction

In order to promote the scalability of the service robot functions, we designed a four level robot control system based on ROS. Figure 2 shows the structure of the four level robot control system. The advanced robotic functions are managed as ROS packages and stored at the service modules level. Therefore, updating the robot functions does not influence the whole software architecture.

(1) **System control level**: This is the top level of the system architecture. When the robot is brought up, functions in this level initialize all the service modules as well as the devices on the robot.

(2) **Mission decision level**: This is the second level of the architecture. Messages from the system control level and the service modules level are fused in this level. Then, the mission decision module forms a command and sends it to the service level.

(3) **Service modules level**: Navigation system, identity recognition system and speech interaction system are managed in this level. This level contains all the service modules of the robot.

(4) **Execution level**: This is the bottom level of the architecture. This level receives commands from the system and service levels. Functions in this level are in charge of

controlling all the devices of the service robot to executes the missions.

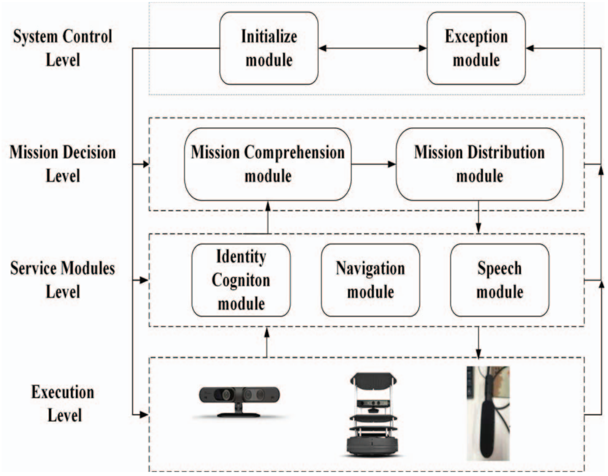


Fig 2. Four Level Robot Control System

## 3 CORE SUBSYSTEMS REALIZATION

### 3.1 Identity Recognition System

Face recognition is used to distinguish the human identity. The robot provides different services based on the recognition result. Therefore, during the design, the recognition accuracy and the response time are the two main concerned properties. The face recognition process consists of the face detection and face recognition. Figure 3 shows the flow chart of the system.

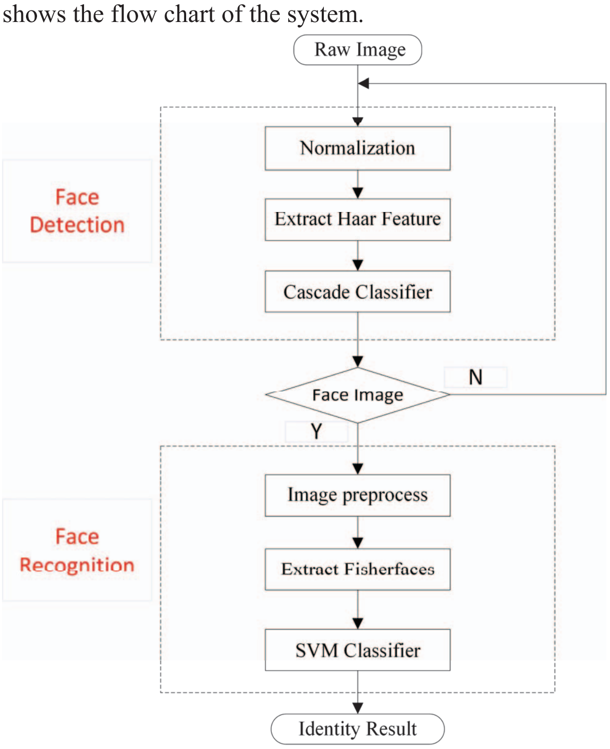


Fig 3. Flow Chart of the Identity Recognition System

The first step is the face detection. In order to prompt the detecting speed, a cascade of boosted classifier is used to detect the human faces in the images. We select the Haar-like feature to describe the human face. Because this feature is simple and can be computed very fast. Integral images are also used to reduce the computational cost in the feature extraction procedure. The structure of this classifier is called Adaboost. Adaboost describes a classifier structure, in which the resultant classifier consists of several simple classifiers and each simple classifier is boosted by different methods [9].

The second step is face recognition. In order to prompt the recognition accuracy, we perform some image preprocessing methods. Firstly, the face images are normalized to the same size and are aligned to the same orientation. Then, histogram equalization and Guassian filter are implemented to reduce the illumination influence. After the image preprocessing, we perform the Linear Distribution Analysis (LDA) on the face images to extract the fisherfaces. The procedure of the feature extraction is explained as the followings. Assume that the number of the face classes is $N$ and the number of the sample in each class is $n$. Therefore, the face database can be defined as

$$\begin{aligned} X &= \{x_1, x_2, \ldots, x_N\} \\ x_i &= \{a_1, a_2, \ldots, a_n\} \quad i = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $x_i$ is a matrix that represents the $i$ class in the $N$ classes and $a_i$ is a column vector that represents a sample in the $i$ class. Then according to the LDA theory, the between-classes scatter matrix is defined as

$$S_B = \sum_{i=1}^{N} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{2}$$

and the within-classes scatter matrix is defined as

$$S_W = \sum_{i=1}^{N} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{3}$$

where $\mu$ is the average value of the whole face set and $\mu_i$ is the average value of the $i$ class and $x_k$ is a sample in a class. Therefore, the projection $W$ that can maximize the ratio of the between-classes to within classes is naturally defined as

$$W_{opt} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} \tag{4}$$

According to the formulas above, the feature extraction is finally an optimization issue, we use the approach proposed by Peter N. Belhumeur [10] to calculate the projection matrix. Once obtained the optimization projection matrix, an image can be projected to the target low-dimensional subspace by the following transition.

$$y = W_{opt}^T x \tag{5}$$

where $x$ is a column vector formed by extending the pixels in the image and $y$ is the coordinate of the face in the low-dimensional subspace. After the feature extraction, we train a Support Vector Machine (SVM) classifier to do the face recognition.

### 3.2 Navigation System

Navigation is the basic function of the service robot. Considering the household environment, which is full of various daily necessities and also space limited, the navigation system of the service robot needs to be accurate, flexible and stable. The navigation system consists of two subsystems. One is a map building system based on the Simultaneous Location and Mapping (SLAM) theory, the other is a navigation system based on ROS. Therefore, the robot can navigate on a pre-established map of the household environment.

Firstly, we try to perform the classical particle filter to establish the environment map. However, it turns out that the mapping process is extremely slow because of the huge number of particles. Then, we test the Rao-Blackwellized particle filter, which turns out to be effective. According to the classical. particle filter method, each particle represents a map of the environment. Thus, numerous particles are generated, which is hard to deal with in the following steps. Unlike the normal particle filters, the Rao-Blackwellized particle filter uses an adaptive approach to reduce the number of the particles. Approaches [11] are used to promote the prediction accuracy and a selective re-sampling approach [12] is implemented to solve the particle depletion issue. Figure 4 specifics the process of SLAM based on the Rao-Blackwellized particle filter.

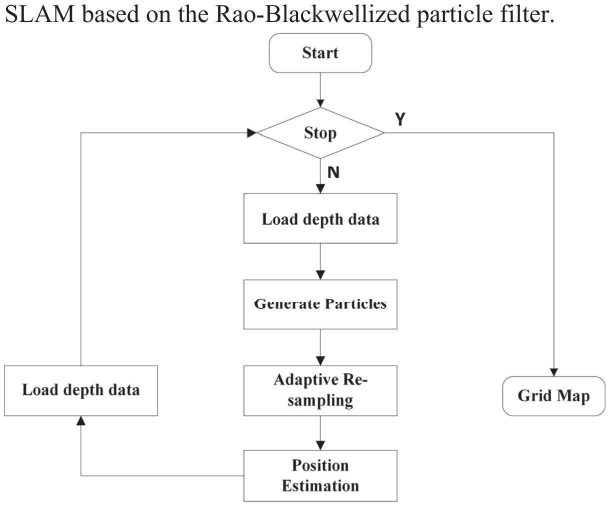


Fig 4. Flow Chart of Rao-Blackwellized Algorithm

The following step is the autonomous navigation. Based on the established map, we use the ROS standard navigation frame to organize the navigation system of the service robot. The ROS standard navigation frame is combined with five internal nodes, which is provided by ROS community. Among the five internal nodes, there are two essential nodes. The global_planner node provides with the global path planning based on the pre-established map, which leads the robot to read the target pose on the map. The local_planner node provides with the local planning, which helps the robot to avoid the static and dynamic obstacles. Figure 5 describes the complete navigation structure.

### 3.3 Speech System

Speech is the most common interaction between human and robots. Therefore, robot speech system is another essential part of the service robot. The key of the robot speech system is the speech recognition problem.

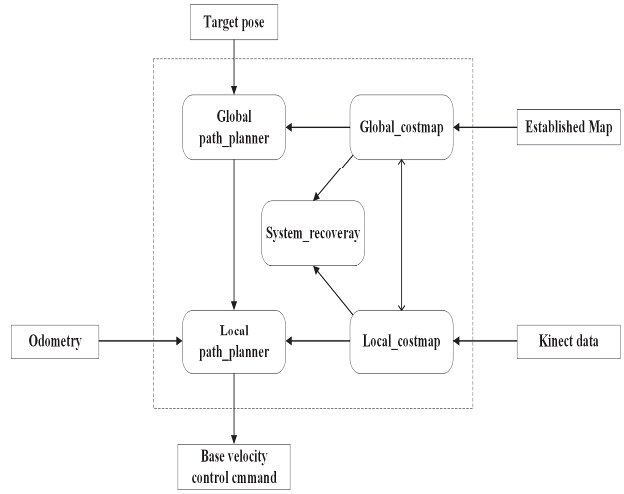


Fig 5. Structure of the Navigation System

It is obvious that the input voice signal contains abundant information such as human voice or noise. Processing the whole signal is computable. There, we extract 39 features within 10ms voice signal, which can represent the main information of the voice signal. A variable named "word error rate" is defined to describe the feature extracting performance. The following equation shows the definition of the variable, in which $N$ represents the whole information amount, $I$ represents the extracted information amount, $S$ represents the replaced information amount and $D$ represents the deleted information amount.

$$WER = \left(I + D + S\right) / N \tag{6}$$

The extracted features are indeed voice segments. After that, the features are compared with the voice clips stored in the voice library within the speech recognition system. This step aims to reduce the words with similar meaning or defect pronunciation. The final stage of the speech recognition is statements combination. In this stage, the processed features are loaded into a hidden Harkov model. HMM is a statistic model, which is essentially a double random process. The probability of the HMM can be computed by the following formulas, in which $Y$ represents the result, $X$ represents the state.

$$P(Y) = \sum_{X} P(Y \mid X) P(X) \tag{7}$$

From the formula, it turns out that the HMM considers the next and last moment of a state, which is similar to the process of the speech analysis based on pronunciation. Because in the speech analysis, the pronunciation of one word within human speech has some relevance with its front word and its back word, which can be accurate modeled by the HMM.

## 4 EXPERIMENTS AND ANALYSIS

### 4.1 Identity Recognition System Testing

In this section, an online experiment is designed to test the correct rate of the face recognition. Face images of five subjects are collected with different face orientation and expression. Table2 shows the information of the five subjects.

Table2. Subject Information

| Subject | Gender | Race |
|---|---|---|
| 1 | M | Asian |
| 2 | M | Asian |
| 3 | M | Asian |
| 4 | F | Asian |
| 5 | M | Asian |

Each subject is collected with 40 face images. The size of each image is 640 x 480 in pixels. Finally, 200 face images are collected in total. Then, 75 percent of the 200 face images are selected as the training set and 25 percent are selected as the testing set. For the reason that the number of the samples in the training set is limited, we implement a five-folder cross validation on the training set. Table3 shows the results in details.

Table3. Five-folder Cross Validation Results

| Training Set | Testing Set | Correct Rate |
|---|---|---|
| 1,2,3,4 | 5 | 93.33% |
| 1,2,3,5 | 4 | 100.00% |
| 1,2,4,5 | 3 | 86.67% |
| 1,3,4,5 | 2 | 100.00% |
| 2,3,4,5 | 1 | 100.00% |
| Average | 96.00% | |

According to the table above, it turns out that the average face recognizing correct rate is 96.00%, which is good enough to recognize a human face. However, there is still a rate below 90% in the Table3. The reason for the low rate is that faces are randomly divided into five groups. The faces images contained in the third groups is almost taken in a much brighter condition. The bad illumination leads to the lower correct rate.

After the cross validation, we use the training set to train a SVM face recognition model and then we use the testing set to test the recognition performance. Table4 shows the confusion matrix of the recognizing result.

Table4. Confusion Matrix of the Face Recognition Results

| Subject | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 10 | 0 | 0 | 0 | 0 |
| 2 | 0 | 10 | 0 | 0 | 0 |
| 3 | 0 | 0 | 9 | 1 | 1 |
| 4 | 0 | 0 | 0 | 9 | 0 |
| 5 | 0 | 0 | 1 | 0 | 9 |

From the Table4, we can see that the recognizing results of the subject 1 and subject 2 are 100%. However, there is still one misclassification in each subject from 3 to 5. The main reason for this still the illumination interference. The illumination in subjects 3 to 5 is bright than that in subjects 1 to 2. Although a face image preprocessing is implemented to promote the recognize rate, it is impossible to eliminate the illumination interference in a real-time system theoretically. However, the final average correct rate is 94%, which can satisfy the requirement of the service robot.

### 4.2 Speech Interaction System Testing

In this section, the performance of the speech recognition system is tested. In order to test the speech interaction performance of the service robot, we design a question and answer (Q & A) experiment. Ten questions are designed and asked randomly by the subjects in Table2 and each question will be asked five times. The experiments are all implemented in a quite circumstance. Because we assume that in a household environment the noise is usually small. Table5 describes the recognition results.

Table5. Speech Recognition Results

| Subject | Accent | Correct Rate |
|---|---|---|
| 1 | Y | 90% |
| 2 | N | 96% |
| 3 | N | 94% |
| 4 | N | 94% |
| 5 | N | 92% |
| Average | 93.2% | |

### 4.3 Navigation System Testing

Figure 6 shows the simulated household environment. The red points in the picture represents the target position in the navigation test. In order to test the obstacle avoidance capability of the robot. We add a static obstacle and a dynamic obstacle to the path of the robot. The robot is designed to navigate from the start point to the rest targets one by one. The test will be repeated for five times.


Fig 6. Simulated household environment consists of living room and hallway.

We also design the following norms to evaluate the performance of the navigation system.

(1) Static obstacle avoiding rate (SOAR) represents the successful rate of avoiding the static obstacle.

(2) Dynamic obstacle avoiding rate (DOAR) represents the successful rate of avoiding the static obstacle.

(3) The navigation time (NT) measures the time of the

navigation process.

(4) Navigation success rate (NSR) represents the final performance of the navigation system.

Table6 shows the navigation results in this experiment.

Table6. Navigation Testing Results

| | SOAR | DOAR | NT(min) | NSR |
|---|---|---|---|---|
| 1 | 100% | 80% | 3.0 | 80% |
| 2 | 100% | 100% | 1.3 | 100% |
| 3 | 100% | 100% | 1.5 | 100% |
| 4 | 100% | 100% | 1.1 | 100% |
| Average | 100% | 95% | 1.725 | 95% |

From the Table6, it turns out that the navigation system has the 100% of avoiding the static obstacles and 95% of avoiding the dynamic obstacles. The average of NSR is 95%, which is same to the ROAR. Therefore, it turns out that the capacity of avoiding the dynamic obstacle affects the whole navigation performance. However, the average successful rates are all above 95%.

### 4.4 Robot Overall Testing

In this section, we designed a final task for the service robot to finish. Firstly, the robot needs to confirm the identity of the people. Then, the robot needs to greet to the people in different ways according to the identity recognition results. Finally, the robot needs to navigate to the target points in Fig. 6. Four subjects are selected. The experiment is repeated for 20 times for each subject. Table7 shows the result of the final test.

Table7. Overall Testing Results

| | Identity Cognition | Speech Interaction | Navigation |
|---|---|---|---|
| 1 | 90% | 90% | 95% |
| 2 | 95% | 95% | 90% |
| 3 | 90% | 90% | 95% |
| 4 | 85% | 90% | 85% |
| Average | 90% | 91.25% | 91.25% |

From Table7, we can conclude that the average successful rate for each subsystem in the final overall performance testing is above 90%. The low rate for subject 4 is caused by the illumination influence and the failure for the robot in avoiding the dynamic obstacles. Above all, the performance of the home service robot satisfied our design target.

## 5 CONCLUSION AND FUTURE WORK

In this paper, we design a low cost home service robot based on Turtlebot and ROS. The hardware used on the robot are cheap to buy and are simple to use. However, with the help of ROS and software optimization, the low cost robot can also have advanced robotic functions like face recognition, navigation and human-robot speech interaction. According to the experiment results, the robot can achieve the advanced robotic functions with an average successful rate above 90%. The software management mechanism in ROS makes it much easier to extend more functions to the service robot without changing the overall architecture of the robot. The final overall performance test shows that the average successful rate is above 90%, which verifies that a low cost robot can also have advanced robotic capacity. The result satisfies our design target.

In the future, we plan to add more robotic functions to the home service robot such as a mechanical arm system. Besides, the robot is low cost but can achieve advanced robotic functions. It is also a good idea to combine the robot with the smart home system.

## REFERENCES


[1] People's Republic of China Ministry of Civil Affairs, http://www.mca.gov.cn

[2] K. W. Lee, H. R. Kim, W. C. Yoon, Y. S. Yoon and D. S. Kwon, Designing a Human-robot Interaction Framework for Home Service Robot, Robot and Human Interactive Communication, Roman, Greece, 2005.

[3] H. M. Do, W. Sheng and M. Liu, Human-assisted Sound Event Recognition for Home Service Robots, Robotics and Biomimetics, Vol.3, No.7, 2016.

[4] T. H. Li, M. C. Kao and P. H. Kuo, Recognition System for Home-Service-Related Sign Language Using Entropy-Based K-Means Algorithm and ABC-Based HMM, IEEE Trans. on Systems, Man, and Cybernetics: Systems, Vol.46, No.7, 150-152, 2015.

[5] T. Taipalus, K. Kosuge, Development of Service Robot for Fetching Objects in Home Environment, IEEE International Symposium on Computational Intelligence in Robotics and Automation, Finland, 2005.

[6] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek and *et al*., Hobbit, a Care Robot Supporting Independent Living at Home: First Prototype and Lessons Learned, Robotics and Autonomous Systems, Vol.75, No.A, 60-78, 2016.

[7] M. Mast, M. Burmester, B. Graf, F. Weisshardt, G. Arbeiter, M. Spanel and *et al*. Design of the Human-Robot Interaction for a Semi-Autonomous Service Robot to Assist Elderly People, Ambient Assisted Living, Part 1, 15-29, 2015.

[8] S. Bedaf, H. Draper, G. Gelderblom, T. Sorell and L. Witte, Can a Service Robot Which Supports Independent Living of Older People Disobey a Command? The Views of Older Peolple, Informal Carers and Professional Caregivers on the Acceptability of Robots, Vol.8, No.3, 409-420, 2016.

[9] P. Viola, M. Jones, Rapid Object Detection Using a Boosted Cascade of Simple Features, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001.

[10] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, NO.7, 711-720, 1997.

[11] G. Grisettiyz, C. Stachniss and W. Burgard, Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Resampling, IEEE International Conference on Robotics and Automation, Barcelona, Spain, 2005.

[12] G. Grisetti, C. Stachniss and w. Burgard, Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters, IEEE Trans. on Robotics, Vol.23, NO. 1, 34-45, 2007.