International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:5, 2008

Visual Tag-based Location-Aware System for Household Robots

Yen-Chun Lin, Yen-Ting Chen, Szu-Yin Lin, and Jen-Hua Wu

***Abstract*—**This paper proposes a location-aware system for household robots which allows users to paste predefined paper tags at different locations according to users' comprehension of the house. In this system a household robot may be aware of its location and the attributes thereof by visually recognizing the tags when the robot is moving. This paper also presents a novel user interface to define a moving path of the robot, which allows users to draw the path in the air with a finger so as to generate commands for following motions.

***Keywords*—finger tip tracking, household robot, location awareness, tag recognition**.

I. Introduction

WITH technological advances, people place more and more importance on their life quality; therefore, establishing a high quality household environment has been a trend of future. In this case, household security robots are playing an indispensable role than ever. For example, these robots are intelligent enough to move autonomously, sweep floors with vacuuming module, detect anomalies via multi-sensing system, monitor houses remotely, send short messages as condition reports--to name but a few. In other words, these robots serve not only as a housekeeper but a house guard. Furthermore, the robots may be living mates if equipped with entertaining functions as well as information querying system.

However, the biggest challenge that household robots are faced with is the complex environment. Different from exhibition fields, offices, factories, and laboratories, indoor design and decorations often change with house owners' consideration. It is quite difficult for a robot to be aware of where it is in an environment with high variance. For this reason, successful technologies including obstacle avoidance, localization, and navigation decide whether or not household are really able to get into household. Present technologies mentioned above are usually carried out by signal processing algorithms with devices adopting infrared light, ultrasonic, GPS, RFID, or laser[1][2]. Navigation, in particular, is one of the biggest challenges that mobile robots confront. One popular technique, dubbed as SLAM (Simultaneous Localization and Mapping), involves having a robot build a map of a local area and tracking the robot's position. While humans find it easy to create "mental maps" in this way, it is difficult and time consuming for a robot to perform the same task. Robots typically use laser scanners and odometers to measure distances for mapping. In order to speed up this process, and to make it more accurate, researchers have previously tried using different algorithms, or set teams of robots to explore an area together.

Although laser-based sensing technologies, such as SLAM, work quite well during localizing a robot itself and establishing the environment map, there are many strong barriers on the way to commercializing household robots: high cost for a laser range-finder, power consumption, and bulky volume of an all-in-one robot. This is why the most acceptable service robots so far are those with vacuuming or lawn-mowing functions which demand much less on localization. In addition to localization, the key for a household robot to offer adequate services is to correctly recognize the attributes of a space. For instance, a house robot in the bedroom might act differently from another in the living room. Especially, for a smaller and safer space like a house, a robot is much more practical when it is able to understand the environment properly rather than to localize itself precisely. Hence, a practicable solution should be considered to make a household robot move according to the path given by householders and recognize where it is at the same time.

In the past few years, solutions to assist robots to be aware of their locations in a sophisticated surrounding based on natural landmarks or artificial tags have been widely proposed. Nevertheless, artificial tags are apparently more applicative than natural landmarks to work with image processing algorithms for reliable localization or navigation. M. Kim *et al.* [3] utilize RFID tags to guide a robot to a specific location, but the RFID tags are more expensive than paper tags. On the other hand, many researchers devoted themselves into various AI algorithms for enhancing visual location awareness. Though paper tags, such as barcode labels, are proposed earlier than RFID tags, they are the most economical solution for location awareness. Thus, many methods for tag design and corresponding recognition are continuously proposed. Briggs *et*

Manuscript received April 30, 2008.

Y. C Lin is with the Mechanical & Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC. (phone: 886-3-5915766; email: VictorLin@itri.org.tw)

Y. Ting Chen is with the Mechanical & Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC. (email: YenTingChen@itri.org.tw)

S. Y. Lin is with the Mechanical & Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC. (email: linssuyin@itri.org.tw)

J. H. Wu is with the Mechanical & Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC. email: JHWu@itri.org.tw)

al. [4] uses self-similar gray pattern landmarks to navigation and localization aids. Yet methods that use peculiar contour or edge information highly depend on the low-level processing results and they are influenced of noise and defocus phenomenon. K. J. Yoon *et al.* [5] apply artificial color tags with symmetry and periodicity for a robot to locate itself. These tags improve the correctness of tag recognition and tracking, as well as the adaptation to noise and focus. H. Wang *et al.* [6] replace 1-D barcodes with simplified 2-D ring-shaped barcodes to perform angle invariance at a better recognition rate.

There is no tag-based localization and navigation solution for general household environments, mainly because of improper tag design. A tag with poor adaptation to various viewing angles, image sensing noise, and lighting conditions significantly limits the self-localization performance of mobile robots.

We believe that it is a breakthrough for navigation if we can have a friendly and visual human-machine interface to do robot path planning as well as a suitable and ordinary tag-based image recognition method used by users in the household environment. Therefore, in this paper, we propose a new household robot system design. Users can post specially made tags in some corners or essential place according to their house environment, and household robot can locate where it is when it is moving. Moreover, users can draw the path they want robot to move by a fingertip in front of a camera to assist the robot to do path planning.

Consequently, our research objective is to integrate visual tag-based image recognition and human-machine interface with the fingertip detection techniques into a household robot system. It is not only to achieve the robot path planning and environment recognition, but also to carry out a whole new interaction way between users and household robots.

The rest of the paper is organized as follows. In Section 2, we describe the household robot system architecture on computer client, visual tag-based image recognition, human-machine interaction, and network communication. In Section 3, we present the proposed system implementation platform for demonstrating our household robot based on human-machine interface by fingertip detection and visual tag-based image recognition. Afterwards, the scenario and experiments are shown in Section 4. Finally, we discuss the conclusion and the future works in the Section 5.

II. System Architecture

The overview of our household robot system architecture is shown in Fig 1. The architecture consists of four important parts: a user's computer as a client, a household robot, human-machine user interface, and network communication. Each part has its own features.

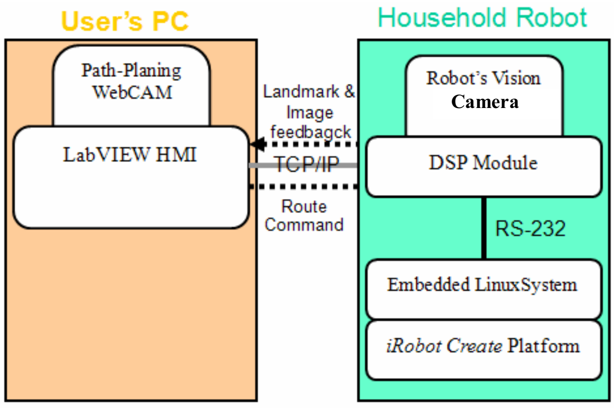


Fig. 1 system architecture

A. Computer Client

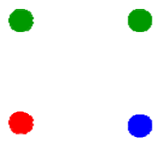
This computer is a local client device for a user to generate commands by human-machine interface. We have a webcam in computer client to show the continuous images and capture the information of the user's fingertip position on human-machine interface. The information will be used to do robot path planning further.

B. Household Robot

The household robot represents a robot platform which has a DSP-based image processing module, an ARM-based embedded module with Linux operating system, and an iRobot create mobile module. In the image processing module, the DSP chip is mainly responsible for tag recognizing, motion command sending, and video transfer back to the computer client. Finally, the motion commands will be sent to the mobile platform, iRobot create, to make it move.

C. Artificial Tag

Another feature of our household robot system is the artificial tag design for providing location information for a mobile robot. These tags, made of paper, are designed with two-dimensional colored patterns, shown as Fig. 2. A single tag contains four colored circles that are located at corners of a rectangle, and each circle is colored as anyone of red, green, and blue. Since each tag comprises four circles with three possible colors, there are 15 possible combinations in all, which is sufficient for general household use. In other words, 15 distinctive locations may be defined by the tags. This tag design enhances the reliability of tag recognition via symmetry and duplication. For example, the tag colors strengthen the robustness to image noise and defocus. Furthermore, the arrangement of the two-dimension, to some extent, substantially improves the adaptability to various viewing angles and distance during recognition.



(Green, Green, Red, Blue in counterclockwise from the upright corner)
Fig. 2 design of a two-dimensional colored tag

The following statements are the image processing flow for visual tag-based image recognition, as show in Fig. 3. Issues on implementation of our household robot system will be discussed in detail in section 3.

Stpe1: To acquire image from a camera
Step2: To filter and separate pixels into color red, blue, and green, and then to calculate the number and ratio of three colors for identifying the type of tag
Step3: To compress and stream the video

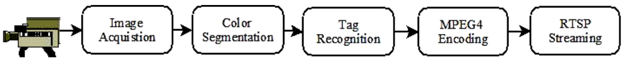
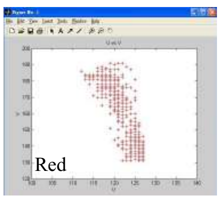


Fig. 3 image processing flow

Prior to the color segmentation phase shown in the above figure, we have to decide the color model on which the analysis of color is based. The reason why we choose the YUV color model is described as follow. First, the Y component in this model stands for luminance whereas U and V represent for chrominance. Among these three components, we employ U and V components for discrimination of different color. Doing so can eliminate the influence caused by varying environmental illumination since Y component is not taken into consideration of color segmentation process. Second, most of the video compression algorithms support and benefit from the YUV/4:2:0 sampling format. This is due to the fact that human eyes are less sensitive to color information compared to the illumination one. As a result, we can reduce the data size occupied by each pixel to improve the overall compression efficiency in advance.

Several color images are captured to construct the image database. Then, we plot the scatter plots for each image on YU, YV, and UV color plane to find out the discriminating rules for red, blue, and green color, respectively. The scatter plot of red, green, and blue for UV color plane is given in Fig. 4. By analyzing the distribution, we can obtain the discriminative rules for each color quickly and easily. Table I provides the final rule.



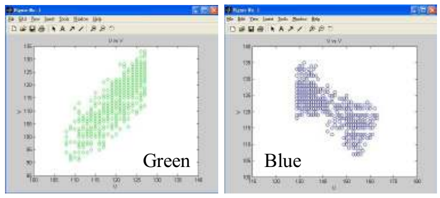


Fig. 4 scatter plot of red, green, and blue for UV color plane

TABLE I
FINAL RULES FOR COLOR RED, GREEN, AND BLUE

Color	Component Range
Red	$60 \le Y \le 220, \quad 85 \le U \le 127, \quad 180 \le V \le 255$
Green	$50 \le Y \le 220, \quad 20 \le U \le 125, \quad 20 \le V \le 110$
Blue	$60 \le Y \le 220, \quad 132 \le U \le 255, \quad 5 \le V \le 125$

We compared the visual tag to the tag database after applying the rule of color segmentation to captured image. In our system, each tag is composed of four circles with same size. The color of each circle can be anyone among red, green, and blue. Therefore, tag [red, blue, blue, green] and tag [red, red, red, green] represent two different tags. The tag design is simple; therefore, the tag recognition process is very efficient. Besides, it is effortless to build the tag. In order to calculate the number of circle of each color, we divide the total pixel number of three colors by 4 to obtain the base number of one single color, which is then used to calculate the proportion of the number of circle. The way we applied to calculate the proportion could reduce influence owing to environmental noise, which is straightforward and computationally efficient.

The image data are fed into the MPEG-4 compression algorithm. The compressed imaged data are then streamed to the website we built. Users connect the website to observe what the robot is now capturing and receive the tag at the same time. To tell the user where the robot is, we associate different visual tags to different physical entities, for example, door, window, and table, etc.

D. Human-Machine Interaction by Fingertip Detection

Users can interact with a system easily through a well-designed HMI (Human-Machine Interactive) which has many solutions to carry out satisfying user experiments. Fingertip tracking is regarded as one of the best solutions because of its intuitive operation. Two achievements are commonly used for detecting gestures or fingertip position of a user: one is based on wearing a glove or sensors on a user's hand [7] [8], and the other is focused on computer vision [9] [10]. It is easier to get precise information of hand gestures by using the glove [7] or infrared devices [8]. Notwithstanding, the cost of implementation is expensive. Some devices are worn on a user's hand necessarily, which bring encumbrances and inconveniences for users. In this paper, the fingertip detection is achieved by a vision-based method. After building up the map of environment, a user can plan the patrol path of a robot in the air by waiving his finger in front of the computer, as illustrated

ISSN: 2517-9942
Vol:2, No:5, 2008

in Fig. 5.

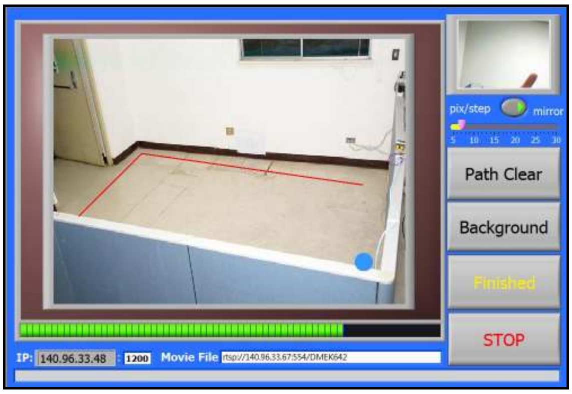


Fig. 5 HMI of fingertip path planning

At first, the up-right corner is the image grabbed by the camera, while the position of the grabbed fingertip will be projected to the environment map on the left side. Next, the user plans the path by moving and pausing his finger intermittently to draw path lines. A confirmed message will be shown if the user moves his finger out of the field of view of the camera, which means the planning is finished. Subsequently, the path information will be translated into motion commands and then sent to the robot through the TCP/IP protocol. Finally, the real-time video captured by the patrolling robot as well as the recognized results will be sent back and displayed on the same HMI. The whole process is shown in Fig. 6.

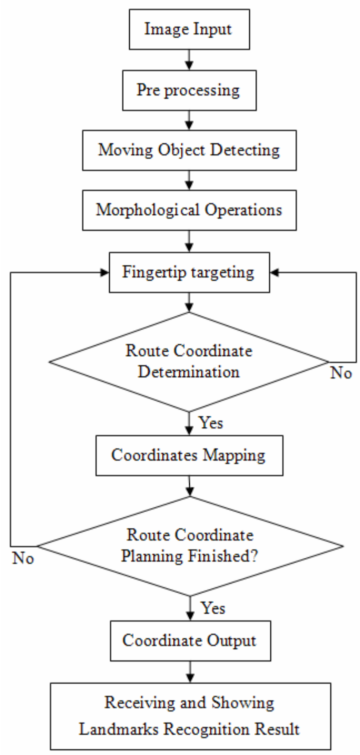


Fig. 6 process flow of fingertip path-planning

Some preprocessing of the input is necessary, such as morphology. The main purpose of the preprocessing is to filter the noise and interference caused by the background. The first frame is treated as the background image. We capture each frame from the camera and then obtain the position of the user's fingertip by applying a simple algorithm as motion detection. A coordinate point is confirmed when the user pause his fingertip for more than 1000 ms. However, a user may do so with slight tremble on his finger; thus, a tolerance is defined so as to ignore the tremble. The final path is the route of coordinate points.

E. Network Communication

The network communication is the medium between computer client and household robot. There are three types can be accomplished in our network communication: one-to-one, one-to-many, and many-to-one.

1) One-to-one: This is the most basic type for one client user to control one household robot, as demonstrated in Fig. 1.

2) One-to-many: This is for the situation that one client user can control and monitor multiple robots simultaneously, and then to further organize collaborative tasks by coordinating individual robot to execute specific task, as shown in Fig. 7.

3) Many-to-one: This type is for multiple client users to monitor but not to control a household robot simultaneously, with the uniqueness of control token taken into account, as shown in Fig. 8.

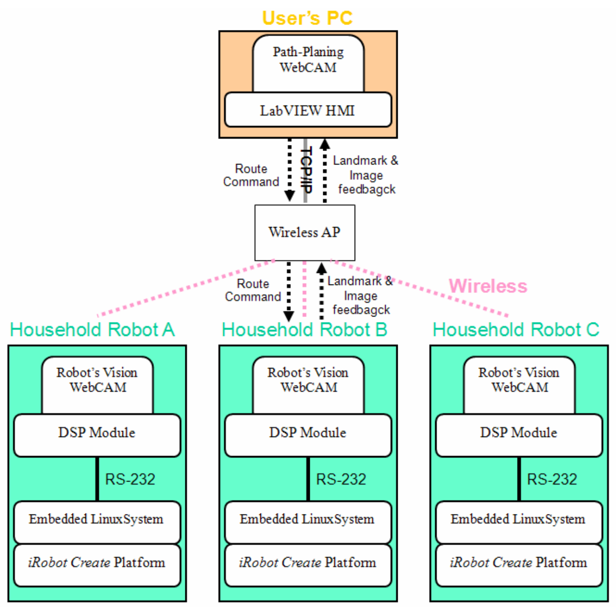


Fig. 7 one-to-many network architecture

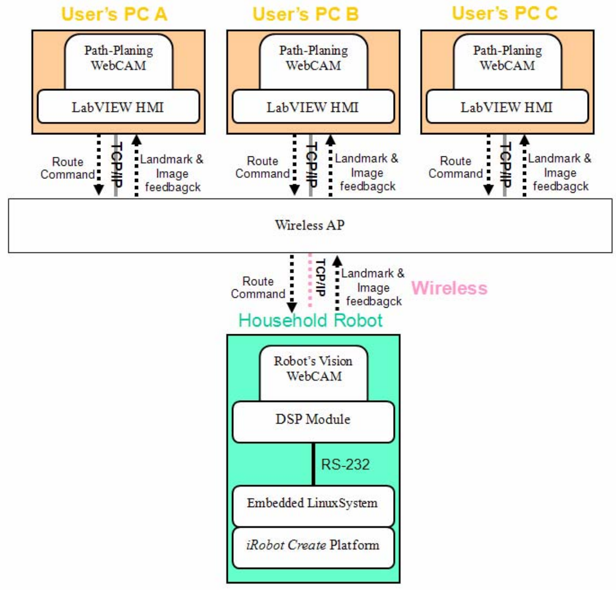


Fig. 8 many-to-one network architecture

III. System Implementation

A. Hardware of Image Processing

We use CCD camera with NTSC TV standards whose sampling rate is 30 frames/sec. ATEME DMEK642 is our development board for image acquisition and processing, and the digital signal processor (DSP) on the board is TI DM642 (600 MHz, 8 * function units, 32 * 32 bits registers, 32KB L1, 256KB internal memory). Besides, the board includes output interface such as video, sound, network, etc. We exploit the powerful capability of this DSP board in the household robot to recognize the tags correctly and quickly in the environment, and then to stream the results of tag recognition to user clients in real time by network communication.

B. Mobile Platform for Household Robot

For effective and rapid household robot system implementation, we choose iRobot Create as the mobile platform for our household robot. The iRobot Create is a mobile out-of-the-box platform. It is not necessary to assemble the driving system or deal with low-level code. It has a lot of open control interface which can be integrated with external sensors, computers, and other functional hardware. In addition, it equipped with some basic sensors like omnibearing infrared sensors, bump sensors, and wheel drop sensors. Developers are also able to program its sensing and control functionalities. Many robot applications use iRobot create as their mobile platform in research and business fields [11].

The design of our modified household robot platform is as Fig. 9. It includes: 1) a power supply module: lead-acid battery provides the source of electric power through this module to distribute voltage into different voltages (12V*3, 5V, and 4V) to support different hardware module in the following; 2) a CCD module: it is for acquiring images of the external environment and artificial tags; 3) a DSP board: images are transferred from the CCD module into the board for tag recognition; 4) an embedded Linux system module: this module is for controlling the mobile platform; and 4) a wireless Ethernet adapter module: the DSP board connects to network via this adapter.

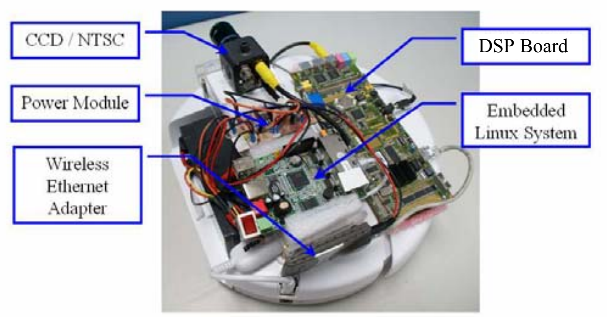


Fig. 9 modified mobile platform

C. Software of Mobile Platform Control

The software architecture of our embedded mobile platform is as below. Two major units are an RS-232 serial I/O unit and a mobile platform control unit, as illustrated in Fig. 10.

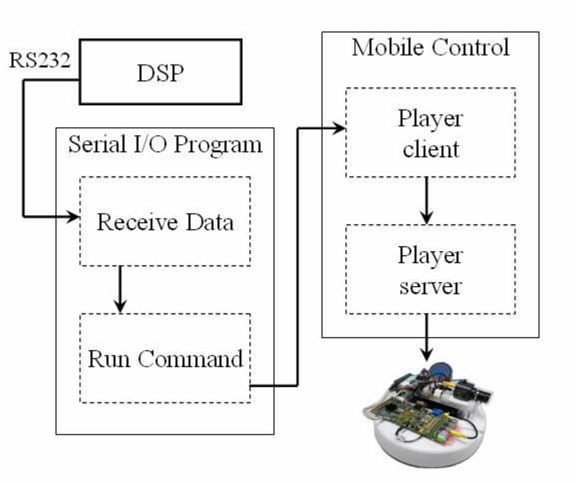


Fig. 10 design of modified household robot platform

The serial I/O unit receives data from the DSP board, and then read commands to trigger the mobile platform control unit to make the mobile platform move. In general, people use a while loop in program to check whether the command data are received or not, which is known as "polling". This polling method is tractable, but its drawback is heavy CPU resource consumption. Therefore, we apply system interrupt functions instead to receive data from serial I/O port for reducing system load.

We use a program package "Player" as the software architecture in our mobile platform [12]. Player is a network server for robot control. Running on a robot, Player provides a simple interface to the robot's sensors and actuators over the IP

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:5, 2008

network. The client program talks to Player over a TCP socket, reads data from sensors, writes commands to actuators, and configures devices on the fly. Player supports a variety of robot hardware. The original Player platform is the ActivMedia Pioneer 2 family, but several other robots, like the iRobot Create that we use, and many common sensors are supported. Player's modular architecture makes it easy to add support for new hardware, and an active user/developer community contributes new drivers.

IV. Scenario and Experiments

A. Household-Robot Control Scenario

The operation scenario is as follows: First of all, a user executes the path planning program on a remote computer. The program keeps capturing images of the user's fingertip waving in front of the webcam, as shown in Fig. 11.

Fig. 11 fingertip detection and tracking

Then, the user moves his finger to draw a trajectory on the pre-established picture of the household environment, as demonstrated in Fig. 12. When the user completes the drawing, the trajectory data are then translated into position coordination as commands for the following robot motion control. These commands are serially transmitted through the wireless network to a DSP board on the robot. This DSP board serves as an IP command server which may receive and transmit data via TCP/IP protocol. Afterwards, the DSP board transfers the same commands to the mobile robot, iRobot, through a RS-232 interface to trigger the motion.

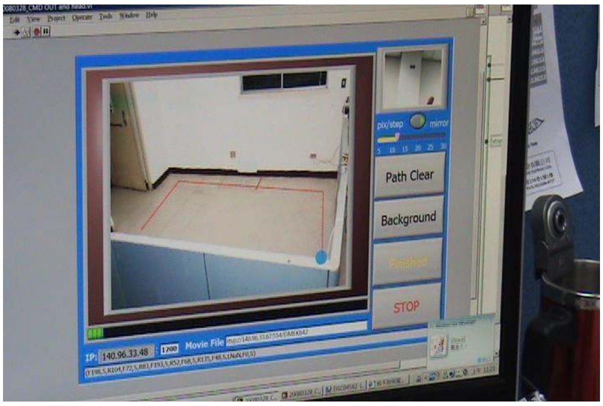


Fig. 12 path planning by the motion of fingertip

During moving, the other camera set up on the robot continuously captures images from the surrounding and transferring the image data to the same DSP board for tag recognition. The results of recognition help the robot to be aware of its location, as displayed in the Fig. 13.

Fig. 13 moving and tag recognition

Meanwhile, the DSP board compresses and transfers the images back to the remote computer via RTSP protocol. The user may watch the streaming video and check the locations the robot just passed by, as shown in Fig. 14.

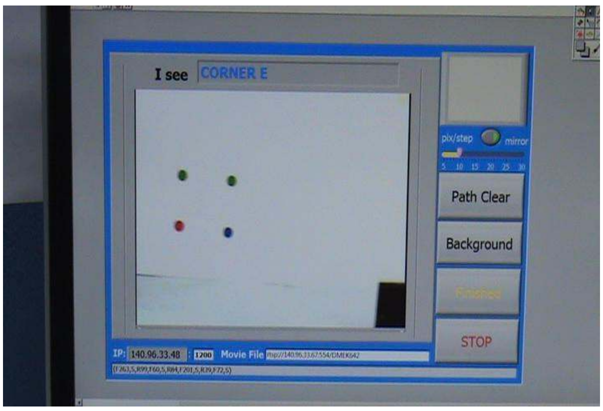


Fig. 14 streaming video and recognition results

B. Experiments

In evaluation of this household robot system, three experiments are performed in this section because there are two important aspects in our system: tag recognition and a mobile platform. The three experiments are: 1) the accuracy of tag recognition in static state; 2) the accuracy of tag recognition at different moving speeds; and 3) the moving error of the mobile platform. There are five artificial tags, “R,B,R,B”, “R,G,R,R”, “B,B,R,B”, “G,G,R,B”, and “G,G,B,B” to be tested in following experiments.

1) The Accuracy of Tag Recognition in Static State

Our design of experiment 1 is as follows. The robot and with a camera is stationary. The distance between the camera and the tag is 45 cm. There are five tags to be recognized and the each tag is tested 50 times in succession. The results are shown in Table II.

TABLE II
STATIC TAG RECOGNITION RESULTS

Tag	Color	Place Symbol	Average Processing Time (ms)	Accuracy
	R, B R, B	DA	1.22	100%
	R, G R, R	DD	1.23	100%
	B, G R, B	CB	1.22	100%
	G, G R, B	CE	1.22	100%
	G, G B, B	LC	1.23	100%

(R: Red; G: Green; B: Blue)

Table II illustrates the results of the five different colored tags recognized by the robot. The average processing time is about 1.22 ms for the robot to recognize a tag and the accuracy rate is 100% for each kind of tags. The place symbols are on behalf of the specific places which are designated in advance by users. For instance, LC may stand for Left Corner.

2) The Accuracy of Tag Recognition in Different Moving Speed

In experiment 2, the robot passes five kinds of tags posted on the wall in succession at different speeds. Each test at specific speed is carried out 3 times. The distance between the camera and the tag is 45 cm. The results are as Table III.

TABLE III
TAG RECOGNITION RESULTS AT DIFFERENT SPEEDS

Speed \ Tag	R, B R, B	R, G R, R	B, B R, B	G, G R, B	G, G B, B
10 cm/s	3	3	3	3	3
15 cm/s	3	3	3	3	3
20 cm/s	3	3	3	3	3
25 cm/s	3	3	3	3	3
30 cm/s	3	1	3	3	3
35 cm/s	3	2	1	3	1

(R: Red; G: Green; B: Blue)

Table III demonstrates the result of the robot passing by five different colored tags in succession at different speeds. It shows that while the speed exceeds 30 cm/s, recognition error occurs. The main reason lies in the fact that image capturing rate is only 30 fps, which is too short for a whole tag to appear in the view of field. Without a whole tag image, the recognition fails. Possible improvements are to pick another camera with higher frame rate or to increase the distance between the camera and the tag. The former will increase the cost of system and the later will lose the meaning of location-aware if the robot is far from the tag too much. Nevertheless, the speed of 25 cm/s is sufficient for most household applications. Moreover, illumination variance might cause misjudgments as well.

3) The Moving Error of Household Robot Platform

The original design is to let power supply module and battery all put in a tank of the mobile platform, with the DSP board, ARM embedded system module, and camera on the top of it. We had trouble in making the robot move straight, so we then figured out that there are greatly relations between the moving direction and the center of gravity of the platform as a whole. Accordingly, we relocate the positions of hardware devices so as to shift the center of gravity and then significantly reduce the moving errors.

The result is as follows. The moving error is up to 120cm/3.6m before adjustment and is within 15cm/3.6m afterwards, as displayed in Fig. 15. To put it another way, only 12.5 percent of the original error remains. The result shows the center of gravity of the platform have a tremendous impact for smaller mobile robots rather than larger ones.

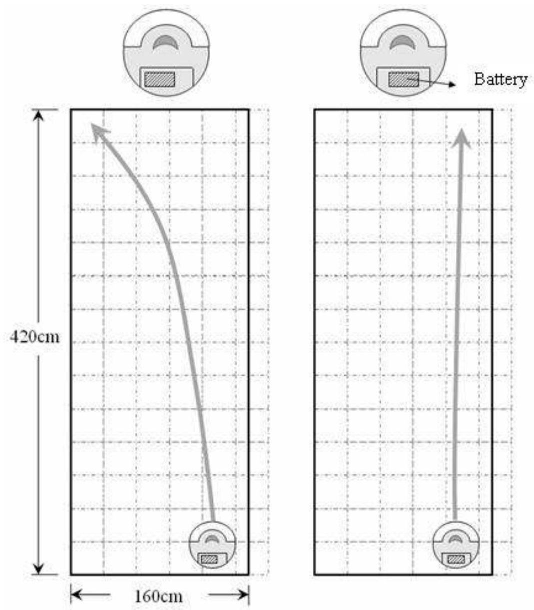


Fig. 15 moving errors with different centers of gravity

V. Conclusion

This paper proposed and carried out a new household robot system with a user interface for path planning. This system allows users to paste predefined tags at different locations according to users' comprehension of the house. The household robot is able to be aware of its location and the attributes thereof by visually recognizing the tags during the moving process. The time consumption for recognizing one tag is about 1.22ms while the moving deviation is down to 15cm/3.6m. The robot identifies all tags without missing anyone at the moving speed up to 25cm/s and therefore percept its location correctly. This household robot system successfully transfers the comprehension of the house from householders to the robot, which realizes not merely a novel interactive path planning technique, but an easy way for location awareness. The future work may focus on two aspects: 1. Estimating the distance between tags and the robot to achieve a more referable tag-based location-aware system; 2. Integrating a feedback mechanism into motion control of the robot so as to enhance the path tracking performance.

Acknowledgment

Many thanks to the patronage and supports provided by the Department of Industrial Technology, of Ministry of Economy, Taiwan.

References

[1] *WORLD ROBOTICS 2006,* European Robotics Research Network, 2006. http://www.euron.org/miscdocs/stats06.pdf

[2] S. S. Manapure , H. Darabi, V. Patel and P. Banerjee, "A Comparative Study of Radio Frequency-Based Indoor Location Sensing Systems," *in Proc. IEEE Control and Sensing Conf.*, Taiwan, 2004.

[3] M. Kim and N. Y. Chong, "RFID-based Mobile Robot Guidance to a Stationary Target," *Mechatronics*, Elsevier Science Ltd., Vol. 17, No. 4-5, 2007, pp. 217-229.

[4] A. J. Briggs, D. Scharstein, D. Braziunas, C. Dima, P.Wall, "Mobile Robot Navigation Using Self-Similar Landmarks," in *Proc. of IEEE International Conference on Robotics and Automation*, New York, 2000, pp. 1428-1434.

[5] K. J. Yoon and I. S. Kweon, "Artificial Landmark Tracking Based on the Color Histogram," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hawaii, USA, 2001, vol. 4, pp. 1918-1923

[6] H. Wang and N. Liu, "Design and Recognition of a Ring Code for AGV Localization," in *Proceedings of IEEE Pacific-Rim Conference on Multimedia (PCM)*, Singapore, Dec. 2003.

[7] D. J. Sturman and D. Zeltzer, "A Survey of Glove-Based Input", *IEEE Computer Graphics and Appl.*, 1994, pp.30-39.

[8] A. Cassinelli, S. Perrin, and M. Ishikawa, "CassinelliSmart Laser-Scanner for 3D Human-Machine Interface", *CHI Interactivity Venue*, USA, 2002, pp.1138-1139.

[9] W. Du and H. Li, "Vision Based Gesture Recognition System with Single Camera", in *Proceedings of 5th International Conference on Signal Processing (WCCC-ICSP 2000)*, Beijing, China, 2000, vol.2, pp.1351-1357.

[10] J. M. Rehg and T. Kanade, "DigitEyes: Vision-Based Hand Tracking for Human- Computer Interaction", in *Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, USA, 1994, pp.16-22.

[11] http://www.irobot.com

[12] M. Kranz, R. B. Rusu, A. Maldonado, M. Beetz, and A. Schmidt, "A Player/Stage System for Context-Aware Intelligent Environments", in *Proceedings of the System Support for Ubiquitous Computing Workshop (UbiSys 2006)*, Orange County, California, September 2006.