Construction of the robot control system which understands voice and pointing action

Yuki TAKENAKA, Norihiro ABE, Yoshihiro TABUCHI
Kyusyu Institute of Technology
680-4 Kawazu, iizuka, Fukuoka 820-8502, Japan
Email: takenaka@sein.mse.kyutech.ac.jp

Hirokazu TAKI
Wakayama University
930 Sakaedani, Wakayama-shi
Wakayama 680-8510, Japan
Email: taki@sys.wakayama-u.ac.jp

Shoujie He
VuCOMP
Richardson, Texas, USA
Email: hesj@computer.org

Abstract: This research is aiming at making the robot that can go to take an object designated by a user. We produce the robot control system that uses pointing and voice. This system is composed of two systems. One system is the object instruction system that uses pointing, another one is the object instruction system that uses voice. An approximate position of a designated object is recognized by the object instruction system that uses pointing, details of information on a designated object and an instruction operation correction are conveyed by the object instruction system that uses voice. This time, the object instruction system that uses pointing has been designed and verified as the first step. A calculation resource to calculate an approximate position of a designated object is obtained when a user directs the object with pointing. A robot is able to be moved around a designated object by using this system. The object instruction system that uses voice will be constructed in the future.

Keywords: Robotics, Image Processing, Binocular Stereo, Pointing, Voice

I. INTRODUCTION

In this research, to develop the nursing mobile robot which every senior person can easily instruct to attain his/her requirement, the robot control system is proposed which is easily instructed with use of voice and pointing action. When a user tells a robot to carry an object to the user using pointing action and voice like " take it to me", the robot must move to the object and bring it near the user.
For the purpose, the robot has to know the position of the user and the object to attain the goal. We propose the approach which makes a robot calculate the position of an object designated by a user and move to the position. When it is difficult to move a robot near the indicated object with only the first indication, information concerning the object such as color, geometry, and orientation is given with voice to the robot to modify the behavior of the robot. In this paper, the design of an indication system using pointing action and voice and the verification is shown as the first step.

II. SYSTEM CONFIGURATION

The system configuration is shown in Fig.1. The image acquired with network cameras on a robot is sent to PC using wireless LAN, and image processing calculate the position of a user and an object, then a command is sent back to the robot.

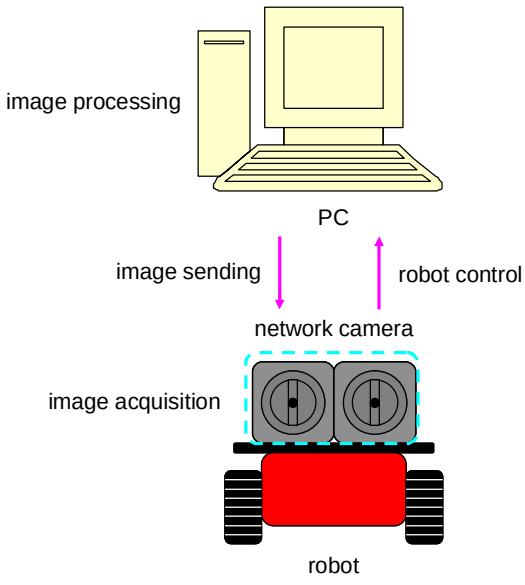


Fig.1.System configuration

III. THE OBJECT INDICATION USING POINTIG ACTION

When a user pointed at an object, it is necessary to make a robot judge where a user is pointing. As it is difficult to calculate the precise decision of position with only pointing action, approximate position of an object is estimated.

Recognizing a face and arm of a user will make it possible to calculate the location of a user and an object in terms of the robot coordination theoretically. Face recognition module realizes the face recognition of a user, and the binocular vision system based on parallax calculated from x coordinate of the center of a face in each camera image will determine the position of the user. An arm vector can be calculated referring to the two characteristic points on an arm based on the binocular vision system in the similar way.

1. Face detection

A face detecting module based on AdaBoost is used because of the speed and precision of the method. Using both features extracted from enormous number of faces saved as data and a classifier learned beforehand, it detects a face from input image.

After a face is recognized, the binocular vision calculates the three dimensional coordinates of the face using center points (two dimensions) of right and left face images.

Fig.3 shows the execution result.

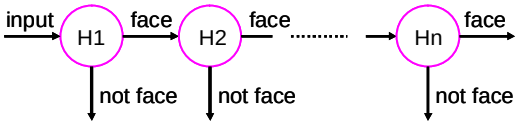


Fig.2.Face detection system

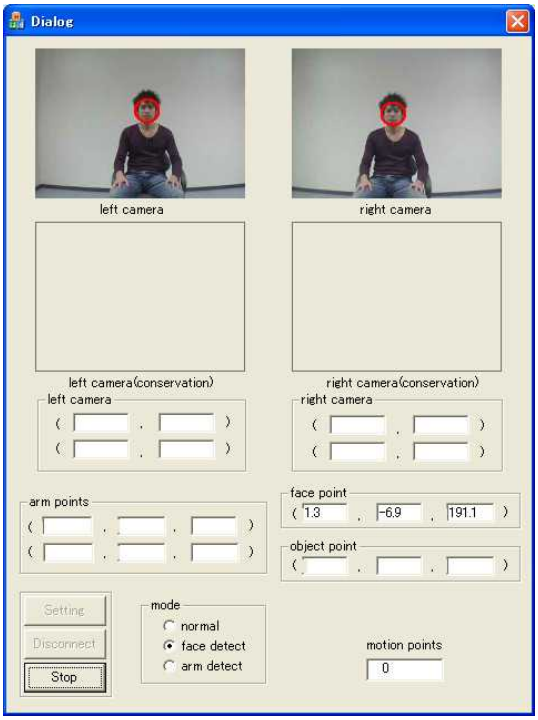
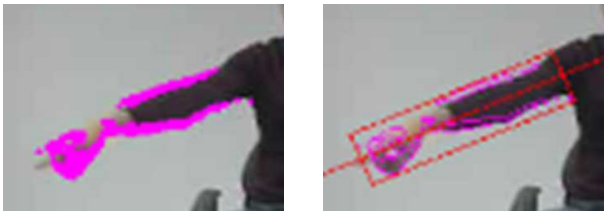


Fig.3.Excution result (Face detection)

2. Arm detection

As parts except for the arm are almost still when pointing action is performed, the arm can be detected by taking difference between consecutive two images.



(a) Difference image (b) Arm vector

Fig.4.Arm detector

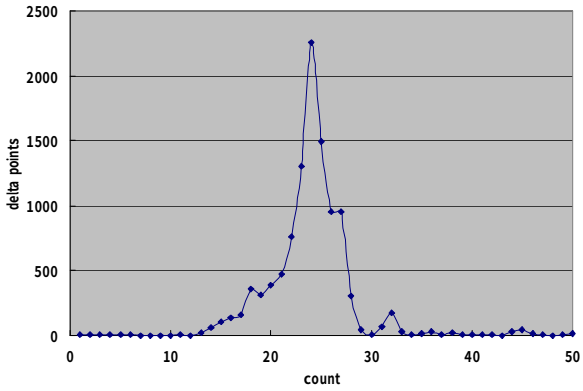


Fig.5.Change in number of difference points

The arm vector is obtained as the line connecting two middle points of two narrow sides of the rectangular which includes difference points. When the number of difference points is quite a few and the difference between consecutive numbers of difference points is negative, pointing action is regarded as completed. Calculating a cross point between the ground and the arm vector obtained when pointing action completed will give the position of the object the user designated by pointing action.

Fig.6 shows the execution result.

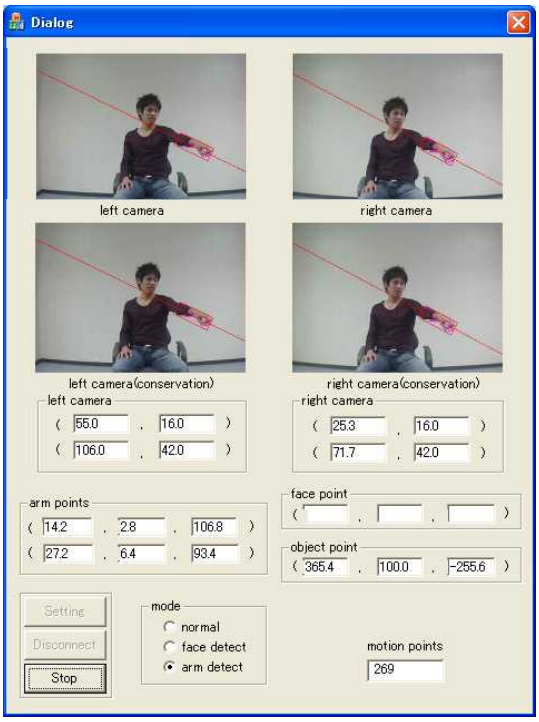


Fig.6.Excution result (Arm detection)

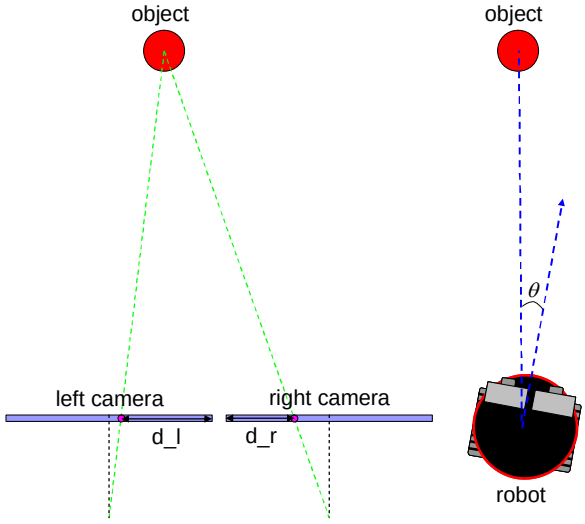
IV. OBJECT RECOGNITION

There is apparently some error in the position of the designated object and it is clear that a robot cannot precisely reach the given goal. These facts make it difficult for a robot reach the position of the object designated. Then additional information on the object must be given to the robot to make the robot modify its behavior. We believe the voice will useful to do this.

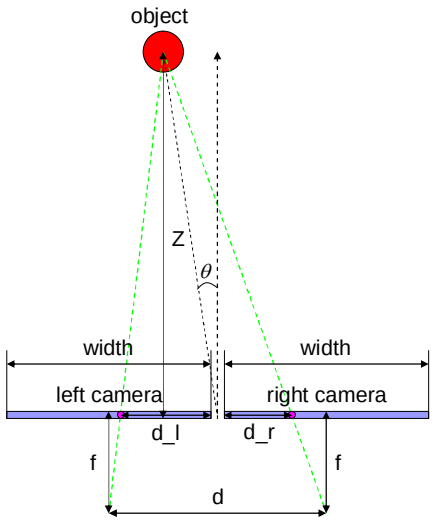
Fig.7 shows the current position of a robot and object. It is unnecessary to correct the behavior of a robot when the robot is facing toward the object. In contrast, a robot must correct it heading when the robot is facing to the right or left of the object. In Fig.7, the target object deviates left by θ from front.

Let the distance from a robot to the object be Z, the distance between cameras d, the Focal distance be f and the width of an image plane be width, then the θ is calculated by the following expression;

$$\theta = \tan^{-1}\left(\frac{width}{2f} - \frac{d\_r}{f} - \frac{d}{2Z}\right) \quad \cdots(1)$$



(a) Physical relationship



(b) Variable definition

Fig.7.Object recognition (d_l>d_r)

Let the position of the object be P, the position of the object in left camera image be PL, the position of the object in right camera image be PR, the distance between cameras d and the Focal distance be f as shown in Fig.8, then P is calculated by the following expression;

$$\begin{cases} X = \dfrac{dXL}{XL - XR} & \cdots(2) \\ Y = \dfrac{dYL}{XL - XR} & \cdots(3) \\ Z = \left(\dfrac{d}{XL - XR} - 1\right) f & \cdots(4) \end{cases}$$

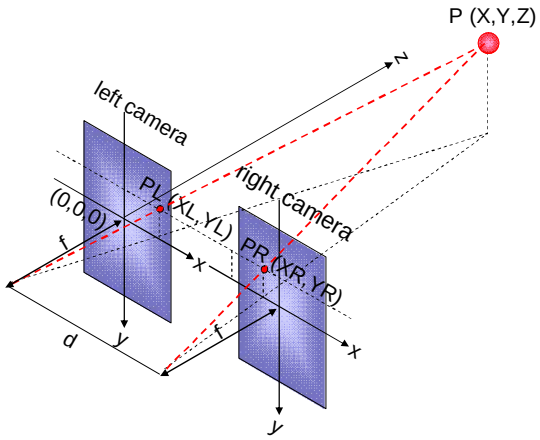


Fig.8.Binocular stereo

According to the deviation from the proper direction, if the behavior of the robot must be corrected, then it will be led to near the position of the designated object.

Fig.9 shows the execution result.

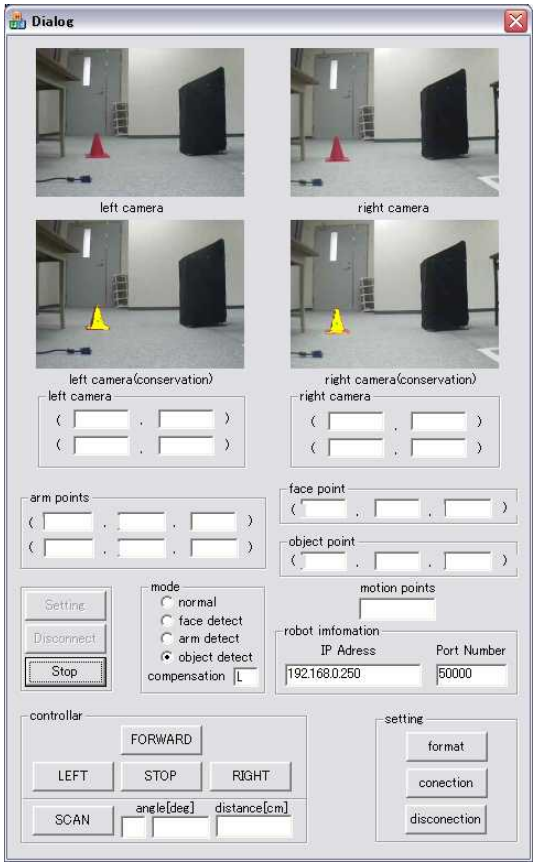


Fig.9.Excution result (Object recognition)

V. CONCLUSION

It is considered possible to calculate the position of the object which a user designates by pointing action. It is confirmed that the proposed method successfully calculates both the user's position by recognizing user's face and the position of the designated object by detecting the user's arm. Calculation results contain errors, but they are considered to be within a tolerance as the behavior of a robot will be corrected with additional voice command from the user when it tries to attain the task given by the user.

The design and implementation of an object indication system using pointing action are generally seemed to be over; the promotion of efficiency of system is attempted while testing it. The validity of the object indication system when the robot is controlled using pointing action, how much correction of robot behavior by voice interruption is necessary must be estimated to aim at developing the robot controlling system using voice and pointing action.

ACKNOWLEDGMENTS

We greatly appreciate the aid of the Grant-in-Aid for Scientific Research (S) and (A).

REFERENCES

[1] M. Inaba, S. Kagami, F. Kanehiro, Y.Hoshino, H. Inoue, Remote-Brained Robot Approach，International Journal of Robotics Research, Vol.19, No.10, pp.933-954, 2000

[2] Yasuhiro Watanabe, Norihiro Abe, Kazuaki Tanaka, Hirokazu Taki, Tetsuya Yagi, Multimodal communication system allowing man and avatar to use voice and beck, 3rd International Conference on Information Technology and Applications(ICITA'2005), pp.161-166

[3] Hiroki Goto, Norihiro Abe, Kazuaki Tanaka, Yoshihiro Tabuchi, Hirokazu Taki, Shoujie He: Usage of Shared Memory for Multimodal Data Processing in an Outdoor Mobile Robot, proc. ISR (36th International Symposium on Robotics) (CD-ROM), 2005

[4] Moez Bellamine, Norihiro Abe, Kazuaki Tanaka, Hirokazu Taki :Remote Diagnosis System for Rotating Machinery Using Virtual Reality, IEICE Transactions,pp.895-903,2005

[5] Toshiaki Oomori, Norihiro Abe, Kazuaki Tanaka, Hirokazu TAKI, Shoujie HE: Concurrent development of Virtual Robots and Real Robots Based on Physical Low, IMEKO / IFAC / IFIP Workshop on Advanced Robot Systems and Virtual Reality.(CD-ROM), . 2005

[6] Norihiro Abe, Yoshihiro Tabuchi, Hirokazu Taki, Shoujie He " Synchronization between Audiovisual and Haptic Feeling for Constructing Edutainment Systems", international journal(Journal of Virtual Reality, accepted, 2006