Construction of the robot control system with use of pointing action and voice

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 action and voice. This system is composed of two systems. One system is the object instruction system that uses pointing action, 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. A robot is able to be moved to a designated object by using this system.

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

I. INTRODUCTION

Recently, robots are expected to help people at a care facility because of labor scarcity. 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 which is easily instructed with use of pointing action and voice is proposed. 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. The method calculates the location of an object designated by a user from the images which are acquired from network cameras installed on the robot. A robot figures out command from a user and object information with a voice recognition system. We control a robot with the pointing information and the voice information.

Not only elderly people but a lot of users can operate a robot easily with this system because it uses easy interface such as pointing action and voice.

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 calculates the position of a designed object. User's voice acquired with a headset which a user wears is sent to PC wirelessly, and the user's instruction is recognized, then a command corresponding to the given instruction is sent back to the robot.

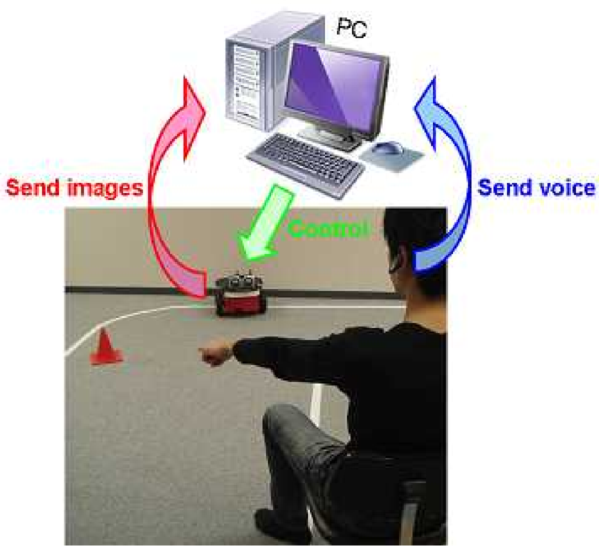


Fig.1. System configuration

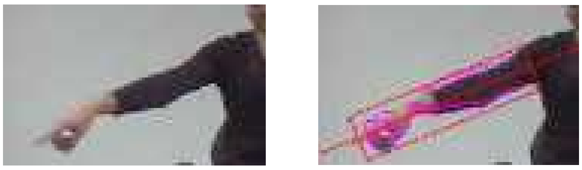
III. THE OBJECT INDICATION USING POINTIG ACTION AND VOICE

When a user designates an object, a robot has to know the position of the object. It is not easy for a robot to calculate the precise location but the approximate one can be determined by referring to the direction of the user's hand if a little error is permitted. We propose the approach which makes a robot calculate the position of the object designated by the user and move to the position. There is, however, the possibility that a user erroneously designates a target by pointing action. When there are similar objects near the target one, the robot may not always succeeds in identifying the target. In the case, as a robot will probably move toward the different object from what the user intends, the user must give an advice using voice to the robot in order to make the robot change its behavior. User's voice captured with a headset is sent with a wireless LAN to the PC which is processing image data sent from a robot with another wireless LAN, and then the voice signals are translated into Japanese character strings with a voice recognition system. The strings are next segmented by referring to grammar rules and a dictionary including words registered beforehand into Japanese word sequence. A set of words available at present includes name, shape or color of objects, operation added to objects and demonstrative pronouns.

1. Image processing system

1.1. Get pointing information

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. The arm vector is obtained as the line connecting two middle points of two narrow sides of the minimum rectangular which includes difference points.



(a) Acquired image (b) Arm vector

Fig.2. Arm detector

Based on the binocular vision system, an arm vector can be obtained from right and left camera images as shown in Fig.2. It is possible to calculate in some measure the position of the designated object from the arm vector.

The binocular vision system is the method which has the position of an object in images calculated from right and left camera images by the use of the triangulation. 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 be d and the Focal distance be f as shown in Fig.3, then P is calculated by the following expression;

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

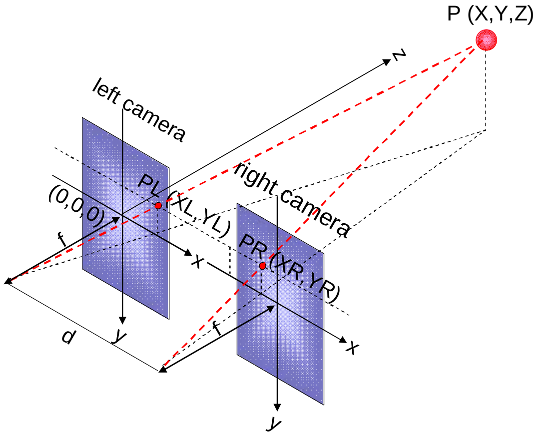


Fig.3. Binocular stereo

When the number of difference points is quite a few and the difference between consecutive numbers of difference points is negative as shown in Fig.4, 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 approximate position of the object designated by pointing action.

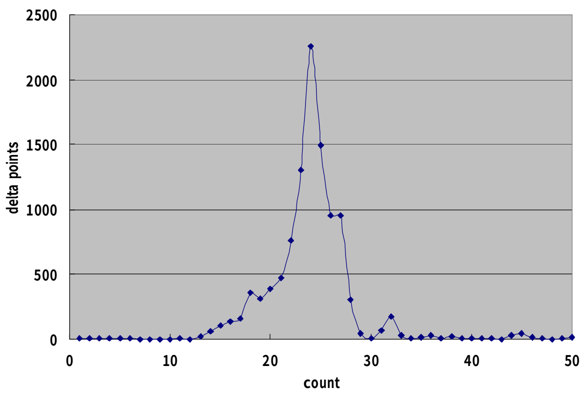


Fig.4. Change in number of difference points

When pointing action comes to an end, pointing information is obtained, and the robot will understand where to go. Fig.5 shows the execution result.

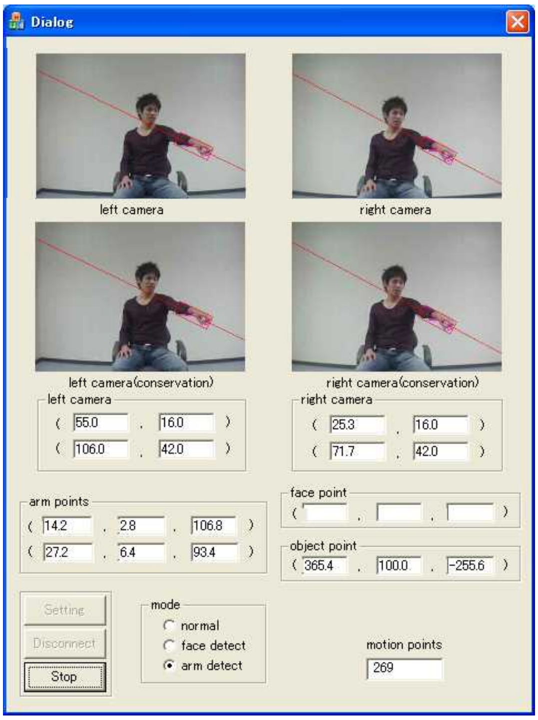


Fig.5. Excution result (Arm detection)

1.2. Motion correction by object correction

There is apparently some error in the calculated 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 to reach the position of the designated object. Then additional information of the object must be given to the robot when a user finds the robot moving toward wrong place.

In the case, the user will utter words or phrases like 'stop' or 'go rightward' to interrupt the motion of the robot to make it notice its wrong behavior. When the robot is interrupted its action, it will stop moving and try to examine the user' utterance to decide what to do next. In the former case, it will need conversation with the user to know why the user interrupted its action, but this needs much complicate discourse analysis. On the other hand, in the latter case, it has only to change its posture using the method shown below.

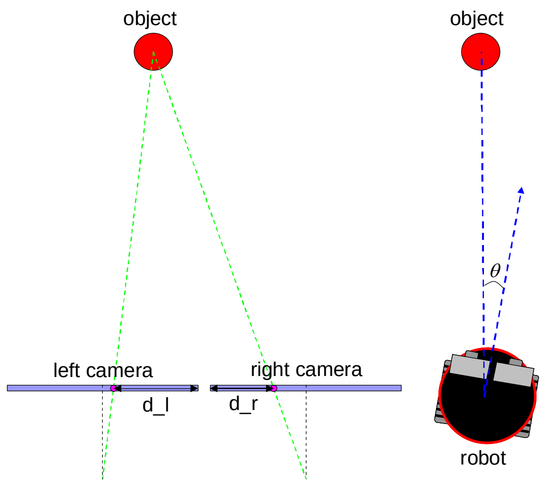
As shown in Fig.6, the object location is recognized using right and left camera images, each position of the center of gravity is calculated and compared to judge the direction which the robot goes in. It is unnecessary to correct the behavior of a robot when the robot is facing toward the object. In 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.

(a) Acquired image (b) Object recognition

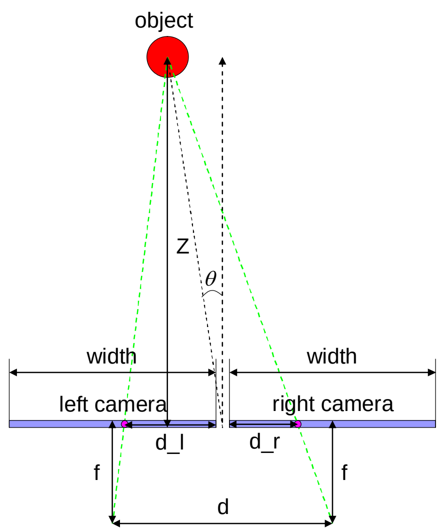
Fig.6. Object detector

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

$$\theta = \tan^{-1}\left(\frac{width}{2f} - \frac{d\_r}{f} - \frac{d}{2Z}\right) \quad \cdots(4)$$



(a) Physical relationship



(b) Variable definition

Fig.7. Object recognition (d_l > d_r)

If the behavior of the robot must be corrected, then it will be led to near the position of the designated object. Fig.8 shows the execution result.

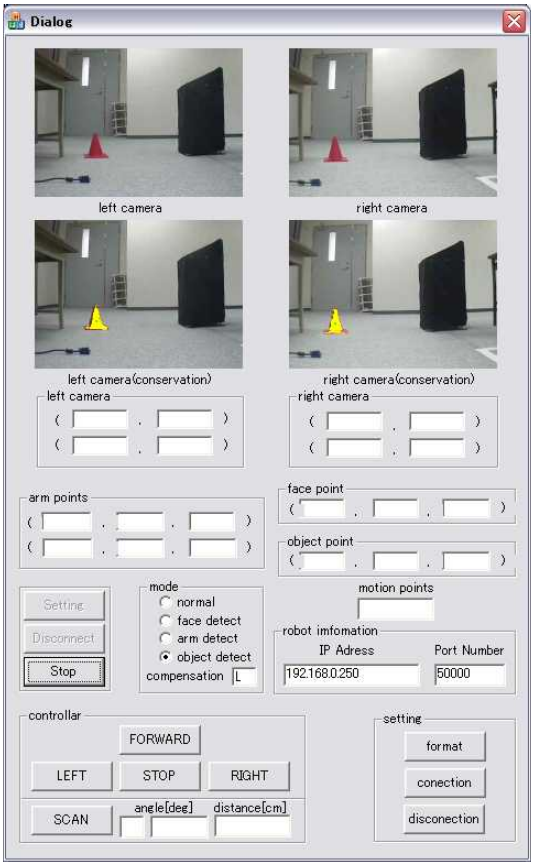


Fig.8.Excution result (Object recognition)

2. Voice recognition

We need two systems to communicate between a user and a robot. One is the system which enables a robot to recognize user's voice; the other is that which permits it to talk with a user. The former method uses Dragon Naturally Speaking 2005, the latter method uses Aques Talk. Fig.9 shows the execution result.

Fig.9.Excution result (Voice recognition)

IV. CONCLUSION

It is considered possible to calculate the position of an object which a user designates by pointing action an d voice. It is confirmed that the proposed method successfully makes a robot find the position of the designated object from pointing and voice information. Calculation results can contain errors, but they are considered to be within a tolerance as the behavior of a robot is corrected with object recognition and additional voice command from the user when it attains the task given by the user.

The design and implementation of an object indication system using pointing action and voice recognition system are nearly complete, but it must be verify whether a robot successfully attain the task in more complicated situation such as the case there are obstacles between the robot and object.

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-
Fig.8.Excution result (Object recognition)
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