Control of Mobile Robot Equipped with Stereo Camera Using the Number of Fingers

Masato Kondo[1] Atsushi Yamashita[2] Toru Kaneko[1] Yuichi Kobayashi[1]
[1]Department of Mechanical Engineering, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka, Japan
[2]Department of Precision Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

***Abstract*—In recent years, introduction of autonomous mobile robots which support humans in environments close to our life is expected. Examples are shopping support robots and porter robots carrying the luggage while accompanying the person. To utilize these robots, real time motion control of the robots is important. In this paper, we propose a mobile robot system which performs the actions by recognizing the number of fingers. The robot recognizes the number of fingers from images obtained with a stereo camera system. Also, as one of the actions, the mobile robot follows human who wears a mark. We verified the validity of the system by using a wheel mobile robot in indoor environment.**

***Keywords*: Mobile Robot, Man-Machine Interface, Gesture Recognition, Human Following, Stereo Camera.**

I. Introduction

In recent years, autonomous mobile robots are desired to support humans in environments close to our life. Examples are shopping support robots [1][2] and porter robots [3] carrying the luggage while accompanying the user. To utilize these robots, it is required to control the robots in real time.

There are studies of controlling a robot using the portable device such as a remote controller [4] and a mobile phone [5]. Although it is possible to reliably the control robot in these methods, the user must hold a device. For the application to the shopping support robot, it is desirable to be hands-free while user is shopping.

On the other hand, there are studies of controlling the robot using a hand gesture. A hand gesture means the shape and movement of the hand, and has been investigated as an recognition problem [6]. Also, the hand gesture has been studied for controlling personal computers and consumer electronics in the intelligent room [7]. These methods usually recognize hand gestures from the images acquired by camera. In most of previous studies of recognition of hand gesture, the moving images are acquired by fixed camera. There are few studies of the recognition of hand gesture from the camera on move. But, it is effective to use a hand gesture which is intuitive and easy for users. By using hand gesture, the user does not need to carry a remote controller or the like and can control the robots more naturally.

There are studies of controlling the mobile robot using the hand gestures. In the method of Hu et al., a hand gesture is detected by analyzing the contour of the hand [8]. Also, in the method of Luo et al., the coordinates system called HFLC (Human-Following Local Coordinate System) is constructed with the origin at the center of gravity [9]. As a result, the trajectory of hand movement becomes simple, is to be extracted fast and stably. However, in these studies, the actual control of robot is not experimented. So, the effectiveness of these methods have not sufficiently verified.

In the method of Loper et al, a hand motion is recognized by searching the silhouette of a person in the depth image [10]. If the user controls the mobile robot by movement of the hand, the more actions can be represented compared to the shapes of hand. However, there are individual differences in movement of the hand. Therefore, it is necessary to register the movements of the hand of the user in advance in order to improve the recognition accuracy, and it is time-consuming.

In this study, we assume a shopping assistance robot. The user controls the robot by the number of fingers as a hand gesture, which is not so much influenced by individual differences of user and movement of robot. In this method, the user does not need to carry a remote controller or the like and can control the robots more naturally. The robot is equipped with a stereo camera for vision and recognizes the number of fingers from acquired images while moving.

II. Outline of processing

In this study, we use a wheeled mobile robot shown in Fig. 1. The robot is equipped with two cameras constructing a parallel stereo system and obtains images in front of the robot. The robot operates in an indoor environment with no steps on the floor. The user wears a clothes whose color is not close to the skin tone, and a mark is equipped on the back of the user. The user shows the hand at the height of the chest in front of the robot when controlling the robot. In the flow of the process first, the image in front of the robot is obtained by the stereo camera. Next, the number of fingers is recognized using the skin color information from the obtained image. Then, the robot performs the action corresponding to the recognized number. By repeating these processes, it is possible to change the actions while the robot is moving.

In this study, we assume a shopping assistance robot. Therefore, as the actions required to transporting the luggage while following a user, we set the actions shown in table I. The robot performs human following when it recognizes one finger. Similarly, the robot turns 180 degrees for three and steps back for five. In human following, a mark attached to the user is detected by template matching and the robot follows the user while keeping certain distance.

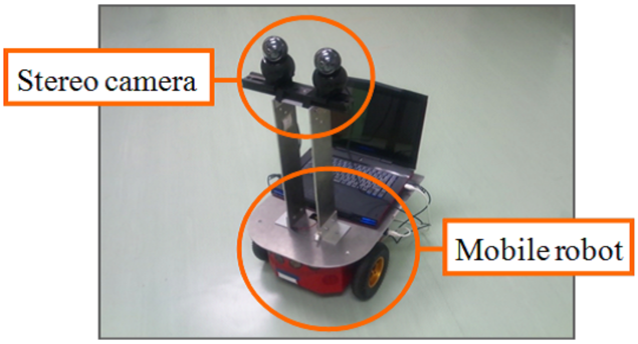


Fig. 1. Mobile robot equipped with stereo camera.

TABLE I
NUMBER OF FINGERS AND ACTIONS.

Number of fingers	Actions
0	Stop
1	Human following
3	180 degrees turn
5	Stepping back

III. RECOGNITION OF THE NUMEBER OF FINGERS

Recognition of the number of fingers is executed for each frame of acquired images. In the recognition process first, the skin color region is extracted first from the left image obtained by the stereo camera. Next, the hand region is extracted from the skin color region and its features are obtained. Finally, the number of fingers is recognized by using these features. These process is performed to the upper part of the acquired image, because the user shows a hand in front of chest.

A. Extraction of Skin Region

The color system of the acquired left image (Fig. 2) is converted from RGB system to HSV system .

Because the color is expressed by hue (H), saturation (S), and lightness (V) in the HSV color system, it is more robust against changing in brightness than RGB color system is. After converting to the HSV color system, pixel values of H, S, and V are examined whether they are within the predetermined ranges of skin color. The extracted skin region is shown in Fig. 3. where skin regions are shown in white and the rest is shown in black.

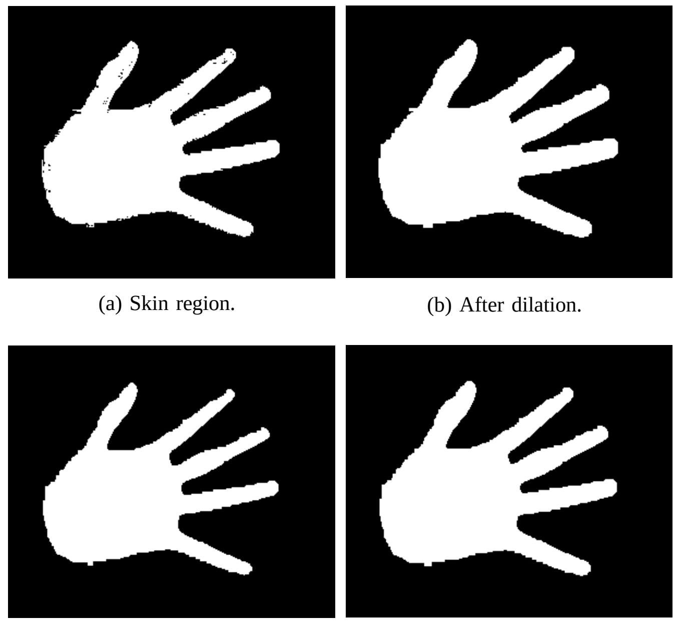
B. Extraction of Hand Region

The extracted skin region (Fig. 3) contains noises such as black pixels in a white area and white pixels in black area. In order to remove such noises, erosion and dilation are applied at the same number of times to the skin region (Fig. 4).

Next, the hand region is extracted from the skin region. Labeling is applied to the skin regions to divide them into connected regions. Then, the connected region that has the maximum area is extracted as a hand region. If the maximum area of the connected region is less a the threshold, the processing is terminated.

Fig. 2. Acquired image.

Fig. 3. Extracted skin region.



(a) Skin region. (b) After dilation.

(c) After erosion. (d) Hand region.

Fig. 4. Process of noise rejection.

There may be small holes still inside the extracted hand region. Therefore, the contour of the hand region is detected, and its inside is filled. The extracted hand region is shown in Fig. 5.

C. Recognition of Number of Fingers

Since the size of the extracted hand region varies according to the distance from the robot and the number of fingers, the area of the finger region also varies. Therefore, a convex hull and a bounding box of the hand region (Fig. 6) is determined and the area ratios of the convex hull to the bounding box of the hand region are calculated. Because these ratios are almost constant irrespective of the distance from the camera, these can be used as features invariant to the distance. The number of times for processing of erosion and dilation is determined from the area ratios. Also, area threshold of the finger region is determined. If the two ratios were less than the thresholds, the process is aborted.

Finally, erosion and dilation are applied to the hand region (Fig. 7 (a)). By repeating dilation more than erosion, a region can be obtained reliably. Fig. 7 (b) shows the region after erosion and Fig. 7 (c) shows the region after dilation. Then, the finger region is obtained by subtracting Fig. 7 (c) from Fig. 7 (a). At last, the number of fingers is recognized by counting the finger regions.

Fig. 5. Extracted hand region.

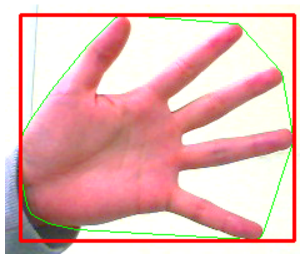


Fig. 6. Areas.

(a) Hand.

(b) After erosion.

(c) After dilation.

(d) Finger region.

Fig. 7. Finger region extraction.

IV. Actions of Robot

If the recognition result of number of fingers is equal to the one shown in Table 1, the robot performs an action corresponding to that number. Then, the robot continues the action until a new action is recognized. However, for safety, the robot steps back only when recognizing five. When the action is changed, the robot outputs the voice corresponding to the action, so that the user can notice it.

V. Human Following

In this study, the mobile robot follows human by detecting a mark attached to the user. The user wears the mark consisting of squares of four colors as shown in Fig. 8. A combination of these selected colors rarely exist in real space and is less affected by brightness variation. Thus, the mark can be detected stably in the acquired image by template matching.

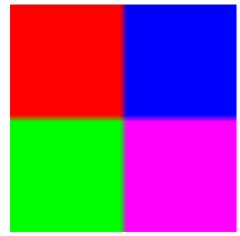


Fig. 8. Mark

Template matching is applied to left and right images obtained by the stereo camera, and positions of the mark are detected. Next, the distance and angle to the user are calculated from the positions of the mark. Finally, the robot follows the user according to the distance and angle.

A. Detection of the mark

A template image is scanned in raster on the left image acquired by the stereo camera and the dissimilarity is calculated. SAD (Sum of Absolute Difference) is used to calculate the dissimilarity. An image size of template is $M \times N$. The coordinates on the acquired image are defined as (x, y) and the coordinates on the template image are defined as (i, j). In this case, the dissimilarity R_{SAD} is calculated by the following equation

$$R_{SAD} = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} |I(x+i, y+j) - T(i, j)|, \qquad (1)$$

where $T(i, j)$ represents the pixel value of the temple image and $I(x+i, y+j)$ represents the pixel value of the input image overlapped by the template image.

In this study, in order to perform template matching faster, the dissimilarity R_{SAD1} is calculated about the hues of the pixels in the four corners of the template image. Then, R_{SAD1} is compared with a threshold value. If R_{SAD1} is below the threshold, dissimilarity R_{SAD} is calculated about the hues of the entire pixels on the template image. The coordinates where R_{SAD} is minimum are detected as the position of the mark on the left image. After performing the template matching to the left image, the template matching is performed similarly on the right image along the epipolar line.

B. Motion of the robot during human following

Let the origin of the coordinate system of the robot be located in the center between left and right camera. In the coordinate system of the robot, the distance and the angle to the user are calculated by using positions of mark detected from the stereo image. The moving speed of the robot changes depending on the distance and angle.

Let the rotational speeds of the right and left wheels be V_R and V_L, respectively. The moving speed of the robot is set according to the distance Z m from the user and angle θ degree. If Z is more than 1 m, the rotational speed of each wheel is as shown in (2). In addition, if Z is less than 1 m and θ is less than -10 degrees, the robot turns around θ-5 degrees. If Z is less than 1m and θ is more than 10 degrees, the robot turns around θ+5 degrees.

$$\begin{cases} V_L = Z + 50 \\ V_R = Z - 50 \end{cases} \quad \text{if} \quad \theta \geq 10 \tag{2}$$

$$\begin{cases} V_L = Z - 50 \\ V_R = Z + 50 \end{cases} \quad \text{if} \quad \theta \leq -10 \tag{3}$$

$$\begin{cases} V_L = Z \\ V_R = Z \end{cases} \quad \text{if} \; -10 \; < \; \theta \; < \; 10 \tag{4}$$

VI. Experiment

In experiment, we evaluated the recognition accuracy of the number of fingers. We also evaluated controllability of the mobile robot by the number of fingers.

A. Experiment Devises

We used a wheeled mobile robot, ActivMedia Robotics Pioneer2. This robot has a mechanism with two drive wheels and a caster. The camera is Logicool Qcam Orbit AF. The computer has CPU of Intel Core i7 X920 2.0GHz and the memory of 8.0GB.

B. Experiment Conditions

The appearance of experimental environment is shown in Fig. 9. The experiment was conducted in an indoor environment with no steps on the floor. The size of the acquired image size was 640 × 480 pixels, and the size of template image was 11 × 11 pixels. Also, the size of the mark attached to the user was 50 × 50 mm. Fig. 10 shows the appearance of the mark.

In evaluation of the recognition accuracy of the number of fingers, the user directed three types of robot control. Assuming the use of real robot, the distance between robot and hand is set to 0.5, 1.0 and 1.5 m. The images were obtained 200 frames per each type and distance. The recognition of the number of fingers was applied to each acquired image. Then, the recognition result of number was confirmed whether it was equal to the number presented by a user.

Fig. 9. Environment.

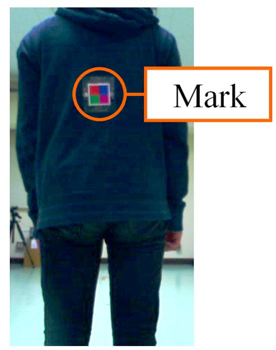
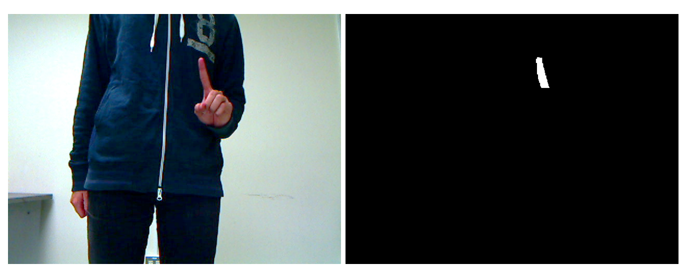


Fig. 10. Mark attaching.

In evaluation of controllability of the mobile robot, the user controlled the robot to follow. After that, the user directed the robot in order of 180 degrees turn, human following, stepping back, and stop.

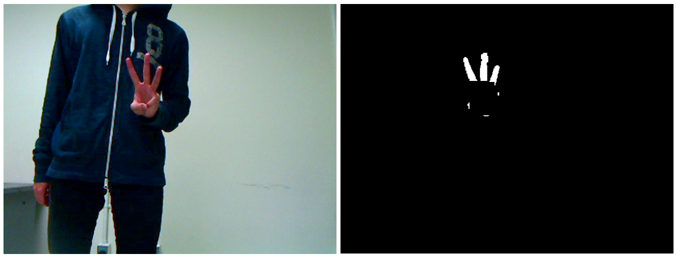
C. Experiment Results

The acquired image and the extracted finger region in each distance are shown in Fig. 11- 13. The left hand was extracted only because the finger recognition process was applied to the height of chest. In each distance, the region of fingers was extracted by our method.



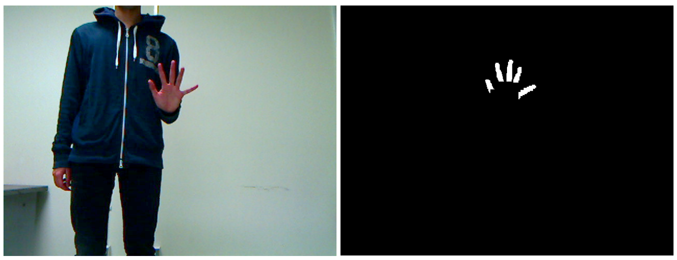
(a) Acquired image. (b) Region of Fingers.

Fig. 11. Appearance in the distance of 50cm.



(a) Acquired image. (b) Region of Fingers.

Fig. 12. Appearance in the distance of 100cm.



(a) Acquired image. (b) Region of Fingers.

Fig. 13. Appearance in the distance of 150cm.

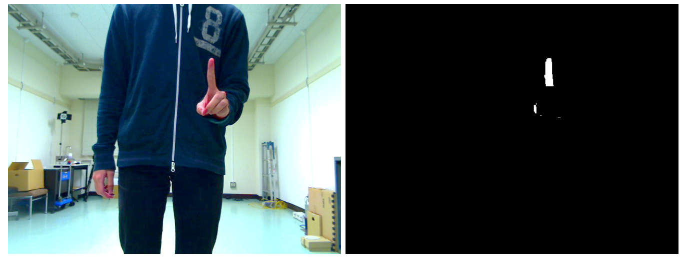
TABLE II
RECOGNITION ACCURACY OF THE NUMBER OF FINGERS.

Number of fingers and actions	Success rate[%]		
	50cm	100cm	150cm
1 (Human following)	94.0	95.5	97.0
3 (180 degrees turn)	92.0	98.5	94.0
5 (Stepping back)	95.0	97.0	98.0

The recognition results of fingers are shown in Table II. Table II shows that the number of fingers was recognized stably. However color variation due to the influence of image blur and lighting was conceivable as cause of false recognition where thresholding did not work well for skin region extraction. In addition, the recognition accuracy was low in 0.5 m. When the distance from the robot was short, the change in the area of the hand region is increased. At this time, there was a case of insufficient repetition number of erosion and dilation. On the other hand, the robot followed the user at distance more than 1m, so the low recognition accuracy due to the closeness of distance is not a problem. Based upon the evaluation of recognition accuracy, the action recognized in three consecutive frames is only performed to prevent malfunction.

The images and the extracted finger regions obtained when action changed are shown in Fig. 14 -18. The experiment showed that the mobile robot recognized the number of fingers stably and performed the corresponding actions.

Images acquired at the human following are shown in Fig. 19. The position of the mark was detected in each of the left and right images.

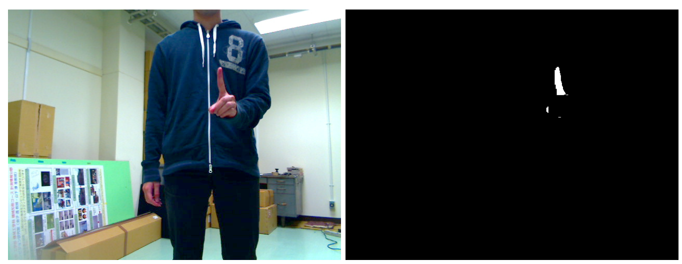


(a) Acquired image. (b) Region of Fingers.

Fig. 14. 1st movement: Human following.

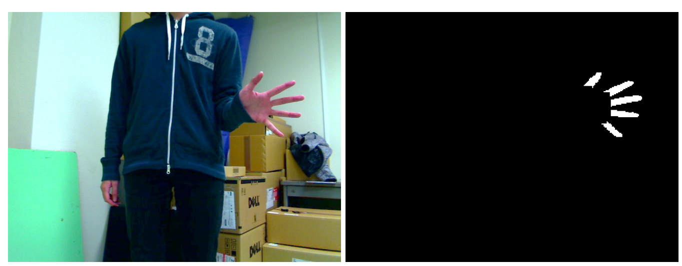
(a) Acquired image. (b) Region of Fingers.

Fig. 15. 2nd movement: 180 degrees turn.



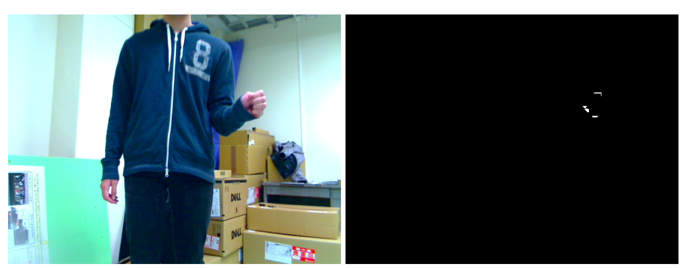
(a) Acquired image. (b) Region of Fingers.

Fig. 16. 3rd movement: Human following.



(a) Acquired image. (b) Region of Fingers.

Fig. 17. 4th movement: Stepping back.



(a) Acquired image. (b) Region of Fingers.

Fig. 18. 5th movement: Stop.

(a) left image. (b) Right image.

Fig. 19. Mark detection in human following.

VII. Conclusion

In this study, we constructed a robot system which recognizes the number of fingers directed by the user and performs the corresponding actions. By using the number of fingers, the user is able to direct the actions accurately to the robot even while it is moving. The robot performs the action which is recognized for three consecutive frames in order to prevent malfunction. In addition, the robot follows the human stably by detecting the proposed template attached to the user. Experimental results showed the validity of the method.

As a future work, we should make the processing of skin color detection more robust because the success rate of fingers number recognition depends on the skin detection result. Further, we should make our system applicable to more complicated scenes, and enhance the ability of hand gestures directing the robot to work in various situation.

References

[1] Fumio Ozaki, Seiji Tokura, Takafumi Sonoura, Tsuyoshi Tasaki, Hideki Ogawa, Masahito Sano, Akiko Numata, Naohisa Hashimoto, and Kiyoshi Komoriya: “Development of Robotic Transportation System - Shopping Support System Collaborating with Environmental Cameras and Mobile Robots -”, *Proceedings of the 2010 41st International Symposium on Robotics and 6th German Conference on Robotics*, pp.1-6, 2010.

[2] Christian Bogdan, Anders Green, Helga Huettenrauch, Minna Rasanen, and Kerstin Severinson Eklundh: “Cooperative Design of a Robotic Shopping Trolley”, *Proceedings of European Cooperation in the Field of Scientific and Technical Research* 2009.

[3] Tsutomu Hasegawa, Kouji Murakami, Ryo Kurazume, Yosuke Senta, Yoshihiko Kimuro, and Takafumi Ienaga: “Robot Town Project: Sensory Data Management and Interaction with Robot of Intelligent Environment for Daily Life”, *Proceedings of the 4th International Conference on Ubiquitous Robots and Ambient Intelligence*, pp.369-373, 2007.

[4] Shandong Wu, Yimin Chen: “Remote Robot Control Using Intelligent Hand-held Devices ”, *Proceedings of the 4th International Conference on Computer and Information Technology*, pp.587-592, 2004.

[5] Myeongjin Song, Baul Kim, Yunji Ryu, Yongdug Kim, and Sangwook Kim: “A Design of Real Time Robot Control System based on Android Smartphone”, *Proceedings of the 7th International Conference on Ubiquitous Robots and Ambient Intelligence*, pp.584-586, 2010.

[6] Michael Van den Bergh, Luc Van Gool: “Combining RGB and ToF Camera for Real-time 3D Hand Gesture Interaction”, *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pp.66-72, 2011.

[7] Kota Irie, Naohiro Wakamura, Kazumori Umeda: “Construction of an Intelligent Room Based on Gesture Recognition”, *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.193-198, 2004.

[8] Chao Hu, Max Qinghu Meng, Peter Xiaoping Liu, and Xiang Wang: “Visual Gesture Recognition for Human-machine Interface of Robot Teleoperation”, *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol.2, pp.1560-1565, 2003.

[9] Dan Luo, Jun Ohya: “Study on Human Gesture Recognition from Moving Camera Images”, *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo*, pp.274-279, 2010.

[10] Mathew M. Loper, Nathan P. Koenig, Sonia H. Chernova, Chris V. Jones, and Odest C. Jenkins: “Mobile Human-robot Teaming with Environmental Tolerance”, *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*, pp.157-164, 2009.