Mecatronics-2014-Tokyo
November 27-29, 2014, Tokyo, Japan

Human Following with a Mobile Robot Based on Combination of Disparity and Color Images

Yuzuka Isobe
Course of Precision Engineering
School of Science and Engineering
Chuo University
Tokyo 112-8551, Japan
Email: isobe@sensor.mech.chuo-u.ac.jp

Gakuto Masuyama
Dept. of Precision Mechanics
Faculty of Science and Engineering
Chuo University
Tokyo 112-8551, Japan
Email: masuyama@mech.chuo-u.ac.jp

Kazunori Umeda
Dept. of Precision Mechanics
Faculty of Science and Engineering
Chuo University
Tokyo 112-8551, Japan
Email: umeda@mech.chuo-u.ac.jp

***Abstract*—This paper presents a target tracking system for a mobile robot in both indoor and outdoor environments. A stereo camera, which is robust to sunlight and illumination changes, is used. Human regions in images are detected from 3D information. Using color information, a target region is discriminated from the detected human regions. Hue and saturation are chosen as features robust to illumination changes. Finally, a mobile robot is controlled based on the 3D information of the detected target region. The effectiveness of the proposed system is verified through human following experiments in both indoor and outdoor environments.**

I. Introduction

Mobile robot systems are expected to be applied in many situations and environments -such as factories, construction sites, and human living spaces [1]. In such real environments, one of the important functions of a mobile robot system is to track a specific person. In order to carry out this function, robots are required to have the ability to recognize the existence of people.

In real and dynamic environments, it is important for robots to have sufficient perceptual capabilities. There have been many works using various sensors for human detection and tracking[2], [3]. Recently, in the area of target tracking, the vision-based method has become mainstream [4], [5], [6]. In [7], Doi *et al.* presented a visual tracking system for a robot using the color information of a target's clothing and skin. In opposition to this color-based method, Satake *et al.* [8] and Chen *et al.* [9] developed tracking methods using distance information as a feature of a target. Most of these methods, however, can only be used in indoor environments because of the lighting conditions (sunshine and shadow) outdoors.

The high illumination of sunshine limits the capabilities of available sensors. Moreover, in sunlight, adjusting back lighting is more necessary than under indoor lights. In addition, sunlight causes objects to have shadows. The darkness of the shadow is changed by how the sun shines against objects and the cloudiness of the skies. However, even under such conditions, target tracking must be carried out. Takemura *et al.* [10] proposed a tracking system that could be applied both indoors and outdoors. Yet this can be difficult, due to the significant adjustment required to accommodate many changes in illumination. One reason is that the process of human detection requires information about the target's color, which is easily affected by illumination changes. When tracking a targeted person in unstructured outdoor environments, the changing light condition is a major problem.

In this paper, we propose a target tracking system using a stereo camera in various illumination environments. A stereo camera can be used in sunlight and, because it is a passive sensor, measurement of human can be taken safely. In addition, the stereo camera has the advantage of capturing both disparity and color images simultaneously.

The rest of the paper is organized as follows. Section II explains the algorithm of our method. Then, Section III presents a human following experiment both indoors and outdoors, and the result is evaluated by investigating how closely it tracks correctly without missing. Finally, conclusions and future works are shown in Section IV.

II. Target Tracking Algorithm

A. System overview

Our system consists of a stereo camera on a mobile robot. The flow of the system is described in Fig. 1. First, disparity and color images are captured using a stereo camera, which means that distance and color information is obtained. By using 3D information that is robust to sunlight and illumination changes, people are detected. Once objects are identified as people, the hue and saturation information of each person is compared with that of a target that has previously been registered. When a person is judged to be a target, a mobile robot moves according to the location information of the target. Repeating these processes, continuous tracking is achieved.

B. Segmentation of objects using a disparity image

In order to extract the object regions by using 3D information, the segmentation method proposed by Ubukata *et al.* [11] is adopted. A disparity image obtained from a stereo camera gives 3D information of the respective pixels. The information is projected onto an overlooked plane called a *projection plane*. Fig. 2(a) is an example of an input image and a *projection plane* obtained from the image (Fig. 2(b)). Then, the plane is divided into cells, and the number of the projected points of each cell is calculated. A histogram of the points in each cell is shown in Fig. 2(c). In the image, the color of the cells depends on the number of points; the more points a cell includes, the

978-1-4799-5717-0/14/$31.00 ©2014 IEEE

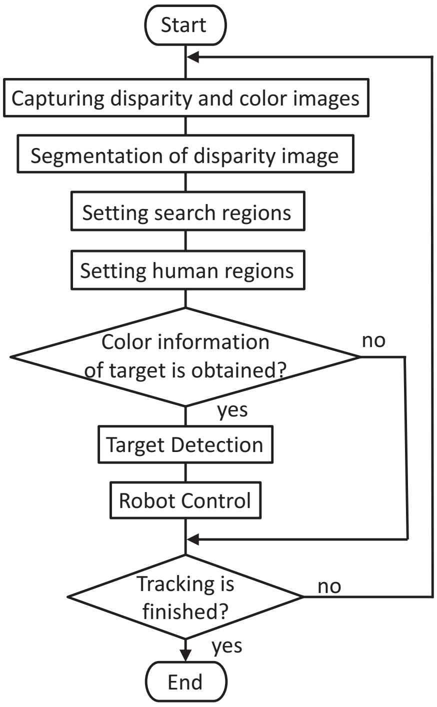
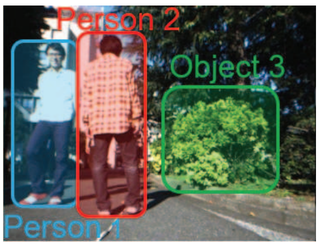
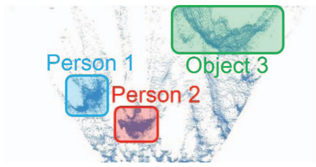


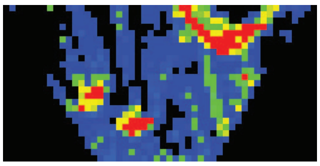
Fig. 1. Flow chart of the proposed method



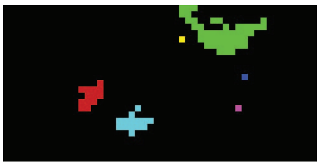
(a) Input image



(b) Projected points



(c) Histogram of projected points in each cell



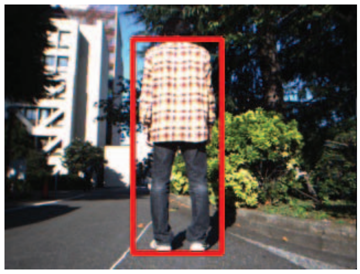
(d) Labeled cells

Fig. 2. Process of setting search regions

closer to red the color becomes. There tend to be many points in the cells in the human region because of its height and width. Therefore, the cells that have the projected points over a threshold are labeled (Fig. 2(d)).

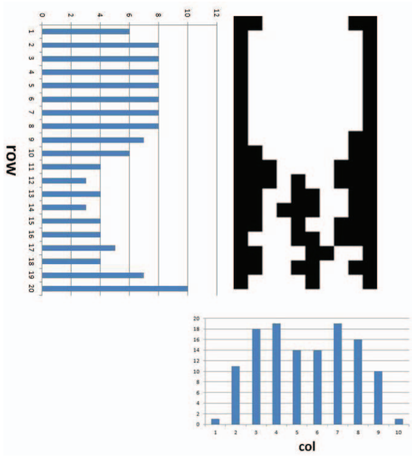
C. Setting search regions

Using the width, depth, and height of each labeled group of cells, a rectangular parallelepiped region is produced. If the region is correctly fixed as human, the length of each side follows into a certain value. Thus, each search region is extracted based on the lengths of the sides. Search regions have the shapes of rectangular parallelepipeds and include objects that have the contours of humans.



(a) Input image

(b) Binary image



(c) Histograms extracted from low resolution binary image

Fig. 3. Process of setting human regions

D. Setting human regions

The result of setting search regions of an input image is shown as a red rectangle in Fig. 3(a). As shown in the figure, each search region is held at the height of one's shoulder from the ground. In a disparity image, each pixel in a rectangular region is binarized so that the human region is shown as white, based on whether the distance value of each pixel is in the search region. Fig. 3(b) is the binary image obtained from the region in Fig. 3(a). The resolution of the binary image is lowered, and then the histograms, which show the values of white pixels at each column and row in the low resolution image, are produced (Fig. 3(c)). These histograms are compared with the template that was previously produced. The similarity is calculated by the Sum of Absolute Difference (SAD) as calculated below:

$$R_{SAD} = \sum_{j=1}^{N} |I_r(j) - T_r(j)| + \sum_{i=1}^{M} |I_c(i) - T_c(i)| \quad (1)$$

In (1), N and M are the row and column numbers in the image, I_r and I_c are respective values of histograms of an input image, and T_r and T_c are respective values of template histograms. When R_{SAD} is less than a threshold, the region is defined as a human region.

E. Target detection from color information

To detect a human region that matches a target, hue and saturation are used as the features of a target. Hue and

TABLE I. EXPERIMENTAL SCENES IN INDOOR ENVIRONMENTS

Scene	Condition	Object	Shadow
1	sun shining from the back	wall	no appearance
2	indirect illumination studded	wall	frequent appearance
3	sun shining from the side	wall and window	occasional appearance

saturation are components of the HSV color space that are robust to illumination changes. If human regions are defined, the information of each region is extracted. Subsequently, the similarity between humans and the template color information of a target acquired beforehand is calculated as follows:

$$R_{Bhat} = \sqrt{1 - \sum_{h}\sum_{s}\sqrt{H_{input}(h,s)H_{template}(h,s)}} \quad (2)$$

where $H_{input}(h,s)$ and $H_{template}(h,s)$ are an H-S histogram of input and template information, where h and s are hue and saturation, respectively.

In order to adjust to lighting changes, the template's color information is updated at intervals of a few frames. However, if mis-detection has occurred, it becomes likely that incorrect color information will be updated as target information. Therefore, to solve this problem, two versions of the color information are held. One version is the information that has been updated continually; the other version remains unchanged from registration. When updating, if the similarity of the current color information to the registered information is calculated to fall under a certain threshold, the color information is considered to belong to the target.

III. TARGET FOLLOWING EXPERIMENT

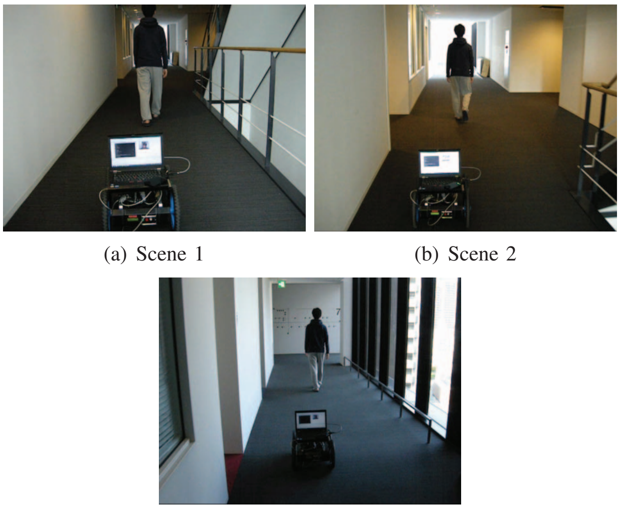
We have implemented the proposed method on a mobile robot (Segway, BlackShip) with a stereo camera (Point Grey Research, Bumblebee2) for the indoor and outdoor experiments. In addition, PID control is used to adjust the distance from a robot to a target and the angle formed by their 3D positions. The effectiveness of the proposed system is verified by two evaluation values, *Precision* and *Recall*.

$$Precision = \frac{\mathrm{A}}{\mathrm{A+B}}, \quad Recall = \frac{\mathrm{A}}{\mathrm{A+C}} \quad (3)$$

A: The number of frames in which the target is correctly detected
B: The number of frames in which a non-target is detected
C: The number of frames in which no objects are detected

A. Indoor target tracking experiment

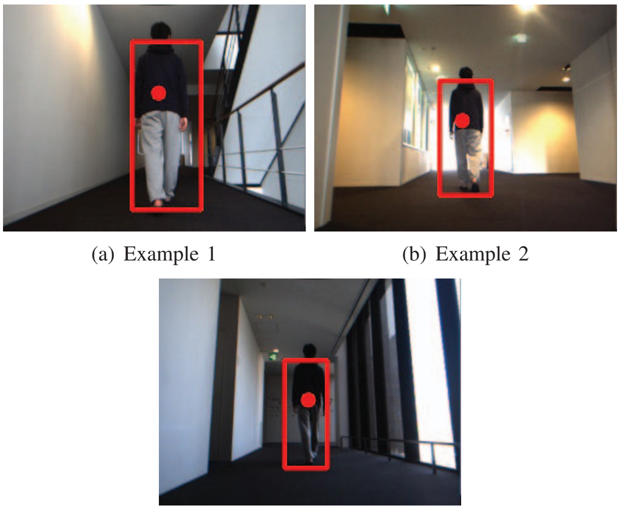
First, the target tracking experiment was performed indoors. The experimental environments are classified into three scenes, as shown in Table I. In Table I, *Condition* means the light condition; *Object* shows what was in the scenes; *Shadow* indicates how often the shadow appeared. Fig. 4 represents these scenes. The target detection examples are shown in Fig. 5. In the figure, the rectangles are the same target regions as are shown in Fig. 3(a), and the dots show the centroids of the white pixels of the low resolution binary images, with reference to Subsection II-D. Fig. 6 indicates the evaluation



(a) Scene 1 (b) Scene 2

(c) Scene 3

Fig. 4. Experimental scenes in indoor environments



(a) Example 1 (b) Example 2

(c) Example 3

Fig. 5. Examples of detecting a target indoors

of the experiment. The experiment has resulted *Precision* and *Recall* values of higher than 90% in all situations. As shown in Fig. 5, the brightness of the target's clothes was changed every moment by indirect illumination and the sunlight through the windows. This experiment has demonstrated that, under three different lighting conditions, a robot could track a target in indoor environments.

B. Outdoor target tracking experiment

In this section, the experiment of tracking a target was conducted outdoors, a more complicated illumination environment.

The experimental environments had five scenes, as indicated in Table II. These scenes and examples of detecting a target are shown in Fig. 7 and Fig. 8, respectively. Fig. 9 shows the results of the experiment.

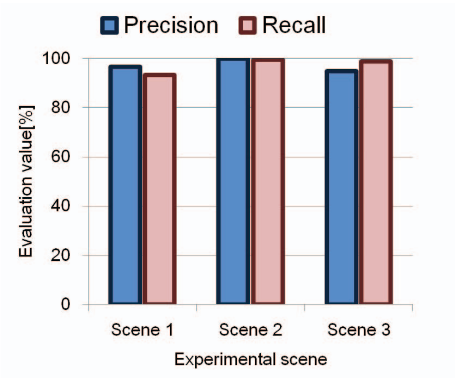
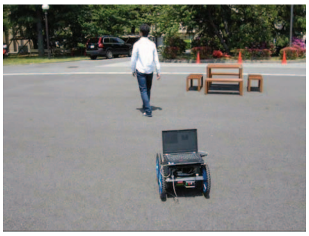


Fig. 6. Evaluation of results in indoor scenes

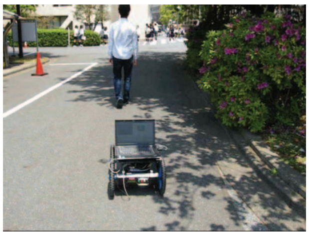
TABLE II. EXPERIMENTAL SCENES IN OUTDOOR ENVIRONMENTS

Scene	Condition	Object	Shadow
1	direct lighting	none	no appearance
2	back lighting	none	no appearance
3	direct lighting	trees and a car	frequent appearance
4	direct lighting	none	constant appearance
5	direct lighting	poles and bikes	occasional appearance



(a) Scene 1

(b) Scene 2



(c) Scene 3

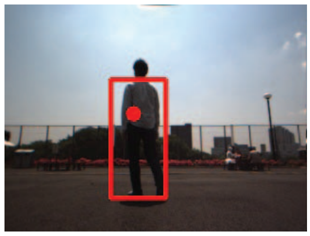
(d) Scene 4

(e) Scene 5

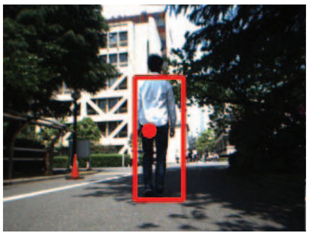
Fig. 7. Experimental scenes in outdoor environments

High *Precision* values could be obtained, and each *Recall* value was higher than 80%. The *Precision* value of Scene 3 is lower than that of other scenes because, as compared to them, there were more objects in the scene whose shapes are like humans. Moreover, mis-detection will increase if

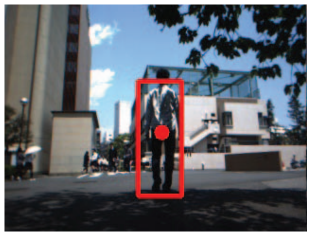
(a) Example 1



(b) Example 2



(c) Example 3



(d) Example 4

(e) Example 5

Fig. 8. Examples of detecting a target outdoors

Fig. 9. Evaluation of results in outdoor scenes

additional people are present. This results from using only color information for a target feature. For safe tracking, mis-tracking should not occur; no mis-detection of a target is desirable, wherever a robot performs.

On the other hand, the *Recall* value of Scene 5 is the lowest of all. This would be because the light-condition changes in the environment of Scene 5 were so frequent and extreme that updates of the color information could not adjust it. Although there were some scenes in which the robot could not find the target, the robot never stopped moving. This is because target detection was repeated in every frame and the color information of the target was updated frequently; therefore, the robot could immediately once again detect the target.

However, if the light condition changes more extremely (e.g., in the case of moving from indoors to outdoors), the robot might be stopped. Therefore, the non-detected scene should be decreased, and the *Recall* value also needs to be higher to use the system in real environments. These goals can be achieved by developing the present system with more features to distinguish humans from other objects and to distinguish a target from other people under varying illuminations.

IV. Conclusion

In this paper, a target tracking system for a mobile robot under varying illuminations using a stereo camera has been proposed. Extraction of a 3D region of a target person by obtaining features that are robust to illumination changes is applied to target tracking in environments with changing light conditions. In indoor and outdoor experiments, nearly 100% of the *Precision* value and more than 80% of the *Recall* value were obtained. Through these experiments, we confirmed that the method is capable of tracking a specific person under various illumination environments.

The current method does not consider environments with multiple people. To cope with such cases, it would be necessary to use time-series information.

References

[1] D. Katz, J. Kenney, and O. Brock, *How Can Robots Succeed in Unstructured Environments?*, In RSS Workshop on Robot Manipulation: Intelligence in Human Environments, 2008.

[2] J. H. Lee, T. Tsubouchi, K. Yamamoto, and S. Egawa, *People Tracking Using a Robot in Motion with Laser Range Finder*, Proc. of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2936-2942, 2006.

[3] L. Mattos and E. Grant, *Passive Sonar Applications: Target Tracking and Navigation of an Autonomous Robot*, Proc. of the 2004 IEEE Int. Conf. on Robotics and Automation, pp. 4265-4270, 2004.

[4] M. Hebert, *Active and Passive Range Sensing for Robotics*, Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation, pp. 102-110, 2000.

[5] C. Hu, X. Ma, and X. Dai, *A Robust Person Tracking and Following Approach for Mobile Robot*, Proc. of the 2007 IEEE Int. Conf. on Mechatronics and Automation, pp. 3571-3576, 2007.

[6] J. Fritsch, M. Kleinehagenbrock, S. Lang, G. A. Fink, and G. Sagerer, *Audiovisual Person Tracking with a Mobile Robot*, Proc. of the 2004 Int. Conf. on Intelligent Autonomous Systems, pp. 898-906, 2004.

[7] M. Doi, M. Nakakita, Y. Aoki, and S. Hashimoto, *Real-Time Vision System for Autonomous Mobile Robot*, Proc. of the 10th IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 442-449, 2001.

[8] J. Satake and J. Miura, *Robust Stereo-Based Person Detection and Tracking for a Person Following Robot*, Proc. of the ICRA 2009 Workshop on People Detection and Tracking, 2009.

[9] Z. Chen and S. T. Birchfield, *Person Following with a Mobile Robot Using Binocular Feature-Based Tracking*, Proc. of the 2007 IEEE Int. Conf. on Intelligent Robots and Systems, pp. 815-820, 2007.

[10] H. Takemura, Z. Nemoto, and H. Mizoguchi, *Development of Vision Based Person Following Module for Mobile Robots In/Out Door Environment*, Proc. of the 2009 IEEE Int. Conf. on Robotics and Biomimetics, pp. 1675-1680, 2009.

[11] T. Ubukata, K. Terabayashi, A. Moro, and K. Umeda, *Multi-Object Segmentation in a Projection Plane Using Subtraction Stereo*, Proc. of the 20th IEEE Int. Conf. on Pattern and Recognition, pp. 32963299, 2010.