

Detection of pedestrians using subtraction stereo

Yuuki Hashimoto*, Yusuke Matsuki*, Tatsuya Nakanishi*, Kazunori Umeda*, Kei Suzuki** and Kazunori Takashio**

Abstract— In this paper, detection of pedestrians using “subtraction stereo” is discussed. Subtraction stereo is a stereo vision method that focuses on the movement of objects to make a stereo camera robust and produces range images for moving regions. Features of pedestrians such as 3D position, height and width are obtained from range images obtained by subtraction stereo. Then a simple method to remove shadows is proposed. The basic algorithm of the subtraction stereo is implemented on a commercially available stereo camera, and the effectiveness of the method to detect pedestrians with removal of shadows is verified by experiments using the stereo camera.

I. INTRODUCTION

A huge number of studies have been carried out for stereo vision until now [1], [2]. These days, several practical stereo vision systems have been reported. Some studies realize real-time acquisition of range images using personal computers (PCs) because the CPUs and Graphics Processing Units (GPUs) are fast enough [3], [4]. In some studies, a Field Programmable Gate Array (FPGA) is used instead of a PC to acquire range images [5]. Some stereo cameras that are connected to a PC are commercially available [6] and widely used. There are stereo cameras that are practically used for automotive cars [7].

We are aiming at developing a practical stereo camera for applications such as surveillance, in which detection of anomalies or measurement of moving people are required. Several systems have been proposed for such kind of surveillance using a single camera [8]. However, a single camera is not sufficient since the size of targets is not obtained. Stereo cameras are more appropriate, for size information can be directly obtained and scalable use for several scenes becomes possible.

We focus on the movement of objects to make a stereo camera robust and have proposed “subtraction stereo” [9]. In this paper, we discuss the detection and measurement of pedestrians using the subtraction stereo.

This paper is organized as follows. In section II, we show the outline of subtraction stereo. In section III, we discuss the measurement of features of pedestrians. Then in section IV, we propose a simple method to remove shadows that are harmful for detection of pedestrians. Experimental results to detect pedestrians using the stereo camera with subtraction stereo algorithm are given in section V. After

some discussions in section VI, this paper is concluded in section VII.

II. BASIC ALGORITHM OF SUBTRACTION STEREO

We show the basic algorithm of the subtraction stereo in Fig.1. In standard stereo vision, two images captured with right and left cameras are matched and disparities are obtained at each pixel. The subtraction stereo adds a step to extract moving regions in the images of each camera, and then applies the stereo matching to the extracted moving regions. The extraction of moving regions is realized by an appropriate subtraction method, in which the simplest one is background subtraction. From a disparity image, a range image can be obtained.

In compensation that the pixels where a disparity is obtained are restricted to moving regions, the subtraction stereo realizes the robustness of the stereo matching with the following reasons.

- Search space for matching can be strongly restricted.
- Motion information as well as the original image can be used for matching.

Note that a target object need not be “moving” literally; a still object can be detected by a subtraction method.

Fig.2 shows an example of the disparity image obtained by the subtraction stereo. Fig.2(a) shows the color image of a target scene, a typical outdoor scene. This image was captured by the right camera of the stereo camera. Fig.2(b) is the obtained disparity image. Color represents the disparity in these images. Bluer color indicates larger disparity, i.e. smaller distance, and redder color indicates smaller disparity, i.e. larger distance. In this experiment, target persons are far and small, and the scene is difficult for standard stereo. Fig.2(b) shows that the target persons are extracted appropriately with distance information. It can be said that subtraction stereo gives enough and better information for applications such as surveillance in which moving (changed) regions are important.

III. MEASUREMENT OF FEATURES OF PEDESTRIANS

In this section, we explain the measurement of features of pedestrians with the subtraction stereo. Three dimensional (3D) position, height, width, etc. of a moving pedestrian can be measured.

*: Course of Precision Engineering, School of Science and Engineering, Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan hashimo@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

** : Faculty of Environmental Information, Keio University / CREST, JST, 5322 Endo, Fujisawa-shi, Kanagawa 252-8520 Japan

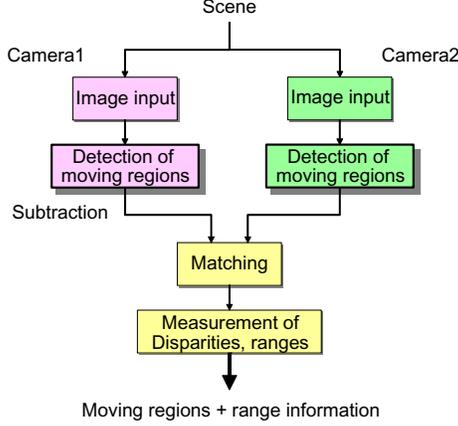
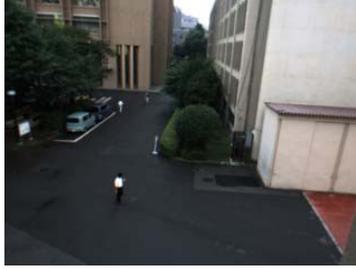


Fig. 1. Flow of subtraction stereo



(a) Color image of the experimental scene



(b) Obtained disparity image

Fig. 2. An example of subtraction stereo

1) *Labeling*: In the subtraction stereo, as the obtained disparity image is originally restricted to moving regions, moving pedestrians can be obtained with the standard labeling technique. Thresholding for the area is effective to remove noises or limit the size of pedestrian to be extracted.

2) *Measurement of 3D position*: When the disparity of a point is given, the corresponding distance along the optical axis and 3D position of the point are calculated. Hereafter, we use the word distance as the meaning of distance along the optical axis. Let disparity be k [pixel] and the distance be z [m]. The distance z is obtained by the following equation.

$$z = \frac{\alpha}{k}, \quad \alpha = \frac{b \cdot f}{p} \quad (1)$$

where b is the baseline length, f is the focal length of the lens, and p is the width of each pixel of the image.

Furthermore, 3D position of the point x is obtained from the distance z and the image coordinates (u, v) of the point in the image (see Fig.3). Assuming that there is no skew and the aspect ratio of each pixel is 1, then x is given as

$$\mathbf{x} = z \cdot \left[\frac{p}{f}(u - u_0) \quad \frac{p}{f}(v - v_0) \quad 1 \right]^T \quad (2)$$

where (u_0, v_0) is the image coordinates of the image center.

3) *Coordinate transformation*: When position and orientation of the camera are known, the point's 3D position \mathbf{x}_w in the world coordinate system can be calculated. Assume that \mathbf{t} and R are camera's position and orientation respectively. \mathbf{t} is a 3D vector and R is a 3×3 rotation matrix. Then,

$$\mathbf{x}_w = R\mathbf{x} + \mathbf{t}. \quad (3)$$

4) *Measurement of height and width*: The height and width of a pedestrian can be measured as follows.

$$height = z \cdot \frac{p}{f}(v_{\max} - v_{\min}) \quad (4)$$

$$width = z \cdot \frac{p}{f}(u_{\max} - u_{\min}) \quad (5)$$

where z is the value that represents the distance to the pedestrian, and the parameters in the image coordinate system are shown in Fig.3.

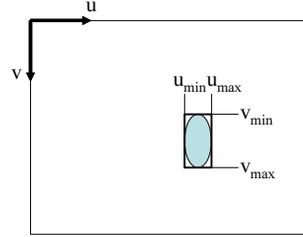


Fig. 3. Parameters of an object in the image coordinate system

IV. METHOD TO REMOVE SHADOWS

In this paper, the targets to extract are pedestrians. If pedestrians are in the sun, their shadows are casted as shown in Fig.4(a). A long shadow is casted on the right-hand side in Fig.4(a) and the regions of the pedestrian and the casted shadow are recognized as one region. Therefore, it is necessary to remove the shadow to get the correct width and height of the pedestrian region.

We propose a simple method to remove shadows. We use the heuristics that shadows appear at the bottom of the pedestrian. First, we obtain the top and the bottom of the region in the image coordinate system. And then we obtain the left and right ends of the region in the row one third from the top as shown in Fig.5(a). Finally, we exclude the regions not between the left and right columns. Fig.4(b) shows the region extracted by this method from Fig.4(a). We can see that shadows are removed appropriately.

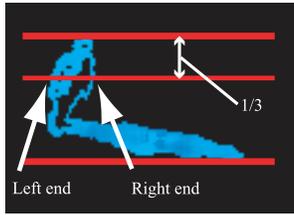


(a) Moving region in the sun

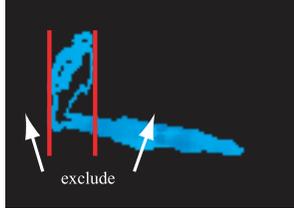


(b) Moving region after the method to remove shadow is applied

Fig. 4. Removal of shadows



(a) Detection of left and right ends



(b) Removal of shadows

Fig. 5. The method to remove shadows

V. EXPERIMENTAL EVALUATION

In this section, we show experimental results to evaluate the method to detect pedestrians, especially, removal of shadows.

A. Implementation of basic algorithm using a commercially available stereo camera

We implemented the basic algorithm of subtraction stereo on a commercially available stereo camera. The stereo camera is Point Grey Research Bumblebee2 (color, $f=3.8\text{mm}$). We set the size of the image to 320×240 .

So far, the simple background subtraction is applied to extract moving objects. The stereo matching procedure of the Bumblebee2 library was applied to the subtraction images of the right and left cameras and a disparity image is obtained. Measurement errors with respect to the distance are evaluated in [9]. The rate to obtain disparity images is about 16[fps] with a PC (Pentium4 3GHz).

The stereo camera was installed so that the pedestrians were looked down on from a building as shown in Fig.6. Fig.7 shows the experimental setup. The camera was set at the height of 8.3[m] with 50° downward tilt.



Fig. 6. Experimental scene

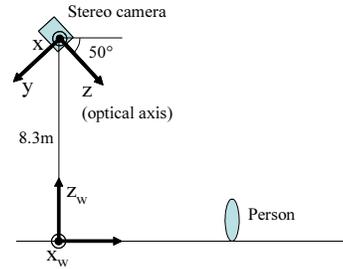


Fig. 7. Experimental setup of the experiments

B. Experiment to evaluate the accuracy of the method to remove shadows

We evaluated the accuracy of the method to remove shadows in the sun. The conditions of the experiment are as follows.

- Pedestrians were not controlled; students walking at a campus were measured.
- The pedestrians were detected in real time.
- Experiments were done 20 times each at four time periods 11:00-12:00, 12:00-13:00, 13:00-14:00, 14:00-15:00.
- If shadows were removed continuously during a pedestrian is in the sun, it is regarded as a success.

The reason we divided the time periods is that the length and direction of shadows change by time. At 11:00-12:00 the pedestrians were mostly in the sun, and at 14:00-15:00, the rate to be in the sun decreased to about 1/4. The result is shown in Table I. This shows that removal of shadows was completely successful at very time period when the number of pedestrians was one. The failures occurred when two or more pedestrians and their shadows overlapped in the image and they were detected as one region, as shown in Fig.8.

C. Measurement of a pedestrian in the sun

We show experimental results to measure a pedestrian in the sun. The 3D position, height and width of a pedestrian were measured sequentially. A person walked from

TABLE I
SUCCESS RATE OF REMOVING SHADOWS

time period	total[%]	one person[%]
11- 12	75	100
12- 13	80	100
13- 14	90	100
14- 15	90	100
total	83.8	100

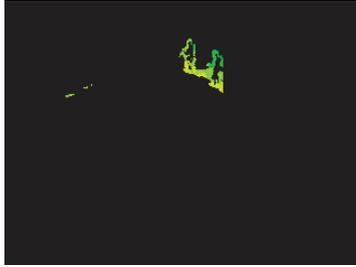


Fig. 8. An example of failure of removing shadows

left to right. Fig.9 show the results. After the 45th frame, the pedestrian appeared in the sun. Fig.9(a) shows the 3D position in the world coordinate system, and Fig.9(b) shows the height and width of the person. The movement of the pedestrian is detected appropriately, and the height and width are measured well. The vibration of the width corresponds to the change of the swing of legs.

The results are not affected by the shadows that appeared after the 45th frame. It shows that the shadows were removed successfully by the proposed method.

VI. DISCUSSION

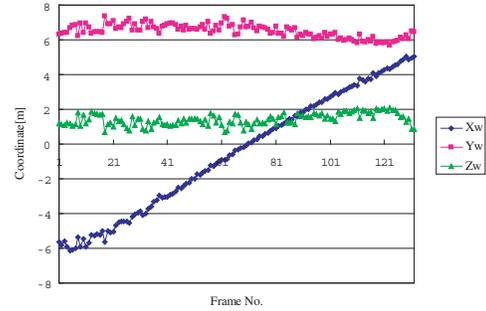
From the experimental results, we showed that detection of pedestrians using subtraction stereo works well and the method to remove shadows is effective. The basic algorithm of subtraction stereo is quite simple and easy to implement. Therefore, our final goal is to construct a stereo camera with the function of subtraction stereo and to apply it for surveillance applications.

There are many open issues as follows.

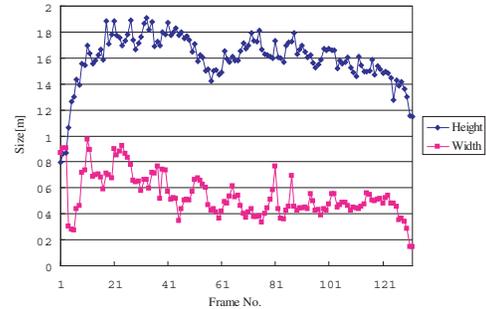
- Improvement of the subtraction stereo itself - the implemented subtraction algorithm is quite simple now.
- Improvement of the detection of pedestrians - when two or more pedestrians overlap, the detection fails.
- Improvement of the removal of shadows - when a shadow is casted vertically, the proposed method fails.
- Extraction of high-level information of the scene with multiple people, etc.

VII. CONCLUSION

In this paper, we discussed the detection of pedestrians using subtraction stereo. Features of pedestrians such as 3D position, height and width are obtained from range images obtained by subtraction stereo. A simple method to remove shadows was proposed. The basic algorithm of the subtraction stereo was implemented on a commercially



(a) 3D Position



(b) Height and width

Fig. 9. Experimental result of measuring a pedestrian in a general outdoor scene

available stereo camera, and the effectiveness of the method to detect pedestrians with removal of shadows was verified by experiments using the stereo camera.

REFERENCES

- [1] R. Hartley and A. Zisserman: "Multiple View Geometry in Computer Vision," Cambridge Univ. Press, 2000.
- [2] M. Hebert: "Active and passive range sensing for robotics," in Proc. of ICRA'00, Vol.1, pp.102-110, 2000.
- [3] S. Kagami, K. Okada, M. Inaba, H. Inoue: "Realtime 3D Depth Flow Generation and its Application to Track to Walking Human Being," Proc. of 15th International Conference on Pattern Recognition (ICPR2000), Vol.4, pp.197-200, 2000.
- [4] T. Ueshiba: "An Efficient Implementation Technique of Bidirectional Matching for Real-time Trinocular Stereo Vision," Proc. of 18th International Conference on Pattern Recognition (ICPR2006), Vol.1, pp.1076-1079, 2006.
- [5] M. Hariyama, Y. Kobayashi, H. Sasaki, M. Kameyama: "FPGA Implementation of a Stereo Matching Processor Based on Window-Parallel-and-Pixel-Parallel Architecture," IEICE Trans. Fundamentals, Vol.E88-A, No.12, pp.3516-3522, 2005.
- [6] Point Grey Research, <http://www.ptgrey.com/>
- [7] K. Hanawa and Y. Sogawa: "Development of Stereo Image Recognition System for ADA," Proc. IEEE Intelligent Vehicle Symposium'01, 2001.
- [8] T. Haga, K. Sumi, Y. Yagi: "Human Detection in Outdoor Scene using Spatio-Temporal Motion Analysis", Proc. of 17th International Conference on Pattern Recognition (ICPR2004), Vol.4, pp.331-334, 2004.
- [9] K. Umeda, M. Wada, Y. Matsuki, Y. Hashimoto, T. Nakanishi, K. Irie: "Subtraction Stereo - A Stereo Camera System That Focuses On Moving Regions -", (under review).