Subtraction stereo
- a stereo camera system that focuses on moving regions -

Kazunori Umeda[a], Yuuki Hashimoto[b], Tatsuya Nakanishi[b], Kota Irie[b] and Kenji Terabayashi[a]

[a]Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan;
[b]Course of Precision Engineering, School of Science and Engineering, Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

ABSTRACT

This study aims at developing a practical stereo camera that is suitable for applications such as surveillance, in which detection of anomalies or measurement of moving people are required. In such surveillance cases, targets to measure usually move. In this paper, "Subtraction stereo" is proposed that focuses on motion information to increase the robustness of the stereo matching. It realizes robust measurement of range images by detecting moving regions with each camera and then applying stereo matching for the detected moving regions. Measurement of three-dimensional position, height and width of a target object using the subtraction stereo is discussed. The basic algorithm is implemented on a commercially available stereo camera, and the effectiveness of the subtraction stereo is verified by several experiments using the stereo camera.

Keywords: Stereo vision, subtraction, motion analysis, detection of moving regions, range image, surveillance camera, error analysis

1. INTRODUCTION

A huge number of studies have been carried out for stereo vision until now.[1–4] 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.[5,6] In some studies, a Field Programmable Gate Array (FPGA) is used instead of a PC to acquire range images.[7] Some stereo cameras that are connected to a PC are commercially available[8,9] and widely used. There are stereo cameras that are practically used for automotive cars.[10]

As explained above, we can say that stereo cameras have already reached a practical level. However, what is called the correspondence problem that stereo matching becomes difficult and not robust for weak textures or recurrent patterns is inevitable for stereo vision. To solve the problem, muti-baseline stereo[11] that uses multiple cameras and make the stereo matching more robust is well known. Another approach is to project some pattern such as random dot pattern to give texture on a scene. However, this approach does not work when targets are far.

In this paper, we aim 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.[12] 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. For example, it becomes possible to distinguish whether the target is a human, a car or a small animal, to know whether the target person is an adult or a child, or to count the number of people in a crowd.

Further author information:
E-mail: {umeda,terabayashi}@mech.chuo-u.ac.jp, {hashimo,tnakani,kotairie}@sensor.mech.chuo-u.ac.jp
http://www.mech.chuo-u.ac.jp/umedalab/index_e.html

Three-Dimensional Imaging Metrology, edited by J. Angelo Beraldin, Geraldine S. Cheok, Michael McCarthy, Ulrich Neuschaefer-Rube, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7239, 723908
 · doi: 10.1117/12.805718

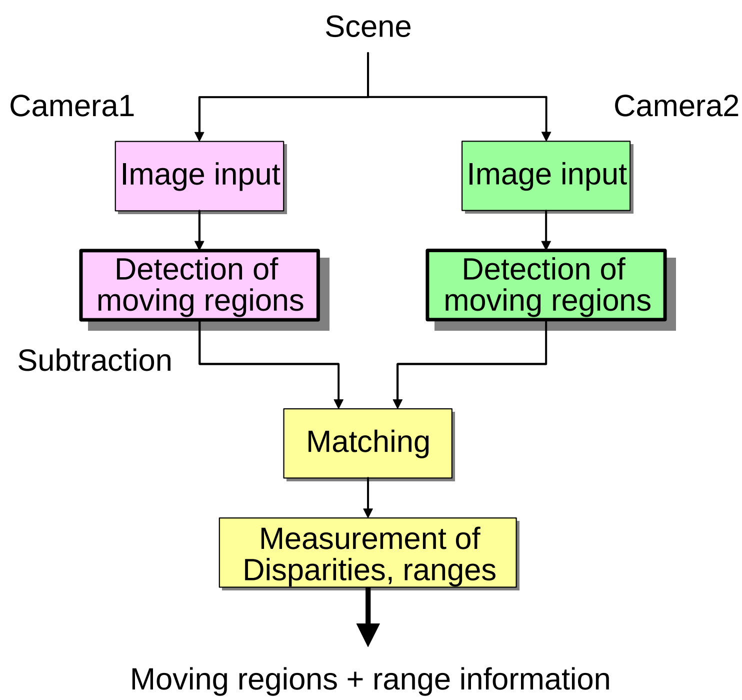


Figure 1. Flow of subtraction stereo.

In such surveillance cases, targets to measure usually move. Therefore, we focus on the movement to make a stereo camera robust, instead of using multiple cameras or an active projection. We propose what we call "subtraction stereo".

The rest of the paper is organized as follows. In Sec. 2, the basic algorithm of the subtraction stereo is proposed. In Sec. 3, measurement of moving objects is discussed, and error analysis is given. Implementation of the basic algorithm on a stereo camera is explained in Sec. 4, and several experiments using the stereo camera are given in Sec. 5 to evaluate the subtraction stereo. Some discussions are given in Sec. 6, and this paper is concluded in Sec. 7.

2. BASIC ALGORITHM

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.

3. MEASUREMENT OF FEATURES OF OBJECTS

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

3.1 Labeling

In the subtraction stereo, as the obtained disparity image is originally restricted to be moving regions, moving objects can be obtained with the standard labeling technique. Thresholding for the area is effective to remove noises or limit the size of object to be extracted.

3.2 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.

$$z = \frac{\alpha}{k}, \qquad \alpha = \frac{b \cdot f}{p} \tag{1}$$

where b is the baseline length, i.e., the interval between the the two lens centers, 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 $\mathbf{x}$ is obtained from the distance z and the image coordinates (u, v) of the point in the image (see Fig. 2). Assuming that there is no skew and the aspect ratio of each pixel is 1, then $\mathbf{x}$ is given as

$$\mathbf{x} = z \cdot \begin{bmatrix} \frac{p}{f}(u - u_0) & \frac{p}{f}(v - v_0) & 1 \end{bmatrix}^T \tag{2}$$

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

In case a target object is near, the disparity becomes large and the distance can be measured precisely. Therefore, 3D coordinates of each point of the object can be obtained precisely and used. On the contrary, in case a target object is far, the measurement error becomes large. In this case, the distance to the object should be represented by one value, for example, the distance at the center of the region. In surveillance applications, target objects are usually far.

The discussion on the error is given in 3.5.

3.3 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,

$$\mathbf{x}_w = R\mathbf{x} + \mathbf{t}. \tag{3}$$

3.4 Measurement of Height and Width

The height and width of an object can be measured as follows.

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

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

where z is the value that represents the distance to the object, and the parameters in the image coordinate system are shown in Fig. 2.

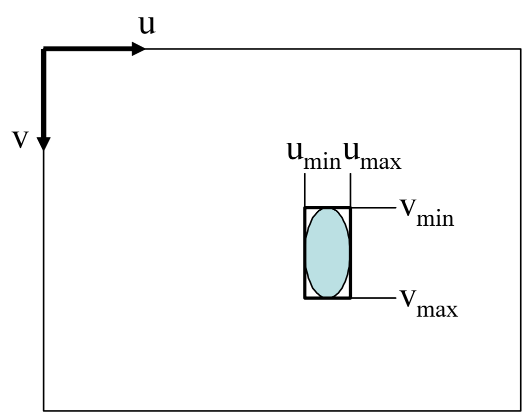


Figure 2. Parameters of an object in the image coordinate system.

3.5 Error Analysis of Range Measurement

Let the precision of disparity be σ_k (standard deviation). By applying the law of propagation of errors to Eq. (1), the precision of the measured distance σ_z (standard deviation) is given by

$$\sigma_z = \left|\frac{\partial z}{\partial k}\right| \sigma_k = \frac{p}{b \cdot f} z^2 \sigma_k. \tag{6}$$

This equation shows that the error of measuring distance is proportional to the square of the distance, which is well known for range measurement with triangulation. In addition, the error is proportional to the width of each pixel, and inverse proportional to the baseline length and focal length. For example, when b is 120 mm, f is 3.8 mm, and p is 14.8 μm, and σ_k is 0.1 pixel,

$$\sigma_z = 3.246 \times 10^{-3} z^2. \tag{7}$$

For z = 1, 10, 25, and 50 m, σ_z = 0.003, 0.3, 2, and 8 m respectively. Note that these parameters roughly correspond to the experimental conditions in the next section.

4. 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=3.8mm). We set the size of the image to 320×240.

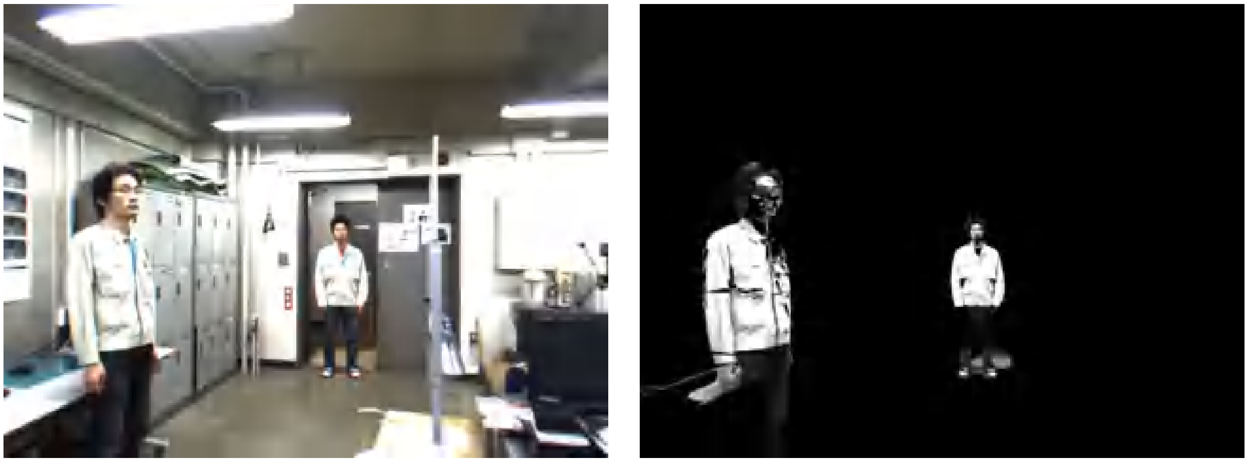
So far, the simple background subtraction is applied to extract moving objects. The stereo matching procedure of the Bumblebee2 library was applied to the subtraction images of the right and left cameras and a disparity image is obtained. The rate to obtain disparity images is about 37 fps with a PC with Core 2 Duo T9300 (2.50GHz).

5. EXPERIMENTAL EVALUATION

In this section, we show several experimental results to evaluate the subtraction stereo.

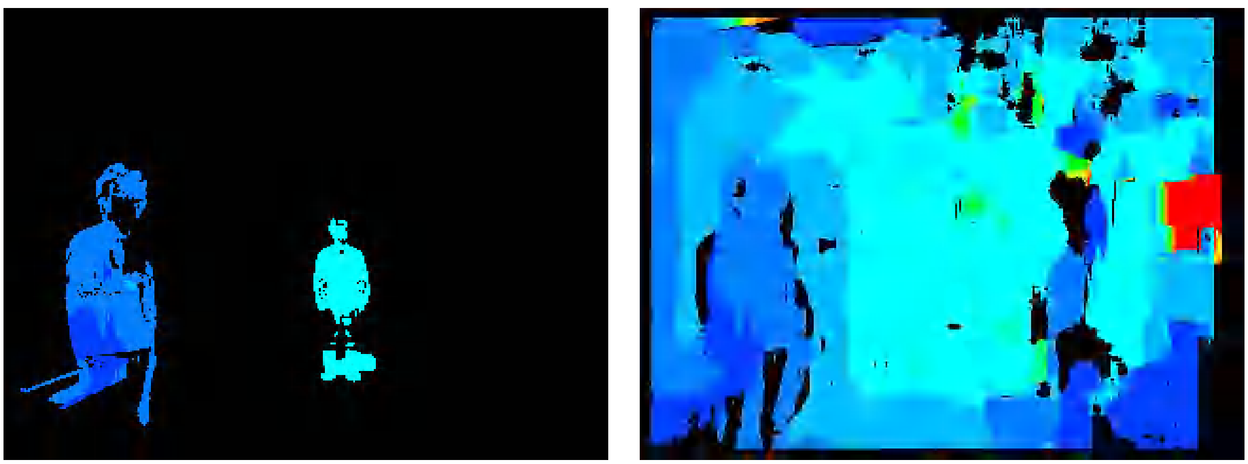
5.1 Experiments to Evaluate Basic Algorithm

Firstly, we show two examples for typical scenes to evaluate the basic algorithm.



(a) Color image of the experimental scene (b) Subtraction image

Figure 3. Experimental scene: indoor.



(a) Subtraction stereo (b) Standard stereo

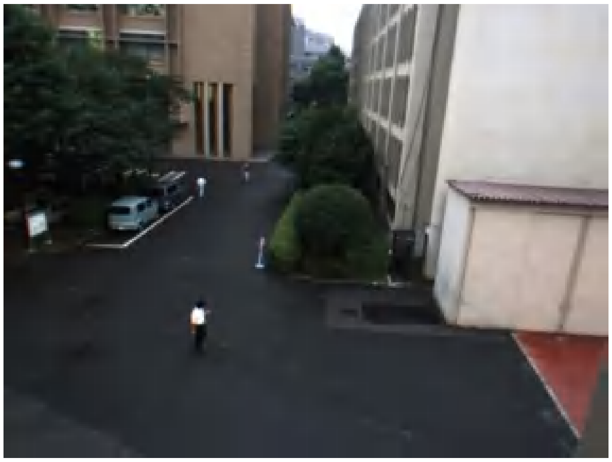
Figure 4. An example of subtraction stereo: indoor.

5.1.1 Indoor experiment

Figure 3(a) shows the color image of a target scene, a typical indoor scene. This image was captured by the right camera of the stereo camera (the same for the following experiments). Background subtraction was applied using the background image that was registered beforehand. The obtained subtraction image is shown in Fig. 3(b). Figure 4(a) is the disparity image obtained by the subtraction stereo from right and left subtraction images. In contrast, Fig. 4(b) is the disparity image obtained by the standard stereo matching. 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. Figure 4(a) shows that appropriate disparity images are obtained for moving objects only. As the stereo matching is restricted to moving regions, calculation cost can be reduced, and wrong correspondences can be suppressed. In fact, some regions with wrong matching are observed in Fig. 4(b) (e.g., right red part) and not in Fig. 4(a).

5.1.2 Outdoor experiment

Figure 5(a) shows the color image of a target scene, a typical outdoor scene. As the indoor experiment, Fig. 5(b) shows the subtraction image, and Fig. 6(a), (b) show the disparity images obtained by the subtraction stereo and standard stereo respectively. In this experiment, target persons are far and small, and the scene is difficult for

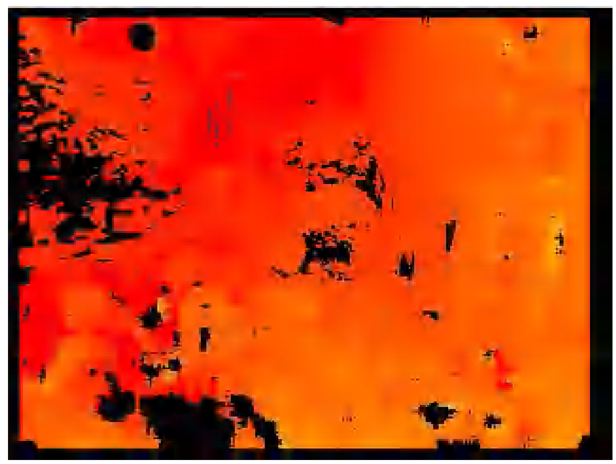


(a) Color image of the experimental scene

(b) Subtraction image

Figure 5. Experimental scene: outdoor.

(a) Subtraction stereo



(b) Standard stereo

Figure 6. An example of subtraction stereo: outdoor.

standard stereo. In fact, it is not easy to detect persons from Fig. 6(b). Figure 6(a) 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.

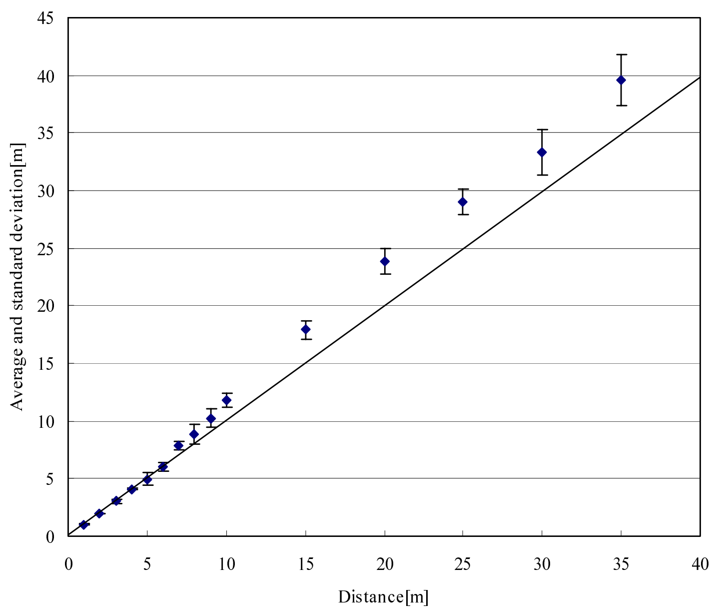
5.2 Experiment to Evaluate Measurement Error

Here we show the result of fundamental experiment to evaluate measurement errors with respect to the distance. A 440 mm×1400 mm flat object with texture was set up on an indoor passage, with known distances from the stereo camera as shown in Fig. 7(a). Measurement was done every 1 m from 1 to 10 m, and every 5 m and from 10 to 35 m. At each distance, the distance at the center of the extracted region was measured 100 times and the average and standard deviation were calculated. Figure 7(b) shows the results. When the measured distance is small, the measurement errors are small. And the errors at large distances are large, and some biases are observed. However, we think that the measurement error even at 35 m, which is 10-20% of the distance, is small enough to be applicable for surveillance use.

5.3 Measurement of a Person Walking Straight Away

We show another fundamental experiment. A person walking straight away was measured at the same passage as shown in Fig. 8(a). Figure 8(b) shows the obtained distances. The horizontal axis represents the frame number.

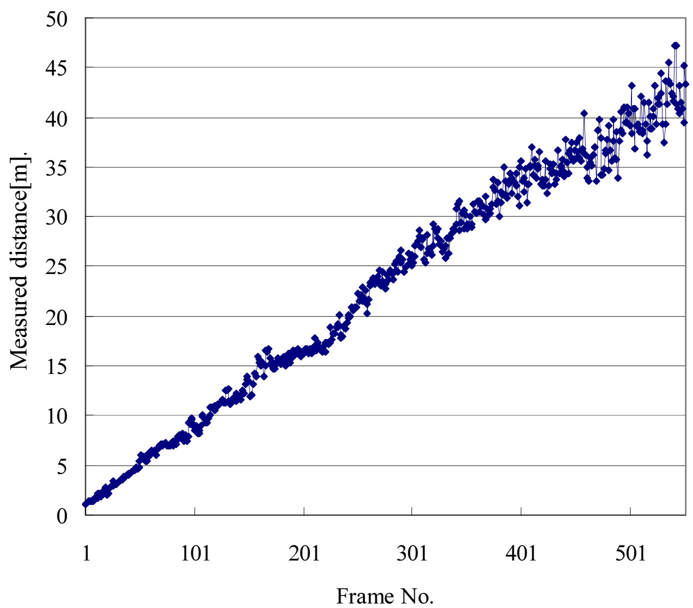
(a) Experimental scene



(b) Results

Figure 7. Experiment to evaluate measurement error: still object.

(a) Experimental scene



(b) Results

Figure 8. Experiment to measure a person walking straight away.

The distance at the center of the extracted person was measured.

It can be said that distances can be measured robustly and appropriately even when a target object is moving. We can see that the measurement error becomes larger with respect to the increase of distance. The size of errors are nearly the same as the still object in Fig. 7. The real distance at the last frame is 37.5 m; some bias is observed again.

5.4 Measurement of a Walking Person in a General Scene

Finally, we show experimental results to measure a walking person in a general outdoor scene. The 3D position, height and width of a walking person were measured sequentially. Figure 9 shows the experimental setup. The camera was set at the height of 8.3 m with 50° downward tilt. Figure 10(a) shows the color image of a target scene taken by the stereo camera. The person in the image, who is 1.73 m tall, walked from right to left.

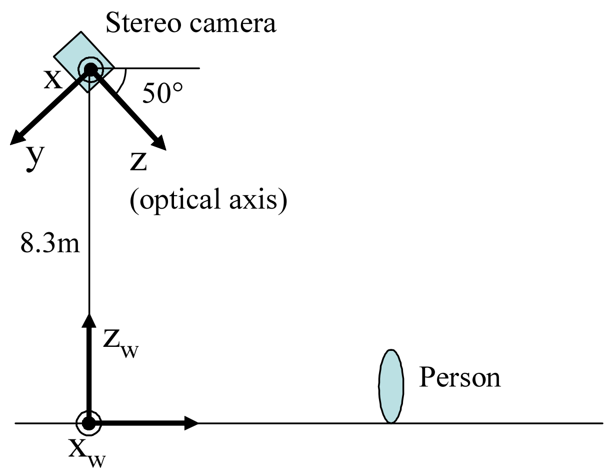


Figure 9. Experimental setup of the outdoor experiments.

Figure 10(b), (c) show the results. Figure 10(b) shows the 3D position in the world coordinate system, and Fig. 10(c) shows the height and width of the person. Since the camera is tilted 50°, the obtained heights are divided by $cos50°$ to compensate for the tilt. The movement of the walking person 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.

6. DISCUSSION

From the experimental results, we think that subtraction stereo works well and is appropriate for applications such as surveillance. It works both for near and far objects, and especially when objects are far, it can detect necessary information appropriately.

Additionally, 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. Figure 11 illustrates the conceptual figure of the stereo camera. The stereo camera should work stand-alone, and should be compact. Pan/tilt mechanism will be effective.

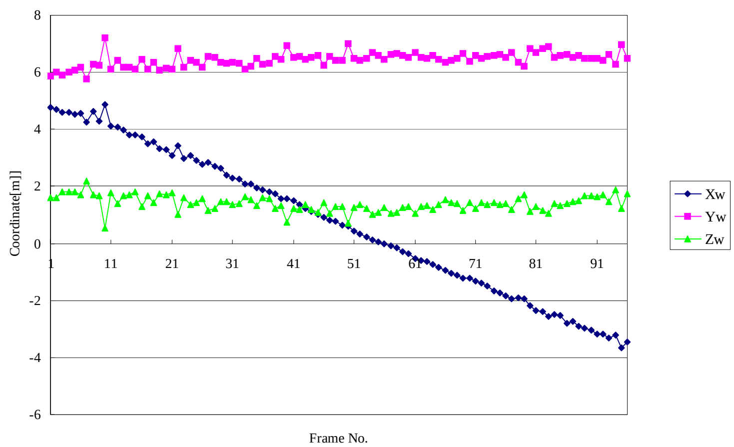
There are many open issues as follows.

- Improvement of the basic algorithm - now the implemented subtraction is quite simple. More effective and robust subtraction method should be adopted.
- Improvement of extracting moving regions - when objects are overlapped or close from the camera, labeling fails.
- Extraction of high-level information of the scene with multiple people
- Construction of the compact, stand-alone stereo camera
- Finding applications for various scales such as indoor, in a building, outdoor, etc.

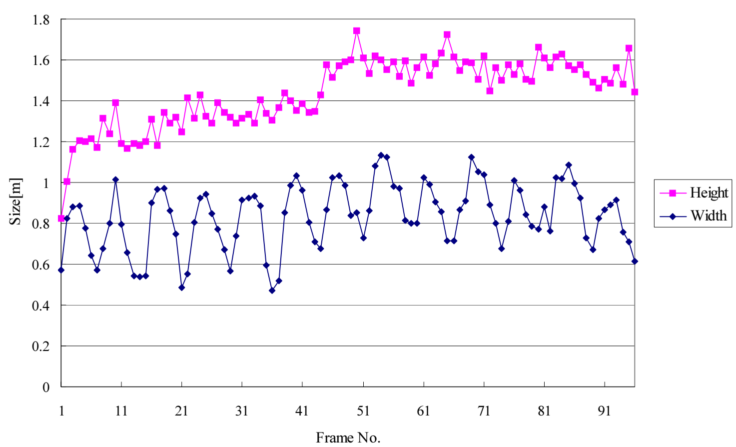
7. CONCLUSION

In this paper, we have proposed the “subtraction stereo” as a practical stereo vision. The method first detects moving regions using subtraction, and then apply stereo matching only to the detected moving regions. Measurement of 3D position, height and width of a moving object was discussed. The basic algorithm of the subtraction stereo was implemented on a commercially available stereo camera, and the effectiveness of the method was verified by several experiments using the stereo camera.

(a) Color image of the experimental scene



(b) 3D Position



(c) Height and width

Figure 10. Experimental result of measuring a walking person in a general outdoor scene.

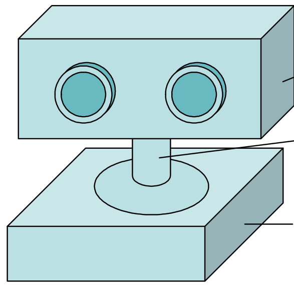


Figure 11. Conceptual figure of our stereo camera.

ACKNOWLEDGMENTS

We thank Yusuke Matsuki and Masaki Wada, former master course students of Chuo University, for their contribution to implementation of the method and experiments.

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*, **1**, pp.102–110, 2000.

[3] M.Z. Brown, D. Burschka, and G.D. Hager: "Advances in computational stereo," *IEEE Trans. PAMI*, **25**, 8, pp.993–1008, 2003.

[4] S.M. Seitz et al.: "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in *Proc. of CVPR'06*, **1**, pp.519–528, 2006.

[5] S. Kagami, K. Okada, M. Inaba, H. Inoue: "Realtime 3D Depth Flow Generation and its Application to Track to Walking Human Being," in *Proc. of 15th International Conference on Pattern Recognition (ICPR2000)*, **4**, pp.197–200, 2000.

[6] T. Ueshiba: "An Efficient Implementation Technique of Bidirectional Matching for Real-time Trinocular Stereo Vision," in *Proc. of 18th International Conference on Pattern Recognition (ICPR2006)*, **1**, pp.1076–1079, 2006.

[7] 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*, **E88-A**, 12, pp.3516–3522, 2005.

[8] Point Grey Research, http://www.ptgrey.com/

[9] Videre Design, http://www.videredesign.com/

[10] K. Hanawa and Y. Sogawa: "Development of Stereo Image Recognition System for ADA," in *Proc. IEEE Intelligent Vehicle Symposium'01*, 2001.

[11] M. Okutomi and T. Kanade: "A Multiple-Baseline Stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **15**, 4, pp.353–363, 1993.

[12] T. Haga, K. Sumi and Y. Yagi: "Human Detection in Outdoor Scene using Spatio-Temporal Motion Analysis," in *Proc. of 17th International Conference on Pattern Recognition (ICPR2004)*, **4**, pp.331–334, 2004.