Measurement of Pedestrian Groups Using Subtraction Stereo

Kenji Terabayashi, Yuki Hashimoto, and Kazunori Umeda

Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
`terabayashi@mech.chuo-u.ac.jp`

Abstract. In this paper, detection of pedestrian groups and counting of the number of pedestrians in each group using "subtraction stereo" are 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. Pedestrian groups are detected with a standard labeling, and three dimensional (3D) features of pedestrian groups are measured from range images obtained by subtraction stereo. Then a method to count the number of pedestrians in a group is proposed. The basic algorithm of the subtraction stereo is implemented on a commercially available stereo camera, and the effectiveness of the method to count the number of pedestrians is verified by experiments using the stereo camera.

1 Introduction

A huge number of studies have been carried out for stereo vision until now [1,2,3,4]. 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] and widely used. There are stereo cameras that are practically used for automotive cars [9].

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 [10,11]. However, a single camera may not be sufficient since the size of a target is not obtained without restriction such that the camera pose is given and the target is on the ground. Stereo cameras are more appropriate, for size information can be directly obtained and scalable use for several scenes becomes possible. It becomes possible to distinguish whether the target is human-size or larger/smaller or to know whether the target person is an adult or a child. For example, it is possible to avoid falsely detecting a bird flying near the camera.

As explained above, we can say that stereo cameras have already reached a practical level. However, stereo vision has a weakness that stereo matching is

inevitably not robust for weak textures or recurrent patterns, which is called the correspondence problem. There have also been many studies that aimed to solve the problem. Muti-baseline stereo [12] 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, though this approach does not work when targets are far. We focus on the movement of objects to make a stereo camera robust and have proposed "subtraction stereo" [13]. In this paper, we choose an issue to detect and measure pedestrian groups, and apply the subtraction stereo to deal with the issue.

This paper is organized as follows. In section 2, we show the outline of subtraction stereo. In section 3, we discuss the detection and measurement of pedestrian groups. Then in section 4, we propose a method to estimate the number of pedestrians in a group. Experimental results to measure pedestrians using the stereo camera with subtraction stereo algorithm are given in section 5. This paper is concluded in section 6.

2   Basic Algorithm of Subtraction Stereo

Fig. 1 shows the basic algorithm of the subtraction stereo. 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 a subtraction method, in which the simplest one is background subtraction. In compensation that the pixels where a disparity is obtained are restricted to moving regions, the subtraction stereo can realize the robustness of the stereo matching by restricting the search space for matching, and by using motion information as well as the original image itself.

Fig. 2 shows an example of the result of subtraction stereo. Fig. 2(b) is the disparity image obtained by the subtraction stereo for the scene Fig. 2(a). Color represents the disparity; bluer color indicates larger disparity, i.e. smaller distance. In contrast to the disparity image obtained by the standard stereo matching Fig. 2(c), appropriate disparity images are obtained for moving objects only.

3   Detection and 3D Measurement of Pedestrian Group Regions

In this section, we explain the measurement of features of pedestrians with the subtraction stereo. We assume a fixed parallel stereo camera in this paper.

3.1   Labeling

In the subtraction stereo, the obtained disparity image is originally restricted to moving regions. Therefore, pedestrian group regions can be obtained with the standard labeling technique.
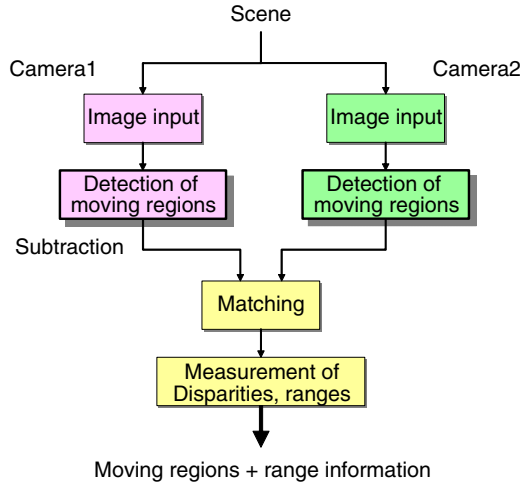
Fig. 1. Flow of subtraction stereo

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 "distance" as the meaning of distance along the optical axis. Let the disparity be k
pixel and the distance be z m. The distance z is calculated by

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

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. As is well known, the distance is inversely
proportional to the disparity.

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

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

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

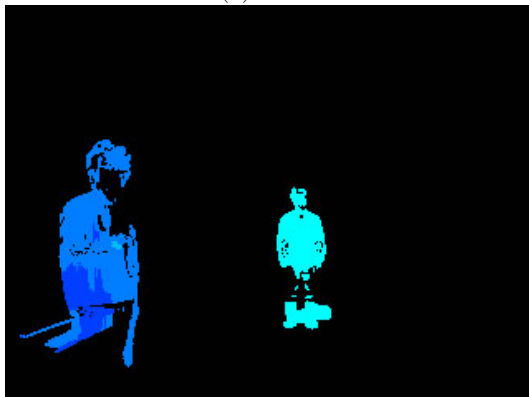## 3.3   Coordinate Transformation

When position and orientation of the camera are given, the 3D position $\mathbf{x}_w$ of
the point in the world coordinate system is calculated as
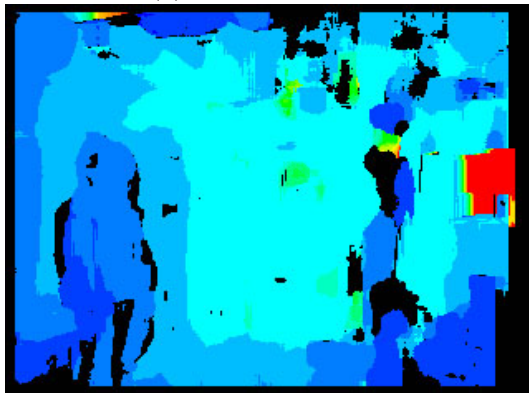
$$\mathbf{x}_w = R\mathbf{x} + \mathbf{t} \tag{3}$$

where $\mathbf{t}$ and R are camera's position and orientation respectively. $\mathbf{t}$ is a 3D vector
and R is a 3×3 rotation matrix.

(a) Scene

(b) Subtraction stereo

(c) Standard stereo

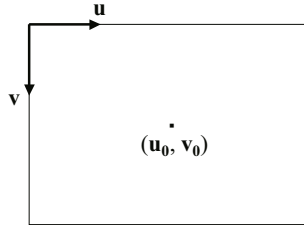**Fig. 2.** An example of subtraction stereo for an indoor scene

Fig. 3. Definition of the image coordinate system

As 3D information of the scene is obtained with a stereo camera, its position and orientation parameters can be measured by observing a scene. When a plane (ground) is observed, its height is measured as a position parameter, and its tilt and roll angles are measured as orientation parameters. It is also possible to obtain positions of pedestrians and use the points in spite of the ground points to calculate the position and orientation parameters of the camera.

3.4   Features of Pedestrian Group Region

It is possible to calculate 3D position of every point in a pedestrian group region using the procedure above. Thinking of rather large measurement errors of stereo vision (which is proportional to the square of distance), we average the distance to each point of the region and use the value z_c as the distance to the region. And by substituting z_c and the image coordinates (u_c, v_c) of the centroid of the region for eq.(2), 3D position $\mathbf{x_c}$ to represent the region is obtained. At the same time, we adopt the area of the region (i.e., the number of the pixels) S as the feature of the region.

4   Estimation of the Number of Pedestrians

In this section, we introduce the method to estimate the number of pedestrians of the detected pedestrian group region.

4.1   Method to Estimate the Number of Pedestrians

The area S and the distance z theoretically satisfies

$$c = S \cdot z^2 \tag{4}$$

where c is a constant. Let c for one person which is obtained by the equation (4) be c_1. If there is no overlap of persons, the number of pedestrians n can be estimated with the following equation.

$$n = \frac{c}{c_1} \tag{5}$$

c_1 can be estimated by calculation, given the size of a person, or it can also be estimated by measuring a person.

4.2   Compensation of the Camera Orientation

The area S is affected by the camera orientation. When the optical axis of a camera is parallel to the ground, the area becomes maximum. A camera is often fixed at a high position with a downward tilt so that the pedestrians are looked down on as illustrated in Fig. 4. In such a case, area S becomes smaller.

When the tilt angle is θ, the area S should be rectified to

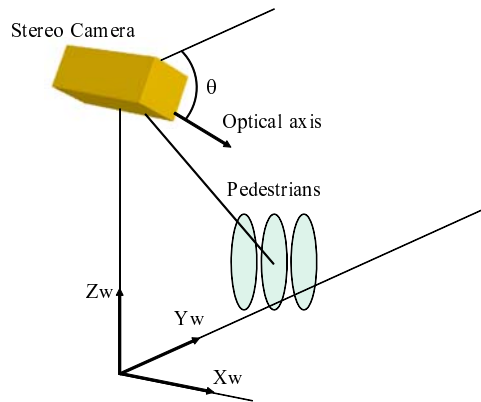$$S' = \frac{S}{cos\theta}. \tag{6}$$

Fig. 4. The relation of the camera and pedestrians

5   Experiments

In this section, we show experimental results to evaluate the proposed method to measure pedestrian groups.

5.1   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 mm). We set the size of the image to 320×240 pixel.

In this implementation, the simple background subtraction is employed 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).

5.2 Experiment to Evaluate the Accuracy of the Method to Estimate the Number of Pedestrians

We evaluated the accuracy of the method to estimate the number of pedestrians. The stereo camera was installed so that the pedestrians were looked down on from a building as shown in Fig. 5. Fig. 6 shows the experimental setup. The camera was set at the height of 8.3 m with 50 deg downward tilt.
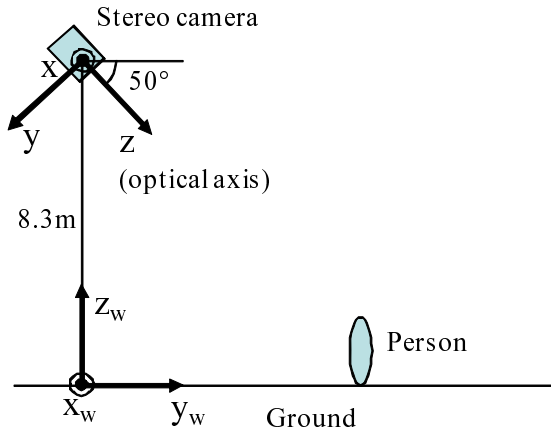

Fig. 5. Experimental scene

Fig. 6. Experimental setup of the experiments

Estimation of the constant. To estimate the constant c_1, we investigated the change of c when a person with a height of 180 cm walked from forefront to back almost vertically to the camera. Fig. 7 shows the result. It is shown that constant c keeps around 140,000 regardless of distance, and therefore, we set constant c_1 to 140,000. If we obtain the distance z and the width p per pixel, the number

of pixels of a region of person in the image coordinate system are obtained from
eq.(2). It means that when the area S' is obtained, then the theoretical value of
constant c is obtained. p of the stereo camera is 7.4 μm. When the theoretical
value of constant c is set to 140,000, it corresponds to the area S' of 0.52 m^2.
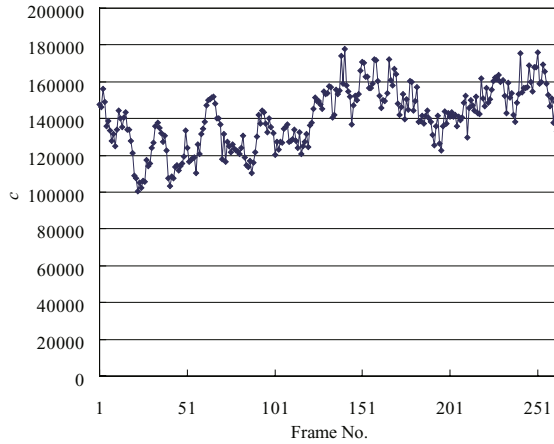This area corresponds to the average width of 29 cm for the height 180 cm.

Fig. 7. Time-series change of c of one person

Examples of estimating the number of pedestrians. We carried out ex-
periments for scenes with different number of pedestrians. Fig. 8, 9, 10 show
the results. In each figure, (a) shows the color image of the scene that is cap-
tured with the stereo camera, and (b) shows the disparity image obtained with
the subtraction stereo and the results to detect pedestrian groups and measure
number of pedestrians of each group from the disparity image. The red rectan-
gles surround detected pedestrian regions, and numbers indicate the estimated
number of pedestrians.

(a) Color image   (b) Result

Fig. 8. An example of detecting pedestrian groups (several groups)

(a) Color image   (b) Result
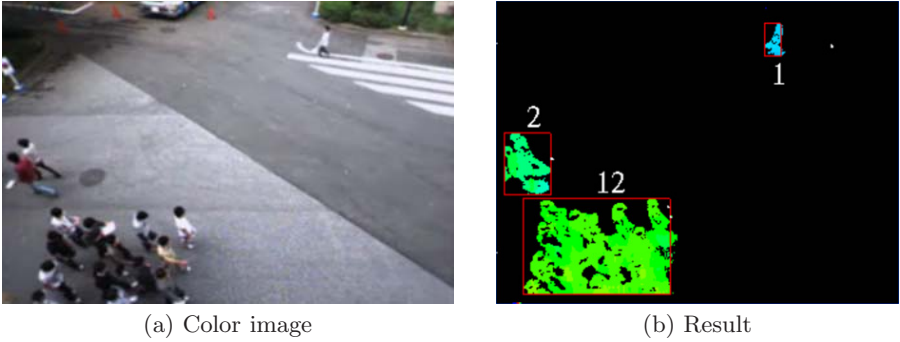
Fig. 9. An example of detecting pedestrian groups (12 persons)

(a) Color image   (b) Result
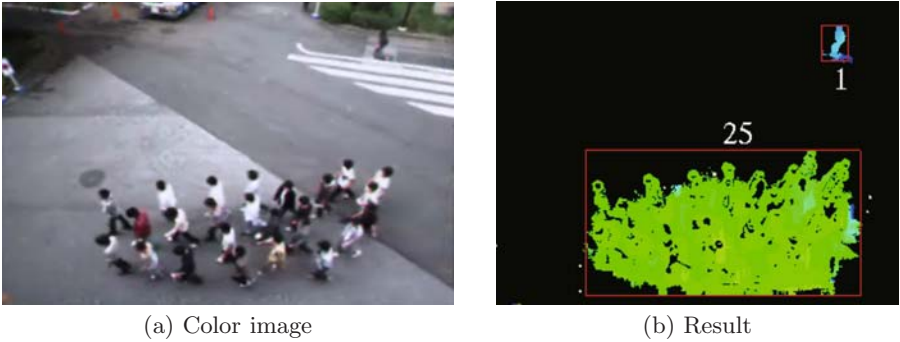
Fig. 10. An example of detecting pedestrian groups (25 persons)

(a) Experimental scene 1 (2 persons)   (b) Experimental scene 2 (10 persons)
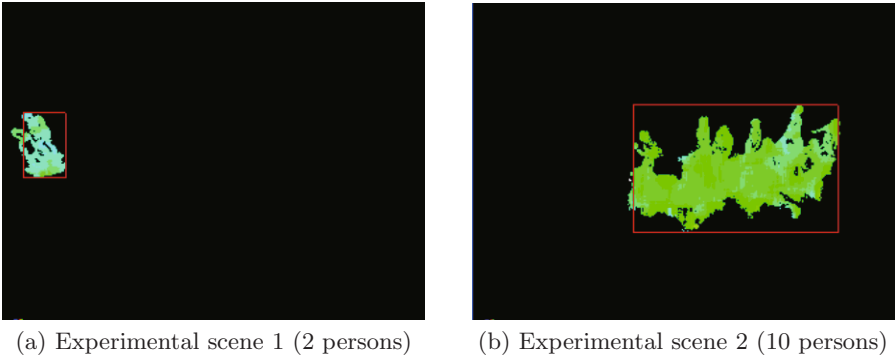
Fig. 11. Disparity image of the experimental scene

The results show that the proposed method works well for several conditions with small (minimum with 1 person) or rather large (maximum with 25 persons) groups; the pedestrian groups are detected appropriately, and the
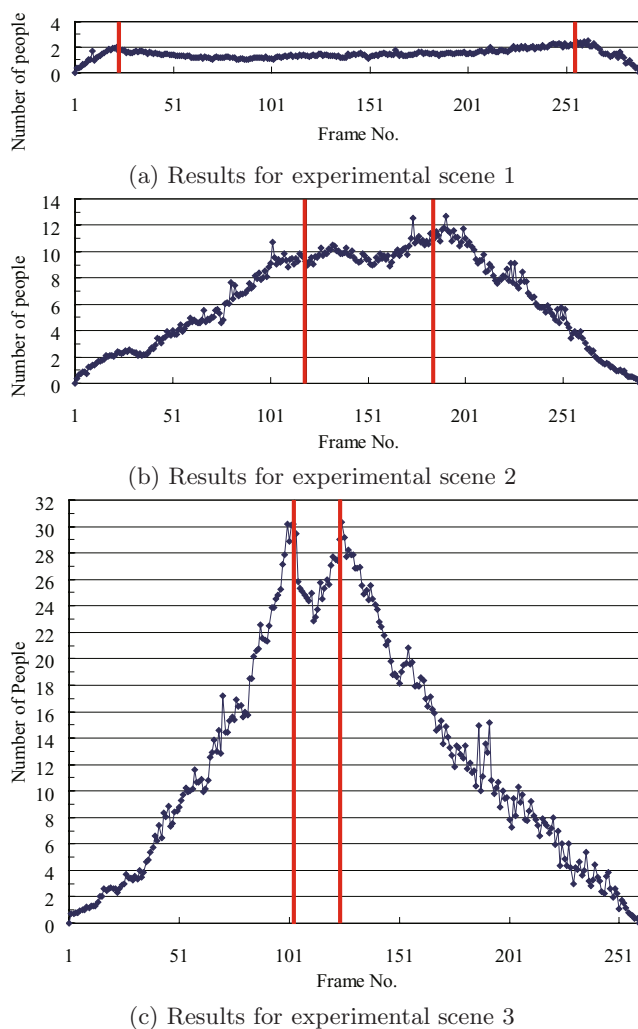
(a) Results for experimental scene 1

(b) Results for experimental scene 2

(c) Results for experimental scene 3

Fig. 12. Experimental result of counting the number of pedestrians

estimated numbers are almost accurate. A video to detect the pedestrian groups and estimate number of pedestrians is attached. As the video shows, the process is realized appropriately in real time.

Evaluation of the accuracy of estimating the number of pedestrians. We carried out experiments to evaluate the accuracy of estimating the number of pedestrians for different number of pedestrians. We applied the proposed method to scenes in which 2, 10, and 25 pedestrians walk from right to left, and estimated the number of pedestrians. Fig. 11(a) and (b) show disparity images of 2 and 10 pedestrians. Again, the red rectangles surround detected pedestrian regions.

A disparity image of 25 pedestrians is given in Fig. 10(b). Fig. 12(a), (b), and (c) show the results to estimate the number of pedestrians for these scenes with 2, 10, and 25 persons respectively. A pedestrian group appeared from the right edge of the image, and disappeared from the left edge of the image. The range between two red lines show the period while every pedestrian was captured by the camera.

From the experimental results, we can see that a certain level of accuracy is realized to estimate the number of pedestrians by the proposed method. Specifically, errors of estimated numbers while every pedestrian was captured are settled as follows. Fig. 12(a) is from -51 to 13 %, Fig. 12(b) is from -11 to 26 %, and Fig. 12(c) is from -8 to 18 %. And at the outside of two red lines in Fig. 12, we can see that estimated numbers change mostly linearly, which are appropriate results.

6   Conclusions

In this paper, we discussed the measurement of pedestrians using subtraction stereo. The method to detect pedestrian groups and estimate the number of pedestrians in each group was proposed. The basic algorithm of the subtraction stereo was implemented on a commercially available stereo camera, and the effectiveness of the proposed method was verified by experiments using the stereo camera.

When the overlapped regions of pedestrians seen from the camera increase, the area becomes smaller and consequently the estimated number of pedestrians becomes less than the actual value. Additionally, as the proposed method uses the area, the estimated number are affected by shadows. We proposed a simple technique to remove shadows in [14], but it is only applicable to separated persons and not applicable to a pedestrian group. Our future works include to deal with these issues. Especially, development of a method to remove shadows in real time is important. And we construct a stereo camera with the function of subtraction stereo and apply it for surveillance applications.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge Univ. Press, Cambridge (2000)
2. Hebert, M.: Active and passive range sensing for robotics. In: Proc. of ICRA 2000, vol. 1, pp. 102–110 (2000)
3. Brown, M., Burschka, D., Hager, G.: Advances in computational stereo. IEEE Trans. Pattern Analysis and Machine Intelligence 25, 993–1008 (2003)
4. Seitz, S., et al.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proc. of CVPR 2006, pp. 519–528 (2006)
5. Kagami, S., Okada, K., Inaba, M., Inoue, H.: Realtime 3d depth flow generation and its application to track to walking human being. In: Proc. of 15th International Conference on Pattern Recognition (ICPR 2000), vol. 4, pp. 197–200 (2000)

6. Ueshiba, T.: An efficient implementation technique of bidirectional matching for real-time trinocular stereo vision. In: Proc. of 18th International Conference on Pattern Recognition (ICPR2006), vol. 1, pp. 1076–1079 (2006)
7. Hariyama, M., Kobayashi, Y., Sasaki, H., Kameyama, M.: Fpga implementation of a stereo matching processor based on window-parallel-and-pixel-parallel architecture. IEICE Trans. Fundamentals, 3516–3522 (2005)
8. Point Grey Research, http://www.ptgrey.com/
9. Hanawa, K., Sogawa, Y.: Development of stereo image recognition system for ada. In: Proc. IEEE Intelligent Vehicle Symposium 2001 (2001)
10. Collins, R., et al.: A system for video surveillance and monitoring. Tech. report CMU-RI-TR-00-12 (2000)
11. Haga, T., Sumi, K., Yagi, Y.: Human detection in outdoor scene using spatio-temporal motion analysis. In: Proc. of 17th International Conference on Pattern Recognition (ICPR 2004), vol. 4, pp. 331–334 (2004)
12. Okutomi, M., Kanade, T.: A multiple-baseline stereo. IEEE Trans. Pattern Analysis and Machine Intelligence 15, 353–363 (1993)
13. Umeda, K., Nakanishi, T., Hashimoto, Y., Irie, K., Terabayashi, K.: Subtraction stereo - a stereo camera system that focuses on moving regions. In: Proc. of SPIE 3D Imaging Metrology, vol. 7239 (2009)
14. Hashimoto, Y., Matsuki, Y., Nakanishi, T., Umeda, K., Suzuki, K., Takashio, K.: Detection of pedestrians using subtraction stereo. In: Proc. of 2nd International Workshop on SensorWebs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2008), pp. 165–168 (2008)