Stereo Surveillance System for Fall Detection

Wei Quan
School of System Design
Tokyo Metropolitan University
Asahigaoka 2-6-6, Hino, Tokyo, Japan
Email: quan-wei1@ed.tmu.ac.jp

Naoyuki Kubota
School of System Design
Tokyo Metropolitan University
Asahigaoka 2-6-6, Hino, Tokyo, Japan
Email: kubota@tmu.ac.jp

***Abstract*—With the technology developing, people enjoy much longer life than ever before. On the other hand, it also bring various problems and inconvenience, one of the most serious is the safety of the elderly people who lives alone. In case of preventing this trouble, we proposed the camera surveillance system for dangerous behavior detecting such as fall down detecting. In this paper, we proposed a new fall down detecting system for stereo version, and it shows much more information than before.**

I. Introduction

As the modern society developing, more and more people enjoy the longer life than ever before. One the other hand, more and more elderly people will live alone and without any first-aid assist. This will cause the safety problem because the situation will be much more serious if the falling down elderly people cannot be given assistance immediately. According to WHO report on global health and aging [2], the percentage of population about elders will increase up to about 17, and this become more serious in Japan already, for this number had already arrived at 14 in 1997(see figure 1, 2). The problems of this population brings could not be ignored.

As the elderly population increase, the referent public service should be proposed for this aged group. One of the popular solutions is to give a healthy guidance system for elderly people, especially for whom are living alone. As we know that the elderly injured people should receive proper health aid as soon as possible, they should get the alert system which can give the alert to others once the accident happen, especially for the elderly who lives alone.

Based on this truth, various kinds of algorithms have been proposed by researchers and engineers. We have also proposed one single camera surveillance system for dangerous activity for example fall detection. Nevertheless, single camera cannot provide sufficient information in some special cases. Based on this truth we proposed advantaged stereo version surveillance system to catch the 3-dimensional position information of the detecting human and make more accurate evaluation.

The following of this paper is organized as follows: section 2 briefly described the fall down detection system and the theorem of stereo geometry, whereas section 3 talked about the proposed system of stereo version, the experiment result showed in section 4, and in the last section, we discussed current research status and gave some extension direction of this system.

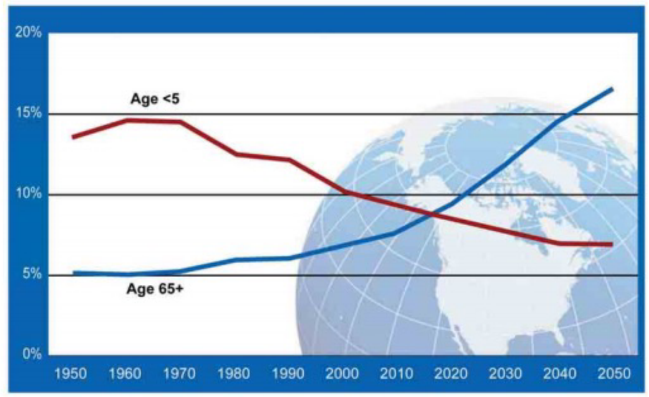

Fig. 1. The growth of population of the people aged above 65 and under 5 from the World Health Organization

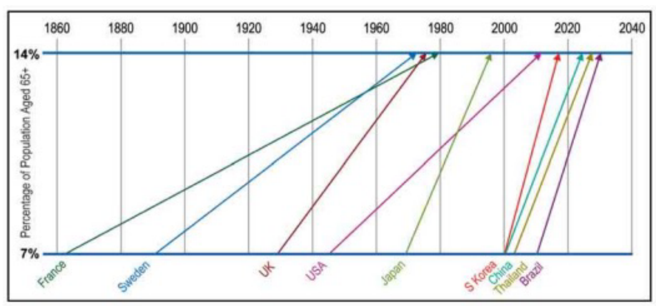

Fig. 2. The growth of the population of the elderly people in major countries

II. Related Work on Fall Detection and Stereo Version

The fall down system has been developed since a long time ago. According to [1] [6], the fall detection system can be broadly categorized into two types: wearable device based system and context-aware system. The first type briefly includes the collaboration between the aware sensor (in most cases the sensor is camera), and the wearable devices. Some of them also incorporate other sensors such as gyroscopes to obtain the information about the target people's position. By wearing these assist device the data will get much more exact and complicated.

And one of the sub-type of the device based system which generally called the ambient-based system has been more and

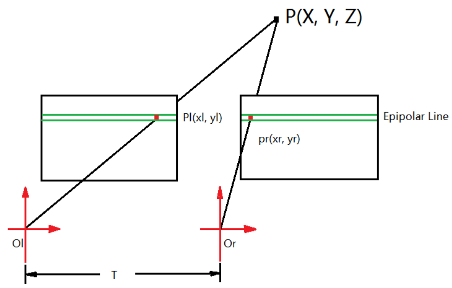

Fig. 3. Explanation of the stereo version

more popular because of the developing of the smart house recently. It combined both audio and video data to evaluate target's current status.

Despite these situations, since fall down system always serve for elderly people, it is a little bit trouble to remind them to wear the device. The detecting result would be imprecise if they forget to wear the device. On the other hand, in most cases the audio data contains a lot of noise which strongly confuses the evaluation. For this reason, we prefer the system with only camera. And we have already made such kind of experiment before.

As the paper we proposed before [2], at first step, we divided the human into head and body, and then detected these two parts respectively. Based on these processing, we can capture the human's status briefly.

Even thought the system with single camera seems much more convenient, it just project object in 3D space into 2D screen, by losing 3D position information. It will make us hard to recognize when in some certain situation. For solving this problem, we just introduce the stereo version system to capture 3D information from the target human.

The theorem of stereo version described in figure 3: the object of real 3D scene will be projected into two screens. nevertheless, the position of projected points in two screen have the difference. This difference of position between two screens is so-called disparity. By introducing 3D geometry, we can evaluate the distance between the camera plane to object as:

$$d(s) = \frac{fT}{x_l - x_r} \tag{1}$$

where f represents the focus length of the cameras, T is for the distance between the focus of the two cameras, and the x^l, x^r are the position of x axis in two camera screens respectively. The detail will describe later.

The formula for calculating the disparity is quite clear, but the process for evaluating these parameters seems much more complicated. We will show how to evaluate these parameters later.

III. Part of Proposed Stereo Surveillance System

The system we proposed process follows as figure 2. There are several main step for construction: Camera calibration; Fundamental matrix calculation; Background construction and foreground detecting; disparity map generation.

A. Camera Rectification and Calibration

The ideal model of camera is pinhole which means that the lens of the camera is just a pinhole and there is no distortion for the camera. Nevertheless, in real situation, the pinhole camera will cost too much time for exposure. Thus they use the glass lens instead it. This just brings the distortion. And because of the structure of the glass lens, the light pass thought the lens will change the direction more or less. This is called radial distortion. And another kind of distortion is called tangential distortion caused by human manufacturing defects. These two kinds of distortion influence the following result and deduce the accurate rate.

$$x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{2}$$

$$y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{3}$$

And for tangential distortion we have

$$x_{corrected} = x + [2p_1 x + p_2(2^2 + 2x^2)] \tag{4}$$

$$y_{corrected} = y + [2p_2 y + p_1(2^2 + 2y^2)] \tag{5}$$

This 5 parameters (r_1,r_2, r_3, p_1 and p_2) is so-called distortion parameters.

And for intrinsic parameters of the camera, b y applying the checkerboard calibration method which proposed by Zhang [3], we get these parameters and rectify images before the other processing.

B. Epipolar Geometry and Fundamental Matrix Calculation

Fundamental matrix is significant for evaluating the relationship between two cameras. It is a necessary step for generating epiline. For any point P which has the position (X, Y, Z), we have two points $p_l = (x_l, y_l)$ and $p_r = (x_r, y_r)$ represent its projection in two camera planes respectively. and p_l p_r follow like:

$$p_l^T F p_r = 0 \tag{6}$$

where F is 3×3 matrix and has the rank of 2. It is so-called fundamental matrix. Because that F is not full rank, for any point in one screen p, there always is a line $l^{'}$ in another screen which satisfy $l^{'} = p\ TF$ (or $l^{'} = Fp$). It represents that for any points in another screen which corresponding to p, it is always on the line l^7. And $l^{'}$ is called the epiline of point p.

By introducing epipolar geometry, we need only search the correspondent points in one dimension, which means that for one point in one camera, we can search the correspondent point in its epiline in another image. It reduce the searching time and improve the accurate rate at the same time. Figure 5 shows the rectified result.

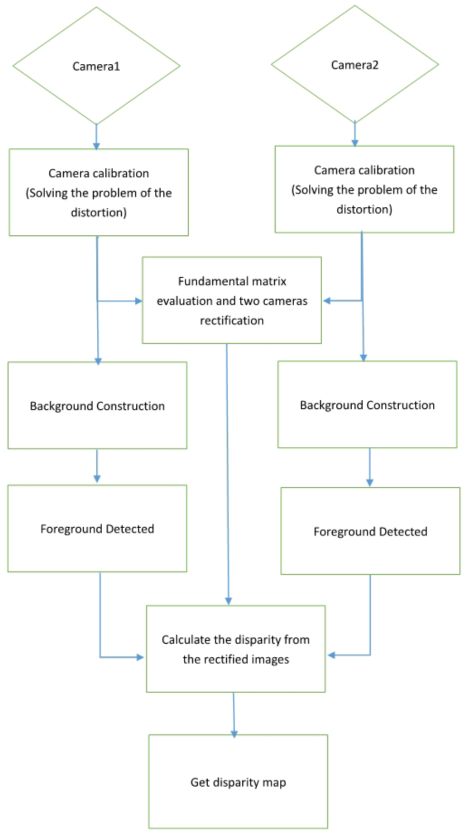

Fig. 4. Processing flow of proposed stereo version system

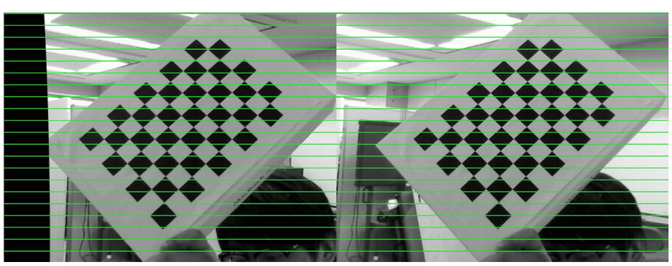

Fig. 5. Rectified left and right image by using checkerboard method

C. Background Construction and Foreground Detecting

Since our target is to build a real-time system, we need to exclude the unnecessary information and focue on the interested only. By building background model we can extract the foreground in a short time. There are various kinds of algorithm for foreground detecting and we are not going to give the detail. For the consideration of computing time and accuracy, in this system we chosed the CodeBook algorithm [4]. The result shows in figure 6.

D. Disparity Map Generation

The disparity map or depth map generation is the most significant part of the whole system. In disparity map, every point has the corresponding points in both two camera screens, and the density value of each point represents the distance from it to the camera plane. By generating disparity map we will use the following methods to evaluate current status of the target human.

Since that we are building real-time system, it is significant to make sure the applied algorithm fast and satisfy the necessary accurate at the same time.

IV. Proposed Fall Detection Method

So far we discussed the theorem of the system, regardless the problem of fall down detection. In this section we will introduce the fall down detection algorithm under the o

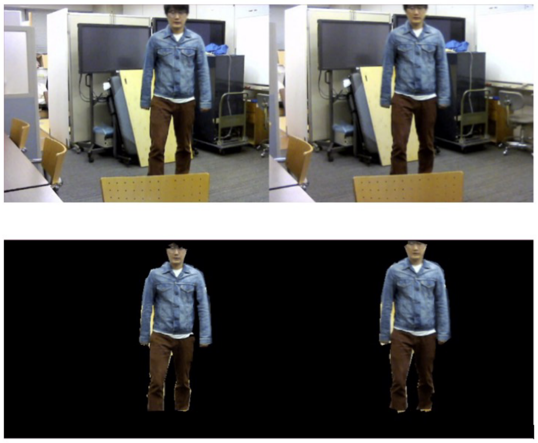

Fig. 6. Result of foreground extraction on left and right camera.

Fig. 7. Stereo cameras constructed by two web cameras

version. From the disparity map we've already generated, the evaluation of fall detection have been applied on it. In this paper, we just combined two methods to evaluate the fall status: aspect ratio and gravity variance.

A. Aspect Ratio

Aspect ratio as one of the basic algorithm for fall detection has been applied in many kinds of fall down detection systems [5]. As human detected in the screen, the aspect ratio of him/her is

$$r = \frac{h}{w} \tag{7}$$

where h and w represent the height and width of the rectangle of target human respectively. If the ratio r is larger than a specified threshold, this person will be regarded as fall down (on the ground). Combined with the speed of falling down, the system can evaluate the status that whether this person is just want to lie or be the accident in some certain situation. It works well with the situation that the human lie parallel to the camera plane.

B. Variance of Disparity

Despite the situation which can be evaluated above, there is another situation that the people fall towards or back to the camera plane. This will cause less change of aspect ratio. For this reason, if we want to lower the threshold for adjusting this situation, it will be much more sensitive for the situation described above. In this situation, we just combined the advantage of stereo version, i.e. the information of the disparity.

Suppose the point p_i ($i = 1...N$) is one of the disparity points that constructing the human shape. We can get the centroid of disparity of the human by

$$c_t = \frac{\sum_{i=1}^{N} w_i p_i}{N} \tag{8}$$

represents the centroid point of the target at time t, where w_i is the weight for every point. In this case the weight is average of $\frac{1}{N}$. And we can get the variance from the centroid by

$$v_t = \frac{\sum_{i=1}^{N} w_i (C - p_i)^2}{N} \tag{9}$$

In normal situation, because that the human stand up or sit on the chair, the disparity of the whole body keeps in the similar value, thus keeps variance in a small value which close to 0. Nevertheless, once fall down accident, v_t changes a lot in a short time.

$$f'_{v_t} = \Delta t = \frac{d(v_t)}{d(t)} \tag{10}$$

If f'_{v_t} at time t is larger than a threshold, it indicates that the disparity of the whole target changing largely, and the fall down toward the direction of normal line of the camera plane will be evaluated.

V. Experiment

In order to make sure the system more suitable, we just chose two normal web camera for the whole system. The type of the camera is Buffalo BSWD06M web camera, the price for each camera is as cheap as 1,000 JPY. We just parallel these two cameras to construct the stereo cameras as figure 7 shows. And in this experiment we ignored sound information.

We first tested the result of the disparity map generation. It showed the different result regarding the different distances. The closer from the target to the cameras, the darker the disparity will be.

From figure 8 we can get the information that the closer from target to the camera, the average disparity value tends to be lower when the target getting closer to the camera. As the disparity map shows, the disparity map could not as accurate as original image, but it is sufficient for us to apply to the referent research.

The fall down detection can be seen in chat of figure 9. These two charts show the change of average disparity value(ADV) and variance of disparity value(VDV). In the first

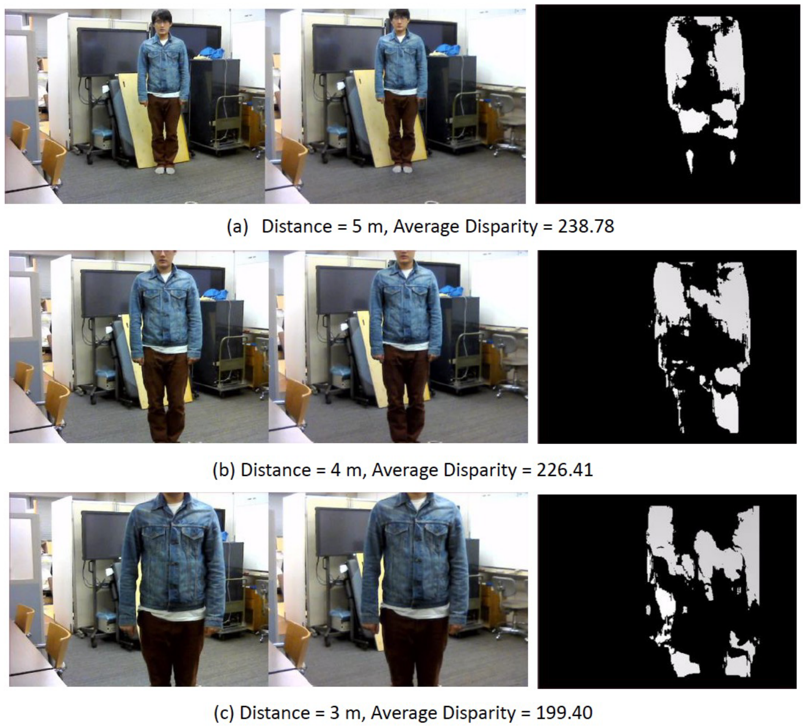

Fig. 8. Experiment result for disparity evaluation. The first two columns represent images of left and right camera, the last column illustrate the disparity map of the referent pair.

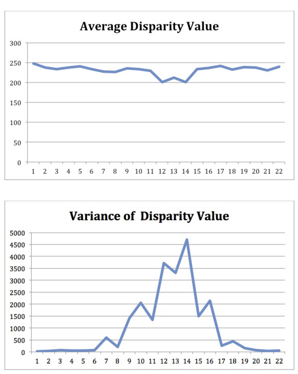

Fig. 9. average disparity value and variance of disparity value during the target's fall down posture. The ADV changed little while VDV changed a lot when the target person falling down

chart we can conclude that the ADV changed only a few since the target fell down on the ground, whereas in the second chart the VDV changed obviously. By defining a certain threshold we can evaluate the current status of target human.

VI. Conclusion

In this paper we proposed the fall detection surveillance system combined with stereo version. The system can be real-time detect, and the accurate rate can satisfy most of the situation. On the other hand, the result still need to be improved.

Based on this system, various kind of other functions can be put in because of the simple structure and good performance of this system. It also has ability to provide much more information such as the 3D construction, and the disparity sequence generation. We would focus on this theme and make the further research on it.

On the other hand, the part of disparity generation still need to be improved. We are currently focus on the efficiency, which make the accurate a little low comparing with other image disparity map generation(generating the disparity map for image only). In the future, we will focus on how to improve the performance and make the system much more robust.

References

[1] Muhammad Mubashir, Ling Shao, LukeSeed: A survey on fall detection: Principles and approaches. J. Neurocomputing, Volume 100, 16 January 2013, Pages 144152.

[2] World Health Organization: World Report on Global Health and Aging. http://www.who.int

[3] May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)

[4] Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)

[5] Miaou SG, Sung PH, Huang CY, : A customized human fall detection system using omni-camera images and personal information, Proceedings of the 1st Distributed Diagnosis and Home Healthcare Conference. Arlington: Institute of Electrical and Electronics Engineers (2006).

[6] Igual,R., Medrano,C., and Plaza,I.: Chanllenges, issues and trends in fall detection systems In: Biome. Eng. Online, 2013.