COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 114 (2014) 172–182

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Fall detection for multiple pedestrians using depth image processing technique

*Shih-Wei Yang**, *Shir-Kuan Lin*

Institute of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 29 September 2013
Received in revised form
28 January 2014
Accepted 5 February 2014

Keywords:
Fall detection
Depth image analysis
Multiple pedestrian detection
Illumination compensation

ABSTRACT

A fall detection method based on depth image analysis is proposed in this paper. As different from the conventional methods, if the pedestrians are partially overlapped or partially occluded, the proposed method is still able to detect fall events and has the following advantages: (1) single or multiple pedestrian detection; (2) recognition of human and non-human objects; (3) compensation for illumination, which is applicable in scenarios using indoor light sources of different colors; (4) using the central line of a human silhouette to obtain the pedestrian tilt angle; and (5) avoiding misrecognition of a squat or stoop as a fall. According to the experimental results, the precision of the proposed fall detection method is 94.31% and the recall is 85.57%. The proposed method is verified to be robust and specifically suitable for applying in family homes, corridors and other public places.

1. Introduction

The accidental fall is a common high-risk injury, especially with regard to the elderly and the physically disabled. Therefore, communities for the elderly, hospitals and other public places are becoming required to install fall detection systems. The most common fall detection techniques can be divided into two types. The first type is such that the pedestrian wears at least one sensor which gathers the information related to changes in the posture of the user. Then, the fall event is judged according to the acquired information of wearable sensors [1–6]. Abbate et al. [7] designed a fall detection system embedded in a smart phone, and thus the smart phone could detect walking, falling and lying down events. However, the wearable sensors may inconvenience the movement of the elderly and the physically disabled. Moreover, if the user forgets to wear the sensors, these methods are completely ineffective.

The second type is the visual surveillance. The two-dimensional (2D) images are captured by a monocular camera, and are used for detecting whether the pedestrian has fallen through the image processing technique [8–13]. Liao et al. [14] combined the features of pedestrians with a Bayesian Belief Network model to detect slip-only events and fall events. However, when more than two pedestrians are partially overlapped, the above 2D methods are likely to cause misjudgments or even completely inapplicable. The depth camera systems [15–19] can be used for solving the occlusion problem. The stereo vision techniques of multi-cameras [16,17] have been proposed for fall detection, but the resulted computation is often huge. The fall detection system based on a Time-of-Flight sensor [18] provides the precise depth information and the guaranty of people privacy, while the system

* *Corresponding author*. Tel.: +886 3 5712121x54423.
E-mail address: swyang.nctu@msa.hinet.net (S.-W. Yang).

http://dx.doi.org/10.1016/j.cmpb.2014.02.001
0169-2607/

setup is very expensive. Rougier et al. [19] obtain the pedestrian centroid height relative to the ground by a structured light sensor (Microsoft Kinect [20]) for detecting falls, but the proposed algorithm is only suitable for detecting a single person and cannot recognize a fall from brutally sitting down or squatting.

Therefore, this study proposes a fall detection technique using the Kinect sensor, which is low cost, so as to improve the above drawbacks. By analyzing the captured depth images, the proposed method identifies a single pedestrian or mutually overlapped pedestrians, and determines whether each pedestrian has fallen. As different from the ellipse fitting technique [14,21], a simplified algorithm of human shape analysis for calculating the pedestrian tilt angle is proposed. In addition, some suitable solutions which can be integrated with the proposed method to enhance the people privacy are also discussed in Section 2. According to the experimental results, the proposed method can distinguish falls, squats and other highly similar motions, thus, avoiding the misjudgments effectively.

2. Methods of fall detection for multiple pedestrians

2.1. *Foreground extraction and overlapped object segmentation*

As the position of the depth camera is fixed, a depth image of the background is built first, so that each of subsequent depth images can be subtracted from this background image. This allows the system to segment the foreground and background, and the segmentation results can be expressed as a binary image (Fig. 1):

$$g_{xy} = \begin{cases} 255, & \left| z_{xy} - z_{xy(b)} \right| \geq T_b \\ 0, & \left| z_{xy} - z_{xy(b)} \right| < T_b \end{cases} \tag{1}$$

where g_{xy} is the gray value of pixel (x,y) in the binary image; $z_{xy(b)}$ is the depth value of pixel (x,y) in the background image; z_{xy} is the depth value of pixel (x,y) in the ith frame; T_b is the binary threshold of depth value.

The segmentation results of Eq. (1) are more accurate than using the color image; the error resulting from the foreground and background colors being too similar can be avoided.

There may be overlapped objects in any isolated area of the foreground, as shown in Fig. 1 (point cloud data). Therefore, the concept of clustering is used to segment overlapped objects:

Step 1 The depth value of every pixel in the isolated area is sorted in ascending order to obtain a depth sequence. The differences $D_{21}, D_{32}, \ldots, D_{i(i-1)}, \ldots, D_{n(n-1)}$ between two consecutive depths in the sequence are calculated, where n denotes the total number of pixels in this isolated area.

Step 2 Ignore the zero differences, and let the average value μ of the minimal non-zero difference and the maximal non-zero difference be the initial threshold. All the non-zero differences are divided by μ into two groups. Those smaller than μ are the first group and the average value μ_1 is calculated; those greater than or equal to μ are the second group and the average μ_2 is calculated. Last, the new threshold $\mu_{new} = (\mu_1 + \mu_2)/2$ is calculated. If μ is equal to μ_{new}, μ is the depth threshold for segmenting overlapped objects; otherwise, let $\mu = \mu_{new}$ and repeat this step until μ converges.

Since there may be a slight depth gap in the pixels of a single object, a threshold T_c is set, and $T_c = 30$ (mm) in this paper. T_c denotes the minimum distance between two overlapped objects that need to be segmented. If $\mu < T_c$, it is not necessary to segment this isolated area. However, if $\mu \geq T_c$ and $D_{i(i-1)} \geq \mu$, as shown in Fig. 2(a), this isolated area contains two overlapped objects. The pixels with a depth value less than or equal to the $(i-1)$th depth value in the sequence are regarded as one object, while the pixels with depth value greater than or equal to the ith depth value in the sequence are regarded as another object. The segmentation of overlapped objects is shown in Fig. 2 (b). This method is also applicable to the segmentation of more than three overlapped objects.

2.2. *Human object detection*

After segmentation, if any object has enough skin colored pixels, the object is recognized as a pedestrian. To avoid the result of skin color detection being influenced by a non-white light source, the illumination compensation should be taken before skin color detection. According to the Gray World Assumption [22,23], the average values of R, G and B components of an image without color deviation should be very close to each other. Therefore, if the image fails to meet the above condition, the R, G and B components of every pixel are corrected to:

$$R'_{xy} = \begin{cases} 255, & R_{xy} \geq R_{5\%} \\ R_{xy} \times \dfrac{255}{R_{5\%}}, & R_{xy} < R_{5\%} \end{cases} \tag{2}$$

$$G'_{xy} = \begin{cases} 255, & G_{xy} \geq G_{5\%} \\ G_{xy} \times \dfrac{255}{G_{5\%}}, & G_{xy} < G_{5\%} \end{cases} \tag{3}$$

$$B'_{xy} = \begin{cases} 255, & B_{xy} \geq B_{5\%} \\ B_{xy} \times \dfrac{255}{B_{5\%}}, & B_{xy} < B_{5\%} \end{cases} \tag{4}$$

where R'_{xy}, G'_{xy} and B'_{xy} are the R, G and B components of pixel (x,y) after compensation; R_{xy}, G_{xy} and B_{xy} are the R, G and B components of pixel (x,y) before compensation; $R_{5\%}$, $G_{5\%}$ and $B_{5\%}$ are the lower bounds of the top 5% R, G and B components of the image.

The silhouette of each object is corresponded to the color image after illumination compensation, and the nonlinear skin color detection method from Ref. [23] is used to calculate the number of skin colored pixels of each object (see [23] for details). As long as an object has a skin colored area of 10% or over, it is recognized as a pedestrian and labeled with