# 3D depth image analysis for indoor fall detection of elderly people

CrossMark

**Lei Yang**[a,b,*], **Yanyun Ren**[a], **Wenqiang Zhang**[a]

[a]*School of Mechatronic Engineering and Automation, Shanghai University, Yangchang Road, Zhabei district, Shanghai 200072, China*
[b]*School of Computer Science and Electrical Engineering, University of Essex, CO4 3SQ, United Kingdom*

**Abstract**
This paper presents a new fall detection method of elderly people in a room environment based on shape analysis of 3D depth images captured by a Kinect sensor. Depth images are pre-processed by a median filter both for background and target. The silhouette of moving individual in depth images is achieved by a subtraction method for background frames. The depth images are converted to disparity map, which is obtained by the horizontal and vertical projection histogram statistics. The initial floor plane information is obtained by V disparity map, and the floor plane equation is estimated by the least square method. Shape information of human subject in depth images is analyzed by a set of moment functions. Coefficients of ellipses are calculated to determine the direction of individual. The centroids of the human body are calculated and the angle between the human body and the floor plane is calculated. When both the distance from the centroids of the human body to the floor plane and the angle between the human body and the floor plane are lower than some thresholds, fall incident will be detected. Experiments with different falling direction are performed. Experimental results show that the proposed method can detect fall incidents effectively.

## 1. Introduction

It is estimated that 17% of the population in China will be over 60 by 2020. The health care of the elderly people living alone is becoming an important problem. One of the risks that the elderly people living alone face is falling down. For instance, there are more than 1.6 million elderly people who suffer fall-related injuries each year [1]. The fall is caused by the fragile bodies of the elderly people and the

*Corresponding author at: School of Mechatronic Engineering and Automation, Shanghai University, Yangchang Road, Zhabei district, Shanghai 200072, China. Tel./fax: +86 21 56331443.
*E-mail address:* yangyoungya@sina.com (L. Yang).

potential surrounding fall hazards (slippery floors, poor lighting, unstable furniture, obstructed ways, etc.) in home environments. If falls can be noticed in time, fall injuries and the associated costs can be reduced dramatically [2].

Most existing fall detection and alarm systems can be classified into three categories: wearable sensor-based, ambient sensor-based, and computer vision-based methods. Wearable sensor-based methods usually rely on accelerometer sensors that are attached to the subject's body, which has high level of obtrusiveness [3]. Ambient sensor-based fall detection systems use external sensors embedded in the environment, including the pressure sensor, acoustic sensors, electromyography sensors and so on [4,6]. With the development of computer vision in recent years, image and video-based methods have become popular in fall detection systems [7,8]. This kind of method is non-obtrusive, and convenient for elders.

In the visual fall detection systems, the status of the scene of the elders can be delivered to an appropriate destination if a fall is detected and alarm signals are triggered. Fall detection systems usually focus on a single elder in practice. If the situation contains two or more people, segmentation and marking module will be used to separate each person and detect and track them individually [9]. However, the fall can be delivered to the destination by the other person. So our work deals with the fall detection of the single elderly person in the home environments.

Since 2D grey or colour images are the projection of 3D targets, the problem of appearance deformation may occur in the fall detection [10]. In order to deal with the problem, this paper presents a new fall detection method based on analyzing the shape in depth images captured by the Kinect sensor. It can be described as follows: 1) depth images are pre-processed by a median filter for both target images and background images; 2) the silhouette of the moving individual in depth images is achieved by the subtraction method of background frames; 3) the floor plane equation is estimated by the least square method and a disparity map converted from the depth image; 4) shape information of the human body in depth images is analyzed by a set of moment functions, and the coefficients of ellipses are calculated to determine the direction and position of the individual. The centroids of the human body and the angle between the human body and the floor plane are further calculated for fall incident detection. When both exceed some thresholds, fall incident will be detected.

The rest of this paper is organized as follows. Section 2 reviews the relative works in the area of fall detection of the elderly people. Section 3 describes the proposed fall detection method in detail. Experimental results are presented in Section 4 to show the feasibility and performance of the proposed method. Finally, a brief conclusion is drawn in Section 5.

## 2. State-of-the-art

This section reviews three categories of sensors adopted in fall detection and two classes of fall detection methods. Applications of depth image and moment function in fall detection are also discussed.

### 2.1. Sensors adopted in fall detection

Fall detection systems have been designed by using either external sensors or wearable sensors. External sensors are deployed in the vicinity of the body in concern, and wearable sensors are connected to the subject of interest (SOI). Most existing fall detection systems can be divided into three categories: wearable sensor-based, ambient sensor-based, and computer vision-based systems [2,11].

Accelerometers are the most common wearable sensor used today. They are small, cheap and can easily be placed on any part of the body [12], which are an alternative to external sensing [2]. The wearable sensor-based approach relies on embedded sensors to detect the movement and position of the subject [11]. Accelerometers are a type of wearable sensors and they are widely used in fall detection systems [13]. There are several main wearable sensors-based methods in the current research, including accelerometer [11], fusion of accelerometer and posture sensors [14], inactivity with accelerometer [15], tri-axial accelerometer [11] and posture-based method [16]. The advantage of wearable sensors is they are generally less expensive than external sensors. However, the main drawback of wearable sensors is high drift.

Ambient sensor-based fall detection methods focus on how to embed sensors into the environment, and how to track the elderly person's movements. Usual features for fall detection and fall tracking are pressure, vibration, sound, infrared array, and so on [2,10]. Zigel et al. proposed a fall detection system based on ground vibration and sound sensors [4]. Infrared arrays can improve the accuracy and efficiency in fall detection [17]. Pressure sensors are commonly used because of their low cost and non-obtrusiveness, which is based on sensor pressure changes. The main disadvantage is the low detection precision of these sensors (below 90%) [2]. Since the pressure is sensed all around the object, they may generate false alarms. In addition, the distance has a direct impact on the accuracy of detection. Even though the sensor is very sensitive, the accuracy will decline if the fall occurred 5 m away [16] In order to obtain higher accuracy, the room should contain more sensors.

In the past 10 years, there have been great advances in computer vision and video cameras. It opens up a new branch of methods for fall detection. In general, fall accidents happen in less than a second of time, usually between 0.45 and 0.85 s [2,18], during which the falling people greatly change in posture and shape. These sudden changes are crucial to determine whether the fall is happening. Compared with other methods, vision-based methods are more robust and are less intrusive [10]. At the same time, more and more cameras are used in a family's daily life. This creates favourable conditions for the application of vision-based methods. Vision-based fall detection systems monitor the position and shape of the subjects which benefit from techniques of pattern recognition and image processing methods.

Vision-based fall detection methods can be broadly divided into three kinds: methods using a single RGB camera, 3D-based methods using multiple cameras, and 3D-based methods that use depth cameras [8,10,18]. Fall

detections using a single RGB camera have been widely studied since the system is easy to set up and inexpensive. Shape relative features of human motion analysis and inactivity detection are usually used as clues for detecting falls. Anderson et al. modelled the proportion of the border around the contour shape with a trained Hidden Markov Models (HMM) [19]. However, this method does not deal with the fall in the optical axis direction and cannot distinguish a fall from other similar activities. For 3D-based methods using multiple RGB cameras, calibration of multi-camera systems allows 3D reconstruction of the object. However a careful and time-consuming calibration process is required. The time-of-flight 3D camera is the earliest depth camera used for fall detection [20]. With the help of the depth cameras, it is simple to calculate the distance from the top of the person to the floor [21,22].

## 2.2.  Moment functions and depth image in fall detection

Moment functions have been adopted in computer vision applications such as shape analysis, shape deformation description, postures estimation, and so on. As for fall detection in computer vision, several groups of researchers have noticed the roles of moment functions for shape and posture description. By using the conception of integrated normalized motion energy image, silhouette motion is modelled for fall detection, and shape deformation quantified from the silhouettes is represented with ellipse fitting, which was determined by a set of moment functions [8]. Information from the ellipse fitting of shape description together with position information is used to provide features to describe the extracted posture silhouettes. Then a One Class Support Vector Machine (OCSVM) Method and Multi-class SVM Classification are applied to distinguish normal daily postures and abnormal postures such as falls [23,24]. Based on the concept of the integrated spatiotemporal energy (ISTE) map, a Bayesian Belief Network Method is proposed for fall detection by using human object shape analysis with moment functions in [25]. This fall detection system is based on an ingenious combination of skeleton features and human shape variations, in which an ellipse is used to approximate the human shape instead of a bounding box [26]. From these works, we noticed that moment functions are powerful tools to describe the human shape which is approximated by an ellipse. An ellipse can provide more precise shape information than a bounding box. So ellipse approximation is adopted in our work. However, the above works are mainly based on grey or colour images.

Although vision-based approach is the most natural way to detect falls, the problem of the projection remains a challenge in the fall detection method with 2D grey or colour images. A monitoring environment with multiple cameras was created in order to obtain the 3D contour representation of human body from multiple cameras [20]. Depth cameras have been used in fall detection more recently because of the merit of the 3D data. The tracked key joints of the human body are analyzed using a depth camera [21]. Based on analyzing a person's vertical state in individual depth images, a decision tree model is

established to compute a confidence of a fall [22]. By using the central line of a human silhouette to obtain the pedestrian tilt angle, depth image is analyzed for fall detection [27]. In 2013 and 2014, fall detection methods using Kinect's depth image sensor are introduced separately in [28,29]. Inspired by above works on shape analysis and the merit of depth image, a new fall detection method is proposed in this paper based on shape analysis of 3D depth images.

## 2.3.  Classification of fall detection methods

People in their daily lives will do variety of different positions and movements. Pattern recognition and classification is an important sort of fall detection method for distinguishing falls from normal activity [2,18]. Feature extraction and selection are the processes by which relevant characteristics or attributes are identified from the collected data [30]. Decision Trees (DTs) are one of the oldest algorithms used in the problem of pattern classification such as falls [21]. Artificial Neural Networks (ANNs) are used to classify falls from daily activities [31]. Different postures are classified via a multi-class Support Vector Machine (SVM) to verify the detected falls [23]. A Bayesian Belief Network (BBS) is used to model the causality of the slip or fall events with other events [25]. Two stages, namely learning and detection, are usually included in those classification methods of fall detection. However, all postures in fall activities cannot be included and learned, no matter how much training data has been used. So threshold methods are a natural choice.

Falls can be easily detected if the depth blob associated with a person is near the floor. However it is difficult to choose different effective features. Head-ground distance gap and head-shoulder distance gap are used as features to detect falls by some thresholds [32]. If the orientation of the ellipse of the human object and the motion of the human object are smaller than the threshold for a period of time, a fall will be confirmed [26]. This work will detect falls by analyzing the distance from the centroids of the human body to the floor plane and the angle between the human body and the floor plane, and it is threshold-based method.

## 3.  The proposed method

In this section, the details of the proposed fall detection method are described. A flow chart will be described briefly. The main processes of the method are then discussed exhaustively.

## 3.1.  Fall detection procedures

In the proposed method, 3D depth images of the room environment and target frame stream containing the elderly individual are captured by a Kinect sensor. Then both are pre-processed by a median filter. Also, the silhouette of the moving individual in each depth image is obtained by background frames subtraction technique. Then the floor plane is estimated by V disparity map and the shape characteristics of the moving individual are analyzed. The

depth images are converted to a disparity map, and the disparity map is obtained by the horizontal and vertical projection statistics histogram. The initial floor plane information is obtained by a V disparity map, and the floor plane equation is estimated by the least squares method. Shape characteristics of the moving individual are described by an ellipse which can be calculated by a set of moment functions. Centroids of ellipses and the angles between the ellipses and the floor plane are calculated. The distances and the angles are then translated to the world coordinate system from the imaging plane coordinate system. In the detection process, when the distance from the centroids of the ellipses to the floor plane and the angle between the ellipses and the floor plane are lower than some thresholds, a fall incident will be detected. Figure 1 shows the flow chart of the proposed fall detection method.
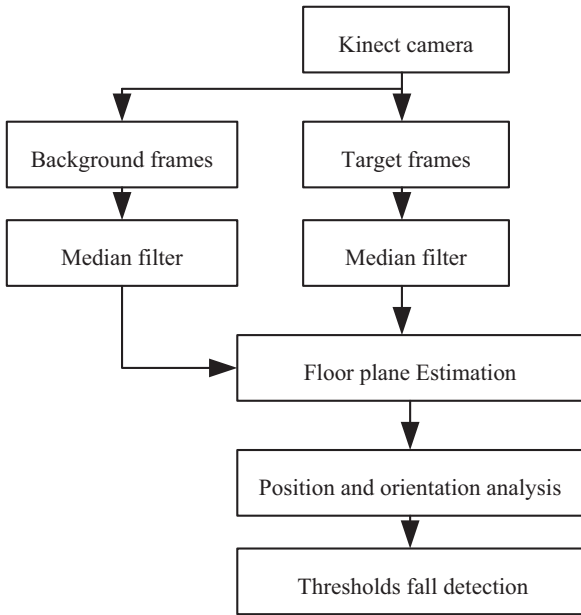


**Figure 1** Flow chart of the proposed fall detection method.

## 3.2. Pre-procession and foreground extraction

The working principle of the Kinect depth sensor is to actively project near infrared spectrum by an infrared projector. When the infrared rays radiate to rough objects, the spectrum is distorted, and form some random reflection spots. Its infrared camera captures these changes in the reflected infrared spectrum. Therefore, in the practical measurement process, speckle noise is produced because of the high coherence of the infrared light from the active projection. This process leads to the loss of the edge information of the object in the image. If an object is transparent as glass medium, the information is seriously missing.

In order to reduce the noise of the depth image, we only deal with the data in the trusted range of the Kinect, i.e. 1.2–3.8 m such as:

$$depth(i,j) = \begin{cases} NaN & depth(i,j) < 1200 \\ depth(i,j) & 1200 \le depth(i,j) \le 3800 \\ NaN & depth(i,j) > 3800 \end{cases}$$

$$(1)$$

where $depth(i,j)$ denotes the depth of the point $(i,j)$ corresponding to the depth of the image. *NaN* denotes that the value is undetermined or unknown, which will be automatically skipped in the programme processing in *Matlab*.

Then the Median Filter Method is adopted for both background images and target images. For a pixel point $(i,j) \in [2, M-1] \times [2, N-1]$ inside the depth image, a two dimensional nine point median filter is used to process the depth image as follow:

$$z(i,j) = median\{depth(m,n), m = i-1, i, i+1; n = j-1, j, j+1\}$$

$$(2)$$

where the resolution of depth image is $M \times N$.

For the indoor environment where the elderly people are living alone, the depth image of the background is generally unchanged or changed a little. Therefore, the background
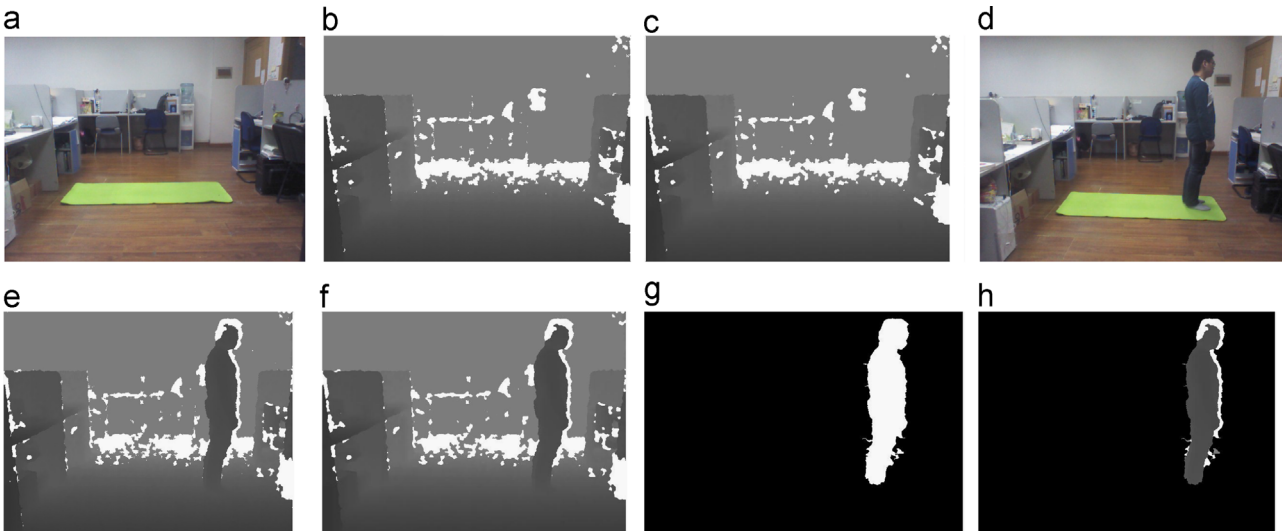


**Figure 2** Pre-procession results of depth images and foreground extraction.
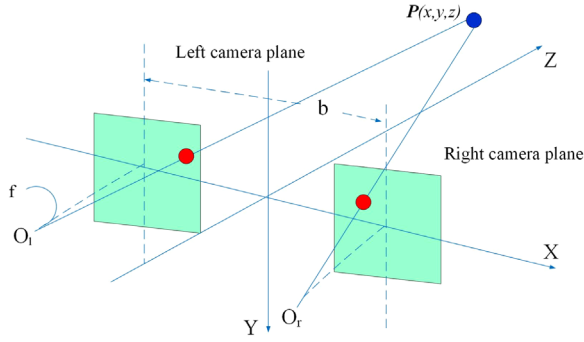
**Figure 3** Kinect imaging system model.

```
Input: disparity map    disp(height,width)
Output: V disparity map  vdisp(height,maxdisp)
for each iᵗʰ row on disp do
for each jᵗʰ line on disp do
    if disp(i,j)>0 then
        vdisp(i,disp(i,j))++
    end
 end
end
```

**Figure 4** The V disparity map algorithm.

frame subtraction method can easily be used to obtain the silhouette of the moving individual in depth images. Figure 2 shows the pre-processing results of the depth images and foreground extraction. Figure 2(a) is the colour image of the background frames; Figure 2(b) is the depth image of the background; Figure 2(c) is the result of the median filter of the background; Figure 2(d) is the colour image of the target; Figure 2(e) is the depth image containing the target; Figure 2(f) is the result using a median filter of the target frame; Figure 2(g) is the silhouette of the moving individual; Figure 2(h) is the extracted result of the moving individual. It can be seen that the moving individual has been extracted.

### 3.3. Solution of floor plane

The disparity map is used to estimate the floor plane. The depth data output from the Kinect is obtained by the disparity of the infrared projector and the infrared camera. Figure 3 shows the Kinect system model. The imaging process can be simulated by a perspective projection model [28].

In a Kinect imaging system, the focal length of the infrared projector and infrared camera are both $f = 580$ pixels. The physical distance named baseline between them is denoted by $b$, and $b$ is 7.5 cm. The distance between one point $P = (X, Y, Z)$ in the space to the centre of the Kinect is $d$, which is the depth data. The imaging of point $P$ in left and right camera are $P_1(x, y)$ and $P_1'(x', y')$. The disparity is $x+x'$ in Figure 3. According to the similar triangles principle, the relationship between the disparity map and the depth image is shown as

$$disparity = \frac{f * b}{d} \tag{3}$$

By using (3), the disparity map can be derived from the depth image. Then the V disparity map is further computed by the statistics histogram of the horizontal direction. Figure 4 shows the V disparity map algorithm.

The floor plane will be exacted from disparity map. There is a noticeable slant and thick straight line in the disparity map shown as in Figure 5(a). The starting point and the slope of the straight line are determined by the height and angle sensor of the Kinect. The length of the straight line corresponds to the position relation of the floor plane in depth image and disparity map. In the indoor environment for fall detection of an elderly adult, the floor plane is the main domain of the scene. So the Otsu algorithm is used to segment the floor plane. Because of the uncertainty of measurement, there is a certain disparity range between $\pm d_t$ of the floor plane. If the disparity of the $y$th line is $d_y$, $d \in (d_y - d_t, d_y + d_t)$ then it will belong to the floor plane.

Most of the plane information is extracted effectively by the V disparity map. Then the floor plane equation is estimated by the least squares method as follows. According to the perspective projection imaging model, there is the following relationship of point $(x, y)$ from pixel coordinates to location $(X, Y, Z)$ of the world coordinate system for the depth image:

$$\begin{cases} X = \frac{Z(x - c_x)}{f} \\ Y = \frac{Z(y - c_y)}{f} \end{cases} \tag{4}$$

where $(c_x, c_y)$ is the centre of imaging, and $f$ is the focus length.

The depth points of the floor plane can be defined as $(X_i, Y_i, Z_i)$, where $X_i, Y_i$ and $Z_i$ are depth point's coordinates in the real world. After the depth point's coordinates have been determined, the floor plane equation can be described as $AX + BY + CZ = 1$. The least squares method is used to obtain coefficients after choosing floor area. The flowing equation is satisfied for these depth points:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{5}$$

And the coefficients can be solved by the least squares method shown as

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \left( \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix}^T \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix}^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{6}$$

In order to show the floor plane, the depth point will be defined as the floor plane if the subtraction result is smaller than 0.05. The criterion is given below.

$$|AX + BY + CZ - 1| < 0.05 \tag{7}$$

Figure 5 shows the result of floor plane estimation from the V disparity map. Figure 5(a) is the V disparity map of the
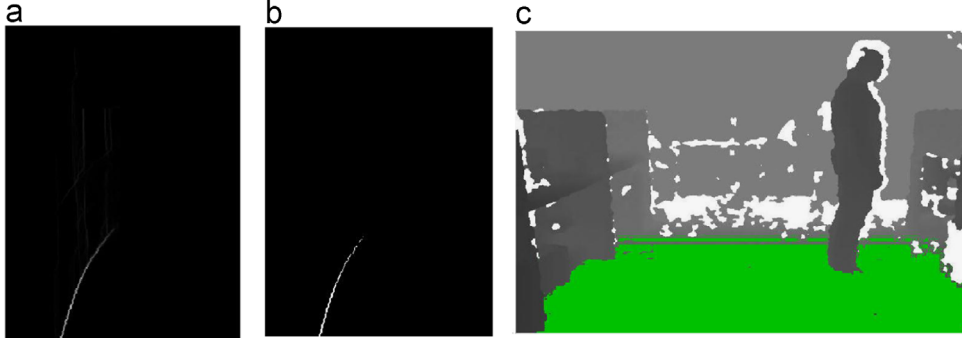
**Figure 5** Extracted floor planes from the depth images.

depth images in Figures 2 and 5(b) is the V disparity map after using the Otsu algorithm; Figure 5(c) is the floor plane domain shown in green. It is clear that the floor plane has been extracted correctly. Then the corresponding depth data is converted to the world coordinate system so that the equation of the floor plane can be determined.

## 3.4.  Orientation analysis of human body

When the depth values of human body region are extracted, an analysis on shape and orientation of the human body is performed to extract effective features for fall detection. In most shape analysis, a bounding box is used to approximate the domain of the body [2,7]. In contrast, an ellipse is used to approximate the segmented human shape in our work which is more accurate than a bounding box [24]. The Moment-based Method is applied to fit the ellipse.

As we know, an ellipsoid is determined by four parameters: centre location $(\overline{x}, \overline{y})$, main axis orientation $\theta$ and the values of major semi-axis $a$ and minor semi-axes $b$. For a segmented discrete depth image $Z(x, y)$, $f(x, y)$ is the corresponding binary image. The origin moments of $f(x, y)$ are given as

$$m_{pq} = \sum_{x,y} x^p y^q f(x, y), p, q = 0, 1. \tag{8}$$

By using the first- and zero-order origin moments, the centre location $(\overline{x}, \overline{y})$ of the ellipse can be computed as:

$$\begin{cases} \overline{x} = m_{10}/m_{00} \\ \overline{y} = m_{01}/m_{00} \end{cases} \tag{9}$$

The central moments for a segmented depth image $z(x, y)$ are given by

$$u_{pq} = \sum_{x,y} (x - x_0)^p (y - y_0)^q f(x, y), p, q = 0, 1, 2. \tag{10}$$

The angle between the major axis of the human body and the horizontal plane denotes the orientation of the ellipse, and it is computed as follows

$$\Theta = \arctan\left(\frac{2u_{11}}{u_{20} - u_{02}}\right) \tag{11}$$

The values of the major semi-axis $a$ and minor semi-axis $b$ can be obtained by calculating the max and min moments of inertia which are denoted as $I_{max}$ and $I_{min}$. Values of both are determined by evaluating the eigenvalues of the covariance matrix:

$$J = \begin{pmatrix} u_{20} & u_{11} \\ u_{11} & u_{02} \end{pmatrix} \tag{12}$$

The max and min moments of inertia are computed as

$$I_{max} = \frac{u_{20} + u_{02} + \sqrt{4u_{11}^2 + (u_{20} - u_{02})^2}}{2} \tag{13}$$

and

$$I_{min} = \frac{u_{20} + u_{02} - \sqrt{4u_{11}^2 + (u_{20} - u_{02})^2}}{2} \tag{14}$$

The major semi-axis $a$ and minor semi-axis $b$ are as follows:

$$a = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{I_{max}^3}{I_{min}}\right]^{1/8} \tag{15}$$

and

$$b = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{I_{min}^3}{I_{max}}\right]^{1/8} \tag{16}$$

The average depth of human domain is computed as follows:

$$\overline{Z} = \frac{\sum\limits_{x,y} Z(x, y)}{\sum\limits_{x,y} f(x, y)} \tag{17}$$

So from the coordinate system translation of (4) for the centroids of the ellipse $(\overline{x}, \overline{y})$, the centroids of human body in the world coordinate system $(\overline{X}, \overline{Y}, \overline{Z})$ can be computed.

Two examples of the approximated enclosing ellipse and the corresponding bounding box are illustrated in Figure 6. Figure 6(a) is the original grey image of a vertical person; Figure 6(b) is the rectangle fitting result; Figure 6(c) is the Ellipse fitting result; Figure 6(d) is the original grey image of an inclined person; Figure 6(e) is the rectangle fitting result; Figure 6(f) is the Ellipse fitting result. As shown in Figure 6, the ellipse performs more accurately when describing human body postures than a bounding box approximation. We can clearly see that the ellipse is more accurate than the bounding box when the human body is inclined. In addition, the orientation of the ellipse provides more accurate information of human body postures.

In Refs. [8,23,25], $\Theta$ in (11) is the orientation of the human body. However, the ellipse determined by $\{(\overline{x}, \overline{y}, \Theta,$
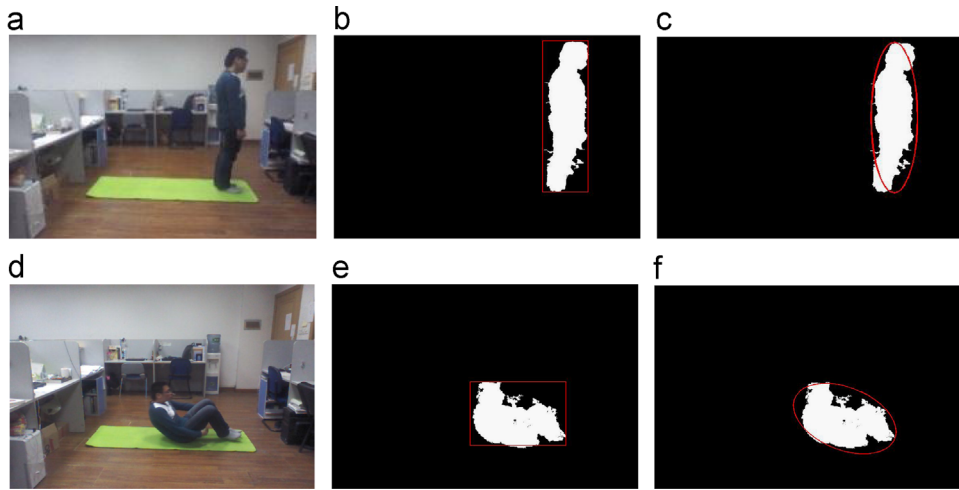
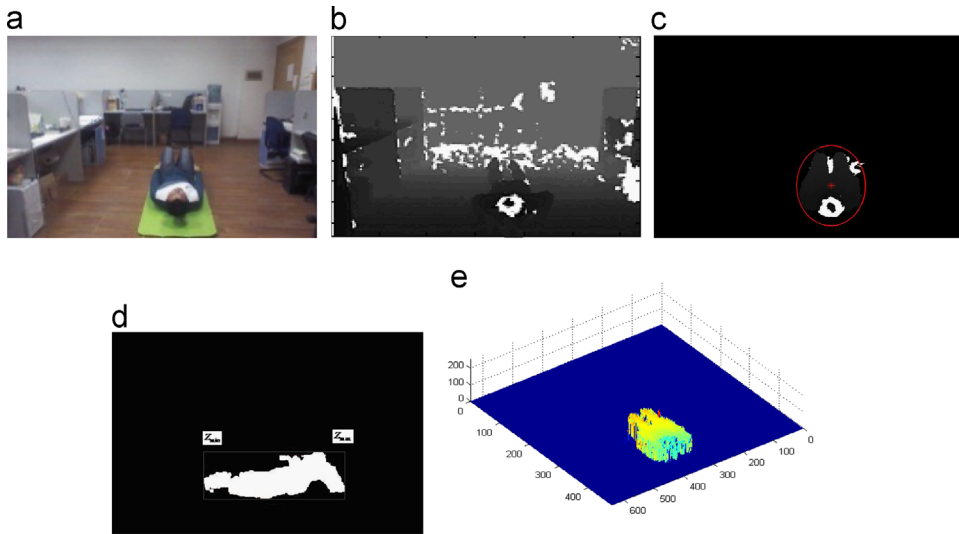**Figure 6**   Rectangle and ellipse estimation results.



**Figure 7**   Cuboid approximation result containing the human body.

*a, b}* is the approximation of the projection of human body on the imaging plane. So the depth values of the human body are used to determine its orientation in our work. $Z_{max}$ and $Z_{min}$ are assumed to be the maximum and minimum depth values of the human body shown in Figure 7(d) in the Z direction. $(x_{mix}, y_{min})$ and $(x_{max}, y_{man})$ are the near point and the far point of a rectangle which contains the ellipse shown in Figure 6. Then $(x_{mix}, y_{min})$ and $(x_{max}, y_{man})$ are translated to the world coordinate system $(X_{mix}, Y_{min})$ and $(X_{max}, Y_{man})$. The orientation of human body is approximated by the following line direction:

$$\frac{X - X_{min}}{X_{max} - X_{min}} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} = \frac{Y - Z_{min}}{Z_{max} - Z_{min}} \quad (18)$$

Figure 7(a) is a colour image of a person; Figure 7(b) is the depth image of the target; Figure 7(c) is the result of the centroids of the fitted ellipse in the imaging projection direction; Figure 7(d) is the projection of the person in the y-z plane. Figure 7(e) illustrates 3D of the human body. We see that the orientation is achieved even though the fall direction is approximately alone with the direction of the axis of imaging. Up to now, the orientation and position of

the human body and the floor plane have been obtained. The fall will be detected by using this information.

## 3.5.   Fall detection

Vision-based fall detection methods are usually based on the distances from the head centroid or human body to the floor plane [17,18]. In [17], the maximum height depth value and threshold are used to detect falls. Note that the Kinect sensor is specifically placed in a top view configuration; otherwise the floor plane cannot be accurately solved. In our proposed method, only the human body and floor plane are in the scene. The distance from the centroid of the human body to the floor plane can be calculated by following the formulation using $(\overline{X}, \overline{Y}, \overline{Z})$ and the plane solved in (6):

$$H = \frac{|\overline{X} * A + \overline{Y} * B + \overline{Z} * C - 1|}{\sqrt{A^2 + B^2 + C^2}} \quad (19)$$

In order to confirm fall detection, the angle between the orientation of the human body and the floor plane is further

computed as follows:

$$\theta = \frac{\pi}{2} - arc \cos$$

$$\frac{(X_{max} - X_{min}) * A + (Y_{max} - Y_{min}) * B + (Z_{max} - Z_{min}) * C}{\sqrt{(X_{max} - X_{min})^2 + (Y_{max} - Y_{min})^2 + (Z_{max} - Z_{min})^2} \sqrt{A^2 + B^2 + C^2}}$$

$$(20)$$

Then the distances from the centroids of the human body to the floor plane and the angles between the body and the floor plane are computed. When the distance from the centroids of the human body to the floor plane and the angle between the ellipses and the floor plane are lower than some thresholds, a fall incident will be detected.

## 4. Experimental results

Since no standard depth image database is available to evaluate the performance of different methods and mechanisms, the performance of the proposed method is evaluated by experiments on our captured dataset in this section. The experiments are performed in a simulated home environment, and a Kinect sensor was used for recording the depth video sequence. The recorded video sequence is processed using Matlab2013 on an Intel(R) Core (TM) i5 3.10 GHz CPU with 3 GB RAM.

Two typical examples of falls in different orientations are shown in Figure 8. Figure 8(a) and (c) are several colour image frames of falling down. Figure 8(b) and (d) are depth images frames of falling down at the same time. The centroids of the human body are marked, and the fitting ellipses are shown in red. In the following frames, centroids of the human body are computed and the distances from the centroids of the human body to the floor plane are computed and compared with the threshold. At the same time, the angles between the body and the floor plane are computed and compared with the threshold. When the double values are both lower than the threshold, the participant will be detected as having fallen. In the last row of each sequence, the fall has been detected. The trajectories of the centroids of the human body to the floor plane and the angles between the body and the floor plane are shown in Figure 9. Correspondingly, the trajectories are marked from green to red when the fall has been detected. The distance threshold is 0.5 m, and the angle threshold is 45°.

As shown in Figures 8 and 9, the fall incidents have been accurately detected. In general methods, falls are hard to detect when the fall orientation is aligned with the optical axis of the vision sensor [3,4]. By using our approach, we get the right detection results in these experiments, although the orientations are almost on the optical axis of the Kinect as shown in Figures 8(c) and (d) and 9(c) and (d).
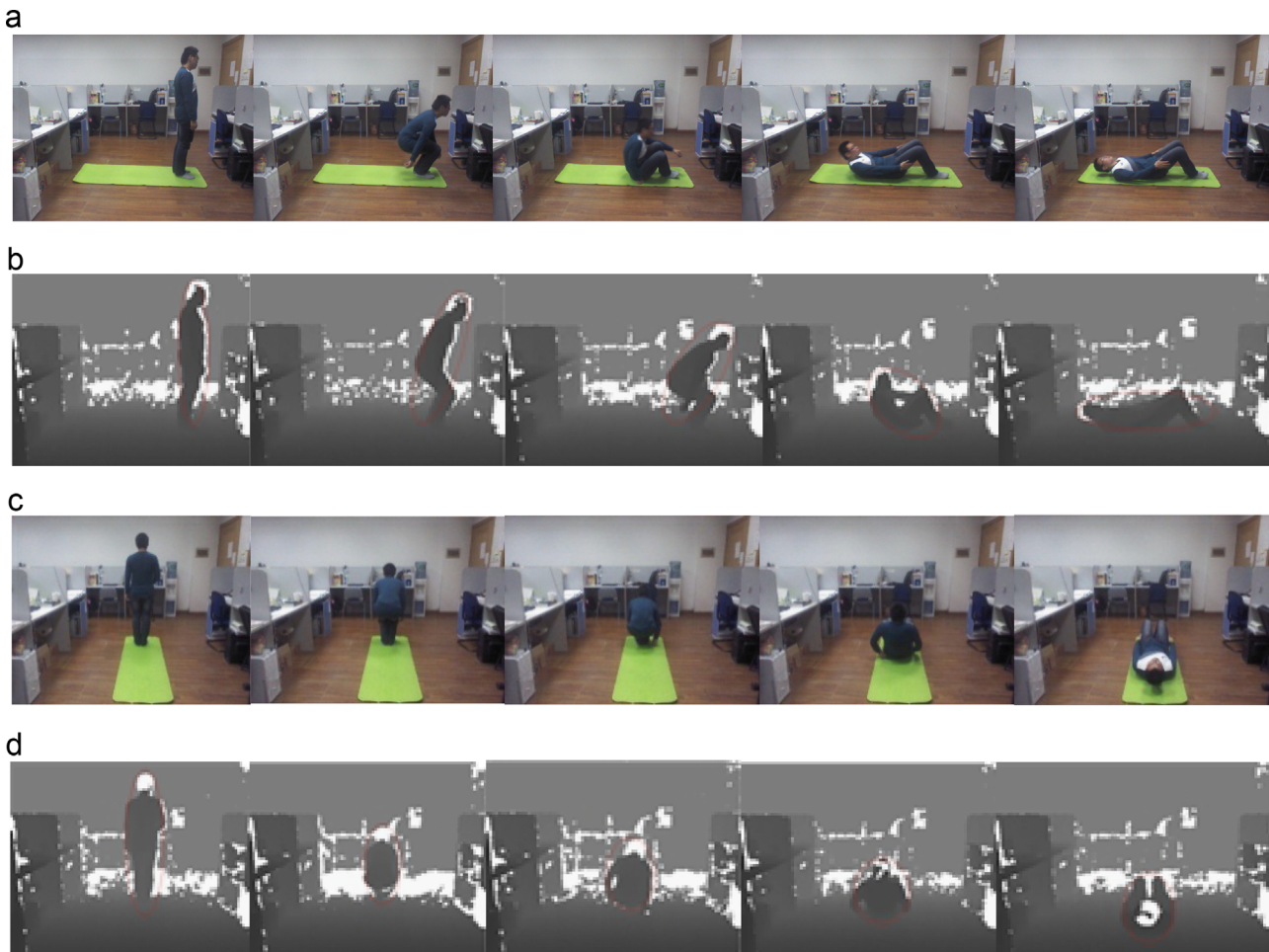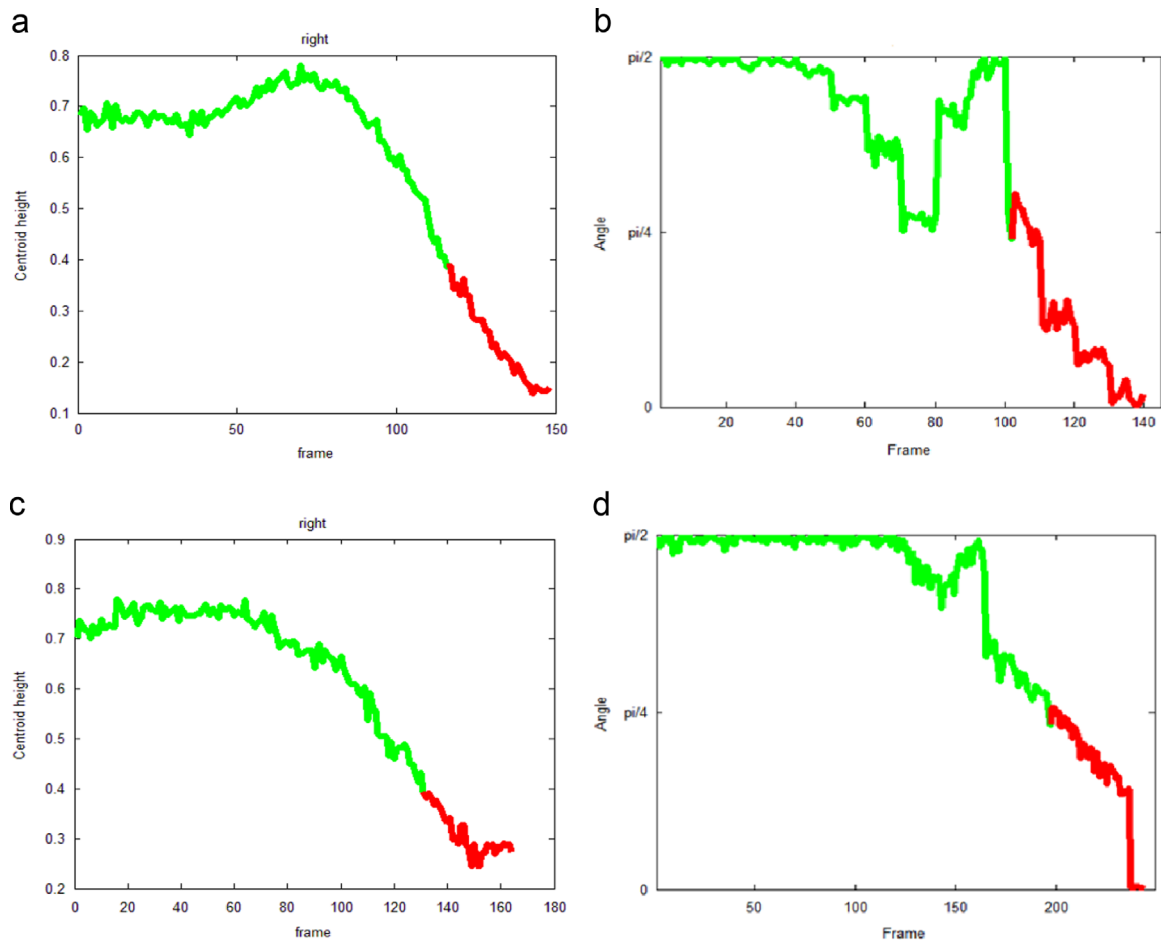


**Figure 8** Fall down in two different orientations.

a



b



c



d



**Figure 9**   Trajectories of distances and angles between the centroids of the human body to the floor plane.

**Table 1**   Time consumption of the proposed method.

| No. | Frame number | Processing time/s | Frame rate (F/s) |
|-----|--------------|-------------------|------------------|
| 1   | 163          | 4.794             | 34               |
| 2   | 246          | 6.844             | 36               |

Fall accidents usually happen in a very short period of time. The computation time of the proposed method is given to show its ability for fast response. The frames of the two videos shown in Figure 8 are 163 and 246. The processing time of each is 4.794 s and 6.844 s separately. Computation results of the frame rate of the proposed method on our PC are shown in Table 1. We see that the frame rate of the proposed method is faster than 30 frames per second. The frame rate of a common video is 30 frames per second. So the method satisfies the real-time requirements.

## 5.   Conclusions

In this paper, we have proposed a new fall detection method for elderly individuals living in a home environment based on shape analysis of depth images. Apart from the existing 2D vision-based fall detections, depth data is used in the proposed method. Depth images are pre-processed by a median filter for background images and target images. The silhouettes of moving individuals in depth images are obtained by a simple subtraction method from background frames. The floor plane equation is estimated by a disparity map with the least squares method. Then the shape information of the human subject in the depth images is analyzed by a set of moment functions. An ellipse model is used to simulate the shape of human subject. The centroids of the human body and the angle between the human body and the floor plane are further calculated. When both values exceed some thresholds, a fall incident can be detected.

The proposed method has the following merits. Firstly, the Kinect sensor is mature and cheap. So the proposed method can be easily applied in a smart home environment. Secondly, the method is based on threshold detection rather than classification techniques that most existing fall detection methods depend on. Only the centroids of the human body and the angle between the target and the floor plane are used as criteria, which avoids feature extraction and classification. At last, shape information of human subject is used by the estimated ellipses derived from the moment functions, which is vivid to describe the shape deformation of the human body. Experimental results show that the proposed method is effective. In this work, the Kinect v1 sensor is used. In our future work, the Kinect v2 sensor will

be deployed to improve the resolution of the depth image. How to improve the accuracy of shape information of the human target will be further investigated.

## Acknowledgements

## References

[1] Nihseniorheatlth: about falls. Available online: ⟨⟨http://nihseniorhealth.gov/falls/aboutfalls/⟩⟩ 01.html (accessed 31.01.15).

[2] Y.S. Delahoz, M.A. Labrador, Survey on fall detection and fall prevention using wearable and external sensors, Sensors 14 (2014) 19806-19842.

[3] R.L. Shen Victor, Lai Horng-Yih, Lai Ah-Fur, The implementation of a smart phone-based fall detection system using a high-level fuzzy Petri net, Appl. Soft Comput. J. 26 (1) (2015) 390-400.

[4] Y. Zigel, D. Litvak, I. Gannot, A method for automatic fall detection of elderly people using floor vibrations and sound proof of concept on human mimicking doll falls, IEEE Trans. Biomed. Eng. 56 (12) (2009) 2858-2867.

[6] Cheng Juan, Chen Xiang, Shen Minfen, A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals, IEEE J. Biomed. Health Inform. 17 (1) (2013) 38-45.

[7] E.B. Nashwa, Q. Tan, F.C. Pivot, L. Anthony, Fall detection and prevention for the elderly: A review of trends and challenges, Int. J. Smart Sens. Intell. Syst. 6 (2013) 1230-1266.

[8] W. Feng, R. Liu, M. Zhu, Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera, Signal Image Video Process. 8 (2014) 1129-1138.

[9] Y.T. Liao, C.-L. Huang, S.-C. Hsu, Slip and fall event detection using Bayesian Belief Network, Pattern Recognit. 45 (2012) 24-32.

[10] L. Yang, Y. Ren, H. Hu, B. Tian, New fast fall detection method based on spatio-temporal context tracking of head by using depth images, Sensors 15 (9) (2015) 23004-23019.

[11] Muhammad Mubashir, Ling Shaon, Luke Seed, A survey on fall detection: principles and approaches, Neurocomputing 100 (1) (2013) 144-152.

[12] A. Sorvala, E. Alasaarela, H. Sorvoja, R. Myllyla, A two-threshold fall detection algorithm for reducing false alarms, in: Proceedings of the 6th International Symposium on Medical Information and Communication Technology, ISMICT, La Jolla, CA, USA, 25-29 March 2012, pp. 1-4.

[13] G. Chen, C. Huang, C. Chiang, A reliable fall detection system based on wearable sensor and signal magnitude area for elderly residents, in: Y. Lee, Z.Z. Bien, M. Mokhtari, J. T. Kim, M. Park, J. Kim, H. Lee, I. Khalil (Eds.), Aging Friendly Technology for Health and Independence, Springer Berlin Heidelberg, Berlin Heidelberg, Germany, 2010, pp. 267-270.

[14] S. Luo, Q. Hu, A dynamic motion pattern analysis approach to fall detection, in: Proceedings IEEE International Workshop on Biomedical Circuits and Systems, 2004, pp. 1-5.

[15] A. Sixsmith, N. Johnson, A smart sensor to detect the falls of the elderly, IEEE Pervasive Comput. 3 (1) (2004) 42-47.

[16] B. Kaluza, M. Lustrek, Fall detection and activity recognition methods for the confidence project: a survey, in: Proceedings of the 12th International Multi-conference Information Society, vol. A, pp. 22-25, 2009.

[17] H. Rimminen, J. Lindström, M. Linnavuo, R. Sepponen, Detection of falls among the elderly by a floor sensor using the electric near field, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 1475-1476.

[18] Zhong Zhang, Christopher Conly, Vassilis Athitsos, A survey on vision-based fall detection, J. Pers. Ubiquitous Comput., 18, 1, pp. 19-26.

[19] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, M. Aud, Linguistic summarization of video for fall detection using voxel person and fuzzy logic, Comput. Vis. Image Underst. 113 (1) (2009) 80-89.

[20] G. Diraco, A. Leone, P. Siciliano. An active vision system for fall detection and posture recognition in elderly healthcare. In Design, Automation and Test in Europe Conference and Exhibition, 2010, pp. 1536-1541.

[21] Zhen-Peng Bian, Junhui Hou, Lap-Pui Chau, Nadia Magnenat-Thalmann, Fall detection based on body part tracking using a depth camera, IEEE J. Biomed. Health Inform. 19 (2) (2015) 430-439.

[22] Erik E. Stone, Marjorie Skubic, Fall detection in homes of older adults using the microsoft kinect, IEEE J. Biomed. Health Inform. 19 (1) (2015) 290-299.

[23] M. Yu, Y. Yu, A. Rhuma, S.M. Naqvi, L. Wang, J.A. Chambers, An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment, IEEE J. Biomed. Health Inf. 17 (2013) 1002-1014.

[24] M. Yu, A. Rhuma, S.M. Naqvi, L. Wang, J. Chambers, A posture recognition based fall detection system for monitoring an elderly person in a smart home environment, IEEE Trans. Inf. Technol. Biomed. 16 (2012) 1274-1286.

[25] Yi. Ting Liao, Chung-Lin Huang, Shih-Chung Hsu, Slip and fall event detection using Bayesian Belief Network, Pattern Recognit. 45 (1) (2012) 24-32.

[26] Yie Tarng Chen, Yu-Ching Lin, Wen-Hsien Fang, A video-based human fall detection system for smart homes, J. Chin. Inst. Eng. 33 (5) (2010) 681-690.

[27] S.-W. Yang, S.-K. Lin, Fall detection for multiple pedestrians using depth image processing technique, Comput. Methods Progr. Biomed. 114 (2014) 172-182.

[28] G. Mastorakis, D. Makris, Fall detection system using Kinect's infrared sensor, J. Real-Time Image Process. 9 (2012) 635-646.

[29] R. Planinc, M. Kampel, Introducing the use of depth data for fall detection, Pers. Ubiquitous Comput. 17 (2013) 1063-1072.

[30] M. Haselman, S. Hauck, The future of integrated circuits: a survey of nanoelectronics, Proc. IEEE 98 (2010) 11-38.

[31] Alhimale Laila, Zedan Hussein, Al-Bayatti Ali, The implementation of an intelligent and video-based fall detection system using a neural network, Appl. Soft Comput. J. 18 (1) (2014) 56-69.

[32] J.L. Chua, Y.C. Chang, W.K. Lim, A simple vision-based fall detection technique for indoor video surveillance, Signal Image Video Process. 9 (3) (2013) 623-633.

**Lei Yang** received his Ph.D degrees in Control Science and Engineering from the School of Electronics and Information Engineering, Xi'an Jiaotong University in 2008. He worked as a post-doctoral researcher in Tongji University from July 2008 to June 2010. Now he is an associate researcher and postgraduate advisor in School of Mechatronics Engineering and Automation, Shanghai University. His research interests are in artificial intelligence and pattern recognition, computer vision and image processing. E-mail: yangyoungya@sina.com.

**Wenqiang Zhang** received his B.S degree in Automation from Changzhou Institute of Technology in 2014. He is currently working towards his M.S degree in Pattern Recognition and Intelligent Systems in Shanghai University. His research interests are image processing and computer vision. Emai:478213479@qq.com

**Yanyun Ren** received his B.S. degree in Architecture Electric and Intelligent from Jilin Architecture University in 2013. He is currently working towards his M.S degree in Pattern Recognition and Intelligent Systems in Shanghai University. His research interests are obstacle and moving target detection. E-mail:973911878@qq.com