Fall Detection Based on Depth Images via Wavelet Moment

Yupeng Ding, Hongjun Li*, Chaobo Li, Ke Xu, Pengzhen Guo
School of Electronic and Information
Nantong University
Nantong, China

Abstract—**Fall detection plays an important role in the detection of human abnormal behaviors. In this paper, a fall detection algorithm based on depth images via wavelet moment is proposed. Firstly, we normalize the image according to each pixel in the image relative to the distance from the centroid, then polar coordinate the normalized image. Secondly, Fast Fourier Transform (FFT) of the picture is performed. Thirdly, the feature vectors of the image are extracted by using wavelet transform. Finally, using the minimum distance and Support Vector Machine (SVM) classification methods recognizes human behaviors. Numerous experiments are carried out on a large number of human behavior samples and the averaged success rate of this algorithm is more than 90%. The experimental results show that the proposed algorithm is robust and has good detecting ability and application prospect.**

Keywords-Fall detection; Depth image; Wavelet moment

I. INTRODUCTION

Fall detection is a hot issue in human abnormal behavior detection [1]. Especially, it is crucial for the elderly living alone. The recent reported methods are as follows: wearable devices use sensors such as accelerometers to detect motion parameters of the objects, but they have poor comfort and expansibility [2-4]. Audio information judges the fall of human by vibration frequency signal, which is produced when someone fell and hit the ground, but it is easy to be disturbed by noise [5-6]. Video capture obtains object information by using image processing technology to test video monitoring system. However, the delay of image transmission leads to the limitation of real-time performance and the insufficient light and shadow leads to the sharp decrease of the success rate [7-8]. In 2010, Kinect equipment provides a possibility for the optimization of the previous human fall detection methods. Based on Kinect detection, it has more advantages of comfort and privacy.

The methods applied to detect human fall on Kinect have mainly two categories. One is the use of Kinect skeleton information for fall detection, such as the algorithm proposed by Kawatsu et al [9] and so on. These methods obtain the spatial position of the joint points of human skeleton by Kinect, design fall detection feature and use the method of threshold detection or support vector machine to detect the fall. The algorithm using skeleton information for the fall detection is simple and efficient, but the joint points data obtained by Kinect is unstable. When the body is blocked by obstacles, some important joint points will be lost, resulting the errors and false reports. The other is based on the color or depth images acquired from Kinect equipment to detect the fall. Rougier [10] segmented the human body image from the foreground image and extracted centroid height. The distance between the centroid of the human body and the ground and the speed of the body centroid are used to identify the falling behavior. The algorithm can recognize slow sitting, squatting and other activities in the blocked scene, but it is easy to get false reports for fast action such as squatting rapidly to pick up things, sitting down quickly etc. Kepski and Kwolek [11-12] combined Kinect and acceleration sensor to detect human fall events. This method detects foreground objects (people) by different processing technology on the depth images, and then calculates the gravity center of human body and the distance between the gravity center of human body and the ground to assess whether the fall event occurred according to the state information and the predefined rules obtained from two kinds of sensors. Mastorakis and Makris [13] established the human 3-diminisional (high, wide, deep) bounding box and calculated the speed of height direction and the depth to width direction to detect fall, which has higher detecting accuracy in the process of walking. Ni et al [14] proposed a method to distinguish and detect human fall events by color depth images.

In this work, we propose a fall detection algorithm from depth images based on wavelet moment. We solve the privacy problem existed in traditional video image processing using depth images acquired by Kinect. And the wavelet moment has the characteristic of translation, scaling and rotation in-variance. In this case, the algorithm not only improves the ability of detecting, but also it is robust.

Our paper is organized as follows: The introduction of wavelet moment is in section 2. Section 3 discusses the depth images used to detect fall based on wavelet moment, while section 4 ends the paper with our conclusion.

II. INTRODUCTION OF WAVELET MOMENT

Wavelet transform inherits and develops the idea of localization in short-time Fourier transform. At the same time, it overcomes the shortcoming of window size

This work is supported by the National Natural Science Foundation of China (NO.61601248, 61401239, 61401241); the University Science Research Project of Jiangsu Province (NO.16KJB510036); the Science and Technology Program of Nantong (NO.MS12016025); Innovation Training program for College Student(NO.201710304021Z).

978-1-5386-1937-7/17/$31.00 ©2017 IEEE

unchanging with the frequency and proposes a "time-frequency" window. Wavelet transform has the characteristic of multiresolution analysis. It has the ability of representing the local features of the signal both in the time domain and frequency domain. Therefore, it is widely used in the field of image processing and pattern recognition, which is one of the powerful tools for signal processing.

If the function $\varphi(x) \in L^1 \cap L^2$ satisfies the condition $C_\varphi = \int_R \frac{|\hat{\varphi}(\omega)|^2}{|\omega|} d\omega < \infty$, and $\varphi_{a,b}(x) = \frac{1}{\sqrt{a}} \varphi\left(\frac{x-b}{a}\right)$, then the wavelet transform of the function is defined as:

$$W_f(a,b) = \langle f, \varphi_{a,b} \rangle = \frac{1}{\sqrt{a}} \int_R f(x) \varphi\left(\frac{x-b}{a}\right) dx \quad (1)$$

The corresponding inverse transformation formula is defined as:

$$f(x) = \frac{1}{C_\varphi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} W_f(a,b) \varphi_{a,b}(x) da db \quad (2)$$

where $\varphi(x)$ is called the mother wavelet function, a is the scale factor, b is the translation factor.

Two dimensional image function $f(x, y)$ using polar coordinates is expressed as $f(r,\theta)$, among $x = r\cos(\theta), y = r\sin(\theta)$. F_{pq} is defined as:

$$F_{pq} = \iint f(r,\theta) g_p(r) e^{jq\theta} r dr d\theta \quad p,q = 0,1,2,\ldots \quad (3)$$

where F_{pq} is p+q moment feature, $g_p(r)$ is the radial component of the transform, $e^{jq\theta}$ is angular component. Formula (3) can be rewritten as:

$$F_{pq} = \int S_q(r) g_p(r) r dr d\theta \quad (4)$$

$$S_q(r) = \int f(r,\theta) e^{jq\theta} d\theta \quad (5)$$

where $S_q(r)$ represents the characteristic distribution of $f(r,\theta)$. Thus, the feature extraction is transformed from 2D to 1D, Assuming the rotated function is $f(r, \theta + \triangle\theta)$, $\triangle\theta$ is rotation angle, rotated moment F'_{pq} is:

$$F'_{pq} = \int\int f(r,\theta_0) g(r) r e^{jq(\theta+\triangle\theta)} dr d\theta = e^{jq\triangle\theta} F_{pq} \quad (6)$$

$\|F_{pq}\| = \|F'_{pq}\|$, the rotation in-variance of the moment is proved.

If $g_p(r)$ has value in the entire definition domain r, $\{0 \le r \le 1\}$, $\{0 \le \theta \le 2\pi\}$, the extracted F_{pq} expresses the global features of an image. If $g_p(r)$ has partial value in the entire definition domain r, the corresponding F_{pq} is the partial feature of an image, so the possibility of overlapping regions of the sample characteristics can be reduced. This is the key idea of using wavelet transform [15] to extract invariant moments from images.

Use wavelet function as $g_p(r)$ to construct wavelet moment. The function of wavelets:

$$\psi_{ab}(r) = a^{-1/2} \psi\left(\frac{r-b}{a}\right) \quad (7)$$

Expansion parameter a and displacement parameter b in $\psi_{ab}(r)$ are taken as integer discrete form, $a = a_0^{\,n}$, $b = b_0 a_0^n$, n is integer. Because the processed image mapped to $0 \le r \le 1$, the wavelet function $\psi_{ab}(r)$ has different frequency components according to the different value of a.

Set $a = 0.5$ and $b_0 > 0$, then when $a = 0.5^n$, $b = m \cdot 0.5^n$, m is integer, discrete wavelet function along the axis is defined as $\psi_{n,m} = 2^{n/2} \psi(2^n r - m)$. Selecting different values of m and n can obtain global and local features of the image. Wavelet moment invariant is defined as:

$$\|F_{n,m,q}\| = \left\| \int S_q(r) \psi_{n,m}(r) r dr \right\| \quad q = 0,1,2,\ldots \quad (8)$$

For a fixed r, $S_q(r)$ represents the qth characteristic quantity of $f(r,\theta)$ in $[0, 2\pi]$. Taking different scale factor n, displacement factor m, $\|F_{n,m,q}\|$ can provide features of $f(r,\theta)$ at different scales.

III. Depth Images to Detect Fall based on Wavelet Moment

A. Depth Images Acquisition

Kinect is a peripheral that Microsoft generated for its Xbox game console, and it is the motion sensing device which takes natural body movements as input. Through projecting Infra-red on the object and calculating the time that every light beam needs to be received by receiver sensor, we can draw a depth map, which make it possible to the motion sensing technology in 3D environment. We solve the privacy problem existed in traditional video image processing by using depth image acquired by Kinect, and improve the robustness of the algorithm. Images acquired and preprocessed are shown in Fig. 1. In this paper, we use the method in binaryzation of depth images before fall detection recognition, which is helpful to reduce the noise and the complexity of the algorithm.

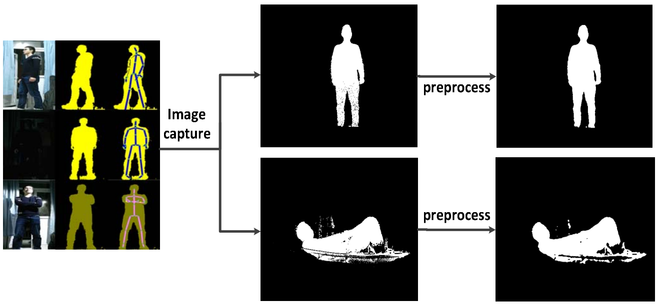


Figure 1. Images acquired and preprocessed

B. Feature Extraction By Wavelet Moments

1) Image Normalization

The first step in constructing wavelet moments is to scale and shift normalization for the image. Concrete methods as follows: First, determine gray scale centroid coordinates $(\bar{x}, \bar{y})$, where $\bar{x}=m_{10}/m_{11}$, $\bar{y}=m_{01}/m_{00}$. Scaling factor $\alpha=\frac{\max(\sqrt{(x-\bar{x})^2+(y-\bar{y})^2})}{M}$, M is normalized radius, and then we normalize according to the distance between the pixel and the centroid, $x=(x-\bar{x})/\alpha$, $y=(y-\bar{y})/\alpha$.

2) Image Polar Coordinates

In the process of image polar coordinates, the centroid of the gray image is used as the coordinate origin to transform. For continuous functions $f(x,y)$, the corresponding polar coordinates are expressed as $f(r,\theta)$, where $x=r\cos(\theta), y=r\sin(\theta)$. Because the sampled two-dimensional digital image is discrete, it is necessary to divide the angle equally when transform from the Cartesian coordinates to polar coordinates, $\theta=2\pi/N$. Normally $N=2^n$, we choose $N=64$.

3) FFT Transform

Theoretically, angle integration $S_q(r)=\int f(r,\theta)e^{jq\theta}d\theta$, it is a One-dimensional sequence about r, what is actually a discrete transform, so need to discrete the integral: select appropriate angle interval $\theta=2\pi/N$, and angle integration is $S_q(r)={}^1/_N\sum_{m=0}^{N-1}f(r,m)e^{-j2\pi mq/N}$. This is the form of Fourier series, which can use FFT to achieve the calculation.

4) Wavelet Moment Feature

Then extract feature from last result using wavelet functions in the radial area $\{0\le r\le 1\}$.Concrete methods as follows: $\left\|F_{j,k,q}\right\|=\left\|\sum_{r=0}^{1}S_q(r)\psi_{j,k}(r)r\right\|$, where $\psi_{j,k}(r)$ represents wavelet function, for a fixed r, $S_q(r)$ represents the characteristic distribution of $f(r,\theta)$ in $[0,2\pi]$. Using different scaling factors j, extract feature step by step. $D^j=\left\|\sum_{r=0}^{1}S_q(r)\psi_{j,k}(r)r\right\|$, $S^j=\left\|\sum_{r=0}^{1}S_q(r)\phi_{j,k}(r)r\right\|$. D^j, S^j and other different combinations record as characteristic quantity to distinguish separately.

We select Daubechies4 wavelet whose scale is 5 to construct wavelet moments. The algorithm flow chart is shown in Fig. 2. Fig. 3 shows the characteristics of different scale factors in fall and non-fall samples. We can know S^4, S^5, D^4 have little difference. In this paper, D^5 is chosen as the characteristic value of the fall detection, in which S represents the low frequency information, and D represents the high frequency information. TABLE Ⅰ lists 4 depth images wavelet moment feature in detail.

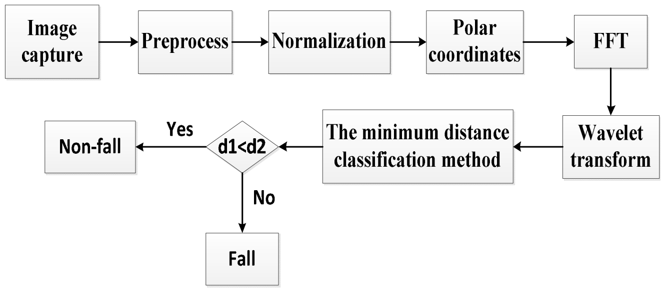
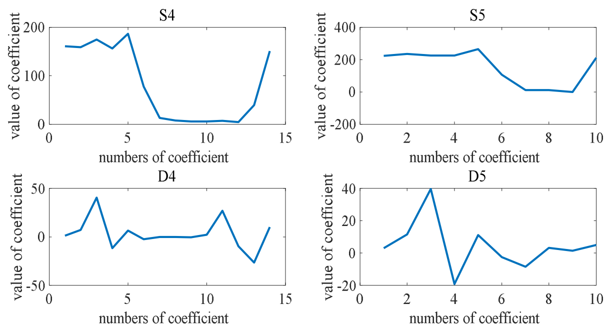
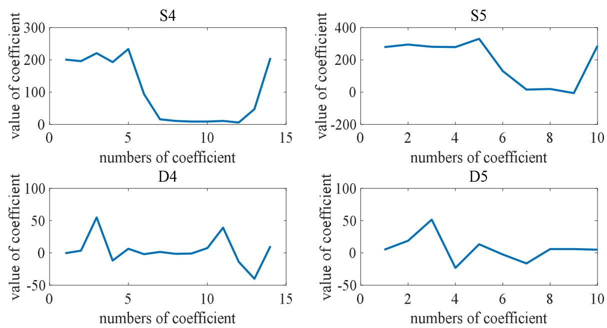


Figure 2. Flow chart of algorithm



(a)Wavelet coefficients at different scales under non-fall behavior



(b) Wavelet coefficients at different scales in fall behavior

Figure 3. Comparison of the characteristics under non-fall and fall behavior

TABLE I. CHARACTERISTIC VALUES IN NON-FALL AND FALL

Feature D^5	Non-fall 1	Non-fall 2	Fall 1	Fall 2
1	2.92	3.42	5.64	4.93
2	11.54	13.26	21.53	18.78
3	39.66	41.20	58.64	51.30
4	-19.10	-19.37	-23.24	-23.03
5	11.17	11.19	13.00	13.24
6	-2.55	-2.41	-2.14	-2.64
7	-8.42	-7.47	-12.48	-16.30
8	3.19	2.75	4.56	6.13
9	1.43	0.92	3.78	5.89
10	4.93	4.54	3.75	5.26
sum	44.77	48.05	73.05	63.57

TABLE II. COMPUTE THE MINIMUM DISTANCE

Feature	Unknown	Non-fall	Fall
1	3.21	2.92	5.64
2	12.54	11.54	21.53
3	40.81	39.66	58.64
4	-19.72	-19.10	-23.24
5	11.34	11.17	13.00
6	-2.25	-2.55	-2.14
7	-7.75	-8.42	-12.48
8	2.92	3.19	4.56
9	0.84	1.43	3.78
10	4.63	4.93	3.75
Distance *d*		1.97	21.31

C. *Fall Detection*

We use minimum distance and Support Vector Machine classification methods to classify the samples separately. Minimum distance classification method is an important method in image recognition. It regards the distance between feature points of an image of an unknown category and the point of the template in the feature space as criteria for classification. For K templates, if the distance between features points of an image of an unknown category and standard sample center is the minimum, it belong this template. The following is an example of fall detection, and the design idea and method of minimum distance classifier are explained. Human behavior can be divided into two categories: non-fall and fall, and use T_1 、 T_2 to represent. Each class has 10 standard sample feature vectors, $(x_1, x_2, \ldots, x_{10})$ and $(y_1, y_2, \ldots, y_{10})$. The Euclidean distance between feature vector $(z_1, z_2, \ldots, z_{10})$ of an image to be detected in an unknown class and T_1 、 T_2 are:

$$d_1 = \sqrt{(z_1 - x_1)^2 + (z_2 - x_2)^2 + \ldots + (z_{10} - x_{10})^2} \tag{9}$$

$$d_2 = \sqrt{(z_1 - y_1)^2 + (z_2 - y_2)^2 + \ldots + (z_{10} - y_{10})^2} \tag{10}$$

The corresponding criteria: calculating the distance of $(z_1, z_2, \ldots, z_{10})$ and feature vectors of two kinds of human behavior image templates can get a distance aggregate d_1 and d_2, and then the unknown human behavior is divided into the nearest category. d_1 and d_2 represent separately the distance between the image to be detected and the template of the non-fall and fall. If $d_1 < d_2$, unknown human behavior categories are not falling, in TABLE Ⅱ. $d_1 = 1.97$, $d_2 = 21.31$, because of $d_1 < d_2$, the man in the picture does not fall.

After testing 100 images of different types of human behavior (including non-fall 58 pieces, fall 42 pieces), the recognition result is: distinguish accurately non-fall images 52 and fall images 37, the correct rate is about 90% and 88%.

The Support Vector Machine is a supervised binary classification approach. SVM consists of cheating an optimal separating hyper plane that increases as much as possible the margin between two points belonging to two different classes. The optimal SVM classification function that describes the decision rule is given by:

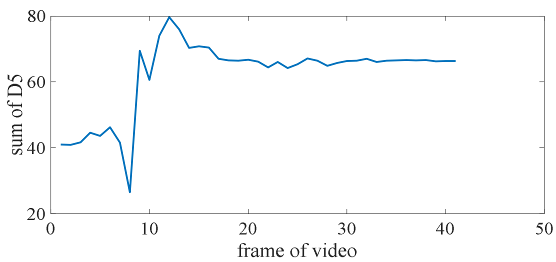
$$f(x) = \mathrm{sgn}[\sum_{i=1}^{N} y_i \alpha_i k(x, x_i) + b] \tag{11}$$

where x_i and $y_i = \{\pm 1\}$ are the feature vector and the class label of training set, respectively. $k(x, x_i)$ is the kernel function of SVM. α is the Lagrange multipliers vector. We tested several kernel functions such as the radial basis function and the linear function, in which the linear function was better. The correct rates of two methods are show in TABLE Ⅲ.

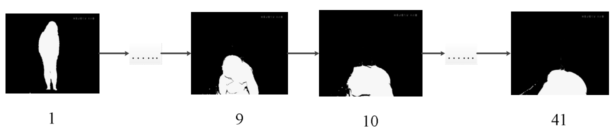
TABLE III. THE CORRECT RATES OF TWO METHODS

Method	Accuracy
Minimum distance	90
SVM	92

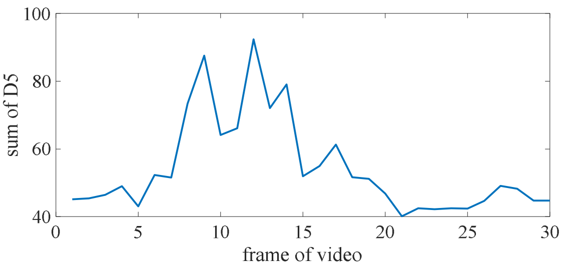
In order to further verify the recognition ability of the algorithm, we analyze 4 videos: (1) from stand to fall; (2) from stand to fall, and then stand up; (3) from stand to squat; (4) from stand to squat, and then stand up. According to Fig. 4, the recognition rate of the algorithm is similar to the recognition rate under static behaviors.



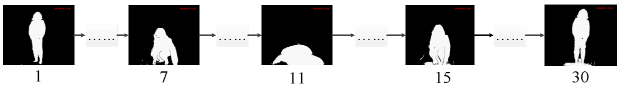
(a1) The sum values of D5 in video (1)



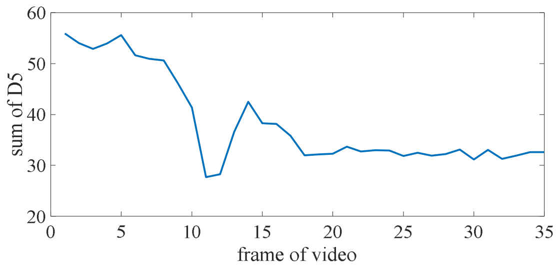
(a2) Images in each frame of the video (1)



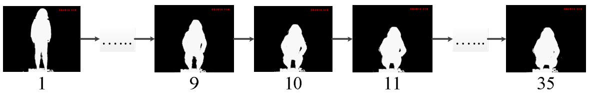
(b1) The sum values of D5 in video (2)



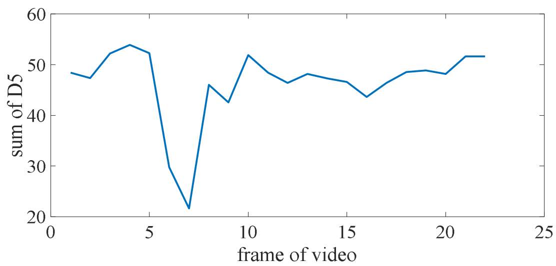
(b2) Images in each frame of the video (2)



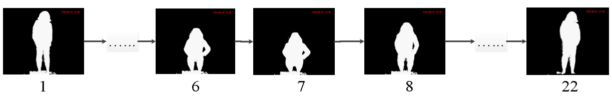
(c1) The sum values of D5 in video (3)



(c2) Images in each frame of the video (3)



(d1) The sum values of D5 in video (4)



(d2) Images in each frame of the video (4)

Figure 4. Comparison of the characteristics under different behaviors

IV. CONCLUSION

According to the human abnormal behavior detection problems, a fall detection algorithm based on depth images via wavelet moment is proposed. First, we solve the privacy problem existed in traditional video image processing using depth images acquired by Kinect. Secondly, the method via Kinect is low cost and robust. The wavelet moment has the characteristic of transformation, scale and rotation invariance. Therefore, the algorithm not only improves the ability of detecting, but also it is robust. We will focus on the study of fall detection in the case of multiple objectives in the future work.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (NO.61601248, 61401239, 61401241); the University Science Research Project of Jiangsu Province (NO.16KJB510036); the Science and Technology Program of Nantong (NO.MS12016025); Innovation Training program for College Student (NO.201710304021Z).

REFERENCES

[1] D. Webster, and O. Celik, "Systematic review of Kinect applications in elderly care and stroke rehabilitation," Journal of Neuroengineering & Rehabilitation, 2014, vol. 11, pp. 1-24.

[2] P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, "A high reliability wearable device for elderly fall detection," IEEE Sensor, 2015, vol. 15, pp. 4544-4553.

[3] F. J. Thilo, S. Bilger, R. J. Halfens, J. M. Schols, and S. Hahn, "Involvement of the end user: exploration of older people's needs and preferences for a wearable fall detection device-a qualitative descriptive study," Patient Preference & Adherence, 2017, vol. 11, pp. 11-22.

[4] E. Casilari, J. A. Santoyoramón, and J. M. Canogarcía, "Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection," 2016, vol. 11, p. e0168069.

[5] Y. Li, K. C. Ho, and M. Popescu, "A microphone array system for automatic fall detection," IEEE Transactions on Biomedical Engineering, 2012, vol. 59, pp. 1291-1301.

[6] D. Litvak, Y. Zigel, and I. Gannot, "Fall Detection of Elderly through Floor Vibrations and Sound," International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 4632-4635.

[7] B. Senouci, I. Charfi, B. Heyrman, J. Dubois, and J. Miteran, "Fast prototyping of a SoC-based smart-camera: a real-time fall detection case study," Journal of Real-Time Image Processing, 2014, pp. 1-14.

[8] C. Maldonado, H. Ríos, E. Mezura-Montes, and A. Marin, "Feature selection to detect fallen pose using depth images," IEEE International Conference on Electronics, Communications and Computers, 2016, pp. 94-100.

[9] C. Kawatsu, J. Li, and C. J. Chung, "Development of a Fall Detection System with Microsoft Kinect," Robot Intelligence Technology and Applications, 2012, vol. 24, pp. 601- 621.

[10] C. Rougier, "Monocular 3D head tracking to detect falls of elderly people," Annual International Conference of the IEEE Engineering in Medicine and Biology-Proceedings, 2006, pp. 6384-6387.

[11] M. Kepski, and B. Kwolek, "Fall detection using ceiling-mounted 3D depth camera," Proceedings of the 9th International Conference on Computer Vision Theory and Applications, 2015, vol. 37, pp. 640-647.

[12] B. Kwolek, and M. Kepski, "Fuzzy inference-based fall detection using kinect and body-worn accelerometer," Applied Soft Computing, 2016, vol. 40, pp. 305-318.

[13] G. Mastorakis, and D. Makris, "Fall detection system using Kinect's infrared sensor," Journal of Real-Time Image Processing, 2014, vol. 9, pp. 635-646.

[14] B. Ni, D. N. Chi, and P. Moulin, "RGBD-camera based get-up event detection for hospital fall prevention," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 1405-1408.

[15] H. Li, and C. Y. Suen, "A novel Non-local means image denoising method based on grey theory," Pattern Recognition, 2016, vol. 49, pp. 237-248.