International Journal of Internet, Broadcasting and Communication Vol.12 No.4 83-92 (2020)
http://dx.doi.org/10.7236/IJIBC.2020.12.4.83

IJIBC 20-4-11

Fall Detection Based on Human Skeleton Keypoints Using GRU

Yoon-Kyu Kang[1], Hee-Yong Kang[2], Dal-Soo Weon[3]

[1]Department of ITPM, Graduate School, Soongsil University, Korea
[2]Adjunct Professor,Information & Science Graduate Schhool, Soongsil University, Korea
[3](Corresponding Aurthor) Professor, Department of Smart IT, Baewha Womens University, Korea
ssme2@naver.com, hykang07@naver.com, dsweon@baewha.ac.kr

Abstract

A recent study to determine the fall is focused on analyzing fall motions using a recurrent neural network (RNN), and uses a deep learning approach to get good results for detecting human poses in 2D from a mono color image. In this paper, we investigated the improved detection method to estimate the position of the head and shoulder key points and the acceleration of position change using the skeletal key points information extracted using PoseNet from the image obtained from the 2D RGB low-cost camera, and to increase the accuracy of the fall judgment. In particular, we propose a fall detection method based on the characteristics of post-fall posture in the fall motion analysis method and on the velocity of human body skeleton key points change as well as the ratio change of body bounding box's width and height. The public data set was used to extract human skeletal features and to train deep learning, GRU, and as a result of an experiment to find a feature extraction method that can achieve high classification accuracy, the proposed method showed a 99.8% success rate in detecting falls more effectively than the conventional primitive skeletal data use method.

Keywords*: Skeleton Key points, GRU, Fall Detection, Deep Learning, PoseNet*

1. Introduction

Falls are a major cause of injuries or deaths in the elderly, incurring high social costs, and also frequently occurring in the manufacturing industry and in the field, requiring accurate detection and prompt action. Accordingly, various detection techniques were introduced, but the sensor-based drop detector devices [1-3] attached to the body were still ineffective due to user inconvenience, response time, and limited hardware resources. For medical and industrial use, there is a need for a vision-based fall detection system that uses an inexpensive general camera that does not require a sensor to be attached to the body.

2. Related Research

A vision-based fall detection system acquires images through video equipment, detects and uses images to

Manuscript Received: August. 15, 2020 / Revised: August. 20, 2020 / Accepted: August. 22, 2020
Corresponding Author: dsweon@baewha.ac.kr
Tel: +82-10-4372-8710, Fax: +82-2-737-6726
Professor, Department of Smart IT, Baewha Womens University, Korea

detect a fall. Deleting the background from the image [1,4] by extracting and characterizing the human skeleton data from the video, it detects a fall by characterizing the shape or silhouette of a person in the video [4,5]. Recently, the method of recognizing human activities based on skeletal data has focused on detecting falls [6]. Skeleton data is extracted using a 3D depth camera such as Kinect developed by Microsoft or 2D RGB video through CNN-based technology such as PoseNet [7]. In other words, PoseNet can conveniently and quickly extract skeleton data from images captured by a camera. Based on the skeleton information of the human body, it has not only high accuracy, but also simple and inexpensive features [1]. The extracted skeleton data is spatial-temporal data that changes with time. Recurrent Neural Network (RNN) is a powerful method of processing time-series data classification, but it takes a long time to learn and a part of the result is a problem in which the weight disappears (vanishing gradient problem) or an exploding gradient problem occurs. To solve this problem, the use of LSTM (Long-short term memory) and GRU (gated recurrent unit) was introduced [1,6]. LSTM does not have the same problem as RNN.

In general, 2D RGB cameras are installed and used by CCTV in hospital rooms and industrial sites of medical facilities. There may be concerns about the protection of some privacy, but it is believed that the camera will not be regulated by the security of privacy regulation because the purpose is to extract the skeleton of the body. GRU has the advantage of short learning time due to its simple structure as a modification of LSTM.

This propose a Fall Detection Method Based on Skeleton Key points Group(FDSG) that classifies and infers human poses from the extracted skeleton data from PoseNet to detects falls pose using artificial intelligence, GRU.

3. Fall Detection Method

A. Algorithm

The architecture of FDSG's fall detection method proposed in this study is as shown in Figure 1. The proposed fall detection method is a combination of HSSC(Head Shoulder Segment key point Coordinates), VHSSC(Velocity of HSSC) and RWHC(Ratio of Width and Height Coordinates) with skeleton key points (SD) extracted by PoseNet.

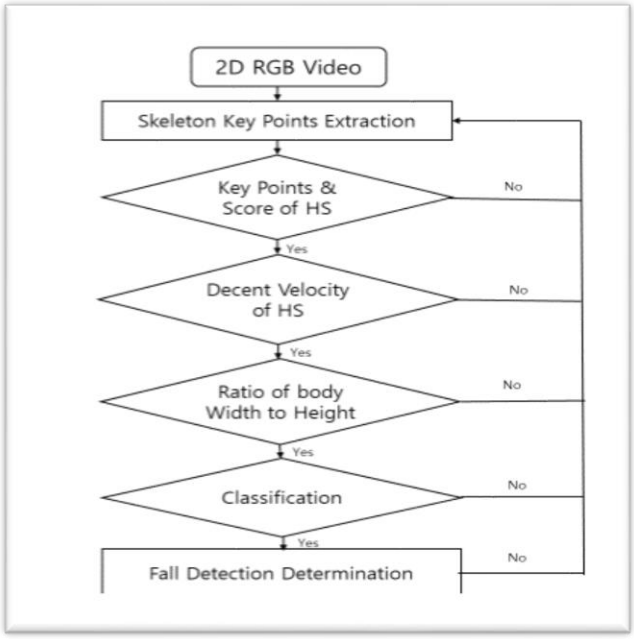

Figure 1. Architecture of FDHG

The proposed algorithm consists of 6 steps.

Step 1, it is a data collection process that collects 2D RGB video data from the camera.

Step 2, raw skeletal data for fall detection is extracted from video or image through PoseNet as shown in Figure 2. 17 key points from the human head to the feet are output of every image frame. The head including the nose, both eyes, both ears and both shoulders combined into one segment (HSS: Head and Shoulder Segment).

17 Key Points

Id	Part
0	nose
1	leftEye
2	rightEye
3	leftEar
4	rightEar
5	leftShoulder
6	rightShoulder
7	leftElbow
8	rightElbow
9	leftWrist
10	rightWrist
11	leftHip
12	rightHip
13	leftKnee
14	rightKnee
15	leftAnkle
16	rightAnkle

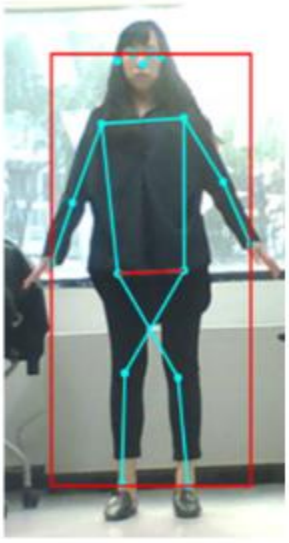

Figure 2. Key points information of human body

Step 3, the descending speed of the extracted HSS key point is calculated and applied.

Step 4, the width-to-height ratio of the body-shaped rectangle connected to the 17 skeleton coordinates which are provided by feature extraction method is applied.

Step 5, it is a posture classification process for determining a fall by enabling the processed input data into the learned GRU for inference.

The final process is the data extracted from the input image is determined whether to fall based on the GRU inference result.

The fall posture learning datasets are domestic AI HUB and URFD(UR Fall Detection) dataset[8]. Sources of GRU training are motion capture datasets and Annotated 2D images.

B. Feature Extraction for HSSC

The proposed feature extraction method used the following two methods. HSSC is as one segment. In the case of a bed fall in the wards, the change in the position of the head and upper body is the most important factor in detecting the fall. HSSC is the x- and y-coordinate corresponding to the body falling from the highest point when the body falls in the direction of gravity.

HSSC consists of 7 XY coordinates which consist of the key point coordinates of 7 parts, including the nose (0) of the face, 1 (1,2) of the eyes, the ears (3,4) and the left (5), right (6) shoulders. Figure 3 shows PoseNet programming for each skeleton key point's XY coordinate extraction.

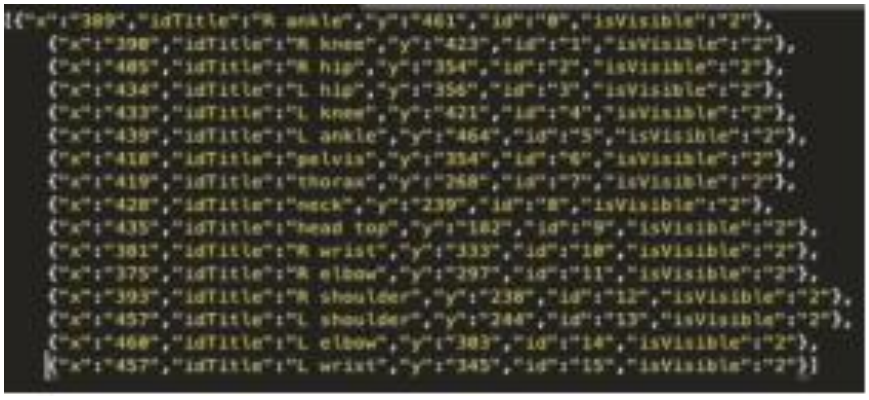

Figure 3. PosNet key point extraction example

C. Velocity of Descending HS

When a fall occurs, the center of gravity of the HS suddenly changes in a vertical direction. The human HS center point represents the center of gravity at the top of the human body and uses this feature. By processing the HS coordinate data obtained from PoseNet, the vertical coordinate of the HSS center point of each frame of the image is obtained. It is a very short process from the standing position to the falling position, and the time used is very short, so it is detected once every 5 adjacent frames at 0.25-second intervals [1].

Figure 4 shows skeleton key points and bounding which contains skeleton key points. This bounding box will be used for calculate the ratio of body width and height to improve the accuracy of fall detection.

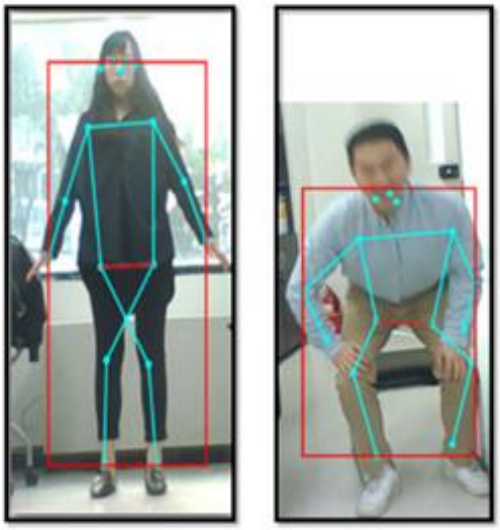

Figure 4. PoseNet's body bounding box

The extraction method is VHSSC. The speed is obtained by subtracting the first position and the end position of a window/frame of a video or image by time [9, 10] and the head position of each frame and the sequential change of HSSC are used as a threshold function.

A sliding window of length n frames is used as a training data set, and the mean values x and standard deviations (SD) of x in a defined frame are as shown in Equation (1).

$$S = \left|\sqrt{(x_{i-1} - x_i)^2} + \sqrt{(y_{i-1} - y_i)^2}\right| \quad (1)$$

Using the Euclidean distance method in Equation (1), Acc is the acceleration of the HSSC position of the X, Y and Z values provided by the Kinect sensor. S in Equation (1) is the acceleration of the head position by applying the Euclidean distance calculation method to the X and Y values provided by the kinetic sensor.

$$v = \frac{1}{n}\sum_{i=1}^{n} Dataset_i \quad (2)$$

In equation (2), v is the average value of the head position velocity in each sliding window, and n is the number of frames in the dataset.

$$Srd = \frac{1}{n}\sum_{i=1}^{n} RDS_i \quad (3)$$

S*rd* in equation (3) is the average value of the head position conversion speed, *n* is the number of frames per second of the sliding window, and real-time acceleration *RDSi* is the head position conversion speed of each sliding window in the series.

D. Ratio of Body Width to Height

This extraction method is RWHC. An eminent feature of the detected fall is the change in the body boundary

(body bounding box), so it can be judged by comparing the horizontal (body width) and the vertical (body height) length of the before and after posture. As for the ratio of the body's width and height, both the width and height of the posture change according to the change of the moving posture and the distance between the camera and the subject, but the ratio is unknown [1].

It is possible to distinguish the motion of falling by changing the width and height of a rectangle representing the outside of a person's body. Figure 5 shows 3 postures including working or stand, falling or setting on chair and lying or fallen. Each posture has its width to height, the ratio of width to height can be a parameter to detect the fall.

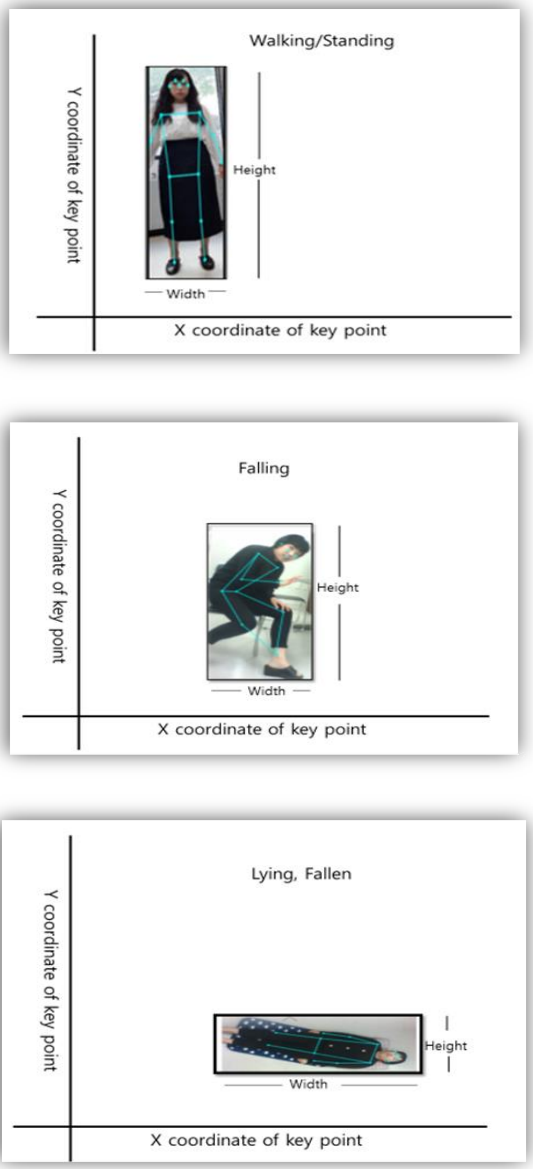

Figure 5. Bounding box width to height by PoseNet

"Walking/standing" in Figure 3 is a daily life (ADL: Activity of Daily Lives). It is possible to distinguish the width and height of the human body outer rectangle, and the ratio of the width to the height, here

R = Width/ Height, $R < 1$.

On the other hand, the ratio of width to height at the bottom figure (Lying/Fallen) is come to be bigger, here $R > 1$.

The proposed feature extraction method detects whether or not the posture is changed by using the variance of the human body proportions.

R(t) in Equation (4) represents the difference between the width and height of the current time t.

$$R(t) = \frac{w(t)}{h(t)} \tag{4}$$

In equation (5), uɤ(t) means the average value of the width and height at the current time t, and the value uɤ(t-1) is the average value of the width and height at the previous time point (t-1).

$$\mu_\gamma(\mathrm{t}) = (1\ \alpha)\mu_\gamma(t-1) + \alpha R(t) \tag{5}$$

$$\delta_\gamma(t) = \gamma(t) - \mu_\gamma(t-1) \tag{6}$$

The variance value of the width vs. height of the dwelling time increases when the posture is changed by Equation (6). If the variance exceeds the threshold, the changed posture is determined whether or not to fall in the learned GRU-based classification process.

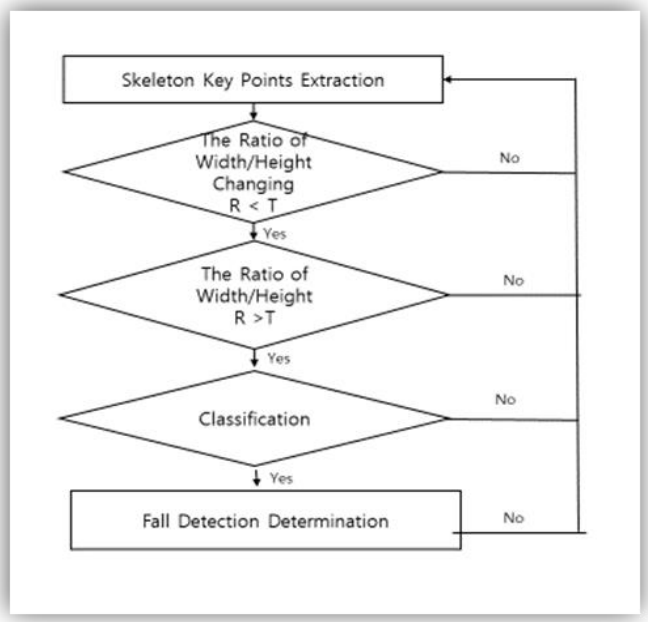

Figure 6. Classification algorithm with width to height ratio

4. Experiment

A. Configuration

The proposed method combines HSSC, VHSSC and RWHC with skeleton data key point (SD) extracted with PoseNet. The coordinates of the joints of each body were extracted using PoseNet as a key point of the

human skeleton, and the LSTM technique was applied among the circulatory neural network techniques to improve the accuracy of human posture classification.

It was set up with two stacks of GRU and 256 hidden layers. The above combination was constructed considering efficiency such as execution time and classification accuracy. The parameters of GRU are batch_size, epoch, and learning_rate, and batch_size is 128 as the size of the data to be input. The parameter Epoch representing the number of repeated learning of the training data was set to 5000. In addition, the learning Learing_rate is Adam[1,5] optimizer, and the initial learning rate is set to 0.0001.

For the dataset, the domestic AI HUB and the published URFD Dataset were used to distinguish falls.

Table 1. PoseNet configuration

algorithm	single-pose
architecture	ResNet50
inputResolution	250
outputStride	32
multplier	1
quantbytes	2

B. Results

The data set used in the experiment was the publicly available AI HUB and UR Fall Dataset (URFD), and walking, sitting, and falling were used for learning. The raw skeleton data (SD) of the input image, the HSSC method with the head and shoulder as one segment, the VHSSC method with the addition velocity of HSSC's descending and RWHC method that added the ratio of (width) to height (height) was experimented.

The coordinates of the key points of the skeleton created with PoseNet were used as the reference data set. A lot of coordinates corresponding to even small change in motion was generated and poses were refined based on four : 1. sitting 2. sitting on chair 3. Falling/lying 4. Fallen pose. Table 2 shows the SD and HSSC converged method extract the x, y coordinates and confidence scores of each key point at the time of fall.

Table 2. HSSC method coordinates and confidence scores

		x(width)	y(height)	score(신뢰점수)
1	head+shoulder	44.8378536542765	29.6835843543787	0.9498823059600
2	leftElbow	65.8753425222545	41.7345342125375	0.9460365844748
3	rightElbow	45.5378215783780	25.9365245745354	0.9245643580977
4	leftWrist	75.5783527858000	57.2156378637370	0.9815611980416
5	rightWrist	55.8353788112353	37.8727542452425	0.9859810826347
6	leftHip	65.4537834527527	44.7837354345378	0.9554020891864
7	rightHip	35.7873737837834	15.2378378227837	0.9418022064184
8	leftKnee	15.6546587373543	97.5828528280000	0.9716006545777
9	rightKnee	65.6546546542765	49.1586374837837	0.9337711781269
10	leftAnkle	45.6783753738765	27.2878245347583	0.9878408746095
11	rightAnkle	85.9375278542535	59.8235348378000	0.9881633553867

Table 3 shows the accuracy of classification and fall detection by applying SD and proposed HSSC, VHSSC and RWHSC methods. SD only feature classification accuracy is on average 97.52%, and the average posture feature classification accuracy of M2 of SD+HSSC, M3 of SD+VHSSC, and M4 with an additional ratio of the width and height of the bounding box is 98.36% and 99.06%, respectively. And with 99.46%, M4 had the highest fall feature classification. In addition, the accuracy of fall detection increased 1.1% in the fall detection accuracy of the proposed method M4 compared to SD without the proposed method.

Table 3. Experiment results

Method	Data Size	Dataset	Classification Accuracy (%)					Fall Detection Accuracy (%)
			Average	Sitting	Standing up	Falling	Fallen	
M1	34	AI HUB URFD	97.52	98.52	97.39	96.93	97.25	98.87
M2	38	AI HUB URFD	98.36	99.78	99.3	97.41	96.95	98.99
M3	40	AI HUB URFD	99.06	99.84	99.84	98.43	98.15	99.00
M4	40	AI HUB URFD	99.46	99.90	99.89	99.05	99.02	99.97

[Legend]

M1 : SD(Skeleton Key point Coordinate)

M2 : SD+HSSC(Head Shoulder Segment Key Point Coordinate)

M3 : SD+VHSSC(Velocity of Shoulder Segment Key Point Coordinate)

M4: HSSC+VHSSC+RWHC(Ratio of Width to Height Key point coordinate)

5. Conclusion

This paper proposes a new fall detection method using two public fall detection data sets, AI HUB and URFD, and a GRU neural network technique for fall classification and detection in video-based motion recognition using PoseNet.

As a result of experimenting the combination of the proposed feature classification method with the extracted human skeleton data, it was proved that the proposed FDSG method improved the accuracy of fall detection. However, it will be difficult to maintain the same high posture classification and fall detection accuracy as the test result even in the actual environment because the various fall data sets are absent and the experiment environment is restricted. In order to apply the vision-based fall detection function to actual distribution, there is room for improvement, such as sharing of technology and information corresponding to the intensity of light, complex environments, and securing various and sufficient data sets.

References

[1] Weiming Chen , Zijie Jiang , Hailin Guo and Xiaoyang, "Fall Detection Based on Key Points of Human-Skeleton Using OpenPose," Symmetry May 2020.
DOI:10.3390 /sym12050744

[2] AK Bourke, JV O'brien, and GM Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm", Gait & Posture, 2007.

[3] KWC Cheng, and DM Jhan, "Triaxial accelerometer-based fall detection method using a self-constructing cascade-Ada Boost-SVM classifier", IEEE Journal of Biomedical and Health Informatics, 2013.

[4] KA. Abobakr, M. Hossny, and S. Nahavandi, "A Skeleton-Free Fall Detection system From Depth Images Using Random Decision Forest", IEEE Systems Journal, vol. 12, Sept. 2018.

[5] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", ICLR, 2015.

[6] Wen-Nung Lie, Anh Tu Le, and Guan-Han Lin, "Human Fall-down Event Detection Based on 2D Skeletons and Deep Learning Approach", International Workshop on Advanced Image Technology, 2018.

[7] Kripesh Adhikari, Hamid Bouchachia, Hammadi Nait-Charif, "Deep Learning Based Fall Dection Using Silplified Human Posture", International Journal of Computer and Systems Engineering, Vol:13, No:5, 2019.

[8] M. D. Solbach and J. K. Tsotsos, "Vision-Based Fallen Person Detection for the Elderly," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, 2017.
DOI: 10.1109/ICCVW.2017.170.

[9] Wu, G, “Distinguishing fall activities from normal activities by velocity characteristics”, Journal of biomechanics, 2000.

[10] M. D. Solbach and J. K. Tsotsos, "Vision-Based Fallen Person Detection for the Elderly," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, 2017.
DOI: 10.1109/ICCVW.2017.170.

[11] Jangmook Kang, Sangwon Lee, “Strategy Design to Protect Personal Information on Fake News based on Bigdata and Artificial Intelligence,” International Journal of Internet, Broadcasting and Communication Vol.11 No.2 59-66 (2019).
DOI: http://dx.doi.org/10.7236/IJIBC.2019.11.2.59

[12] Jae-Jeong Hwang, Joon Moon, “Inductive Sensor and Target Board Design for Accurate Rotation Angle Detection,” International Journal of Internet, Broadcasting and Communication Vol.9 No.1 64-71 (2017).
DOI: https://doi.org/10.7236/IJIBC.2017.9.1