IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON RECENT ADVANTAGES OF COMPUTER VISION BASED ON CHINESE CONFERENCE ON COMPUTER VISION (CCCV) 2017

Received November 27, 2017, accepted January 1, 2018, date of publication January 18, 2018, date of current version March 13, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2795239

Detection of Human Falls on Furniture Using Scene Analysis Based on Deep Learning and Activity Characteristics

WEIDONG MIN, (Member, IEEE), HAO CUI, HONG RAO, ZHIXUN LI, AND LEIYUE YAO
School of Information Engineering, Nanchang University, Nanchang 330031 China

Corresponding author: Weidong Min (minweidong@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61762061 and in part by the Natural Science Foundation of Jiangxi Province, China under Grant 20161ACB20004.

ABSTRACT Automatic human fall detection is one important research topic in caring for vulnerable people, such as elders at home and patients in medical places. Over the past decade, numerous methods aiming at solving the problem were proposed. However, the existing methods only focus on detecting human themselves and cannot work effectively in complicated environments, especially for the falls on furniture. To alleviate this problem, a new method for human fall detection on furniture using scene analysis based on deep learning and activity characteristics is presented in this paper. The proposed method first performs scene analysis using a deep learning method faster R-CNN to detect human and furniture. Meanwhile, the space relation between human and furniture is detected. The activity characteristics of the detected people, such as human shape aspect ratio, centroid, motion speed are detected and tracked. Through measuring the changes of these characteristics and judging the relations between the people and furniture nearby, the falls on furniture can be effectively detected. Experiment results demonstrated that our approach not only accurately and effectively detected falls on furniture, such as sofa and chairs but also distinguished them from other fall-like activities, such as sitting or lying down, while the existing methods have difficulties to handle these. In our experiments, our algorithm achieved 94.44% precision, 94.95% recall, and 95.50% accuracy. The proposed method can be potentially used and integrated as a medical assistance in health care and medical places and appliances.

INDEX TERMS Activity characteristics, deep learning, faster R-CNN, human fall detection, medical assistance, scene analysis.

I. INTRODUCTION

Over the past decade, more and more elderly people had to live alone due to the development of seriously aging society. In the report of literature [1], the old-age dependency ratio will sharply increase from 22% in 2010 to 37% by 2050. The attendant problem is that falls have been one of the major health hazards among the population of age over 60 living alone, among whom accidental falls have become a widespread accident. Nowadays, falls are threatening the health and lifestyle of victims. Furthermore, falls are considered as the eighth leading cause of death in the U.S. [2]. A lot of falls occur in our daily activities. Besides falls during walking, many falls occur in the process of sitting or lying. Compared to healthy people, patients most likely have difficulties to control the balance of the body and hence fall, therefore human falls often happen in medical places. If a person falls unconsciously without getting emergency treatments, irreversible consequences such as fracture, stroke, disability and even death may occur. Unfortunately, the existing methods cannot effectively detect falls in complicated environments, especially for the falls on furniture which have different features from falls on floor because of involving furniture. In order to automatically detect the human fall in real time and provide medical rescue timely, it is extremely important to achieve highly accurate fall detection in complicated environments, especially for the falls on furniture which are big challenges for the existing methods.

For the reasons described above and other reasons, more and more researchers are keen on fall detection and activity classification, and have published many literatures on fall detection. The fall detection methods can be roughly classified into two types, i.e. the Auxiliary Equipment-based Method and the Vision-based Method.

A. THE AUXILIARY EQUIPMENT-BASED METHOD

Fall detection methods based on auxiliary equipments are mainly reflected in the wearable equipments and ambient sensors. As for wearable equipment, people mainly use them to monitor human and record some movement parameters. Pierleoni *et al.* [3] proposed a fall detection system consisting of an inertial unit that included tri-axial accelerometer, gyroscope and magnetometer to extract movement parameter. Their experiments obtained excellent accuracy, sensitivity and specificity, but they must place the sensors on the waist of a person. Ozcan *et al.* [4] placed a sensor camera in a person, and compared the changes in perspective to determine the fall behavior. Based on a wearable device, Ejupi *et al.* [5] developed a wavelet-based algorithm to detect and assess quality of sit-to-stand movements and assessment of fall risk in older people. Literatures [6]–[11] also described fall detection based on wearable devices such as smart phone, accelerometer or gyroscope. As for ambient sensors, people used them to collect environmental information to detect pedestrian falls. In the literatures [12], [13], researchers collected floor information including vibration and pressure to detect falls. Andò *et al.* [14] detected ADLs (Activities of Daily Living) by smart phone equipped with many embedded sensors to collect environment information. Stone *et al.* [15] used Kinect camera to extract personal depth images, and then calculated the vertical state to detect the fall. Zhuang *et al.* [16] proposed a method collecting audio signals from a microphone to detect fall behavior. Droghini *et al.* [17] used sound as input signal and proposed a semi-supervised framework that distinguished normal people from falling people based on a combination of OCSVM (One-Calss Support Vector Machine) and template matching classifier. There are other literature [18]–[21] to detect fall by collecting ambient information through smart camera and Kinect.

The Auxiliary Equipment-based Method as discussed above has some obvious shortcomings such as poor robustness and accuracy, and requirement of placing multiple sensors at selective and strategic positions. The above methods cannot effectively detect falls in complicated environments, especially for the falls on furniture. The methods using pressure, vibration and other ambient signals are very sensitive to external factors and hence have poor anti-noise capability. Another limitation is that these approaches are confined to where the sensors are installed. Besides, wearing these auxiliary equipment may be inconvenient and uncomfortable for people.

B. THE VISION-BASED METHOD

Due to the defects of the Auxiliary Equipment-based Method, more and more researchers have begun to study the Vision-based Method. Compared with the Auxiliary Equipment-based Method, this method needs no auxiliary equipment and achieves fall detection by intelligent algorithm analyzing video stream. Wang *et al.* [22] presented a novel model to segment foreground and detect pedestrian, then calculated the change of human contour to detect fall behavior. According to the changes of human body contours, Zerrouki *et al.* [23] identified human posture by SVM (Support Vector Machine) and classified fall behavior by HMM (Hidden Markov Model). Sun *et al.* [24] proposed a fall detection model based on thresholds according to the three stages of human fall: free fall, hitting the ground and stationary. This type of method mainly includes feature analysis approaches [25], [26], shape change analysis approaches [27], [28], posture analysis approaches [29], [30], and position analysis approaches [31], [32]. These methods perform well in normal fall behavior, but robustness and reliability were poor in complicated environment, especially for the falls on furniture. Since many researchers used various methods to detect fall, what is the difference of our work?

C. DIFFERENCE OF OUR WORK

Literature [33] pointed out that activities of elderly could be divided into four states, i.e. (1) the resting state such as standing, sitting, lying, or sleeping, (2) the movement state such as walking or running, (3) the emergency state such as falling, (4) the transition state such as standing to sitting, standing to lying, and so on. A lot of research work has been done to classify the basic activities of elderly mentioned above. Toreyin *et al.* [34] used HMM to track human, and used video sensor and shape parameters to detect fall behavior. This method could distinguish a person sitting on a floor from a person stumbling and falling. But this method used a sound sensor that was susceptible to external disturbances and could classify only limited activities. Rougier *et al.* [28] detected fall behavior based on a combination of motion history and human shape variation. Due to rely on history motion it couldn't classify special fall behaviors. Nait-Charif *et al.* [36] used a coarse ellipse model as a cue for fall detection, while Weidong *et al.* [37] used human shape aspect ratio to judge falls toward different directions. However, they did not discuss how to process falls on furniture. Ozcan *et al.* [38] used wearable camera to perform fall detection, but using an auxiliary equipment was inconvenient.

To alleviate the problems of the existing methods discussed above, a new video detection method for human fall detection on furniture using scene analysis based on deep learning and activity characteristics is presented in this paper. This method first performs scene analysis using a deep learning method, i.e. Faster R-CNN, to detect human and furniture such as sofa. The space relation between human and furniture are detected in scene analysis. The activity characteristics of the detected people such as human shape aspect ratio, centroid, motion speed are detected and tracked. By means of measuring the changes of these characteristics and judging the relations between people and furniture nearby, the falls on furniture can be effectively detected.

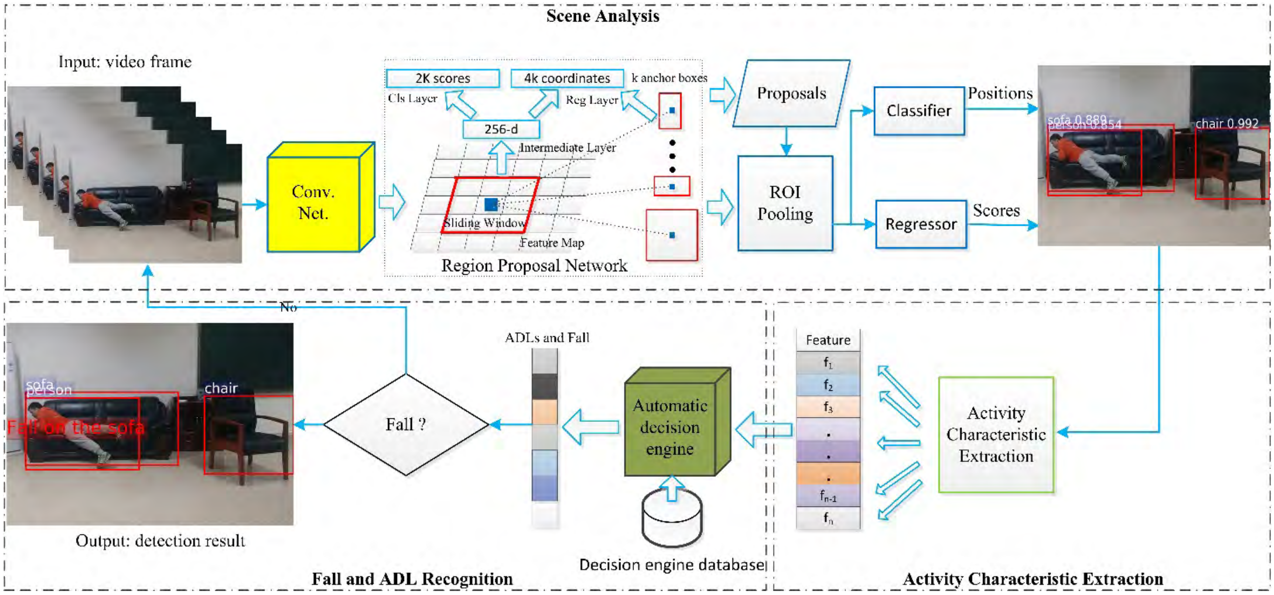


FIGURE 1. Overview of our proposed method for human fall detection.

The rest of this paper is organized as follows. Section II is an overview of our proposed method in the whole detection framework. In section III, we describe object detection and scene analysis in details. Then fall detection experiments are discussed in section IV. The paper is concluded in section V.

II. OVERVIEW OF OUR PROPOSED METHOD FOR HUMAN FALL DETECTION

As shown in Fig. 1, our proposed method for human fall detection on furniture consists of three parts. The first part is the scene analysis module based on Faster R-CNN to obtain the information of locations and objects in the scene. For detecting falls in complicated environments, we first put forward scene analysis using a deep learning method Faster R-CNN to measure the space relation between human and furniture. The second part is the Activity Characteristics Extraction module which calculates activity features of the detected people such as human shape aspect ratio, centroid, motion speed during detecting and tracking people. The third part is the fall and ADL recognition module. According to the features extracted from the second part, this module uses an automatic decision engine and a series of criteria to distinguish falls from ADLs by measuring the changes of these characteristics and judging the relation between people and furniture.

III. FALL DETECTION USING SCENE ANALYSIS BASED ON DEEP LEARNING AND ACTIVITY CHARACTERISTICS

A. SCENE ANALYSIS USING FASTER R-CNN

Scene analysis is an important procedure in our detection framework. In order to extract the accurate information of locations and objects in the scene, it's essential to choose an excellent object detection algorithm. Compared with R-CNN [39], Fast R-CNN [40] and other methods [41], [42], Faster R-CNN [43] has higher accuracy and faster speed of object detection. The object detection method Faster R-CNN mainly consists of two modules. The first module is the Fast R-CNN detector [40] that uses the proposed regions. The second module is a deep fully convolutional network that proposes regions. So, Faster R-CNN can be simply seen as a method combining region proposal network with Fast R-CNN. In other words, it uses region proposal network to replace the selective search method in Fast R-CNN. As shown in Fig. 2, from R-CNN to Fast R-CNN, and to Faster R-CNN, four steps (candidate region generation, feature extraction, classification, and location refinement) are finally unified into a depth network framework. All the calculations are not repeated and completely run in the GPU, greatly improving the speed.

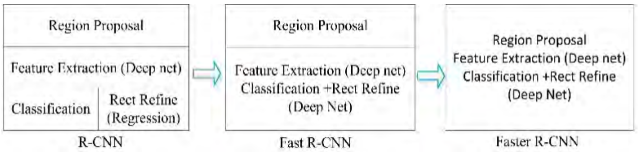

Region Proposal
Feature Extraction (Deep net)
Classification
Rect Refine (Regression)
R-CNN
Region Proposal
Feature Extraction (Deep net)
Classification +Rect Refine (Deep Net)
Fast R-CNN
Region Proposal
Feature Extraction (Deep net)
Classification +Rect Refine (Deep Net)
Faster R-CNN

FIGURE 2. The difference of R-CNN, Fast R-CNN and Faster R-CNN.

As shown in Fig. 1, the region proposal network slides a small network over the *n-by-n* convlutional feature map, and chooses a small sliding window as the input of the samll network. Each sliding window is decreased to 256-dimension features contained in two sliding full-connected layers, i.e. a regression layer (*reg layer*) and a classification layer (*cls layer*). We locate each sliding window and predict multiple region proposals to find the number of maximum possible proposals for each location (denoted as k). Therefore, $4k$ outputs encode the coordinates of k boxes in *reg layer*, and the

$2k$ outputs estimate probability of object for each proposal in *cls layer* [43]. The k proposals are relative to k reference boxes, which is defined as anchors. An anchor is centered at the sliding window in question, and is connected with a scale and aspect ratio as shown in the Region Proposal Network in Fig. 1. The RPN is trained end-to-end to generate accurate region proposal. Following literature [43], the loss function is defined as

$$L\left(\{p_i\},\{t_i\}\right) = \frac{1}{N_{cls}}\sum\nolimits_i L_{cls}\left(p_i, p_i^*\right) + \frac{1}{N_{reg}}\sum\nolimits_i p_i^* L_{reg}\left(t_i, t_i^*\right) \quad (1)$$

Here, p_i is the predicted probability of judging anchor i as an object. The parameter p_i^* is euqal to 1 or 0 which means the anchor is positive or negative. The 4 parameterized coordinates of the perdictrd bounding box is denoted as t_i, and t_i^* is the bounding box associated with a positive anchor. The classification loss L_{cls} is log loss between being this object and not this object. We use $L_{reg} = \left(t_i, t_i^*\right) = R\left(t_i - t_i^*\right)$ where R is the robust loss function (smooth L_1) to represent the regression loss. The two terms are normalized by N_{cls} and N_{reg} and weighted by a balancing parameter λ.

Although RPN and Fast R-CNN have different achitecture including different convolutional layers and training ways, both them have a raw feature extraction net to produce feature. Faster R-CNN develops an technique which shares convolutional layers between the two networks. RPN and Fast R-CNN are merged into a single network to share their convolutional features. As for classification and location refinement, the feature can be viewed as a 256-channel image with a scale of $51 * 39$. For each position of the image, the feature can be considered as nine possible candidate windows named anchors. The classification layer outputs the probability that each of the nine locations belongs to the foreground and background. The window regression layer outputs the position of each anchor, and the 9 anchors' corresponding windows should shift the scaled parameters. For each location, the classification layer outputs the probability of the foreground and background from the 256-dimensional feature. The window regression layer outputs four translation scaling parameters from the 256-dimensional feature. In partial cases, the two layers are fully-connected networks. In global cases, because the network in all locations ($51 * 39$) has the same parameters, using the actual size of 1×1 net achieves the convolution network.

In order to detect human falls on furniture, we must analyze the objects in scene. We can extract the accurate information of locations and objects by Faster R-CNN. Literature [33] addresses mainly activities of daily living. Therefore, it is necessary for us to detect some objects associated with ADLs such as chair and sofa. We carry out object detection with five videos, each including 600 frames. As shown in the Table 1, we analyze *recall, precision* and *accuracy* of the algorithm for persons, chairs and sofas. We have rec*all* $= \frac{TP}{TP+FN}$, *precision* $= \frac{TP}{TP+FP}$ and *accuracy* $= \frac{TP+TN}{TP+TN+FP+FN}$. Table 1 demonstrates that Faster-RCNN can detect the objects effectively. Based on Faster R-CNN to analyze scene, we can obtain the accurate information of locations and objects.

TABLE 1. The result of scene analysis.

Object	Real result	Prediction result		Recall	Precision	Accuracy
		Positive	Negative			
Person	TRUE	2807	81	99.80%	96.32%	96.27%
	FALSE	107	5			
Chair	TRUE	2372	369	94.88%	94.76%	91.36%
	FALSE	128	131			
Sofa	TRUE	2411	295	92.73%	95.83%	90.20%
	FALSE	189	105			

B. ACTIVITY CHARACTERISTICS OF HUMAN FALLS

There are various causes for fall behavior. Besides many falls occurred during walking, falls often happen when people sit or lie down on the furniture such as chair and sofa. Because characteristics of special falls such as falling on the sofa are similar to characteristics of some human ADLs, the traditional method has difficulty to distinguish special falls and fall-like activities from ADLs. Even if some methods can detect falls on the furniture, these methods cannot classify and identify fall behaviors. Therefore, detection of fall on furniture is a challenging problem. To solve the above problem, we take the space relation between human and objects (as shown in Fig. 3, left) into consideration. Then, we first propose a new feature (denoted as Dn) to measure the spatial relation between human and furniture.

We extract the accurate information of location and object in scene analysis. We assume that the given video is represented as V, and N $(n|n \in Z)$ denotes the total frames in the video V. The human center ($\vec{Hc}$), human width (Hw), human height (Hh), the location information of each person can be represented as $Person_V = \{(\overrightarrow{Hc}_i, Hw_i, Hh_i)|i \in\}$. The object center ($\vec{Oc}$), object width ($Ow$), object height ($Oh$), location information of each object in the video V can be represented as $Object_V = \{\left(\overrightarrow{Oc}_i, Ow_i, Oh_i\right)|i \in N\}$. If having no object, the value will be given null. We define the spatial distance (D) between human and furniture as Formula (2). Due to the different size of each furniture, using distance directly cannot be a unified measure. Therefore we need to normalize the distance between human and furniture. The space relation between human and furniture (Dn) is defined as Formula (3):

$$D^V = \left\{D_i^V \mid D_i^V = \left|\overrightarrow{Hc}_i - \overrightarrow{Oc}_i\right|, i \in N\right\} \quad (2)$$

$$Dn^V = \left\{Dn_i^V \mid Dn_i^V = \frac{\left|\overrightarrow{Hc}_i - \overrightarrow{Oc}_i\right|}{\sqrt{Ow_i^2 + Oh_i^2}}, i \in N\right\} \quad (3)$$

First of all, the feature Dn is tested using a video with two chairs and a sofa, in which, we firstly walk through the chair2 and sofa, then sit in the chair1. By making the experiment, we get the changes of Dn shown in Fig. 4.

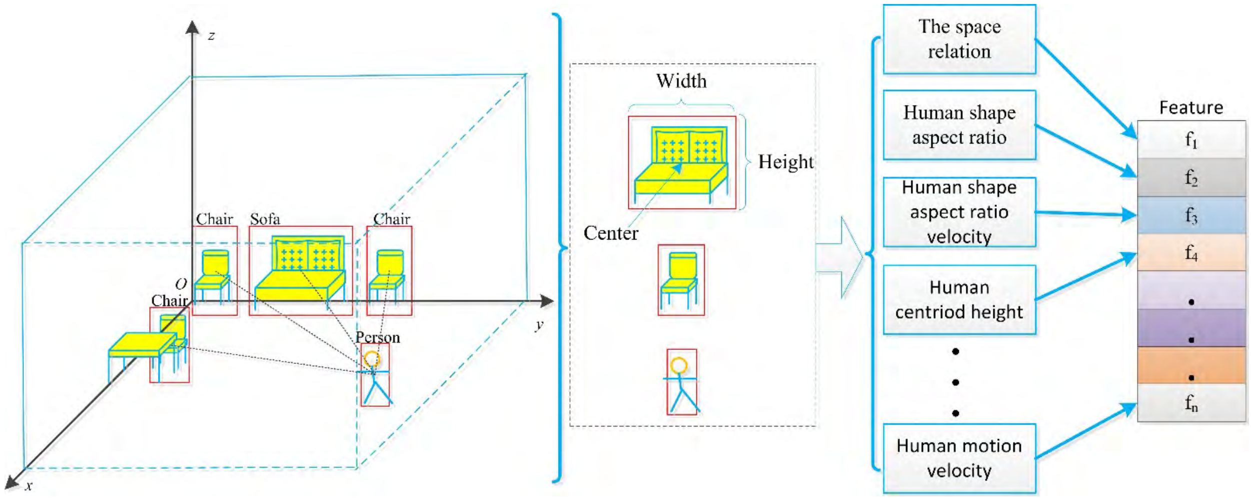


FIGURE 3. The framework of activity characteristic extraction.

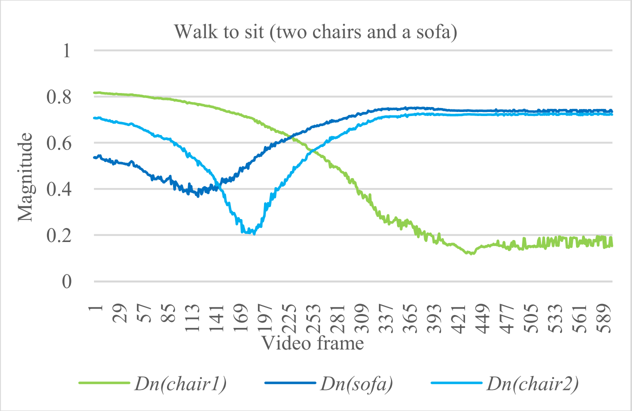


FIGURE 4. The change of *Dn* during walking for different objects.

When people pass the chair2 and the sofa, the magnitude of *Dn*(chair2) and *Dn*(sofa) is decreasing. From the 350th frame to the 440th frame, the person *Dn*(chair2) and *Dn*(sofa) become larger, but *Dn*(chair1) is decreasing. Therefore, it is determined that the behavior is relevant to chair1. A large number of methods were investigated in literatures [25]–[32] such as using human shape aspect ratio (R), the standard deviation of each ten adjacent frames, shape aspect ratio (σ), human centroid (Hc) height (Hh) and human motion velocity (Hv). We extract some activity features following the Formula (4)-(7). The direction vector in the vertical direction is denoted as $\vec{e_z}$:

$$R^V = \left\{ R_i^V \mid R_i^V = \frac{Hw_i}{Hh_i}, i \in N \right\} \tag{4}$$

$$\sigma^V = \left\{ \sigma_i^V \mid \sigma_i^V = \sqrt{\frac{1}{10}\sum_{i}^{i-10}\left(R_i^V - \overline{R_i^V}\right)^2}, i \in N \right\} \tag{5}$$

$$Hh^V = \left\{ Hh_i^V \mid Hh_i^V = \overrightarrow{Hc_i} \cdot \vec{e_z}, i \in N \right\} \tag{6}$$

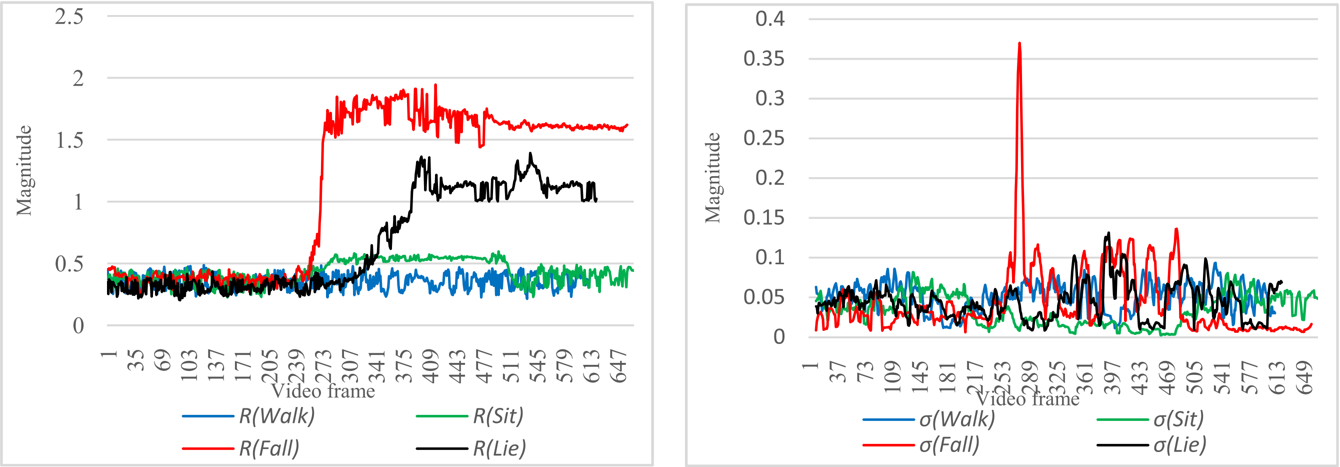
$$Hv^V = \left\{ Hv_i^V \mid Hv_i^V = \left|\overrightarrow{Hc_i} - \overrightarrow{Hc_{i-1}}\right|, i \in N \right\} \tag{7}$$

Secondly, we extract activity characteristics of human fall following procedures in Fig. 3. Four activities such as walking, walking to sit, walking to lie down and walking to fall are contained in our test videos. According to Formula (4)-(6), we extract the three features, hence the magnitude trends of Dn, R and σ with each frame of the video are plotted in Fig. 5.

In Fig. 5(a), people walk from begin to end (Fig. 5(a), blue line), sit on the chair from 253th frame to 503th frame (Fig. 5(a), green line), fall on the ground from 255th frame to end (Fig. 5(a), red line) and lie on the sofa from 315th frame to end (Fig. 5(a), black line). In Fig. 5(b), the magnitude of σ also changes as the magnitude of R changes instantaneously. We can clearly see that the magnitude of R is very small and smooth when people walk normally and the magnitude of R has undergone great changes when people's activities change. Obviously, the feature show different values in different behaviors and have a good distinction.

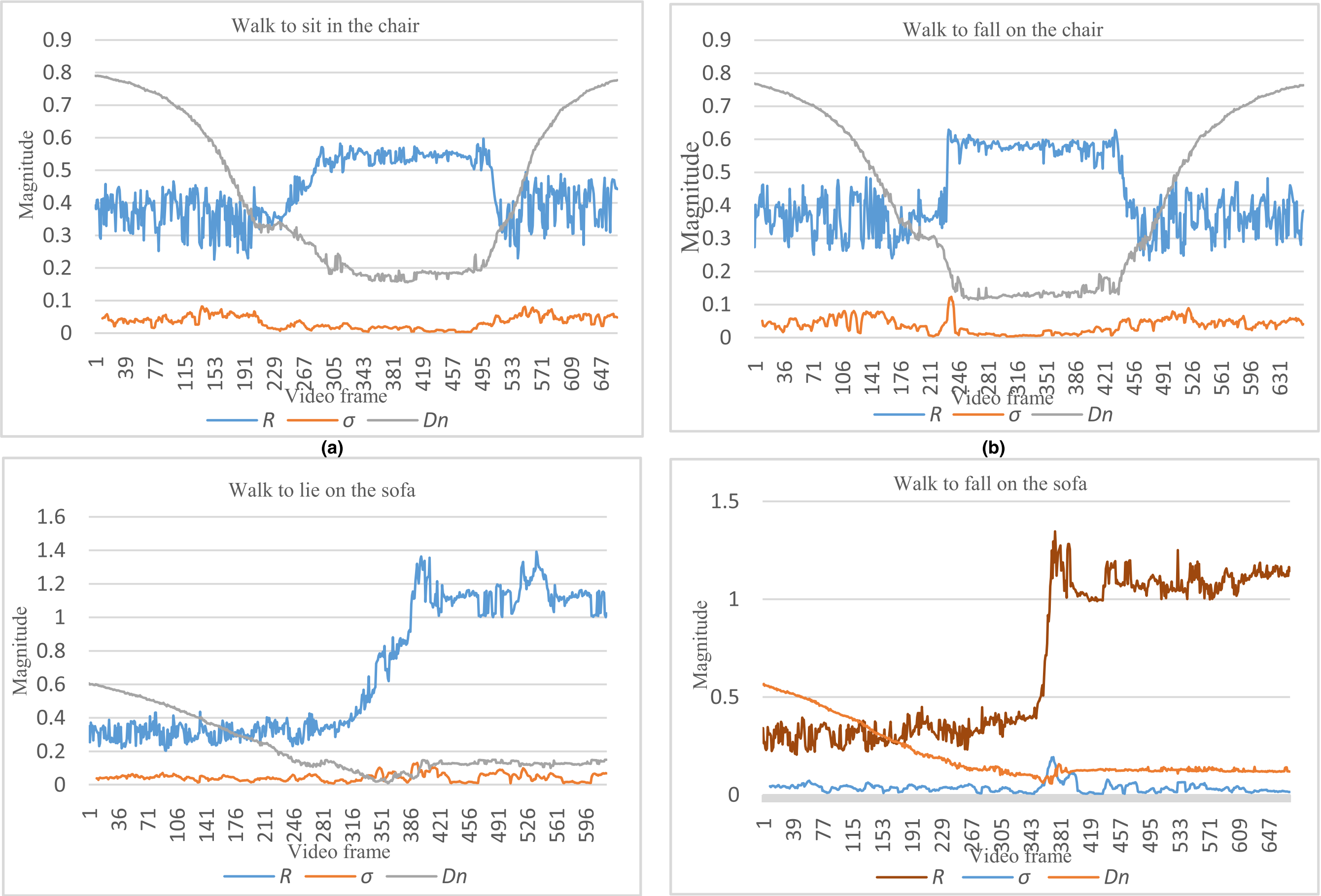
Thirdly, we discuss some special conditions of the fall such as falling on a chair or sofa. Our algorithm is able to effectively distinguish lying from falling. In the experimental sample videos, people's activities mainly include sitting in a chair, falling on a chair, lying on the sofa, and falling on the sofa. As shown in Fig. 5, we calculate and plot the changes of three features for four activities. In Fig. 6(a) and Fig. 6(b), we can see that it is difficult to distinguish between sitting in the chair and falling on the chair by R and Dn. But σ has a good effect for distinguishing both activities. Similarly, in Fig. 6(c) and 6(d), σ also has a good effect for distinguishing lying on the sofa from falling on the sofa.

According to the above experiments, the features of each behavior are within a fixed range. We can summarize the range of three parameters for different activities as shown in Table 2. Through a series of experiments, we record the ranges of each feature for different behaviors. We design an algorithm for recognizing falls and classifying ADLs, as shown in Fig.7. Through the range of different



(a) (b)

FIGURE 5. Some examples about the characteristics of ADLs. (a) and (b) are the changes of R and σ for four different activities.



(a) **(b)** **(c)** **(d)**

FIGURE 6. Some examples about the characteristics of special fall and fall-like activities. (a) and (b) are the changes of *R*, σ and *Dn* for sitting in the chair and falling on the chair. (c) and (d) are the changes of *R*, σ and *Dn* for lying on the sofa and falling on the sofa.

behavioral features, our method performs well in behavior classification and fall recognition. We perform a series of experiments to verify the performance of our algorithm in section IV.

C. ALGORITHM OF FALL AND ADL RECOGNITION

Based on the features extracted from activities and some feasibility analysis in activity characteristic of falls, we propose an automatic engine (as shown in Fig. 7) to

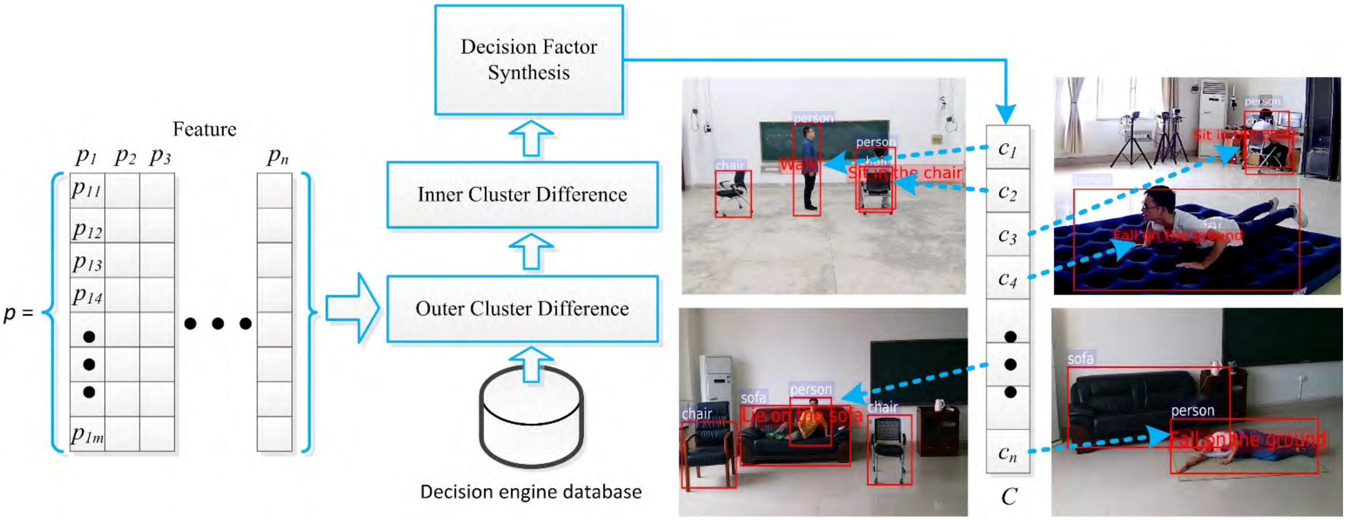


FIGURE 7. The procedure of our algorithm to classify activities.

TABLE 2. The range of three features for different activities.

Activity	R	σ	Dn(sofa)	Dn(chair)
Walk	< 0.50	< 0.10	\	\
Sit in the chair	0.50- 0.70	< 0.10	\	< 0.25
Sit in the sofa	0.50- 0.70	< 0.10	< 0.25	\
Lie on the sofa	0.60- 2.00	< 0.15	< 0.25	\
Fall	0.60- 2.00	> 0.20	\	\
Fall on a chair	0.50- 1.00	> 0.10	\	< 0.25
Fall on a sofa	0.60- 2.00	> 0.18	< 0.25	\

distinguish falls from ADLs, given that there are n persons in the scene. We definite $p = \{p_1, p_2, p_3, \ldots, p_n\}$ as the sample collection, and each sample has m characteristics $p_i = \{p_{i1}, p_{i2}, p_{i3}, \ldots, p_{im}\}$, where p_{ij} represents j^{th} features in i^{th} sample. n samples are divided into k classes which are denoted as $C = \{c_1, c_2, c_3, \ldots, c_k\}$. The minimum square error of k classes is denoted as

$$E = \sum\nolimits_{i=1}^{k} \sum\nolimits_{x \in C_i} \|x - \mu_i\|_2^2 \tag{8}$$

where $\mu_i = \frac{1}{|C_i|} \sum\limits_{x \in C_i} x$ is the mean vector of c_i. μ_i describes the compactness of the inner class mean vector. The smaller value of E means the higher similarity in outer class. Our algorithm in detail is shown in the Algorithm 1.

As shown in Fig. 7, the feature of each person is regarded as the input, and the features for different classes are stored in the decision engine database. Then, based on the decision engine database, we calculate the correlation coefficient between p_j and p_i' to judge the differences of outer class and calculate the distance between p_j and μ_i to judge the differences of inner class. Next, the decision factor is synthesized by the differences between outer class and inner class. Finally, we choose the minimum decision factor to classify input feature. Some results of classification are shown in Fig. 7. We can clearly see that our algorithm has a good distinction of falls from some activities of daily living such as sitting, lying down and falling. Here, we extract three kinds of features (R, σ and Dn) for three kinds of behaviors (walk, sit and fall). To prove our algorithm, we can intuitively see the result of classification in three features. One hundred and fifty samples are divided into 3 classes, denoted as $C = \{c_1, c_2, c_3\}$. Our original features are shown in Fig. 8, where the red points, green points and blue point represent the features of walking, sitting and falling, respectively. After 150 epochs, we obtain the result of classification in Fig. 9, where the red points, green points and blue point represent the behavior of walking, sitting and falling, respectively. The cluster centers are marked with black ×.

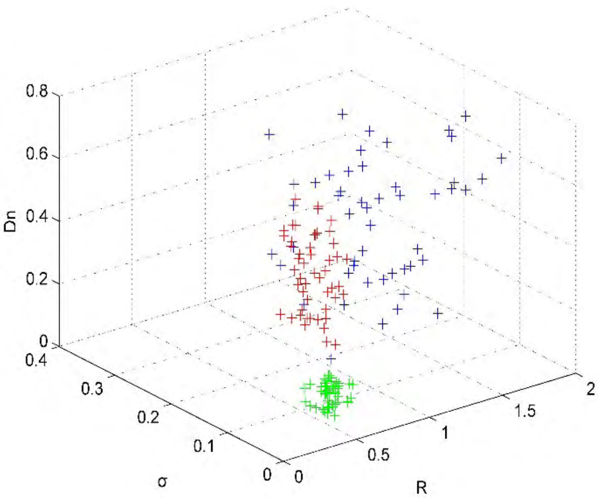


FIGURE 8. The distribution of original features' magnitude.

Algorithm 1 Automatic Decision Engine

```
Input: sample collection $p = \{p_1, p_2, p_3, \ldots, p_n\}$.
       The total number of class $k$.
Output: $C = \{c_1, c_2, c_3, \ldots, c_k\}$.
Steps:
$K$ samples are randomly selected from $C$ as the initial samples $\{p'_1, p'_2, \ldots, p'_k\}$, and its mean vector $\{\mu_1\mu_2, \ldots, \mu_k\}$
repeat
  $c_i = \emptyset\ (1 \le i \le k)$
  for $j = 1, 2, \ldots, m$ do
    Calculating the correlation coefficient between $p_j$ and $p'_i$:
      $r_{ji} = \frac{p_j \bullet p'_i}{\|p_j\|^2 + \|p'_i\|^2 - p_j \bullet p'_i}$
    Calculating the distance between $p_j$ and $\mu_i$:
      $\mathrm{d}_{ji} = \left\|p_j - \mu_i\right\|_2$
    Calculating decision factor $\psi_{ji}$: $\psi_{ji} = \mathrm{d}_{ji} + \kappa \frac{1}{r_{ji}}$
    Generating the classification mark of $p_j$:
      $\lambda_j = \arg\min_{i\epsilon\{1,2,\ldots,k\}} \psi_{ji}$
    Add $p_j$ to responding class: $\mathrm{c}_{\lambda_j} = \mathrm{c}_{\lambda_j} \bigcup \{p_j\}$
  for i = 1, 2, ..., k do
    Calculating the new mean vector: $\mu'_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$
    if $\mu'_i \ne \mu_i$ then
      Update $\mu_i = \mu'_i$
    else
      Don't update $\mu_i$
Until: all mean vectors don't need to update.
```

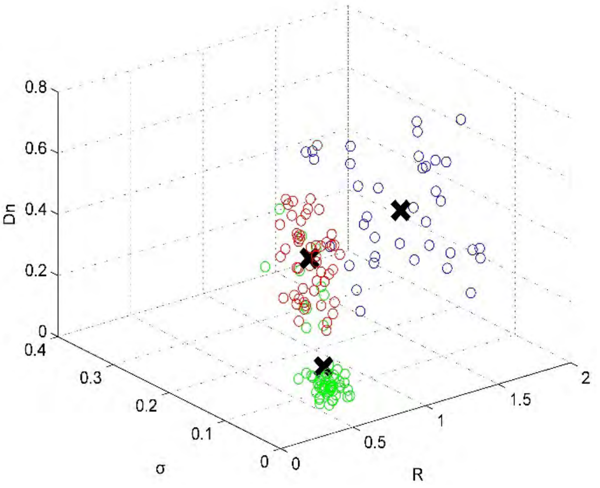


FIGURE 9. The distribution of classification result in our algorithm.

IV. EXPERIMENTS

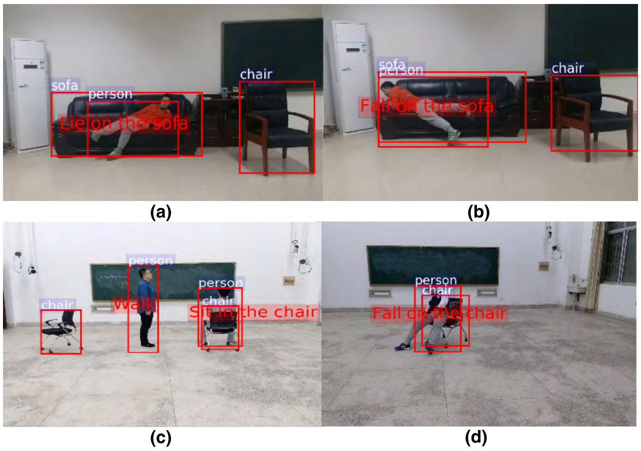
Our method was implemented using Ubuntu 16.04LTS + Tensorflow 1.0.0 + Opencv2.4.9 + on a PC using an Intel Core i7-4790 3.60 GHz processor and Quadro M4000, 8G RAM. Our main purpose is to distinguish falls in daily activities which contains walking, sitting, sleeping, lying down, falling on the ground, sitting in chair, sitting in sofa, falling on sofa and falling on chair. Most of the existing activity datasets lack the data we need. For example, the KTH Dataset and the Weizmann Dataset are commonly used to recognize human action, while the KTH Dataset contains 6 kinds of activities, such as walking, jogging, running, boxing, hand waving and hand clapping. The Weizmann Dataset contains 10 activities, such as bending, jacking, jumping, running, skipping, walking, wave1 and wave2 which do not provide all action samples we need. Ultimately, we chose UR fall detection dataset [44] to test the performance of our algorithm. UR fall detection dataset contains 70 video sequences (30 falls + 40 activities of daily living). Fall events are recorded with 2 Microsoft Kinect cameras and corresponding accelerometric data. ADL events are recorded with only one device and accelerometer. Since the approach we proposed is a vision based algorithm and no auxiliary equipment is required, RGB images recorded with camera 0 are used in our experiments.

We also collected our own datasets using HIKVISION DS-2D3304IW-D4 Webcam, which was fixed 1.6 meters above the ground. The distance between camera and object is about 4 meters. Person and furniture are included in our video samples. The self-collected dataset mainly contains seven actions, including walking, falling on the ground, falling on the furniture such as sofa or chair, sitting, lying down and so on. We collected 200 videos in various scenes. Each video contains 400 to 800 frames. The duration of each video is 20 to 30 seconds. There are total 100 fall videos (50 videos of falling on the ground and 50 videos of falling on furniture) and total 100 no-fall videos (25 videos of walking, 25 videos of sitting in chair, 25 videos of sitting in sofa and 25 videos of lying on sofa). For safety and realistic performance considerations, subjects performed the fall actions on a 5 cm-thick cushion.

A. QUALITATIVE ANALYSIS

Qualitative analysis was done first in our experiment. Since the existing datasets cannot provide enough data we need, especially for sitting or falling on furniture such as chair and sofa, our qualitative experiments were carried out using self-collected dataset according to previous experience. As discussed above, we propose a new detection method for human fall detection on furniture using scene analysis based on deep learning and activity characteristics. Unlike classical approaches only focusing on detecting human themselves, the proposed method first determines the spatial relation between human and furniture detected in scene analysis. To demonstrate the effectiveness of our method and difference from other methods, we will take some experiment examples to compare with other methods.

The detection results of our algorithm on self-collected dataset are shown in Fig. 10. According to Fig. 10(a) and Fig. 10(b), our method had a good performance to classify



(a) (b) (c) (d)

FIGURE 10. The detection results of our algorithm on self-collected dataset.

lying down and falling on sofa, whereas the shape aspect ratio based method [37] and the height based method [45] did not work well for this condition, both generating false detection and incorrectly judging lying down on sofa as fall behavior. As for Fig. 10(c) and Fig. 10(d), our method performed well in distinguishing the behaviors of sitting and falling on chair, whereas the height based method and the shape aspect ratio based method had failure of missed detection. Although the shape aspect ratio velocity based method [46] and the height velocity based method [47] sometimes correctly detected falls happened on furniture, the accuracy of these methods are lower than our method. However, even when these methods only detected falls happened, they could not identify and recognize that the falls were falling on furniture. Therefore, compared with other methods, our proposed method has better capability of distinguishing some fall-like behaviors and some special fall behaviors such as falling on furniture.

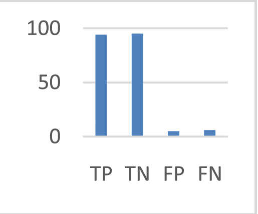
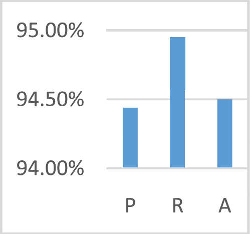
B. QUANTITATIVE ANALYSIS

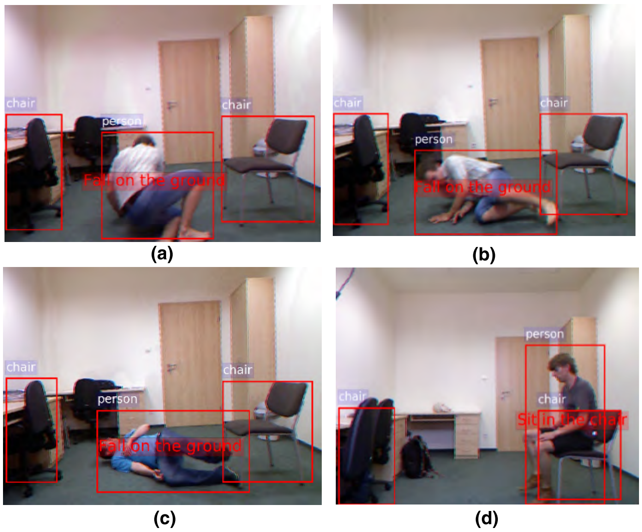
Quantitative analysis was done first on our self-collected dataset. By our proposed method, the number of true detection for different behaviors is recorded in Table 3. We obtained a good performance to distinguish some special falls and some fall-like behaviors. Seven kinds of activities in our self-collected dataset were classified accurately, especially for distinguishing falling on sofa from lying down, and distinguishing sitting in chair from falling on chair. Here, TP (True Positive) means the fall samples judged as falls. TN (True Negative) means the no-fall samples judged as no-falls. FP (False Positive) means the no-fall samples judged as fall. FN (False Negative) means the fall samples judged as no-fall. We have R($recall$) $= \frac{TP}{TP+FN}$, P($precision$) $= \frac{TP}{TP+FP}$ and A($accuracy$) $= \frac{TP+TN}{TP+TN+FP+FN}$.

To further verify the robustness of our proposed method, the method was compared with the existing advanced methods using quantitative analysis. The fall detection accuracy of the proposed method was compared with other methods using ROC curves and AUC (Area Under ROC Curve). In recent years, some methods such as shape aspect ratio based method, height based method, shape aspect ratio velocity based method and height velocity based method were proposed. Although they have some variations, their major ideas still just focus on detection of human themselves. We have conducted our comparison experiments against the shape aspect ratio based method [37], the height based method [45], the shape aspect ratio velocity based method [46] and the height velocity based method [47].

Both the standard dataset and our self-collected dataset were used to construct our comparative experiments. Thirty falls and 40 ADLs which were shuffled and collected in the UR fall detection dataset [44] are regarded as the standard dataset. Some detection results of our algorithm on standard dataset are shown in Fig. 11. Two hundred self-collected videos which contain seven actions including walking, falling

TABLE 3. Experimental results in our proposed method.

Activity	The total number of videos	The number of detected as fall videos	
Walk	25	0	
Sit in the chair	25	1	
Sit in the sofa	25	2	
Lie on the sofa	25	1	
Fall on the ground	50	48	
Fall on the chair	25	22	
Fall on the sofa	25	24	



(a) (b) (c) (d)

FIGURE 11. The detection results of our algorithm on standard dataset.

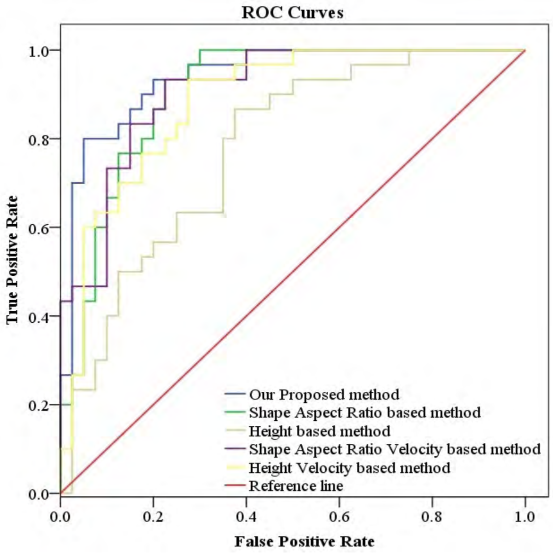


FIGURE 12. ROC curves of fall detection methods on the self-collected dataset.

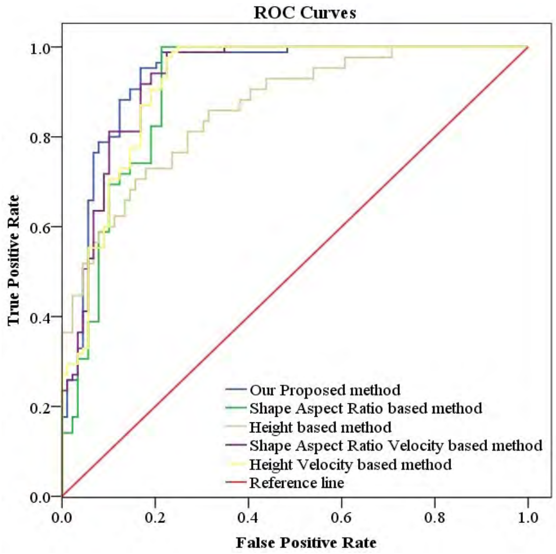


FIGURE 13. ROC curves of fall detection method on the standard dataset.

on the ground, falling on furniture such as sofa or chair, sitting and lying down, are regarded as the self-collected dataset. The detection results of our algorithm on self-collected dataset are shown in Fig. 10. The standard dataset was utilized to verify the universality of the method. Because the self-collected images are more complex and closer to the fall condition, they are used to verify the superiority of our proposed method.

As we all know, if the ROC curve of algorithm A is encased by algorithm B, it means the performance of B is better than A. It is shown from Fig. 12 and Fig. 13 that the ROC curve of our proposed method almost encases the other four methods on self-collected dataset and standard dataset. Fig. 12 and Fig.13 also demonstrate that our proposed method has the higher true positive rate than the other four methods on the two testing datasets. We can conclude that our proposed method performs better than other four methods in both the self-collected dataset and the standard dataset.

TABLE 4. The magnitude of AUC for different methods.

Algorithms	The magnitude of AUC	
	Self-collected dataset	Standard dataset
Our proposed method	0.935	0.941
Shape aspect ratio based method	0.904	0.905
Height based method	0.862	0.777
Shape aspect ratio velocity	0.927	0.909
Height velocity based method	0.917	0.884

Besides, the magnitude of AUC also can determine the performance of algorithm. The larger the area of AUC is, the better the performance of the method [48]. The magnitude of AUC is calculated based on ROC curves and shown in the Table 4. It is shown from the Table 4 that the proposed method has the largest AUC area on the two testing dataset. All of the experimental results can demonstrate that our method has a much better effect in detecting human fall behaviors, especial for distinguishing the falls on furniture from not only ADLs but also sitting or lying on the furniture.

V. CONCLUSION

In this paper, we propose a new method for detecting human fall on furniture using scene analysis based on deep learning and activity characteristics. The proposed method first performs scene analysis using a deep learning method Faster R-CNN to detect human and furniture such as sofa. Unlike classical approaches only focusing on detecting human themselves, the spatial relation between human and furniture are detected in the scene. A series of activity characteristics are detected and tracked. By means of measuring the changes of these characteristics and judging the relations between people and furniture nearby, the falls on furniture can be effectively detected. In our experimental result on the self-collected dataset and the standard dataset, falls can be accurately and effectively distinguished from other fall-like activities such as sitting or lying down. Some special fall behavior such as falling on sofa or chair can be distinguished accurately, which are very difficult for other existing approaches to differentiate. Besides, the qualitative and quantitative experiments demonstrate that our proposed method improves accuracy and overall performance in fall detection. As for the fall-like behaviors and special fall behaviors like falling on furniture, our method achieved excellent robustness to distinguish them.

REFERENCES

[1] C. Wang, S. J. Redmond, W. Lu, M. C. Stevens, S. R. Lord, and N. H. Lovell, "Selecting power-efficient signal features for a low-power fall detector," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2729–2736, Nov. 2017.

[2] M. Heron, "Deaths: Leading causes for 2007," *Nat. Vital Stat. Rep.*, vol. 59, no. 8, pp. 1–95, Aug. 2011.

[3] P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, "A high reliability wearable device for elderly fall detection," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4544–4553, Aug. 2015.

[4] K. Ozcan, S. Velipasalar, and P. K. Varshney, "Autonomous fall detection with wearable wameras by using relative entropy distance measure," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 31–39, Feb. 2017.

[5] A. Ejupi, M. Brodie, S. R. Lord, J. Annegarn, S. J. Redmond, and K. Delbaere, "Wavelet-based sit-to-stand detection and assessment of fall risk in older people using a wearable pendant device," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1602–1607, Jul. 2017.

[6] L. N. V. Colón, Y. DeLaHoz, and M. Labrador, "Human fall detection with smartphones," in *Proc. IEEE Latin-Amer. Conf. Commun. (LATINCOM)*, Cartagena de Indias, Colombia, Nov. 2014, pp. 1–7.

[7] M. J. Mathie, A. C. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Meas.*, 25, no. 2, pp. R1–R20, Apr. 2004.

[8] T. T. Nguyen, M. C. Cho, and T. S. Lee, "Automatic fall detection using wearable biomedical signal measurement terminal," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Minneapolis, MN, USA, Sep. 2009, pp. 5203–5206.

[9] A. Sixsmith and N. Johnson, "A smart sensor to detect the falls of the elderly," *IEEE Pervasive Comput.*, vol. 3, no. 2, pp. 42–47, Apr. 2004.

[10] M. Kangas, I. Vikman, J. Wiklander, P. Lindgren, L. Nyberg, and T. Jämsä, "Sensitivity and specificity of fall detection in people aged 40 years and over," *Gait Posture*, vol. 29, no. 4, pp. 571–574, Jun. 2009.

[11] B. Andò, S. Baglio, C. O. Lombardo, and V. Marletta, "A multisensor data-fusion approach for ADL and Fall classification," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 1960–1967, Sep. 2016.

[12] M. Alwan, P. J. Rajendran, S. Kell, and D. Mack, "A smart and passive floor-vibration based fall detector for elderly," in *Proc. IEEE Inf. Commun. Technol. (ICTTA)*, Apr. 2008, pp. 1003–1007.

[13] G. Feng, J. Mai, Z. Ban, X. Guo, and G. Wang, "Floor pressure imaging for fall detection with fiber-optic sensors," *IEEE Pervasive Comput.*, vol. 15, no. 2, pp. 40–47, Apr./Jun. 2016.

[14] B. Andò, S. Baglio, C. O. Lombardo, and V. Marletta, "An event polarized paradigm for ADL detection in AAL context," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1814–1825, Jul. 2015.

[15] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the Microsoft Kinect," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 290–301, Jan. 2015.

[16] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 69–72.

[17] D. Droghini, D. Ferretti, E. Principi, S. Squartini, and F. Piazza, "A combined one-class SVM and template-matching approach for user-aided human fall detection by means of floor acoustic features," *Comput. Intell. Neurosci.*, vol. 2017, May 2017, Art. no 1512670, doi: 10.1155/2017/1512670.

[18] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "Radar and RGB-depth sensors for fall detection: A review," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3585–3604, Jun. 2017.

[19] Q. Zhang and M. Karunanithi, "Feasibility of unobstrusive ambient sensors for fall detections in home environment," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 566–569.

[20] D. Liciotti, G. Massi, E. Frontoni, A. Mancini, and P. Zingaretti, "Human activity analysis for in-home fall risk assessment," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, U.K., 2015, pp. 284–289.

[21] Y. Nizam, M. N. H. Mohd, and M. M. A. Jamil, "Human fall detection from depth images using position and velocity of subject," *Procedia Comput. Sci.*, vol. 105, pp. 131–137, Dec. 2017.

[22] X. Wang, H. Liu, and M. Liu, "A novel multi-cue integration system for efficient human fall detection," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Qingdao, China, Dec. 2016, pp. 1319–1324.

[23] A. Zerrouki and A. Houacine, "Combined curvelets and hidden Markov models for human fall detection," *Multimedia Tools and Applications*, Mar. 2017, pp. 1–20, doi: 10.1007/s11042-017-4549-5.

[24] Z. W. Sun and X. W. Sun, "A human fall detection algorithm based on acceleration sensor," *Comput. Eng. Sci.*, to be published.

[25] H. Liu, D. Liu, X. Sun, F. Wu, and W. Zeng, "On-line fall detection via a boosted cascade of hybrid features," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Hong Kong, Jul. 2017, pp. 249–254.

[26] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth-based human fall detection via shape features and improved extreme learning machine," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1915–1922, Nov. 2014.

[27] Y. Yun and I. Y. H. Gu, "Human fall detection via shape analysis on Riemannian manifolds with applications to elderly care," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 3280–3284.

[28] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Proc. 21st Int. Conf. Adv. Inf. Netw. Appl. Workshops (AINAW)*, Niagara Falls, ON, Canada, 2007, pp. 875–880.

[29] C.-F. Lai, Y.-M. Huang, J. H. Park, and H.-C. Chao, "Adaptive body posture analysis for elderly-falling detection with multisensors," *IEEE Intell. Syst.*, vol. 25, no. 2, pp. 20–30, Mar./Apr. 2010.

[30] S. W. Abeyruwan, D. Sarkar, F. Sikder, and U. Visser, "Semi-automatic extraction of training examples from sensor readings for fall detection and posture monitoring," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5406–5415, Jul. 2016.

[31] L. Malheiros, G. D. A. Nze, and L. X. Cardoso, "Fall detection system and body positioning with heart rate monitoring," *IEEE Latin Amer. Trans.*, vol. 15, no. 6, pp. 1021–1026, Jun. 2017.

[32] P. V. G. F. Dias, E. D. M. Costa, M. P. Tcheou, and L. Lovisolo, "Fall detection monitoring system with position detection for elderly at indoor environments under supervision," in *Proc. 8th IEEE Latin-Amer. Conf. Commun. (LATINCOM)*, Medellín, Colombia, Nov. 2016, pp. 1–6.

[33] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 156–167, Jan. 2006.

[34] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "HMM based falling person detection using both audio and video," in *Proc. IEEE 14th Signal Process. Commun. Appl. Conf.*, Antalya, Turkey, Apr. 2006, pp. 211–220.

[35] S. Gasparrini, E. Cippitelli, S. Spinsante, and E. Gambi, "A depth-based fall detection system using a kinect sensor," *Sensors*, vol. 14, no. 2, pp. 2756–2775, 2014.

[36] H. Nait-Charif and S. J. McKenna, "Activity summarisation and fall detection in a supportive home environment," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4. 2004, pp. 323–326.

[37] W. D. Min, L. S. Wei, Q. Han, and Y. Z. Ke, "Human fall detection based on motion tracking and shape aspect ratio," *Int. J. Multimedia Ubiquitous Eng.*, vol. 11, no. 10, pp. 1–14, 2016.

[38] K. Ozcan, A. K. Mahabalagiri, M. Casares, and S. Velipasalar, "Automatic fall detection and activity classification by a wearable embedded smart camera," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 3, no. 2, pp. 125–136, Jun. 2013.

[39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[40] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[41] L. Huang, Y. Yang, Y. Deng, and Y. Yu, "DenseBox: Unifying landmark localization with end to end object detection," *Comput. Sci.*, 2015. [Online]. Available: https://arxiv.org/abs/1509.04874

[42] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2874–2883.

[43] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[44] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 489–501, Dec. 2014.

[45] L. Yang, Y. Ren, H. Hu, and B. Tian, "New fast fall detection method based on spatio-temporal context tracking of head by using depth images," *Sensors*, vol. 15, no. 9, pp. 23004–23019, Sep. 2015.

[46] G. Mastorakis and D. Makris, "Fall detection system using Kinect's infrared sensor," *J. Real-Time Image Process.*, vol. 9, no. 4, pp. 635–646, 2014.

[47] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "3D head tracking for fall detection using a single calibrated camera," *Image Vis. Comput.*, vol. 31, no. 3, pp. 246–254, 2013.

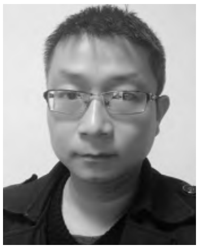
[48] S. Yu, Y. Wu, L. Wei, Z. Song, and W. Zeng, "A model for fine-grained vehicle classification based on deep learning," *Neurocomputing*, vol. 257, pp. 97–103, Sep. 2017.

WEIDONG MIN (M'12) received the B.E., M.E., and Ph.D. degrees in computer application from Tsinghua University, China, in 1989, 1991, and 1995, respectively. He was an Assistant Professor at Tsinghua University from 1994 to 1995. From 1995 to 1997, he was a Post-Doctoral Researcher at the University of Alberta, Canada. From 1998 to 2014, he was a Senior Researcher and Senior Project Manager at Corel and other companies in Canada. In recent years, he cooperated with School of Computer Science & Software Engineering, Tianjin Polytechnic University, China. Since 2015, he has been a Professor with the School of Information Engineering, Nanchang University, China. He is a Member of The Recruitment Program of Global Expert of Chinese Government. He is the Executive Director of China Society of Image and Graphics. His current research interests include computer graphics, image and video processing, distributed system, software engineering, and network management.

HAO CUI received the B.E. degree in communication engineering from Huaqiao University, China, in 2016. He is currently pursuing the master's degree in behavior detection and object detection at Nanchang University, China.

HONG RAO received the B.S., M.S., and Ph.D. degrees from Nanchang University in 1994, 2004, and 2009, respectively. She is currently a Professor with the School of Information Engineering, Nanchang University. Her research interests are in the definition of intelligent systems adopting machine learning, computational intelligence, and large-scale data analysis. She has authored over 30 publications in these areas.

ZHIXUN LI received the M.Sc. degree in computer science and technology from Université Toulouse III, Toulouse, France, in 2006, and the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, China, in 2017. He is currently a Lecturer with the School of Information Engineering, Nanchang University, China. His primary research interests are medical image processing and pattern recognition.



LEIYUE YAO received the master's degree from Nanchang University in 2006, where he is currently pursuing the Ph.D. degree with the School of Information Engineering. He has been conducting research in the field of information processing and computer vision.