



Fall detection approach based on combined two-channel body activity classification for innovative indoor environment

Anurag De¹ · Ashim Saha¹ · Praveen Kumar² · Gautam Pal³

Received: 17 June 2021 / Accepted: 10 January 2022 / Published online: 24 January 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Human fall detection plays a vital role in monitoring senior citizens safely while being alone. In recent works, vision-based techniques provide favorable and effective results. In this paper, a combined two-channel fall detection approach is proposed using Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) classification models based on the displacement of significant spatial features of the foreground image. Initially, training of both fall and daily activity scenarios is done using a standard fall detection dataset. Keyframes consisting of significant body shape features are then obtained from the surveillance video subjected to the two-channel classification model. We consider the classification results if both the channels generate similar outputs, failing which, additional intelligence is used to classify the fall and daily activity event. The keyframe selection is based on the displacement in height-to-width ratio and displacement in horizontal and vertical centroid movement of the object having a threshold higher than a preset value. The proposed fall detection system achieves a peak accuracy of 98.6% and sensitivity of 100% in detecting falls. The proposed model achieves satisfactory performance in comparison to existing state-of-the-art techniques.

Keywords Fall detection · Human shape · Daily life activity · Keyframe selection · Classification

1 Introduction

Fall incidents constitute a significant risk factor for older people and senior citizens living alone. As per the United Nations (2017) report, the world population is overgrowing, and also the number of senior citizens of age 60 years and above due to improved healthcare facilities. It is forecasted that the population growth of senior citizens will be exceedingly fast, surpassing 1000 million by 2030 and 2000 million by 2050. In a recent study, Vollset et al. (2020) funded by the Bill and Melinda Gates Foundation, the population of senior citizens above 80 years of age is predicted to touch 866 million by the end of the century from around 141 million at 2017. As most elderly citizens remain indoors and

most of the time lonely, a fall incident can eventually result in a severe injury. It could be fatal if not taken care of in time. Subsequently, various fall detection methodologies have been evolving and gaining significance in recent times (Ramachandran and Karupiah 2020; Wang et al. 2020). These systems can detect a fall incident and proactively send notifications so that necessary measures can be taken to save precious lives. The usage of surveillance devices for different applications has increased over the years with the advancement in technology. Surveillance data is stored and fetched based on the requirement. Tracking surveillance data manually in real-time is a time-consuming job Sulman et al. (2008). To address this issue, detecting abnormal human activity like falls automatically without any human intervention is the solution (Li et al. 2018; Shu and Shu 2021; Chelli and Patzold 2019; Han et al. 2020; Kalinga et al. 2020). This paper underlines one such method, i.e., automatic detection of human fall concerning normal daily living activities.

Generally, fall detection methods are divided into two groups. The first group is device-based, automatically detecting possible falling activities from normal daily living activities using electronic devices. This group is further subdivided into wearable device-based systems and

✉ Anurag De
anurag.de111@gmail.com

¹ Department of Computer Science and Engineering, NIT Agartala, Agartala, India

² Department of Computer Science and Engineering, VNIT Nagpur, Nagpur, India

³ Department of Computer Science and Engineering, TIT Narsingarh, Agartala, India

non-wearable device-based systems. Wearable devices (Xi et al. 2020; Makhoulf et al. 2018; Hussain et al. 2019) gather information from electronic instruments like accelerometers, gyroscopes, magnetometers, etc., to recognize a fall event. Even though these instruments are cost-effective, they need to be worn at all-time, which is a significant disadvantage. This may result in difficulty for elderly persons, and they might also forget to wear them every time. Also, using the help button Directalert (2021) is of no use if the person becomes immobilized or remains unconscious after the fall. Moreover, rechargeable batteries are needed for these devices, which require to be recharged regularly for effective functioning. Non-wearable devices are mostly antenna-based equipment that is generally installed in indoor conditions. They use built-in sensors to measure parameters such as the floor vibration, the pressure exerted on the floor to recognize a falling behavior successfully, mapping the amplitude of a wireless signal to human motion (Sadreazami et al. 2019; Huang et al. 2019). However, these systems have the disadvantage of experiencing a high false alarm rate. A minor fluctuation in the indoor situation can give rise to a pressure difference or introduce noise that may not be through human beings.

The second group belongs to the vision-based methods, which became very popular in recent years. It has a benefit that does not require wearing any instrument all the time as it requires only video surveillance cameras to be installed for detecting a fall event in real-time.

The proposed method in this paper is based on the assumption that the body shape of a person varies during any activity. The change in shape is significant during a fall incident in comparison to any other daily life activity. Falls generally occur accidentally, due to some kind of weakness (Rubenstein 2006; WHO 2018), or epileptic seizures (Rubenstein 2006; Russell-Jones and Shorvon 1989; Krumholz and Hopp 2008; WHO 2018), etc. Activities of daily living (ADL) are like sitting down, bend to pick up something, or lie down, etc. This variation in the body shape of the individual helps to identify a fall from a daily living activity. The primary contributions of the proposed work are summarized as follows:

- This paper achieves novelty through the extraction of keyframes represented by significant shape-change features. Instead of selecting frames randomly for classification, keyframes representing different fall and fall-like activity events are chosen. The keyframe selection helps distinguish a stable or stationary phase from an activity phase and further restricts the start and end of a fall or daily activity event in a video sequence. Selecting keyframes for classification also helps achieve better time complexity as it filters the redundant frames for training the system (Makandar et al. 2015; Luo et al. 2018).

- For reducing the huge processing time, foreground movement features from the input video sequence are estimated at intervals of five frames instead of all frames. This strategy helps reduce the computation time without affecting the keyframes for classification.
- A combination of machine learning-based and threshold-based approaches is used to design the fall detection model to make the system more robust. The proposed method outperforms state-of-the-art models on detecting both falling and daily activity instances.
- A low-cost system is designed based on RGB frames as input. A simple and economical RGB surveillance camera is sufficient for the video acquisition of the proposed approach.

The remainder of the paper is organized as follows: Sect. 2 presents a detailed survey of existing approaches related to human fall and daily living activity detection. Section 3 presents the proposed fall detection methodology. Section 4 shows the experimental results and the performance evaluation of the proposed system, followed by a comparison with existing methods. Section 5 concludes the paper along with the scope for future enhancements.

2 Related work

In this section, the literature survey of fall detection methods is discussed. This survey is focused on fall detection using wearable devices, non-wearable devices, and vision-based approaches.

Zitouni et al. (2019) designed an innovative sole embedded with a fall detection algorithm based on a single tri-axis accelerometer. Thresholds depending on acceleration, position, and duration parameters are used by the algorithm to detect a fall concerning activities of daily living. Although the authors claim the system to be unobtrusive and the sole is comfortable to wear, the footwear fitted with the smart sole needs to be worn around the clock, which does not seem convenient for the elderly. Sabatini et al. (2016) discussed a fall detection approach based on a barometric altimeter. This system differentiates among various kinds of falls and daily activities. The technique is devised based on a bi-fold decision mechanism. In the first stage, significant high-impact falls are detected by estimating the vertical velocity. The second stage uses impact features, posture, and height change features, alone or combining, for post-impact fall detection or soft fall detection. The technique has a low sensitivity of 80% for significant high-impact fall detection and 100% sensitivity for soft-fall detection. Kerdjijdj et al. (2020) implemented an automatic fall and daily activity detection system. This system used a lightweight and easy-to-wear platform. An accelerometer, magnetometer, gyroscope, and

electrocardiogram (ECG) are used to provide a large amount of data. However, frequent battery charging is needed to keep it in operating condition.

The literature related to fall detection discussed above is based on sensors and devices attached to the body. These methods give good results but have few drawbacks in real-time scenarios like frequent recharge is needed, the extra burden to wear, especially for old aged personals. The use of non-wearable and surveillance-based devices overcomes the limitations of wearable devices. Some of the existing works based on non-wearable device-based approaches are mentioned below.

Wang et al. (2017) presented a real-time, contactless, low-cost indoor fall detection technique. This technique relies on the fine-grained channel state information (CSI) phase and amplitude found in commodity Wi-Fi devices. The CSI phase difference is used here to observe fall and fall-like typical actions. Simultaneously, the sharp power profile decline pattern in the time–frequency domain is used to attain accurate recognition. The approach gives good performance with indoor light intensity changes. But its performance degrades while moving into a new environment with an old setting. Junior and Adami (2018) presented a hybrid fall detection method based on analysis of thresholds, analysis of the device's orientation, and decision trees algorithm to detect falls from normal daily life activities. The data in this study was collected using a smartphone through its tri-axis accelerometer. A different approach is presented by Huang et al. (2019). The authors propose a fall detection system based on geophones that receive the floor vibration signals. It analyzes the vibration signals to extract time-dependent features to detect a potential fall concerning daily life activities using the Hidden Markov Model (HMM). The system works effectively for a single individual residing at home. However, it struggles to detect falls of multiple persons as it lacks significantly in processing multiple floor vibration signals.

Though non-wearable techniques are better in comparing with wearable based approaches, it also has few limitations. Environmental noise can alter the wireless signals used by the non-wearable process, which is one of the main limitations. These drawbacks of non-wearable systems are overcome using vision or surveillance-based methods and gained popularity in real-life fall detection systems.

Peng et al. (2019) proposed a vision-based fall detection approach based on the human point cloud. Depth data is input to the system using Kinect. The method further maps the depth information into a point cloud image which represents humans using color spectrum. The height change acceleration feature is used to identify fall behavior. The system can recognize potential falls and activities such as sitting, squatting, walking, and bending. However, it lacks significantly in distinguishing between falls and controlled

lying, such as sleeping or lying on the floor or any surface. Htun et al. (2020) proposed a vision-based surveillance system based on image processing techniques. A hidden Markov model is used for detecting falls and daily living activities. Human shape-based features such as Silhouette surface area, centroid height, and bounding box aspect ratio are used to analyze the person in the frame. The system shows a recall of 98.37% using experiment videos containing both regular and non-abnormal events, including falls. Ample scope is present for incorporating multiple persons fall detection in a frame as the scope of the work is limited to a single person. Geertsema et al. (2019) presented a vision-based fall and daily activity detection system using vertical speed and acceleration obtained from optical flow vectors. It also uses sound amplitude as a feature to enhance specificity. A specificity of 92% and 99.7% is achieved by the algorithm using a public dataset and real-life data sequences, respectively. The system accurately detects high-impact falls but lags significantly in the case of low-impact falls.

Kepski and Kwolek (2015) designed a low-cost fall detection system. Authors use motion information from accelerometer data and depth images acquired using Kinect sensors. Spatial features extracted from the depth images are analyzed only when the person's movement is above a predefined threshold. This strategy helps reduce the computation cost. It also reduces the false alarms by combining both motion features and spatial features of the depth images. Even though the system achieves 95.71% accuracy and 100% sensitivity for detecting falls, it retains the limitations of wearing a body-worn device to assist in the fall detection process. Merrouche and Baha (2017) proposed a vision-based fall detection system using a depth camera based on the combination of human shape analysis, head, and centroid tracking to detect falls and normal daily life activities. The system aims to upgrade itself to address more complicated movements such as backward fall and fall from chair using an appropriate dataset. A vision-based fall detection system is designed by Wang et al. (2019) using Convolutional Neural Networks (CNN). The authors implement the transfer learning concept to train the VGG-16 network to identify a fall movement in a frame. The frames are pre-processed using background subtraction and morphological operations. Although the algorithm performs well in normal lighting conditions, it lacks significantly in low-light environments.

In this paper, a new approach for detecting human fall is proposed based on significant foreground features of the surveillance video using a combination of classic machine learning-based and threshold-based techniques. Video frames with the moving person are selected for a two-channel classification depending upon the feature displacement concerning a predefined threshold. The classification results help identify a fall from a regular daily life activity. Section 3 presents the detailed methodology.

3 Methodology

The proposed approach comprises the following steps: video acquisition and frame extraction, moving object/human detection through foreground segmentation, morphological noise reduction, model design based on training samples, keyframes selection, and a keyframe-based two-channel classification. Figure 1 shows the schema of the proposed method. Figure 2 illustrates the flowchart of the proposed fall detection methodology.

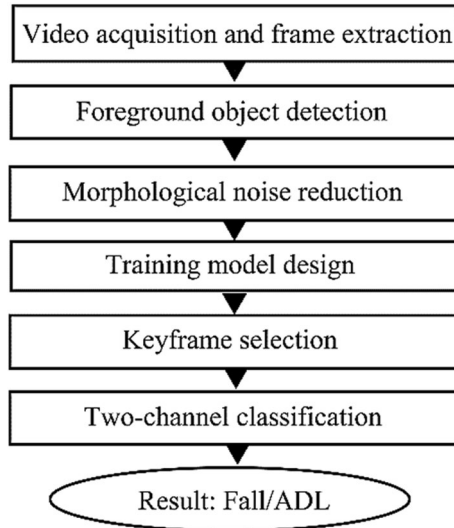


Fig. 1 Schema of the proposed methodology

3.1 Video acquisition and frame extraction

Video acquisition refers to obtaining a video sequence in real-time. The first step is to process a particular frame to detect a falling activity by interpreting the motion sequence. In this method, the University of Rzeszow Fall Detection Dataset (URFD Dataset) consisting of 30 falls and 40 daily activity video sequences are used Kepski and Kwolek (2014). It is a standard publicly available dataset. It is rarely possible to gather real-time fall sequences of elderly persons. Thus, they used simulated fall and daily life activity video sequences of young volunteers for processing. Each video sequence has a frame rate of 30 frames/second and a frame resolution of 640×240 pixels.

We observed that an individual's body shape does not alter much in every adjacent frame of a video sequence. Instead, it changes at an interval of few frames based on the type of activity. Hence, it becomes necessary to filter the redundant frames representing similar body shapes for a set of frames. This approach will significantly improve the overall processing time of the algorithm. A threshold interval of five frames is chosen based on logical reasoning and through evaluation and observation of the dataset sequences. The spatial information obtained from the first frame is used to recover at least the next five frames, which helps reduce the time complexity of processing the frames with similar body shapes. We choose a threshold of five frames as the maximum filtering threshold, and anything above it may compromise the valuable information required. It helps retain the effective posture change frames and filters the redundant frames with similar postures.

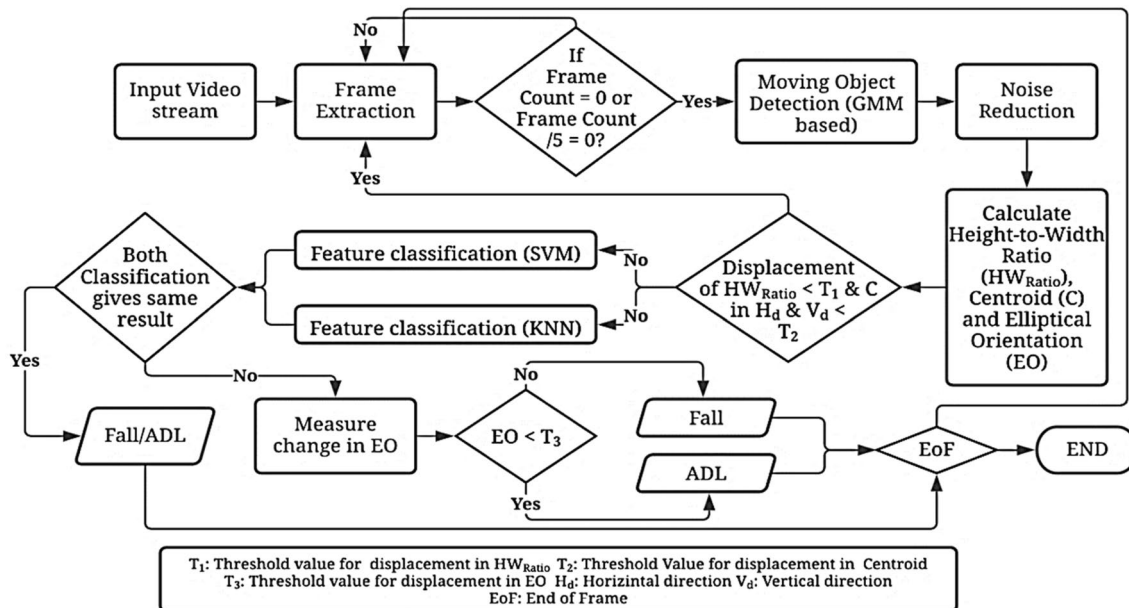


Fig. 2 Flowchart of the proposed fall detection system

3.2 Moving object detection

The frames extracted from the input video stream are subject to detection of the moving object in the frame. Locating the moving region in the frame is the initial work in distinguishing the motion behavior of the object. We use an adaptive background mixture model (Stauffer and Grimson 1999; Paul et al. 2013) to isolate the moving object in real-time from the video stream. It is quite robust against various lighting conditions. The algorithm models every pixel as a mixture of Gaussian and accordingly updates the background information using an approximation. This update technique helps the system in adapting itself to changes in illumination and objects that stopped moving. It tracks the transformation of the corresponding pixel's state from one frame to another. The pixels experiencing no state change are assigned with weight-0 (black), i.e., background pixels, and those changing states are given weight-1 (white), i.e., foreground pixels. The background pixels mostly do not change state. Hence, the foreground pixels represent the moving object in the video frame.

3.3 Morphological noise reduction

The foreground image contains the motion region of interest (RoI). It denotes the location of the moving object from the rest of the image frame. It is separated using both binary statistical morphological operations, i.e., erosion and dilatation Jamil et al. (2008). Binary statistical erosion is used to remove the noise from the image by eliminating the isolated noisy pixels. Dilatation recovers the loss caused by erosion by filling holes to retrieve the effective pixels removed during erosion. It unites the areas split during the binarization of the image frame. Finally, the obtained ROI is brought together into a moving object region through a connected

component criterion, customarily called a blob. It helps cluster multiple moving areas considered part of a single moving object or other moving regions into a single moving object. Hence, it plays an essential role in detecting moving humans in a frame and during the target object's occlusion.

3.4 Training fall and daily activity postures

In this section, the fall detection model based on spatial components is designed by training the posture samples of these two data classes. As discussed earlier, it is considered that a fall and daily activity movement possess different body stances. In this study, the URFD Fall Detection dataset Kepski and Kwolek (2014) is used for training purposes. This dataset consists of 30 fall events and 40 daily activity video sequences (such as bending to pick, sitting on the chair, sitting on knees, crouching, lying, etc.) respectively. We have designed the training model using the frontal URFD video sequences. Figures 3 and 4 show example training samples of fall and ADL. Representatives for both fall and daily activities are obtained from selected surveillance sequences. These samples are then trained to generate a spatial model using support vector machine (SVM) and K-nearest neighbors (K-NN) classifiers.

3.5 Keyframe selection

Random selection of frames for classification may not result in optimum classification results. Thus, frames with the displacement of the significant foreground features above a certain threshold are a suitable choice. It leads to effective classification and, simultaneously helps in achieving an improved time complexity.

Here the selection of keyframes is based on the observation that the horizontal or vertical or both the person's



Fig. 3 Training samples of falling events

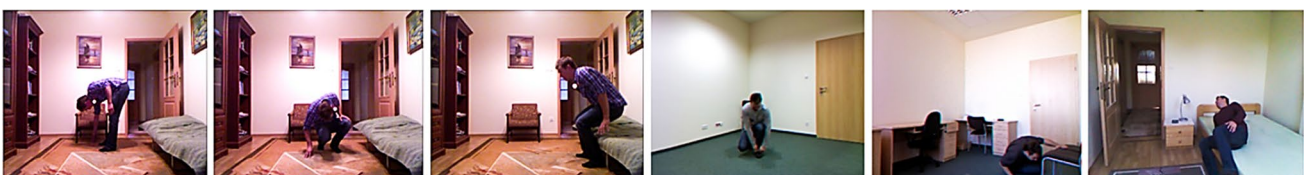


Fig. 4 Training samples of daily life activities

displacement is significantly higher during a fall and fall-like daily activity than during a non-falling or inactivity. Hence, based on this concept, we analyze the frames by measuring three most significant geometric shape-based features: blob height-to-width ratio, horizontal centroid displacement, and vertical centroid displacement. These parameters are described as follows.

Blob height-to-width ratio is the aspect ratio of the detected foreground object and can be calculated as shown in Eq. (1):

$$hwr(t) = h(t)/w(t), \tag{1}$$

where $hwr(t)$ denotes the height-to-width ratio of the object derived by dividing the value of the height (h) by that of the width (w) for the object at time t .

To define the centroid of the foreground object, we consider p to be a position at time t on the detected foreground object. Then, the centroid $C(t)$ is defined in Eq. (2) as:

$$C(t) = (x_{Ch}(t), y_{Cv}(t)) \tag{2}$$

Here, $p(t) = (x(t), y(t))$ and x_{Ch} represents the horizontal centroid component as shown in Eq. (3):

$$x_{Ch} = \sum_{i=1}^N (x_i/N), \tag{3}$$

where N is the number of foreground pixels. Similarly, y_{Cv} represents the vertical centroid component as shown in Eq. (4):

$$y_{Cv} = \sum_{i=1}^N (y_i/N) \tag{4}$$

The parameters are illustrated diagrammatically in Fig. 5.

These spatial features are chosen because we use the frontal URFD sequences where the person in the frame is facing the camera. The frontal sequences are represented by the *cam0* feed of the URFD dataset containing both fall and ADL sequences. Therefore, a person

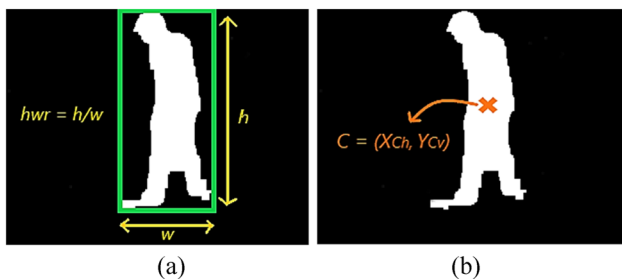


Fig. 5 **a** Height-width ratio (hwr) of the foreground object and **b** Centroid (C) of the foreground object (horizontal centroid component (x_{Ch}) and vertical centroid component (y_{Cv}))

falling parallelly to the camera optical axis experiences a significant change in height-to-width ratio and vertical centroid movement. Whereas, if he falls perpendicularly to the camera optical axis, the centroid in the horizontal direction moves significantly. Hence, the displacement of these parameters in the horizontal or vertical direction or both is observed to be significantly higher during a fall while compared to any other regular life activity. The absolute difference in the displacement of these parameters is measured. For example, the absolute difference in the displacement of height-to-width ratio can be represented as $hwr_{frame(n)} - hwr_{frame(n-1)}$ where, $hwr_{frame(n)}$ and $hwr_{frame(n-1)}$ represents the height-to-width ratio of the foreground object in the current frame and previous frame respectively. Figure 6 shows the absolute difference in the displacement of the height-to-width ratio and the person’s centroid in both horizontal and vertical directions during a fall and daily activity sequences.

As multiple features are analyzed, the variation in these features’ displacement differs between a steady phase and an activity phase. A steady phase represents body activities like walking, standing, lying, or any steady or inactive/stationary posture. Thus, separate feature thresholds are required to segregate the activity phase and the stationary phase. The thresholds are set depending on the displacement variance of these features. The variance in the displacement of the height-width ratio is calculated as follows.

$$\mu_{hwr}(t) = (1 - \alpha)\mu_{hwr}(t - 1) + \alpha HWR(t) \tag{5}$$

$$\sigma_{hwr}(t) = HWR(t) - \mu_{hwr}(t - 1) \tag{6}$$

In Eq. (5), $HWR(t)$ and $\mu_{hwr}(t)$ denotes the displacement in height-width ratio and its mean value at time t , respectively. While, $\mu_{hwr}(t - 1)$ denotes the mean value at a time ($t - 1$). Value α represents the updated parameter. $\sigma_{hwr}(t)$ represents the variance at time t as shown in Eq. (6). Similarly, displacement variance of centroid in the horizontal and vertical direction is calculated.

$$\mu_{(Ch\ or\ Cv)}(t) = (1 - \alpha)\mu_{(Ch\ or\ Cv)}(t - 1) + \alpha(CH\ or\ CV)(t) \tag{7}$$

$$\sigma_{(Ch\ or\ Cv)}(t) = (CH\ or\ CV)(t) - \mu_{(Ch\ or\ Cv)}(t - 1) \tag{8}$$

In Eq. (7), $(CH\ or\ CV)(t)$ and $\mu_{(Ch\ or\ Cv)}(t)$ denotes the centroid displacement in the horizontal or vertical direction and its mean value at time t , respectively. While, $\mu_{(Ch\ or\ Cv)}(t - 1)$ denotes the mean value at a time ($t - 1$). Value α represents the updated parameter. $\sigma_{(Ch\ or\ Cv)}(t)$ represents the variance at time t as shown in Eq. (8).

The variance remains low when there is a minor change in the person’s shape, such as in a steady or inactive phase.

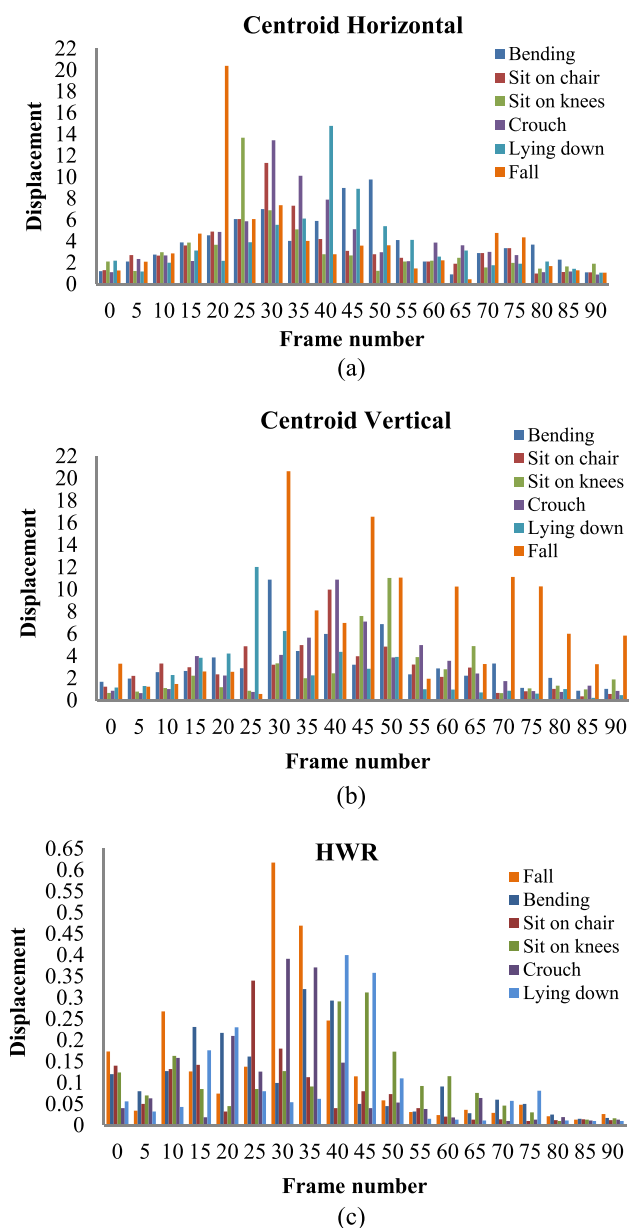


Fig. 6 Absolute difference in the displacement of **a** horizontal centroid movement, **b** vertical centroid movement and **c** height-to-width ratio (HWR) for fall and ADL sequences

It is higher for a change in shape during a falling movement or a daily life activity such as sitting, bending, crouching, or any other fall-like activity. When the variance exceeds a threshold, it can detect a posture change from a stable phase to an activity phase which is then subjected to an SVM and K-NN classification. Figure 7 illustrates the keyframe selection process.

Here, the threshold for the displacement in height-width ratio T_{hwr} , horizontal centroid displacement T_{Ch} , and vertical centroid displacement T_{Cv} are set to 0.18, 2.5, and 2.9, respectively. These thresholds are selected by evaluating the

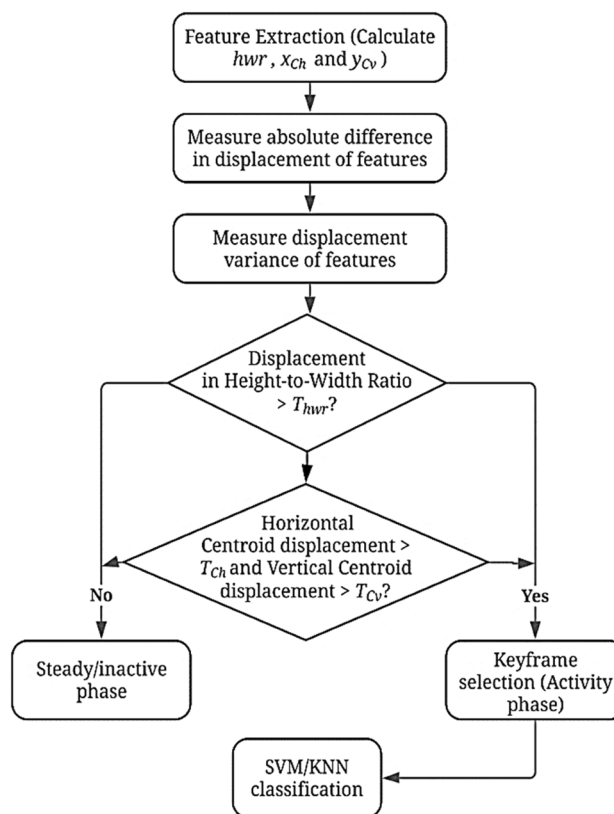


Fig. 7 Flowchart illustrating the keyframe selection process

training sequences based on the displacement of the corresponding spatial features. Accordingly, we can detect a fall or daily activity in the middle of a video sequence based on these threshold values. However, if it is set too high, fall and fall-like activities may not get detected, and when it is set too low, the false alarm rate goes up. Hence, we consider the frames covering the activity phase as the keyframes.

3.6 Keyframe-based classification

After the keyframes are extracted depending on the displacement of the shape-based features, they are classified for activity recognition under two classes, namely, fall and daily living activity.

Although multiple classification models are available, we choose to use two classic machine learning models, a linear SVM, and K-NN, to classify the keyframes. As the number of keyframes extracted is limited from the training video sequences, SVM is a better choice because linear SVM can be modeled with limited training samples, and its competency is higher for a two-class linear classification scenario (Cortes and Vapnik 1995; Boateng et al. 2020) Also, the number of independent variables or features in our approach is limited as most significant foreground features are chosen. Hence, we use the K-NN classification to

enhance the system's accuracy for the limited number of features compared to the training data Boateng et al. (2020). The two-channel classification using both SVM and K-NN can result in a classification disparity for a particular keyframe, like delivering different outputs. Hence, to resolve this issue, we use additional information to present the final decision. Here, we choose the elliptical orientation of the foreground object for this purpose. Because we use the frontal URFD sequences to design the classification model, displacement in the person's orientation concerning the floor is selected as a significant foreground feature of the moving object. It is the threshold to distinguish between a fall and an ADL. We can observe from the training samples that the orientation displacement tends to be large enough during a fall sequence when compared to any other activity of daily living. The displacement happens in the horizontal or vertical direction or both. After that, we can combine both SVM and K-NN results based on the decision obtained from the elliptical orientation displacement for a keyframe. As a result, we can resolve any classification disparity between the two channels. Figure 8 shows the absolute difference in the displacement of the elliptical orientation during fall and ADL sequences.

4 Experiment and results

This section evaluates the efficacy of the proposed fall detection system. Experiments are carried out upon the fall and ADL video sequences of the UR Fall Detection (URFD) dataset Kepski and Kwolek (2014) to evaluate the performance of the approach. It is a publicly available dataset. The dataset consists of thirty video sequences with falls, thirty video sequences with typical daily life activities such as sitting down, bending to pick up something from the floor, crouching down to look for something beneath, and ten video sequences with fall resembling daily activity as lying on the floor and lying on the couch or bed. Two types of falls were simulated by

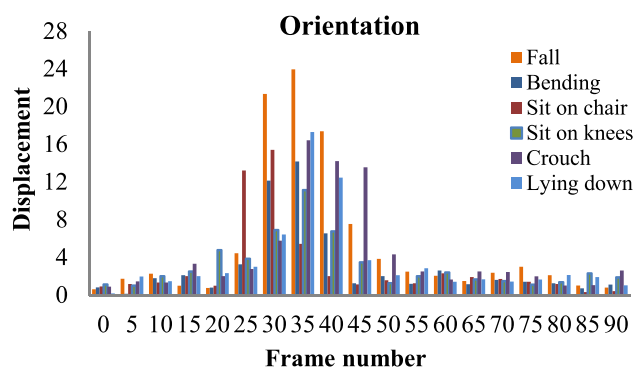


Fig. 8 Absolute difference in the displacement of elliptical orientation for fall and ADL sequences

five volunteers: from standing posture and sitting posture. The video sequences are represented using depth maps and corresponding RGB images acquired by a static Kinect sensor. The Kinect was placed at the height of 1 m from the floor which captured the frontal URFD sequences. The frontal sequences referring to the *Cam0* feed is evaluated because it represents both fall and daily activity video sequences. All the fall and activity sequences were recorded at a 30 Hz frame rate with a frame resolution of 640×240 pixels. We considered the RGB image frames as it is a vision-based technique and, simultaneously any low-cost RGB surveillance camera can be used for the video acquisition of the proposed approach.

Experiments are conducted using MATLAB on a system having a configuration of Intel(R) Core (TM) i5-1135G7 @ 2.42 GHz processor with 8 GB of RAM and 256 GB SSD. The classification result of SVM and K-NN are combined with the help of the orientation decision based on which the final output is recognized to be a success or failure. Table 1 shows the recognition result for a few sample keyframes. It depicts the two-channel combined result of SVM and K-NN classification based on the orientation output. Here, the first five keyframes represent fall sequences and the remaining represent ADL sequences. Figure 9 shows the example keyframes (1, 6–10) presented in Table 1 with their corresponding motion RoI or blob. An example of a stable movement frame is shown in Fig. 10. Here, the person walks and hence displays no huge change in body shape. HWR displacement is ($hwr=0.16$), Centroid horizontal and vertical displacement are ($Ch=2.09$ and $Cv=1.23$). The displacement of these parameters is below the threshold to distinguish between a steady and an activity phase. Hence, this frame is not considered as a keyframe and the algorithm stops at this step and fetches the next frame for processing.

4.1 Performance evaluation

We use a few widely used performance metrics in fall detection methods to evaluate the performance of the proposed approach, as shown in Eqs. (9)–(12).

$$\text{Sensitivity/recall}(\%) = \frac{TP}{TP + FN} \quad (9)$$

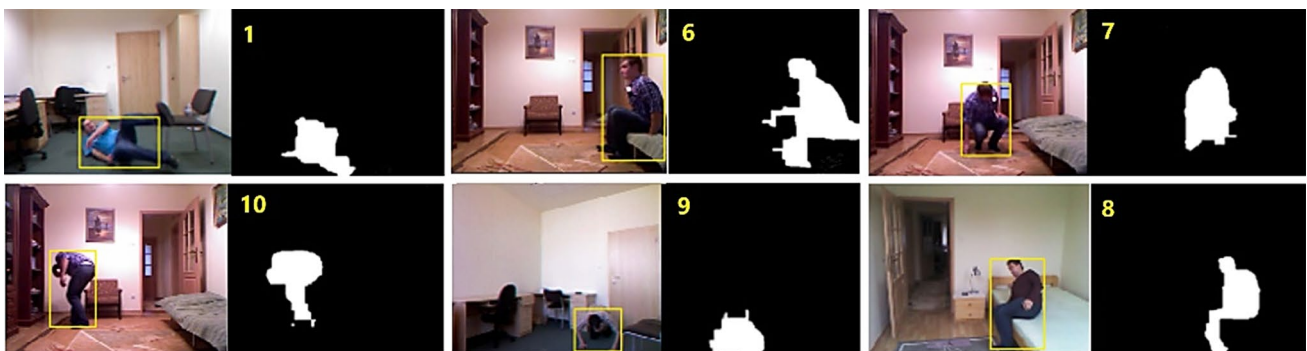
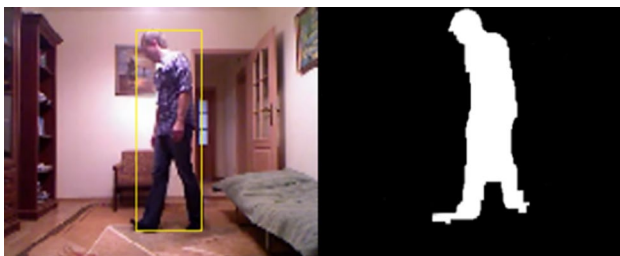
$$\text{Specificity}(\%) = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Precision}(\%) = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Accuracy}(\%) = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Table 1 Recognition result of sample keyframes

Key frame	Event	<i>hwr</i>	<i>Ch</i>	<i>Cv</i>	SVM decision	K-NN decision	Orientation decision	Combined result	Recognition
1	Fall	0.47	19.36	20.12	Fall	Fall	Fall	Fall	Success
2	Fall	0.54	17.21	18.22	Fall	Fall	Fall	Fall	Success
3	Fall	0.39	13.62	17.4	ADL	ADL	Fall	Fall	Success
4	Fall	0.31	12.7	11.7	ADL	ADL	ADL	ADL	Failure
5	Fall	0.4	15.4	18.1	ADL	Fall	Fall	Fall	Success
6	ADL	0.34	11.31	9.98	ADL	ADL	ADL	ADL	Success
7	ADL	0.29	10.67	11.03	ADL	ADL	ADL	ADL	Success
8	ADL	0.38	14.78	13.9	Fall	ADL	Fall	Fall	Failure
9	ADL	0.37	13.43	12.89	ADL	ADL	ADL	ADL	Success
10	ADL	0.27	8.98	9.87	ADL	ADL	ADL	ADL	Success

**Fig. 9** Sample keyframes (1, 6–10) presented in Table 1 with their corresponding motion RoI (clockwise-direction)**Fig. 10** Example of stable movement frame ($hwr=0.16$, $Ch=2.09$ and $Cv=1.23$)

True positive (TP) refers to an object experiencing a fall activity, and the system detects it correctly. However, a system that fails to catch a falling entity is represented by the parameter false negative (FN). True negative (TN) refers to an entity carrying out a regular daily activity, and the system can detect it accurately. An event in which an object performs a daily action, but the system determines as a falling movement is represented by false positive (FP). These four parameters TP, FN, TN, and FP determine the performance metrics such as sensitivity/recall, specificity, precision, and accuracy.

Sensitivity/Recall evaluates the capacity of the technique to detect fall activity. Specificity evaluates the ability of the method to see daily life activities. Precision represents a positive predictive value and accuracy represents the overall recognition rate of the system.

A two-channel classification model is designed for the proposed fall detection approach using SVM and K-NN. Both depend upon the selected spatial features based on which the training for each keyframe is carried out. The UR fall detection dataset consists of 70 video sequences (30 falls + 40 daily activities), of which 50 video sequences comprising of 20 falls (i.e., 66.66% of the fall sequences) and 30 daily activities (i.e., 75% of the daily activity sequences) are used as the training set. Both the models are evaluated based on the keyframes that are subject to classification. Table 2 reports the cumulative confusion matrix representing the quantitative performance of the proposed fall detection method based on video sequence classification of fall and ADL sequences.

We evaluate the performance of the system in terms of sensitivity, specificity, precision, and accuracy. The classification models considered are SVM, K-NN, orientation threshold-based, and SVM-K-NN combination based on

Table 2 Cumulative confusion matrix

Classification model	Event	Fall detected	ADL detected	Recall (%)	Specificity (%)	Precision (%)	Accuracy (%)
SVM	Fall	28	02	93.33	92.5	90.32	92.85
	ADL	03	37				
K-NN	Fall	29	01	96.66	95	93.54	95.71
	ADL	02	38				
Orientation threshold	Fall	29	01	96.66	97.5	96.66	97.14
	ADL	01	39				
Combined approach	Fall	30	00	100	97.5	96.77	98.6
	ADL	01	39				

Table 3 Performance comparison of our approach with existing techniques based on the frontal URFD sequences

Approach	Features used	Classification	Input signal	Recall (%)	Sp (%)	Prec (%)	Acc (%)
Kepski and Kwolek (2015)	HWR, Blob-height to physical height ratio, Centroid distance to the floor	SVM	Depth,	100	90	88.24	94.28
		K-NN	accelerometer data	100	92.5	90.90	95.71
Merrouche and Baha (2017)	Height, Centroid tracking	Feature threshold	Depth	76.92	80	–	78.68
Wang et al. (2019)	Background subtracted frames	CNN	RGB	94.86	–	99.64	–
Proposed method	HWR displacement, Horizontal centroid displacement, Vertical centroid displacement, Orientation displacement	SVM	RGB	93.33	92.5	90.32	92.85
		K-NN		96.66	95	93.54	95.71
		Orientation threshold		96.66	97.5	96.66	97.14
		Combined approach		100	97.5	96.77	98.6

Rec: Recall/Sensitivity, Sp: Specificity, Prec: Precision, Acc: Accuracy

orientation. Table 3 shows the comparison between the existing and the proposed approach. It is done considering the classification approaches and the features used for classification. All the methods are compared based on their performance on the frontal URFD sequences. Kepski and Kwolek (2015) use both SVM and K-NN classifiers to evaluate the system's performance based on spatial features such as blob height-to-width ratio, blob-height-to-physical-height ratio, and the centroid distance to the floor. They achieved 100% recall in detecting falls. However, the proposed fall detection model too achieves a 100% sensitivity in detecting fall movements and simultaneously outperforms Kepski and Kwolek (2015) in terms of specificity, precision, and accuracy. In Merrouche and Baha (2017), the features evaluated are the threshold to distinguish between a fall and daily life activity. Here, spatial features like height and centroid tracking are used to assess the fall detection system. The proposed approach clearly outperforms Merrouche and Baha (2017) concerning all the evaluated performance measures. Wang et al. (2019) use a CNN-based deep learning approach to classify the events. Background subtracted frames are input to the CNN network. The proposed method outperforms the CNN-based approach concerning the fall detection capacity with a recall of 100%. Although CNN is a state-of-the-art

classification model, CNN learns features effectively and delivers optimum performance when the dataset used for training the network is large enough with more labeled samples. The URFD dataset used in this method consists of limited training samples. However, Wang et al. (2019) display a better precision of 99.64% than the proposed fall detection system. Although, a higher precision deals with the significant cost of false-positive affecting the performance of the system, a fall detection system rather benefit from a higher recall. When the performance cost associated with false negatives is high, a better recall should be achieved. Such as, a fallen person (true positive) is predicted as an ADL (false negative), the consequences will be severe and result in a significant loss in performance. The comparison is made using performance measures, namely sensitivity, specificity, precision, and accuracy. It is to be noted, '–' symbol indicates the data is not available. Based on these performance measures, the proposed method using the two-channel combined approach outperforms the existing techniques.

Overall, using all the classification strategies, the proposed fall detection model delivers a very satisfactory performance. It achieves an accuracy of 92.85% and 95.71% using SVM and K-NN, respectively. 97.14% accuracy is achieved based on the orientation threshold. Although

Table 4 Processing time observation (System configuration: Intel(R) Core (TM) i5-1135G7 @ 2.42 GHz, 8 GB RAM, and 256 GB SSD)

Keyframe (640×240)	Processing time (ms)	Frame rate (FPS)
1	31.14	32.11
2	32.25	31.00
3	32.43	30.83
4	30.82	32.44
5	31.12	32.13
6	32.87	30.42
7	31.98	31.26
8	33.74	29.90
9	31.90	31.34
10	32.48	30.78
Avg. processing time (100 keyframes)	31.16 Frames/second	

combining both SVM-KNN classification output based on the elliptical orientation output, the proposed method achieves an enhanced accuracy of 98.6%, which is significantly better than the classification performance of the SVM and K-NN model individually. The proposed algorithm takes RGB images as input. RGB images increase the scope for feature extraction in an economical way relating to the camera used. It can be captured using any simple RGB surveillance camera, which can eventually reduce the implementation cost of the fall detection system.

The proposed method reports the total processing time of a keyframe in a few milliseconds. Table 4 reports the running time of the example keyframes presented in Table 1 in milliseconds and corresponding frame rate (FPS). The resolution of each keyframe is 640×240 pixels.

5 Conclusion and future directions

In this paper, we design a human fall detection system by analyzing significant spatial features of the moving object. Keyframes represented by these features are extracted. Keyframes help detect activity in the middle of a video sequence by filtering stationary or steady movement frames, leading to reduced processing time. These frames are then classified using two-channel machine learning techniques: SVM and K-NN. The orientation threshold is used to remove any classification disparity between the two channels in allocating a label to a keyframe as a fall or daily activity. Finally, the two-channel classification output is combined based on the orientation decision to label the body as a potential fall or ADL. The proposed approach achieves 92.85% accuracy using SVM classification, 95.71% accuracy using K-NN classification, 97.14% accuracy using orientation threshold, and a combined accuracy of 98.6%. The proposed fall

detection system results in robust performance using limited training samples compared to existing state-of-the-art techniques. The analysis of the proposed approach is based on artificial lighting conditions, and its evaluation under natural light needs to be done in the future. The public dataset used in this method contains one actor per video sequence. Hence, executing multiple persons' fall detection in an image frame can also be considered as future work. Also, we expect applying modern deep learning techniques can improve our proposed method in the future.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The used dataset is the publicly available URFD dataset (<http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html>).

References

- Boateng EY, Otoo J, Abaye DA (2020) Basic tenets of classification algorithms *K*-nearest-neighbor, support vector machine, random forest and neural network: a review. *J Data Anal Inf Process* 8:341–357
- Chelli A, Patzold M (2019) A machine learning approach for fall detection and daily living activity recognition. *IEEE Access* 7:38670–38687. <https://doi.org/10.1109/ACCESS.2019.2906693>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Directalert (2021) Wireless emergency response system [online]. <http://www.directalert.ca/emergency/help-button.php>. Accessed 5 June 2021
- Geertsema EE, Visser GH, Viergever MA, Kalitzin SN (2019) Automated remote fall detection using impact features from video and audio. *J Biomech* 88:25–32
- Han Q, Zhao H, Min W et al (2020) A two-stream approach to fall detection with MOBILEVGG. *IEEE Access* 8:17556–17566
- Htun SN, Zin TT, Tin P (2020) Image processing technique and hidden markov model for an elderly care monitoring system. *J Imaging* 6:49
- Huang Y, Chen W, Chen H, Wang L, Wu K (2019) G-Fall: device-free and training-free fall detection with geophones. In: 2019 16th annual IEEE international conference on sensing, communication, and networking (SECON), Boston, MA, USA, pp 1–9
- Hussain F, Hussain F, Ehatisham-ul-Haq M, Azam MA (2019) Activity-aware fall detection and recognition based on wearable sensors. *IEEE Sens J* 19(12):4528–4536
- Jamil N, Sembok TMT, Bakar ZA (2008) Noise removal and enhancement of binary images using morphological operations. In: 2008 international symposium on information technology, Kuala Lumpur, pp 1–6
- Junior CLB, Adami AG (2018) SDQI—fall detection system for elderly. *IEEE Latin Am Trans* 16(4):1084–1090
- Kalinga T, Sirithunge C, Buddhika A, Jayasekara P, Perera I (2020) A fall detection and emergency notification system for elderly. In: Proceedings of the 2020 6th international conference on control, automation and robotics (ICCAR), Singapore, pp 706–712

- Kepski M, Kwolek B (2015) Embedded system for fall detection using body-worn accelerometer and depth sensor. In: Proceedings of IEEE 8th international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS), pp 755–759
- Kepski M, Kwolek B (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. *Comput Methods Programs Biomed* 117(3):489–501
- Kerdjadj O, Ramzan N, Ghanem K, Amira A, Chouireb F (2020) Fall detection and human activity classification using wearable sensors and compressed sensing. *J Ambient Intell Humaniz Comput* 2020:1–13
- Krumholz A, Hopp J (2008) Falls give another reason for taking seizures to heart. *Neurology* 70:1874–1875
- Li X, Nie L, Xu H, Wang X (2018) Collaborative fall detection using smartphone and kinect. *Mobile Netw Appl* 23:775–788. <https://doi.org/10.1007/s11036-018-0998-y>
- Luo Y, Zhou H, Tan Q et al (2018) Key frame extraction of surveillance video based on moving object detection and image similarity. *Pattern Recognit Image Anal* 28:225–231. <https://doi.org/10.1134/S1054661818020190>
- Makandar A, Mulimani D, Jeevor M (2015) Preprocessing step – review of key frame extraction techniques for object detection in video. *Int J Curr Eng Technol* 5(3):2036–2039
- Makhlouf A, Boudouane I, Saadia N, Amar RC (2018) Ambient assistance service for fall and heart problem detection. *J Ambient Intell Humaniz Comput* 10(4):1527–1546
- Merrouche F, Baha N (2017) Fall detection using head tracking and centroid movement based on a depth camera. In: Proceedings of ACM international conference on computing for engineering and sciences (ICCES), pp 29–34
- Paul M, Haque SME, Chakraborty S (2013) Human detection in surveillance videos and its applications—a review. *EURASIP J Adv Signal Process* 2013(1):1–16
- Peng Y, Peng J, Li J et al (2019) Design and development of the fall detection system based on point cloud. *Procedia Comput Sci* 147:271–275
- Ramachandran A, Karuppiyah A (2020) A survey on recent advances in wearable fall detection systems. *Biomed Res Int* 2020:1–17. <https://doi.org/10.1155/2020/2167160>
- Rubenstein LZ (2006) Falls in older people: epidemiology, risk factors and strategies for prevention. *Age Ageing* 35:ii37–ii41
- Russell-Jones DL, Shorvon SD (1989) The frequency and consequences of head injury in epileptic seizures. *J Neurol Neurosurg Psychiatry* 52:659–662
- Sabatini AM, Ligorio G, Mannini A, Genovese V, Pinna L (2016) Prior-to- and post-impact fall detection using inertial and barometric altimeter measurements. *IEEE Trans Neural Syst Rehabil Eng* 24(7):774–783
- Sadrezami H, Bolic M, Rajan S (2019) TL-FALL: contactless indoor fall detection using transfer learning from a pretrained mode. In: Proceedings of the 2019 IEEE international symposium on medical measurements and applications (MeMeA), Istanbul, Turkey, pp 1–5
- Shu F, Shu J (2021) An eight-camera fall detection system using human fall pattern recognition via machine learning by a low-cost android box. *Sci Rep* 11:2471. <https://doi.org/10.1038/s41598-021-81115-9>
- Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: Proceedings of the 1999 IEEE Computer Society conference on computer vision and pattern recognition (Cat. No PR00149), pp 246–252
- Sulman N, Sanocki T, Goldgof D, Kasturi R (2008) How effective is human video surveillance performance? In: 19th international conference on pattern recognition (ICPR 2008), pp 1–3
- United Nations DoEaSA, Population Division (2017) World Population Ageing 2017—Highlights (ST/ESA/SER.A/397)
- Vollset SE, Goren E, Yuan C-W, Cao J et al (2020) Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30677-2](https://doi.org/10.1016/S0140-6736(20)30677-2)
- Wang H, Zhang D, Wang Y, Ma J, Wang Y, Li S (2017) RT-fall: a real-time and contactless fall detection system with commodity wifi devices. *IEEE Trans Mob Comput* 16(2):511–526
- Wang X, Ellul J, Azzopardi G (2020) Elderly fall detection systems: a literature survey. *Front Robot AI*. <https://doi.org/10.3389/frobt.2020.00071>
- Wang H, Gao Z, Lin W (2019) A fall detection system based on convolutional neural networks. In: Proceedings of ACM the international conference on robotics, intelligent control and artificial intelligence (RICA I 2019), Shanghai, China, pp 242–246. <https://doi.org/10.1145/3366194.3366236>
- WHO (2018) Falls. <https://www.who.int/news-room/fact-sheets/detail/falls>. Accessed 16 Jan 2018
- Xi X, Jiang W, Lü Z, Miran SM, Luo Z-Z (2020) Daily activity monitoring and fall detection based on surface electromyography and plantar pressure. *Complexity*. <https://doi.org/10.1155/2020/9532067>
- Zitouni M, Pan Q, Brulin D, Campo E (2019) Design of a smart sole with advanced fall detection algorithm. *J Sens Technol* 9:71–90

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.