Fall detection method based on Spatio-temporal feature fusion using combined two-channel classification

Anurag De[1] · Ashim Saha[1] · Praveen Kumar[2] · Gautam Pal[3]

Abstract
Nowadays, the growing population of senior citizens is a challenge for almost all developing countries. New technologies can help monitor elderlies at home by providing an innovative and secure environment and further enhancing their quality of living. Vision-based systems offer promising results in analyzing human posture and detecting abnormal events like falls. Falls appear to possess the most considerable risk for seniors living alone. In this article, a new fall detection method is proposed based on a fusion of motion-based and human shape-based features. Motion History Images (MHI) represent the temporal feature in our approach. Simultaneously, the height-to-width ratio and centroid of the moving person represent the spatial features. A two-channel classification model is designed using a threshold-based and a keyframe-based approach. The two channels are further combined based on any classification disparity for which more information is used to classify between falls and daily activities. Keyframes are selected based on the displacement of the spatial features having a threshold higher than a preset value. Keyframes are subject to a K-NN classification. The proposed algorithm delivers promising results on the UR fall detection dataset's simulated fall and daily activity sequences. It provides satisfactory performance compared to existing state-of-the-art methods and shows a peak accuracy of 98.6% and recall of 100% in detecting falls. Specificity and precision are over 96%.

✉ 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

1 Introduction

Fall incidents are significant danger events for frail individuals and senior citizens living alone. According to the United Nations report [40], the world population is overgrowing. Because of advanced medical facilities, older people of age 60 and more are also increasing day by day. It is estimated that the population of elderly people will cross 1000 million by 2030 and 2000 million by 2050. Bill and Melinda Gates Foundation sponsored one study related to the population of senior citizens [41], where it is estimated that people above 80 years of age group will cross 866 million at the end of this century from 141 million in the year 2017. As most of the senior people of age, more than 60 prefer to stay inside the home and due to business of other family members often those older people remain alone within the home. Under this circumstance, fall incidents can bring a genuine injury and could be fatal [37] if not dealt with quickly. Consequently, researchers proposed many fall detection techniques in recent days [10, 30, 44]. These proposed methodologies can detect a fall incident and communicate with the intended person by sending messages so that necessary intervention can be taken to protect the older people's lives. Surveillance system technology has become more advance nowadays for detecting various abnormal incidents. The information saved in the surveillance system is accessed and processed on a needed basis. Processing the stored data in a surveillance system by the human being in real-time consumes more time [38]. So, finding unusual human activities like falls automatically in the surveillance video is the solution to this problem [11, 17, 35, 47]. In this paper, an automatic detection technique of human falls for older people is proposed.

Broadly, fall detection techniques are divided into two categories. In the first category, electronic devices automatically detect probable falling incidents from normal daily living activities. Again, this approach has two types, wearable device-based and non-wearable device-based systems. Wearable devices [15, 25, 29, 33, 46] use electronic components such as accelerometer, gyroscope, magnetometer, body-worn barometric altimeter, surface electro-myograph, etc., for collecting information that helps to detect fall incidents.

The main limitation of this kind of approach is that the electronic components always need to be worn, though this approach is cost-effective. Senior people may face problems with these wearable devices, and they often may not remember to wear those electronic gadgets. Non-wearable systems are mostly fitted inside the house. This approach uses sensors to gather data about the value of floor vibration, the pressure exerted on the floor, map amplitude of wireless signals to human motion, etc. to detect falling behavior [7, 14, 34] effectively. These systems also have some limitations. A small amount of fluctuation in the indoor environment can lead to a pressure difference or generate noise, though humans may not do this.

The second category became very popular, nowadays which is based on a vision system. This approach does not need to wear any electronic equipment always or use a help button [6] to raise the alarm. Here only video surveillance cameras are used to detect fall incidents in real-time. It generally requires a conventional wall-mounted power supply and a backup battery for round-the-clockreal-time surveillance. Apart from surveillance, a vision-based system can also provide a wide range of information about the person's behavior, location, sleep tracking, meal tracking, medication tracking, etc. Figure 1 shows a schematic diagram representing the classification of fall detection approaches.

The proposed system is designed to know that the motion and posture of a person change significantly during a fall compared to other regular activities like walking, sitting down,
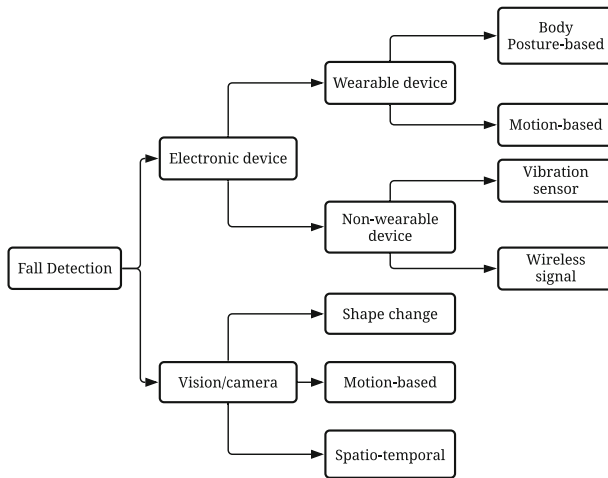
Fig. 1 Classification of fall detection approaches

bending to pick up something, or lying down. A fall can occur accidentally or due to weakness [31, 45], epileptic seizures [22, 31, 32, 45], etc. The individual's motion and body shape variation help to distinguish a fall event from a regular living activity. The primary contributions of the proposed methodology are summarized as follows:

– This paper achieves novelty by presenting a feature fusion of body motion and significant changes in human shape to detect falls. The combination of the temporal and spatial features helps to analyze and provide crucial information on human activities.
– A combination of threshold-based and machine learning-based classification strategies is used to evaluate the performance of the fall detection model to make the system more robust. The proposed approach has proven its robustness on real-time video sequences of simulated falls and Activities of Daily Living (ADL).
– Instead of selecting activity frames randomly for classification, keyframes representing fall and fall-like daily activities are chosen. Keyframes help to identify an activity posture from a stationary or inactive posture. It further leads to restricting a fall or fall-like daily activity in the middle of a video stream. Keyframe selection also improves the overall time complexity of the fall detection algorithm by filtering redundant frames for training the system [23, 24].
– A budget-friendly system is designed using RGB frames as input. A single low-cost, conventional, wide-angle RGB surveillance camera will be sufficient to carry out real-time indoor surveillance.

The rest of the paper is organized as follows. Section 2 presents a detailed literature survey of existing work related to human fall and daily life activity detection. Section 3 presents the system overview. Section 4 illustrates the proposed fall detection methodology. Section 5 discusses the experimental results along with performance evaluation followed by a comparison with existing work. Section 6 concludes the paper and mentions the scope for future enhancements.

2 Related work

This section highlights the literature review of existing fall detection approaches using wearable device-based, non-wearable device-based, and vision-based techniques.

Zitouni et al. [48] proposed an intelligent sole embedded with a fall detection technique based on a single tri-axis accelerometer. The proposed method chooses the acceleration, position, and duration parameters-based thresholds to find the fall incidents. In this approach, the intelligent sole's footwear always needs to be worn by older adults is the main drawback. Chelli and Patzold [4] presented a machine learning-based fall and daily activity detection technique. An accelerometer and gyroscope were used to extract the acceleration and angular velocity data subject to classification. Machine learning algorithms such as KNN, ANN, QSVM, and EBT were implemented to reach an accuracy of 85.8%, 91.8%, 96.1%, and 97.7%, respectively. Xi et al. [46] designed a fall detection and daily activity monitoring system based on surface electromyography (sEMG) and plantar pressure signals. The system achieved above 96% accuracy, sensitivity, and specificity for different posture transition, gait, and fall events. Reducing the number of sensors and attaining a high recognition rate is considered as future work. Kerdjidj et al. [20] proposed a similar approach where fall and daily activity is automatically detected based on wearable fall detection type. This algorithm used a lightweight and easy-to-wear system. An accelerometer, magnetometer, gyroscope, and electrocardiogram (ECG) helped generate a large amount of data. However, repeated battery charging is required to keep it in operating mode.

The fall detection literature stated above is based on sensors and wearable devices attached to the body. These algorithms are better but have some limitations during practical implementation. Some limitations are frequent recharge, extra load to wear devices always mainly for elderly people, etc. So, to overcome these limitations, non-wearable and surveillance-based devices are used. Few existing works based on non-wearable device-based techniques are stated below.

Tian et al. [39] introduced a fall detection system named Aryokee, which is multi-function in nature. It can detect falls, stand-up events, and fall duration based on Radio Frequency (RF) signals. It uses an FMCW radio fitted with two antenna arrays for separating the reflections from multiple objects in the surroundings. More than 140 volunteers performed 40 types of actions in various conditions to collect data for the evaluation of the system. A convolutional neural network was used for classification, which resulted in a recall of 94% and a precision of 92%. Wang et al. [42] presented a real-time, contactless, and low-cost indoor fall detection approach based on the phase and amplitude of the fine-grained channel state information (CSI) found in stock Wi-Fi devices. Fall and similar incidents are observed using the CSI phase difference. In addition, the sharp power profile decline pattern in the time-frequency domain is used for better fall detection. This technique performs better with indoor light intensity changes but poorly while moving into a new environment with an old setting. In [12], the authors proposed a similar technique based on wireless signals. They proposed a fall detection method based on channel state information for a 5G environment. It maps the amplitude data in the wireless signal to detect a fall event. The method utilizes the 5 GHz signal for better subcarrier frequency domain information. It improves the relationship between human motion and wireless signals for the effective detection of falls and normal activities. The system achieves a peak accuracy of 92.3%. However, its performance is highly susceptible to transmitter-receiver distance and multipath interference. Huang et al. [14] presented a fall detection method based on geophones that receive the floor vibration signals. It extracts time-dependent features

by analyzing the vibration signals. These features help the system recognize a potential fall incident based on the Hidden Markov Model (HMM). This approach works better for older people staying alone at home. But this strategy cannot detect the fall of multiple people effectively due to the lack of processing of multiple floor vibration signals.

Non-wearable systems are better in comparison to wearable-based methods, yet it has few drawbacks. The environmental noise that can change the signals used by the non-wearable devices is the major drawback of this kind of approach. The use of vision or video surveillance-based systems in real-life fall incident detection solutions to non-wearable system limitations.

Peng et al. [28] mentioned a vision-based fall detection method based on a human point cloud. Depth data is given to the system using Kinect as input. In addition, this technique maps the depth information into a point cloud image where a human is presented using a color spectrum. Fall behavior is estimated based on the height change acceleration feature. This technique can detect potential falls and activities like sitting, squatting, walking, and bending. It shows poor performance in differentiating between fall incidents and controlled lying, such as sleeping or lying on the floor or on any surface. Gracewell and Pavalarajan in [8] designed a fall detection model based on a two-stream spatial and temporal classification. The authors select keyframes that are subject to the classification streams. Keyframes are extracted by analyzing the displacement in the centroid of the moving object concerning a threshold value. Optical flow vectors for the selected keyframes are extracted, which are subject to temporal classification. The authors evaluate the performance of the model using the publicly available UR Fall Detection dataset. It shows an accuracy of 88.57% using the spatial features as input and 97.14% by combining the features. However, the authors have not mentioned the performance measures based on the temporal classification. In [19], the authors presented a low-cost fall detection technique. Motion data is used from an accelerometer and depth images are captured using Kinect sensors for the implementation. Spatial features are obtained from the depth images. These features are analyzed only when a person's movement is above a predefined threshold. This algorithm helps in minimizing the computation cost. It also minimizes the false alarms by combining both motion and spatial features of the depth images. The approach delivers 95.71% accuracy and 100% sensitivity for detecting fall incidents. However, electronic sensors that need to wear by people are the main drawback of this system.

In [26], the authors suggested a vision-based fall detection method using a depth camera. It combines human shape, head, and centroid tracking analysis to recognize fall incidents concerning normal activities. The technique needs to upgrade to address more complicated movements such as backward fall and fall while sitting on a chair using an appropriate dataset. Wang et al. [43] designed a vision-based fall detection system using Convolutional Neural Networks (CNN). They train the VGG-16 network to identify a fall movement in a frame by implementing the transfer learning concept. The frames are pre-processed, implementing background subtraction and morphological operations. Although the algorithm displays promising results in normal lighting conditions, it lacks significantly in low-light surroundings. Htun et al. [13] proposed a vision-based monitoring system based on image processing technology. A Hidden Markov Model is used for detecting falls and regular activities. It uses human shape-based features like Silhouette surface area, centroid height, and bounding box aspect ratio to analyze the person in the frame. The system shows a sensitivity of 98.37% using experiment videos containing both normal and non-abnormal events, including falls. There remains ample scope for the system, by incorporating multiple persons fall detection in a frame as the scope of the work is limited to a single person. The following section presents the system overview of the proposed fall detection system.

3 System overview

A new fall detection method is presented here by combining Spatio-temporal features of the input video sequence. Motion History images (MHI) and significant human shape changes are considered temporal and spatial features. The proposed method notes that the motion is large enough during a fall compared to any other regular activity. Hence, it is necessary to detect a significant movement of the person in the frame. This task is the first step of the system and is carried out using Motion History Image.

Once a motion is detected, the next step is to analyze the human shape. During a fall, there is an occurrence of significant shape changes. This change in the person's shape can distinguish the enormous motion detected is from a regular activity like walking, sitting, lying down, etc., or from a fall.

3.1 Video acquisition and frame extraction

It is necessary to acquire a surveillance video sequence in real-time and extract its frames to generate Motion History Images and take out and analyze the human shape. We used the University of Rzeszow Fall Detection Dataset (URFD Dataset) consisting of 30 falls and 40 ADL sequences [18]. It is challenging and hardly possible to acquire real-time fall sequences of elderlies. Hence, in [18], young volunteers' fall and daily living activity video sequences are simulated for processing. Each video sequence has a frame rate of 30 frames/s and a frame resolution of 640 × 240 pixels.

3.2 Motion History Image generation

It is observed that the motion of a person provides vital information during a fall, as no potential fall takes place without significant body movement. Depending on this observation, temporal or motion information is extracted from the video stream. Generally, optical flow outputs [2] are considered to extract motion information from a video stream. But it has some limitations in real-time fall detection systems as it is prone to errors during large movements during a fall. Also, we don't need to predict the direction of the fall. Our objective is to estimate the motion of the person. The motion is large during a fall event. Hence, we consider generating Motion History Image (MHI) [3], as it is one of the simple and efficient ways to represent motions in surveillance videos by creating a motion template. It provides the temporal information of motion in a video in the form of an image. The pixels in this image are brighter where the motion took place recently, and its intensity reduces where the motion took place earlier. The creation of MHI and its different variants is discussed in detail in [1, 3, 21]. As discussed in [1], MHI $H\tau_{(x,y,t)}$ is attained from an update function $\Psi_{(x,y,t)}$, as shown in Eq. (1).

$$H\tau_{(x,y,t)} = \begin{cases} \tau, & if\ \Psi_{(x,y,t)} = 1 \\ \max\left(0, H\tau_{(x,y,t-1)} - \delta\right), & otherwise \end{cases} \quad (1)$$

Here, $\Psi_{(x,y,t)}$ denotes motion or object in the current video frame. Variables (x, y), t and δ represent the pixel location, time, and decay parameter, respectively. As considered in [1], different values of δ provide somewhat additional motion information and must be determined empirically. For our UR Fall detection dataset experimentation, we set the decay parameter

ranging from 25 to 30. The duration τ controls the temporal extent of the movement. The update function $\Psi_{(x,y,t)}$ mentioned above with a threshold ξ, is measured using frame subtraction as shown in Eq. (2).

$$\Psi(x,y,t) = \begin{cases} 1, & if \quad D_{(x,y,t)} \geq \xi \\ 0, & otherwise \end{cases} \tag{2}$$

As expressed in [1, 21], a distance threshold Δ is imposed on the function $D_{(x, y, t)}$ for the frame subtraction process. The function $D_{(x, y, t)}$ carries out the background subtraction. It is represented in Eq. (3).

$$D_{(x,y,t)} = \mid I(x,y,t) - I(x,y,t \pm \Delta) \mid \tag{3}$$

Here, $I(x, y, t)$ denotes the pixel intensity value at respective pixel location (x, y), at given time t. Here, we have set the distance threshold $\Delta = 1$ for the dataset experimented. The MHI generated from Eq. (1) is in the gray level form.

3.3 Human shape extraction

To obtain the person's shape from the extracted frames of the video sequence, first, the foreground of the image frame is segmented using a background subtraction algorithm to detect the moving object in the frame. Next, the motion Region of Interest (RoI) which denotes the moving object, is located and approximated into a connected component criterion called a blob.

3.3.1 Moving object detection through foreground segmentation

The input video is split into a sequence of image frames. These frames are the input for detecting a moving object within the video. The motion behavior of the person is analyzed by finding the moving region in the video frames. An adaptive background mixture model [27, 36] separates the moving object in real-time from the video. Considering various lighting conditions, the algorithm is found to be highly robust. The algorithm updates the background information using an approximation by modeling each pixel as a Gaussian mixture. This update technique results in adapting the system to variations in illumination and objects that stopped moving. It further tracks the evolution of the corresponding pixel's state from one frame to another. The pixels experiencing no state change are assigned with weight-0, which represents black color pixels. These are the background pixels and the ones changing states are assigned weight-1, which represents white color pixels. These are the foreground pixels. The background pixels hardly change their state. Hence, the moving object in the frame is represented by the foreground pixels.

3.3.2 Motion RoI approximation through morphological noise reduction

The foreground image contains the motion Region of Interest (RoI) which denotes the location of the moving object from the rest of the image frame. It is represented based on binary statistical morphological operations, namely erosion and dilatation [16]. The noise from the image is removed by eliminating the isolated noisy pixels using binary statistical erosion.

Dilatation helps in recovering the loss caused by erosion by filling holes. It leads to retrieving the essential pixels done away with during the process. It further unites the areas split during the binarization of the image frame. As a result, the obtained RoI is combined into a moving object region represented by a connected component criterion, generally called a blob. It clusters different moving regions considered part of a single moving object. It also connects other moving areas into a single moving object. Hence, approximation of the motion Region of Interest (RoI) plays a major role in detecting persons moving in a frame and also at times of occlusion of the target object.

3.4 Training fall and daily activity sequence

The fall detection model based on the fusion of Spatio-temporal components is designed by training the activity samples of these two data classes, namely, fall and daily activities. As mentioned, it is observed that a fall and regular activity movement possess different motions and body shape changes. In our approach, we considered the UR Fall Detection dataset (http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html) [18] for training purposes. It contains 70 video sequences of 30 are fall samples and 40 samples of Activities of daily living (ADL). We have designed the training model evaluating the frontal camera sequences of the UR Fall Detection dataset. Training sample frames of falling action (**a–f**) and training sample frames of daily activities (**g–l**) are shown in Fig. 2. Samples for both fall and daily activities are taken from selected dataset sequences. These samples are trained to generate a Spatio-temporal model using a threshold-based and machine learning-based classifier.

4 Methodology

The proposed fall detection system is based on the fusion of significant Spatio-temporal features. Estimated motion based on Motion History Images and notable spatial features of the foreground image is combined here to design the proposed system. A two-channel classification of fall and daily life activities are carried out. One channel is a feature threshold-based classification, and the other is a keyframe-based classification using K-NN classifier. Further, these classification results are combined using some additional knowledge to make the system more robust and efficient. The flowchart of the proposed fall detection methodology is illustrated in Fig. 3. Figure 3a represents the two-channel fall detection system
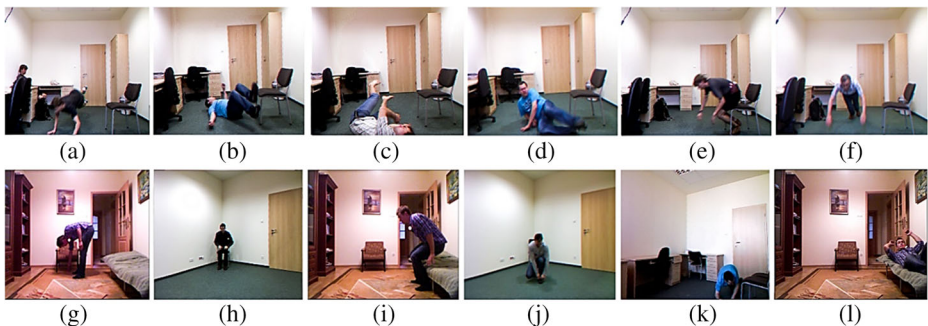
Fig. 2 (**a–f**) Training samples of fall sequences; (**g–l**) Training samples of activities of daily living (ADL) sequences

and Fig. 3b represents fall detection based on the combination of the two classification channels. The main steps of the proposed algorithm are Motion estimation, Human shape analysis, and Classification.

4.1 Motion estimation

Motion estimation enables the detection of rapid body movements like falls. To estimate the individual's motion in the surveillance video sequence, a coefficient MHI_{motion} is computed [9]. The computation is done using the motion history representing the most recent movement of the person. MHI_{motion} is shown in Eq. (4).

$$MHI_{motion} = \sum pixel(x,y) \in blob\left(H\tau_{(x,y,t)}\right) / Pixels \in blob \qquad (4)$$

Here, the *blob* represents the connected component criterion extracted using foreground segmentation. $H\tau$ represents the Motion History Image. The coefficient is ranged to a percentage of motion between 0 to 100% where 0% denotes no-motion and 100% for high-motion. The largest blob is considered here as it eliminates smaller motions. As the duration of a fall is generally very short, typically in milliseconds, we measure the MHI by collecting
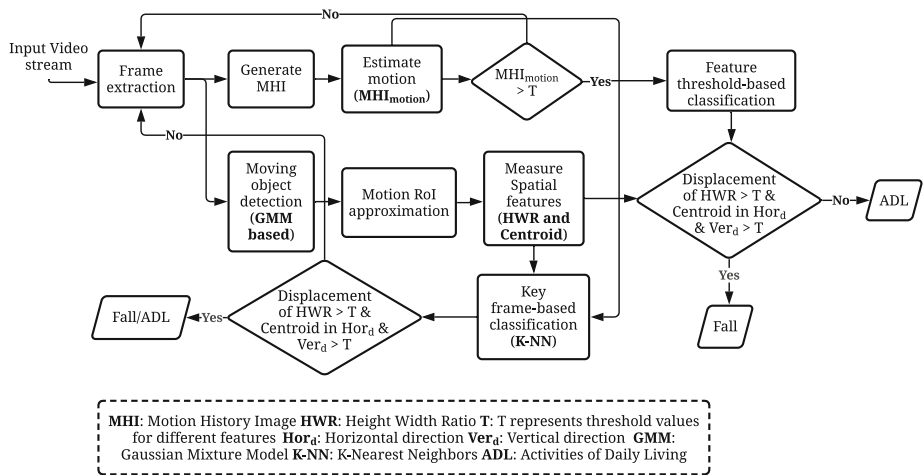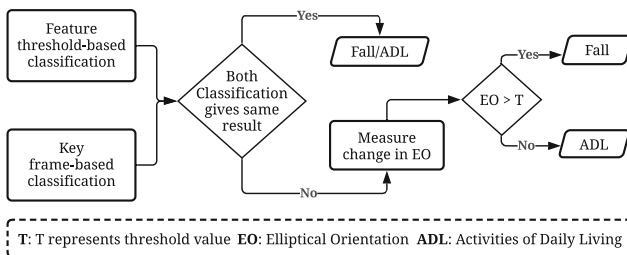
Fig. 3 Flowchart of the proposed fall detection methodology (a) Two-channel fall detection system (b) Fall detection based on the combination of the two classification channels

motion information for 350 ms. A motion or a rapid body movement is considered as a possible fall if MHI_{motion} is larger than 60%. Daily activities like walking, sitting abruptly, crouching, etc. can possess large and quick body movements. Hence, further analysis is essential to effectively distinguish a fall from an ADL.

4.2 Human shape analysis

Once we estimate the person's motion and a high motion ($MHI_{motion} > 60\%$) is detected, the significant changes in the human shape are analyzed precisely to distinguish a fall from other daily life activities. It is observed that during a fall, the horizontal or vertical or both the person's displacement in the frame is significantly higher than during any other regular life activity. Based on this observation, we measure three most important spatial features of the human shape to analyze the frames. One, the blob height-to-width ratio, the centroid displacement in the horizontal and vertical direction. These features are selected because we use the frontal video sequences of the UR fall detection dataset. Here the person in the frame faces the camera. The *cam0* feed of the dataset acquiring both fall and ADL sequences represents the frontal data. Therefore, a person falling parallelly to the camera's optical axis experiences a significant change in height-to-width ratio and vertical centroid movement. Whereas if he falls perpendicularly, the centroid in the horizontal direction moves significantly. As a result, we measure the absolute difference in the displacement of the chosen features. The absolute difference in the displacements of blob height-to-width ratio, horizontal centroid movement, and vertical centroid movement of the moving person during a fall and different ADL sequences are shown in Fig. 4.

The variance in the displacement of the features is calculated. It acts as the threshold to distinguish a potential fall from both fall-like daily activities (sitting, bending, crouching, lying down, etc.) and non-fall-like daily activities (walking, standing, etc.). Also, can be used to extract keyframes from a video sequence for efficient classification of fall and fall-like daily activities. The variance in the displacement of the height-to-width ratio can be calculated as follows [5].

$$\mu_{ar}(t) = (1 - \alpha)\mu_{ar}(t - 1) + \alpha AR(t) \tag{5}$$

$$\sigma_{ar}(t) = AR(t) - \mu_{ar}(t - 1) \tag{6}$$

In Eq. (5), AR(t) and μ_{ar}(t) denotes the displacement in aspect ratio or height-width ratio and its mean value at time t, respectively. While, μ_{ar} (t − 1) denotes the mean value at a time (t – 1). Value α represents the updated parameter. σ_{ar}(t) represents the variance at time t as shown in Eq. (6). Similarly, the variance in the centroid displacement in horizontal and vertical directions can be calculated as follows [5].

$$\mu_{(Chor\ or\ Cver)}(t) = (1 - \alpha)\mu_{(Chor\ or\ Cver)}(t - 1) + \alpha(CHOR\ or\ CVER)(t) \tag{7}$$

$$\sigma_{(Chor\ or\ Cver)}(t) = (CHOR\ or\ CVER)(t) - \mu_{(Chor\ or\ Cver)}(t - 1) \tag{8}$$

In Eq. (7), ($CHOR$ or $CVER$)(t) and $\mu_{(Chor\ or\ Cver)}$(t) denotes the centroid displacement in the horizontal or vertical direction and its mean value at time t, respectively. While, $\mu_{(Chor\ or\ Cver)}$(t − 1) denotes the mean value at a time (t – 1). Value α represents the updated parameter. $\sigma_{(Chor\ or\ Cver)}$(t) represents the variance at time t as shown in Eq. (8).
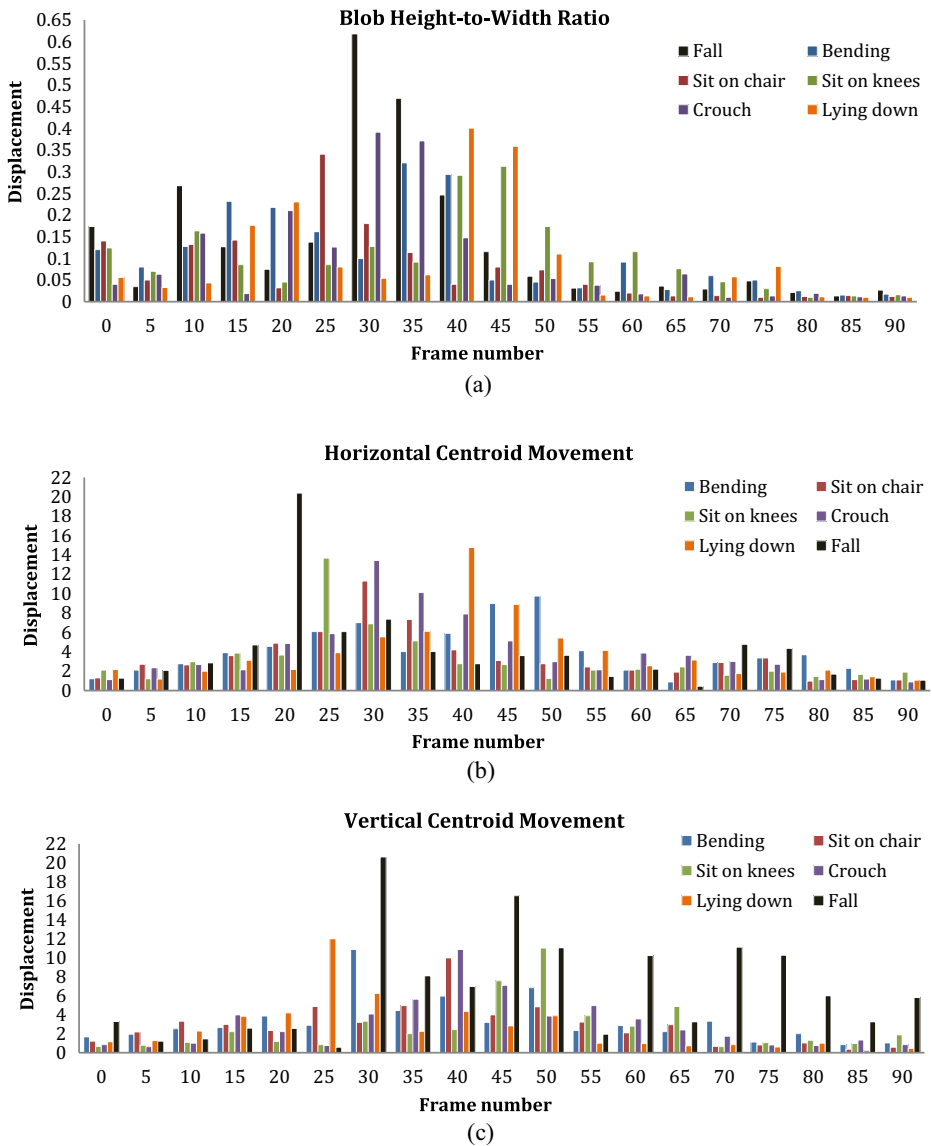
Fig. 4 Absolute difference in the displacement of (**a**) Blob height-to-width ratio (HWR) (**b**) Centroid movement in the horizontal direction and (**c**) Centroid movement in the vertical direction for fall and daily activity sequences

4.3 Classification

In this section, the fall and daily activity sequences are classified using a two-channel combined strategy. Channel one classifies the frames based on feature threshold and channel two selects keyframes which are then classified using a machine learning model. The two channels are then combined based on additional knowledge, which enhances the system's overall accuracy. The classification techniques are elaborated as follows.

4.3.1 Feature threshold-based classification

The estimated motion and the displacement in selected spatial features to analyze the human shape are the thresholds to distinguish between a potential fall and ADL. The need to threshold these parameters occurs once there is a substantial movement of the person in the frame, as discussed earlier in Section 3. Hence, when $MHI_{motion} > 60\%$, the variance in the centroid displacement in the horizontal and vertical direction, and the variance in the displacement of height-to-width ratio are the threshold to detect a fall in a surveillance video among any other activities of daily living. The variance is observed to be significantly high during a falling movement than any other normal activity. The threshold for the displacement in the height-to-width ratio: T_{ar}, horizontal centroid displacement: T_{Chor}, and vertical centroid displacement: T_{Cver} is set. We consider a large motion in the blob as a fall if T_{ar}, T_{Chor}, and T_{Cver} are higher than 0.4, 16.5, and 17.2 respectively. These thresholds were chosen through logical reasoning and by evaluating and observation of the training sequences. Accordingly, a potential fall can be detected in the middle of a video sequence using this threshold set. However, if it is set too high, it sometimes may get unnoticed, and when it is set too low, fall resembling activities with large motion such as sudden sitting, crouching, lying down can get detected, and the false alarm rate goes up. Similarly, more falls can be seen by reducing the threshold for MHI_{motion}. But this will lead to false detections such as during a sudden sitting or crouching when the motion is significantly high and if there is a sharp change in height-to-width ratio and centroid displacement in both directions. Examples of different activity frames (**a–g**), corresponding Motion History Images (**h–n**), and corresponding motion RoIs (**o–u**) are presented in Fig. 5, respectively. MHI_{motion} and the displacement in the selected human shape-based features, namely *AR*, *CVER*, and *CHOR* have been measured for the activity frames presented in Fig. 5 and shown in Table 1.

The results shown in Table 1 represent the estimated motion and the displacement in the spatial features for the respective activity frames as presented in Fig. 5. The first frame represents a fall event. Here fall is detected because MHI_{motion} is significantly high at 84.6%. Also, the displacement in centroid (both *CVER* and *CHOR*) and height-to-width ratio (*AR*) is substantial and above the threshold. The next frame represents a bending activity. The person is bending quickly to pick up something. Here, MHI_{motion} is just higher than the threshold as the person bends quickly, *AR* is also above the threshold but *CHOR* and *CVER* are below the threshold, so fall is not detected and is considered as an ADL. The next is a Sit-on-chair activity. Here, MHI_{motion} is lower than the threshold as no large motion is present because the person sits slowly on the bed. The algorithm stops due to lack of motion and iterates to check the next frame in the sequence. The next activity is Squatting or Sit-on-knees. The volunteer in the frame simulates this action by sitting on his knees quickly. It generates a high MHI_{motion} at 71.35% and simultaneously a higher than threshold displacement of the shape-based features. As a result, the activity is falsely classified as a falling movement. The next two frames represent identical fall-like activities such as crouching and lying down. In both the frames, the MHI_{motion} is below the threshold as the volunteers perform the actions casually by crouching on the floor and lying down on the bed. Also, the displacement in the spatial features is below the threshold. Hence, these activities are considered ADLs. The last example frame shows a walking event. Here the person walks fast, so there is a good amount of motion and is above the threshold. A possible falling event is considered. *CHOR* also represents significant horizontal displacement but remaining spatial feature displacements are much below the threshold. So, it is labeled as an ADL.
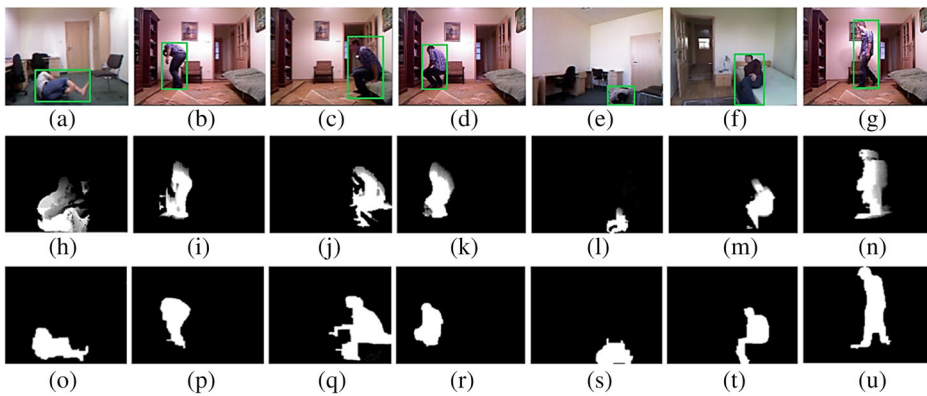
Fig. 5 (**a–g**) Examples of activity frames representing Falling, Bending, Sit-on-chair, Sit-on-knees/Squatting, Crouching, Lying down, and Walking; (**h–n**) corresponding Motion History Images; (**o–u**) corresponding motion RoIs

4.3.2 Keyframe-based classification

Random selection of frames for classification may not result in an optimum result. Thus, frames with the displacement of the chosen shape-based features for a certain threshold are a suitable choice. Classification of these selected frames will deliver an enhanced classification accuracy and simultaneously improve time complexity [23, 24]. The keyframes are chosen based on the observation that when a fall and fall-like daily activity occurs in the activity video sequence, the horizontal or vertical or both the person's displacement is higher than during a non-falling activity or inactivity. For analyzing the keyframes, we consider displacement in the height-to-width ratio and the centroid's displacement of the person concerning the floor in the horizontal and vertical direction. Frames with variance in displacement above a certain threshold in the horizontal or vertical direction or both are selected for classification. This threshold helps to restrict an activity phase (fall and fall-like) from a stationary phase (non-falling or inactive) in the video sequence.

The variance tends to be low when there is little change in displacement of the person, such as in a steady or inactive phase. It is higher for a variation in displacement during a fall or a fall-like activity such as sitting, bending, crouching, etc. Upon exceeding a threshold, it can be used to detect a significant variation in displacement from a steady phase to an activity phase. The frames with change in the person's displacement represent an activity phase if $T_{ar}, T_{Chor,}$ and T_{Cver} are higher than 0.2, 2.7, and 2.9, respectively. This threshold setting enables the

Table 1 Results of estimated motion and the displacement in the selected human shape-based features for the activity frames presented in Fig. 5

Activity frames	MHI_{motion} (%)	AR	CVER	CHOR	Threshold-based output
Falling	84.6	0.47	20.63	20.38	Fall
Bending	62.5	0.41	10.87	9.76	ADL
Sit on a chair	52.7	0.34	9.98	11.31	ADL
Squatting	71.35	0.42	17.4	16.9	Fall
Crouching	40.2	0.37	10.81	13.43	ADL
Lying down	58.5	0.35	12.01	14.78	ADL
Walking	76.2	0.21	5.98	18.01	ADL

detection of a fall or fall-like daily activity in the middle of a video sequence. However, if it is set too high, fall and fall-like activities may not get detected, and setting it too low increases the false alarm rate. Hence, the keyframes can be considered as those spanning the activity phase. Figure 6 represents the activity phase detection.

The keyframes are classified under two classes, namely fall and activities of daily living (ADL). For training and classifying the keyframes, we have chosen the K-NN classification model. As the proposed system has limited features for training, K-NN classification tends to enhance the system's accuracy for a limited number of features compared to the training data. The keyframes' displacement variance and estimated motion (MHImotion) are input to the K-NN classification model. The two-channel classification using both threshold-based and keyframe-based approaches can result in a classification disparity for a particular frame subject to classification, producing different outputs. As a result, we use additional knowledge to resolve the issue and present the final decision. We select the displacement in the elliptical orientation of the foreground moving object as additional information. As the classification model is designed using the frontal URFD video sequences, the person's orientation concerning the floor is selected as a significant foreground feature. It is observed from the training samples that the displacement in the orientation of the person is much higher during a falling movement compared to any other regular life activity. The displacement takes place in the horizontal or vertical direction or both. It is the threshold to distinguish between a fall and an ADL. After that, the two classification channels are combined using the decision obtained from the orientation displacement for a keyframe. Figure 7 represents the absolute difference in the displacement of the elliptical orientation of a person during a fall and different ADL sequences.

5 Experiment and results

This section evaluates the efficiency of the proposed fall detection system. Experiments are conducted on 30 falls and 40 ADL sequences of the UR Fall Detection Dataset to evaluate the
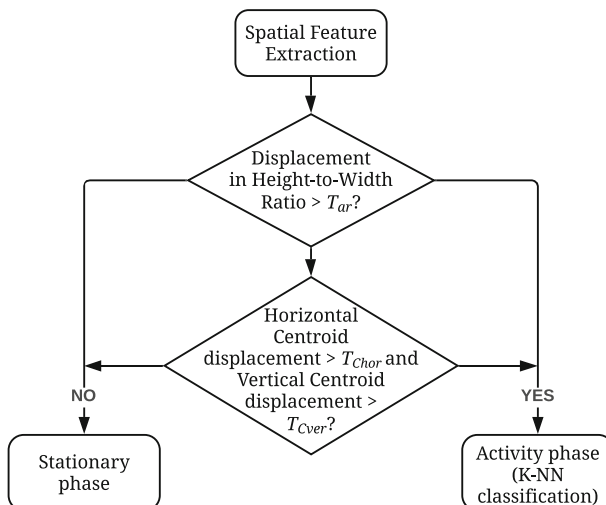

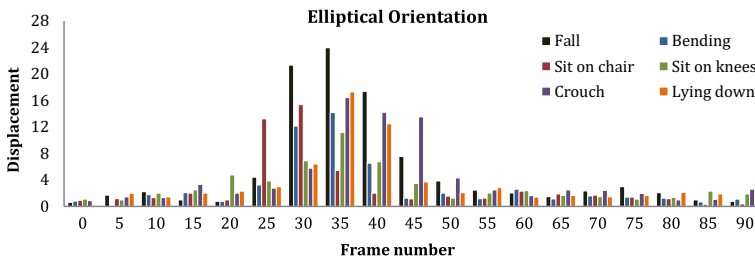Fig. 6 Flowchart representing activity phase detection

Fig. 7 Absolute difference in the displacement of elliptical orientation for fall and daily activity sequences

performance of the proposed system. The *cam0* data referring to the frontal URFD sequences are evaluated as we considered both fall and ADL video sequences. Being a vision-based technique, we considered only the RGB frames of the UR Fall Detection dataset. Experiments were conducted using MATLAB on a system having a configuration of Intel Core i5 2.42 GHz processor with 8 GB of RAM.

5.1 Performance evaluation

Performance metrics widely used in fall detection methods are used here to evaluate the performance of the proposed system as shown in Eqs. (9), (10), (11), and (12).

$$Sensitivity/Recall~(\%) = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity~(\%) = \frac{TN}{TN + FP} \tag{10}$$

$$Precision~(\%) = \frac{TP}{TP + FP} \tag{11}$$

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

Sensitivity/Recall, Specificity, Precision, and Accuracy are represented by four performance parameters: TP, FN, TN, and FP. True Positives (TP) determines a person is falling, and the system detects it correctly. However, a system that fails to see the fall is denoted by False Negative (FN). True Negatives (TN) determines a person carrying out daily activities, and the system detects them correctly. False Positives (FP) defines an everyday activity as a fall event. Sensitivity/Recall denotes the capability of the system to detect falls, and specificity indicates the ability of the system to see ADLs. Precision is a positive predictive value. Accuracy determines the overall classification rate of the system.

Eighty percent of the UR Fall Detection dataset comprising of 24 fall sequences and 32 daily activity sequences are used as training data to design the two-channel classification model. Both the models are evaluated based on the threshold-set and keyframes, respectively, that are subject to classification. Tables 2, 3, and 4 show the confusion matrix of the proposed

🖄 Springer

Table 2 Confusion matrix of the binary classification based on feature threshold-based classification

	Actual Fall	Actual ADL
Predicted Fall	28 **(TP)**	03 **(FP)**
Predicted ADL	02 **(FN)**	37 **(TN)**

Table 3 Confusion matrix of the binary classification based on the keyframe-based classification

	Actual Fall	Actual ADL
Predicted Fall	29 **(TP)**	02 **(FP)**
Predicted ADL	01 **(FN)**	38 **(TN)**

fall detection method based on binary classification of the video sequences of the URFD dataset into two classes: Fall and ADL. The confusion matrix of the two-channel fall detection system based on feature threshold-based classification and keyframe-based classification are presented in Tables 2 and 3 respectively. Table 4 shows the confusion matrix of the proposed approach by combining the output of the two classification channels. We evaluated the system's performance in terms of Sensitivity/Recall, specificity, precision, and accuracy. Table 5 presents the quantitative performance of the proposed system using different classification techniques, namely feature threshold-based, keyframe-based, and two-channel combined approaches. In Table 6, the performance of the proposed methodology is compared with state-of-the-art existing fall detection techniques based on the frontal camera sequences of the UR Fall detection dataset. The comparison is made based on the performance parameters, namely Specificity, Recall/Sensitivity, Precision, and Accuracy. It is to be noted, '–' symbol indicates the data is not available. In [26], features evaluated are the threshold to distinguish between a fall and regular activity. Human shape-based features like height and centroid are used for the evaluation of the technique. In [8], the authors use centroid displacement and optical flow vectors to design the fall detection system. Classification of the events into two classes is done using SVM. Authors in [43] implement a CNN-based deep learning technique to classify the fall and daily activity events. Pre-processed foreground frames are input to the CNN network. Based on the performance parameters evaluated, the proposed method using

Table 4 Confusion matrix of the binary classification by combining the two classification channels

	Actual Fall	Actual ADL
Predicted Fall	30 **(TP)**	01 **(FP)**
Predicted ADL	00 **(FN)**	39 **(TN)**

Table 5 Quantitative performance of the proposed approach using different classification techniques

Classification	Recall (%)	Specificity (%)	Precision (%)	Accuracy (%)
Feature threshold-based	93.33	92.5	90.32	92.85
Keyframe-based	96.66	95	93.54	95.71
Two-channel combined	100	97.5	96.77	98.6

Table 6 Performance comparison of the proposed method with existing fall detection techniques based on the frontal sequences of the URFD dataset

Approach	Features used	Classification	Input signal	Recall (%)	Sp (%)	Prec (%)	Acc (%)
Merrouche and Baha [26]	• Height • Centroid tracking	Feature threshold	Depth	76.92	80	–	78.68
Gracewell and Pavalarajan [8]	• Centroid displacement • Optical flow vectors	SVM	RGB	100	95	93.75	97.14
Wang et al. [43]	• Pre-processed fore-ground frames	CNN	RGB	94.86	–	99.64	–
Proposed approach	• MHI • HWR displacement • Centroid displacement (Horizontal and vertical) • Orientation displacement	Feature threshold Keyframes (K-NN) Two-channel combined	RGB	93.33 96.66 100	92.5 95 97.5	90.32 93.54 96.77	92.85 95.71 98.6

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

the combined two-channel classification outperforms the existing techniques in fall detection capacity. The proposed algorithm achieves 100% sensitivity in detecting falls using the two-channel combined classification. At the same time, our method is a hybrid Spatio-temporal technique using a combined threshold-based and machine learning-based system resulting in robust performance. The overall performance of the proposed method is found to be very impressive. It shows 92.85% and 95.71% accuracy using the threshold-based and keyframe-based classification, respectively. Combining the two classification streams results in a significant jump at 98.6% accuracy. All the compared approaches use the publicly available UR Fall Detection dataset [18] comprising video sequences recorded in simulated indoor environments. RGB frames are considered as the input signal to the proposed method. It supports the implementation of a conventional RGB camera to design an economical fall detection system.

6 Conclusion and future directions

This paper proposes a new approach to elderly fall detection by integrating the input frames' motion and significant human shape-based features. A two-channel classification strategy is adopted to detect falls from regular life activities: threshold-based classification and keyframe-based classification. Also, when there is any classification disparity between the two channels in deciding a fall or regular activity event, additional knowledge is used to classify the frames. A two-channel combined classification technique is used on integrated motion and shape-based data. Experiments show the proposed algorithm displays promising results. It achieves robust performance based on the frontal camera feed of the URFD sequences. Our approach is evaluated based on real-time sequences of simulated falls and ADLs by young volunteers. Behavioral characteristics like body posture, time is taken, human gait, etc., can be considered between young and elderly to improve the visual interpretation of human behavior. The analysis of the proposed approach is based on artificial lighting conditions, and it should be enhanced for dark environments too for use in real-life situations. A single camera is used for capturing the video sequences. Incorporating multiple cameras to view the person from

separate angles can enhance feature extraction. Moreover, we expect our proposed method to be improved by applying modern deep learning techniques in the future.

Declarations

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

References

1. Ahad MAR, Tan JK, Kim H, Ishikawa S (2012) Motion history image: its variants and applications. Mach Vis Appl 23(2):255–281
2. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical flow techniques. Int J Comput Vis 12: 43–77. https://doi.org/10.1007/BF01420984
3. Bobick AF, Davis JW (2001) The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 23(3):257–267
4. 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
5. Chen M-C, Liu Y-M(2013) An indoor video surveillance system with intelligent fall detection capability. Math Probl Eng 2013:1–8
6. Directalert. Wireless emergency response system [Online]. Available: http://www.directalert.ca/emergency/help-button.php. Accessed 2 July 2021
7. Garripoli C, Mercuri M, Karsmakers P, Soh PJ, Crupi G, Vandenbosch GA, Schreurs D (2014) Embedded DSP-based tele health radar system for remote in-door fall detection. IEEE J Biomed Health Inform 19(1): 92–101
8. Gracewell JJ, Pavalarajan S (2019) Fall detection based on posture classification for smart home environment. J Ambient Intell Humaniz Comput 2019:1–8
9. Gunale K, Mukherji P (2018) Indoor human fall detection system based on automatic vision using computer vision and machine learning technique. J Eng Sci Technol 13(8):2587–2605
10. Gutiérrez J, Rodríguez V, Martin S (2021) Comprehensive review of vision-based fall detection systems. Sensors 21:947. https://doi.org/10.3390/s21030947
11. Han Q, Zhao H, Min W, Cui H, Zhou X, Zuo K, Liu R (2020) A two-stream approach to fall detection with MobileVGG. IEEE Access 8:17556–17566
12. Hao Z, Duan Y, Dang X, Xu H (2019) KS-FALL: indoor human fall detection method under 5 GHz wireless signals. IOP Conf Ser Mater Sci Eng 569(3):8
13. Htun SN, Zin TT, Tin P (2020) Image processing technique and hidden Markov model for an elderly care monitoring system. J Imaging 6(49)
14. 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
15. Hussain F, Hussain F, Ehatisham-ul-Haq M, Azam MA (2019)Activity-aware fall detection and recognition based on wearable sensors. IEEE Sensors J 19(12):4528–4536
16. 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
17. 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
18. Kepski M, Kwolek B (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Prog Biomed 117(3):489–501
19. 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
20. Kerdjidj O, Ramzan N, Ghanem K, Amira A, Chouireb F (2019) Fall detection and human activity classification using wearable sensors and compressed sensing. J Ambient Intell Humaniz Comput 2020: 1–13
21. Khaire P, Kumar P, Imran J (2018) Combining CNN streams of RGB-D and skeletal data for human activity recognition. Pattern Recogn Lett 115:107–116

22. Krumholz A, Hopp J (2008) Falls give another reason for taking seizures to heart. Neurology 70:1874–1875
23. Luo Y, Zhou H, Tan Q, Chen X, Yun M (2018) Key frame extraction of surveillance video based on moving object detection and image similarity. Pattern Recog Image Anal 28:225–231. https://doi.org/10.1134/S1054661818020190
24. 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
25. 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
26. 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
27. Paul M, Haque SME, Chakraborty S (2013) Human detection in surveillance videos and its applications—a review. EURASIP J Adv Signal Proc 176
28. Peng Y, Peng J, Li J, Yan P, Hu B (2019) Design and development of the fall detection system based on point cloud. Procedia Comput Sci 147:271–275
29. Queralta JP, Gia T, Tenhunen H, Westerlund T (2019)Edge-AI in Lora-based health monitoring: fall detection system with fog computing and LSTM recurrent neural networks. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP) (IEEE), pp 601–604. https://doi.org/10.1109/TSP.2019.8768883
30. Ramachandran A, Karuppiah 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
31. Rubenstein LZ (2006) Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 35:ii37–ii41
32. Russell-Jones DL, Shorvon SD (1989) The frequency and consequences of head injury in epileptic seizures. J Neurol Neurosurg Psychiatry 52:659–662
33. 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
34. Sadreazami 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
35. 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
36. 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). Fort Collins, CO, USA, pp 246–252
37. Sterling DA, O'connor JA, Bonadies J (2001) Geriatric falls: injury severity is high and disproportionate to mechanism. J Trauma Inj Infect Crit Care 50:116–119
38. 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
39. Tian Y, Lee G-H, He H, Hsu C-Y, Katabi D (2018)RF-based fall monitoring using convolutional neural networks. Proc ACM Interact Mob Wearable Ubiquit Technol 2(3):1–24. https://doi.org/10.1145/3264947
40. United Nations (2017) World population ageing 2017—Highlights (ST/ESA/SER.A/397). Department of Economic and Social Affairs, Population Division
41. Vollset SE, Goren E, Yuan C-W, Cao J, Smith AE, Hsiao T, Bisignano C, Azhar GS, Castro E, Chalek J, Dolgert AJ, Frank T, Fukutaki K, Hay SI, Lozano R, Mokdad AH, Nandakumar V, Pierce M, Pletcher M, … Murray CJL (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 396:1285–1306. https://doi.org/10.1016/S0140-6736(20)30677-2
42. 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
43. 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 (RICAI 2019), Shanghai, China, pp 242–246. https://doi.org/10.1145/3366194.3366236
44. Wang X, Ellul J, Azzopardi G (2020) Elderly fall detection systems: a literature survey. Front Robot AI 7:71. https://doi.org/10.3389/frobt.2020.00071
45. WHO (2018) Falls. https://www.who.int/news-room/fact-sheets/detail/falls. Accessed 16 Jan 2018
46. 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 2020:1–12. https://doi.org/10.1155/2020/9532067

47. Zhang J, Wu C, Wang Y (2020) Human fall detection based on body posture Spatio-temporal evolution. Sensors 20(3):946. https://doi.org/10.3390/s20030946
48. Zitouni M, Pan Q, Brulin D, Campo E (2019) Design of a smart sole with advanced fall detection algorithm. J Sensor Technol 9:71–90

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