IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 15, 2018, accepted November 8, 2018, date of publication November 14, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2881237

Supporting Independent Living for Older Adults; Employing a Visual Based Fall Detection Through Analysing the Motion and Shape of the Human Body

AHMAD LOTFI[1], (Senior Member, IEEE), SUAD ALBAWENDI[1], HEATHER POWELL[1], KOFI APPIAH[2], (Member, IEEE), AND CAROLINE LANGENSIEPEN[1]
[1]School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, U.K.
[2]Department of Computing, Sheffield Hallam University, Sheffield S1 2NU, U.K.

Corresponding author: Ahmad Lotfi (ahmad.lotfi@ntu.ac.uk)

ABSTRACT Falls are one of the greatest risks for older adults living alone at home. This paper presents a novel visual-based fall detection approach to support independent living for older adults through analysing the motion and shape of the human body. The proposed approach employs a new set of features to detect a fall. Motion information of a segmented silhouette when extracted can provide a useful cue for classifying different behaviours, while variation in shape and the projection histogram can be used to describe human body postures and subsequent fall events. The proposed approach presented here extracts motion information using best-fit approximated ellipse and bounding box around the human body, produces projection histograms and determines the head position over time, to generate 10 features to identify falls. These features are fed into a multilayer perceptron neural network for fall classification. Experimental results show the reliability of the proposed approach with a high fall detection rate of 99.60% and a low false alarm rate of 2.62% when tested with the UR Fall Detection dataset. Comparisons with state of the art fall detection techniques show the robustness of the proposed approach.

INDEX TERMS Ambient intelligence, ambient assisted living, smart homes, image motion analysis, machine learning, identification of persons, image classification.

I. INTRODUCTION

Falls in the older adults are a relatively common occurrence that can have dramatic health consequences. Studies have shown that 28 − 34% of the older adults have at least one fall every year. In addition, falling is the second biggest cause of accidental death for more than 87% of older adults. New technologies like wearable fall detectors are used to support independent living and provide security for the older adults [1]. Several techniques for automatically detecting falls have been proposed [2], [3]. The existing technologies can be classified into three main groups of fall detectors, namely: ambient device-based [4], wearable sensor-based [5] and computer vision-based techniques [6].

Ambient device-based techniques use vibration or pressure sensors which are installed on the floor surface or under the bed. Sensors are used to capture the sound and vibration to detect the presence and position of a person [4]. Although these devices are inexpensive and do not disturb the user [7], the detection rate is rather low and many false alarms are generated [8].

Wearable devices use different sensors such as microswitches, accelerometer and gyroscopes. These sensors are widely used to capture the human body movement information and generate an alarm when the orientation or acceleration of the person reaches a predefined threshold [7]. There are some wearable devices which can be activated by pushing an alarm when there is a fall. However, such an alarm cannot be activated if the person is unconscious after the fall. Kepski and Kwolek [9] employed a tri-axial accelerometer to measure a person's acceleration and a depth sensor to extract depth features such as the ratio of bounding box and the distance of the person's centroid to the floor. Extracted features were then fed into a support vector machine classifier which obtained an accuracy of 98.33% with sensitivity and specificity of 100% and 96.67% respectively.

Much work has been undertaken investigating the use of visual-based sensors for fall detection using single [10], multiple [11] and omni-directional [12] cameras. Recently, depth cameras such as the Microsoft Kinect 2.0 [13] and ASUS XtionPro Live [14] have been used for detecting falls. The Kinect is a motion sensing technology which combines an RGB camera and a depth sensor to track the moving object in 3D [15]. Hbali *et al.* [16] proposed a new skeleton-based approach to describe spatio-temporal aspects of human activity for monitoring older adults using three-dimensional (3D) depth images. Two feature channels were calculated from the 3D joint position including the spatial and the temporal aspect of the activity. These features were used as inputs to an Extremely Randomized Trees algorithm to train the human activity model. The experiments were conducted using the Microsoft Research MSR 3D Action dataset [17]. The trained classifier model achieved an accuracy of 80.92%. Similarly, Ding *et al.* [18] proposed a method for skeleton-based human action recognition. Motions between rigid bodies are used to describe human posture, drawing movement components and mapping them to the points on a Grassmannian manifold. Then representative postures are extracted through spectral clustering. An action is represented by a symbol sequence generated with a global linear eigenfunction. Finally, a Hidden Markov Model (HMM) is used to classify these action sequences. The recognition rate is over 80%.

Thermal cameras are also used to track a thermal target and analyze its motion to detect a fall. The work in [19] presents a new approach for fall detection with an Infrared (IR) thermal array sensors. Features including the maximal temperature difference, the maximal motion distance, the duration of motion and variance of the maximal thermal difference between foreground and background over time are extracted for fall detection. The system achieves a fall recognition rate over 94% at room temperature (up to $24°C$).

A disadvantage of wearable sensors is that they are invasive. They require wearing and carrying various devices which could be forgotten, particularly by older adults with dementia. For computer vision-based methods, there is no need for older adults to wear an accelerometer-enabled device and its use is not affected by ambient noise. In addition, video-based solutions offer several advantages over ambient and wearable devices as they are capable of detecting other activities like sitting and bending.

In this paper, a novel computer vision-based fall detection approach suitable for a home environment is proposed. The proposed approach starts by detecting a moving object using a background subtraction algorithm. The second step is to extract robust features which describe the change in human shape and allow discrimination of falls from other activities like lying and sitting. These are based on motion, change in the human shape, projection histogram and temporal change of head position. These features thus extracted from the human silhouette are fed into a multilayer perceptron (MLP) neural network for fall detection. The main novelties of this paper include:

- Using the major semi-axis and minor semi-axis of the ellipse fitting around the human silhouette as features to increase the accuracy of the proposed fall detection.
- Finding a combination of features based on the way fall motion occurs that is capable of effectively detecting a fall.
- Using an MLP neural network to classify whether a set of features indicates a fall or not.

The rest of the paper is organised as follows: a brief review of publications related to the application of the vision-based system in fall detection is presented in Section II followed by an overview of the proposed system in Section III. Section IV describes in detail the features to be used, and discusses the use of a neural network for identification of falls. Section VI gives information about the dataset used for this research and the experimental process. Section VII presents results and discusses the performance of the proposed fall detection algorithm in comparison with state of the art visual-based fall detection approaches. Pertinent conclusions and future work are presented in Section VIII.

II. RELATED WORK

The Motion History Image (MHI) technique is widely used to extract motion from video sequences. The MHI indicates the speed of movement: if a person conducts an unusual activity such as a fast walk or run, it returns a high MHI value. The main limitation of MHI is that it only encodes the time from last observed motion at every pixel [20]. To overcome MHI constraints, a variant of it, such as timed Motion History Image (tMHI) method can be used. The tMHI method is used in [21] for motion segmentation to track objects in real time. This method makes the representation independent of camera speed or frame rate. Thus for a given movement, tMHI can cover the same MHI area at different capture rates.

One of the conventional methods for detecting a fall from video sequences is to represent the human shape using the bounding box method. This method is simple and easy to implement [22]. The bounding box attributes of height and width are used to represent the human physical shape. The approach proposed in [23] extracted four features from the bounding box around the human silhouette to describe a fall. The features were the aspect ratio, fall angle, centre speed and head speed. A Support Vector Machine (SVM) classifier was then employed to detect the fall using these features. However, The major drawback is the inadequate description of human motion by simply using a bounding box [11].

The other effective model is to represent the human in the video using an ellipse. The study in [24] presents a novel method to detect falls which combines the orientation angle and the ratio of a fitted ellipse around the human body, motion coefficient and silhouette threshold features. The extracted features are then used as inputs to a KNN classifier to classify fall events. The accuracy of the system was 95%. The authors in [25] used fourteen features extracted from the bounding box including height, width, aspect ratio, centroid coordinates of the box and ellipse orientation for

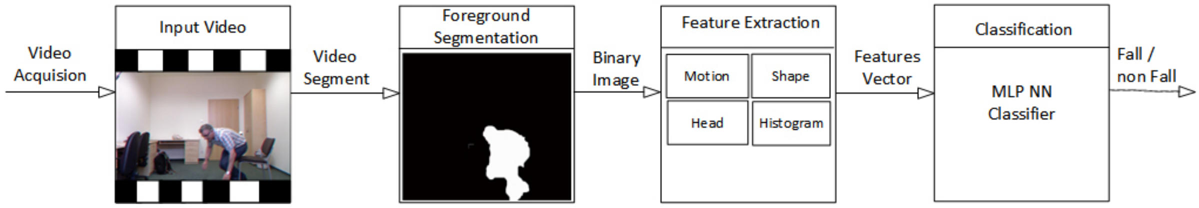

FIGURE 1. **Flow diagram of the proposed human fall detection approach.**

fall detection. Fourier and wavelet transformations were then applied to these features before fall detection by using either SVM or AdaBoost classifiers. Using the Le2i fall detection dataset, they achieved specificity of 100%, an accuracy of 99.9% and recall of 98%. The work presented in [26] proposes a fall detection approach using a Gaussian mixture background model to build the background. MHI is applied to analyse the fall behaviour and the orientation and the ratio of the ellipse are computed to represent the variation in the shape of the human object. In addition, two extra features, acceleration and angular acceleration, are computed to improve fall detection accuracy. However detailed performance data was not provided.

Fitting an ellipse is insufficient to describe the posture of the human body in detail and it might be hard to differentiate two postures by using only the global information [27]. Therefore more information from local features is required to describe different postures. A widely used feature to describe such detailed information is the projection histogram. The projection histogram features are computationally efficient to derive and produce a good performance for posture classification [3], [28]. The work in [28] presents fall detection by using the information from ellipse fitting and a projection histogram along the axes of the ellipse to distinguish different postures of the person. The system achieves a high fall detection rate of 97.08% in a simulated home environment. Similarly, the approach presented in [29] is based on measuring a temporal variation of pose change and body motion to detect falls. Several measures such as centroid velocity, head-to-centroid distance, a histogram of oriented gradients and optical flow were computed. The system can correctly classify 90.6% of falls. A comparison amongst several fall detection systems was performed in [30], showing sensitivities from 71 to 100%, specificities from 73% to 100% and accuracies of 84 to 94%.

Doulamis [31] proposes a fall detection framework based on the joint estimation of foreground object and motion information. The motion vectors on particularly selected points on the image plane and the vertical velocity of the upper boundary of the foreground object are estimated to detect falls in different directions from the camera position. In addition, a new method for fall detection has been proposed by [11] through analyzing dynamic shape and motion of human body regions on Riemannian manifolds. The method represents human activities by dynamic shape points and motion points moving on two simple Riemannian manifolds. Afterwards, the velocity statistics on the two manifolds are computed. The test results show a high detection rate of 99.38%. Min *et al.* [32] propose a fall detection method for falls against furniture such as sofas and chairs. Their method is based on the activity characteristics of the detected people such as motion speed and human shape aspect ratio. A deep learning method called R-CNN is employed to obtain the information of locations and objects in the scene. The method can distinguish falls from other activities with an accuracy of 95.50%.

Our proposed approach reported in the remainder of this paper makes an attempt to overcome some of the shortcomings of the existing research published in this area.

III. SYSTEM OVERVIEW

This paper proposes a method for monitoring human activities in a home environment and detecting a fall event based on motion information, changes in shape orientation and position of human head and projection histograms. An overview of the proposed fall detection system is shown in Figure 1. The proposed fall detection system includes four steps: data collection, foreground segmentation, feature extraction and fall detection. Background subtraction is implemented to segment out moving objects [33]. This method uses the difference between the current image and the background image to detect moving objects, which is in our case used to extract the human silhouette. The next step is to track the moving object to recognise the motion trajectory of an object as video frames progress by identifying the object position in every frame. Afterwards, useful features such as motion information, shape orientation, temporal change of the head and histograms for detecting a fall from other daily activities are extracted. The proposed method exploits motion, histogram and shape features based on the observation that human falls often involve drastic shape changes and abrupt motions as compared to other activities.

The first stage of the system is to analyse the motion occurring in a given time window, using tMHI [34]. The proposed system aims to detect high degrees of motion of the person in the video sequence using tMHI. The motion is quantified by calculating the pixel value of the motion history image blob in the current frame, which is then divided by the number of pixels in the human blob. The second stage is to analyse the change of the human shape to identify any fall. Analysis of the moving object is performed by fitting an approximate ellipse around the human body. The orientation of the fitted ellipse provides information about the body posture [35]. After ellipse fitting, the orientation of the ellipse and the ratio between the major semi-axis a and the minor semi-axis b are taken as features to describe human body posture in a general way. However, these features alone cannot describe postures in detail for distinguishing different activities [27], therefore more features are needed. A bounding box was used to surround the foreground object, then the y-coordinate of the top left point of the bounding box was computed and the absolute difference of y-coordinates in successive frames were used as features.

It is assumed that human falls have a higher acceleration than other daily activities. However, focusing only on fast acceleration can result in many false alarms during fall-like activities like sitting down quickly [1]. Therefore combining motion with other features extracted from the fitted ellipse around the human body helps to discriminate actual fall from other activities. After ellipse fitting, the orientation, ratio, major semi-axis and minor semi-axis are taken as features to describe human body posture. The motion feature C_{motion} indicates the changing rate of human motion and the orientation feature indicates the changing human shape [36].

Our previous study highlighted the situation when a human fall can occur at low speed. A typical example is when a person loses their balance, holds onto furniture to prevent a fall and yet still falls on the ground. Therefore an additional feature, the projection histogram, is applied to confirm a fall event [37]. Occlusion occurs when a relevant area of the person is covered or when the person moves behind an object and consequently part of his/her body disappears [38]. To deal with the occlusion problem, the proposed fall detection approach applies another feature - tracking the human head in consecutive frames. Tracking the head position of the person can provide useful information in such instances, as the head tends to be visible most of the time. Finally, the extracted features are used as input vectors to the MLP neural network for fall and non-fall event classification.

IV. ENHANCED FEATURES FOR FALL DETECTION

The feature extraction process includes the application of background subtraction algorithms to extract the human silhouette, and generation of features from the shape of the human silhouette. The ability to distinguish a fall action depends mainly on the quality of the classifier input, so the features of the extracted human silhouette play a key role in the effectiveness and robustness of detecting human falls [39]. This section presents details on how the features are extracted.

A. TIMED MOTION HISTORY IMAGE

The MHI is generalized by directly encoding the actual time in a floating point format, which is called Timed Motion History Image (tMHI) [21]. With the tMHI, the representation can be used to determine the current pose and also measure the motion of an object. In general, the tMHI is updated by time stamps of the video sequence, rather than the frame numbers [40]. The motion history image contains the trajectory information of the action being performed and recent motion is emphasized more than past motion [10]. This technique was used in our previous work [36].

When the human body region is extracted, the motion activity of the segmented foreground object is measured by generating a timed motion history image (tMHI). The tMHI image is computed as:

$$tMHI_{\delta}(x, y) = \begin{cases} \tau & \text{if } \textit{current silhouette at } (x, y), \\ 0 & \text{if } tMHI_{\delta}(x, y) < (\tau - \delta). \end{cases} \tag{1}$$

where τ is the current time-stamp and δ is the maximum time duration constant (typically a few seconds) associated with the template [21], [40].

B. QUANTIFY THE MOTION

To quantify human motion, it can be calculated using the pixel values of the motion history image, divided by the number of pixels in the human 'blob'.

In the proposed method, a coefficient C_{motion} is computed based on the tMHI using:

$$C_{motion} = \frac{\Sigma pixel(x, y) \in blob\ tMHI_{\delta}(x, y)}{\# pixels \in blob} \tag{2}$$

The term *blob* refers to the silhouette of a person extracted using the background subtraction method, and $tMHI_{\delta}$ means the timed Motion History Image for th emaximum time duration δ. The value of C_{motion} is a percentage ranging from 0% (no movement) to 100% (maximum movement) [41].

C. APPROXIMATED ELLIPSE

To represent the motion or the shape of the human body in a video sequence, a single object is surrounded by an ellipse. The approximated ellipse offers information in relation to the shape and orientation of the person in the image [35]. There are three important parameters of the ellipse. They are a) the vertical or current angle of the object, b) the major axis of the object, and c) the minor axis of the object. In [28], analysis of the moving object is performed to detect a change in human shape; particularly in orientation and proportion.

The moving object is approximated by an ellipse using a moment-based method which is detailed in the work by Forougi *et al.* [42]. The zeroth moment represents the total area of all the white pixels in the image. The coordinates x and y of the centre of the image are described by the spatial moments of first order m_{10} and m_{01} divided by the zeroth order moment m_{00}. Higher order moments may be calculated by similar summations. The zeroth, first and second order moments are then used with the centroid $(\bar{x}, \bar{y})$ to compute the central moments μ_{11}, μ_{20} and μ_{20} as follows:

$$\mu_{pq} = \sum_{pq} (x - \bar{x})^p \, (y - \bar{y})^q f(x, y) \tag{3}$$

where $f(x, y)$ denotes the pixel value at (x,y) in a binary image, and $p, q = 0, 1, 2, 3\ldots$. This leads to:

$$\mu_{11} = \frac{m_{11}}{m_{00}} - \bar{x} * \bar{y} \tag{4}$$

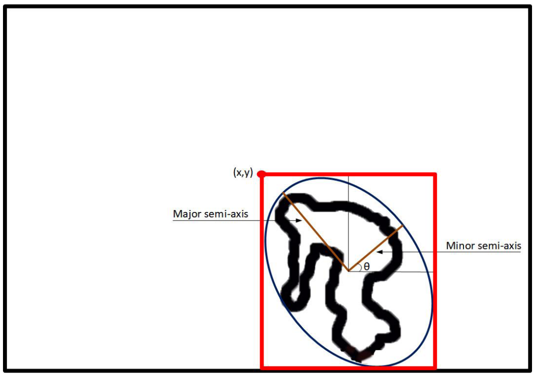

FIGURE 2. Ellipse fitting and bounding box around human body.

$$\mu_{20} = \frac{m_{20}}{m_{00}} - \bar{x}^2 \tag{5}$$

$$\mu_{02} = \frac{m_{02}}{m_{00}} - \bar{y}^2 \tag{6}$$

The central moments μ_{pq} of the image are required to calculate the orientation of the ellipse. The angle between the major axis of the person and the horizontal axis gives the ellipse orientation and can be computed with the central moments of second order. The orientation describes the direction of the major axis and is within the range: $-\pi/4 \leq \theta \leq \pi/4$.

$$\theta = \frac{1}{2} arctan(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}) \tag{7}$$

The eigenvalues I_{min} and I_{max} are given by:

$$I_{min} = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \tag{8}$$

$$I_{max} = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \tag{9}$$

Then, the major semi-axis, a, and the minor semi-axis, b, of the best fitting ellipse are given by:

$$a = (4/\pi)^{1/4} \left[\frac{(I_{max})^3}{I_{min}} \right]^{1/8} \tag{10}$$

$$b = (4/\pi)^{1/4} \left[\frac{(I_{min})^3}{I_{max}} \right]^{1/8} \tag{11}$$

The ratio of the ellipse is computed as $\rho = a/b$ [35]. Figure 2 illustrates features extracted from ellipse fitting and enclosed bounding box.

D. PROJECTION HISTOGRAM

The horizontal and vertical projection histogram of foreground object is obtained by calculating the number of foreground pixels row wise and column wise. The foreground F image is donated as cloud of $2D$ points with (x_p, y_p) as the pixel coordinates. The horizontal projection histogram $Hz(y)$ of foreground F can be defined as the cardinality of a set of points as follows:

$$Hz_{(y)} = \left|(x_p, y_p) \in F, (y_p = y)\right| \tag{12}$$

Similarly, the vertical projection histogram $Vt_{(x)}$ can then be computed as follows:

$$Vt_{(x)} = \left|(x_p, y_p) \in F, (x_p = x)\right| \tag{13}$$

For each activity, the horizontal and vertical projection histograms are computed for each frame. Then, the maximum values of foreground pixels in the horizontal and vertical histograms, as well as the difference between maximum values are calculated to effectively discriminate falls from other activities [42].

E. TEMPORAL CHANGES OF HEAD POSITION

The reason to track the head is mainly because the head is usually visible in the scene and has a large movement during the fall. When the fall occurs the head moves abruptly and its displacement would be large. Thus the proposed algorithm aims to estimates a person's head position in every frame of a video sequence. In order to determine the head position, the top left detected point of the silhouette is marked. Firstly, the silhouette is enclosed by a minimum bounding box and then the top left detected point of the bounding box is marked in each frame. In addition, the absolute difference values of the top left point of the head over successive frames are obtained and form a feature vector, which represents the vertical displacement of the head point [42].

In the moment when a fall occurs, the y-coordinate of the person's head increases significantly, which leads to considerable variance in the vertical velocity. Consequently, the y-coordinate of the person's head was selected as a feature for fall identification. In addition, the standard deviation of y-coordinate σ_y is calculated as:

$$\sigma_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left|y_i - \mu_y\right|^2} \tag{14}$$

where y_i is the y coordinate which is calculated from the i^{th} frame; μ_y represents the average value of y within the specified number of frames; N is number of frames [43].

The absolute difference of subsequent y-coordinates is also used as a feature for fall detection. The difference of the y-coordinate is computed by:

$$d_y = (y_i - y_{i-1}). \tag{15}$$

where y_i is the y coordinate of the human head in the ith frame and y_{i-1} is the y coordinate of the human head in the $(i-1)^{th}$ frame. The standard deviation of the absolute difference in the y-coordinate of the head is calculated using a similar equation to Eqn.14, where y_i is replaced by d_i.

F. SELECTED FEATURES

Based on the information provided in the preceding sections, for each video sequence, the moving object is detected and 10 unique features represented in the feature vector F are used to identify falls. The selected feature vector is given by:

$$F = [C_{motion}, \theta, \rho, a, b, Hz_{(y)} - Vt_{(x)}, y, \sigma_y, \left|y_i - y_{i-1}\right|, \sigma_{\left|y_i - y_{i-1}\right|}] \tag{16}$$

where the individual features areas follows:

- coefficient of motion C_{motion} - defined in Eqn. 2,
- the orientation of the ellipse θ - defined in Eqn. 7,
- the major semi-axis a,
- the minor semi-axis b of the ellipse and,
- the ratio of the ellipse ρ - i.e. the ratio of a/b where a and b are defined by Eqn. 10, 11,
- the difference between the horizontal and vertical projection histograms $Hz_{(y)} - Vt_{(x)}$ - which are defined by Eqn. 12 and 13
- the y-coordinate of the head point y,
- the standard deviation of y-coordinate σ_y - defined by Eqn. 14
- the absolute difference of y-coordinate d_y - defined by Eqn. 15
- the standard deviation of the absolute difference of y-coordinate $\sigma_{|y_i - y_{i-1}|}$ - defined by Eqn. 14 as applied to d_y.

For every video sequence, 20 frames per window with an overlap of 19 frames per window are chosen. For example, the first window contains frames from frame 1 to frame 20, the second window will contain frames from frame 2 to frame 21 and so on. Features are extracted for every frame, and every sequence of 20 frames is used to generate one feature vector which is used as an input to the neural network to detect the fall.

V. FALLS CLASSIFICATION AND DETECTION BY NEURAL NETWORK

In order to classify between falls and non-fall activities, the feature vector was fed as input to a three-layered Multilayer Perceptron (MLP) Neural Network, which is shown schematically in Figure 3. The size of the hidden layer is ten since there are ten features in the input feature vector. The hidden layer makes use of the scaled conjugate gradient algorithm. The use of this algorithm enables the network to perform well despite the dynamic range of the inputs, by reducing the number of iterations required when some features are much larger than the others. The outputs of the hidden layer are passed to the output layer, from where the decisive outputs are generated. A 10-fold cross-validation method was used in which the whole data was divided into 10 sets, where 9 were used to train and one to validate the system. The training and testing process was repeated for all possible combinations of the ten sets.

The data has already been normalised with zero mean and standard deviation equal to 1, so that it is consistent with the transfer function. A mean squared error function was chosen as the evaluation criterion. This function minimises the mean of the squares of the errors produced in each iteration and updates the network weights and biases accordingly. The performance goal for the training is decided using this error function only.

VI. EXPERIMENTAL DATASET AND PROCESS

The experiments were conducted using the UR Fall Detection Dataset [44], which contained 30 fall scenarios recorded by

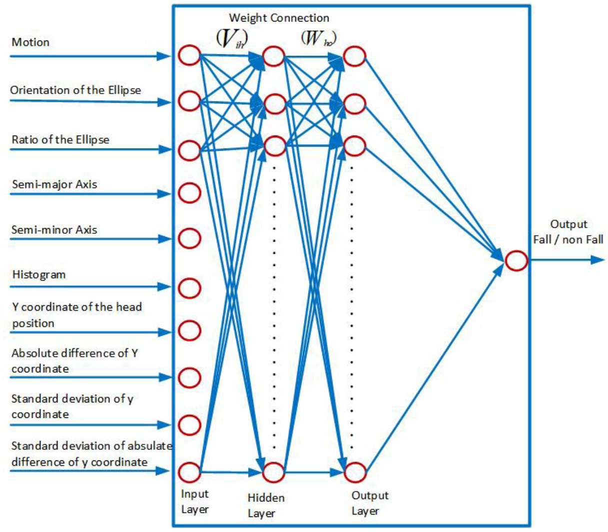

FIGURE 3. Fall detection neural network architecture.

two Kinect sensors and an accelerometer and 40 normal activities using one Kinect sensor parallel to the floor. The normal daily activities included walking, sitting, lying down, bending and crouching down. Some of these videos were recorded in low light conditions and some videos present examples of occlusion. Falls were simulated in different directions with respect to the camera view. Different types of fall incidents were recorded to include forward, backward and sideways falls. Frame samples taken from UR Fall Detection datasets are shown in Figure 4. In our tests, only RGB data was used.

In order to segment out the foreground object from the background, a simple background subtraction algorithm [33] was performed to extract the silhouette from the background. After the silhouettes were acquired through segmentation, the second step was to extract useful features from the human silhouette to detect falls. These are the features discussed and defined in section IV. The values associated with each feature during normal activity and fall are shown in Figure 5.

The whole database was divided into training and testing data. The training data contained 5087 × 10 samples and the testing data contained 2090 × 10 samples. The experiments were done using 10-fold cross-validation. The training samples were created from the selected features vectors and stored along with the target output values corresponding to the specific input patterns. The fully trained neural network was then used on the test data to classify the patterns associated with the activities of the humans and generate the outputs corresponding to the respective activities, which in our case were either fall or non-fall activities.

VII. PERFORMANCE EVALUATION

There are four possible outcomes for testing a sequence as a fall event which are defined as follows:

- True positive (TP): a video segment contains a fall, and is correctly detected as a fall.
- False positive (FP): a video segment does not contain falls, but is incorrectly detects as a fall.

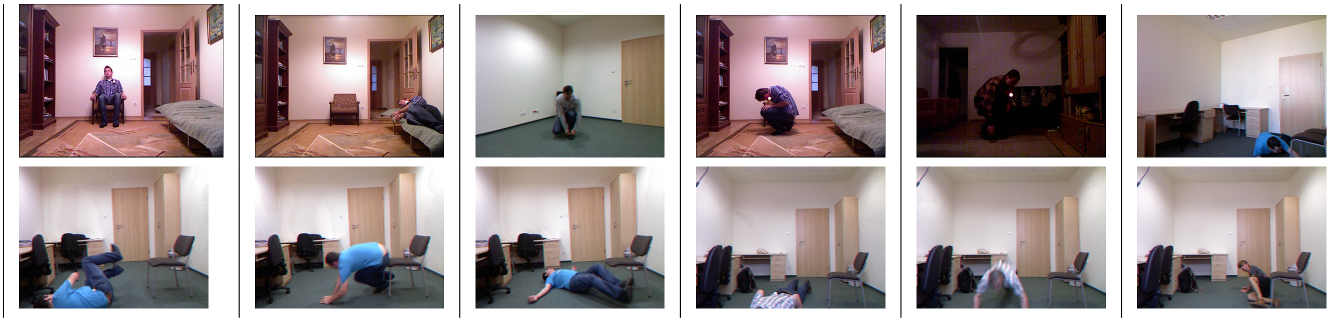

FIGURE 4. Sample frames from the UR Fall Detection Dataset. Upper row images are normal daily activities including sitting, lying down, crouching down and bending over. Lower row images are human falls in various ways.

TABLE 1. Recognition results of the proposed fall detection approach using different combination of features.

Method	Accuracy	F-Score	FPR	FNR	Precision	Sensitivity	Specificity
Case A - 4 features	88.69%	93.82%	79.02%	0.91%	89.08%	99.09%	20.98%
Case B - 6 features	92.77%	95.94%	44.75%	1.46%	93.47%	98.54%	55.25%
Case C - 8 features	96.35%	97.91%	16.89%	1.61%	97.43%	98.39%	83.11%
Case D - 4 key features	92.13%	95.54%	41.07%	0.0277%	93.91%	97.23%	58.93%
Case E - 10 features	99.24%	99.56%	02.62%	0.48%	99.60%	99.52%	97.38%

- True negative(TN): a video segment does not contain falls, and is correctly detects as non-fall.
- False negative(FN): a video segment contains a fall, but is incorrectly detected as not a fall.

The performance of the fall detector was evaluated with respect to accuracy, F-score, false positive rate (FPR), false negative rate (FNR), precision,sensitivity and specificity. They were calculated as follows:

- Accuracy = (TP+TN)/(TP+TN+FP+FN);
- F-score = 2TP/ (2TP + FP + FN);
- The false positive rate = FP/(FP+TN);
- The false negative rate = FN/(FN+TP);
- Precision = TP/(TP+FP);
- Sensitivity = = TP/(TP+FN);
- Specificity = TN/(TN+FP);

In order to evaluate the performance of the fall detection approach, subsets of input features were used.

- Case A: The first row of the Table1 presents the recognition results in which the input features of MLP network used are motion, the orientation and ratio of the ellipse, and projection histogram features. This subset of features corresponds to those we used in previous work on our own dataset, and are typically the main 4 motion features considered by other authors [37].
- Case B: The second row of the Table1 presents the recognition results where the input features of MLP network include motion, the orientation, the ratio, the major semi-axis and the minor semi-axis of the ellipse, and projection histogram features. Thus this row adds the 2 semi-axes of the ellipse, giving more information on orientation, size and change of proportion (potentially showing someone crumpling as they fall)
- Case C: The third row of the Table1 presents the recognition results obtained using features of motion, the orientation, the ratio of the ellipse, projection histogram, y-coordinate of the head point, the standard deviation of y-coordinate, the absolute difference of y-coordinate and the standard deviation of absolute difference of y-coordinate. This combines the basic motion characteristics with the detailed behaviour of the head in terms of its position and position variance.
- Case D: In order to further validate our feature choices, the data in Figure 5 was examined, and only those features which exhibited obviously large changes were used. These were the motion, the y coordinate, the ratio and the major semi-axis (as shown in graphs a,c, e and g).
- Case E: The last row of the Table1 shows the recognition results obtained using all features - motion, the orientation, the ratio, the major semi-axis and the minor semi-axis of the ellipse, projection histogram features, y-coordinate of head point, the standard deviation of y-coordinate, the absolute difference of y-coordinate and the standard deviation of absolute difference of y-coordinate.

Using all the features, the algorithm achieves 97.38% specificity and this means that most normal daily activities are assigned to the non-fall class. A high sensitivity implies that most falls are recognised as a fall event. The accuracy of human fall detection is 99.24%. The proposed system shows a high detection rate of 99.60% while maintaining a low rate of false alarms.

From the curves in Figure 5, it is possible to infer that the more significant features characterising the fall events are: motion, ratio, the major-semi axis and the y-coordinate of the head point. Table 1 show that these key features as in group D achieved similar accuracy to features group B which includes six features.

However, using only these features the system achieves an accuracy of about 92.13% and low specificity (true negative

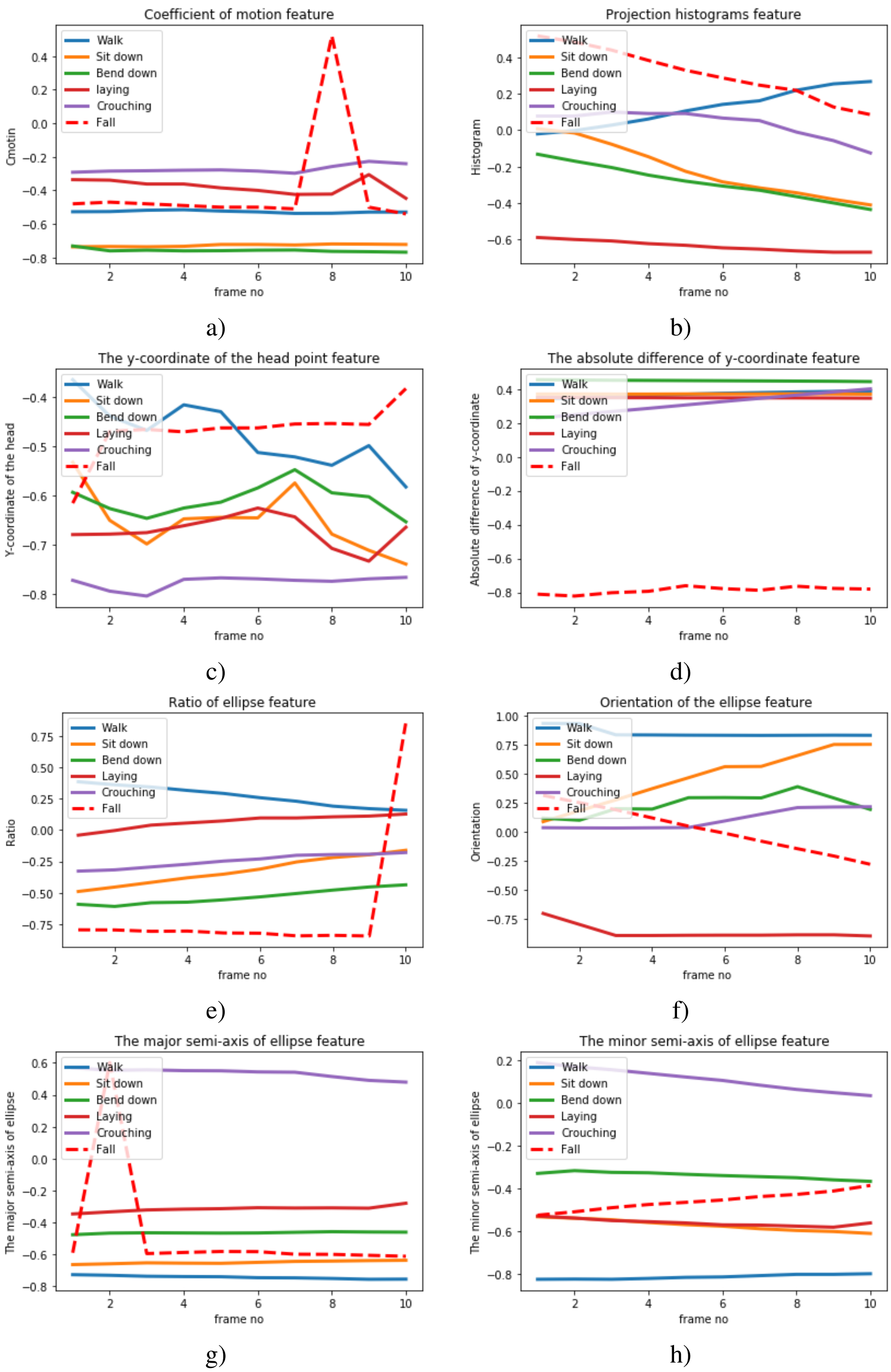

FIGURE 5. Features used to distinguish between normal daily activities and fall. Normal daily activities include walking, sitting, bending, crouching and lying down. Selected features are a) coefficient of motion b) the orientation of the ellipse c) the ratio of ellipse d) the major semi-axis e) the minor semi-axis f) the difference between the horizontal and vertical projection histograms g) the y-coordinate of the head point h) the absolute difference of y-coordinate.

rate) 58.93%. This means that some normal activities such as sitting down and lying are detected as falls, and these features are not sufficient to discriminate a real fall from a person lying or sitting down. It appears that these features cannot always distinguish between activities when there is a high degree of similarity in terms of high motion, or when the ratio of the approximated ellipse is nearly the same for both non-fall and fall activities. This can give rise to false positives.

In order to evaluate the performance of the proposed 10 feature system, a comparison with other vision-based fall detector systems is conducted. Compared to similar systems mentioned in Section II, our performance values (sensitivity

TABLE 2. Comparison of Existing Methods using the same database. Where no values were given by the authors, the fields are blank.

Method	Accuracy	Precision	Sensitivity	Specificity
Marcos (2017) Convolutional NN	95%	–	100%	92%
Kwolek and Kepski (2014) SVM	90%	83.30%	100%	80%
Harrou et al (2017) SVM	96.66%	93.55	100%	94.93%
Yun and Gu (2016)	–	–	100%	100%
Charfi (2013) SVM/Adaboost	99.42%	95.91%	92.15%	99.79%
Bourke (2007)	–	–	100%	100%
Proposed	99.24%	99.60%	99.52%	97.38%

of 99.52%, specificity of 97.38%, precision of 99.60% and accuracy of 99.24%) seem to be good considering that we are using only RGB images and modest hardware. The comparison between the presented fall detection approach and common methods to detect falls is shown in Table 2.

VIII. CONCLUSIONS AND FUTURE WORK

The work presented in this paper is mainly focused on investigating a relatively low-cost and reliable fall detection approach for older adults based on computer vision techniques. The approach presented here employs enhanced features which are extracted from the human silhouette. These features are the motion information, orientation, ratio, the major semi-axis and the minor semi-axis of the fitting ellipse, the projection histogram, the y-coordinate of the head point, the standard deviation of y-coordinate, the absolute difference of y-coordinate and the standard deviation of absolute difference of y-coordinate.

Experimental results show that the proposed algorithm is reliable for fall detection. The proposed approach is based on a combination of timed motion history and variation in human shape. The combination of motion and change in the human shape offers crucial information about human activities. Firstly, timed Motion History Image is implemented to quantify the motion of the person, the person is approximated by an ellipse using moments to detect a change in the human shape by computing the orientation, the ratio, the major semi-axis and the minor semi-axis of the ellipse [22]. In addition, the local feature which is based on projection histograms is applied to identify fall among other daily activities. Tracking of head position features was used to improve performance of the presented system. These features are then fed into MLP Neural Network for fall classification. It can be observed that the proposed algorithm produces a high recognition rate of 99.60% while maintaining a low false alarm rate of 02.62%.

The combination of features is significant. The histogram feature helps in identification of sudden changes in human body shape. Adding the orientation feature helps in charactерising a human's fall because it has less variation during the frames of the video sequence until a fall happens. Then the orientation will decrease suddenly until the human body reaches a prone position after the fall.

When a large motion is detected, the orientation and ratio of the ellipse change suddenly and at the same time, the histograms decrease rapidly. This means a fall event is considered possible. In the proposed fall detection approach more features are employed to decide the human fall, the results have shown that the use of major and minor semi-axis of the ellipse contributes towards accurate fall detection. Additionally, head features are applied for reliable classification of motion and determination of a fall event.

From the analysis, the system achieved the best performance and lowest false alarm rate after adding the difference between the horizontal and vertical projection histograms, the standard deviation of y-coordinate and the absolute difference of y-coordinate as input features. Detection results obtained using a different combination of features have shown that the best performance regarding accuracy and specificity is obtained when using all ten features together.

In future work, various machine learning methods including SVM and KNN which have been documented to be appropriate to study human activities will be applied to the aim of fall classification. Moreover, to enhance our system capability to recognise more activities, the plan is to use other datasets and assess its robustness on a wider range of data.

ACKNOWLEDGMENT

Authors would like to thanks Dr Mufti Mahmud's for his advice and comments.

REFERENCES

[1] P. Kostopoulos, T. Nunes, K. Salvi, M. Deriaz, and J. Torrent, "F2D: A fall detection system tested with real data from daily life of elderly people," in *Proc. 17th Int. Conf. E-health Netw., Appl. Services (HealthCom)*, Oct. 2015, pp. 397–403.

[2] H. Foroughi, A. Rezvanian, and A. Paziraee, "Robust fall detection using human shape and multi-class support vector machine," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, Dec. 2008, pp. 413–420.

[3] C.-L. Liu, C.-H. Lee, and P.-M. Lin, "A fall detection system using k-nearest neighbor classifier," *Expert Syst. Appl.*, vol. 37, no. 10, pp. 7174–7181, 2010.

[4] Y. Zigel, D. Litvak, and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2858–2867, Dec. 2009.

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

[6] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Monocular 3D head tracking to detect falls of elderly people," in *Proc. 28th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBS)*, Aug. 2006, pp. 6384–6387.

[7] B. Mirmahboub, S. Samavi, N. Karimi, and S. Shirani, "Automatic monocular system for human fall detection based on variations in silhouette area," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 427–436, Feb. 2013.

[8] R.-D. Wang, Y.-L. Zhang, L.-P. Dong, J.-W. Lu, Z.-Q. Zhang, and X. He, "Fall detection algorithm for the elderly based on human characteristic matrix and SVM," in *Proc. 15th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2015, pp. 1190–1195.

[9] M. Kepski and B. Kwolek, "Event-driven system for fall detection using body-worn accelerometer and depth sensor," *IET Comput. Vis.*, vol. 12, no. 1, pp. 48–58, 2017.
[10] M. Hagui, M. A. Mahjoub, and F. Elayeb, "Coupled hidden Markov model for video fall detection," in *Proc. 11th Int. Conf. Natural Comput. (ICNC)*, Aug. 2015, pp. 675–679.
[11] Y. Yun and I. Y.-H. Gu, "Human fall detection in videos by fusing statistical features of shape and motion dynamics on Riemannian manifolds," *Neurocomputing*, vol. 207, pp. 726–734, Sep. 2016.
[12] S.-G. Miaou, P.-H. Sung, and C.-Y. Huang, "A customized human fall detection system using omni-camera images and personal information," in *Proc. 1st Transdiscipl. Conf. Distrib. Diagnosis Home Healthcare (D2H2)*, Apr. 2006, pp. 39–42.
[13] Z.-P. Bian, J. Hou, L.-P. Chau, and N. Magnenat-Thalmann, "Fall detection based on body part tracking using a depth camera," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 430–439, Mar. 2015.
[14] A. Barriga, J. M. Conejero, J. Hernández, E. Jurado, E. Moguel, and F. Sánchez-Figueroa, "A vision-based approach for building telecare and telerehabilitation services," *Sensors*, vol. 16, no. 10, p. 1724, 2016.
[15] J. Shotton *et al.*, "Efficient human pose estimation from single depth images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2821–2840, Dec. 2013.
[16] Y. Hbali, S. Hbali, L. Ballihi, and M. Sadgal, "Skeleton-based human activity recognition for elderly monitoring systems," *IET Comput. Vis.*, vol. 12, no. 1, pp. 16–26, 2017.
[17] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 9–14.
[18] W. Ding, K. Liu, H. Chen, and F. Tang, "Human action recognition using similarity degree between postures and spectral learning," *IET Comput. Vis.*, vol. 12, no. 1, pp. 110–117, 2017.
[19] A. A. Trofimova, A. Masciadri, F. Veronese, and F. Salice, "Indoor human detection based on thermal array sensor data and adaptive background estimation," *J. Comput. Commun.*, vol. 5, no. 4, p. 16, 2017.
[20] J. M. Carmona and J. Climent, "Human action recognition by means of subtensor projections and dense trajectories," *Pattern Recognit.*, vol. 81, pp. 443–455, Sep. 2018.
[21] A. Senior and S. Tosunoglu, "Hybrid machine vision control," in *Proc. 18th Florida Conf. Recent Adv. Robot.*, Gainesville, FL, USA, 2005, pp. 1–6.
[22] H. Rajabi and M. Nahvi, "An intelligent video surveillance system for fall and anesthesia detection for elderly and patients," in *Proc. 2nd Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Mar. 2015, pp. 1–6.
[23] G. Debard *et al.*, "Camera-based fall detection on real world data," in *Outdoor and Large-Scale Real-World Scene Analysis*. Springer, 2012, pp. 356–375.
[24] K. G. Gunale and P. Mukherji, "Fall detection using k-nearest neighbor classification for patient monitoring," in *Proc. Int. Conf. Inf. Process. (ICIP)*, Dec. 2015, pp. 520–524.
[25] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: Comparison of support vector machine and Adaboost-based classification," *J. Electron. Imag.*, vol. 22, no. 4, p. 041106, 2013.
[26] C.-Y. Lin, S.-M. Wang, J.-W. Hong, L.-W. Kang, and C.-L. Huang, "Vision-based fall detection through shape features," in *Proc. IEEE 2nd Int. Conf. Multimedia Big Data (BigMM)*, Apr. 2016, pp. 237–240.
[27] M. Yu, Y. Yu, A. Rhuma, S. M. R. Naqvi, L. Wang, and J. A. Chambers, "An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 6, pp. 1002–1014, Nov. 2013.
[28] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274–1286, Nov. 2012.
[29] Y. Yun, C. Innocenti, G. Nero, H. Lindén, and I. Y.-H. Gu, "Fall detection in RGB-D videos for elderly care," in *Proc. 17th Int. Conf. E-health Netw., Appl. Services (HealthCom)*, 2015, pp. 422–427.
[30] K. Chaccour, R. Darazi, A. H. El Hajjam, and E. Andrès, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors J.*, vol. 17, no. 3, pp. 812–822, Feb. 2017.
[31] N. Doulamis, "Iterative motion estimation constrained by time and shape for detecting persons' falls," in *Proc. 3rd Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, 2010, p. 62.
[32] W. Min, H. Cui, H. Rao, Z. Li, and L. Yao, "Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics," *IEEE Access*, vol. 6, pp. 9324–9335, Jan. 2018.
[33] S. C. Sen-Ching and C. Kamath, "Robust techniques for background subtraction in urban traffic video," *Proc. SPIE*, vol. 5308, pp. 881–892, Jan. 2004.
[34] Z. Li, P. Li, X. Yu, and M. Hashem, "Real-time tracking by double templates matching based on timed motion history image with HSV feature," *Sci. World J.*, vol. 2014, Jan. 2014, Art. no. 793769.
[35] W. Feng, R. Liu, and M. Zhu, "Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera," *Signal, Image Video Process.*, vol. 8, no. 6, pp. 1129–1138, 2014.
[36] S. Albawendi, K. Appiah, H. Powell, and A. Lotfi, "Video based fall detection with enhanced motion history images," in *Proc. 9th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ. (PETRA)*, 2016, pp. 29:1–29:7.
[37] S. Albawendi, K. Appiah, H. Powell, and A. Lotfi, "Video based fall detection using features of motion, shape and histogram," in *Proc. 11th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ. (PETRA)*, New York, NY, USA, 2018, pp. 529–536.
[38] K. de Miguel, A. Brunete, M. Hernando, and E. Gambao, "Home camera-based fall detection system for the elderly," *Sensors*, vol. 17, no. 12, p. 2864, 2017.
[39] K. Wang, G. Cao, D. Meng, W. Chen, and W. Cao, "Automatic fall detection of human in video using combination of features," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1228–1233.
[40] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, "Motion history image: Its variants and applications," *Mach. Vis. Appl.*, vol. 23, no. 2, pp. 255–281, 2012.
[41] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Proc. 21st Int. Conf. Adv. Inf. Netw. Appl. Workshops (AINAW)*, vol. 2, May 2007, pp. 875–880.
[42] H. Foroughi, B. S. Aski, and H. R. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Proc. 11th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Jan. 2008, pp. 219–224.
[43] N. Worrakulpanit and P. Samanpiboon, "Human fall detection using standard deviation of C-motion method," *J. Autom. Control Eng.*, vol. 2, no. 4, pp. 388–391, 2014.
[44] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 489–501, Dec. 2014.

AHMAD LOTFI (M'96–SM'08) received the B.Sc. degree in control systems from the Isfahan University of Technology, Iran, and the M.Tech. degree in control systems from IIT, India, and the Ph.D. degree in learning fuzzy systems from The University of Queensland, Australia, in 1995. He is currently a Professor of computational intelligence at Nottingham Trent University, where he is leading the research group in computational intelligence and applications. His research focuses on the identification of progressive changes in behavior of elderly people suffering from Dementia or any other cognitive impairments. Accurate identification of progressive changes through utilization of unobtrusive sensor network or robotics platform will enable carers (formal and informal) to intervene when deemed necessary. Research collaboration is established with world-leading researchers. He has worked in collaboration with many healthcare commercial organizations and end-users, including Tunstall Healthcare Group and Nottingham Adult Care. He has received external funding from Innovate U.K., EU, and industrial companies to support his research. He has authored and co-authored over 100 scientific papers in the area of computational intelligence, abnormal behavior recognition, and ambient intelligence in highly prestigious journals and international conferences. He has been invited as an Expert Evaluator and a Panel Member for many EU Framework Research Programmes, including FP6/FP7 ERC and Horizon 2020.

SUAD ALBAWENDI received the M.Sc. degree from Nottingham Trent University in 2010, where she is currently pursuing the Ph.D. degree with the School of Science and Technology. She is working on her Ph.D. thesis "Automated Behavioural Recognition From Visual Data". Her current research interests include image and video processing, computer vision, and pattern recognition. As part of her research she is trying to identify falls for older adults living alone in a home environment. Her research focuses mainly on using low cost visual sensors.

HEATHER POWELL graduated (Hons.) in electronic engineering from Sheffield University in 1982. She undertook work for her Ph.D. in impedance imaging for medical application within the Royal Hallamshire Hospital, Sheffield, gaining the award from the University of Sheffield in 1989. She then spent five years in industry developing computer control systems and networking software. He has worked in academia for 28 years with experience as the Head of the Computer Science Department before roles as Acting Associate/Deputy Dean and the Head of STEM developments within the School of Science and Technology. She has supervised/co-supervised nine Ph.D. students to completion and has authored/co-authored 38 scientific papers. Her research interests have covered artificial neural networks for natural language processing, vision and information extraction, intelligent tutoring/knowledge provision systems, and interactive systems for social inclusion. She is a Chartered Engineer, a member of the BCS, and a Senior Fellow of the Higher Education Academy, U.K.

KOFI APPIAH (M'04) received the B.Sc. degree in computer science from the University of Science and Technology, Ghana, the M.Sc. degree in electrical and electronic engineering from the Royal Institute of Technology, Stockholm, Sweden, the M.Sc. degree in computer science from the University of Oxford, U.K., and the Ph.D. degree from the University of Lincoln, U.K. He is currently a Senior Lecturer at Sheffield Hallam University after working as a Lecturer/Senior Lecturer at Nottingham Trent University from 2013 to 2017. His current research interests include the use of computer vision in support of independent living, bio-inspired computer vision on embedded architectures, and highly parallel software and hardware architectures.

CAROLINE LANGENSIEPEN is a Chartered Engineer and has had substantial industrial software experience prior to becoming an academic. She has designed and implemented many systems that process large-scale physical data sets, both in real time and non-real time settings. She has been a software design authority on such systems and has experience in taking projects (both large scale and small scale) from initial requirements analysis through to delivery. This experience has since been enhanced by academic research and consultancy in the area of software development, and research in data interpretation/ numerical analysis from physical systems. She has published more than 50 papers, undertaken consultancy with local companies to improve their software analysis and design skills and procedures, and is an active researcher in various areas related to real-time monitoring and control.