

Article

Image Processing Technique and Hidden Markov Model for an Elderly Care Monitoring System

Swe Nwe Nwe Htun ^{1,*}, Thi Thi Zin ² and Pyke Tin ³

¹ Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, Miyazaki 889-2192, Japan

² Graduate School of Engineering, University of Miyazaki, Miyazaki 889-2192, Japan; thithi@cc.miyazaki-u.ac.jp

³ International Relation Center, University of Miyazaki, Miyazaki 889-2192, Japan; pyketin11@gmail.com

* Correspondence: nc17010@student.miyazaki-u.ac.jp

Received: 19 May 2020; Accepted: 10 June 2020; Published: 13 June 2020



Abstract: Advances in image processing technologies have provided more precise views in medical and health care management systems. Among many other topics, this paper focuses on several aspects of video-based monitoring systems for elderly people living independently. Major concerns are patients with chronic diseases and adults with a decline in physical fitness, as well as falling among elderly people, which is a source of life-threatening injuries and a leading cause of death. Therefore, in this paper, we propose a video-vision-based monitoring system using image processing technology and a Hidden Markov Model for differentiating falls from normal states for people. Specifically, the proposed system is composed of four modules: (1) object detection; (2) feature extraction; (3) analysis for differentiating normal states from falls; and (4) a decision-making process using a Hidden Markov Model for sequential states of abnormal and normal. In the object detection module, background and foreground segmentation is performed by applying the Mixture of Gaussians model, and graph cut is applied for foreground refinement. In the feature extraction module, the postures and positions of detected objects are estimated by applying the hybrid features of the virtual grounding point, inclusive of its related area and the aspect ratio of the object. In the analysis module, for differentiating normal, abnormal, or falling states, statistical computations called the moving average and modified difference are conducted, both of which are employed to estimate the points and periods of falls. Then, the local maximum or local minimum and the half width value are determined in the observed modified difference to more precisely estimate the period of a falling state. Finally, the decision-making process is conducted by developing a Hidden Markov Model. The experimental results used the Le2i fall detection dataset, and showed that our proposed system is robust and reliable and has a high detection rate.

Keywords: normal and abnormal states; fall detection; Mixture of Gaussians; graph cut; virtual grounding point; Hidden Markov Model

1. Introduction

The risk of falling is a huge issue, not only for older adults in aging societies, but also for younger adults with drug and alcohol problems and for patients with chronic diseases. Falling has serious health consequences and is a leading cause of death. According to a survey by the World Health Organization [1], falling occurs most frequently in the 28–35% of individuals between the ages of 65 and 70 and in the 42% who are over 70 years old. Falling is particularly dangerous for persons who live alone in an indoor environment, because much time can pass before they receive assistance. In this

situation, many countries are adopting policies to increase the life expectancy, by providing extra care to people living independently.

For this reason, much research has focused on developing a robust fall-detection process in smart home systems using their respective specialized technologies. Progress in developing such intelligent technologies holds the promise of improving the quality of life for the aged and infirm. Popular fall detection and prediction systems have emerged based on wearable monitoring devices, ambient devices, vision-based devices, and portable devices. The most commonly used wearable devices feature accelerometers and gyroscopes embedded in belts, watches, or pendants that are reasonably comfortable to wear. However, some of these devices have disadvantages that limit their usability, such as excessive power consumption. Some of these devices produce false alarms when triggered by normal body movements, and some require the manual activation of an alarm after a fall. In addition, the elderly often forgets to wear the devices. However, wearable devices based on machine learning and intelligent systems are effective for monitoring people in both indoor and outdoor environments. Various interesting approaches for fall detection systems using wearable devices are discussed in [2–5], in which accelerometers and gyroscopes are employed to collect data on the emotional state and body movements of subjects.

The processing of data collected using ambient devices provides information without demanding user intervention. Commercially available devices using advanced technologies include presence sensors, motion sensors, and vibration sensors, which can be embedded in an indoor environment, such as on furniture. Presence sensors can detect the tiniest movements of subjects, such as the movement of a finger, with a high resolution and precision. They can easily be set on high ceilings, as well as the floor, to detect the smallest movement. Compared to presence sensors, motion sensors are useful in perceiving the arm movements involved in walking in detection zones. Such zones are selected as high-traffic areas for detecting moving objects in busy indoor or outdoor areas. Vibration sensors are useful for detecting falling events, and can distinguish the activities through vibrations. Vibration sensors installed on floors or beds acquire vibration data that can be analyzed by determining the fall detection ratio [6]. Although such sensors do not disturb the people involved, they can generate false alarms.

In recent years, the development of smart phone technologies that are incorporated with accelerometers has provided some very interesting approaches for healthcare monitoring systems, including fall detection. Mobile phone-based fall detection algorithms have been presented in [7–10]. However, these devices are not useful if the monitored subject does not always have them in their hand.

For these reasons, video monitoring systems based on computer vision and machine learning are potentially more beneficial and reliable for fall detection. Vision-based devices utilizing cameras have some of the same limitations as ambient devices, such as the fact that the devices must be installed in several places to provide full coverage of the required areas over a long period of time. However, video surveillance systems can effectively predict specific human activities, such as walking, sitting down, going to bed, and getting up from bed [11,12], as well as detecting fall events [13]. Moreover, a large amount of visual information is captured in the video record of falling events. The various definitions of falling events, as well as the reasons and circumstances behind such events and other abnormal occurrences in real-world environments, should be visually analyzed in monitoring systems. The vision-based detection of abnormal or fall events could become an important tool in elderly care systems.

In developing better vision-based video monitoring systems, abnormal event detection holds great promise, but faces many challenges in real-world environments. In developing applications for this purpose, establishing a fall detection system presents difficulties because the dynamic conditions involved in falls are not well-understood. As an important focus, researchers must investigate the universal features common to all falling events. Therefore, we must differentiate falls from normal activities in a comprehensive way that reflects the scene of falling scenarios. Any system developed must be robust in the face of changing postures and positions, such as a loss of balance, and abrupt changes in direction. These conditions must be considered in reliably assessing abnormal events.

However, research trends and current best practices indicate that the prospects are great for using vision-based monitoring to improve the quality of health care.

Therefore, we propose a vision-based system for fall detection by differentiating abnormal behavior or falls from normal states. This system includes simple and uncomplicated feature extraction or effective statistical analyses. Key to detecting abnormal or falling events is the recognition that they involve a loss of balance. The three research objectives of this proposed system are as follows:

- Develop a monitoring system that provides a visual understanding of a person's situation and can judge whether the state is abnormal or normal based on video data acquired using a simple and affordable RGB camera;
- Develop an individualized and modified statistical analysis on each of the extracted features, providing trustworthy information, not only on the definite moment of a fall, but also on the period of a fall;
- Develop an efficient way of using a Hidden Markov Model (*HMM*) for the detailed detection of sequential abnormal and normal states for the person being monitored.

Our proposed system intends to establish a long-term monitoring system for facilitating independent living. Major steps in the system include (1) feature extractions in estimating positions and postures of the person by utilizing the virtual grounding point (*VGP*) concept, and related visual features inspired by our previous research [13]; (2) modified statistical analysis to estimate the time interval or period for falls and normal states through extensive feature observation; and (3) the establishment of a Hidden Markov Model (*HMM*) to detect the sequential normal and fall states of the person. The rest of the paper is organized as follows: Section 2 presents related works; Section 3 presents theoretical analysis and methodologies of the proposed system; Section 4 presents and evaluates experimental results, comparing the robustness and limitations of the proposed systems with existing algorithms; and finally, Section 5 presents the conclusions of this work.

2. Related Works

Related state-of-the-art fall detection systems will be discussed in this section. In order to effectively define the postures assumed by a human object, it is very important to perform feature extraction or selection, analyze the selected features, and set detection rules in a vision-based video monitoring system, especially for monitoring in health care. The most common feature extraction methods used in fall detection systems involve the human shape and motion-history images. The fall detection system presented in [14] is a method based on motion history images (*MHI*) and changes in the human shape. In this system, information on the history of motion occurring over a fixed interval can be obtained from *MHI*. Then, the human shape is obtained by constructing a blob using an approximate ellipse. Finally, fall detection is achieved by considering three factors: motion quantification, analysis of the human shape, and the lack of motion after a fall. Motion quantification allows sudden motion to be detected when a person falls. The approximated ellipse constructed on the object can provide information about changes in the human shape, more precisely, changes in orientation. The final analysis provides a moving ellipse, which indicates whether the person is moving after a fall. The decision confirming a fall is made when the ellipse stops moving for five seconds after a fall. The video sequences are captured using wall-mounted cameras to cover wide areas, and the results are presented in 2D motion and shape information. Extended work on the system presented in [15] involves both 2D and 3D information for fall detection, as the researchers intended to recover localization information on the person relative to the ground. This extended process of feature extraction involves computing the 3D head trajectories of a person, and a fall is detected if the velocity of the head exceeds a certain value and the position of the head is too close to the floor or the ground.

Similarly, a process of detecting unnatural falls has been developed [16], which includes background subtraction using the frame difference method, feature extraction using *MHI*, and a change in human shape and classification using a support vector machine (*SVM*). Compared with

the previous approach [15], the system involves constructing three specific features on the human shape, namely, the orientation of the approximated ellipse, its aspect ratio, and silhouette detection below a threshold line. The aspect ratio of the ellipse describes changes in major and minor axes, and differentiates a fall from normal activities. After a fall, the previously moving object lies on the ground. For this reason, the threshold line is set by considering a suitable height from the ground. After that, fall and non-fall objects can be differentiated according to the height of the silhouette object. Finally, classifications are performed using the support vector machine (SVM), k-nearest neighbor (KNN) classifier, Stochastic Gradient Descent (SGD), Decision Tree (DT), and Gradient Boosting (GB). The DT algorithm consistently outperforms the rest, with a high detection rate, as confirmed using the Le2i fall detection dataset.

Another fall detection system for elderly care is proposed in [17]. This system firstly conducts background subtraction to segment out the silhouette moving object, and then tracks the object to determine the trajectory. Secondly, a timed motion history image (*tMHI*) is constructed to detect high velocities. Then, the motion is quantified to acquire pixel values for the *tMHI*, which are divided by the number of pixels in the detected silhouette object. Similar to the approach presented in [16], this system provides a useful definition of the human body posture using the following combined features as the input: the ratio, orientation, and major and minor semi-axes of the fitted ellipse. In this system, both *MHI* and a projection histogram are applied to confirm that a falling event has occurred. In addition, the position of the head can be tracked in sequential frames in order to obtain useful information, since the trajectory of the head is visible most of the time. Finally, a multilayer perceptron (*MLP*) neural network is applied on the extracted features to classify falls and non-falls, with an accuracy of 99.24%, and a precision of 99.60%, as confirmed using the UR fall detection dataset. In addition, an extensive automated human fall-recognition process is proposed in [18] to support independent living for the elderly in indoor environments using the Le2i fall-detection dataset. This system is based on motion, orientation, and histogram features, and achieves an overall accuracy of 99.82%. The approaches in [14–18] discussed above focus on distinguishing falls from normal activities.

However, any monitoring system should take into account consecutive daily activities in a real-world environment. Such activities include walking, standing, and sitting, as well as transitioning between these activities. In this regard, our previous work [19] proposed human action analysis based on motion history and the orientation of the human shape. As its main purpose, this system takes into consideration a prediction of the degree of mobility for the elderly in daily activities, such as getting into and out of bed. These activities include the following consecutive actions: sitting, transitioning from sitting to lying, lying, transitioning from lying to sitting, transitioning from sitting to standing, and walking. We firstly conducted background subtraction [20,21] for the proper separation of foreground and background objects. Among the features used in this system are *tMHI* and the orientation of the approximated ellipse. However, the construction of one ellipse is not enough for detecting the human object region. As a key part of our design, two approximated ellipses are constructed using horizontal and vertical histogram values. The vertical histogram is employed for the whole body region, and the horizontal histogram is employed for the upper body region. Then, motion quantification is used to analyze these human activities. In considering detailed sequential actions, multiple threshold values are observed, which depend on the shape orientation and the coefficient of motion. Rather than using fall scenes, the experiments were conducted using the normal scenes in the 14 videos of the Le2i dataset. After that, we extended our system [22] to analyze not only normal activities, but also falls. In this system, the virtual grounding point (*VGP*) is introduced for feature extraction, and analyses are performed by the combined features of *tMHI* and *VGP*. The overall accuracy of the proposed system for the detection of falling events was 82.18%, as experimentally confirmed using 15 videos from the Le2i fall detection dataset. To improve the accuracy of detecting falls in a given video sequence, we extended our research [13] by incorporating feature selection using the *VGP* concept with its related features and statistical analysis for estimating the falling period, as well as two classification methods, namely the support vector machine (SVM) and period detection

(PD). A comparison with existing approaches using the Le2i dataset showed that our SVM approach outperforms the rest, with a precision of 93%, a recall of 100%, and an accuracy of 100%. However, our previous system exclusively concentrated on differentiating falling events from normal events; in other words, classifying the videos which contain abnormal or normal events. A long-term monitoring system for home care requires the effective detection of sequential normal and abnormal states.

Another vision-based fall detection system using a convolutional neural network has been developed by Adrian [23]. This system was designed to work on human motion, avoiding any dependence on the appearance of an image. An optical flow image generator is utilized to efficiently represent human motion. A set of optical flow images are stacked and then set as inputs in a convolutional neural network (CNN) that can learn longer time-related features. A fully connected neural network (FCNN) receives these features as inputs, and produces fall and non-fall outputs. The best overall accuracy for this system is 97%, as confirmed using the Le2i fall-detection dataset. Moreover, the approach used in [24] proposes fall detection based on body keypoints and sequence-to-sequence architecture. In this system, a skeleton framework is modeled to receive a sequence of observed frames. Then, the coordinates of keypoints of the object are extracted from observed frames. The bounding boxes of the detected object are given to the tracking algorithm for clustering body keypoints belonging to the same person in different video sequences. A keypoint-vectorization method is exploited to extract salient features from associated coordinates. Next, the pose prediction phase is conducted, predicting the vectors of future keypoints for the person. Finally, falls are classified using the Le2i dataset, achieving an accuracy of 97%, a precision of 90.8%, a recall of 98.3%, and an F1-score of 0.944.

A related approach is presented in [25] in an automatic fall-detection system with an RGB-D camera using a Hidden Markov Model (HMM). Background subtraction is firstly performed by averaging the depth map to learn the background. The Kinect optical parameters for factory use are employed to obtain a real-world coordinate system, and an Open Natural Interaction (OpenNI) is used for a Kinect-to-real-world transformation. After that, the center of mass of the person is extracted to calculate the vertical speed from that point. The standard deviation for all points belonging to a person are then calculated. After that, these three features are used as inputs to calculate the probability of HMM. Finally, the forward-backward and Viterbi algorithms are applied to classify the states of normal activities and falls. The experiments were conducted on young and healthy subjects, and occlusions were not included. In future research, this system will be tested using real-life unhealthy subjects with occlusions. In addition, other fall detection systems for shop floor applications [26] were modeled using an HMM based on the vertical velocity, area variance, and height of a person, using cameras positioned to provide a top view. This incident detection method focuses on two things: detecting people in restricted areas and detecting falls. Potential fall events are analyzed based on specific features and on circumstances such as whether the person can get up or not. Analysis is also based on an allocation of status according to the location, whether the event occurs in an area that is restricted to all personnel, where work is ongoing, or where maintenance is being performed. Such considerations are valuable in identifying health and safety issues. In future work, this system will be improved by incorporating more incidents, such as collisions.

Vision-based monitoring systems for detecting falls or abnormal events can be powerful tools in various applications. These new technologies have great potential as intelligent monitoring systems. Therefore, we propose a detection system for indoor environments that include normal activities; one that is based on the consecutive states of abnormal or fall events using image processing techniques and a Hidden Markov Model. This detection system will be ideal for elderly care monitoring systems.

3. Proposed Architecture of the System

The proposed monitoring system provides a way to detect normal, abnormal, and falling states for persons being monitored, such as seniors with a limited mobility, patients with chronic diseases, and those recovering from surgical procedures. The proposed system is composed of the following four main modules: (1) object detection; (2) feature extraction using the virtual grounding point (VGP)

concept; (3) analysis of normal, abnormal, or falling events; and (4) establishment of decision-making rules utilizing a Hidden Markov Model (*HMM*). The major work flow for our proposed system is illustrated in Figure 1, and specialized theoretical analyses for each module are described in the following sections.

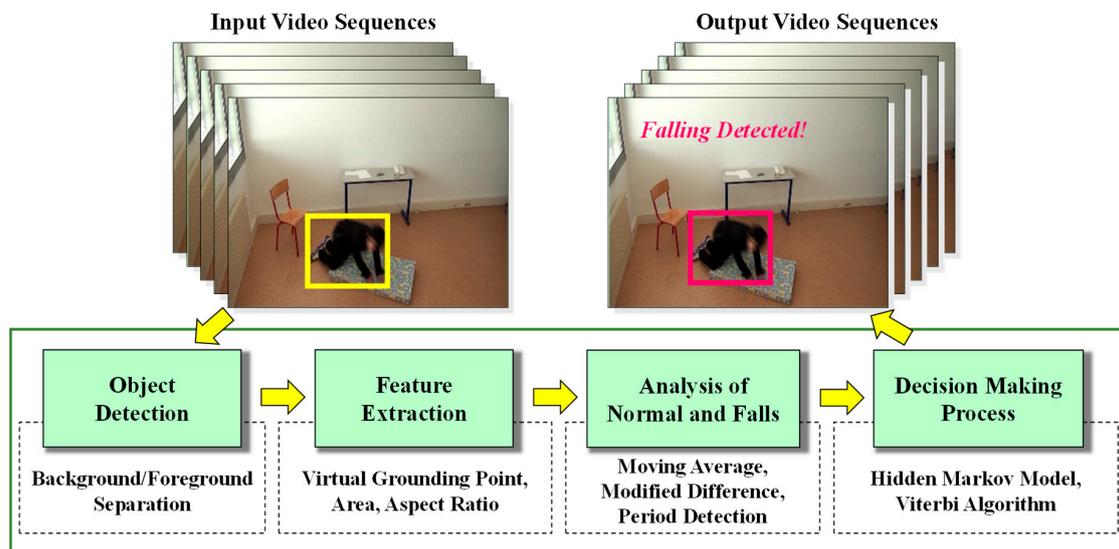


Figure 1. The major work flow of our proposed system.

3.1. Object Detection

This module features an object detection procedure similar to that used in our previous work [13], and is briefly described here. We select a specific Mixture of Gaussians (*MoG*) distribution [20] for modeling the foreground, which is updated frame by frame. We also use specific low-rank subspace learning for modeling the background for each successive frame in the video sequences. After that, an expectation and maximization (*EM*) algorithm is used for updating the foreground and background parameters of each new frame. However, the quality of the foregrounds is unsatisfactory because ghost effects are included in the resultant foregrounds. In real-life video sequences, as well as in simulated video sequences, much redundant data occurs, such as when a person moves very slowly or stays in place for a long period of time. In such situations, most systems have trouble recognizing a person as the foreground. To overcome this problem, a graph cut algorithm is applied to refine the foreground [21]. A set of the resulting foreground and background frames is given as inputs for the video sequences. Then, we seek binary labels regarded as the foreground, which is set to 1, and the background, which is set to 0. These foreground and background labels are calculated by constructing a graph structure, $G = (v_{vertex}, \epsilon)$, where v_{vertex} is the set of vertices or pixels and ϵ is the set of edges linking the closest four-connected pixels [21]. The maximum-flow and minimum-cut method is used to find the vertex label with a minimum energy function. Most traditional graph cut algorithms work on manual drawings of the scribbles and region of interest for the targeted object in every frame. In applying graph cut theory for refining the foreground, we obtain the associated *MoG* foreground mask in every 100th frame, instead of repeatedly re-drawing the scribbles and region of interest. The combination of these two effective methods optimizes the solution for object detection.

3.2. Feature Extraction

In this module, feature extraction is performed by applying the virtual grounding point (*VGP*), as well as the area and aspect ratio of the object [13]. The computation strategies for feature extraction are simple and effective. Using these strategies, we adapted our previously proposed method of feature extraction [13] in this extension of our work. Here, we briefly describe our previous research using the

VGP concept. A virtual grounding point is a point which is constructed for exploring the position and posture of humans. Specialized technical details are described as follows.

Let p be the position at time t on the detected silhouette object. Then, the centroid of object $C(t)$ is extracted as defined in Equation (1):

$$C(t) = (x_c(t), y_c(t)), \tag{1}$$

where t represents $p(t) = (x(t), y(t))$, and x_c and y_c denote that $x_c = \sum_{i=1}^N (x_i/N)$ and $y_c = \sum_{i=1}^N (y_i/N)$ respectively.

Then, a vertical line from the top-most to the bottom-most row corresponding to the x axis of $C(t)$ is constructed on the detected object. Similarly, a horizontal line from the left to the right column corresponding to the y axis of $C(t)$ is constructed. Finally, a central point where the vertical line along the x axis crosses the horizontal line along the y axis is marked as a VGP, which can be defined as in Equation (2):

$$VGP(t) = (x_{VGP}(t), y_{VGP}(t)), \tag{2}$$

where the VGP is a virtual grounding point at time t .

Then, the point distance (d) between the centroid $C(t)$ and its corresponding virtual grounding point $VGP(t)$ is formulated as shown in Equation (3):

$$d(t) = |y_{VGP}(t) - y_c(t)|, \tag{3}$$

where $d(t)$ can be denoted as the point distance between centroid $C(t)$ and its corresponding virtual grounding point $VGP(t)$ on the detected object at time t to extract the features for the position and posture of the person.

After that, the area of the object shape (a) can be derived to obtain knowledge for the shape regularity of the moving object, as described in Equation (4):

$$a(t) = \left(\sum_{i=1}^N \sum_{j=1}^N I(i, j), t \right), \tag{4}$$

where $a(t)$ is the measurement of the area indicating the relative size of the object obtained by summing all the pixels in the object at time t .

Finally, the aspect ratio (r) with respect to the object is simply calculated to estimate the human posture as shown in Equation (5):

$$r(t) = w(t)/h(t), \tag{5}$$

where $r(t)$ is the aspect ratio of the object found by dividing the value of the width (w) by that of the height (h) for the object at time t . The concepts for deriving the individualized features are illustrated in Figure 2.

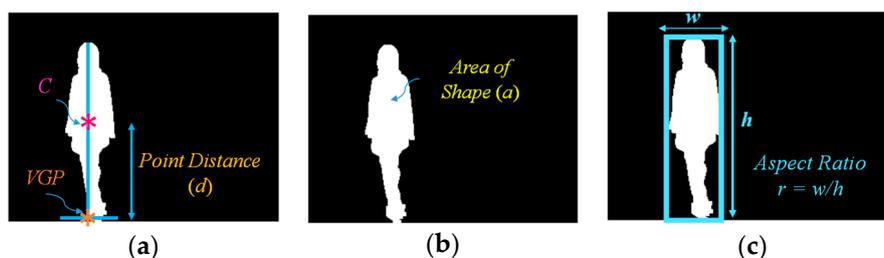


Figure 2. Feature extraction: (a) Point distance (d); (b) area of the object (a); (c) aspect ratio of the object (r).

3.3. Analysis of Abnormal and Normal Events

For detecting abnormal or falling events, we first computed the moving average (MA) [13] using features extracted from sequential data. In our previous research [13], we used two features (point distance (d) and aspect ratio (r)) for analyzing the data points of the moving object. In this extension of our work, statistical analysis of the area of the object shape on the observable data series is used with one additional feature. As a result, our proposed formulation uses the following three features: (1) the point distance (d) between the centroid and the corresponding virtual grounding point; (2) the area of the object shape (a); and (3) the aspect ratio of the object (r), as shown in Equation (6):

$$MA(t, F, N) = \frac{1}{2N + 1} \sum_{f=t-N}^{f=t+N} F(t), F(t) = d(t), a(t), r(t), \quad (6)$$

where $MA(t, F, N)$ is an output value for the average period in N at time t , N represents the number of periods, and $F(t)$ is the calculation on the three extracted features: the point distance (d), the area of the object shape (a), and the aspect ratio (r) at time t . The optimized threshold (Th) value is considered in detecting abnormal or falling events. Here, we set the threshold value of N at $Th(MA(t, d, Th))$, $Th(MA(t, a, Th))$, and $Th(MA(t, r, Th))$, where $Th = 51$, which depends on the video frame rate. The process for selecting the optimal threshold is similar to that used in our observations in previous research [13]. Calculating the moving average for the sequential data approximates irregular and falling events by observing the crossing point on primary and average feature points.

After that, we define the difference calculation, specifically the modified difference (MD), for more precisely estimating abnormal and falling states on the crossing point of the moving average (MA), as shown in Equation (7):

$$MD(t, F, N_0, N_1) = MA(t, F(t + N_0 + N_1), N_1) - MA(t, F(t - N_0 - N_1), N_1), F(t) = d(t), a(t), r(t), \quad (7)$$

where $MD(t, F, N_0, N_1)$ denotes the modified difference for the three selected features (d , a , and r) at time t and here, the selected optimal value of the threshold for $N_0 = 0$ and $N_1 = 51$.

In analyzing a series of data based on the point distance (d), the modified difference (MD) reached a minimum point, which indicates a high possibility of a fall. The modified difference observed for the aspect ratio (r) of the person indicates a high possibility of a fall with a maximum point. Here, we say that the highest point is the local maximum (l_{max}) and the lowest point is the local minimum (l_{min}). The concepts of discovering a point which is important in referring an abnormal state or a fall by utilizing r and d are illustrated in Figures 3 and 4, respectively.

When we observe stationary points for the modified difference on the area of the object shape (a), we make an interesting discovery that both l_{max} and l_{min} occur with respect to the falling posture and direction of the object. The observations for l_{max} and l_{min} which correspond to the different falling postures and positions (sideways falls, forward falls, and backward falls) are described in Figures 5–7, respectively. This observation and analysis of the area of the object shape is part of the extension of our previous work [13], and can be properly used in further observations to develop decision-making rules.

Then, we analyze the interval or period of the abnormal event by computing the half width value (v_{hw}) [13] on the observed curve of modified difference (MD), as formulated in Equation (8). The half width value (v_{hw}) more accurately estimates the period of abnormal and normal events. At the halfway point on the largest curve obtained from the computation of MD , the starting point (f_1) and ending point (f_2) are set to represent irregular fall events. After that, the periods of events are observed by calculating the distance of the two points f_1 and f_2 .

$$v_{hw} = |f_1 - f_2|, \quad (8)$$

where v_{hw} is the estimated period for falling assumed by the half of the curve including l_{max} or l_{min} , and f_1 and f_2 represent the starting and ending points of a fall, respectively. Illustrations for developing a better understanding of v_{hw} are embedded in Figures 3–7.

In Figure 3, the analysis for a high possibility of abnormal (falling) points is illustrated based on the aspect ratio (r) with respect to the crossing point of the moving average (MA), local maximum (l_{max}), and half width value (v_{hw}) obtained from the distance between f_1 and f_2 . Figure 4 describes a high possibility of abnormal points based on the point distance (d) with respect to the crossing point of the moving average (MA), local minimum (l_{min}), and half width value (v_{hw}) obtained from the distance between f_1 and f_2 . By observing the area of the object shape (a), we noticed that the local minimum (l_{min}) occurs in the posture involved in falling sideways and backwards, as demonstrated in Figures 5 and 6, respectively. In Figure 7, the local maximum (l_{max}) occurs for the posture involved in falling forward when analyzing a high possibility of falling points based on the area of the object shape.

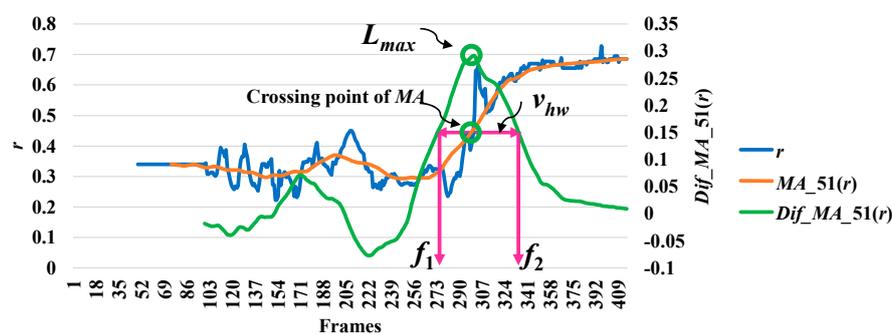


Figure 3. Analysis for a high possibility of abnormal (falling) points based on the aspect ratio (r).

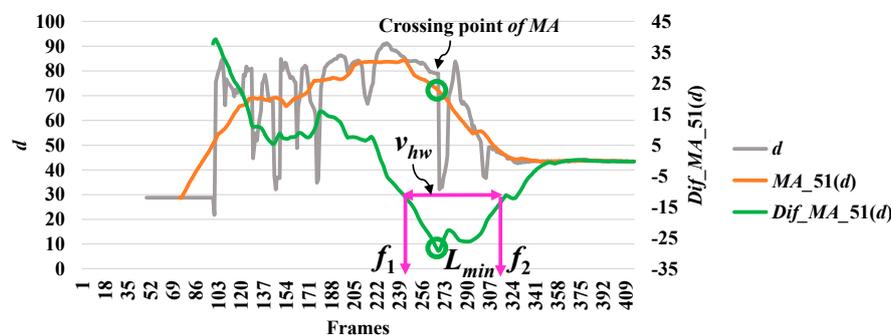


Figure 4. Analysis for a high possibility of abnormal (falling) points based on the point distance (d).

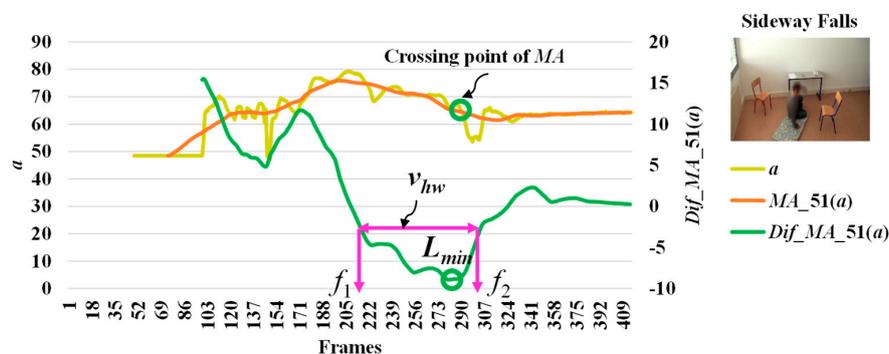


Figure 5. Analysis for a high possibility of abnormal (falling) points based on the area of the object shape (a) by falling sideways.

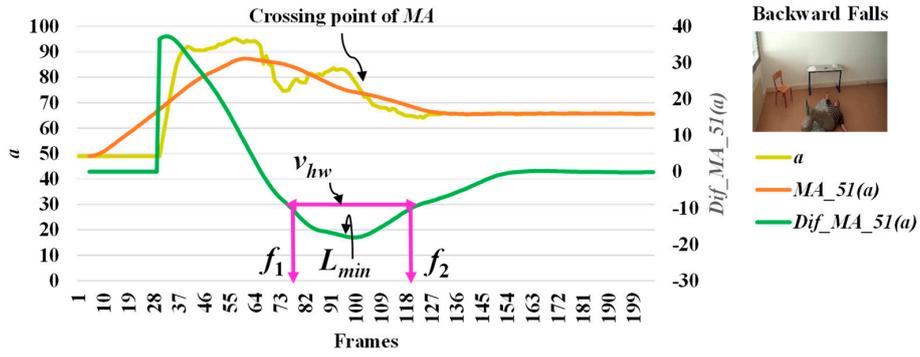


Figure 6. Analysis for a high possibility of abnormal (falling) points based on the area of the object shape (a) by falling backwards.

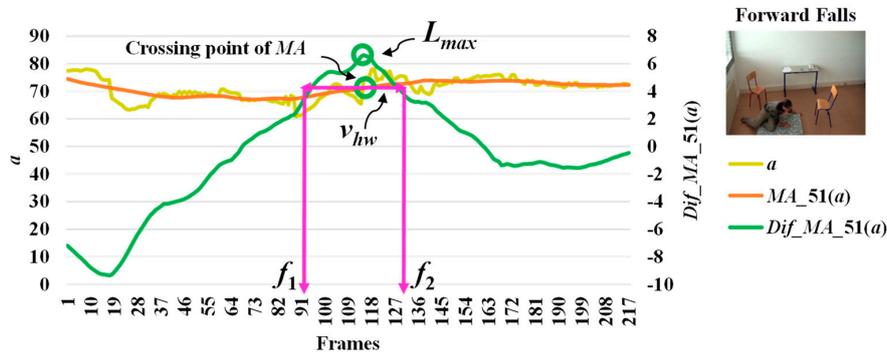


Figure 7. Analysis for a high possibility of abnormal (falling) points based on the area of the object shape (a) by falling forwards.

Finally, we consider the threshold values of the time intervals for abnormal and normal events based on the half width value (v_{hw}) in a given video sequence. Let $A = \{a_1, a_2, \dots, a_k\}$ represent the videos that include the abnormal events, and $N = \{n_1, n_2, \dots, n_k\}$ represent the videos composed of the normal events. After that, we calculate

$$\alpha_1 = \min(v_{hw}) \in A, \alpha_2 = \max(v_{hw}) \in N, . \tag{9}$$

Then, the mid value or falling period detection (PD) is calculated for three features, namely, the point distance (d) between the centroid and its corresponding virtual grounding point of the object, the area of the object shape (a) as an extensive feature observation, and the aspect ratio (r), which are formulated as shown in Equations (10)–(12), respectively.

$$PD(d) = \frac{(\alpha_1(d) + \alpha_2(d))}{2}, \begin{cases} l_1 & \text{if } v_{hw} \geq PD(d) \\ l_2 & \text{otherwise} \end{cases}, \tag{10}$$

$$PD(a) = \frac{(\alpha_1(a) + \alpha_2(a))}{2}, \begin{cases} l_1 & \text{if } v_{hw} \geq PD(a) \\ l_2 & \text{otherwise} \end{cases}, \tag{11}$$

$$PD(r) = \frac{(\alpha_1(r) + \alpha_2(r))}{2}, \begin{cases} l_1 & \text{if } v_{hw} \geq PD(r) \\ l_2 & \text{otherwise} \end{cases}, \tag{12}$$

where α_1 and α_2 represent the minimum and maximum threshold values used to estimate the period of abnormal and normal events, respectively; l_1 represents the class label for the given video that includes an abnormal event; and l_2 represents the class label which indicates that the given video does not include an abnormal event.

3.4. Decision-Making Rules

As the main contributions of this research, we developed considerations for decision-making rules. Abnormal and normal actions for the person being monitored will be detected for every detailed moment by employing a Hidden Markov Model (HMM).

In order to describe Hidden Markov Chains, we define *two states*: $S = \{S_1, S_2\}$. Let S_1 be an abnormal state which includes “falling,” and let S_2 be a normal state, including a set of actions, {“falling,” “lying,” “lying to sitting,” “sitting to standing,” “walking,” “walking to sitting,” “sitting,” and “sitting to lying”}. The process would be started in one of these two states, and proceed smoothly from one state to another. If the chain were presently in state S_1 , then it would proceed to the next state S_2 with a specific probability. However, it should be noted that this probability will not be based on which state the chain was in before the current state.

In order to form the Markov Transition matrix, we firstly observe each of the given videos by developing fixed intervals. The video sequence will become a sequence of intervals, such as $t, t + d, t + 2d, t + 3d, \dots, t + nd$. Here, d is the length of the interval to be predefined, in which only one abnormal or normal state occurs. The video sequences are manually observed to obtain the occurrence state symbol for every interval. These video sequences provide the sequences of the occurrences of the states, $S = \{S_1..S_2\}$, as illustrated in Figure 8. In this figure, S_1 represents the abnormal state and S_2 represents the normal state.

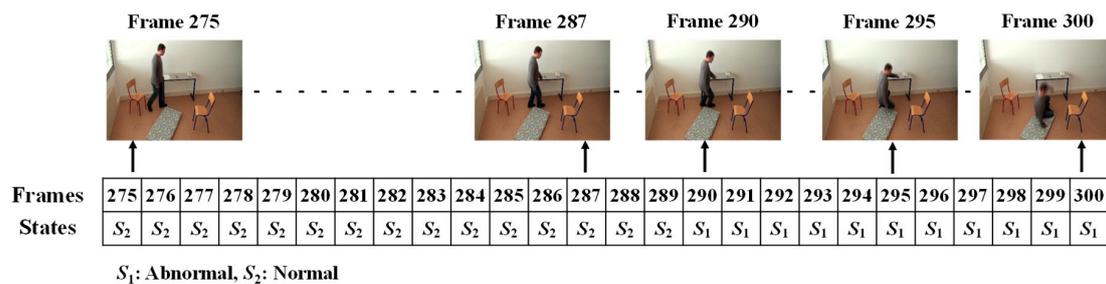


Figure 8. Observed sequence of the occurrences of states.

Then, the co-occurrence matrix M is constructed as shown below:

$$M = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{13}$$

where C_{11} represents the number of pairs (S_1, S_1) , C_{12} represents the number of pairs (S_1, S_2) , C_{21} represents the number of pairs (S_2, S_1) , and C_{22} represents the number of pairs (S_2, S_2) . After that, each row of matrix M is summed as shown in the following:

$$C_1 = \sum (C_{11} + C_{12}), C_2 = \sum (C_{21} + C_{22}), \tag{14}$$

To form the Markov transition matrix, we defined the following:

$$a_{11} = \frac{C_{11}}{C_1}, a_{12} = \frac{C_{12}}{C_1}, a_{21} = \frac{C_{21}}{C_2}, a_{22} = \frac{C_{22}}{C_2}, \tag{15}$$

Then, the state transition matrix of the Markov Chain for the Hidden Markov Model (HMM) is obtained as shown in Equation (16):

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{16}$$

where the elements in the first row of matrix A represent the probabilities for the actions following an abnormal action or fall. Similarly, the elements in the second row represent the probabilities for the actions following normal actions.

For each of the possible states, emission probabilities are calculated. We define the observable features with respect to the distance between the centroid and the corresponding virtual grounding point (d), the area of the object shape (a), and the aspect ratio (r). To define the observable symbols, a suitable threshold value for each of the features is selected. In doing so, we firstly calculate the features for extracting the falling period for a person, as discussed in Section 3.3. These falling periods can determine that each of the observed videos includes an abnormal or normal event. The frame index numbers through the video sequences represent that a fall can be extracted between starting point f_1 and ending point f_2 , as illustrated in Figure 9.

In the upper part of Figure 9, the frame numbers marked in green represent the normal activities of a person, such as walking, standing, sitting, and lying. The frame numbers marked by black represent abnormal or falling events. The period of a fall might be obtained when the half width value (h_{wv}), referring to the distance between a starting point (f_1) and an ending point (f_2), is greater than or equal to period detection (PD), as formulated in Section 3.3. Similarly, in the lower part of Figure 9, the frame numbers represent the normal activities of a person. When the person is involved in activities such as walking, walking to sitting, and sitting to standing, the analyzed half width value (h_{wv}) is less than the detected period.

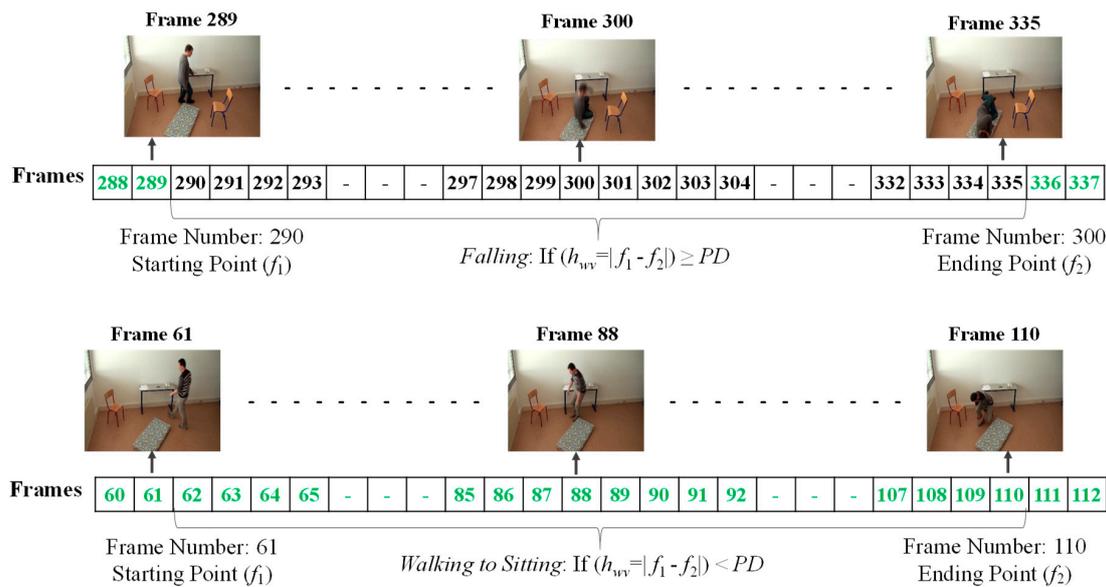


Figure 9. Observations in the period of abnormal and normal states. Upper part: falling; lower part: walking to sitting.

After that, it is determined whether the action is abnormal or normal by observing the six possible feature values defined below. Satisfying the criteria for this determination involves passing the selected value for threshold F in the consecutive frames from starting to ending points for abnormal and normal events based on the observed half width value, which determines the period of events.

$$\begin{aligned}
 F_1 &= \{f_1 \dots f_2\} \text{ where } v_{hw} \geq PD(d), & F_2 &= \{f_1 \dots f_2\} \text{ where } v_{hw} < PD(d), \\
 F_3 &= \{f_1 \dots f_2\} \text{ where } v_{hw} \geq PD(a), & F_4 &= \{f_1 \dots f_2\} \text{ where } v_{hw} < PD(a), \\
 F_5 &= \{f_1 \dots f_2\} \text{ where } v_{hw} \geq PD(r), & F_6 &= \{f_1 \dots f_2\} \text{ where } v_{hw} < PD(r),
 \end{aligned}$$

where the symbol F denotes that for the observed features for HMM , the set $\{f_1 \dots f_2\}$ represents the sequential falling frames from starting point f_1 to ending point f_2 . $PD(d)$, $PD(a)$, and $PD(r)$ denote the period detection for a fall or an abnormal event with respect to the distance between the centroid and the corresponding VGP , the area of the shape, and the aspect ratio of the person, respectively, and v_{hw} means the half width value employed to determine the period of the abnormal event.

After obtaining thresholds for features, calculations of emission probabilities are performed for developing the HMM model. The hierarchical tree structure used to form features for independent observable symbols O is illustrated in Figure 10.

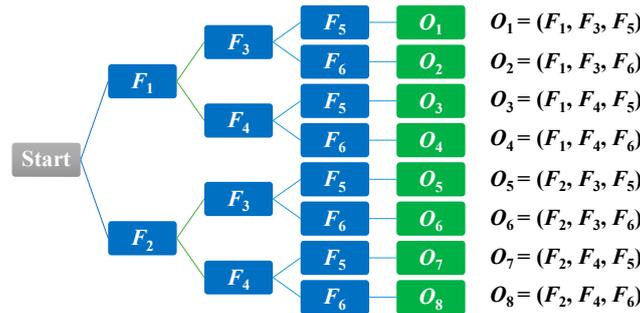


Figure 10. Hierarchical tree structure of independent observable symbol O .

When studying the observable symbols, we can see that F_1 and F_2 , F_3 and F_4 , and F_5 and F_6 are all disjointed. The occurrence of an observable symbol is then labeled based on the eight observable symbols $O = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ for the video sequence. The emission probability of observed symbol o_k ($k = 1, 2, \dots, 8$) in state S_j ($j = 1, 2$) is denoted as in Equation (17).

$$b_j(k) = \frac{\text{Expected Number of Occurrences of } o_k \in \text{States } S_j}{\text{Expected Number of Occurrences of States } S_j}, \quad (17)$$

The emission probability matrix B is then obtained by using the following:

$$B = \begin{bmatrix} b_1(1) & b_1(2) & b_1(3) & b_1(4) & b_1(5) & b_1(6) & b_1(7) & b_1(8) \\ b_2(1) & b_2(2) & b_2(3) & b_2(4) & b_2(5) & b_2(6) & b_2(7) & b_2(8) \end{bmatrix}. \quad (18)$$

Then, the initial probability states are assumed to be $Prob(S_1) = 0.8$ and $Prob(S_2) = 0.2$. Denoting the initial probability vector as π gives the Hidden Markov Model (HMM), as shown in Equation (19):

$$\lambda = (A, B, \pi), \quad (19)$$

where π is the initial probability vector.

After that, the Viterbi algorithm [27] is applied to solve the HMM model to detect consecutive moments in the actions. The idea is to compute the best hidden state sequence (abnormal and normal states) in the given observation sequences and in an HMM .

4. Experiments

4.1. The Dataset

In order to demonstrate a sub-sequential experiment, we used the publicly available Le2i fall detection dataset [28], which simulates the activities of the elderly in a home environment, as well as in an office. This dataset was developed for detecting falls using artificial vision, which can be useful in helping the elderly. These simulated videos illustrate the difficulties in creating realistic video sequences that include variable illumination, occlusions, and cluttered or textured backgrounds.

The actors performed various normal daily activities and falls, in an attempt to simulate elderly behavior. In addition, the actors did not perform acts that would be unnatural for the elderly, such as running or walking quickly. The video sequences were captured using a single RGB camera, in which the frame rate was 25 frames per second, and the resolution was 320×240 pixels. From this dataset, 20 videos taken in an office were selected, representing normal daily activities and falls as performed by four different actors.

Acquiring useful qualitative data for surveillance systems is very important in measuring distances, positions, velocities, and directions for the objects in the scenes, and is a necessary consideration in developing reliable systems [29]. As an important contribution of the current work, we have developed guidelines for setting up and calibrating cameras. First of all, it is essential to try to confirm whether or not the camera height and angle are approximately the same for all videos. We can derive an approximate linear equation for camera calibration. To do so, two extracted features for the object are employed, namely, the aspect ratio (r), and y_{VGP} , corresponding to the y coordinate. These two features can provide information on steps in the walking pattern for a person, as described in Figure 11. In Figure 11, the person exhibits the same posture corresponding to y_{VGP} and r , as illustrated by the blue and orange dashed lines. The interval of the step can be obtained from the place between where the same postures occur. Then, we can observe the index numbers for frames and the corresponding values for the point distance (d), area of the object shape (a), and aspect ratio (r), with respect to starting and ending frames in the step intervals. In order to formulate a linear equation for camera calibration, we assume three kinds of points: (y_{VGP}, d) , (y_{VGP}, a) , and (y_{VGP}, r) , as shown in Figure 12. Then, by observing the testing points, Figure 12 shows that no significant differences exist between the variation within a scene and the variation between all scenes.

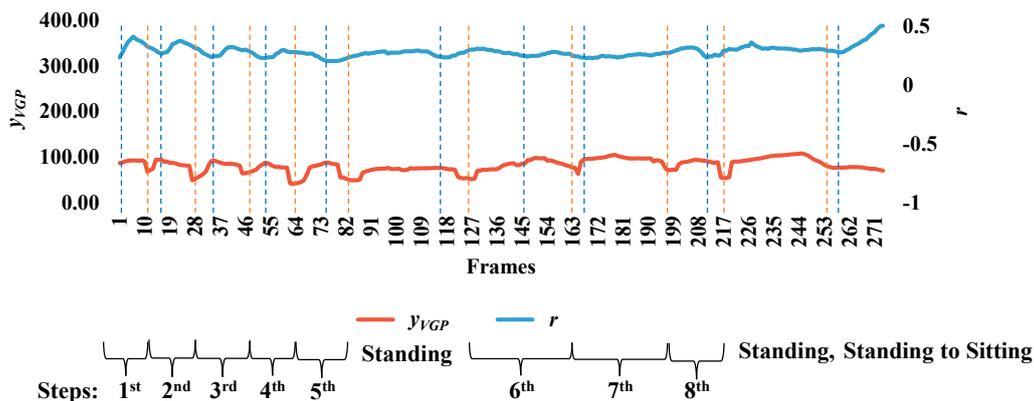


Figure 11. Two extracted features— y_{VGP} (y coordinate of the virtual grounding point (VGP)) and aspect ratio (r)—which can provide information on the number of steps for a person.

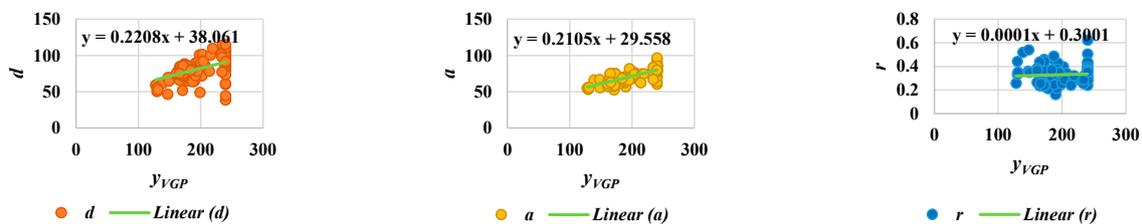


Figure 12. Analysis of camera calibration for all scenes.

4.2. Experimental Results

The given video was firstly analyzed by setting a fixed interval. Here, we made the interval every frame in which only one state occurs (S_1 : abnormal or S_2 : normal). Manual observations of the video sequences were performed to provide the occurrence state symbol for every interval or frame. Such observations for a video sequence used in these experiments are illustrated in Figure 13.

For effectively demonstrating the consequences of the sub-sequential experimental results, here, we denote the name of the person in the video sequence as “Mr. ONE.”

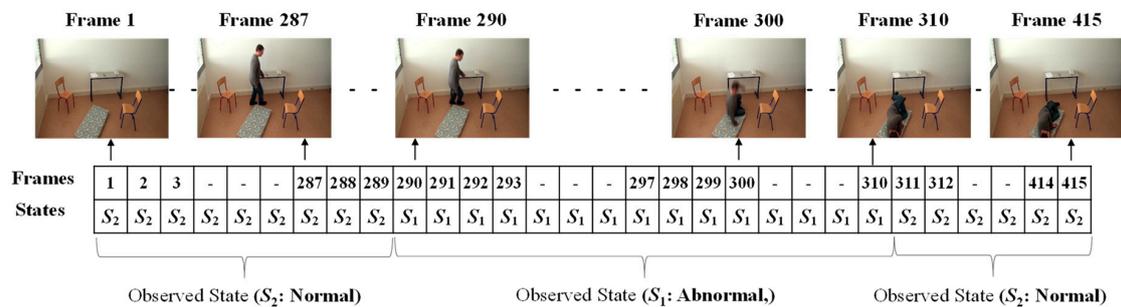


Figure 13. Observations of the occurrences of states for Mr. ONE (Video 1).

After that, as discussed in Equation (16), the state transition matrix of the Markov Chain for the HMM was obtained by applying Equations (13)–(15).

After making observations of the occurrence state symbols (S₁,S₂), the co-occurrence matrix was constructed by applying Equation (13), as shown below.

$$M = \begin{bmatrix} 900 & 11550 \\ 11550 & 148225 \end{bmatrix}$$

Then, the following values were obtained using Equations (14) and (15):

$$a_{11} = c_{11}/c_1 = 0.07229, a_{12} = c_{12}/c_1 = 0.92771, a_{21} = c_{21}/c_2 = 0.07229, a_{22} = c_{22}/c_2 = 0.92771.$$

After that, we obtained the transition probability matrix A(a_{ij}), which represents the probability of proceeding from state S_i to state S_j, by computing Equation (16).

$$A = \begin{bmatrix} 0.07229 & 0.92771 \\ 0.07229 & 0.92771 \end{bmatrix}$$

By employing Equations (17) and (18), the independent observable symbols O were obtained as shown in Figure 14. After that, the occurrence of an observable symbol was identified using one of the labels for the eight observable symbols {o₁,o₂,o₃,o₄,o₅,o₆,o₇,o₈} for the video sequence. The emission probability matrix B was then obtained as shown below:

$$B = \begin{bmatrix} 0.75 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0.020672 & 0.062016 & 0.015504 & 0 & 0 & 0.069767 & 0.054264 & 0.777778 \end{bmatrix}.$$

Matrix B represents a sequence of observation likelihoods in which each element expresses the probability of an observation o_k (k = 1,2, ... ,8) being generated from a state S_j (j = 1,2).

Finally, the Viterbi algorithm [27] was employed for solving the HMM model to recognize sequential abnormal and normal states for “Mr. ONE.” The input parameters for implementing the Viterbi algorithms were summarized as two states—abnormal and normal—including eight observable action symbols; {“falling,” “lying,” “lying to sitting,” “sitting to standing,” “walking,” “walking to sitting,” “sitting,” and “sitting to lying”}, a state transition probability matrix (A), a sequence of observation likelihood or emission probability matrix (B), and an initial probability distribution over states. The final outcome results for “Mr. ONE” are illustrated in Figure 15. Additional experimental results for sequential abnormal and normal states in each of the video sequences simulated by Ms. TWO are demonstrated in Figures 16–18, respectively.

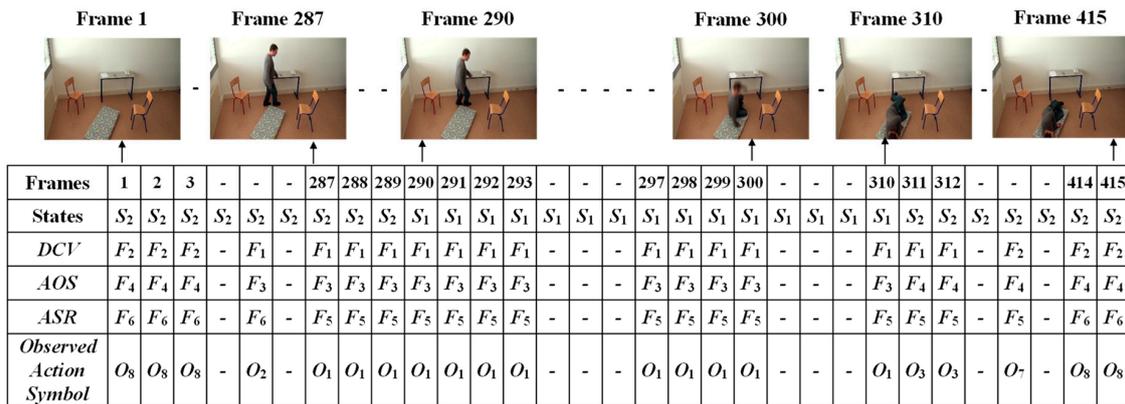


Figure 14. Observations for independent action symbol O for Mr. ONE (Video 1).

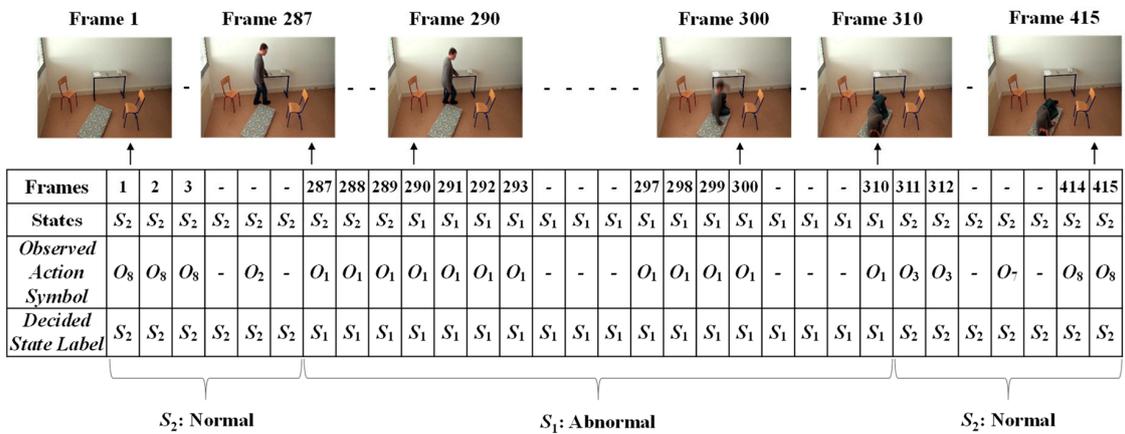


Figure 15. Results for deciding on the sequential abnormal and normal states for Mr. ONE (Video 1).

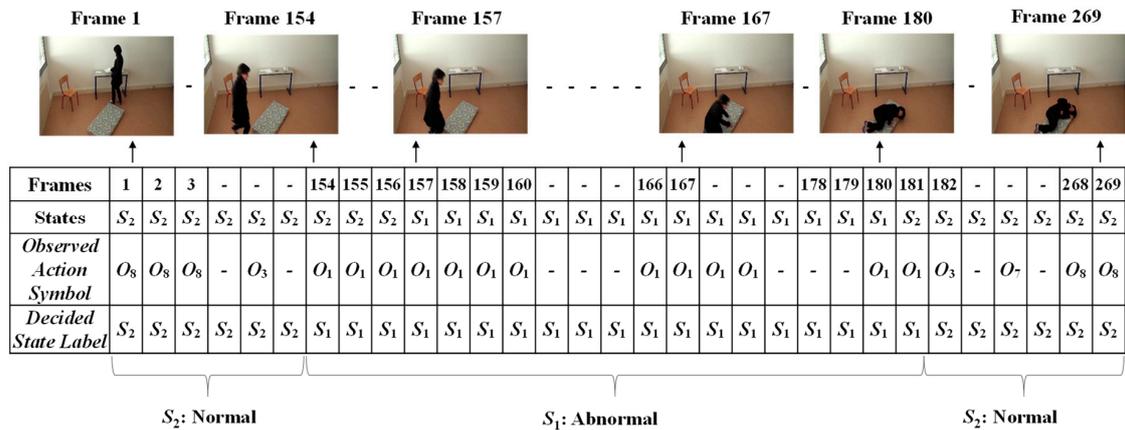


Figure 16. Results for deciding on the sequential abnormal and normal states for Ms. TWO (Video 1).

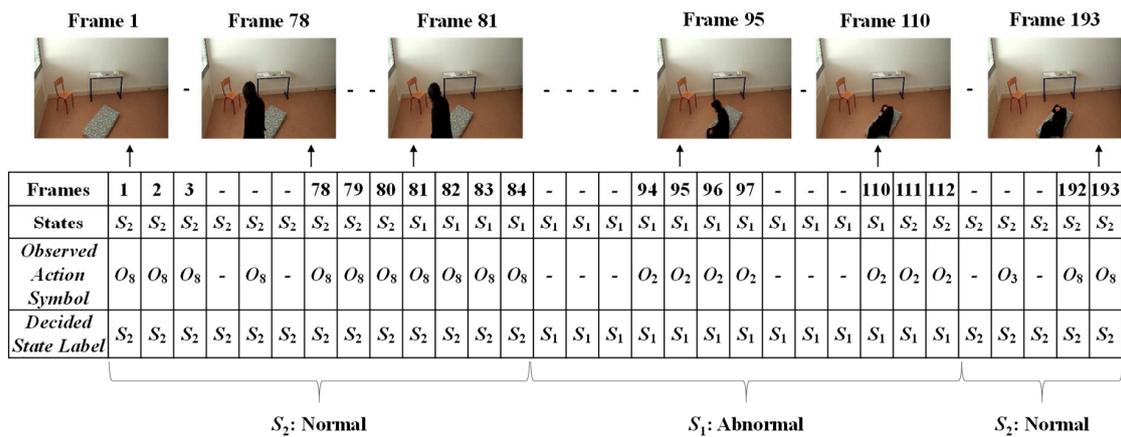


Figure 17. Results for deciding on the sequential abnormal and normal states for Ms. TWO (Video 2).

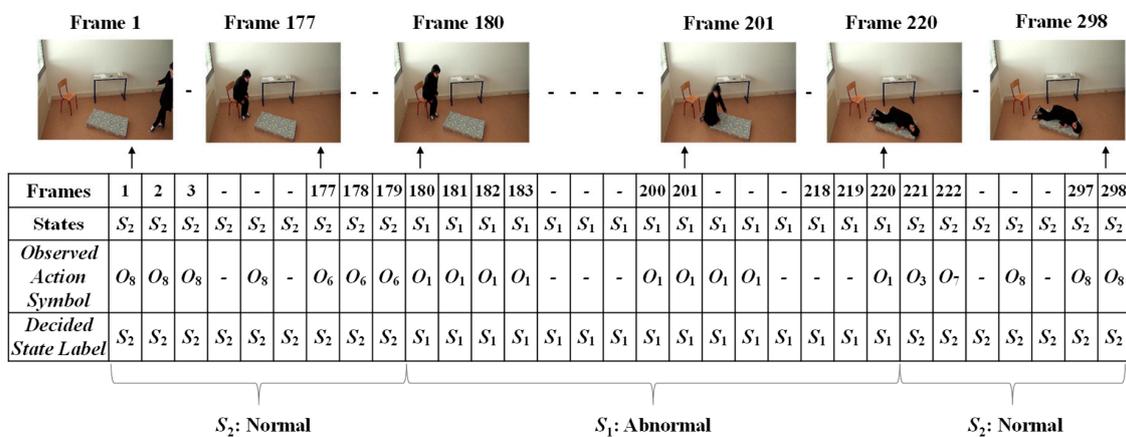


Figure 18. Results for deciding on the sequential abnormal and normal states for Ms. TWO (Video 3).

4.3. Performance Evaluation of the Proposed Approach

In order to evaluate the performance of our proposed system, we firstly used the detection process for falling and normal events using three-fold cross-validation, which presents the swapped variables for learning and testing used in our previous work [13]. The given videos were classified as either falling or non-falling videos. Our proposed period detection (PD) achieved a precision of 93.33% and a recall of 93.33%. In the current work, we enhanced our detection system “from video to consecutive video sequences” by utilizing the analyzed features working for the PD. The features were used as inputs to calculate the probability for HMM. There are four possible outcomes in discriminating abnormal and normal states for the consecutive states of a person, and definitions and relative symbols are described in the following:

- Detected Abnormal State (As_1): A video frame represents an abnormal state, and is correctly classified as “Positive Abnormal”;
- Undetected Abnormal State (As_2): A video frame represents an abnormal state, and is incorrectly classified as “Negative Normal”;
- Normal State (Ns_1): A video frame does not represent an abnormal state, and is correctly classified as “Negative Normal”;
- Mis-detected Normal State (Ns_2): A video frame does not represent an abnormal state, and is incorrectly classified as “Positive Abnormal”.

The following Equations (20)–(24) describe as a performance evaluation the calculated precision, recall, accuracy, specificity rate, and negative predictive value, respectively. In evaluating each of the

given videos, an “undecided class” is nominated in order to save the unnecessary failed states. In other words, where we consider the risk to be low for a given video, we ignore it when a normal case is misclassified as abnormal.

$$\text{Precision} = \frac{A_{s1}}{(A_{s1} + N_{s2})} * 100, \quad (20)$$

$$\text{Recall} = \frac{A_{s1}}{(A_{s1} + A_{s2})} * 100, \quad (21)$$

$$\text{Accuracy} = \frac{(A_{s1} + N_{s1})}{(A_{s1} + N_{s1})} * 100, \quad (22)$$

$$\text{SpecificityRate} = \frac{N_{s1}}{(N_{s1} + N_{s2})} * 100, \quad (23)$$

$$\text{NPV} = \frac{N_{s1}}{(N_{s1} + A_{s2})} * 100, \quad (24)$$

Table 1 provides the precision, recall, accuracy, specificity rate, and negative predicted value (NPV) for 13 videos, including both abnormal and normal states. The remaining seven videos that include a normal state provide a precision of 100% and a recall of 100%. Figure 19 shows receiver operating characteristic (ROC) curve analyses for the results of true positive and false positive fractions, for the 13 videos that include abnormal events. The results of AUC (area under ROC curve) show that our proposed method is quite capable of distinguishing between the classes. A comparison of the performance of the proposed system for each video and related existing methods is shown in Table 2.

Our proposed system was implemented in MATLAB 2019a on an academic license. A series of experiments were performed on Microsoft Windows 10 Pro with an Intel (R) Core (TM) i7-4790 CPU@3.60 GHz and 8GB RAM. The overall computation time in this system is 0.93 s per frame. In order to achieve real-time monitoring, we expect that the GPU implementation could be further fine-tuned to our system.

Table 1. Performance evaluation for 13 fall video sequences.

Videos	Precision (%)	Recall (%)	Accuracy (%)	Specificity (%)	NPV (%)
1	100	100	100	96.88	96.88
2	100	100	100	96.07	96.09
3	100	86.66	98.32	99.04	97.17
4	100	100	100	97.07	97.06
5	100	100	100	90.40	90.40
6	82.98	97.50	99.76	97.89	97.63
7	98.04	100	100	99.36	99.36
8	100	100	100	94.70	94.70
9	100	83.33	97.41	90.80	99.09
10	100	100	100	97.72	97.72
11	100	100	100	98.54	98.54
12	100	100	100	96.38	96.38
13	100	100	100	88.48	88.48

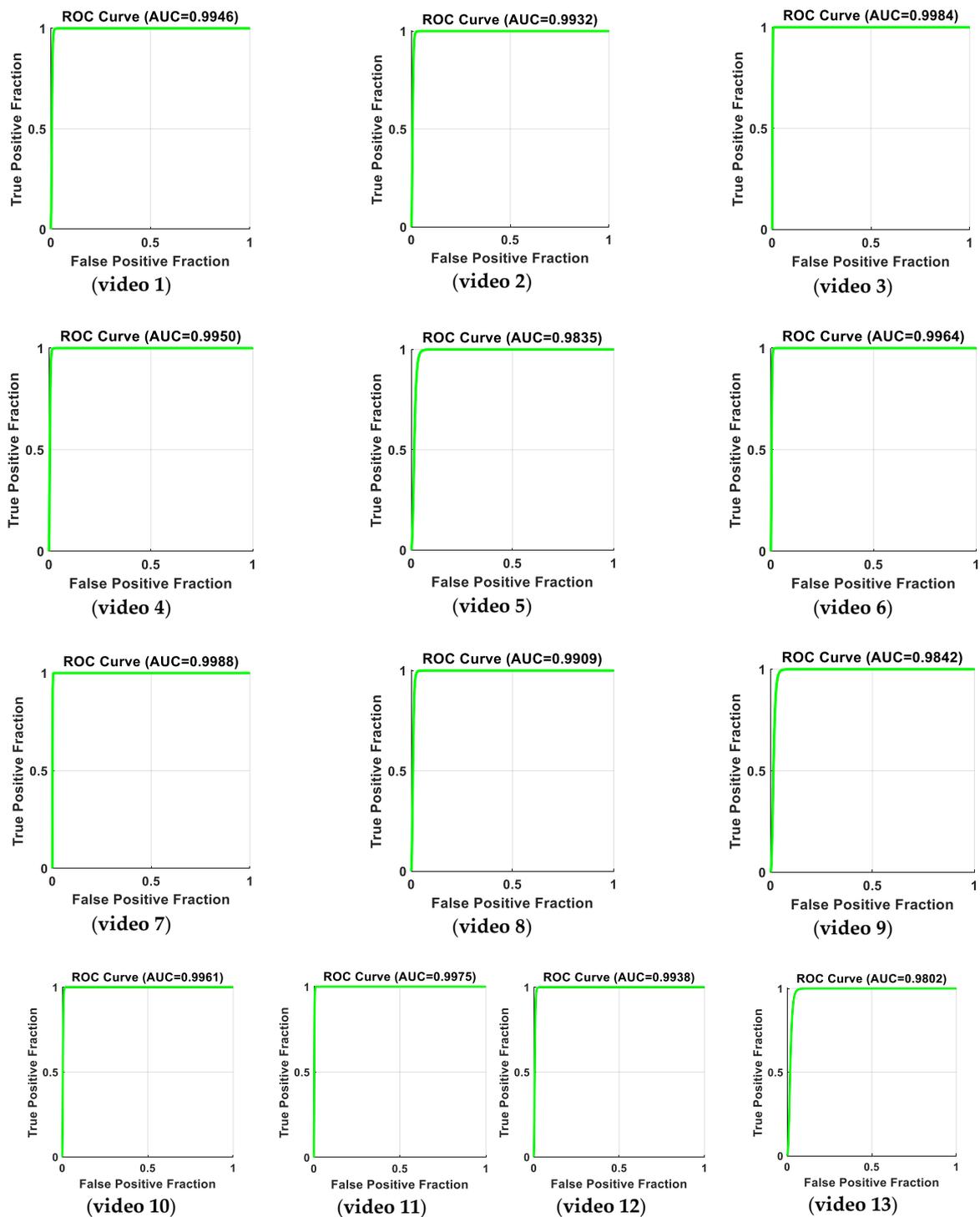


Figure 19. Receiver operating characteristic (ROC) curve analysis for the results of the true positive fraction ($A_{s1}/(A_{s1} + A_{s2})$), and false positive fraction ($N_{s2}/(N_{s2} + N_{s1})$), for 13 videos including abnormal and normal events. The results of AUC (area under ROC curve) with respect to the true positive fraction and false positive fraction show that our proposed method is quite capable of distinguishing between the abnormal and normal states.

Table 2. Performance of our system compared with existing methods using the same dataset.

Methods	Precision (%)	Recall (%)	Accuracy (%)
Kishanprasad [16]	94	95	— ¹
Suad [18]	100	95.27	99.82
Adrián [23]	— ¹	99.00	97.00
Minjie [24]	90.8	98.3	97.8
Ours	99.05	98.37	99.8

Note: —¹ means the provided data is not available.

4.4. Comparative Studies of the Merits and Demerits of Our Proposed System

Kishanprasad presented a fall-detection system [16] providing a precision of 94% and a recall of 95%, as tested in an office environment from the Le2i dataset. The merits of this system were described as robust, fast, and computationally efficient. The system could be enhanced using deep learning models in future work. Demerits of the system were not discussed in detail. However, our proposed system achieved higher precision and recall rates, as shown in Table 2.

Suad [18] proposed a fall detection approach using threshold-based methods and a neural network. The system achieved a high detection rate in monitoring a single person in a home environment. However, their proposed system can only detect events in high-intensity light. They mentioned ways to enhance their proposed system, such as using it to detect multiple people and installing multiple cameras to make sure that the object is visible in at least one camera view. Moreover, they have mentioned adopting enhanced background subtraction for better object segmentation. Other improvements can be achieved by detecting additional features using sensors such as an accelerometer. Capturing 3D information using a depth camera could also improve the detection rate. Using simple algorithms, we achieved an accuracy similar to that of Suad’s proposed system, not attaining the accuracy of our proposed system.

Adrian [23] presented a vision-based fall detection system using *CNN* architecture. As one drawback of the system, using optical flow images involves a heavy computational burden in generating consecutive sequences. Furthermore, the system is not robust when lighting changes. Therefore, a more reliable network could be modeled to gain a better representation of hierarchical motion from the images. In addition, more research involving multiple people in falling events would be useful before implementation in public areas. Compared to this method, our proposed system gives more accurate rates.

Minjie’s work [24] investigated an approach for predicting falling events in advance using monocular videos. The system has credibility as it provided excellent performance results using a complex system architecture. However, they described two failures due to the non-detection of keypoints, which occurred when the person’s upper body was not visible from the perspective of the camera. Moreover, falls cannot be predicted in advance if the precursor of the fall is not detected. In comparison with our proposed system, we achieved a higher precision, recall, and overall accuracy, as described in Table 2.

In this paper, we have proposed a system for detecting normal, abnormal, and falling states using video sequences for people being monitored. In detecting normal states, our system outperformed other approaches, with an accuracy of 99.8%, using experimental videos containing non-abnormal events. In addition, our proposed system proved to be reliable and effective, with a precision of 99.05% and recall of 98.37%, as confirmed using 20 video sequences that contained abnormal and normal events. Though the system was adapted to a real-world environment, the scope of research was limited to detecting a single person. The computation time for the object detection module is slow in a real-time monitoring system. Using multiple cameras to view the human body from different perspectives would improve feature extraction. Currently, the system is only effective during the day, and should be improved for night-time use in the real world of elderly care. Moreover, the system functionality should be expanded by exploring feature analyses such as pre-fall states, dangerous

post-fall situations, and abnormal gaits (walking patterns). We expect that smart image processing technologies have great potential for improving the quality of life of the targeted population.

5. Discussion and Conclusions

The research discussed in this paper is concentrated on applied statistical analysis for developing an advanced image-processing technology in vision-based elderly care monitoring systems, facilitating health care for elderly people living independently. In brief, our proposed approach involves the following. Background subtraction is first used to detect moving and motionless objects. Simple and effective feature extraction is then performed using the virtual grounding point concept (*VGP*), as well as the area of the shape and the aspect ratio of the detected object. In addition, these three features are observed in detailed considerations of the moving average and modified difference calculations. Careful observations of the critical features of the local maximum or local minimum are made to predict the high possibility of an abnormal state for the point distance of the *VGP*, area, and aspect ratio. Moreover, the optimal threshold values are exploited by calculating the mid value or period detection for the probability distribution of the Hidden Markov Model. Finally, the consecutive abnormal and normal states are differentiated using the Viterbi algorithm. Contributing modules include an analysis of the state with an extensive observable area of shape features and making decision rules using the Hidden Markov model. Moreover, a simple analysis of camera calibration is performed using our observed *VGP* and its inclusive features. Experimental results indicate that our proposed system is reliable and effective for detection. In order to confirm the validity of our proposed methods, we have drawn comparisons with existing methods described in the literature. Specifically, comparisons were made between our approach and three similar methods: (1) threshold-based fall detection; (2) neural network algorithms [16,18]; and (3) the *CNN* approach [23,24]. Compared to these methods, our proposed method achieved a higher performance, as confirmed using the public dataset of videos taken in simulated environments that closely approximate real life. This paper focuses on detecting falls, which are common among elderly people. In future research, we will improve our analysis by applying it to behavioral features, such as timing, posture, and gait. Using the results of surveys, we will also take into consideration behavioral differences between healthy young adults and the elderly. Though human behavior is complex and inconsistent between different persons, situations, and environments, our system can improve the interpretation of human behavior through an analysis of its visual manifestations in different situations. However, further testing using large amounts of data is required for system optimization. Furthermore, we expect that our proposed system can be improved by applying techniques related to modern deep learning. In fact, future experiments involving simulations of abnormal events should prove a fruitful area for applying such techniques.

Author Contributions: Conceptualization, S.N.N.H. and P.T.; methodology, S.N.N.H.; software, S.N.N.H.; validation, T.T.Z. and P.T.; formal analysis, S.N.N.H.; investigation, T.T.Z.; resources, S.N.N.H., T.T.Z., and P.T.; data curation, S.N.N.H.; writing—original draft preparation, S.N.N.H.; writing—review and editing, S.N.N.H., T.T.Z., and P.T.; visualization, S.N.N.H.; supervision, T.T.Z.; S.N.N.H. conducted a series of experiments and all three authors analyzed the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to express sincere thanks to Hiromitsu Hama who is an emeritus Professor at Osaka City University, Japan, for his kind and valuable suggestions and advice during this research work. The authors also would like to thank reviewers and editors of this special issue for all valuable suggestions and comments for improving this research paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization Regional Office for Europe. Healthy Ageing: Policy. Available online: <http://www.euro.who.int/en/health-topics/Life-stages/healthy-ageing/policy> (accessed on 13 December 2018).

2. Lindemann, U.; Hock, A.; Stuber, M.; Keck, W.; Becker, C. Evaluation of a fall detector based on accelerometers: A pilot study. *Med. Biol. Eng.* **2005**, *43*, 548–551. [[CrossRef](#)] [[PubMed](#)]
3. Wang, C.-C.; Chiang, C.-Y.; Lin, P.-Y.; Chou, Y.-C.; Kuo, I.-T.; Huang, C.-N.; Chan, C.-T. Development of a Fall Detecting System for the Elderly Residents. In Proceedings of the 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, Shanghai, China, 16–18 May 2008; pp. 1359–1362.
4. Bagalà, F.; Becker, C.; Cappello, A.; Chiari, L.; Aminian, K.; Hausdorff, J.M.; Zijlstra, W.; Klenk, J. Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls. *PLoS ONE* **2012**, *7*, e37062. [[CrossRef](#)]
5. Mathie, M.J.; Coster, A.C.F.; Lovell, N.H.; Celler, B.G. Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiol. Meas.* **2004**, *25*, R1–R20. [[CrossRef](#)] [[PubMed](#)]
6. Rimminen, H.; Lindström, J.; Linnavuo, M.; Sepponen, R. Detection of falls among the elderly by a floor sensor using the electric near field. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 1475–1476. [[CrossRef](#)] [[PubMed](#)]
7. Abbate, S.; Avvenuti, M.; Bonatesta, F.; Cola, G.; Corsini, P.; Vecchio, A. A smartphone-based fall detection system. *Pervasive Mob. Comput.* **2012**, *8*, 883–899. [[CrossRef](#)]
8. Albert, M.V.; Kording, K.; Herrmann, M.; Jayaraman, A. Fall Classification by Machine Learning Using Mobile Phones. *PLoS ONE* **2012**, *7*, e36556. [[CrossRef](#)] [[PubMed](#)]
9. Mao, A.; Ma, X.; He, Y.; Luo, J. Highly Portable, Sensor-Based System for Human Fall Monitoring. *Sensors* **2017**, *17*, 2096. [[CrossRef](#)]
10. Chaccour, K.; Darazi, R.; El Hassani, A.H.; Andres, E. From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems. *IEEE Sens. J.* **2017**, *17*, 812–822. [[CrossRef](#)]
11. Sugimoto, M.; Zin, T.T.; Takashi, T.; Shigeyoshi, N. Robust Rule-Based Method for Human Activity Recognition. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2011**, *11*, 37–43.
12. De Miguel, K.; Brunete, A.; Hernando, M.; Gambao, E. Home Camera-Based Fall Detection System for the Elderly. *Sensors* **2017**, *17*, 2864. [[CrossRef](#)]
13. Htun, S.N.N.; Zin, T.T.; Hama, H. Virtual Grounding Point Concept for Detecting Abnormal and Normal Events in Home Care Monitoring Systems. *Appl. Sci.* **2020**, *10*, 3005. [[CrossRef](#)]
14. Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. Fall Detection from Human Shape and Motion History Using Video Surveillance. In Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), Niagara Falls, ON, Canada, 21–23 May 2007; pp. 875–880. [[CrossRef](#)]
15. Rougier, C.; St-Arnaud, A.; Rousseau, J.; Meunier, J. Video Surveillance for Fall Detection. In *Video Surveillance*; IntechOpen: London, UK, 3 February 2011. [[CrossRef](#)]
16. Kishanprasad, G.; Prachi, M. Indoor Human Fall Detection System Based on Automatic Vision Using Computer Vision and Machine Learning Algorithms. *J. Eng. Sci. Technol.* **2018**, *13*, 2587–2605.
17. Lotfi, A.; Albawendi, S.; Powell, H.; Appiah, K.; Langensiepen, C. Supporting Independent Living for Older Adults; Employing a Visual Based Fall Detection Through Analysing the Motion and Shape of the Human Body. *IEEE Access* **2018**, *6*, 70272–70282. [[CrossRef](#)]
18. Suad, G.A. Automated Human Fall Recognition from Visual Data. Ph.D. Thesis, Nottingham Trent University, Nottingham, UK, February 2019; pp. 1–160.
19. Htun, S.N.N.; Zin, T.T. Motion History and Shape Orientation Based Human Action Analysis. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 754–755.
20. Yong, H.; Meng, D.; Zuo, W.; Zhang, K. Robust Online Matrix Factorization for Dynamic Background Subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1726–1740. [[CrossRef](#)] [[PubMed](#)]
21. Marki, N.; Perazzi, F.; Wang, O.; Sorkine-Hornung, A. Bilateral Space Video Segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 743–751.
22. Htun, S.N.N.; Zin, T.T.; Hama, H. Human Action Analysis Using Virtual Grounding Point and Motion History. In Proceedings of the 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), Kyoto, Japan, 10–12 March 2020; pp. 249–250.
23. Núñez-Marcos, A.; Azkune, G.; Arganda-Carreras, I. Vision-Based Fall Detection with Convolutional Neural Networks. *Wirel. Commun. Mob. Comput.* **2017**, *2017*, 1–16. [[CrossRef](#)]

24. Minjie, H.; Yibing, N.; Shiguo, L. Falls Prediction Based on Body Keypoints and Seq2Seq Architecture. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshop), Seoul, Korea, 27–28 October 2019; pp. 1251–1259. [[CrossRef](#)]
25. Amandine, D.; François, C. Automatic Fall Detection System with a RGB-D Camera using a Hidden Markov Model. In Proceedings of the ICOST-11th International Conference on Smart Homes and Health Telematics, Singapore, 19–21 June 2013; pp. 259–266. [[CrossRef](#)]
26. Triantafyllou, D.; Krinidis, S.; Ioannidis, D.; Metaxa, I.; Ziazios, C.; Tzovaras, D. A Real-time Fall Detection System for Maintenance Activities in Indoor Environments—This work has been partially supported by the European Commission through the project HORIZON 2020-INNOVATION ACTIONS (IA)-636302-SATISFACTORY. *IFAC-PapersOnLine* **2016**, *49*, 286–290. [[CrossRef](#)]
27. Daniel, J.; James, H.M. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed.; Stanford University: Stanford, CA, USA, 16 October 2019.
28. Le2i Fall Detection Dataset. Available online: <http://le2i.cnrs.fr/Fall-detection-Dataset?lang=fr> (accessed on 27 February 2013).
29. Cattaneo, C.; Mainetti, G.; Sala, R. The Importance of Camera Calibration and Distortion Correction to Obtain Measurements with Video Surveillance Systems. *J. Phys. Conf. Ser.* **2015**, *658*, 012009. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).