



Article Virtual Grounding Point Concept for Detecting Abnormal and Normal Events in Home Care Monitoring Systems

Swe Nwe Nwe Htun ^{1,*}, Thi Thi Zin ² and Hiromitsu Hama 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 Graduate School of Engineering, Osaka City University, Osaka 558-8585, Japan; hama@ado.osaka-cu.ac.jp
- * Correspondence: nc17010@student.miyazaki-u.ac.jp

Received: 31 March 2020; Accepted: 22 April 2020; Published: 25 April 2020



Featured Application: This research can be applied to home care monitoring systems to assist in providing care for dependent persons by analyzing abnormal events or falls.

Abstract: In this paper, an innovative home care video monitoring system for detecting abnormal and normal events is proposed by introducing a virtual grounding point (VGP) concept. To be specific, the proposed system is composed of four main image processing components: (1) visual object detection, (2) feature extraction, (3) abnormal and normal event analysis, and (4) the decision-making process. In the object detection component, background subtraction is first achieved using a specific mixture of Gaussians (MoG) to model the foreground in the form of a low-rank matrix factorization. Then, a theory of graph cut is applied to refine the foreground. In the feature extraction component, the position and posture of the detected person is estimated by using a combination of the virtual grounding point, along with its related centroid, area, and aspect ratios. In analyzing the abnormal and normal events, the moving averages (MA) for the extracted features are calculated. After that, a new curve analysis is computed, specifically using the modified difference (MD). The local maximum (l_{max}) , local minimum (l_{min}) , and half width value (v_{hw}) are determined on the observed curve of the modified difference. In the decision-making component, the support vector machine (SVM) method is applied to detect abnormal and normal events. In addition, a new concept called period detection (PD) is proposed to robustly detect the abnormal events. The experimental results were obtained using the Le2i fall detection dataset to confirm the reliability of the proposed method, and that it achieved a high detection rate.

Keywords: virtual grounding point; abnormal and normal events; mixture of Gaussians; moving average; modified difference; half width value; support vector machine

1. Introduction

In recent years, 24/7 monitoring systems for dependent persons who are living alone at home have become an important topic of research in the field of image technology. Here, the term dependent persons not only includes older adults who require care with regular, long-term monitoring, but also disabled persons, and patients with chronic diseases [1]. Such people may have problems with mobility that can affect their health, quality of life, and expected life span. Falls are the most common injuries facing dependent persons. Accordingly, much research has focused on fall detection. Most systems in use for detecting abnormal events or falls are classified into three groups: those using wearable-sensors [2],

or ambient sensors [3], and those using computer-vision and image processing [4–6]. Wearable sensors are commonly used to collect information related to body movements and provide notification when a fall occurs. However, movement-based systems cannot provide notification if the person is already unconscious after falling. Ambient sensors can be installed under the bed or floor to capture the vibration that occurs when a person falls. Although such sensors do not disturb the person, they can generate false alarms. For these reasons, a system based on computer vision is more beneficial and reliable. In addition, visual surveillance systems can detect specific human activities, such as walking, sitting, and lying down [7,8].

Therefore, we propose a vision-based system for home care monitoring that detects normal as well as abnormal events, including falls. The main contributions of this paper are described in the following developments.

- A detection system for visible abnormal and normal events based on data gathered using an RGB video camera;
- A modified method of statistical analysis involving virtual grounding point features that provides reliable information, not only on the exact time of a fall, but also on the pre-impact period of a fall.

In this study, our proposed approach makes an effort to improve the detection rates for abnormal and normal events in a home-care monitoring system. Since our intention is to develop a long-term monitoring system for an assisted-living environment, we take great care in considering the extracted features for the details of human posture for precisely detecting falls. In this study, we consider new features using the concept of a virtual grounding point for the human body, and related visual features. We conduct abnormal event analysis including modified statistical analysis. Finally, the decision-making process is performed by applying a support vector machine, and a new consideration that involves detecting the period of a fall.

The following provides a step-by-step description of system and method. We firstly conduct foreground and background separation to detect both moving and motionless people in the video scenes. Secondly, we perform feature extraction, including the construction of a virtual grounding point and its associated visual features. Thirdly, we obtain analytical information for fall events by analyzing moving averages for the extracted features, and computing differences for the observed moving averages for the extracted features. Finally, a support vector machine is used to set the rules for decision making and detect the period of the event in effectively distinguishing abnormal and normal events. The rest of this paper is organized as follows: Section 2 presents related works; Section 3 presents the methodologies of our proposed system; Section 4 presents and evaluates experimental results, and finally, Section 5 presents the conclusion and speculates on future trends in our approach.

2. Related Works

In most video monitoring systems, the most fundamental step is background subtraction, which assumes that the distribution of background pixels can be separated from that of foreground pixels in detecting a silhouetted object. Methods used for this purpose involve statistical measures such as median and mean [9] to model the background. In addition, a more complicated distribution for background pixels can be obtained using models such as mixture of Gaussians (*MoG*) [10] and mixture of generalized Gaussians (*MoGG*) [11]. As these methods do not always provide great performance due to not taking a knowledge of video structure into consideration, a low-rank subspace learning approach has been proposed [12] to take account of the video structure, including the temporal similarity of the background scene, and the spatial contiguity of foreground objects. However, most of these methods focus on solving for detection of moving objects in video scenes. When the foreground objects move very slowly, the redundant data occurs, resulting in serious outliers. To solve this problem, methods are applied for motion detection and frame differencing, eliminating redundant data [13]. However, this technique can result in a loss of useful information in real-life video sequences. Therefore, we have applied a graph cut theory as a solution for refining the results of background subtraction [14,15].

In this study, we achieved foreground refinement by combining *MoG* with the low-rank subspace learning method [12] for background subtraction and the graph cut algorithm [15].

In order to represent human objects in video sequences, shape analysis is performed using the bounding box method, which gives the attributes of height and width to calculate the aspect ratio in defining a fall. Our previous method for action analysis first involved creating a fitted ellipse bounding box around the object body [16]. The moment is then computed for the continuous images and the ellipse's center. Its major axis, minor axis, and its orientation are used as the observed features for human actions. In addition, horizontal and vertical histograms are constructed to obtain good performance for posture detection of the object. An additional method in our previous work for feature extraction involved the variation of motion using timed motion history images (*tMHI*). This had provided the basic facts for detecting great motion in abnormal scenes. However, these techniques required fixed threshold values for analyzing and detecting the events. In addition, the movement characteristics in the walking patterns of the human body are studied by determining the features of temporal variability, such as joint angle (ankle, knee, hip, torso) movements. However, additional observations are needed in which values for temporal variations with the pre-selected variables might deviate during investigation, and useful information for analysis on potential features might be discarded. In this regard, the proposed method [17] uses movement variability throughout the whole body, which need not consider pre-selected variables. However, this system uses a full-body marker set, which consists of 28 markers placed on human body segments for gait analysis. Requiring the elderly and dependent persons who have chronic diseases to wear markers 24 h per day is unrealistic.

Moreover, we propose a three-dimensional model of the human body [18] based on the visual appearance of the human subject that changes over time in aspects such as self-shadowing, and clothing deformation. The idea is to develop an adaptive, appearance model for articulated body parts by using a mixture model that includes an adaptive template and a frame-to-frame matching process. As motivation for this approach, research featuring background subtraction from foreground silhouettes has not provided reliable tracking. This suggests that a three-dimensional articulated model of the human body must be developed without using a high-quality silhouette provided by background subtraction. However, the system has limitations in requiring high-quality resolution of image data and multiple cameras to cope with self-occlusions. In addition, a method of estimating real-time, multi-person, two-dimensional poses is proposed [19] by using part affinity fields of skeleton images. This architecture is designed to jointly study human-body part locations to recognize human poses by utilizing a convolutional neural network (CNN) trained with a large amount of data. However, the results which are tested on different datasets show that the detection of body parts could not adequately differentiate human subjects from objects similar to the human body. Therefore, a suggestion for improving this system involves embedding an improved background subtraction technique during the pre-processing stage. This kind of extended body part model might provide the research for the detection of abnormal or falls and normal daily activities. Moreover, the features based on the body shape of the silhouette images are investigated for patterns of human movement in the literature survey [20]. The boundary points of the body are extracted and the distances are calculated from the centroid of the object. Moreover, several physical features could be observed for the gait period, stride length, and height, as well as the ratio of chest width to body height with respect to analyzing the patterns of walking slow, fast and normal. However, most of the research that focused on an analysis of walking patterns mentions that each individual person has a unique walking style, and it is necessary to pre-select according to age, and gender, as well as whether subjects are healthy or not.

In analyzing abnormal and normal events, most systems utilize the coefficient of motion and prefixed threshold features [5,7,16,21]. The proposed method [22] uses four features as the inputs to a k-nearest neighbor (*K*-*NN*) classifier to detect a fall: orientation angle, ratio of fitted ellipse, coefficient of motion, and silhouette threshold. The overall accuracy of the system is 95% in real time video sequences. Similarly, a fall-detection method is proposed [23] that depends on the activity patterns of the detected person, such as the aspect ratio and speed of motion of the human object. The region based

convolutional neural network (*R*-*CNN*) deep learning algorithm is then selected to obtain information on the object's position in the video sequences during a fall. This method can be used to classify falls during normal daily activities with an accuracy of 95.50% using the simulated video sequences. Moreover, Charfi 2013 [24] observed fourteen features extracted from the detected bounding box, which includes the aspect ratio, centroid, and ellipse orientation for the detection of falls. The combination of Fourier and wavelet transforms, using first and second derivatives, is also utilized in determining these features. An evaluation is then performed using the support vector machine (SVM) and adaptive boosting (AdaBoost) classifiers. The system achieves an accuracy of 99.42%, a precision of 95.91%, and a recall of 92.15% using the Le2i fall detection dataset. In addition, another method for the detection of unnatural falls is proposed in 2018 [25] by extracting the features of aspect ratio, orientation angle, motion history image, and objects below the threshold line. The obtained features are utilized as the inputs for the detection system by applying SVM, K-NN, Stochastic Gradient Descent (SGD), Decision Tree (DT) and Gradient Boosting (GB). The observed performance of the detection system using Decision Tree (DT) provides an accuracy of 95%, a precision of 94%, and a recall of 95% in the Le2i fall detection dataset. Then, The proposed fall-detection system [5], in support of independent living for older adults, generates features for classifying falls by extracting motion information using a best-fit ellipse and a bounding box around the silhouetted object, projection histograms, and estimates for head position over time. A multilayer perceptron (MLP) neural network is used to generate the extracted features for fall classification. This method shows the reliability of the approach with a high fall-detection rate of 99.60% when tested with the UR (University of Rzeszow) fall detection dataset. The Le2i fall-detection dataset was also used to extend [26] the performance evaluation of the fall detection method. The accuracy of this method is 99.82% with a precision of 100% and a recall of 95.27%.

3. Proposed System Architecture

In this section, we provide an overview of the proposed home care monitoring system to detect abnormal events or falls occurring to ambulant people living independently. The term abnormal represents the falls and normal represents normal daily activities such as walking, standing, sitting and lying down. The proposed system uses a virtual grounding point concept, and its observable visual features are as shown in Figure 1. The proposed system includes four main components: object detection, feature extraction, analysis of abnormal and normal events, and the decision-making process for event detection. The technical details for each component are described in the following sub-sections.



Figure 1. Overview of the proposed system.

3.1. Object Detection

The main purpose of object detection is to properly separate foreground objects from the background in the scene. There are two main parts in object detection, described as follows.

3.1.1. Mixture of Gaussians (MoG) Model

In this part, the mixture of Gaussians (*MoG*) using the low-rank matrix factorization model [12] is selected to perform foreground and background segmentation. Using this model, knowledge from previous frames is learned and updated frame by frame. The probabilistic model of *MoG* noise distribution in the low-rank matrix factorization form in each successive frame is briefly introduced in Equation (1):

$$x_i^t \sim \prod_{k=1}^K \left(x_i^t \Big| (u_i)^T v_i, \sigma_k^2 \right)^{z_{i_k}^t}, z_i^t \sim \text{Multi} \left(z_i^t \big| \Pi \right)$$
(1)

where x_i^t signifies the *i*th pixels of x_t , $k = 1 \dots$, K represents the number of Gaussians, N means the number of random variables, u_i means the *i*th row vectors of low-rank properties; U(basis), and v_i are the V(coefficient) matrices. In addition, σ and Π are the variances and mixture rate, respectively. Multi refers to the multinomial distribution. Then, the *MoG* model is utilized by implementing the expectation maximization (*EM*) algorithm on a new frame sample x_t for updating parameters for foreground and background [12]. However, the resultant *MoG* cannot give the optimal solution for background subtraction, and ghost effects occur around the foreground object. In real-life video sequences, much redundant data occurs, as when foreground objects move very slowly or remain in place for a long period. In such situations, the system cannot always recognize a person as the foreground when the person comes into the frame and sits there for a long time. This has been a recurrent problem, but to address this issue, graph theory is used to refine the foreground.

3.1.2. Graph Cut

The resultant foreground and background pixels are given as a set of inputs for the video sequences. We now seek binary labels that mark each vertex v_{vertex} as the foreground, set to 1, and the background set to 0. Then, these labels are computed by constructing a graph $G = (v_{vertex}, \varepsilon)$, where v_{vertex} is the set of vertices (i.e., pixels), and ε is the set of edges linking nearby 4-connected pixels [15]. Finally, the maximum-flow and minimum-cut algorithm is applied to find the vertex labeling with a minimum energy function [27]. In applying the theory of graph cut for refining the foreground, we here focus on the user-assisted case, but note that an *MoG* mask is given every 100th frame, instead of manually re-drawing the scribbles and region of interest (*ROI*) for every frame. A comprehensive research problem and solution is illustrated in Figure 2.



Figure 2. Foreground refinement by using graph cut.

3.2. Feature Extraction

In this component, the features are extracted from the detected foreground object, including the centroid (*C*), area, height, width, aspect ratios (*r*), and the virtual grounding point (*VGP*). Specialized terminology and notations for feature extraction are provided in the following.

With the use of *VGP*, we aim to define new parameters describing patterns of human action. Four steps are involved in constructing *VGP*, and the technical details are described as follows.

1. Firstly, the position *p* at time *t* of the detected foreground object is defined as in Equation (2).

$$p(t) = (x(t), y(t))$$
 (2)

Then, the centroid of the object is obtained by Equation (3), as shown in Figure 3a.

$$C(t) = (x_c(t), y_c(t))$$
 (3)



Figure 3. Constructing a virtual grounding point by (**a**) centroid of object, (**b**) vertical line from top row to bottom row across the *x* axis, (**c**) horizontal line from the left to right columns across the *y* axis, (**d**) virtual grounding point.

Specifically, each x_c and y_c is simply formulated as in Equation (4).

$$x_c = \sum_{i=1}^{N} \frac{x_i}{N}, \quad y_c = \sum_{i=1}^{N} \frac{y_i}{N}$$
 (4)

- 2. Secondly, a vertical line from the top-most row to the bottom-most row is created along the *x* axis of the centroid, as shown in Figure 3b.
- 3. Thirdly, a horizontal line from the left column to the right column is created along the *y* axis of the centroid at the bottom-most row, as shown in Figure 3c.
- 4. Finally, a point for *VGP*(*t*) is marked on the horizontal line of the bottom row along the *y* axis where the vertical line on the *x* axis extends from the centroid. Figure 3d describes the final result for *VGP*, which can be formulated as in Equation (5).

$$VGP(t) = (x_{VGP}(t), y_{VGP}(t))$$
(5)

We noticed that the virtual grounding point (*VGP*) can be simply obtained from the object centroid. In addition, the patterns of posture can be analyzed by observing pairs of changes in *C* and *VGP*, as shown in Figure 4. The underlying pattern in Figure 4a indicates that the distance between *C* and *VGP* is initially quite short, and then lengthens as the pattern of the person's position changes from lying down to getting up. The distance between points *C* and *VGP* in the pattern for sitting down is quite short, and it shortens during the transition from standing to sitting, as shown in Figure 4b. Therefore, the differences between *VGP* and *C* are regarded as values for the supportive features for abnormal and normal patterns, as formulated in Equation (6).

$$d(t) = y_{VGP}(t) - y_c(t) \tag{6}$$

where *d* is the distance of *VGP* from *C* along the *y* axis, $y_{VGP}(t)$ means the virtual grounding point along the *y* axis at time *t*, and $y_c(t)$ represents the centroid along the *y* axis at time *t*.



Figure 4. Example action patterns obtained from constructing the virtual grounding point. (**a**) Getting up pattern by changing centroid (*C*) and virtual grounding point (*VGP*). (**b**) Pattern for sitting down by changing centroid (*C*) and virtual grounding point (*VGP*).

Moreover, information on the regularity of the object shape related to VGP is obtained by calculating the area. Finally, the aspect ratio (r) of the object is simply calculated to predict the posture as in Equation (7). The concepts for calculating area and aspect ratio are shown in Figure 5.

$$r(t) = w(t)/h(t) \tag{7}$$

where r(t) represents the aspect ratio of the object at time t, and w and h refer to the width and height of the object, respectively.



Figure 5. Illustration for area, width and height of object. (**a**) Area of object. (**b**) Width (*w*) and height (*h*) of object.

3.3. Abnormal and Normal Event Analysis

In the analysis determining whether events are abnormal or normal, we first observe the features of VGP on $x_{VGP}(t)$, $y_{VGP}(t)$, d, area and r, starting with observed features for $x_{VGP}(t)$ and $y_{VGP}(t)$, as illustrated in Figure 6a. In the analysis for walking as shown in this Figure, $x_{VGP}(t)$ decreases at each pixel location before the turning point. The turning point indicates where the person is walking from, or standing to, the right or the left, and then turns to the left or the right. After this turning point, $x_{VGP}(t)$ again increases significantly. At that time, $y_{VGP}(t)$ also decreases during a finite period before the turning point, and then increases for an extended period after the turning point. Therefore, the period of actions can be clearly analyzed. Comparisons between distance d and $y_{VGP}(t)$ can be considered supportive VGP features in analyzing changes in the object's position, as shown in Figure 6b. Then, aspect ratio (r) is added as a feature to efficiently analyze the object's posture. The person remains

in the same posture for a period of the time, as shown by orange and blue dashed lines in Figure 6c. Therefore, determining whether events are abnormal or normal depends on distance (d) and aspect ratio (r), as performed using the three steps described in the following sub-sections.



Figure 6. Concepts for observation of human posture. (a) Observation on the features $x_{VGP}(t)$ and $y_{VGP}(t)$. (b) Observation on the features $y_{VGP}(t)$ and d(t). (c) Observation on the features d(t) and r(t).

3.3.1. Moving Average (MA)

In analyzing the data points statistically, the moving average is first calculated on the series of data. An odd length symmetric moving average (*MA*) is computed, which can be utilized at points to smooth time series data in order to estimate the expected trend of abnormal events. We here propose a formula for moving average (*MA*) as in Equation (8).

$$MA(t,F,N) = \frac{1}{2N+1} \sum_{f=t-N}^{f=t+N} F(f), \quad F(t) = d(t), r(t)$$
(8)

where MA(t,F,N) represents the average period in N at time t, N means the number of time periods, F(t) represents the computation on two features, namely point distance (d) and aspect ratio (r). We here set the optimal value of N at Th (MA(t, d, Th), and MA(t, r, Th)), to determine the detection rule by analyzing the crossing point. The optimal value of Th depends on the frame rate, and here we set Th = 51. The idea behind making observations from the crossing point of the moving average is demonstrated in Figure 7a,b.



Figure 7. Analysis for possibility of abnormal event by using moving average. (**a**) Moving average calculation on aspect ratio. (**b**) Moving average calculation on point distance.

3.3.2. Modified Difference (MD)

The difference calculus is formulated to determine stationary points on the moving average of aspect ratio (r) and point distance (d), as shown in Equation (9). The observation can clearly provide information on the high possibility of an abnormal point, confirmed according to the crossing point of moving average and the maximum or minimum stationary point.

$$MD(t, F, N_0, N_1) = MA(t, F(t+N_0+N_1), N_1) - MA(t, F(t-N_0-N_1), N_1)$$
(9)

where $MD(t,F,N_0,N_1)$ represents the modified difference for the selected features (*r* and *d*) at time *t* between the predefined moving averages. The selected optimal value for N_0 is 0 and N_1 is 51.

Since the possibility of an abnormal point is estimated when the aspect ratio decreases relative to the pixel's location, the maximum difference value (l_{max}) on the moving average of aspect ratios can give the highest abnormal action point. When a person falls, point distance *d* and its moving average immediately increases relative to the pixel's location. In that case, a minimum difference value (l_{min}) must be considered in order to detect an abnormal event. Concepts for consideration are sketched out in Figure 8a,b.



Figure 8. Analysis for possibility of abnormal event by using modified difference. (a) Calculation of modified difference on moving average of aspect ratio. (b) Calculation of modified difference on moving average of point distance.

3.3.3. Half Width Value (v_{hw})

In order to more precisely detect the periods of abnormal events, we here consider a parameter called half width value (v_{hw}) on the curve of the modified difference. The starting point (f_1) and the ending point (f_2) are set at the half of the largest curve which can represent the irregular event. Then,

the periods for abnormal and normal events are obtained by calculating the distance of f_1 and f_{2} , as in Equation (10), and the consideration of v_{hw} is described in Figure 9.

$$v_{hw} = |f_1 - f_2| \tag{10}$$

where v_{hw} is the half width value of MD, f_1 and f_2 mean the estimated starting and ending periods of a fall event, respectively.



Figure 9. Analysis for possibility of abnormal period by calculating half width value. (**a**) Half width value (v_{hw}) on aspect ratio. (**b**) Half width value (v_{hw}) on point distance.

3.4. Decision Making Process

We first apply a support vector machine (*SVM*) in order to classify the abnormal and normal events. The main reason for selecting an *SVM* approach is that it can work well if the training dataset is small, or occupies a high dimensional space. For the extracted aspect ratio (r), l_{max} and v_{hw} are used as inputs to *SVM*. Then, for the extracted point distance (d), l_{min} and v_{hw} are used as inputs to *SVM*. Then, for the extracted point distance (d), l_{min} and v_{hw} are used as inputs to *SVM*. Then, for the extracted point distance (d), l_{min} and v_{hw} are used as inputs to *SVM*. Then, for the extracted point distance (d), l_{min} and v_{hw} are used as inputs to *SVM*. Then, for the observed aspect ratio (r) is

$$D(r) = a * r(l_{\max}) - b * r(v_{hw}) + c, \begin{cases} l_1 & \text{if } D(r) \ge 0\\ l_2 & otherwise \end{cases}$$
(11)

where D(r) represents the distance between a point (support vector) and a linear straight line for the feature of aspect ratio r. l_{max} and v_{hw} are local maximum and half width values from the observed modified difference, respectively. c means the *SVM* optimization value to avoid misclassifying each training example. l_1 and l_2 represent "abnormal" and "normal" events, respectively.

Then, the feature called point distance (d), from VGP to C is formulated as,

$$D(d) = a * d(l_{\min}) - b * d(v_{hw}) + c, \quad \begin{cases} l_1 & \text{if } D(d) \ge 0\\ l_2 & otherwise \end{cases}$$
(12)

where D(d) represents the distance between a point (support vector) and a linear straight line for the observed point distance between virtual grounding point and centroid. l_{min} and v_{hw} are local minimum and half width values from the observed modified difference, respectively.

When we observe v_{hw} for abnormal and normal events, we also notice that the period of an abnormal event is longer than that of a normal event. The reason is that a person who has fallen may take time to recover. If such a person does not get up for a long time, that would indicate a dangerous situation. An evaluation of the period of an event can be used to detect a fall.

To do this, we define the sets of abnormal events as $A = \{a_1, a_2, ..., a_k\}$, and of normal events as $N = \{n_1, n_2, ..., n_k\}$. We can then compute,

$$\alpha_1 = MIN(v_{hw}) \in A \tag{13}$$

$$\alpha_2 = MAX(v_{hw}) \in N \tag{14}$$

After that, the period detection (PD) of a fall for the observed aspect ratio (r) is obtained by,

$$PD(r) = (\alpha_1(r) + \alpha_2(r))/2, \begin{cases} l_1 & \text{if } v_{hw} > PD(r) \\ l_2 & \text{otherwise} \end{cases}$$
(15)

Then, the period detection (PD) for point distance (d) is computed by,

$$PD(d) = (\alpha_1(d) + \alpha_2(d))/2, \begin{cases} l_1 & \text{if } v_{hw} > PD(d) \\ l_2 & \text{otherwise} \end{cases}$$
(16)

where *PD*(*r*) and *PD*(*d*) mean period detection for abnormal and normal events for two features (aspect ratio *r* and point distance *d*, respectively). α_1 represents the minimum period value for abnormal events. α_2 represents the maximum period value for a normal event. l_1 and l_2 are the class labels for "abnormal" and "normal" events, respectively.

In setting the decision-making rules, the "undecided" class is nominated in order to save the failed states. For example, when a "normal" case is misclassified as an "abnormal" case, it can be considered low risk. Therefore, the decision rules are set to include the undecided class. In doing so, the rules for abnormal and normal event classification are verified with the ground truth, which refers to the information provided by direct observation. If the value for label l_1 equals that for the ground truth, l_1 is considered an "abnormal" event. If the value for label l_2 equals that for the ground truth, l_2 is considered a "normal event", otherwise it is "undecided."

4. Experiments

4.1. Dataset

In order to illustrate the proposed system, the experiments were conducted using the Le2i fall detection dataset [28], which represents a realistic video surveillance setting taken by a single RGB camera. The frame rate was 25 fps and the size was 320×240 . The video sequences presented typical difficulties, such as occlusions, clutter, and textured backgrounds. In the video scenes, falls and normal daily activities were simulated at different locations, such as home and office. Different types of fall events were recorded to include falls caused by a loss of balance, as well as forward and backward falls. In the dataset, 20 videos were randomly selected to confirm the effectiveness of the proposed system. In the video sequences, four healthy subjects including three males and one female performed the simulated falls. Some of the video sequences used in the experiments are as shown in Figure 10.



Figure 10. Example video sequences which include the falls.

In the experimental work, step-by-step procedures for object detection were conducted. Then, the acquired silhouetted objects were used to extract useful features through *VGP*. After that, the extracted features from the human body were analyzed to detect falls. At this point, we stress the importance of using step-by-step methodologies of statistical analysis for precisely detecting abnormal and normal events. We first calculated the moving average (*MA*) by observing details of human behavior and posture. For estimating the possibility of an abnormality through the crossing point, we performed approval calculations for the modified difference (*MD*), including its local maximum (l_{max}), and local minimum (l_{min}). In addition, the period of the falling event was analyzed using the half width value (v_{hw}). In order to classify events as abnormal or normal, the l_{max} and v_{hw} of aspect ratio (r), l_{min} and v_{hw} of point distance (d) were used as the input features into a support vector machine (*SVM*). To visualize the input features, Figures 11 and 12, respectively, illustrate r and d conducting the linear discriminating line for classification. Then, period detection (*PD*) using the half width value was performed to confirm falls. In addition, some of the experimental results for distinguishing abnormal from normal events are illustrated in Figures 13–15, respectively. The analytical results for fall trajectories were demonstrated as shown in Figure 16.



(a)

(**b**)

Figure 11. Results for abnormal and normal event classification using aspect ratio. (**a**) Event classification based on local maximum (l_{max}) and v_{hw} . (**b**) Linear discrimination for class categories.



Figure 12. Results for abnormal and normal event classification using point distance. (a) Event classification based on local minimum (l_{min}) and v_{hw} . (b) Linear discrimination for class categories.



Figure 13. Analyzing abnormal and normal events in scenario 1. (a) Abnormal event analysis based on l_{max} . (b) Abnormal event analysis based on l_{min} .



Figure 14. Analyzing abnormal and normal events in scenario 2. (a) Abnormal event analysis based on l_{max} . (b) Abnormal event analysis based on l_{min} .



Figure 15. Analyzing abnormal and normal events in scenario 3. (a) Abnormal event analysis based on l_{max} . (b) Abnormal event analysis based on l_{min} .





Figure 16. Experimental results for trajectory of fall scenarios where blue line indicates *MA* of the person, red line shows the period of a fall event by v_{hw} , and orange point represents a falling point by l_{max} or l_{min} .

For Figure 16, the following explains fall scenarios obtained from the trajectories.

- Scenario 1: In the video scene, the person is walking from the right side to the chair near the window. Then, the person turns back and immediately falls down.
- Scenarios 2, 3 and 8: In the video scenes, the person immediately falls down while walking from right to left, falling sideways, forward, and backwards, respectively.
- Scenario 4: The person is standing near the window and then immediately falls while turning back.
- Scenario 5: The person is walking from the left and then immediately falls down.
- Scenario 6: The person walks from the right, and sits on a chair. While getting up from the chair, she immediately falls down.
- Scenario 7: The person walks from the left, stands near a table and walks toward a chair near the window. After that, he immediately falls on the bedsheet.

4.3. Performance Evaluation

To evaluate the performance of the proposed methods, 3-fold cross-validation was conducted in which variables for learning and testing were swapped. We here suppose that the abnormal $A = \{a_1, a_2, \ldots, a_{13}\}$, and the normal $N = \{n_1, n_2, \ldots, n_7\}$. We then classified abnormal into three groups: $A_1 = \{a_1, \ldots, a_4\}$, $A_2 = \{a_5, \ldots, a_8\}$, $A_3 = \{a_9, \ldots, a_{13}\}$, and also for normal: $N_1 = \{n_1, n_2\}$, $N_2 = \{n_3, n_4\}$, $N_3 = \{n_5, \ldots, n_7\}$. Then, we performed 3-trials with L_1, L_2, L_3 representing the learning process, and T_1, T_2, T_3 signifying the testing process for trials 1, 2 and 3, respectively. In trial 1, we assumed that L_1 was $(A_2 \cup A_3) \cup (N_2 \cup N_3)$ for learning, and T_1 was $A_1 \cup N_1$ for testing. In trial 2, L_2 was $(A_1 \cup A_3) \cup (N_1 \cup N_3)$ for learning, and T_2 was $A_2 \cup N_2$ for testing. In trial 3, L_3 was $(A_1 \cup A_2) \cup (N_1 \cup N_2)$ for learning and T_3 was $A_3 \cup N_3$ for testing. After computing the learning process for each trial, L_1, L_2 and L_3 , the detection rate was obtained by using the testing T_1, T_2 and T_3 , respectively. The overall accuracies of the system were finally computed for each of the features by applying *SVM* and *PD*.

There are four possible results in classifying abnormal and normal events, and the definitions and symbols are described as follows.

- Detected Abnormal (*P*₁₁): A video includes an abnormal event, and is correctly classified into class "Positive Abnormal."
- Undetected Abnormal (*P*₁₂): A video includes an abnormal event and is classified into class "Negative Normal."

- Normal (*N*₁₁): A video does not include an abnormal event, and is correctly classified into class "Negative Normal."
- Misdetected Normal (*N*₁₂): A video does not include an abnormal event, and is incorrectly classified into class "Positive Abnormal."

The precision, recall, and accuracy were used for evaluating performance, calculated as follows.

$$Precision = \frac{P_{11}}{P_{11} + N_{12}} * 100$$
(17)

$$\text{Recall} = \frac{P_{11}}{P_{11} + P_{12}} * 100 \tag{18}$$

Accuracy =
$$\frac{P_{11} + N_{11}}{P_{11} + N_{11}} * 100$$
 (19)

where accuracy was considered by including the undecided area. The concepts for calculating precision, recall, and overall accuracy are illustrated in Figure 17.



Figure 17. Illustration of performance evaluation from the classified abnormal and normal events. In this figure, (**a**) represents four different class labels, (**b**) represents the consideration for precision calculation $(P_{11}/(P_{11}+N_{12}))$, (**c**) represents the recall calculation $(P_{11}/(P_{11}+P_{12}))$, (**d**) represents the calculation of accuracy $(P_{11}+N_{11}/P_{11}+N_{11})$ where the undecided area is an empty set.

The detection precision achieved 93.33% by utilizing *SVM* and *PD* for each of the features: aspect ratio and point distance. The percentage of recall by applying *SVM* and *PD* were 100% and 93.33% for

the two features, respectively. Table 1 provides a comparison of our proposed detection method for abnormal and normal events versus existing approaches.

Methods	Precision	Recall	Accuracy
Charfi 2013 [24]	95.91%	92.15%	99.42%
Gunale 2018 [25]	95%	94%	95%
Suad 2019 [26]	100%	95.27%	99.82%
Ours (SVM)	93.33%	100%	100%
Ours (PD)	93.33%	93.33%	100%

Table 1. Comparison with existing approaches using the same dataset.

The proposed system was implemented in MATLAB 2018b on an academic license using C++. All of the experiments were performed on a Microsoft Windows 10 Pro with an Intel (R) Core (TM) i7-4790 CPU@3.60 GHz and 8GB RAM. Comparing runtime with existing systems is difficult, due to the various programming and optimization levels in use. The overall average computation time for our proposed system is 0.72 s per frame. We expect that implementation on a tuned GPU would be faster, and could provide real-time monitoring.

4.4. Comparative Studies of the Effectiveness and Limitations of the Proposed System

Charfi, 2013 [24] proposed an approach to detect falling events in a simulated home environment. The processes of the system are motion detection and tracking using background subtraction. The extracted binary image is used to construct the coordinates of the bounding box, aspect ratios, and ellipse orientation. Then, *SVM* and the Adaboost-based method are utilized to classify falls. This system is robust regarding the location changes and taking into account a tolerance on the instant of detection.

The proposed system by Gunale, 2018 [25] extracts four visual features: motion history image (*MHI*), aspect ratio, orientation angles using ellipse approximation, and the thresholds below the referenced line. After that, these features are inputted into five different machine learning algorithms (i.e., *SVM*, *K*-*NN*, *SGD*, *GB* and *DT*) to recognize falling problems. The *DT* provides the optimal result to confirm the effectiveness of the system. The limitations of the applied methods are not widely discussed and the deep learning models could be utilized to improve the detection rates in future work.

The system proposed by Suad, G. A, 2019 [26] investigated the effectiveness of motion information by using *tMHI*, the variations of shape which are fixed in the approximated ellipse around the silhouette body and the standard deviation of the difference for both horizontal and vertical histograms. Then, the neural network is applied to detect falls. The limitations are that the performance of the system depends heavily on multiple fixed thresholds. Thus, it is essential to judge the thresholds which are the best for detecting falls. In addition, these threshold parameters are needed to observe the adaptation for different persons in the monitoring system.

In our proposed system, background subtraction using *MOG* and foreground refinement using graph cut are performed to obtain the silhouette images with low loss of useful information. Then, the concept of *VGP* and the related visual features are properly presented to retrieve significant abnormal and normal action patterns. In the analysis component, we emphasize that detection of a falling point within the falling period depends on the modified difference of moving average. Finally, these features are put into *SVM* and *PD*. The proposed system shows the effectiveness of the system with high detection rates. The system scope is limited to attaining a real-time monitoring system due to the time-consuming object detection techniques that provide good foreground images. The current research works focuses on day-time visual abnormal and normal event detection. However, providing a 24-h service requires extending the monitoring period to include night-time monitoring. Moreover, a better understanding of the environment and of human–object interaction should be developed to create an improved home-care monitoring system.

5. Conclusions

The research work reported in this paper mainly focused on image processing technologies to assist in providing care for dependent persons, using a new approach for reliably detecting a fall or abnormal event. In brief, this approach uses an enhanced background subtraction method for object detection, as well as a simple and effective feature extraction method incorporating the new concept of a virtual grounding point. Moreover, a step-by-step approach is used for detailed considerations in correctly selecting useful features by employing moving average and difference calculus. Finally, abnormal and normal events are classified using a machine learning method employing a support vector machine and our proposed period detection for events. Experimental results indicate that the detection rate using the proposed approach achieves 100% with a low risk for error. However, our purpose is not only to detect abnormal events such as falls, but also the details of human behavior during normal activities by embedding an understanding of the environment. Future research in developing a robust home care monitoring system will be enhanced by considering when and where normal and abnormal activities occur.

Author Contributions: Conceptualization, S.N.N.H. and H.H.; methodology, S.N.N.H.; software, S.N.N.H.; validation, T.T.Z. and H.H.; formal analysis, S.N.N.H.; investigation, T.T.Z.; resources, S.N.N.H.; T.T.Z. and H.H.; 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 H.H.; visualization, S.N.N.H.; supervision, T.T.Z., S.N.N.H. preformed the 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.

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

References

- Bourennane, W.; Charlon, Y.; Bettahar, F.; Campo, E.; Esteve, D. Homecare monitoring system: A technical proposal for the safety of the elderly experimented in an Alzheimer's care unit. *IRBM* 2013, 34, 92–100. [CrossRef]
- 2. Karantonis, D.M.; Narayanan, M.R.; Mathie, M.; Lovell, N.H.; Celler, B.G. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 156–167. [CrossRef]
- 3. Zigel, Y.; Litvak, D.; Gannot, I. 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.* **2009**, *56*, 2858–2867. [CrossRef]
- 4. Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. Monocular 3D Head Tracking to Detect Falls of Elderly People. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 6384–6387. [CrossRef]
- 5. 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]
- 6. 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.
- Rougier, C.; Meunier, J.; 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 (AINAW07), Niagara Falls, ON, Canada, 21–23 May 2007; Volume 2, pp. 875–880. [CrossRef]
- 8. Tin, P.; Zin, T.T.; Hama, H.; Toriu, T. Challenges and Promises in Human Behavior Understanding Research. *ICIC Express Lett.* **2011**, *5*, 3761–3766.
- 9. McFarlane, N.J.; Schofield, C.P. Segmentation and tracking of piglets in images. *Mach. Vis. Appl.* **1995**, *8*, 187–193. [CrossRef]
- Stauffer, C.; Grimson, W.E.L. Adaptive background mixture models for real-time tracking. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 23–25 June 1999; Volume 2.

- Allili, M.S.; Bouguila, N.; Ziou, D. A Robust Video Foreground Segmentation by Using Generalized Gaussian Mixture Modeling. In Proceedings of the Fourth Canadian Conference on Computer and Robot Vision (CRV '07), Montreal, QC, Canada, 28–30 May 2007; pp. 503–509. [CrossRef]
- 12. Yong, H.; Meng, D.; Zuo, W.; Zhang, L. Robust Online Matrix Factorization for Dynamic Background Subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, 40, 1726–1740. [CrossRef] [PubMed]
- 13. Sobral, A.C. Robust Low-Rank and Sparse Decomposition for Moving Object Detection: From Matrices to Tensors. Ph.D. Thesis, Université de La Rochelle, La Rochelle, France, 11 May 2017; pp. 1–165.
- 14. Carsten, R.; Vladimir, K.; Andrew, B. GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314.
- Maerki, N.; Perazzi, F.; Wang, O.; Sorkine-Hornung, A. Bilateral Space Video Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 12 December 2016; pp. 743–751.
- 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. [CrossRef]
- Federolf, P.; Tecante, K.; Nigg, B. A holistic approach to study the temporal variability in gait. *J. Biomech.* 2012, 45, 1127–1132. [CrossRef] [PubMed]
- Alexandru, O.B.; Michael, J.B. An Adaptive Appearance Model Approach for Model-based Articulated Object Tracking. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 17–22. [CrossRef]
- Cao, Z.; Simon, T.; Wei, S.; Sheikh, Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1302–1310.
- 20. Connor, P.; Ross, A. Biometric recognition by gait: A survey of modalities and features. *Comput. Vis. Image Underst.* **2018**, *167*, 1–27. [CrossRef]
- 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 2020), Kyoto, Japan, 10–12 March 2020; pp. 249–250.
- Gunale, K.G.; Mukherji, P. Fall detection using k-nearest neighbor classification for patient monitoring. In Proceedings of the 2015 IEEE International Conference on Information Processing (ICIP), Pune, India, 16–19 December 2015; pp. 520–524.
- 23. Doulamis, N. Iterative motion estimation constrained by time and shape for detecting persons' falls. In Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments, New York, NY, USA, 23–25 June 2010; Volume 62, pp. 1–8. [CrossRef]
- 24. Charfi, I.; Miteran, J.; Dubois, J.; Atri, M.; Tourki, R. Optimized spatiotemporal descriptors for real-time fall detection: Comparison of support vector machine and adaboost-based classification. *J. Electron. Imaging* **2013**, *22*, 041106. [CrossRef]
- 25. 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.
- 26. Suad, G.A. Automated Human Fall Recognition from Visual Data. Ph.D. Thesis, School of Science and Technology, Nottingham Trent University, Nottingham, UK, February 2019; pp. 1–160.
- 27. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]
- 28. Le2i Fall Detection Dataset. Available online: http://le2i.cnrs.fr/Fall-detection-Dataset?lang=fr (accessed on 27 February 2013).



© 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/).