

Article

Bed-Exit Behavior Recognition for Real-Time Images within Limited Range

Cheng-Jian Lin ^{1,2}, Ta-Sen Wei ³, Peng-Ta Liu ³, Bing-Hong Chen ¹ and Chi-Huang Shih ^{1,*}

¹ Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung 411, Taiwan; cjlin@ncut.edu.tw (C.-J.L.); zejipa5218@gmail.com (B.-H.C.)

² College of Intelligence, National Taichung University of Science and Technology, Taichung 404, Taiwan

³ Fall Prevention Center, Department of Physical Medicine & Rehabilitation, Changhua Christian Hospital, Changhua 500, Taiwan; tase@ccch.org.tw (T.-S.W.); 105546@ccch.org.tw (P.-T.L.)

* Correspondence: chshih@ncut.edu.tw

Abstract: In the context of behavior recognition, the emerging bed-exit monitoring system demands a rapid deployment in the ward to support mobility and personalization. Mobility means the system can be installed and removed as required without construction; personalization indicates human body tracking is limited to the bed region so that only the target is monitored. To satisfy the above-mentioned requirements, the behavior recognition system aims to: (1) operate in a small-size device, typically an embedded system; (2) process a series of images with narrow fields of view (NFV) to detect bed-related behaviors. In general, wide-range images are preferred to obtain a good recognition performance for diverse behaviors, while NFV images are used with abrupt activities and therefore fit single-purpose applications. This paper develops an NFV-based behavior recognition system with low complexity to realize a bed-exit monitoring application on embedded systems. To achieve effectiveness and low complexity, a queueing-based behavior classification is proposed to keep memories of object tracking information and a specific behavior can be identified from continuous object movement. The experimental results show that the developed system can recognize three bed behaviors, namely off bed, on bed and return, for NFV images with accuracy rates of 95~100%.

Keywords: behavior recognition; images; bed exit



Citation: Lin, C.-J.; Wei, T.-S.; Liu, P.-T.; Chen, B.-H.; Shih, C.-H.

Bed-Exit Behavior Recognition for Real-Time Images within Limited Range. *Sensors* **2022**, *22*, 5495.

<https://doi.org/10.3390/s22155495>

Academic Editors: Shih-Chia Huang, Benjamin C. M. Fung, Cheng Zhang and Yan-Tsung Peng

Received: 31 May 2022

Accepted: 21 July 2022

Published: 23 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human behavior recognition is a process consisting of object detection and tracking and has become the core component of intelligent applications such as security surveillance, bed-exit monitoring, and fall detection [1]. According to the techniques used to identify the target object, object detection approaches can be divided into two categories: sensor and image. In sensor-based techniques, the object can be detected by means of either fixed or mobile modules. Fixed modules, such as infrared (IR) sensors [2,3], hybrid IR and pressure sensors [4], and radio frequency identification (RFID) [5,6], detect the object in a limited range. On the other hand, mobile modules based on accelerometers and gyroscopes can be integrated into wearables, which are attached to the detected objects [7,8]. Image-based techniques are typically combined with deep learning model to detect a target object after a data training process. The common source of images can be video sequences [9,10], IR thermal images [11], and integrated images from RGB and IR cameras [12,13]. Compared with sensor-based techniques, image-based techniques can better preserve the spatial information for object detection and have the potential to achieve multi-object detection.

Based on deep learning techniques, three main stages are included in an image-based behavior recognition process, namely object detection, object tracking, and behavior recognition. Three stages can be associated with their respective deep learning techniques. In the object detection stage, a convolution-based model is generally used to extract features of the target object from an image. Popular object detection models include You Only

Look Once (YOLO) [14] and Region-based Convolution Neural Network (R-CNN) [15,16]. Furthermore, the lightweight version of YOLO (e.g., Yolov3-tiny [17] or Yolov4-tiny [18]) is designed to realize real-time object detection in embedded systems and mobile devices. The object tracking stage utilizes the object detection results to monitor the object temporally and establishes a spatial track for the target object. The most well-known object tracking models are Simple Online and Real-time Tracking (SORT) [19] and its extended version, Deep SORT [20]. In SORT/Deep SORT, a Kalman Filter is adopted to predict and update the object track [21], while the Hungarian algorithm solves the matching problem between the predicted track and the current detected track in the presence of short-term object occlusion [22]. As the Kalman Filter aims to provide an object tracking solution based on a linearity assumption, particle filter solutions, such as Correlative Mean-Field (CMF) filter, can be used to track objects in a nonlinear system [23,24]. In addition to the typical CNN-based object detection and tracking models, the Plain Vision Transformer detector (ViTDet) uses a non-hierarchical backbone network for object detection purposes [25], and the Spike Transformer Network (STNet) performs single object tracking in an event-driven manner [26]. In the final stage, the spatial tracks recorded for target objects can correspond to a behavior, and a memory-based technique is required to recognize behaviors from the long-range dependencies of sequential data (i.e., spatial tracks). Typically, Recurrent Neural Network (RNN) [27] and Long Short-Term Memory (LSTM) [28] are useful models for predicting important events over time. Consequently, a behavior recognition system can be regarded as a combination of deep learning models with their respective data training, parameters, and computation.

This paper aims to design an image-based behavior recognition system for a bed-exit monitoring application. In the World Health Organization's (WHO) "Global Report on Falls Prevention in Older Age", almost 30~50% of people living in long-term care institutions fall every year, and 40% of them suffer from recurrent falls [29]. Other reports show that approximately one in three community-dwelling people aged over 64 falls every year [30–32]. Bed-exit monitoring can be the first line of defense in preventing the subsequent fall, injuries, and even death [33]. For caregivers, a well-designed bed-exit monitoring system is expected to reduce their additional burden, especially for patients with high fall risk [34]. In addition to the effectiveness of behavior recognition, efficiency is essential for a bed-exit monitoring system in the smart ward environment. Specifically, caregivers and medical institutions prefer a rapid system deployment without construction in the ward. The requirements of such rapid deployment include (1) installing, removing, and transferring the system as required; (2) staying connected with the smart ward system when installed; (3) a single-purpose application designed only for bed-exit monitoring to achieve fast personnel training and easy use in the institutions. These requirements imply that bed-exit behavior recognition should be done using a single device with mobility and wireless connectivity. The existing behavior recognition solutions rely on a complex system consisting of three stages. When the bed-exit monitoring device is required to be small and lightweight for rapid deployment, the complex system can impede real-time behavior recognition due to its complexity and overhead. Furthermore, data training for multiple stages can be a long process and this easily causes difficulties with follow-up system maintenance.

In this paper, a simplified behavior recognition solution designed for a low-cost embedded system is presented for bed-exit monitoring. As a branch of behavior recognition, bed-exit monitoring focuses on identifying bed-related behaviors, such as laying on the bed and leaving the bed, to prevent potential falls and the resultant injuries. To ensure efficiency, this simplified system considers the images captured by a narrow field of view (NFV). In NFV images, the object detection range is limited to a specific area (e.g., a bedside) or even to a part of the user's body in such a way that only some specific behaviors (e.g., on bed and bed exit) are monitored. Compared with wide field of view (WFOV), in which more spatial information can be obtained, NFV images have simpler scenes to facilitate the training and operation of the behavior recognition process. NFV images,

however, are associated with more object detection failures and abrupt scene changes can incur a negative effect on effectiveness. To achieve a tradeoff between effectiveness and efficiency, the proposed system develops an NFV-based behavior recognition process with the following stages: object detection, status classification, and behavior recognition. In the object detection stage, a highly efficient model, namely Yolov4-tiny, is adopted to attain real-time object detection in the embedded systems. Next, the object position information is retrieved and different activity statuses can be classified in accordance with the current position track. The behavior recognition stage utilizes a status classification queueing (SCQ) system to keep the memories of input statuses among pre-defined behavior queues. As a result, the negative effects, such as discontinued object detection and abrupt position change, are eliminated, and behavior can be identified according to a series of continued activities. Performance is evaluated via experiments conducted in the ward environment and time complexity analysis. Furthermore, a system prototype is developed in the experiments to capture NFV images with four camera angles: horizontal high/low and vertical high/low. From the performance results, the proposed system can achieve a real-time bed-exit behavior recognition with low complexity, and the results associated with horizontal angles outperform those of vertical ones.

This paper is organized as follows. Section 2 describes the concept of behavior recognition using NFV images, the proposed simplified behavior recognition system in embedded systems, and its bed-exit monitoring application. The experimental and analytical results are reported in Section 3. Section 4 gives the conclusions and future works.

2. Materials and Methods

2.1. Images with a Narrow Field of View

Figure 1 illustrates different object tracking ranges to monitor bed-related behaviors in the ward environment. A wide field of view tends to obtain more activity information by covering as wide an area as possible. Consequently, WFV images can contribute to the effectiveness of behavior recognition and are beneficial to tracking various behaviors such as bed exit and falling. In contrast, a narrow field of view aims to cover an area of interest where only the target is active and can obtain a simple scene for behavior recognition purposes.

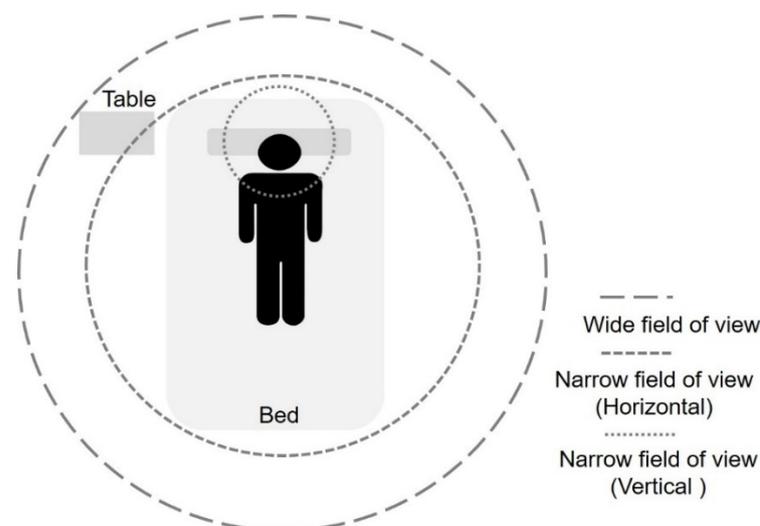


Figure 1. Object (human body) tracking range.

Figure 2 shows the NFV system prototype used in this paper. The main components consist of a primary device, a secondary device, and a mobile stand. The primary device is in charge of user input (buttons and microphone), user output (display and speaker), and communication (Wi-Fi and call bell socket). On the other hand, the secondary device is

connected to a camera and receives images to conduct the behavior recognition process. The recognition results are then sent to the primary device for the potential follow-up notification to the institution. Both primary and secondary devices are attached to the mobile stand. Accordingly, this prototype provides a post-installation feature and can be deployed or removed as required. Furthermore, the camera device can be adjusted as shown in Figure 2b,c to capture NFV images from various angles. Figure 3 presents images captured from four different angles: horizontal high, horizontal low, vertical high, and vertical low. Referring to Figure 1, the horizontal angles cover the bed region, while only a portion of the body is covered in the images captured by vertical angles. From Figure 3, it can be seen that NFV images with limited spatial information can impede object identification and tracking, and user activities become unpredictable.

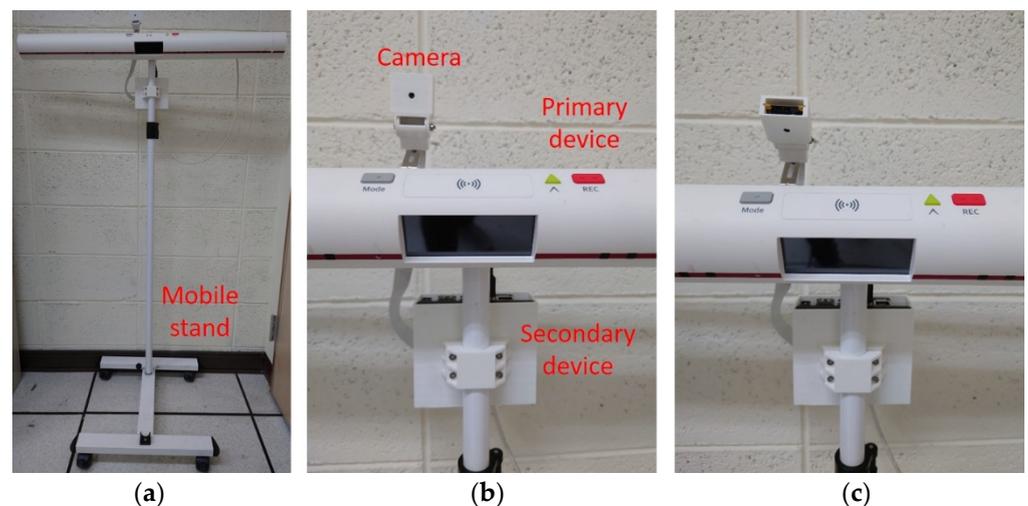


Figure 2. System prototype: (a) system with mobile stand; (b) camera capturing a horizontal view; (c) camera capturing a vertical view.

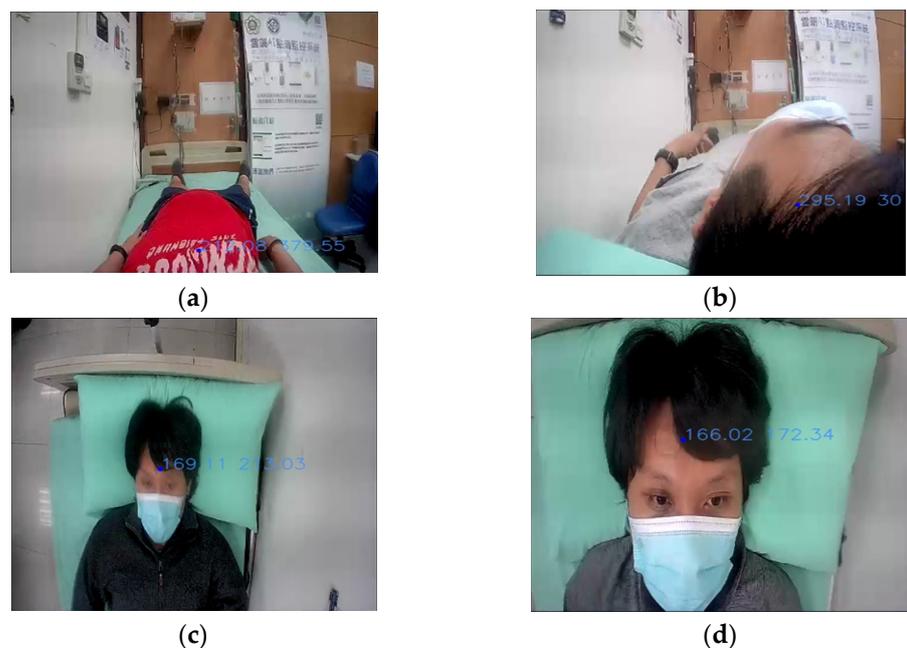


Figure 3. Images with four angles: (a) horizontal high; (b) horizontal low; (c) vertical high; (d) vertical low.

2.2. Behavior Recognition for NFV Images

In this paper, a three-stage behavior recognition process is developed for NFV images: object detection stage, status classification stage, and behavior recognition stage. The object detection stage aims to identify the application-specific objects from a series of images. In order to adapt to the narrow field of view, it is beneficial to detect multiple objects in accordance with the scenarios defined in the application. In bed-exit monitoring applications, for instance, the detected objects can be “head” and “body” of a target user. When the object is successfully detected, its coordinate in the image is generally labeled to indicate the position of the detected object. Based on the object detection results, the status classification stage can further track the object in accordance with its position changes, and an activity status is determined. Both object detection stage and status classification stage are performed to generate their respective outputs per image. The behavior recognition stage continuously monitors the status output to identify a user behavior within a specified duration.

Figure 4 presents a status classification queueing (SCQ) system proposed in this study to realize the behavior recognition stage for NFV images. The proposed SCQ system consists of four components: input queue, classifier, output queue, and controller. The input queue and output queue are data buffers with a queue structure. That is, the buffered data are processed in a first-in-first-out manner. The data unit is the activity status from the status classification stage. Each status result enters the input queue and is then assigned by the classifier to one among N output queues according to its status. That is, the status i is in the output queue i with a length L_i . Meanwhile, the controller drops one status from the head of its output queue to manage the number of buffered statuses in SCQ. The number of statuses, M , can be regarded as a status window size specified by a time slot of T seconds:

$$M = R \times T, \text{ subject to } M \leq \sum_{i=1}^N L_i \text{ and } M \geq \max\{L_i\}, \quad (1)$$

where R is the frame rate estimated in frames per second (fps). The queue length determines the recognition sensitivity for the behavior associated with it. Generally, a shorter queue length corresponds to a faster activity or a critical behavior requiring early notification. The controller continuously monitors N output queues. When one of them is full, the SCQ system can determine a user behavior associated with the full queue. Immediately after the behavior is identified, the controller empties all output queues to start a new recognition process. The next sub-section presents a bed-exit monitoring application based on this simplified behavior recognition process, including the proposed SCQ, to identify bed-related behaviors from NFV images.

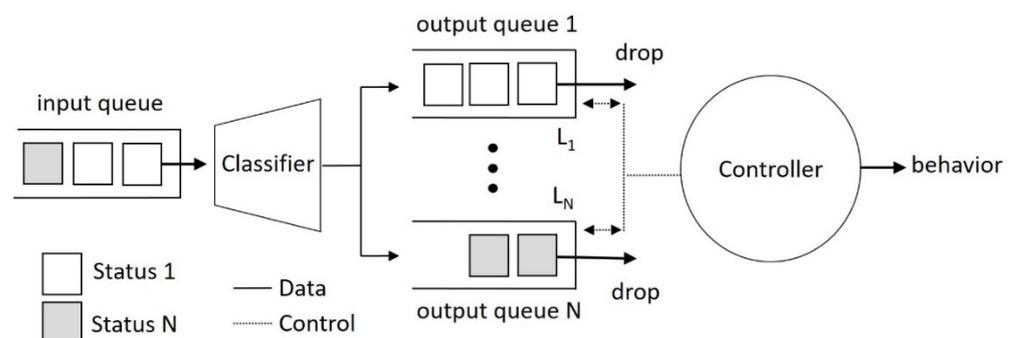


Figure 4. Status classification queueing system.

2.3. Bed-Exit Application

This section describes a bed-exit application based on a three-stage behavior recognition system, for NFV images captured by the prototype shown in Figure 2. In the object detection stage, Yolov4-tiny is adopted to identify the head and trunk of a monitored user. As the object is detected in an NFV image, a bounding box with the position information of

X and Y coordinates can be available to the status classification stage. Figure 5 shows the object detection results for NFV images captured from the horizontal low angle. Figure 5a–d indicates a series of bed-related activities including laying on the bed, turning over, getting up, and walking away from the bed. In Figure 5a–c, the detected objects are the head labeled in the pink bounding boxes, while Figure 5d detects the human trunk labeled in the green box.

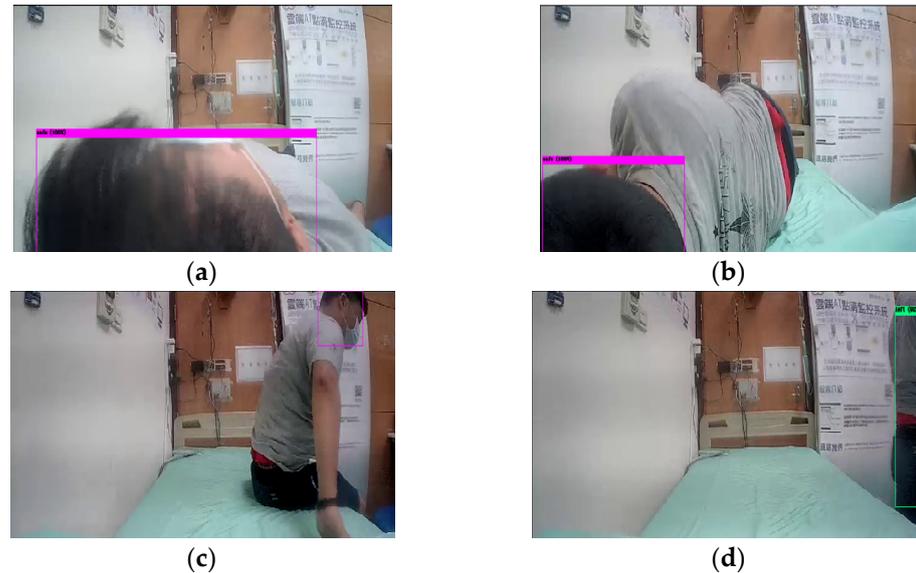


Figure 5. Object detection with bounding box in pink (a–c) and green (d).

Denote the X and Y coordinates centered in the bounding box as x_center and y_center , respectively. The center position of the bounding box is

$$P = \{x_center, y_center\}. \quad (2)$$

Given a time slot, W , in seconds, the number of images within W for an image sequence with R frames per second can be computed as

$$uF = W \times R. \quad (3)$$

Then the center position vector for bounding boxes of uF images is expressed as

$$FP = \{P_1, P_2, \dots, P_{uF-1}, P_{uF}\}. \quad (4)$$

In order to observe the FP changes, Figure 6 presents a real-time trace of the image sequence captured from the horizontal low angle. By inspecting the image sequence, we can differentiate user behaviors into four categories: laying on the bed, exiting the bed, nobody, and back to bed. These four behaviors are repeated twice and finally end with the “laying on the bed” behavior. Accordingly, a total of 9 behaviors are numbered from 1 to 9 in Figure 6. In the case of “nobody” behavior (numbered as 3 and 7 in Figure 6), the bounding box is absent and therefore no information about X and Y coordinates is available. In the behavior case labeled as 1 and 2 (i.e., laying on the bed and exiting the bed), however, three events labeled as A, B, and C are observed to have the same situation as “nobody” behavior. This is because the head and the trunk can be missed or be incomplete using a narrow field of view. Figure 7 shows the object tracking results for another image sequence captured from the vertical low angle. Following similar observations as in Figure 6, two object missing events occur in Figure 7: event A in the “back to bed” behavior numbered as 4, and events E–G in the “laying on the bed” behavior numbered as 9. Additionally, abrupt position changes are common for NFV images. In Figure 6, the “laying on the bed” behavior, numbered as 1, has

frequent position level fluctuations, especially for the X coordinate. In Figure 7, position changes become even more abrupt. From Figure 7, it can be seen that (1) for the behaviors numbered as 4, 5, and 6 (i.e., “back to bed”, “laying on the bed”, and “exiting the bed”), three behaviors are performed in order within a short duration; (2) events B~D exhibit fast XY curve fluctuation indicating that the detected object appears and disappears suddenly; (3) the “laying on the bed” behavior numbered as 9, has position level fluctuations. This implies that diverse user activities such as turning over and getting up exist when the target user is on the bed.

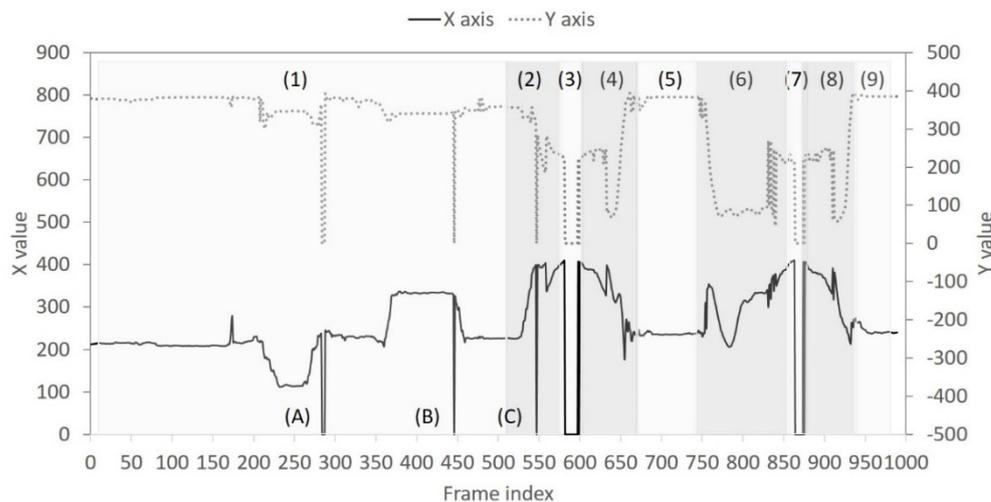


Figure 6. Real-time trace of XY coordinates for the bounding box from the horizontal low angle: numbers (1)~(9) stand for different statuses; letters (A)~(C) represent detection failure events.

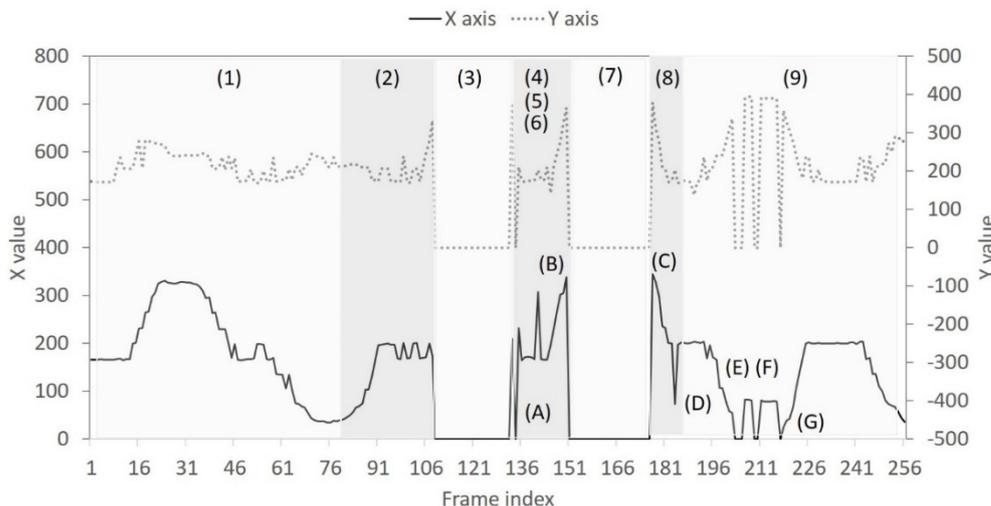


Figure 7. Real-time trace of XY coordinates for the bounding box from the vertical low angle: numbers (1)~(9) stand for different statuses; letters (A) and (E)~(G) represent detection failure events; letters (B)~(D) represent scene change events.

In the status classification stage, a status, \hat{S} , can be defined in accordance with the distance between two positions as follows:

$$\hat{S} = \begin{cases} 1, & \text{if } \overline{P_i P_0} - \overline{P_{i-1} P_0} \geq \epsilon, \\ 0, & \text{if } |\overline{P_i P_0} - \overline{P_{i-1} P_0}| < \epsilon, \\ -1, & \text{if } \overline{P_{i-1} P_0} - \overline{P_i P_0} \geq \epsilon, \\ -2, & \text{No bounding box,} \end{cases} \quad (5)$$

where ε is the distance threshold and P_0 represents an initial object position. It is noted that the origin of an image (i.e., (0,0)) is located at the top left point/pixel. Referring to Figure 5a, the user laying on the bed initially has a smaller value of the X coordinate and a larger value of the Y coordinate. During the period of exiting the bed, the value of the X coordinate is gradually increased, and the value of the Y coordinate becomes smaller. The status of $\hat{S} = 1$ indicates that the target user moves away from the bed since the distance between the current position and the initial position is larger than the distance between the previous position and the initial position by ε . On the contrary, the status of $\hat{S} = -1$ means that the user moves back to the bed when the distance between the current position and the initial position decreases from image $i-1$ to i . When the distance between two neighboring positions is smaller than ε , the intention for a user to move toward/back the bed stays neutral and the status \hat{S} is set to 0. Finally, the status of $\hat{S} = -2$ indicates that the bounding box has been missed. In Equation (5), the initial position (P_0) can be either a predefined value when the user remains stationary on the bed, or just set to the origin.

From the above-mentioned discussion, the status classified by a short-term position change may not correspond to the user behavior in the presence of abrupt position changes. In the behavior recognition stage, SCQ is utilized to differentiate the status associated with current behavior from that induced by abrupt position changes. In this study, four statuses defined in Equation (5) are injected into queues $\{Q_j\}$, where $j = 1\sim 4$. For $Q_1\sim Q_4$, the behaviors associated with their corresponding queues are "off bed", "on bed", "return", and "nobody", respectively. According to the queueing discipline of SCQ, a behavior is recognized when buffered statuses in their corresponding queue are full. Let the queue length of Q_j be L_j and given the number of statuses M , the behavior recognition of SCQ can be expressed as

$$\hat{E} = \begin{cases} \text{off bed, if } \sum_{k=i}^{i-M+1} \hat{S}_k = L_1, \\ \text{on bed, if } \sum_{k=i}^{i-M+1} (1 - \hat{S}_k) = L_2, \\ \text{return, if } \sum_{k=i}^{i-M+1} -\hat{S}_k = L_3, \\ \text{nobody, if } \sum_{k=i}^{i-M+1} -(\hat{S}_k/2) = L_4. \end{cases} \quad (6)$$

Based on (6), the statuses associated with a specified behavior can be accumulated to obtain an accurate observation of recent user activities. Meanwhile, the statuses representing abrupt position changes caused by a missed bounding box or object detection failure are omitted by means of putting them into other queues. All queues are emptied to restart another behavior recognition process after a behavior is determined.

3. Results

3.1. Experiment

In this study, a series of experiments were conducted in the ward environment. The system prototype presented in Figure 2 was deployed at the head of the bed, and four different camera angles were adopted to capture images with a frame rate of 15 frames per second and a resolution of 640×360 . The input images were then adjusted to 416×416 for object detection by Yolov4-tiny. Two participants, each of whom wore clothes in five different colors, were involved in the experiments to establish a data set of 10 video sequences for every camera angle. Consequently, a total of 40 sequences were available for data training and performance evaluation purposes. The data set had 1213, 1319, 1124, and 861 images for horizontal high, horizontal low, vertical high and vertical low angle, respectively. The image ratio used for data training and evaluation is 8:2. In our system prototype, an Nvidia Jetson Xavier NX with 384 CUDA core and 8G memory operates as the secondary device to realize the three-stage behavior recognition process presented in Section 2.2. In order to evaluate the timely recognition of NfV images, lengths of all queues and the number of buffered statuses (M) are fixed to 15 (images/statuses) in the SCQ. That means the minimum response time to recognize a behavior will be one second.

For NfV images, multi-object detection facilitates the subsequent object tracking and behavior recognition stages. As mentioned in Section 2.3, the main detected objects in this study are the head and human trunk. Typically, the detection results can be divided into

four classes: true positive (*TP*), false positive (*FP*), true negative (*TN*), and false negative (*FN*). Based on these detection classes, the detection accuracy (*A*) can be computed

$$A = \frac{TP + TN}{TP + TN + FP + FN}. \quad (7)$$

Table 1 presents the basic object detection results for different camera angles. In addition to the four NFV angles, the “diagonal high” angle is further considered as a WFV case for performance comparison purposes. Figure 8 shows the snapshots from a “diagonal high” video sequence. The data set of diagonal high angle has 2460 images and the ratio of data training to evaluation is 8 to 2. For horizontal angles, both the high and low angles can attain an accuracy rate of 85% and even higher. In the case of horizontal high angle, the successful detection rate for the human trunk is higher than that of the horizontal low angle because the higher angle can better capture the trunk as the user walks around the bed. For vertical angles, the accuracy rate of successfully detecting the head can be higher than 95%. The vertical low angle, however, fails to detect the human trunk as the user gets up from the bed. Compared with NFV images captured from horizontal and vertical angles, the diagonal high angle covers an area as wide as possible to obtain better object detection performance. Specifically, accuracy rates of 99% and 96% are attained for head detection and human trunk detection, respectively.

Table 1. Head and trunk detection results for different camera angles.

Camera Angle	Object	TP	FP	A
Horizontal High	Head	159	46	86%
	Trunk	103	5	93%
Horizontal Low	Head	149	48	85%
	Trunk	112	25	85%
Vertical High	Head	218	15	96%
	Trunk	67	20	85%
Vertical Low	Head	154	1	95%
	Trunk	x	x	x
Diagonal High	Head	1817	2	99%
	Trunk	763	39	96%

Table 2 shows the more detailed results of the object detection stage. For each class among TP, FP, and FN, the detection results of “head” and “trunk” are counted together to observe the object detection performance of five different camera angles. The detection precision (*P*), Recall (*R*), and F1-score (*F1*) are given in order by:

$$P = \frac{TP}{TP + FP}. \quad (8)$$

$$R = \frac{TP}{TP + FN}. \quad (9)$$

$$F1 = \frac{2 \times P \times R}{P + R}. \quad (10)$$

From Table 2, the vertical angles have higher values than horizontal angles in terms of *P*, *R*, and *F1-score*. This is based on a fact shown in Table 1, that the vertical angles can better detect the head as the user lays on the bed. Except for the vertical low angle, two horizontal angles (i.e., high and low) and a vertical high angle achieve nearly the same mean average precision (*mAP*) value. The vertical low angle has a lower *mAP* value of 49% due to the poor detection of the human trunk. Although the three angles with high

mAP values are able to support object detection for NFV images, the diagonal high angle achieves the highest values among all performance metrics $\{P, R, F1, mAP\}$.

Table 2. Object detection statistical results for different camera angles.

Camera Angle	P	R	F1	TP	FP	FN	mAP
Horizontal High	84%	85%	85%	262	51	45	90%
Horizontal Low	84%	85%	82%	261	73	45	88%
Vertical High	89%	92%	90%	285	37	24	90%
Vertical Low	99%	90%	95%	154	1	17	49%
Diagonal High	98%	98%	98%	2580	41	51	98%

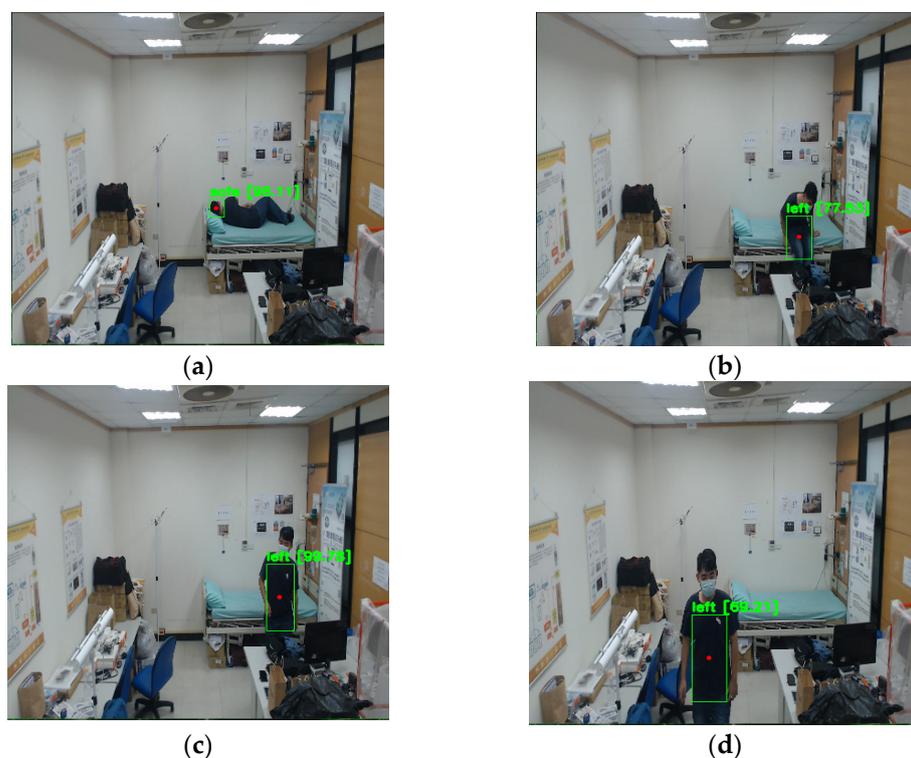


Figure 8. Images from the diagonal high angle: (a) lay on the bed; (b) get up and turn around; (c) sit on the edge of bed; (d) exit the bed.

As discussed above, NFV images can impede successful object detection and are expected to become an obstacle to behavior recognition accordingly. Obviously, the object detection failure would spread to the status classification stage where the object is further tracked and classified into different activity statuses. Consequently, the status track becomes more abrupt or discontinuous. The proposed SCQ aims to keep the memories of statuses based on a multi-queue architecture and predicts the user behavior labeled by a full queue. In the bed-exit monitoring application, we focus on three main user behaviors: on bed, off bed, and return to the bed. For each video sequence in our data set, the target user first lays on the bed, then walks away from the bed, and finally returns to the bed. This scenario is repeated once. Figure 9 shows the behavior recognition results for the video sequence presented in Figure 6. In order to ensure a good presentation quality, only the X-axis trace is plotted, and the three bed-related behaviors are expressed using behavior indexes. “on bed”, “off bed”, and “return” behaviors are numbered in order as 0, 1, and -1 . The “nobody” behavior is indexed as 2. In the behavior case numbered as 1 (i.e., laying on the bed), SCQ gives a behavior index of 0 in the presence of object detection failure

events A and B. In the end of behavior case numbered as 2 (i.e., exiting the bed), SCQ can successfully recognize “off bed” behavior with index = 1, while identifying “return” behavior (index = −1) in the beginning period of the behavior case numbered as 4 (i.e., back to bed). Similar results can be observed in behavior cases 6 and 8. In order to evaluate the corresponding performance for three bed behaviors, each video sequence is further divided into four “on-bed” clips, two “off-bed” clips, and two “return” clips. In the experiment, the amount of “on-bed” clips used to evaluate performance is $10 \times 4 = 40$. On the other hand, both the total amount of “off-bed” clips and “return” clips are the same as $10 \times 2 = 20$. The final behavior recognition results are presented in Table 3. From Table 3, the proposed SCQ can identify “on-bed” and “off-bed” behaviors with a successful recognition rate of 100% for all horizontal and vertical angles. In identifying “return” behavior, SCQ can achieve a successful recognition rate of 100% for horizontal angles, while obtaining a successful recognition rate of 95% for vertical angles. The images captured by vertical angles have a narrower field of view and fail to detect the object (head or trunk) for SCQ during a “return” behavior. Based on a similar deployment scenario, where the monitor device is installed at the head of the bed, the sensor-based work presented in [34] adapts an infrared array to differentiate the “off bed” behavior from body activities on the bed. According to the results reported in [34], an accuracy rate of about 99% is attained in recognizing “off bed” behavior. In addition to the fact that the proposed SCQ achieves an accuracy rate of 100% for “off bed” behavior, SCQ can recognize more bed-related behaviors, such as “on bed” and “return”, thanks to the better spatial information obtained by images. To conclude, SCQ is effective in identifying bed-related behaviors from NFV images and the horizontal angles are the preferred configuration in deploying an NFV-based bed-exit monitoring system.

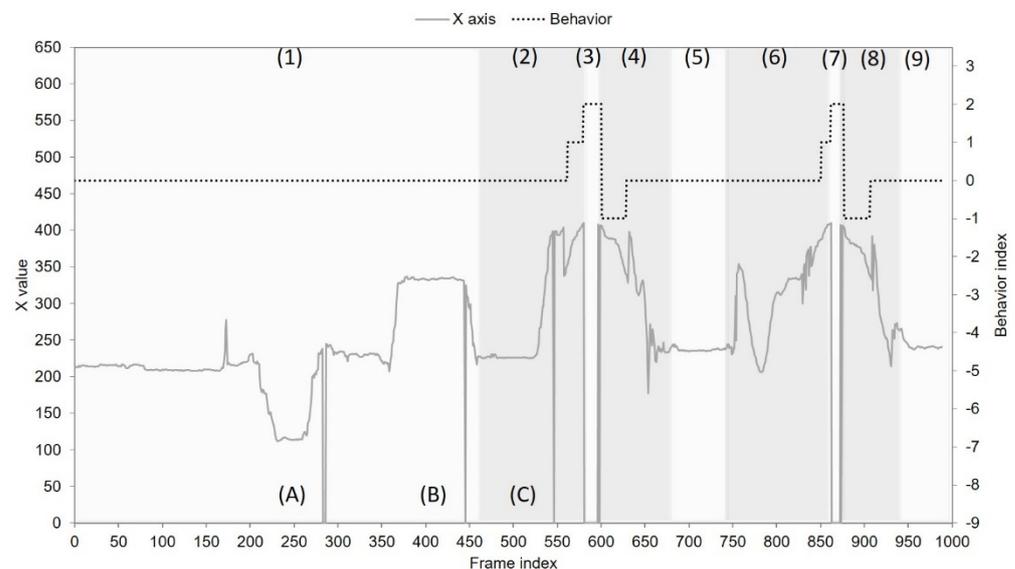


Figure 9. Behavior index trace from the horizontal low angle: numbers (1)–(9) stand for different statuses; letters (A)–(C) represent detection failure events.

Table 3. Behavior recognition statistical results for different NFV camera angles.

Camera Angle	On Bed	Off Bed	Return
Horizontal High	100% (40/40)	100% (20/20)	100% (20/20)
Horizontal Low	100% (40/40)	100% (20/20)	100% (20/20)
Vertical High	100% (40/40)	100% (20/20)	95% (19/20)
Vertical Low	100% (40/40)	100% (20/20)	95% (19/20)

3.2. Complexity Analysis

In this section, the time complexity of the proposed behavior recognition system is calculated and compared with that of the general one. For general behavior recognition systems, the object detection stage adopts a CNN-based method to find the features of the target object. The complexity of CNN is mainly caused by the convolution operation, the number of kernels, and the memory access [35]. The number of convolution operations per input grows quadratically with an increased kernel size. Accordingly, the complexity of d convolution layers can be estimated as

$$O\left(\left(\sum_{i=1}^d C_{i-1} \times S_i^2 \times F_i \times M_i^2\right) \times l \times e\right), \quad (11)$$

where C_{i-1} is the number of input channels of layer i , S_i is the spatial size of the filter, F_i is the number of filters of layer i , M_i is the spatial size of the output feature map for layer i , l is the input length, and e is the number of epochs. In the object tracking stage, some general systems, such as SORT or Deep SORT, use the advanced technique, to facilitate behavior recognition. In Deep SORT, the Kalman Filter and Hungarian Algorithm are two primary components used to obtain the spatial track from the object detection results. According to the work reported in [36], the complexity of the Kalman Filter in an iteration is given by

$$O\left(9 \times H^{2.367} + 10 \times H^2 + 5 \times H\right), \quad (12)$$

where H is the state size in the Kalman Filter. For the Hungarian Algorithm, its complexity was proved to be strongly polynomial, $O(n^4)$ [37], while Edmonds and Karp presented a faster version of the algorithm to achieve a running time of $O(n^3)$ [38].

In the behavior recognition stage, LSTM is considered a popular method of identifying user behavior based on sequential spatial data. The traditional LSTM contains memory blocks in the recurrent hidden layer, and each block is controlled by memory cells, gates, and peephole connections. Among LSTM gates, an input gate controls the flow of input into the memory cell, an output gate controls the present flow of output (i.e., weight) into the LSTM network, and a forget gate can reset the cell state. The peephole connection records the timing of output by linking gates in the same block. The idea behind LSTM is to allow temporal information to persist or be forgotten in its internal state. Generally, LSTM can be trained using the Back Propagation Through Time (BPTT) technique. For each time step, LSTM requires $O(1)$ per weight to obtain its computational results [28,39]. Denote w as the number of weights and the total complexity of LSTM becomes:

$$O(w \times l \times e). \quad (13)$$

In the proposed three-stage behavior recognition process, the object detection stage employs CNN-based Yolov4-tiny and obtains the same time complexity as (10). As shown in Equation (5), the status classification stage calculates the distance between two positions twice, and its complexity can be $O(1)$ based on the fact that the input size of the distance formula is fixed to two in each time step. Note that the advanced object tracking technique in general behavior recognition systems also involves distance computation, and the corresponding complexity can be ignored when compared to the large amount of computation contributed by the Kalman Filter and Hungarian Algorithm. Finally, the proposed SCQ performs an enqueue/dequeue operation or empties the queue in each time step, and its complexity is $O(1)$ accordingly. From the statistics from the experiment results, the three-stage behavior recognition process conducted for bed-exit monitoring can operate in NX at an average frame rate of 34.65 frames per second. An average value of 35.88 fps is attained when only the object detection stage operates. Consequently, the performance gap caused by the complexity of the status classification stage and the behavior recognition stage is 1.23 fps on average.

4. Discussion

This paper aims to establish an image-based behavior recognition system in a real-world ward to identify bed-related behaviors when dealing with practical issues such as fast deployment and system maintenance. This means that for a typical behavior recognition process, its data training and computational cost need to be limited so as to realize the bed-exit system on an embedded system. In order to achieve efficiency, images are captured with a narrow field of view to obtain a simple scene for behavior recognition. Furthermore, a simplified three-stage behavior recognition process is presented to ensure recognition effectiveness: (1) in stage 1, multi-object detection based on Yolov4-tiny is conducted to enhance the detection performance when using a narrow field of view; (2) in stage 2, the changes in object position are classified into different statuses to achieve simple object tracking; and (3) a queue-based behavior recognition approach, namely SCQ, is proposed in stage 3 to deal with the problems associated with a narrow field of view, such as discontinuous and abrupt position changes, by means of keeping the memories of object tracking statuses.

The experiment results show that the object detection of wide-area images outperforms that of images with a narrow field of view, and the proposed SCQ can efficiently recognize bed-related behaviors from NFV images. For images captured from horizontal angles, the successful recognition rate of 100% can be attained in terms of “on bed”, “off bed”, and “return” behavior. On the other hand, the successful recognition rate associated with the vertical angles is 100% for “on bed” and “off bed” behavior, and 95% for “return” behavior. In addition, the complexity analysis indicates that stages 2 and 3 in the proposed behavior recognition process are conducted with very low complexity, and therefore, the resultant bed-exit monitoring system can achieve real-time processing at nearly 35 fps on average using an Nvidia NX testbed.

Author Contributions: Conceptualization, C.-J.L., T.-S.W. and C.-H.S.; methodology, C.-H.S. and B.-H.C.; software, B.-H.C.; validation, C.-J.L., T.-S.W. and C.-H.S.; formal analysis, C.-H.S.; investigation, C.-J.L. and C.-H.S.; resources, T.-S.W. and P.-T.L.; data curation, B.-H.C. and P.-T.L.; writing—original draft preparation, B.-H.C.; writing—review and editing, C.-H.S.; visualization, C.-H.S.; supervision, C.-J.L. and T.-S.W.; project administration, C.-H.S.; funding acquisition, C.-J.L. and C.-H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Taiwan Ministry of Science and Technology, grant number MOST 110-2637-E-167-008.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to thank Gadgle Creative Tech. Ltd. for technically supporting the prototype implementation.

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

References

1. Aggarwal, J.K.; Ryoo, M.S. Human Activity Analysis: A Review. *ACM Comput. Surv.* **2011**, *43*, 1–43. [[CrossRef](#)]
2. Suzuki, R.; Otake, S.; Izutsu, T.; Yoshida, M.; Iwaya, T. Monitoring Daily Living Activities of Elderly People in a Nursing Home Using an Infrared Motion-Detection System. *Telemed. e-Health* **2006**, *12*, 146–155. [[CrossRef](#)] [[PubMed](#)]
3. Shin, J.H.; Lee, B.; Park, K.S. Detection of Abnormal Living Patterns for Elderly Living Alone Using Support Vector Data Description. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 438–448. [[CrossRef](#)] [[PubMed](#)]
4. Lu, C.; Huang, J.; Lan, Z.; Wang, Q. Bed Exiting Monitoring System with Fall Detection for the Elderly Living Alone. In Proceedings of the 2016 International Conference on Advanced Robotics and Mechatronics, Macau, China, 18–20 August 2016.
5. Chesser, M.; Jayatilaka, A.; Visvanathan, R.; Fumeaux, C.; Sample, A.; Ranasinghe, D.C. Super Low Resolution RF Powered Accelerometers for Alerting on Hospitalized Patient Bed Exits. In Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications, Kyoto, Japan, 11–15 March 2019.
6. Mori, Y.; Kido, S. Monitoring System for Elderly People Using Passive RFID Tags. *J. Robot. Mechatron.* **2014**, *26*, 649–655. [[CrossRef](#)]
7. Atallah, L.; Lo, B.; King, R.; Yang, G.Z. Sensor Placement for Activity Detection using Wearable Accelerometers. In Proceedings of the International Conference Body Sensor Networks, Biopolis, Singapore, 7–9 June 2010.

8. Vilas-Boas, M.C.; Correia, M.V.; Cunha, S.R.; Silva, P. Monitoring of Bedridden Patients: Development of a Fall Detection Tool. In Proceedings of the IEEE 3rd Portuguese Meeting in Bioengineering, Braga, Portugal, 20–23 February 2013.
9. Chen, T.; Hsiao, R.; Kao, C.; Liao, W.; Lin, D. Bed-exit Prediction based on Convolutional Neural Networks. In Proceedings of the 2017 International Conference on Applied System Innovation, Sapporo, Japan, 13–17 May 2017.
10. Chen, T.; Hsiao, R.; Kao, C.; Lin, H.P.; Jeng, S.; Lin, D. Vision-Assisted Human Motion Analysis for Bed Exit Prediction Model Construction. In Proceedings of the 2017 International Conference on Information, Communication and Engineering, Fujian, China, 17–20 November 2017.
11. Chiu, S.; Hsieh, J.; Hsu, C.; Chiu, C. A Convolutional Neural Networks Approach with Infrared Array Sensor for Bed-Exit Detection. In Proceedings of the 2018 International Conference on System Science and Engineering, New Taipei City, Taiwan, 28–30 June 2018.
12. Inoue, M.; Taguchi, R.; Umezaki, T. Bed-Exit Prediction Applying Neural Network Combining Bed Position Detection and Patient Posture Estimation. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Berlin, Germany, 23–27 July 2019.
13. Cao, Z.; Simon, T.; Wei, S.E.; 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.
14. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
15. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [[CrossRef](#)]
17. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *IEEE Trans. Pattern Anal.* **2018**, *15*, 1125–1131.
18. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-YOLOv4: Scaling Cross Stage Partial Network. *arXiv* **2020**, arXiv:2011.08036.
19. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016.
20. Wojke, N.; Bewley, A.; Paulus, D. Deep SORT: Simple Online and Realtime Tracking with a Deep Association Metric. *arXiv* **2017**, arXiv:1703.07402.
21. Kalman, R. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
22. Kuhn, H.W. The Hungarian Method for the Assignment Problem. *Nav. Res. Logist. Q.* **1955**, *2*, 83–97. [[CrossRef](#)]
23. Gao, J.; Tembine, H. Correlative Mean-field Filter for Sequential and Spatial Data Processing. In Proceedings of the IEEE International Conference on Smart Technologies, Ohrid, Macedonia, 6–8 July 2017.
24. Gao, J.; Tembine, H. Distributed Mean-Field-Type Filters for Big Data Assimilation. In Proceedings of the IEEE the International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems, Sydney, Australia, 12–14 December 2016.
25. Li, Y.; Mao, H.; Girshick, R.; He, K. Exploring Plain Vision Transformer Backbones for Object Detection. *arXiv* **2022**, arXiv:2203.16527.
26. Zhang, J.; Dong, B.; Zhang, H.; Ding, J.; Heide, F.; Yin, B.; Yang, X. Spiking Transformers for Event-Based Single Object Tracking. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21–24 June 2022.
27. Medsker, L.R.; Jain, L.C. Recurrent Neural Networks. *Des. Appl.* **2001**, *5*, 64–67.
28. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
29. World Health Organization. WHO Global Report on Falls Prevention in Older Age. Available online: <https://apps.who.int/iris/handle/10665/43811> (accessed on 31 May 2022).
30. Sherrington, C.; Fairhall, N.; Kwok, W.; Wallbank, G.; Tiedemann, A.; Michaleff, Z.A.; Ng, C.A.C.M.; Bauman, A. Evidence on Physical Activity and Falls Prevention for People Aged 65+ Years: Systematic Review to Inform the WHO Guidelines on Physical Activity and Sedentary Behavior. *Int. J. Behav. Nutr. Phys. Act.* **2020**, *17*, 1–9. [[CrossRef](#)] [[PubMed](#)]
31. Campbell, A.J.; Borrie, M.J.; Spears, G.F.; Jackson, S.L.; Brown, J.S.; Fitzgerald, J.L. Circumstances and Consequences of Falls Experienced by a Community Population 70 Years and Over During a Prospective Study. *Age Ageing* **1990**, *19*, 136–141. [[CrossRef](#)]
32. Tinetti, M.E.; Speechley, M.; Ginter, S.F. Risk Factors for Falls among Elderly Persons Living in the Community. *N. Engl. J. Med.* **1988**, *319*, 1701–1707. [[CrossRef](#)]
33. Peel, N.M.; Kassulke, D.J.; McClure, R.J. Population based Study of Hospitalised Fall related Injuries in Older People. *Inj. Prev.* **2002**, *8*, 280–283. [[CrossRef](#)]
34. Lin, C.J.; Shih, C.H.; Wei, T.S.; Liu, P.T.; Shih, C.Y. Local Object Tracking Using Infrared Array for Bed-exit Behavior Recognition. *Sens. Mater.* **2022**, *34*, 855–870. [[CrossRef](#)]
35. He, K.; Sun, J. Convolutional Neural Networks at Constrained Time Cost. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
36. Montella, C. The Kalman Filter and Related Algorithms: A Literature Review. ResearchGate. 2011. Available online: https://www.researchgate.net/publication/236897001_The_Kalman_Filter_and_Related_Algorithms_A_Literature_Review (accessed on 31 May 2022).
37. Munkres, J. Algorithms for the Assignment and Transportation Problems. *J. Soc. Ind. Appl. Math.* **1957**, *5*, 32–38. [[CrossRef](#)]

-
38. Edmonds, J.; Karp, R.M. Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems. *J. ACM* **1972**, *19*, 248–264. [[CrossRef](#)]
 39. Tsironi, E.; Barros, P.; Weber, C.; Wermter, S. An Analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for Gesture Recognition. *Neurocomputing* **2017**, *268*, 76–86. [[CrossRef](#)]