

Article

Simultaneous Indoor Tracking and Activity Recognition Using Pyroelectric Infrared Sensors

Xiaomu Luo ¹, Qiuju Guan ², Huoyuan Tan ¹, Liwen Gao ¹, Zhengfei Wang ¹ and Xiaoyan Luo ^{1,*}

¹ School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou 510000, China; woodwood2000@163.com or lxm@gzucm.edu.cn (X.L.); tanhuoyuan@gmail.com (H.T.); glw@gzucm.edu.cn (L.G.); wzf@gzucm.edu.cn (Z.W.)

² College of Mechanical and Electrical Engineering, Zhongkai University of Agriculture Engineering, Guangzhou 510000, China; qiujuguan@zhku.edu.cn

* Correspondence: lxy@gzucm.edu.cn; Tel.: +86-20-39358087

Received: 20 June 2017; Accepted: 25 July 2017; Published: 29 July 2017

Abstract: Indoor human tracking and activity recognition are fundamental yet coherent problems for ambient assistive living. In this paper, we propose a method to address these two critical issues simultaneously. We construct a wireless sensor network (WSN), and the sensor nodes within WSN consist of pyroelectric infrared (PIR) sensor arrays. To capture the tempo-spatial information of the human target, the field of view (FOV) of each PIR sensor is modulated by masks. A modified partial filter algorithm is utilized to decode the location of the human target. To exploit the synergy between the location and activity, we design a two-layer random forest (RF) classifier. The initial activity recognition result of the first layer is refined by the second layer RF by incorporating various effective features. We conducted experiments in a mock apartment. The mean localization error of our system is about 0.85 m. For five kinds of daily activities, the mean accuracy for 10-fold cross-validation is above 92%. The encouraging results indicate the effectiveness of our system.

Keywords: Wireless Sensor Network (WSN); Pyroelectric Infrared (PIR) sensor; random forest; simultaneously tracking and recognition

1. Introduction

With the population of the world increasing, the number of older people is growing inevitably. For personal comfort and due to limited medical resources, most of them live alone within their own house instead of nursing houses [1]. However, in their private space, emergency situations may not be noticed by others in time. For example, lying on the floor for a long time after a fall is one of the most dangerous situations. This will cause complications, and even death for the elderly [2]. Hence, how to assist them to live conveniently and safely has become an important social issue.

To achieve the automatic recognition of human daily activities for healthy aging, the methods proposed by scientists can be roughly divided into three categories [3]. The first category is based on vision sensors. Vision-based systems can monitor the entire scenario and capture the detailed movement of the human target [4,5]. However, because of the data association problem, it is challenging to handle the huge volume of vision data effectively [6]. Environmental factors such as occlusions and poor illumination conditions will deteriorate this problem. Besides, many people are uncomfortable living with cameras, which make them feel infringement on their privacy [7,8]. The second category is based on wearable sensors. Compared to vision sensors, the acceptance of wearable sensors is more preferable, and the volume of data to process is much less; there is also no data association problem [9,10]. However, the attachment of wearable sensors on the human body—even only one sensor—will feel obtrusive and uncomfortable to the resident [11]. What is more, people usually change their clothes daily and forget to attach the wearable sensors again or sometimes are not sufficiently clothed

to wear sensors when the indoor temperature is high [12]. Even after the careful design of the power management unit, the batteries inside the wearable sensors need to be recharged or changed regularly, which feels inconvenient for the users [13]. The third category is dense sensing-based [3]. Dense sensing refers to the deployment of numerous low-cost low-power sensors in the ambient intelligent environment. These sensors include microphones, vibration sensors, switch sensors, pressure mat sensors, etc. The interaction between the human and the object with sensors attached often provides powerful clues about the activity being undertaken. However, compared with wearable sensors, each of these dense sensors needs “fine tuning” after deployment, which means that they are hardly used ubiquitously [2]. To sum up, there is a great demand to “fill in the blanks” when these three categories of sensors are unsuitable for use in daily life.

Pyroelectric infrared (PIR) sensors are an excellent candidate for pervasive sensing. They are well accepted, because they appear in numerous places as part of security systems, including homes, banks, libraries, etc. [14]. They are inexpensive and can be attached to any indoor environment, which makes them “invisible” to the occupants. They also do not need to be worn or carried, which avoids the problems of forgetting to carry sensors and recharge batteries. They are a kind of passive infrared sensor; their performances are not affected changes of illumination [15]. However, for a single PIR sensor, its output is a raw sine-like signal and can only be used to detect whether or not human motion occurs. We have to carefully design the sensing paradigm and classification algorithm to develop their full potential.

In this paper, we propose an approach to extract and fuse the location information and motion information from the PIR sensor data stream simultaneously. To monitor indoor environments, we built a wireless sensor network (WSN). In the WSN, sensor nodes consist of pyroelectric infrared (PIR) sensors. The field of view (FOV) of each PIR sensor is modulated by a two degrees of freedom (DOF) segmentation, including bearing segmentation and circle segmentation, which provide the spatio-temporal information of the human target. The sensor nodes are attached to the ceiling; data fusion of adjacent sensor nodes will improve the localization accuracy. The speed of human locomotion could also be acquired. To achieve human activity classification, we propose a two-layer random forest (RF) classifier. Based on the location and moving speed of the human, the first layer of RF will label the activity type for each data frame. To boost the performance of our system, we incorporate prior knowledge of human activities. Because the duration of each kind of activity is a useful feature for activity classification, we employ a finite-state machine (FSM) to record the duration of the same activity for successive data frames. All of the features—including location, speed, and duration—will be input to the second-layer RF for final activity classification.

The main contributions of this paper are two-fold:

1. We propose a scalable framework that can decompose basic individual activities (“walking”, “lying”, “sitting”, “standing”, and “transitional activities”) into simple PIR data streams. With relatively low communication burden, our system could be expanded to cover any size indoor environment and fulfill the real-time processing. Non-intrusive PIR sensors are embedded in the indoor environment to achieve ambient intelligence, which can reduce the feeling of obtrusiveness to the minimum.
2. We propose a two-layer RF algorithm that leverages three simple yet powerful features (“location”, “speed”, and “duration of successive activity”) to achieve the recognition of human activities. Our approach is validated using data gathered from a mock apartment to make our results more confident. No human effort is needed to segment the monitored region for different activities.

2. Related Work

To avoid complicated data processing, especially feature extraction from the continuous vision sensor data stream [6], some researchers apply wearable or binary sensors instead for human daily activity recognition. Wilson et al. [16] proposed the simultaneously tracking and activity recognition of the occupants. Four kinds of binary sensors were employed to capture the human motion within each room. A dynamic Bayes net was used to infer the human location and achieve activity recognition

by fusing heterogeneous sensor data. The Rao–Blackwellised particle filter (RBPF) was employed to solve the data association problem. Zhu et al. [11] integrated the location information and motion information to infer the daily human activity. An optical motion capture system was installed on the corner of the ceilings to provide the human location information. The target human body had an inertial sensor attached to capture the human motion. Neural networks were used to achieve coarse granularity activity classification, and hidden Markov networks were utilized to refine the fine granularity activity classification result. Finally, the location and motion information was fused based on Bayes' rule.

Due to their simplicity and robustness to illumination variance, PIR sensors have recently been gaining increasing attention. In [17], Hao et al. proposed the use of side-view-based PIR sensor nodes to locate human targets. Within its FOV, each sensor node can detect the angular displacement of a moving human target; multiple sensor nodes enhanced the localization accuracy. They applied the same hardware setting for multiple human tracking [18]. Their sensor nodes were deployed to facilitate the data association problem. An expectation-maximization-Bayesian tracking scheme was proposed to enhance the system performance.

To avoid the region partitioning and region classifier, Yang et al. [19] proposed a special optical cone to model the FOVs of the PIR sensors into petal shapes. Intersections of detection lines formed by these petal shapes defined the measurement points, which would be assigned credits to represent the probability of the human target falling within the FOV. The data association problem of multiple human targets can also be addressed by this credit-based method after cluster analyzing.

However, in the research mentioned above, the PIR sensors were oriented in side-view or placed on the ground, which means that they were easily occluded by furniture or other obstacles in the real deployment. To overcome this drawback, Tao et al. [20] attached binary infrared sensors to the ceiling of an office. Weak evidence such as people location, moving direction, and personal desks was synthesized to achieve soft tracking. They declared that their system can track up to eight persons with high accuracy. To increase the space resolution and improve the deployment efficiency of sensor nodes, Luo et al. [21] applied another scheme of FOV modulation to realize human indoor localization. In their system, the FOVs of PIR sensors were modulated by two degrees of freedom (DOF) of spatial segmentation, which provided the flexible localization schema for information fusion. The Kalman filter and smoother were utilized to refine the human motion trajectory.

Some researchers have been devoted to exploiting the potential of PIR sensors for activity recognition. In [22], Liu et al. proposed the employment of pseudo-random code, based on compressive sensing, to modulate the FOVs of the PIR sensors. The human activity within a confined region was cast into low-dimensional data streams, and could be classified by the Hausdorff distance directly. Luo et al. [23] proposed a method for abnormal behavior detection by investigating the temporal feature of the sensor data stream. Modified Kullback–Leibler (KL) divergence accompanied with self-tuning spectral clustering were leveraged to profile and cluster similar normal activities. Feature vectors were formed by hidden Markov models (HMMs). Finally, one-class support vector machines (SVMs) were employed to detect abnormal activities.

Guan et al. [24] employed PIR sensors to capture the thermal radiation changes induced by human motion. Three sensing nodes were utilized to construct a multi-view motion sensing system, including one ceiling-mounted node and two nodes on tripods facing each other. HMMs and SVMs were employed to classify six types of activities.

In summary, the above-mentioned research of PIR sensors focused on human localization or human activity recognition separately. This paper endeavors to provide a framework to address these two synergy problems simultaneously.

3. Sensor Node Design

The sensor node is the foundation and basic element of WSN. In this section, we will introduce the rationale behind the design of the sensor node, including the sensing model design and sensor node implementation.

3.1. Sensing Model

Our sensing model originated from the reference structure tomography [25]. The illustration of our sensing model is shown in Figure 1. The monitored space is the space where the human target moves and performs different activities. The measurement plane is the plane where the PIR sensors are located. The reference structures are located between the monitored space and measurement plane and are used to modulate the visibility of the sensor array.

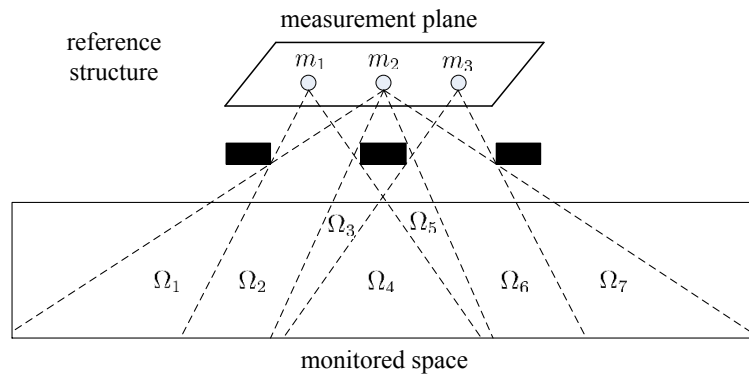


Figure 1. Our sensing model: reference structure, measurement plane, monitored space. $\Omega_1 \dots \Omega_7$ are sampling cells. Each sampling cell has the same visibility to the sensor array.

Assume that there are M PIR sensors in the measurement plane, and the monitored space can be divided into L sampling cells. The visibility of the same sampling cell to the sensor array is identical [26]. Thus, the output of the PIR sensor array can be represented in vector form as:

$$\mathbf{m} = \mathbf{V}\mathbf{s} \quad (1)$$

where $\mathbf{m} = [m_j(t)] \in \mathbb{R}^{M \times 1}$ is the output vector of the PIR sensor array and $m_j(t)$ the output of the j th PIR sensor. $\mathbf{s} = [s_i(t) * H(t)] \in \mathbb{R}^{L \times 1}$ is the radiation status vector of the sampling cells; $H(t)$ denotes the impulse response function of the PIR sensor; $*$ is the convolution operator; $\mathbf{V} = [v_{ji}] \in \mathbb{R}^{M \times L}$ is the measure matrix. If the i th sampling cell is visible to the j th sensor, $v_{ji} = 1$; $v_{ji} = 0$ otherwise.

Thus, the thermal variation of the monitored space is represented by the output of the PIR sensor vector. The reference structure plays the key role of radiation embedding; the spatio-temporal characteristics of the human target activities are cast into the low-dimensional sensor data stream. Because the PIR sensors can only detect the movement of thermal objects, non-thermal objects such as balls or infrared remote controllers will not trigger the PIR sensors.

3.2. Reference Structure Specification

The sensor node design is based on the sensing model mentioned above. In our design, there are nine PIR sensors on one sensor node, as shown in Figure 2. To enhance the sensibility of the PIR sensor array, each PIR sensor is equipped with a hemisphere Fresnel lens array. Before visibility modulation, the FOVs of all of the sensors are multiplexed, forming a cone-shaped monitored space. The opaque mask plays its role as the reference structure. There are two types of masks in our design for FOV segmentation; they are used for bearing segmentation and radial segmentation, respectively.

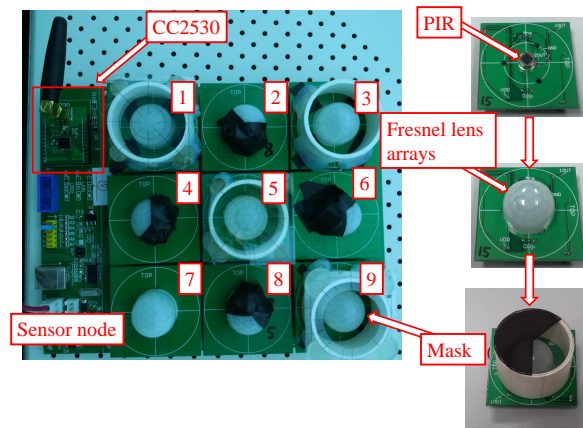


Figure 2. The pyroelectric infrared (PIR) sensor node. The sensor node consists of nine PIR sensors arranged in a grid shape. Four of them using the Type I mask, the rest using the Type II mask. CC2530 is used to sample the PIR signals and communicate with the sink node.

The Type I mask is used for bearing segmentation, as shown in Figure 3a. In our node design, four out of nine sensors are equipped with the Type I mask. The sweep angle ϕ of the FOV is 120° , as shown in Figure 3b. All of the FOVs of these four PIR sensors are overlapping and then are rotated 90° one by one. As a result, such a multiplexing scheme segments the monitored region of the sensor node into eight sampling cells. The coding scheme of corresponding angle θ for each sampling cell is listed in Table 1.

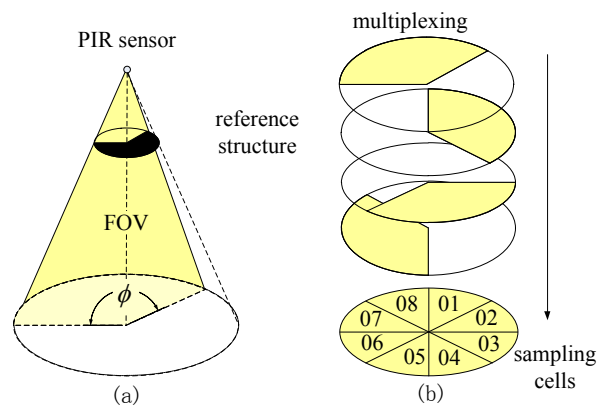


Figure 3. Type I mask: (a) bearing segmentation; (b) multiplexing of four PIRs forms eight sampling cells. FOV: field of view.

Table 1. The coding scheme of angle θ .

Sampling Cells	PIR 6	PIR 2	PIR 4	PIR 8	Angle θ
01	1	0	0	1	67.5°
02	1	0	0	0	22.5°
03	1	1	0	0	337.5°
04	0	1	0	0	292.5°
05	0	1	1	0	247.5°
06	0	0	1	0	202.5°
07	0	0	1	1	157.5°
08	0	0	0	1	112.5°

The Type II mask is used for radial segmentation, as shown in Figure 4a. After being masked, the sensing region of each PIR sensor is still a full cone, but the cone angle β is modulated by the radius of the mask. There are five sensors on the sensor node equipped with the Type II mask. The multiplexing of these five sensors segments the monitored region of the sensor node into five sampling cells, which correspond to different radii r . The coding scheme of radius r is listed in Table 2, where the preferential coding strategy is employed.

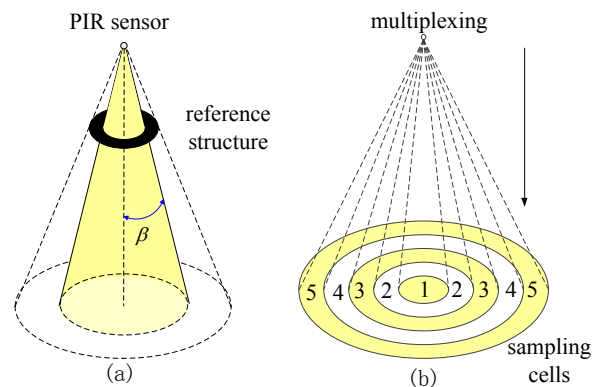


Figure 4. Type II mask: (a) radial segmentation; (b) multiplexing of five PIRs forms five sampling cells.

Table 2. The coding scheme of radius r .

Sampling Cells	PIR 1	PIR 5	PIR 9	PIR 3	PIR 7	Radius r (m)
1	1	x	x	x	x	0.0
2	0	1	x	x	x	0.8
3	0	0	1	x	x	1.3
4	0	0	0	1	x	2.3
5	0	0	0	0	1	3.0

“x” denotes 0 or 1.

By integrating these two kinds of masks, the monitored space is segmented into several sampling cells, as shown in Figure 5a. The center of the sampling cell (x, y) is represented by polar coordinates r/θ , as shown in Figure 5b:

$$\begin{cases} x = r \times \cos \theta \\ y = r \times \sin \theta \end{cases} \quad (2)$$

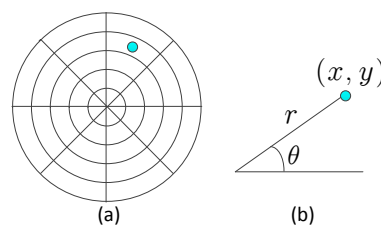


Figure 5. (a) The floor projection of the sampling cells; the segmentation of the monitored space within one sensor node. (b) The center of the cell; the blue point is denoted by polar coordinates.

When the human object moves within the monitored space, several PIR sensors will be triggered. According to the coding scheme, we could decode the location of the human. The center of the

sampling cell triggered by the human will be regarded as the position estimation of the human target. In other words, the accuracy of one sensor node is related to the size of the sampling cell.

4. Localization

4.1. Signal Feature Extraction

As the impulse response function of the PIR sensor is not linear, the output of the PIR sensor $m_j(t)$ is a sine-like waveform, as shown in Figure 6. Because the human target is not rigid and deformable, the $m_j(t)$ is not a smooth curve. Thus, it is not suitable to use the amplitude of the signal directly as the feature for localization.

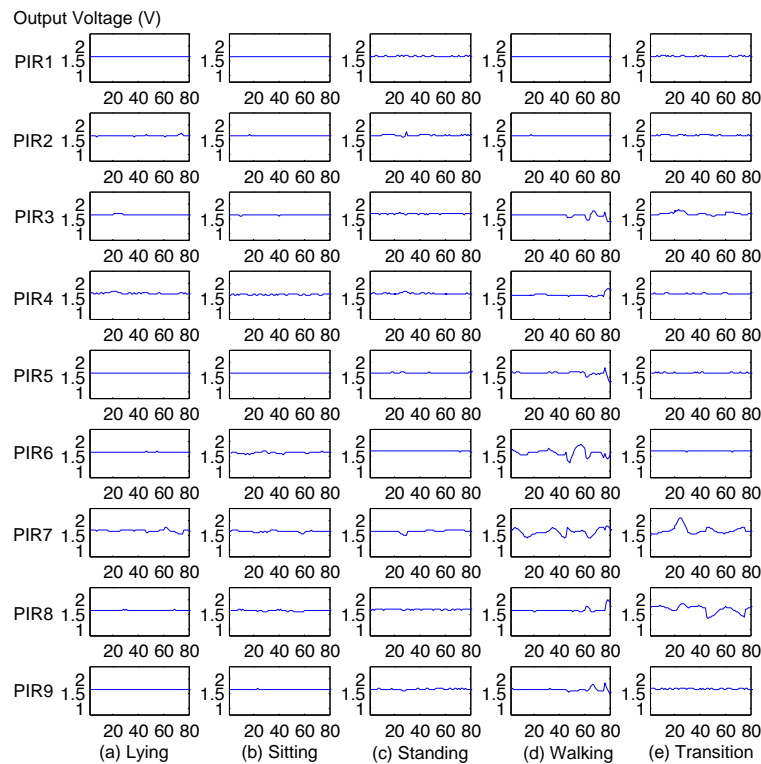


Figure 6. The output voltage of nine PIR sensors in one sensor node for five kinds of activities: (a) lying; (b) sitting; (c) standing; (d) walking; and (e) transitional activities.

Even though the outputs of PIR sensors are not stationary, their energy is stable for a short period of time. In the speech recognition domain, short time energy (STE) is widely used for discriminating voiced and unvoiced segments for speech signals [27]. Inspired by its effective performance, we employ STE as the signal feature to classify whether the PIR sensors are fired. For the j th PIR sensor, the n th window of STE is defined as:

$$p_j(n) = \sum_{k=0}^{Z-1} |m_j(k) - avSTE_j(n)| \quad (3)$$

$$\text{with } avSTE_j(n) = \frac{1}{Z} \sum_{k=0}^{Z-1} m_j(k) \quad (4)$$

where $m_j(k)$ is the output voltage (V) of the k th sampling point and $avSTE_j(n)$ is the average voltage of all sampling points within the n th window. In each window, there are Z sampling points. In our system deployment, $Z = 15$, as shown in Figure 7.

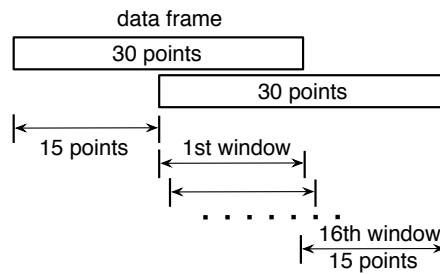


Figure 7. Data frame and window. The size of each data frame is 30 sampling points, and the size of each window is 15 sampling points. The overlap between two successive data frames is 15 sampling points.

Certain threshold th_j is set for $p_j(n)$ to indicate whether or not the PIR is triggered:

$$m_j(n) = \begin{cases} 1 & p_j(n) > th_j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Based on the output of the PIR sensor array $\mathbf{M}(n) = [m_1(n), \dots, m_9(n)]$, we lookup Tables 1 and 2 to figure out which sampling cells are fired and calculate the location estimation of the human target according to Equation (2). Because the output of the PIR sensor is proportional to the surface size and moving speed of the thermal object, the movement of a domestic animal will not trigger the PIR sensors by setting the appropriate threshold for each PIR sensor [28].

4.2. Data Fusion

In our deployment, there are five sensor nodes attached on the ceiling, forming a WSN as shown in Figure 8. For the k th window, $\mathbf{z}_k = (x_k, y_k)$ is defined as the position estimation of the human target. The monitored space of each sensor node is overlapped to enhance the localization precision by data fusion. The data fusion strategy is based on the number of sampling cells triggered within the same time window:

- Two sampling cells: the midpoint of these two centers is regarded as \mathbf{z}_k , as shown in Figure 9a;
- More than two cells: the maximum likelihood estimation algorithm is used to figure out \mathbf{z}_k , as shown in Figure 9b.

However, this strategy will fail in some situations, as shown in Figure 9c. The distances between \mathbf{z}_k and three triggered sampling cells are the same, but this is obviously not correct. Thus, we have to judge the relationship between \mathbf{z}_k and the sampling cells: if \mathbf{z}_k is outside the convex region formed by the centers of sampling cells, the geometric center of these cell centers will be the \mathbf{z}_k estimation.

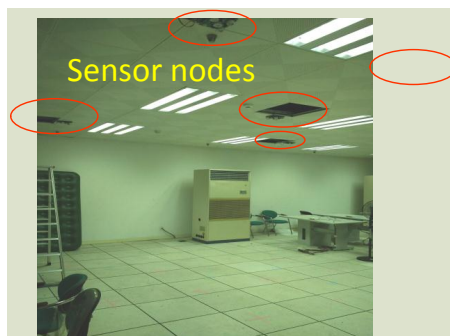


Figure 8. Deployment of the sensor nodes. The red circles are the locations of sensor nodes.

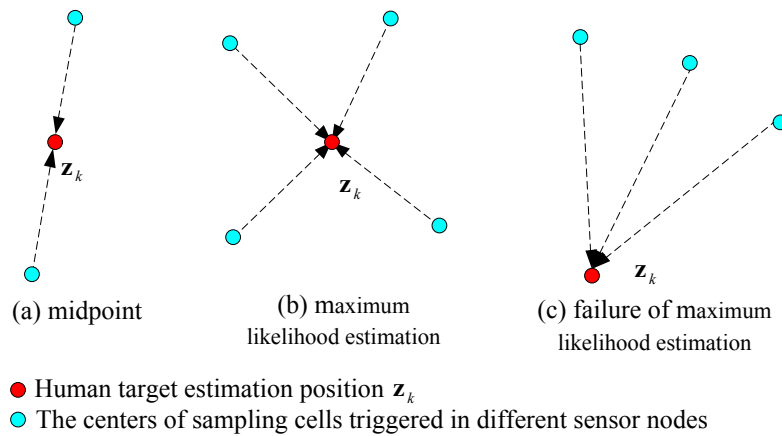


Figure 9. Data fusion strategy of adjacent sensor nodes.

4.3. Particle Filter

In our system, the sampling rate of the PIR sensors is 15 Hz. The data stream will be segmented into data frames before further processing, as shown in Figure 7. Each data frame is 30 sampling points, which is about two seconds. The overlap between two successive data frames is 15 sampling points. In each data frame, the localization algorithm is applied to consecutive windows. The size of the window is 15 sampling points. As a result, there will be a maximum of 16 localization results in each data frame.

Assume that there are L localization results in the t th data frame, denoted as $\mathbf{y}_t^l = (x_t^l, y_t^l)$, where $l = 1, \dots, L$. The previous location of the human target is denoted as $\mathbf{Z}_{t-1} = (x_{t-1}, y_{t-1})$. Based on the idea of the particle filter (PF) [29], the weight of the l th sample can be defined as:

$$w_t^l = p(\mathbf{y}_t^l | \mathbf{Z}_{t-1}) \quad (6)$$

In our system model, we assume that the probability of current location \mathbf{y}_t^l given the previous location \mathbf{Z}_{t-1} is two-dimensional Gaussian, then Equation (6) can be rewritten as:

$$w_t^l = N(\mathbf{y}_t^l | \mathbf{Z}_{t-1}, \sigma) \quad (7)$$

where σ is the variance of the observation model. γ is the threshold to validate \mathbf{y}_t^l . If $w_t^l < \gamma$, \mathbf{y}_t^l will be discarded. If w_t^l is too small, this means that the probability of \mathbf{y}_t^l generated by a false alarm is high. After discarding invalid localization results, $L \leq 16$.

Assume there are total L valid localization results in the same data frame; the weight of each result should be normalized as follows:

$$\tilde{w}_t^l = \frac{w_t^l}{\sum_{l'=1}^L w_t^{l'}} \quad (8)$$

Thus, the current location of the human target can be represented as:

$$\mathbf{Z}_t = \sum_{l=1}^L \mathbf{y}_t^l \tilde{w}_t^l \quad (9)$$

where \tilde{w}_t^l is the normalized weight of \mathbf{y}_t^l .

5. Two-Layer Random Forest

The design of the machine learning algorithm is the key consideration of our system. The algorithm must be able to incorporate heterogeneous features such as location, speed, etc.

5.1. Random Forest

In our system, random forest (RF) is the basic classifier. For many multi-class recognition tasks, RF has shown its effectiveness [30,31]. The RF consists of a number of decision trees. Each tree is developed from bootstrap samples from the training data [32]. When constructing individual trees, a random subset of input features is selected. Based on these features, the trees will be split when the largest information gain is achieved, as shown in Figure 10. Each decision tree is grown to the largest extent without pruning. The number of decision trees and the number of input features are the most important variables [33].

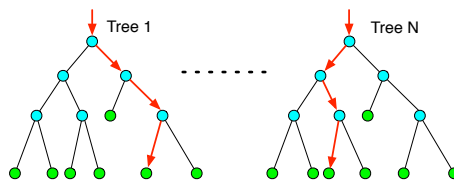


Figure 10. Random forest (RF). The RF consists of a number of decision trees. After splitting, each tree contains green leaf nodes and blue split nodes. For a particular input, the red arrows indicate the different paths from the root node to the leaf nodes along different trees. The final result is based on the maximum voting among all of the trees.

For classification, the final result of the forest is based on the maximum voting among all of the decision trees. In general, the random forest algorithm is an ensemble classifier having a fast training time and very high generalization accuracy without special feature selection [34].

5.2. Overview of Data Processing

The overview of data processing of our system is shown in Figure 11. The sensor data stream generated from all of the sensor nodes is segmented into data frames, as shown in Figure 7. In each data frame, we calculate the location estimation of the human target according to Equation (9). Based on the distance between the locations of two successive data frames, (x_t, y_t) and (x_{t-1}, y_{t-1}) , the moving speed V_t of the human target in the t th data frame is calculated as:

$$v_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} / T \quad (10)$$

where T is the time interval between two data frames. Then, the vector $\{x_t, y_t, v_t\}$ is used as the input feature for the first layer RF; the output of the first layer RF is denoted as L_t^1 .

To acquire the duration of the same successive activity, a two-state finite state machine (FSM) is used to indicate the change of L_t^1 . Whenever L_t^1 is changed, the duration of the same successive activity is known.

$$\begin{cases} C_t = C_{t-1} + 1, & \text{if } L_t^1 = L_{t-1}^1 \\ C_t = 1 \quad \text{and} \quad D_{t-1} = C_{t-1}, & \text{if } L_t^1 \neq L_{t-1}^1 \end{cases} \quad (11)$$

where C_t is the duration counter of data frame t and D_t is the duration of the consecutive activity.

For example, data frame $\{m, m+1, \dots, t\}$ is the same activity classified by the first-layer RF; that is, $L_m^1 = L_{m+1}^1 = \dots = L_t^1 \neq L_{t+1}^1$, then $D_m = D_{m+1} = \dots = D_t = C_t$.

Based on the output of the first-layer RF, we construct the feature vector $\{x_t, y_t, v_t, D_t\}$ for the second-layer RF. The output of the second RF is the final classification of our system.

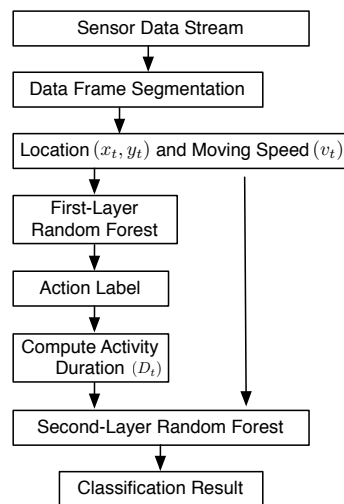


Figure 11. The overview of data processing. The feature vector for the first-layer RF is $\{x_t, y_t, v_t\}$. The feature vector for the second-layer RF is $\{x_t, y_t, v_t, D_t\}$, where (x_t, y_t) is the location of the human target at the t th data frame, v_t is the moving speed, and D_t is the activity duration time.

6. Experiments

6.1. Environmental Setup

We conducted experiments in a mock apartment, as shown in Figure 8. The monitored space of each sensor node is approximately a cone with a 3-m radius. There are in total five sensor nodes attached on the ceiling 3 m above the ground, forming a star topology WSN [35]. The positions of these five sensor nodes are $(2,2)$, $(2,-2)$, $(-2,-2)$, $(-2,2)$, and $(0,0)$, respectively. The sensor nodes can be regarded as being located on four corners and the center of a square. Such a deployment is based on the consideration that the sensor nodes should cover as much of the monitored region while having some region overlap to enhance localization accuracy. The monitored region of our WSN is $6\text{ m} \times 6\text{ m}$. The sampling rate of each PIR sensor is 15 Hz. The sensor data collected by the CC2530 on sensor nodes will be sent to the sink node based on the ZigBee (IEEE 802.15.4) protocol with a 250-kbps data rate [36]. The sink will transport the data to the PC host for further data processing.

We collected 67 datasets from three volunteers. The first volunteer is a female, age 23, 42 kg, 160 cm; the second volunteer is a male, age 22, 64 kg, 169 cm; the third volunteer is a male, age 37, 174 cm, 70 kg. Each volunteer performed five kinds of daily activities: walking, sitting, standing, lying and transitional activities. The transitional activities include: sit-to-lie, lie-to-sit, sit-to-stand, stand-to-sit, etc. The duration of each dataset is about four minutes. There are a total of 16,328 data frames for all datasets. Our system does not need to be re-parameterized for different human subjects. The same system configuration is able to recognize the activities performed by volunteers with their own style. We used a web cam to record the process of experiments as the ground truth.

6.2. Recognition Result

In our experiments, the accuracy was calculated based on the classification result of each data frame, as shown in Figure 7. The recognition result was compared with the ground truth labeled manually for every second according to the video. Through a random partition, 90% of the datasets were selected as the training set, and 10% of the datasets were selected as the testing set. We employed 10-fold cross-validation (CV) to evaluate the performance of our system. In each CV, the training set included 14,841 data frames, and the testing set included 1487 data frames.

Some typical frames of video and PIR data frames are shown in Figure 12. The right column of each subfigure is the snapshot captured from the video. The middle column is the human moving trace projected on the floor. The mean localization error of our system is about 0.85 m. The red square

is the region of chairs. The green square is the location of the bed, where sitting, lying, and transitional activities may occur. The left column is the ground truth labels of each second, the classification results of the first-layer RF and second-layer RF, respectively.

In the left column of each subfigure, labels for the ground truth and activity classification results are as follows: 1. lying, 2. sitting, 3. standing, 4. walking, and 5. transitional activities. In Figure 12a, the human subject stands up and then walks towards the bed. The activities mainly include sitting and walking. In Figure 12b, she walks around and then sits down on the chair. The activities include walking and sitting. In Figure 12c, she walks to the bed, sits on the bed, and then lies on the bed. The activities mainly include walking, transitional activities, and lying. In Figure 12d, she walks around and then stands still for a while. The activities indicate walking and standing. In Figure 12e, she lies on the bed for a while and then sits on the bed, which represents the transitional activities of lie-to-sit.

In order to compare the effect of different parameter settings, we calculated the mean accuracy and standard deviation of 10-fold CV for different numbers of decision trees in RF, as shown in Figure 13. The improvement of the second-layer RF compared to the first-layer RF is obvious; for the second-layer RF, the accuracy is much higher, and the standard deviation is much lower.

According to Figure 13, 20 decision trees represent a good trade-off between accuracy and algorithm complexity, and we use this setting to calculate the confusion matrix of all of the activities, as listed in Tables 3 and 4. The mean accuracy of the first-layer RF and the second layer RF is 82.47% and 92.51%, respectively. The standard deviation of the first-layer RF and the second-layer RF is 5.12% and 1.46%, respectively. For a specific type of activity, the value in bold is the percentage of correct classification results, and other numbers are for incorrect classification results. Comparing these two tables, the addition of the duration of successive activity as the input feature to the second layer RF is quite helpful; the recognition accuracy is higher, and the standard deviation is lower. The mean accuracy of 10-fold CV is above 92%.

Table 3. Mean accuracy of the first-layer random forest (%).

Ground Truth	Classification Result					Test Accuracy
	Walking	Sitting	Lying	Standing	Transitional	
Walking	62.17	37.00	0.74	0.00	0.09	62.17
Sitting	8.67	90.40	0.80	0.00	0.13	90.40
Lying	0.33	2.21	93.80	3.43	0.22	93.80
Standing	0.00	0.00	0.27	90.25	9.48	90.24
Transitional	0.06	0.11	0.03	16.73	83.08	83.08

20 Decision trees: mean accuracy is 82.47%, and standard deviation is 5.12%.

Table 4. Mean accuracy of the second-layer random forest (%).

Ground Truth	Classification Result					Test Accuracy
	Walking	Sitting	Lying	Standing	Transitional	
Walking	98.75	1.16	0.00	0.00	0.09	98.75
Sitting	2.01	97.61	0.29	0.00	0.08	97.61
Lying	2.21	1.33	93.47	2.77	0.22	93.47
Standing	0.00	0.00	0.38	87.50	12.13	87.50
Transitional	0.06	0.14	0.19	9.77	89.84	89.84

20 Decision trees: mean accuracy is 92.51%, and standard deviation is 1.46%.

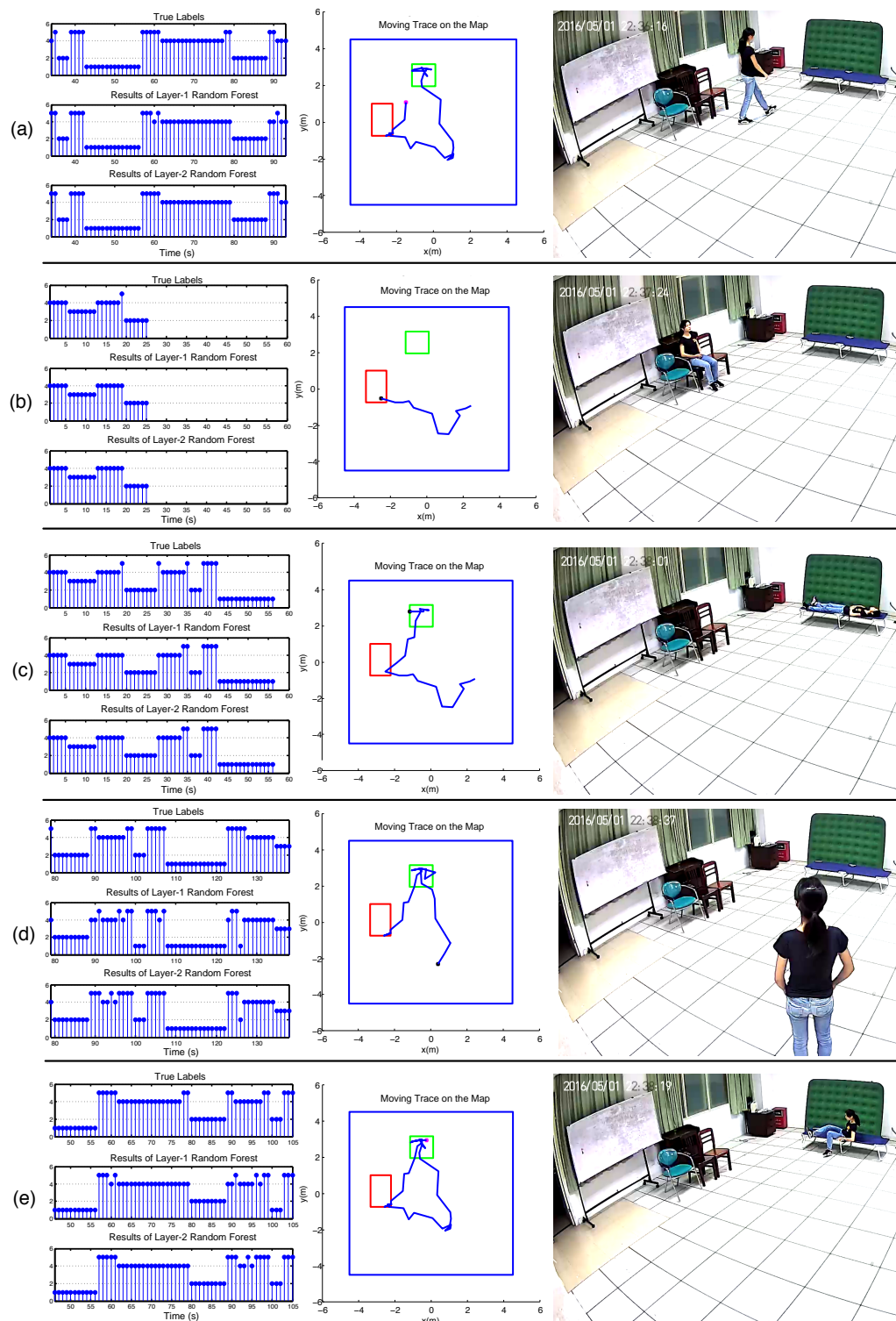


Figure 12. Typical snapshots of video, moving trace of the human target, and the classification results of our two-layer RF. (a–e) The left column represents the ground truth labels of each second, the results of the first-layer RF, and the results of the second-layer RF, respectively. The y-axis represents the activity classification result of each second (s): 1. lying, 2. sitting, 3. standing, 4. walking, 5. transitional activities. The middle column is the moving trace of the human target on the map. The red square is the region of chairs, and the green square is the location of the bed. The right column is the snapshot captured from the video, where (a–e) represent walking, sitting, lying, standing, and lie-to-sit, respectively.

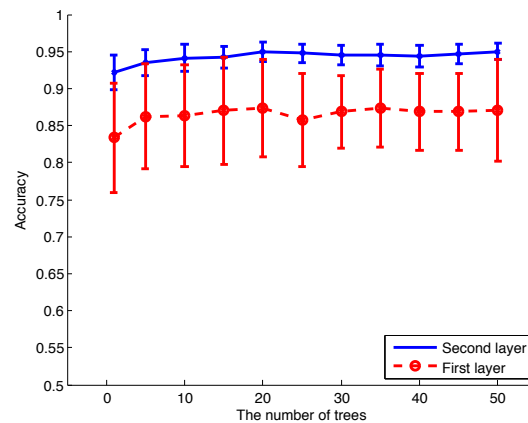


Figure 13. Mean accuracy and standard deviation of 10-fold cross-validation (CV) for different numbers of decision trees in RF. The number of trees includes 1, 10, 15, 20, 25, 30, 35, 40, 45, 50. The blue line is the performance of the second-layer RF, and the dash with circle is the performance of the first-layer RF.

7. Discussion

To compare the effectiveness of RF with other classifiers, we used SVM and naive Bayes to replace RF in both layers of our classifier framework and ran the experiments again [29,30]. The mean accuracy and standard deviation of 10-fold CV are listed in Table 5. It shows that the accuracy of SVM and naive Bayes is lower than RF. The reason lies in that the decision boundary of naive Bayes is linear, which is not consistent with the fact that the locations of the activities that occurred inside the mock apartment were regional, as shown in the middle column of Figure 12. Within the regions of the chairs and the bed, sitting and lying will occur with high probabilities; outside these regions, walking and standing will occur with high probabilities. Thus, the location of the human target (x_t, y_t) is a good indicator. However, using naive Bayes, after assigning the fixed weights to x -axis and y -axis separately, the input feature is not linearly separable. The performance of SVM is better than naive Bayes, because the hyperspace produced by the inner product of the input feature can model the non-linearly separable feature. To enhance the performance of naive Bayes and SVM, much effort is needed to segment the activity region manually and to calculate the probability of each activity occurring at each location.

Table 5. Comparison of different algorithms.

	First-Layer RF	Second-Layer RF	SVM	Naive Bayes
Mean Accuracy	0.82	0.93	0.79	0.66
Stand Deviation	0.05	0.01	0.03	0.05
Training Time	-	1839.14	1428.41	995.72
Testing Time	-	0.12	5.25×10^{-4}	0.06

In each cross validation (CV), there is a total of 14,841 data frames for training, 1487 data frames for testing. The unit of training time is the total seconds per CV, and the unit of testing time is seconds per data frame. SVM: support vector machine.

The RF can model the activity region inside the mock apartment better than SVM and naive Bayes. RF is composed of many decision trees, which are more interpretable to model the square region where different activities may happen; the decision thresholds applied to x -axes and y -axes can be different for each decision tree. The majority vote scheme can boost the decision accuracy. Thus, the decision boundary of RF can be square, which is consistent with the layout of the mock apartment. Furthermore, the RF is also a useful framework to incorporate heterogeneous features such as location, speed, and the duration of successive activities.

We recorded the training time and testing time of different algorithms, as listed in Table 5. In the host PC, we used MATLAB 2013b for data processing. The CPU of the host PC was an Intel(R) Core(TM) i5-6400 2.70 GHz and 8.00 GB RAM. For each CV, the total training time of naive Bayes was the least, and the RF was the most. In the testing phase, the testing time of SVM was the least, and the RF was the most. However, the testing time of RF for each data frame (two seconds) was 0.12 s, which fulfills the requirement of real-time processing.

Because our PIR sensors are ceiling mounted, they will not be easily affected by the existence of obstacles such as the furniture in the mock house. However, when the position of a certain piece of furniture is modified, the location information of the human activities will change as well. We must re-train the RF again, because the statistical distributions of the features are shifted. In such a situation, one of the advantages of our system is that we only need to label the type of activities being performed, with no need to segment the activity region manually and assign different probability for each region.

The performance of our system was compared with some recent existing systems based on wearable sensors or video sensors, as listed in Table 6. The recognition accuracy of our system is comparable to or even higher than other systems in some activity types. However, because the experimental configurations and the daily activities to be classified are not identical, the mere comparison of accuracy is not comprehensive enough. The method proposed in this paper is focused mainly on daily basic activities, which are the elements for more complicated activities' recognition. Our system could not only work independently, but could also cooperate with existing systems; it could be regarded as complementary to the wearable or vision-based sensors. Our approach will increase the robustness of the smart home system. In future work, we will focus on how to recognize complex activities such as "house keeping" and "cooking" by leveraging more sophisticated algorithms to capture the spatio-temporal features of human activities. The quality of activities (e.g., the quality of walking after sitting) will also be investigated.

Table 6. Comparison of human activity recognition (HAR) systems.

Methods	Sensor Type	Activity Types	Mean Accuracy
Zhu et al. [11]	Wearable + Optical Tracker	Lying, Sitting, Standing, Walking, Transitional Activities	about 85%
Jalal et al. [37]	Depth Video	Smart Home Activities Smart Office Activities Smart Hospital Activities	92.33 % 93.58 % 90.33%
Liu et al. [38]	Wearable	Housekeeping Tasks Activity Level Classification	90.67% 94.35%
Maglogiannis et al. [39]	Fisheye Cameras	Walking, Standing, Sitting	about 94%
Brdiczka et al. [40]	Video	Walking, Standing, Sitting, Interaction with Table, Sleeping	77.86%
Proposed Method	PIR Sensors	Lying, Sitting, Standing, Walking, Transitional Activities	92.51%

8. Conclusions

In this paper, we proposed a method to leverage the synergy between location and motion information to solve the problem of human simultaneous tracking and activity recognition (STAR) [16]. To show the potential of simple PIR sensors for automatic surveillance, the coding scheme of the FOV was designed to capture the spatio-temporal information of the human subject. We also designed useful features for human activities recognition, including human location, locomotion speed, and

the duration of the successive activity. A two-layer RF framework was used to model these features and to output the final classification result. We conducted experiments in a mock apartment. The accuracy and standard deviation of activity recognition results were evaluated and compared with other algorithms. Besides high recognition accuracy, our system can significantly reduce the burden of data communication and the complexity of data processing. It is crucial in resource-deprived scenarios, such as WSNs. In our future work, we will focus on multiple human localization and fine granularity activity recognition. The coding scheme of the PIR sensors will be further investigated, and the reference structure will be redesigned to facilitate the data association problem of multiple targets. More sophisticated and hierarchical classification models such as conditional random field (CRF) [41] and HMM [42] will be employed to model the sequential constraints of successive activities.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61301294 and 61601523, the Natural Science Foundation of Guangdong under Grant No. 2016A030310238 and Youth Elite Project of Guangzhou University of Chinese Medicine.

Author Contributions: Study concept and design: Xiaomu Luo and Xiaoyan Luo. Acquisition of data: Xiaomu Luo, Qiuju Guan and Huoyuan Tan. Analysis and interpretation of data: Xiaomu Luo and Huoyuan Tan. Drafting of the manuscript: Xiaomu Luo and Qiuju Guan. Discussion and reviewing the manuscript: Liwen Gao and Zhengfei Wang. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Morgan, L.A.; Perez, R.; Frankowski, A.C.; Nemeč, M.; Bennett, C.R. Mental Illness in Assisted Living: Challenges for Quality of Life and Care. *J. Hous. Elder.* **2016**, *30*, 185–198.
2. Wickramasinghe, A.; Torres, R.L.S.; Ranasinghe, D.C. Recognition of falls using dense sensing in an ambient assisted living environment. *Pervasive Mob. Comput.* **2017**, *34*, 14–24.
3. Chen, L.; Hoey, J.; Nugent, C.D.; Cook, D.J.; Yu, Z. Sensor-based activity recognition. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 790–808.
4. Turaga, P.; Chellappa, R.; Subrahmanian, V.; Udrea, O. Machine recognition of human activities: A survey. *IEEE Trans. Acoust. Speech Signal Process.* **2008**, *18*, 1473–1488.
5. Liu, A.A.; Su, Y.T.; Nie, W.Z.; Yang, Z.X. Jointly learning multiple sequential dynamics for human action recognition. *PLoS ONE* **2015**, *10*, e0130884.
6. Chaquet, J.M.; Carmona, E.J.; Fernández-Caballero, A. A survey of video datasets for human action and activity recognition. *Comput. Vis. Image Underst.* **2013**, *117*, 633–659.
7. Ryoo, M.S.; Rothrock, B.; Fleming, C.; Yang, H.J. Privacy-Preserving Human Activity Recognition from Extreme Low Resolution. *arXiv* **2016**, arXiv:1604.03196.
8. Delibasis, K.K.; Goudas, T.; Maglogiannis, I. A novel robust approach for handling illumination changes in video segmentation. *Eng. Appl. Artif. Intell.* **2016**, *49*, 43–60.
9. Liu, K.C.; Chan, C.T. Significant Change Spotting for Periodic Human Motion Segmentation of Cleaning Tasks Using Wearable Sensors. *Sensors* **2017**, *17*, doi:10.3390/s17010187.
10. Wang, Z.; Yang, Z.; Dong, T. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors* **2017**, *17*, doi:10.3390/s17020341.
11. Zhu, C.; Sheng, W. Motion-and location-based online human daily activity recognition. *Pervasive Mob. Comput.* **2011**, *7*, 256–269.
12. Storm, F.A.; Heller, B.W.; Mazzà, C. Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS ONE* **2015**, *10*, e0118723.
13. Zhu, C.; Sheng, W.; Liu, M. Wearable Sensor-Based Behavioral Anomaly Detection in Smart Assisted Living Systems. *IEEE Trans. Acoust. Speech Signal Process.* **2015**, *12*, 1225–1234.
14. Chodon, P.; Adhikari, D.M.; Nepal, G.C.; Biswa, R.; Gyeltshen, S.; Chencho. Passive Infrared (PIR) Sensor Based Security System. *Int. J. Electr. Electron. Comput. Syst.* **2013**, *14*, 1–5.
15. Lee, M.; Guo, R.; Bhalla, A.S. Pyroelectric sensors. *J. Electroceram.* **1998**, *2*, 229–242.

16. Wilson, D.H.; Atkeson, C. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In Proceedings of the International Conference on Pervasive Computing, Munich, Germany, 8–13 May 2005; Springer: Berlin/Heidelberg, Germany, 2015; pp. 62–79.
17. Hao, Q.; Brady, D.; Guenther, B.; Burchett, J.; Shankar, M.; Feller, S. Human tracking with wireless distributed pyroelectric sensors. *IEEE Sens. J.* **2006**, *6*, 1683–1696.
18. Hao, Q.; Hu, F.; Xiao, Y. Multiple human tracking and identification with wireless distributed pyroelectric sensor systems. *IEEE Syst. J.* **2009**, *3*, 428–439.
19. Yang, B.; Zhang, M. Credit-based Multiple Human Location for Passive Binary Pyroelectric Infrared Sensor Tracking System: Free from Region Partition and Classifier. *IEEE Sens. J.* **2017**, *17*, 37–45.
20. Tao, S.; Kudo, M.; Pei, B.N.; Nonaka, H.; Toyama, J. Multiperson locating and their soft tracking in a binary infrared sensor network. *IEEE Trans. Hum. Mach. Syst.* **2015**, *45*, 550–561.
21. Luo, X.; Liu, T.; Shen, B.; Chen, Q.; Gao, L.; Luo, X. Human indoor localization based on ceiling mounted PIR sensor nodes. In Proceedings of the 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2016; pp. 868–874.
22. Liu, T.; Liu, J. Design and implementation of a compressive infrared sampling for motion acquisition. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 1–15.
23. Luo, X.; Tan, H.; Guan, Q.; Liu, T.; Zhuo, H.H.; Shen, B. Abnormal Activity Detection Using Pyroelectric Infrared Sensors. *Sensors* **2016**, *16*, doi:10.3390/s16060822.
24. Guan, Q.; Yin, X.; Guo, X.; Wang, G. A Novel Infrared Motion Sensing System for Compressive Classification of Physical Activity. *IEEE Sens. J.* **2016**, *16*, 2251–2259.
25. Brady, D.; Pitsianis, N.; Sun, X. Reference structure tomography. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2004**, *21*, 1140–1147.
26. Luo, X.; Liu, T.; Liu, J.; Guo, X.; Wang, G. Design and implementation of a distributed fall detection system based on wireless sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2012**, *2012*, 1–13.
27. Jalil, M.; Butt, F.A.; Malik, A. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In Proceedings of the 2013 International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), Konya, Turkey, 9–11 May 2013; pp. 208–212.
28. Hossain, A.; Rashid, M. Pyroelectric detectors and their applications. *IEEE Trans. Ind. Appl.* **1991**, *27*, 824–829.
29. Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, UK, 2012.
30. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
31. Ooi, S.Y.; Tan, S.C.; Cheah, W.P. Temporal sampling forest (TS-F): An ensemble temporal learner. *Soft Comput.* **2016**, 1–14, doi:10.1007/s00500-016-2242-7.
32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
33. Feng, Z.; Mo, L.; Li, M. A Random Forest-based ensemble method for activity recognition. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 5074–5077.
34. Shotton, J.; Sharp, T.; Kipman, A.; Fitzgibbon, A.; Finocchio, M.; Blake, A.; Cook, M.; Moore, R. Real-time human pose recognition in parts from single depth images. *Commun. ACM* **2013**, *56*, 116–124.
35. Mukhopadhyay, S.C. Wearable sensors for human activity monitoring: A review. *IEEE Sens. J.* **2015**, *15*, 1321–1330.
36. Han, D.M.; Lim, J.H. Smart home energy management system using IEEE 802.15.4 and zigbee communication. *IEEE Trans. Consum. Electron.* **2010**, *56*, 1403–1410.
37. Jalal, A.; Kamal, S.; Kim, D. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* **2014**, *14*, 11735–11759.
38. Liu, K.C.; Yen, C.Y.; Chang, L.H.; Hsieh, C.Y.; Chan, C.T. Wearable sensor-based activity recognition for housekeeping task. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Eindhoven, The Netherlands, 9–12 May 2017; pp. 67–70.
39. Maglogiannis, I.; Delibasis, K. Video Analytics for Activity Recognition in Indoor Environments Using Fisheye Cameras. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*; Springer: Berlin, Germany, 2016; pp. 1105–1110.

40. Brdiczka, O.; Reignier, P.; Crowley, J.L. Detecting individual activities from video in a smart home. In Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Vietri sul Mare, Italy, 12–14 September 2007; Springer: Berlin, Germany, 2007; pp. 363–370.
41. Liao, L.; Fox, D.; Kautz, H. Hierarchical conditional random fields for GPS-based activity recognition. In *Robotics Research*; Springer: Berlin, Germany, 2007; pp. 487–506.
42. Raman, N.; Maybank, S. Activity recognition using a supervised non-parametric hierarchical HMM. *Neurocomputing* **2016**, *199*, 163–177.



© 2017 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/>).