EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article **EAI.EU**

A Multimodal Human Sensing System for Assisted Living

Sonia[1,*], Tushar Semwal[2]

[1]University of Petroleum and Energy Studies, India
[2]The University of Edinburgh, United Kingdom

Abstract

Introduction: Advancements in sensor technology have resulted in smart systems that can analyze, interpret and understand their surroundings for decision making. Using such intelligent systems for environments such as smart homes, device automation, and hospitals, makes sensing human presence an essential requirement. Human sensing becomes critical, especially in assisted living scenarios such as in elderly care. However, mechanisms for human sensing are not foolproof because of the dynamic nature and ability of human beings to deliberately mislead the sensors.
Objective: Objective of paper is to detect human presence in varying environment using non-intrusive sensors.
Method: Besides, sensors have inherent limitations, due to either their mechanism of sensing or the environmental conditions, which can cause them to fail in human detection. These limitations can, at times, cause sensors to provide useless data to the system. In this paper, we propose an adaptive multi-modal human sensing mechanism which can autonomously identify and ignore unnecessary data from a set of sensors, thereby reducing computation complexity, reducing false alarm rate and yielding better performance. The effect of sensing when the human being is in motion has also been studied.
Results: The results portrayed in the paper prove the efficacy of the proposed multi-modal system over its single sensor counterparts when used in changing environments and other proposed multi-modal human sensing.
Conclusions: Human sensing is vital to many smart applications such as smart homes, traffic management systems, human-computer interfaces, etc. Since human beings, in general, are always on the move, the use of a dedicated sensor could fail to detect human presence, especially when the ambient parameters around the sensor change. In this paper, a multi-modal human sensing approach has thus been prescribed for overcoming this issue. This work described focused on automating the identification of inappropriate data relayed by some sensors under certain environmental conditions. Experiments reported include both cases - when the sensors are mounted on a static unit (door frame) and also on a mobile robot. The corresponding results reveal that a combination of sensors outperforms the use of individual dedicated sensors for human detection. Analyses of the walking speed of human being have also been studied which endorse the robustness of the approach. However, different directions of motion need exploration in future.

Received on 20 May 2020; accepted on 19 October 2020; published on 26 November 2020

Keywords: Multi-modal human sensing, Adaptive human sensing, autonomous sensor selection

doi:10.4108/eai.26-11-2020.167285

1. Introduction

A system incorporated with knowledge and intelligence can analyze, interpret and understand its surrounding environment and take appropriate actions or decisions. In the last two decades, a significant amount of prototypes and solutions based on Machine Learning (ML) techniques have been proposed by researchers to empower machines with intelligence. In the current era of the Internet of Things (IoT), sensors play a

[*]Corresponding author. Email: sonia.iitg@gmail.com

dominant role. They have become more robust, cost-effective and smaller in size. This has stimulated their deployment on a large scale in factories, offices and even living environments. The concept of IoT has given rise to smart environments wherein a range of sensors embedded within are used to monitor a variety of parameters. While such habitats allow enhancement in lifestyles, they are also capable of monitoring the health status of the inhabitant(s), intruder detection, fall detection, smart lighting, device automation, etc. Realization of smart systems for spaces inhabited by human beings is highly dependent on mechanisms used to detect human beings. *Human Sensing* can be defined as the process of differentiating human beings from non-humans using data received via sensors [1]. It encompasses issues from the lowest level instantaneous sensing challenges up to large-scale data mining. Such sensing mechanisms are not foolproof since the motion and behaviors of human beings are dynamic. Worse, human beings can also purposefully evade detection by hiding from sensors. All this makes human sensing a tough problem. While cameras have been widely used for this purpose, they suffer from the issue of privacy. There is thus a dire need to find non-intrusive ways for human detection.

Unlike, vision-based sensors, non-intrusive sensors such as an Ultrasonic Sensor (US) or a Pyro Infra-Red (PIR) sensor, can sense only low-level features such as color, heat difference, the energy of the reflected ultrasonic wave, etc. Detection of these features is dependent on the environmental factors, for instance, light intensity, temperature, humidity, the distance of human being from sensors, speed of motion of human being, etc. Therefore, human detection with the help of non-intrusive sensors is a challenging task. A PIR sensor is one of the widely used sensors to detect the presence of a human being in an indoor environment [2]. They have also been used for other purposes such as detection of a falling of a person [3], to count the number of human beings [4] and human tracking [5]. A PIR works on the principle of sensing the IR waves emitted by living beings. Unfortunately, this sensor cannot differentiate human beings from other animate objects, including animals. This can mislead the system into generating a false alarm. Vibration sensors, which work on the amount of force applied, have been used to learn walking patterns [6], fall detection[7], etc. As on date, not much work has been carried out to explore the use of a US for human detection. A US uses the difference between the energies of the emitted and reflected ultrasonic waves. Waves of a specified ultrasonic frequency are bombarded on target objects (which are placed at equal distance from the source of ultrasonic waves). The reflected waves from these objects are heterogeneous in nature. This heterogeneity arises from the fact that the objects have different absorption coefficients for ultrasonic waves. Such sensors are thus widely used in detecting deformity in metals [8] and differentiating surfaces [9]. However, every sensor has its own limitations [1]. For instance, while a PIR sensor cannot differentiate between human beings and animals. Similarly, a vibration sensor is an infrastructure dependent sensor and its implementation costs are high. Sensors thus need to complement one another to overcome the limitations. Multi-modal sensing is widely accepted where sensors complement one another and overcome their own limitations to quite an extent. An example of such multi-modal sensing is when a camera, which needs appropriate light to function correctly, is used in conjunction with a PIR sensor to detect human presence. When the sensing area becomes dark [10] the camera output deteriorates, and the PIR takes over. Vibration and PIR sensors have also been coupled to reduce the false alarm rates [11] in fall detection. However, in multi-modal sensing, the decision of when and which sensor data should be processed plays an important role. Processing the data from all the sensors every time increases the computation overhead. With enhanced communication and machine learning algorithms, learning patterns from sensory data has become more accessible. Raw data received from sensors need to be processed further based on the kind of expected output. In general, models based on Machine Learning (ML) techniques are built to analyze or categorize the various objects/situations based on the sensory data.

The key contributions of this paper can be listed as:

1. Combine data from multiple sensors situated in different environments (indoor and semi-open).

2. Analyze the in-built characteristics of every sensor so as to automate the process of finding a blinked sensor(s), thus reducing computation time.

3. Propound an algorithm for human detection based on multi-modal sensing in both scenarios where sensors are static and mounted on a mobile robot. It also covers both the cases of stationary human as well as moving human.

4. Analyze the walking speed (which is a part of the dynamic nature of human) versus accuracy of human detection.

In the following section, a literature survey on multi-sensors approach for human sensing is presented. In section 3, problem definition is provided, following which methodology is explained. Towards the end of this paper, experiments and results obtained were discussed, which is followed by conclusions.

2. Literature Survey

Researchers have used several parameters such as body temperature, weight, gait pattern, heartbeat, uniqueness of the structure, vibration, scent, etc., to distinguish a human from non-humans. These typical traits of a human being are captured with the help of available sensors and are used to model and develop various applications. Guo et al. [12] present a survey on PIR sensors used for human detection to automate the switching on and off of lights in an indoor scenario. PIR sensors have also been used by researchers to count the number of persons in a room [13], to estimate the occupancy of a given indoor space [4], intruder detection [14], etc. Kessler et al. [6] report the use of vibration sensors to detect human beings. An ultrasonic sensor has been used by Sonia et al. [15] to differentiate human from non-human things. However, no one claims 100% accuracy to detect the human presence in every possible scenario with a single sensor. In recent years vision-based systems have advanced substantially and are capable of detecting human beings with relatively high accuracies. Chen et al. in [16] present a survey on human motion analyses. They emphasize the lack of realistic data sets which covers a vast number of human beings in numerous poses to train the system for use in a human sensing system. Dynamism in human nature such as different walking speeds, different body structures, different clothing styles, etc. makes this sensing task challenging to perform. This contributes to one of the probable reasons responsible for the failure of human detection.

Teixeira et al. [1] presents a survey of different sensors used for human sensing. They state the limitations and failures (such as PIR fails to detect stationary human being, the camera fails to detect human being in the dark, etc.) of individual sensors used for sensing, and stress the need of using a combination of multiple heterogeneous sensors to improve sensing accuracy. Bellotto et al. [17], use multi-modal human sensing to sense and track the human being in a cluttered indoor environment. However, this combination is not cost-effective. Vision sensor, wearable devices and audio sensors have been combined to reduce the false alarm of the system build for human detection in [18] and [19]. This combination can fail in a foggy environment if a human fails to wear the sensor and walk silently. A survey by Avci et al. [20] cite the use of multiple sensors such as PIR, ultrasonic, vibration, temperature sensor, etc., to monitor the daily activities of human beings. Similarly, a survey on wearable devices presented by Lara et al. in [21] emphasizes the utility of multiple sensors to monitor the health and wellness of human beings. While adhering to the basic concepts of building a system for human detection through successive layers, the task-achieving behavior of a system can be fragmented into many smaller decision-making units [22]. Each unit has an input that needs to be converted to the output using analytical data techniques. Data received via different resources need to be processed using various techniques [23]. Matthews et al. [24] describe a system where different algorithms are implemented to process the data from different sensors. A voting-based approach is explored in [25] to process the data from both ultrasonic and PIR sensors to detect the human presence. One may thus conclude from the literature that data received via a variety of sensors need to be processed in different ways so as to detect human presence reliably.

Yang et al. in [26] have used the face, body appearance and silhouette via a Kinect sensor and multiple color cameras to detect human beings in a health-care application. Using a Kalman filter, they claim their multi-modal approach to be more effective than those reported by others. However, their sensor combination has been tested only in a controlled environment. Human detection has also been performed in robotics where sensors mounted on a robot try to detect a human presence for assistance. Human detection has been performed by fusing the data from a laser sensor (to detect leg structure) and a camera [17]. Jin et al. [27] have used unattended ground sensors for human sensing. In the approach proposed herein, the primary level features from individual sensor signals are extracted and then combined to form composite patterns based on relational dependencies between them. The advantage of this approach is that system can function (with reduced accuracy) even in case of a failure of a sensor. Vibration sensors come to the rescue when the PIR sensor fails to differentiate between a human being and animals. Occupancy detection which solely depends on human detection has been demonstrated by Candanedo et al. [28]. They have presented a detailed analysis of all the sensors and pairwise sensor combinations via a correlation matrix. Analysis shows the performance of the different statistical models concerning different combinations of sensors to sense the human presence. The paper focuses on choosing the combination of sensors with the highest accuracy. Table 1 presents a survey of some of the multi-modal/sensor fusion approaches to detect various human activities. But, it is hard to find exact related word.

In contrast to the applications described, the proposed algorithms are not tested in dynamic environments. The work described in this paper focuses on scenarios where the situation could change and thus affect system performance. Under such condition the system automatically detects which sensors have failed in that environment and ignores their inputs accordingly, thus saving computational time while also enhancing performance. This paper also explores the

Sensors used	Features used	Technique
Kinect sensor and multiple color cameras [26]	face, body appearance and silhouette	Using a Kalman filter
vibration and passive infrared [29]	Temperature and gait	wavelet based signal processing methods based sensor fusion
Smartphone sensors [30]	skin temperature, galvanic skin response, heat flux, and a 2-d accelerometer	Multi-modal Sensing for Human Activity Modeling in the Real World
Eight body-worn Inertial Measurement Units [31]	Multiple properties	Deep Learning
Foot-mounted inertial sensors and multi-sensor fusion [32]	Inertia, vibration	Hybrid NN/HMM model

Table 1. Approaches for sensor fusion

effect of height and speed of motion of a human being on True Positive Rate (TPR) of human detection. This aspect avoids further processing of incoming data from those sensors which provide data that can degrade the performance of human detection. The system has been tested in both indoor and outdoor environments.

3. Problem Definition

- For a given classification task T, let $S = S_1, S_2, ..., S_{N_S}$ be a set of N_S sensors used to collect the data. In this paper, the task T is to sense human presence.

  $$S_i^U = S - S_i^B \tag{1}$$

  where, S_i^U is the set of sensors from which data extracted can be made use of accomplishing task T, S_i^B is the set of blinked sensors from which data extracted cannot be used to accomplish task T .

  It may be noted that, as environment changes, S_i^B will also differ, causing corresponding changes in S_i^U. Thus, the problem herein is to find the set of blinked sensors ($S_i^B \in S$) based on the raw data received from all sensors in S without any human intervention.

- A sensor (S_i) is defined as a *blinked sensor* in a particular environment (E_j) if information obtained from it cannot be used to accomplish task T.

3.1. Elimination and Decision-making Features

Every sensor has its limitations due to which it could fail to sense a human being in a particular environment [1]. For example, in dark areas, information retrieved from a camera cannot be used to detect a human being. Under this condition, the maximum and minimum pixel values within the concerned frame are equal to zero. Since the image is entirely black, one may conclude that the frame cannot be used to detect human presence. Likewise, if the raw data from a PIR sensor frame cannot be used for human detection when the maximum and minimum values lie in the range between 150 and 500, since it indicates that either the human being is stationary or not present.

To eliminate the processing of information contained in a frame F_j^k of the sensor S_k, we need to extract some typical features. Let, $\mathcal{F}_j^k$ represents the feature vectors of eliminating features of j^{th} frame of the k^{th} sensor.

The data from the remaining non-eliminated frames are used to extract features that will aid in accomplishing the task of human detection. For example, color could be the feature used to decide on a camera-based model that differentiates between an apple and an orange. However, some of these features could be redundant and may mislead the overall decision process. Decision features($\mathbb{F}_j^k$) obtained from a non-eliminated frame can be represented as a feature vector $\mathbb{F}_j^k$. where, $\mathbb{F}_j^k$ is the decision making feature vector of the j^{th} frame obtained from the k^{th} sensor .

3.2. Decision Model Selection

Decision features are used to train an ML model so as to detect human presence or absence. The heterogeneity of the data received from different sensors causes different ML models to perform differently. Selection of the best ML model for data from a sensor needs to be done based on its performance. For instance, to find an appropriate model for the PIR sensor, one could use SVM, Linear regression, Decision tree for training and then choose the best performing one. Thus if models $M_1^k, M_2^k, \ldots M_p^k$ are initially considered for sensor S_k and $\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_p$ are their respective performance measures. Then the model with the highest value of $\mathcal{P}$ is selected for the detection process.

3.3. Online Clustering

Clustering techniques group similar objects to form a cluster. Clustering thus distinguishes two different objects/things. The distribution of data in an n-dimensional space can be represented in the form of clusters. This forms the basis of classification algorithms such as SVM, k-means, decision trees, etc. In the proposed approach, we have used the online clustering technique proposed in [33] and used the elimination characteristics for deciding sensors that should not be taken into consideration while arriving at a decision. The used approach fix the radius of the cluster at the initial stage and update the cluster centers in an online manner.

3.4. Methodology

This proposed methodology is based on the one-class classification to detect the presence of a human being in different environments. Figure 1 shows an overview of the proposed method. As shown in the figure, during the training phase, both eliminating and decision-making features are extracted. The output of eliminating features are clustered via clustering algorithm. However, decision making features extracted from each sensor S_k are used to train their respective machine learning models. It can also be seen from the diagram that during the testing phase, first eliminating features are extracted. The extracted feature vector is used to find blinked sensors. From the remaining sensors decision making features are extracted to be used for final decision making using their respective ML models. If greater than 50 per cent of sensors are detecting a human than the final decision is considered as positive for human detection.

Training Phase. As discussed earlier, since a single sensor is not sufficient for human sensing, multi-modal human sensing was considered. For such sensing, the data obtained from the sensors need to be preprocessed and models generated, individually. For the proposed multi-modal sensing, the training phase can be divided into two steps:

- **1. Selection and training of ML model for each sensor respectively:** Let for $S_k \in S$, d_k be the data stream which is buffered in term of frames, F_i^k. In the current multi-modal human sensing method, for every sensor $S_k \in S$, an ML model needs to be generated for the classification process. The manner by which the appropriate model (M_k) for sensor (S_k) is chosen has been described earlier in the *Decision model selection* section. While capturing the data required for training the respective models, all sensors should point to the same human being and sense concurrently.

  From the buffered frames, the individual feature vectors used for elimination and decision making, viz. $\mathcal{F}_j^k$ and $\mathbb{F}_j^k$ are calculated for all the sensors. $\mathbb{F}_j^k$ are used for training the selected ML model for a sensor.

- **2. Obtain clusters:** The clustering technique [33] used was provided with all the feature vectors used for elimination, $\mathcal{F} = \mathcal{F}_j^1, \mathcal{F}_j^2, ..., \mathcal{F}_j^N$, so as to determine the radii and centers of the clusters generated.

  $\mathcal{F}_j^k$ represents the eliminating feature vector of j^{th} frame of the k^{th} sensor.

Algorithm 1 Algorithm for elimination of blinked sensors

Input: Prior calculated Cluster centers and radii (which are output of clustering technique).
Input: Sensor dependent range of the features.
Output: Eliminated sensors.

```
for all Data frames received from N sensors do
    calc_features;
end for

if $\mathcal{F}_j$ lies in any of the clusters then
    No sensor can be eliminated.
    Update_Clusters($\mathcal{F}_j$)
else if $\mathcal{F}_j$ does not lies in any of the clusters then
    for all $S_k$ do
        Check_range()
        if ($S_k$ is out of defined range then
            Add_sensor_$S_i^B$($S^B$)
        else if No sensor is out of range then
            Update_Clusters($\mathcal{F}_j$)
        end if
    end for
end if
```

Distinction Phase. Concluding on the presence or absence of a human being involves:

- Elimination of processing of frame received from the blinked sensors
- Concluding on whether the target is a human being

Algorithms 1 and 2 explain these two processes. Algorithm 1 takes in the cluster centers and radii computed in the training phase along with the sensor-specific ranges for deciding whether or not to categorize a sensor as blinked. Also, as mentioned earlier, based on the values within the frames, the corresponding sensor is either eliminated (blinked) or considered. Similar to the training phase, here too, data from N_S sensors is buffered in the form of frames. The data within

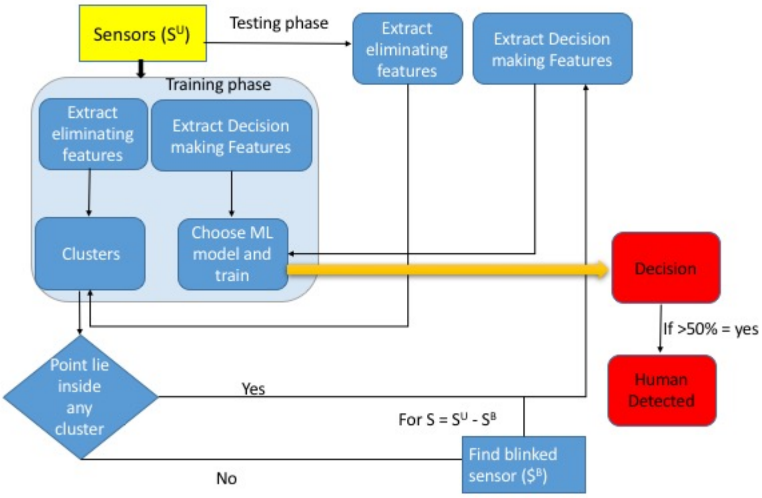
Sensors (S^U)
Testing phase
Training phase
Extract eliminating features
Extract Decision making Features
Extract eliminating features
Extract Decision making Features
Clusters
Choose ML model and train
Decision
If >50% = yes
Human Detected
Point lie inside any cluster
Yes
For S = S^U - S^B
Find blinked sensor ($^B)
No

Figure 1. Overview of the system

Algorithm 2 Algorithm for detection of Human Presence

```
Input: Pre-trained models ($M_k$) for sensor ($S_k$)
Output: Human presence either yes or no.

$S^B$ = Algorithm for identification of blinked sensors.
$S^U = S - S^B$
for all Data frames ($F_j^k$) received from $S - S^U$ sensors do
    $\mathbb{F}_k = calc\_features(F_j^k)$
    $Output = Predict\_output(M^k, \mathbb{F}_k)$
end for
$Count = Calculate\_positive\_count(Output)$
if Count is > 1/2(Count of sensors in $$^U$) then
    Human detected.
    $Update\_Positive\_Models(M^k, \mathbb{F}_k)$
else if Count is <= 1/2(Count of sensors in $$^U$) then
    Human not detected
    $Update\_Negative\_Models(M^k, \mathbb{F}_k)$
end if
```

these is used to ascertain the elimination features using $calc\_features()$ which in turn outputs $\mathcal{F}^t$

If $\mathcal{F}^t$ lies within any of the clusters then frames from all the N_S sensors are taken into consideration for human sensing. $\mathcal{F}^t$ is also used to update the cluster centers as in [33] using $Update\_Clusters()$, to facilitate dynamic evolution of the clusters.

On the contrary, if $\mathcal{F}^t$ lies outside these clusters, the values, the algorithm 1 finds the sensor(s) whose data was responsible for making it an outlier. This is done by inspecting the associated features within the relevant frames based on the sensor-specific ranges that are already available to the algorithm. If any of the value(s) of any of the feature(s) pertaining to a sensor S_k within a frame is out of the predefined feature-specific range, then this sensor is deemed to be blinked and added to the set of blinked sensors, S^B. If $\mathcal{F}^t$ does not lie within any of the clusters and none of the associated sensors have been deemed to be blinked then a new cluster if formed using $Create\_Cluster()$.

The Algorithm 2 depicts the process of detection of human presence. It takes the already available pre-trained models for each sensor as its input and uses the set of blinked sensors, S^B obtained from algorithm 1 to eventually find the set of useful sensors(S^U). The associated frames from the sensors in S^U are used to compute the decision-making features that is $\mathbb{F}_t^k$ These features are used by the sensor-specific ML model to decide whether or not the target is a human being. A human being is said to be detected if the majority of sensor-specific models report this detection to be positive. As its last step, the algorithm 2 update the sensor-specific models based on the final output by re-training, on-the-fly.

4. Experiments and Results

Experiments to validate the efficacy of the proposed human sensing methodology were conducted in two different indoor setups.

4.1. Experiment 1

The hardware used in the experimentation included - 1. **Analog Ultrasonic Sensor (XL-MaxSonar -EZ/AE) (AUS):** The output of this sensor is analog voltage envelope of return acoustic waveform. Data is buffered in the form of frames. If the standard deviation of such

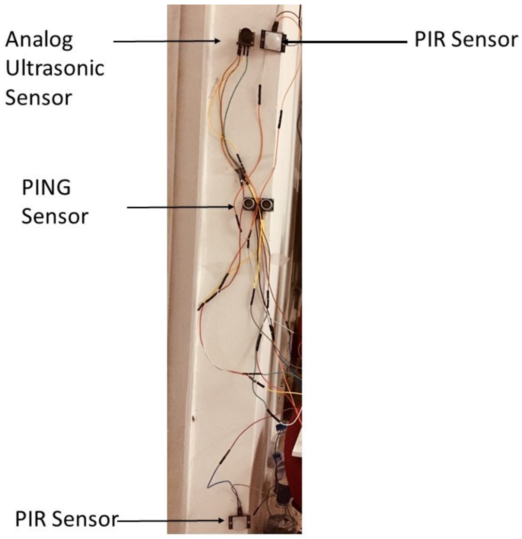

Figure 2. Sensors setup on a wall

a frame is not equal to zero, then data of this frame can be used to sense the presence or absence of human being. 2. **Two Pyro infrared Sensor(PARALLAX PIR sensor (Rev B)) (PIR):** The analog output of the sensor is buffered in frames. For a moving object (man or pet), which is in the sensing range of a PIR sensor, the maximum and the minimum values output of the frame are higher than 500 and less than 150, respectively. 3. **Ping Sensor (PARALLAX PING)))) (US):** This sensor outputs the distance of the obstacle in front. The data from AUS and PIR sensors are processed depending on the output of US sensor. 4. **Arduino Mega 2560:** All the sensors are connected with the Arduino board to read the sensory data. The sensors - Ping, AUS and PIR(PIR_1) - one each, were placed on a wall (near to a door) at a height of 72 cm from the ground as shown in the Figure 2. Another PIR(PIR_2) sensor was placed on the same wall at a height of 20 cm from the ground. All sensors were placed in such a way that they all pointed at the same target. In this setup PIR_1 is prioritized over PIR_2 sensor PIR_1 is placed at an height of 72 cm which is above average height of cats and dogs. That is why, for the final decision of human detection PIR_1 is prioritized. Therefore, PIR_1, AUS, PIR_2 is the sequence of sensors in the decreasing order of priority for the finalization of decision of human sensing.

Apart from the distance of the obstacle in front, from analog AUS signal, a comprehensive analysis can be performed to differentiate human from non-humans. Similarly, unlike binary output of PIR sensor analog signal of the PIR sensor can be analyzed for the direction of motion and speed of motion.

The data of PIR sensors and AUS is processed only when an object is detected at a distance of 60-70 cm from the sensors. To train the system, the incoming data was buffered into data frames. One data frame of a sensor consisted of all the data obtained from a sensor in a time period of 4 seconds. When human being was standing at a distance of 60-70 cm, the sensory data of PIRs and AUS was buffered in the form of frames. Features were extracted from every frame of AUS and both PIR sensors. For training purpose, both were eliminated, and decision-making features were extracted from each frame. Equations 3 and 2 represent the eliminating features for the PIR sensors while equation 3 through 10 represent the decision making features. Equation 10 represents the eliminating feature for AUS . However, equations 3 through equation 10 represent the decision making features for the same. Table 2 represents the eliminating features and their specified ranges for different sensors.

Sensor	Eliminating Features	Human sensing range
PIR_1	Max and Min	> 500 and <= 0
PIR_2	Max and Min	> 500 and <= 0
AUS	Standard Deviation	> 0

Table 2. Eliminating features and their specified ranges for PIR sensors and AUS for Experiment 1

$$Minimum = Min\,(a_x)_{x=1}^{X} \tag{2}$$

$$Maximum = Max\,(a_x)_{x=1}^{X} \tag{3}$$

$$Median = median\,(a_x)_{x=1}^{X} \tag{4}$$

$$Mean = mean(\Sigma_{x=1}^{X} a_x) \tag{5}$$

$$Kurtosis = X\frac{\Sigma_{x=1}^{X}\,(a_x - average)^4}{\left(\Sigma_{x=1}^{X}\,(a_x - average)^2\right)^2} \tag{6}$$

$$Energy = \Sigma_{x=1}^{X}\,(a_x * a_x) \tag{7}$$

$$CrestFactor = \frac{\frac{1}{2}\,(Maximum - Minimum)}{RootMeanSquare} \tag{8}$$

$$RootMeanSquare = \sqrt{\frac{1}{X}\Sigma_{x=1}^{X} a_x} \tag{9}$$

$$StandardDeviation = \sqrt{\frac{1}{X-1}\Sigma_{x=1}^{X}\,(a_x - Mean)^2} \tag{10}$$

where, X is the number of data instances in the the data frame and a_x represents a single data instance of the buffered file. Therefore,
$\mathcal{F} = max(PIR_1), min(PIR_1), max(PIR_2),$
$min(PIR_2), Standarddeviation$

For training purposes, the data obtained when 15 different human beings stood/moving at a distance of 60-70 cm. away from the sensors were collected. Human beings wore different types of clothing and

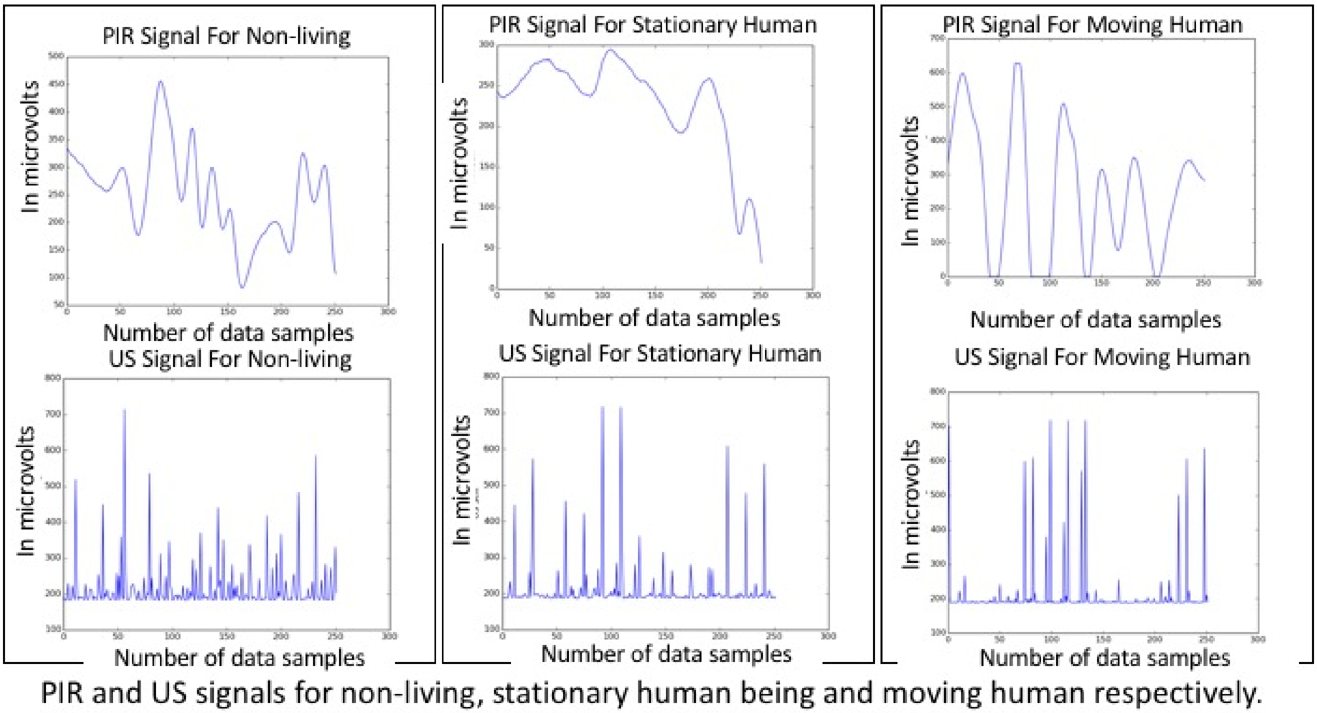

Figure 3. Input to the system under various conditions

stood in different postures. 100 data frames were collected for each human being in order to train the system. Human presence was confirmed in a supervised way. After every frame, human presence is confirmed by pressing a physical switch. If the switch was pressed then that particular frame was retained else discarded. As per the methodology section $\mathcal{F}$ and $\mathbb{F}$ were calculated. Initially, based on the literature survey, for PIR and AUS, multiple ML algorithms were considered. Models showing highest accuracies from among conventional machine learning algorithms were experimentally selected for each of the sensors. To select such a model for PIRs and AUS, extracted decision making ($\mathbb{F}$) were considered. The collected data was divided into the ratio 7:3 to train and test for a particular model. Also, it should be emphasized that for the testing purpose data frames for cats and dogs were also included. The graph in figure 4 shows the initially considered models along with their respective accuracies for PIR sensors and AUS.

It can be seen from figure 4 that kNN performs better for PIR data as compared to SVM, VBA, and fuzzy logic. Similarly, it also reveals that k-NN clustering performs better for AUS data. Therefore, k-NN was the selected ML model for PIRs and AUS. In parallel, centers of the clustered were also calculated using extracted feature vectors $\mathcal{F}$. The cluster centers were calculated in an online manner, as described in the methodology section.

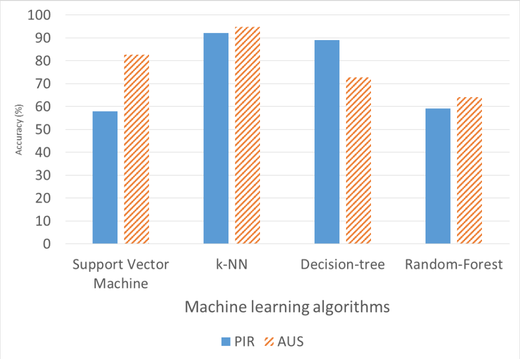

Figure 4. Accuracies of different ML algorithm for AUS and PIR sensor

Radius	Number of clusters	Average cluster density
0.01	12	55.96
0.05	10	45.96
0.1	10	39.06
0.15	9	36
0.2	7	28.34

Table 3. Cluster numbers and their average density at varying radius (r)

Table 3 shows the number of clusters and the average cluster density for different chosen values of r.

For the detection of human sensing in the given set up, the value of r was taken to be 0.05mm. Thus, a trained multi-modal system (which consists of two PIRs, one AUS and One Ping sensor) was deployed for testing. The system was tested for human presence detection. When the Ping sensor reported an obstacle at a distance of 60 to 70 cm, the data was buffered in the frames for each sensor. The two features viz. $\mathcal{F}_t$ and $\mathbb{F}_t$ were extracted for all sensors from the respective data frames of the sensors.

If the $\mathcal{F}_t$ (which represents the vector of eliminating features of all the sensors) was inside the clusters built a priori during the training, then the decision making features($\mathbb{F}_t$) were extracted from the data frames of all the sensors. If $\mathcal{F}_t$ was outside all the clusters then eliminating features of all the sensors were checked to find whether they lie within the prior defined sensor-specific sensing range or not. If the value of any of the eliminating features of a sensor S_k is found to be out of sensor-specific sensing range, then the data received from that sensor was ignored in the decision making process for human detection and sensor is added to S^B. For the remaining sensor (S^U), the decision making features were extracted from their data frames. After the decision-making features were extracted, the output was predicted using the sensor-specific models. The outputs of these models are either a 0 or a 1 (0 indicates that the data of the processed data frame does not belong to a human being and 1 suggests that it belongs to the human class).

It is possible that no sensor has failed (i.e. S^B is empty) and $\mathcal{F}_t$ does not lie in any of the already available clusters. In such a situation, a new cluster is formed, and the corresponding decision-making features are extracted from all the sensors. The output of all the respective models is then considered for human sensing.

For the current setup, if AUS is turned off, PIR sensors fail to sense the presence of a stationary human being. Similarly, if PIR sensors are turned off, AUS fails to sense the human being if he walks out at high speed in front of the sensor. However, if all sensors (AUS, PIRs) are turned on, stationary human beings (when PIR cannot sense) are identified by AUS. Similarly, human walking at high speed can be sensed by PIRs. The above statement is supported by the results presented in figure 5. Results show that that PIRs failed to detect the presence of stationary human 45 times from a total of 45 and succeed to detect the moving human an all the cases. Also, AUS detected stationary human 37 times from a total of 45 and detected moving human being 35 times from a total of 45. However, using a combination of PIR and AUS 43 time stationary human was detected correctly out of 45 times, and moving human was detected correctly in all the cases.

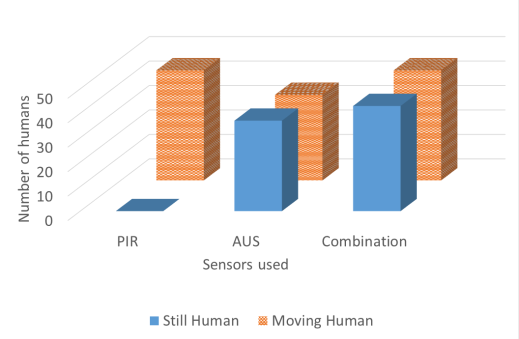

Figure 5. Number of human correctly classified (total number = 45)

4.2. Experiment 2

For the second experiment following hardware was used:
1. **Pioneer Amigobot :** This robot has eight sonar sensor covering an angle of 360 degree, which makes obstacle avoidance possible while moving in an environment.
2. **Analog Ultrasonic Sensor (XL-MaxSonar - EZ/AE)(AUS):** This sensor outputs the envelope of the reflected wave which is used to analyze the presence/absence of a human being. 3. **Pyro Infrared Sensor(Parallax PIR sensor (Rev B)):** As prior explained, this sensor works on infrared radiations emitted from the body of living beings. This sensor is helpful to find a living being in motion. 4. Camera: The vision-based sensor is used to analyze a given environment. The camera clicks the pictures of the objects in front at a defined frequency to investigate the presence or absence of the human being in front.

In this experiment, the sensors were mounted on a mobile robot. The experiments were performed both in an indoor as well as an outdoor environment. All the sensors (AUS, PIR and camera) were mounted on the robot at a height of 70 cm from the ground as shown in figure 6. The mounting was done in a manner that all sensors point to the same obstacle at a given instant of time. The data was processed only when an obstacle was encountered by the robot at a distance of 60-70 cm. The robot was programmed in a way that whenever an obstacle was encountered, its speed decreased to 0.05m/sec. The speed of the robot was changed based on the inference derived from the outputs of the sensor data frames. Similar to the previous experiment, a data frame consists of data buffered from the sensors for 4 seconds. All the sensors operated concurrently. Thus, a data frame of all the sensors consisted of information of the same obstacle at a given instant of time. The models chosen for the experiment for PIR, AUS and camera sensors were SVM, KNN and CNN, respectively.

The CNN for the camera was trained off-line for the human class using 1000 pictures of human beings in

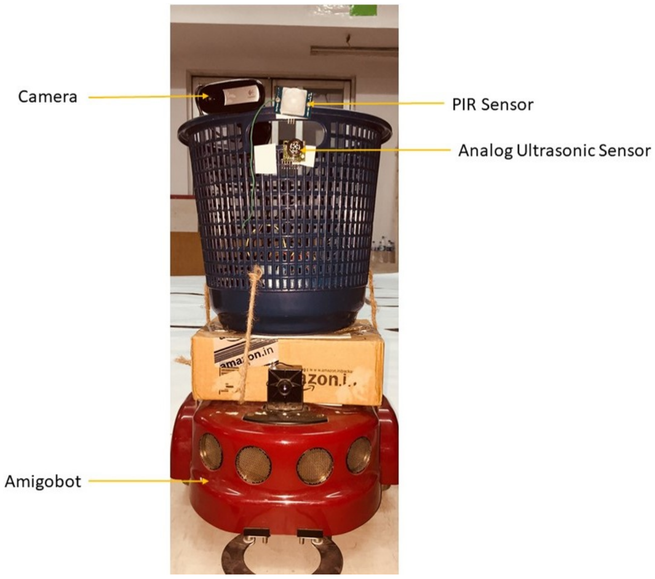

Figure 6. Sensors Setup on a Mobile Robot

different poses and dresses. The training of the other models for the PIR and AUS sensors was performed online where the robot was made to move in an environment which was congenial to the camera (i.e., in daylight) as also the other sensors. As soon as an obstacle was detected at a distance of 60-70 cm, the camera was enabled to capture its image. The data obtained from other sensors(AUS, PIR) was buffered into frames of 4 second while the camera clicks every fourth second when conditions are met.

Just as in the previous experiment, the eliminating and decision-making features were extracted from the frames obtained from all sensors.

Sensor	Eliminating Features	Range for human sensing
PIR	Maximum and Minimum	> 500 and $<= 0$
AUS	Standard Deviation	> 0
Camera	Maximum Pixel Intensity and Minimum Pixel Intensity	> 0

Table 4. Eliminating features and their specified ranges for PIR sensors, AUS and Camera for Experiment 2

Table 4 shows the eliminating features for PIR, AUS and camera with their respective ranges. As per entries in the table 4, PIR data should be analyzed if the calculated maximum value of a data frame is above 500, and the minimum value is less than or equal to zero. Similarly, for the ultrasonic sensor, the calculated value of the standard deviation of a data frame should be higher than zero. However, for a camera, an image captured cannot be analyzed for human presence if it is completely dark that is both minimum and maximum pixel intensity are equal to zero.

For the PIR sensor, equations 2 and 3 were used to find the eliminating feature(s) while the equations 3 through 10 were used likewise for the decision making features. Similarly, for the AUS, the equation 10 were used to calculate the eliminating features while equations 3 through 10 were used for the decision making features. Similar to the experiment 1, multiple machine learning algorithms were tested to select the best one for each of the sensor (to be used for experiment) respectively.

The graph in figure 7 shows the calculated accuracies obtained while using different machine learning models for the PIR sensor and AUS, respectively. Unlike experiment 1, for experiment 2 sensors were mounted on the robot even for the preprocessing phase of the experiment. This includes a selection of a machine learning algorithm which outperforms other chosen algorithms for each of the sensors respectively.

It can be seen from the graph in figure 7 that K-NN performs better for both PIR as well as AUS as compared to SVM, Decision-tree and Random forest. For the camera too, the eliminating features were extracted for every image whose output was a human class. The eliminating features of an image captured with the help of a camera are maximum pixel intensity and minimum pixel intensity. Therefore for the complete system = maximum, minimum, Standard deviation, maximum pixel intensity, minimum pixel intensity.

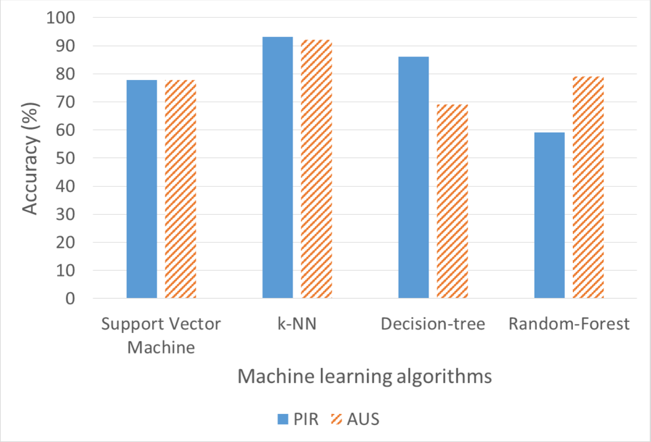

Figure 7. Accuracy of different ML algorithms for both AUS and PIR sensor mounted on a mobile robot

To train the complete system, the mobile robot was made to move in an environment where the rate of human beings to be detected was high. Data was captured for 150 human beings and the eliminating features extracted from the buffered data frames of each sensor were fed to an online clustering methodology so as to form clusters of a defined radius r. Cluster centers were found using 0.1mm as the radius. Decision-making features were used to train the models for PIR sensor and AUS, respectively. A system with three different sensors (viz. PIR, AUS and Camera) and their respective trained models was built tested both indoors as well as outdoors with varying light intensities. For obvious reasons, in dark area, the camera fails to detect anything. However, in that scenario, the decision is made by PIR and AUS readings as per the proposed approach. For the comparison purpose, the robot is tested in the same environment with PIR sensor alone, AUS alone and camera alone. Accuracies of the three experiments are shown in figure 8.

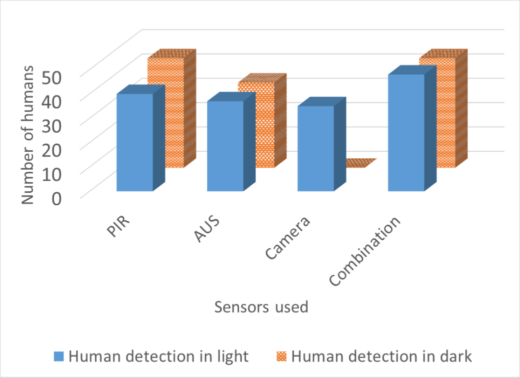

Figure 8. Number of human correctly classified (total number = 45)

4.3. Experiment 3

Human being shows dynamism in behavior such as he can walk at various speeds; he can wear different items of clothing, he can have shown multiple poses, etc. Therefore, in this experiment, considering the varying walking speed of a human, the robustness of the system was tested. For this experiment, the pedometer was used to count the number of steps per minute. To perform the experiment, humans were made to walk at different speeds in front of the sensors in a particular direction, as shown in figure 9 and 10. Figure 9 shows the movement of human being w.r.t. sensor setup of experiment 1 and figure 10 shows the direction of movement of both robot and a human being with respect to the sensors setup of experiment 2. Figure 10 also shows the sensing zone, i.e. when sensory data is considered for further processing.

Results are compiled in a graph as shown in figure 11 and 12. It can be concluded from the graph in figure 11 that as the walking speed of human increases, True Positive Rate (TPR) decreases. Similar pattern can be observed from the graph in figure 12.

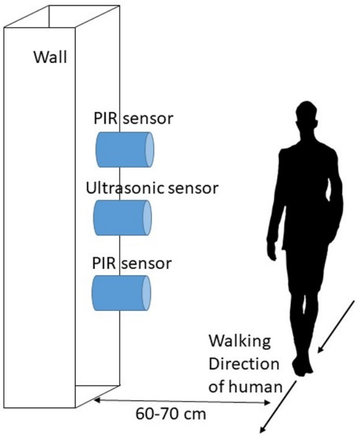

Figure 9. Direction of motion of human being during experiment

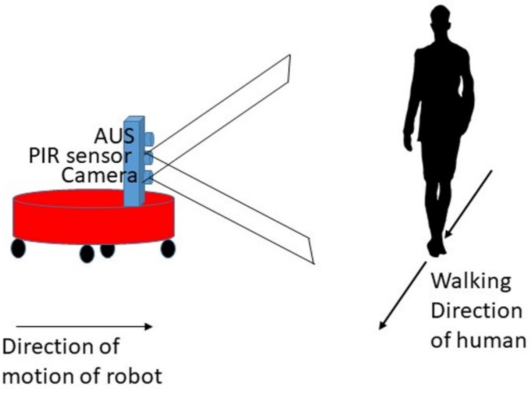

Figure 10. Direction of motion of the mobile robot and human being during experiment

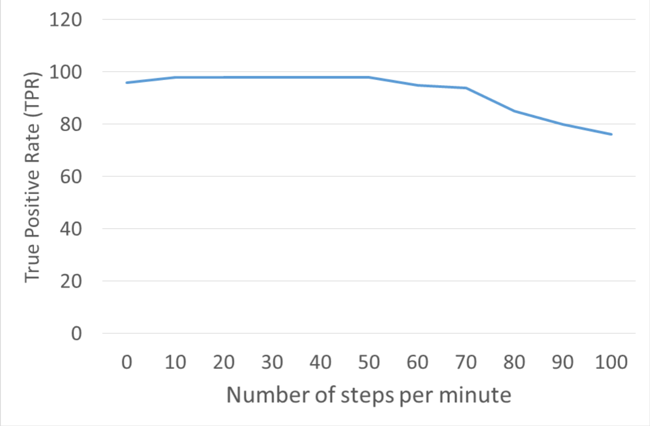

Figure 11. Direction of motion of human being during experiment

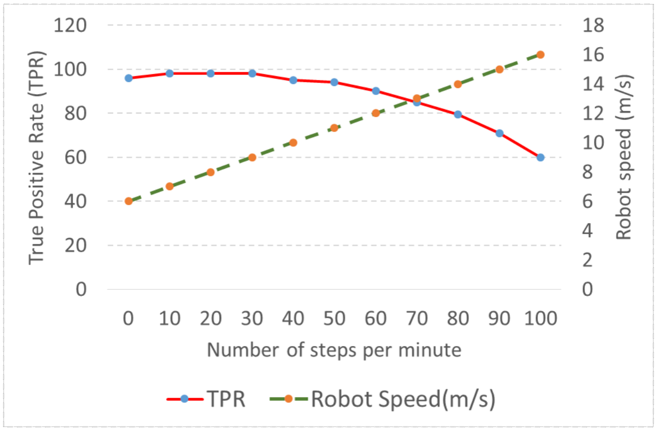

Figure 12. Direction of motion of mobile robot and human being during experiment

The reason for the decreased TPR as observed is that that with high walking speed human move away from sensors so quickly that feature values extracted from the data frame are not able to detect human presence. Thus, the proposed approach has an upper bound on the walking speed of human for human detection.

5. Conclusion

Human sensing is vital to many smart applications such as smart homes, traffic management systems, human-computer interfaces, etc. Since human beings, in general, are always on the move, the use of a dedicated sensor could fail to detect human presence, especially when the ambient parameters around the sensor change. In this paper, a multi-modal human sensing approach has thus been prescribed for overcoming this issue. This work described focused on automating the identification of inappropriate data relayed by some sensors under certain environmental conditions. Experiments reported include both cases - when the sensors are mounted on a static unit (door frame) and also on a mobile robot. The corresponding results reveal that a combination of sensors outperforms the use of individual dedicated sensors for human detection. Analyses of the walking speed of human being have also been studied which endorse the robustness of the approach. However, different directions of motion need exploration in future.

References

[1] Teixeira, T., Dublon, G. and Savvides, A. (2010) A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. *ACM Computing Surveys* **5**(1): 59–69.

[2] Chuah, F.K. and Teoh, S.S. (2020) Thermal sensor based human presence detection for smart home application. In *2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (IEEE): 37–41.

[3] Vallabh, P. and Malekian, R. (2018) Fall detection monitoring systems: a comprehensive review. *Journal of Ambient Intelligence and Humanized Computing* **9**(6): 1809–1833.

[4] Raykov, Y.P., Ozer, E., Dasika, G., Boukouvalas, A. and Little, M.A. (2016) Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*: 1016–1027.

[5] Zappi, P., Farella, E. and Benini, L. (2010) Tracking motion direction and distance with pyroelectric ir sensors. *IEEE Sensors Journal* **10**(9): 1486–1494.

[6] Kessler, E., Malladi, V.V.S. and Tarazaga, P.A. (2019) Vibration-based gait analysis via instrumented buildings. *International Journal of Distributed Sensor Networks* **15**(10): 1550147719881608.

[7] Clemente, J., Li, F., Valero, M. and Song, W. (2019) Smart seismic sensing for indoor fall detection, location, and notification. *IEEE journal of biomedical and health informatics* **24**(2): 524–532.

[8] Ibrahim, S., Zahidin, N.S. and Yunus, M.A.M. (2014) Determination of pipe deformity using an ultrasonic system. *Proceeding of the Electrical Engineering Computer Science and Informatics* **1**(1): 228–231.

[9] Khishe, M. and Mohammadi, H. (2019) Passive sonar target classification using multi-layer perceptron trained by salp swarm algorithm. *Ocean Engineering* **181**: 98–108.

[10] Prati, A., Vezzani, R., Benini, L., Farella, E. and Zappi, P. (2005) An integrated multi-modal sensor network for video surveillance. In *Proceedings of the third ACM international workshop on Video surveillance & sensor networks* (ACM): 95–102.

[11] Yazar, A., Erden, F. and Cetin, A.E. (2014) Multi-sensor ambient assisted living system for fall detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'14)* (Citeseer): 1–3.

[12] Guo, X., Tiller, D., Henze, G. and Waters, C. (2010) The performance of occupancy-based lighting control systems: A review. *Lighting Research & Technology* **42**(4): 415–431.

[13] Wahl, F., Milenkovic, M. and Amft, O. (2012) A distributed pir-based approach for estimating people count in office environments. In *Computational Science and Engineering (CSE), 2012 IEEE 15th International Conference on* (IEEE): 640–647.

[14] Moghavvemi, M. and Seng, L.C. (2004) Pyroelectric infrared sensor for intruder detection. In *TENCON 2004. 2004 IEEE Region 10 Conference* (IEEE), **500**: 656–659.

[15] Sonia, Tripathi, A.M., Baruah, R.D. and Nair, S.B. (2015) Ultrasonic sensor-based human detector using one-class classifiers. In *2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*: 1–6. doi:10.1109/EAIS.2015.7368797.

[16] Chen, L., Wei, H. and Ferryman, J. (2013) A survey of human motion analysis using depth imagery. *Pattern Recognition Letters* **34**(15): 1995–2006.

[17] Bellotto, N. and Hu, H. (2009) Multisensor-based human detection and tracking for mobile service robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **39**(1): 167–181.

[18] Harrison, B.L., Consolvo, S. and Choudhury, T. (2010) Using multi-modal sensing for human activity modeling in the real world. In *Handbook of Ambient Intelligence and Smart Environments* (Springer), 463–478.

[19] Bruno, B., Grosinger, J., Mastrogiovanni, F., Pecora, F., Saffiotti, A., Sathyakeerthy, S. and Sgorbissa, A. (2015) Multi-modal sensing for human activity recognition. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*: 594–600. doi:10.1109/ROMAN.2015.7333653.

[20] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R. and Havinga, P. (2010) Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *Architecture of computing systems (ARCS), 2010 23rd international conference on* (VDE): 1–10.

[21] Lara, O.D. and Labrador, M.A. (2013) A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* **15**(3): 1192–1209.

[22] Rosenblatt, J.K. and Payton, D. (1989) A fine-grained alternative to the subsumption architecture for mobile robot control. In *Proceedings of the IEEE/INNS international joint conference on neural networks*, **2**: 317–324.

[23] Gandomi, A. and Haider, M. (2015) Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* **35**(2): 137–144.

[24] Matthews, D.A., Lerner, M.R., De Vorchik, D.G., Sechrest, S., Zou, S. and Anderson, B.P. (2016), Search tool using multiple different search engine types across different data sets. US Patent 9,323,867.

[25] Sonia, Singh, M., Baruah, R.D. and Nair, S.B. (2017) A voting-based sensor fusion approach for human presence detection. In Basu, A., Das, S., Horain, P. and Bhattacharya, S. [eds.] *Intelligent Human Computer Interaction* (Cham: Springer International Publishing): 195–206.

[26] Yang, M.T. and Huang, S.Y. (2014) Appearance-based multimodal human tracking and identification for healthcare in the digital home. *Sensors* **14**(8): 14253–14277.

[27] Jin, X., Gupta, S., Ray, A. and Damarla, T. (2011) Multimodal sensor fusion for personnel detection. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on* (IEEE): 1–8.

[28] Candanedo, L.M. and Feldheim, V. (2016) Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models. *Energy and Buildings* **112**: 28–39.

[29] Zigel, Y., Litvak, D. and Gannot, I. (2009) A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls. *IEEE Transactions on Biomedical Engineering* **56**(12): 2858–2867.

[30] Xu, H., Yang, Z., Zhou, Z., Shangguan, L., Yi, K. and Liu, Y. (2016) Indoor localization via multi-modal sensing on smartphones. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*: 208–219.

[31] Chung, S., Lim, J., Noh, K.J., Kim, G. and Jeong, H. (2019) Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors* **19**(7): 1716.

[32] Zhao, H., Wang, Z., Qiu, S., Wang, J., Xu, F., Wang, Z. and Shen, Y. (2019) Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. *Information Fusion* **52**: 157–166.

[33] Baruah, R.D. and Angelov, P. (2012) Evolving local means method for clustering of streaming data. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on* (IEEE): 1–8.