

Adaptive Thermal Sensor Array Placement for Human Segmentation and Occupancy Estimation

Abdallah Naser , Ahmad Lotfi  and Junpei Zhong 

Abstract—Thermal sensor array (TSA) offers privacy-preserving, low-cost, and non-invasive features, which makes it suitable for various indoor applications such as anomaly detection, health monitoring, home security, and monitoring energy efficiency applications. Previous approaches to human-centred applications using the TSA usually relied on the use of a fixed sensor location to make the human-sensor distance and the human presence shape fixed. However, placing this sensor in different locations and new indoor environments can pose a significant challenge. In this paper, a novel framework based on a deep convolutional encoder-decoder network is proposed to address this challenge in real-life deployment. The framework presents a semantic segmentation of the human presence and estimates the occupancy in indoor-environment. It is also capable to segment the human presence and counts the number of people from different sensor locations, indoor environments, and human to sensor distance. Furthermore, the impact of the distance on the human presence using the TSA is investigated. The framework is evaluated to estimate the occupancy in different sensor locations, the number of occupants, environments, and human distance with classification and regression machine learning approaches. This paper shows that the classification approach using the adaptive boosting algorithm is an accurate approach which has achieves an accuracy of 98.43% and 100% from vertical and overhead sensor locations respectively.

Index Terms—Thermal sensor array, occupancy estimation, sensor placement, semantic segmentation, deep learning, shallow neural network, adaptive boosting, human-centred approach, adaptive system

I. INTRODUCTION

Occupancy estimation, which refers to counting the number of people in a given area, has a significant impact on facing essential challenges in various sectors. For example, occupancy estimation systems have been applied to manage social distancing during emergencies such as the coronavirus COVID-19 pandemic period. This is primarily crucial in older adult homes to ensure social distancing policies are observed, and alerts are triggered when needed. The need for an accurate occupancy estimation will become more apparent for older adults who prefer to live in their own homes.

Long-term care for older adults is costly and will increase due to the increasing population in the ageing community [1]. Besides, there is low acceptability of nursing homes among older adults [2]. This raises the need for alternative solutions that provide them with independent living in their homes. Despite the fact that homes occupied by an average of 3.14 people per household [3], most of the previous research work

on assisted living to help older adults to live independently in their own homes assumed they live in a single residential environment [4]. Therefore, it is essential to add a new functional layer to distinguish between single and multi-occupancy status in smart home solutions to make them applicable to real-life scenarios. Occupancy estimation systems are also applicable in other areas; for instance, energy used in buildings have at least 40% of the total energy consumed [5]. This percentage can be reduced by adopting the consumption of energy systems based on the number of people participating in the environment at the time these systems work.

Unlike sensing approaches for the outdoor environment, the privacy preservation of the sensor is crucial for deployable domestic applications. However, the trade-off between the performance and the privacy of the visual sensors commonly used to monitor indoor occupancy is one of the deployment limitations of such systems on a larger scale. For example, the vision-based sensor offers high performance in occupancy estimation for the indoor environment. However, it violates user privacy which makes the acceptability of vision-based systems very low by end-users. In contrast, ambient-based sensors such as the infrared sensor (PIR) provide high privacy capability. Nevertheless, intelligent solutions based on the PIR sensors require prior information of the home layout [6], which makes the usability feature of these sensors low. The PIR sensors lack the capability of distinguishing human subjects.

Lately, there have been some works using the Thermal Sensor Array (TSA) to estimate the occupancy in the domestic

Abdallah Naser is a PhD student within the Computational Intelligence and Applications Research Group at Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom (email: abdallah.naser@ntu.ac.uk).

Ahmad Lotfi is the Head of Computational Intelligence and Applications (CIA) Research Group at Nottingham Trent University, Nottingham, NG11 8NS, United Kingdom (email: ahmad.lotfi@ntu.ac.uk).

Junpei Zhong is an independent research fellow within the Computational Intelligence and Applications (CIA) Research Group, School Of Science and Technology, Nottingham Trent University, NG11 8NS, United Kingdom (email: joni.zhong@ntu.ac.uk).

environment. The motivation behind using this sensor is that it features privacy-friendly, high-performance, low-cost, and low-power capabilities. The main gap for most research works utilising this type of sensor is that the sensor's algorithm is not adaptive to operate in different locations (e.g. walls, ceiling). This is because the temperature of the captured thermal objects varies with the distance between these objects and the sensor. Therefore, previous works based on this sensor to detect human existence were usually fixed at a location that has a constant distance between human objects and the sensor [7]. Such poor adaptability of the TSA's location raises significant concerns about the deployability of this sensor applications, for example, placing the sensor on the ceiling of the room reduces the sensor's field of view, which means that more sensors will be needed to cover a wider area.

This paper proposes and investigates an adaptive framework for TSA for human segmentation and occupancy estimation, which has the capabilities to work from different sensor-location configurations, and in a noisy thermal environment. In summary, the main contributions of this paper include:

- a robust and adaptive occupancy estimation framework, able to estimate the occupancy from different sensor locations, human-sensor distance, human-human distance, and in an unseen noisy domestic environment;
- the use of a deep encoder-decoder convolutional neural network to semantic segment the human from a low-resolution TSA's output;
- an investigation of the impact of the sensor operating distance on the human presence;
- the capability of the proposed framework to segment the human presence from a distance of up to 9 meters.

The remaining parts of this paper are organised as follows: in Section II a summary of the related work regarding occupancy estimation in the smart environment is presented. Section III explains the proposed framework architecture. Experimental results are presented and discussed in Section IV, and V followed by pertinent conclusions drawn in Section VI.

II. RELATED WORK

Research has been conducted to investigate the methods for counting the number of occupants in a home environment for different purposes using different sensing approaches [8], [9]. Nevertheless, the usage of the TSA for occupancy estimation is relatively low [10] compared to other sensing methods. [11] have utilised the PIR sensor to distinguish between the single and multi-occupancy environment in the aim of determining the visit time of the older adults in a single inhabitant environment by measuring the randomness of the PIR-based binary data using different entropy measures. The PIR sensor has been used in multi-occupancy indoor environment [12]–[15]. However, these works were only able to identify whether more than one person occupies the environment without providing an exact estimation of the number of people. Furthermore, they relied heavily on the sensor layout and the ground-truth annotated sensor data.

Other privacy-friendly sensing approaches reported in [16]–[18] have proposed multi-modal systems to solve the problem

of counting the number of people in the home environment. They were based on the usage of multiple environmental sensors such as lighting, temperature, movement, CO, CO₂, and humidity. Although the multi-modal approach increases these systems' performance, it raises serious questions about the applicability of these systems in real-use case scenarios. This is because these works have assumed that ventilation does not affect the performance of their proposed systems. However, ventilation may alter the level of humidity, CO, and CO₂ in the home environment, resulting in a wrong estimate of the occupancy.

Recently, [19] overcame the PIR sensor's deployability problem to count the number of people in a home environment by introducing a new algorithm, based on continually tracking each person's location in the home, without requiring other additional information such as the ground-truth annotated sensor data. However, it is not clear whether the algorithm will work when two people simultaneously walk alongside each other, for example, the caregiver who helps the older adult to walk.

Beltran et al. [20] suggested a multi-modal system, consisting of a PIR sensor and a TSA to estimate the occupancy. The purpose of using the PIR sensor was to detect the empty occupancy environment and the TSA to count the number of people in the environment. However, the proposed system may fail to estimate the occupancy when a person has been inactive for a long time, for example, sleeping as the probability of identifying human radiation as the background temperature increases. Furthermore, the TSA is placed in a fixed location, which results in the system not being work at a different sensor location. Other works [7], [10], [21]–[27] that used the TSA to estimate the occupancy also contain similar sensor location adaptation problems.

Naser et al. [28] proposed a framework capable of distinguishing between human radiation and animal pet radiation acquired by a TSA. It is able to estimate the entropy point of animal pet and human radiations as a feature to be classified. The motivation behind this work was to remove any noise heat such as hot objects similar to the human temperature in order to only keep the human presence in the thermal scene to further human activity recognition. Other TSA applications based on human detection have also been proposed [29]–[33] with the same problem of sensor location adaptability.

The use of semantic segmentation using the deep convolutional encoder-decoder network is not explored yet. To the best of our knowledge, the deep learning-based semantic segmentation of the human presence from different TSA placements has been proposed for the first time in this paper methodology.

III. HUMAN-CENTRED OCCUPANCY ESTIMATION USING THERMAL SENSOR ARRAY

The data acquisition stage of the proposed framework is based on using the TSA. Specifically, the MLX90640 sensor [34] is used. The sensor measures the temperatures of objects in its field of view and returns the information as a 32×24 matrix. This sensor is chosen to estimate the occupancy as it

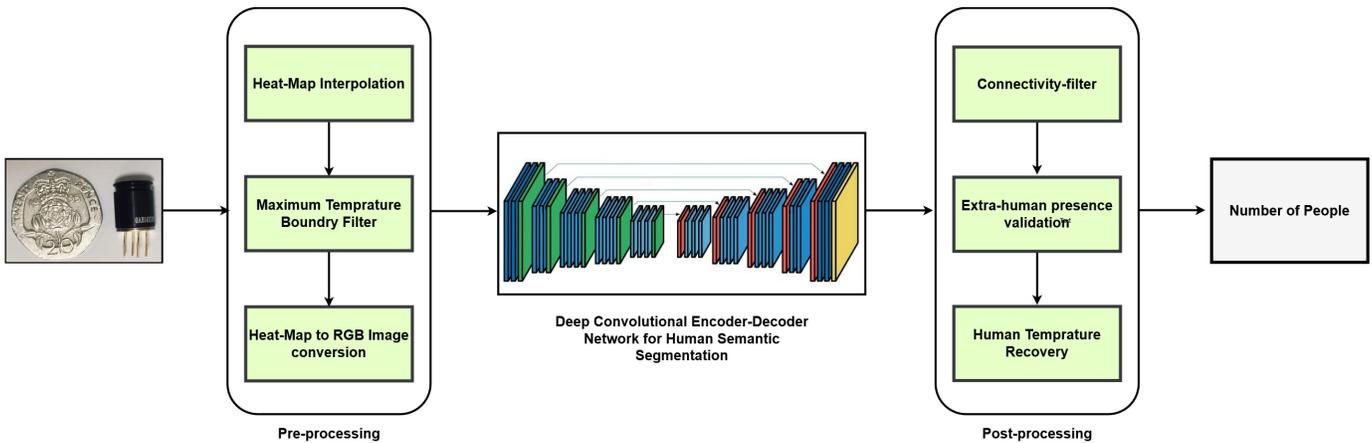


Fig. 1: The proposed framework to estimate the number of people in the thermal scene obtained using the TSA after applying a set of pre-processing techniques, a deep convolutional encoder-decoder network to semantic segment the human presence, and post-processing techniques that consider the characteristics of the used sensor.

is a privacy-preserving and low-cost sensor [35], which makes it suitable for home environment applications. Moreover, the refresh rate of the sensor is between 0.5 Hz and 64 Hz , and this makes it capable of detecting swift human movements.

A schematic diagram of the proposed system designed to suit the characteristics of the TSA is shown in Fig 1. For example, the TSA is not light sensitive compared to the camera sensor. However, the TSA is sensitive to environmental temperature and of low resolution. Therefore, it is crucial to develop a systemic framework that depends on the type of the used sensor per itself. Besides, the proposed framework segments the human presence from a noisy heat-map using a deep convolutional encoder-decoder network. A set of pre-processing and post-processing techniques are introduced to make the sensor output applicable to the proposed segmentation technique. A detailed description of the proposed framework stages is provided below.

A. Pre-processing

The pre-processing stage consists of three sequential phases. The first phase is to increase the resolution of the original heat-map obtained by the TSA through interpolating the original 32×24 temperature matrix to 96×72 by repeating refined temperature values 3 times in each dimension. Fig 2(a) shows an example of the raw heat-map and Fig. 2(b) shows the result of interpolating the heat-map by a factor of 3. This factor is chosen because it provides the best visual resolution of the TSA sensor-based heat-map and is proportional to the size of the input image of the used convolutional encoder-decoder network described in Sec. III-B to segment the human presence.

The detected human temperatures vary depending on the distance between the human and sensor locations. Also, the covered parts of the human body have a higher temperature than the uncovered parts. The normal maximum human temperature detected using the MLX90640 sensor from a nearby point is 33°C . Thus, the second phase of the pre-processing stage is to filter any value that has a higher-temperature than

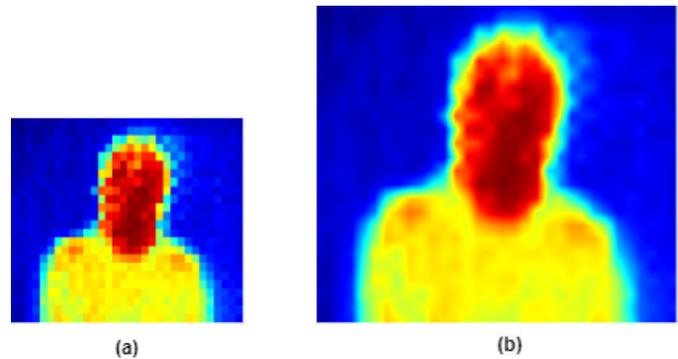


Fig. 2: Heat-maps visualisation of (a) original heat-map, (b) interpolated heat-map.

33°C by converting it to the minimum temperature value in the temperature matrix. By doing so, thermal noises such as a hot kettle will be removed, and the human presence will remain as a foreground object in the thermal scene.

It should be noted that Fig 2 shows a visualisation of the temperature matrix using colour mapping to facilitate the reader's visibility of sensor data visually. Moreover, the mentioned maximum filter converts any temperature above the set limit to the minimum temperature in the acquired temperature matrix to maintain the variance between the temperature values. If the filter converts the high temperatures to zeros, this will cause the variance to be high, which results in a different colour scheme. Continually, the third phase of the pre-processing stage is to export the colour mapped matrix to an RGB image.

B. Semantic Segmentation for the Human Heat-Map

To separate a human subject from the RGB image produced after the pre-processing step, a semantic segmentation is applied. Semantic segmentation aims to classify each pixel in the image into a corresponding class. In contrast, object detection classifies the regions of the image into a different class and

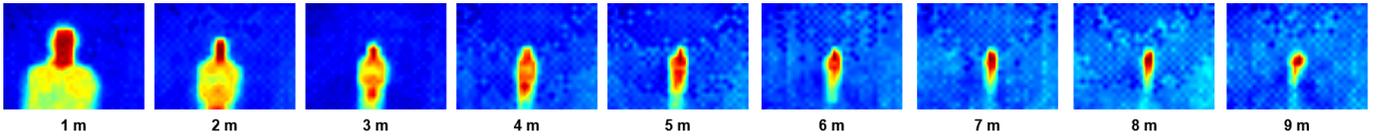


Fig. 3: The Human presence using the thermal sensor array at different distances after applying the proposed pre-processing techniques.

draws a bounding box around the object of interest. In order to make TSA adaptive to different locations, object detection may not work well due to the high intra-class variation of the human object in the thermal scene at different sensor locations and human-sensor distances. Fig 3 shows the human presence in the thermal scene after applying the pre-processing techniques at distances from 1 m to 9 m. It can be observed from the Fig 3 that the human presence changes its size and topology with respect to the distance. Therefore, instead of detecting the human object, this paper proposes to use a deep convolutional encoder-decoder network to classify each pixel in the thermal scene acquired by the TSA to either human or background classes.

The convolutional network architecture proposed in [36] is used here. The first path of this network, the encoder, is used to capture the context of the thermal image. The encoder consists of a typical stack of convolutional and max-pooling layers. The following part is the decoder part, which is the symmetric expanding part that enables the precise localisation using transposed convolutions. In total, the architecture of this network has 23 convolutional layers. The reason behind choosing this network architecture with the TSA is that it is designed for low-resolution images and does not require an extensive dataset as it performs excessive data-augmentation techniques per itself.

The network is optimised using Adaptive Moment Estimation (Adam) [37] to compute the adaptive learning rates for each parameter using gradient descent optimisation approach. This optimiser computes the first squared gradients m_t (the mean) and the second squared gradients v_t (the uncentered variance) as follow:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2)$$

m_t is the estimate of the first moment of the gradient, where v_t is the estimate of the second moment of the gradient. These estimates are biased towards zero, particularly during the initial time steps when the decays rates are small (i.e. β_1 and β_2 are close to 1). To compute the bias-corrected first and second moment estimates:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (4)$$

Then, the network weight update as follow:

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (5)$$

The initial default value for β_1 is 0.9, β_2 is 0.999, and 10^{-8} for ϵ .

The network is trained with a dataset containing 47 labelled thermal images acquired from one human object from a vertical position at distances from 0.5 m to 9 m. The output of this network is a matrix that shows the class (human or background) of each pixel, i.e. a binary mask that shows the human presence in the scene.

C. Post-processing

The semantic segmentation technique proposed in the previous section has one drawback which comes from the low-resolution thermal sensing methodology itself. Unlike the RGB camera, the TSA also senses the thermal noises left by humans even when they have left the thermal scene, which has a similar temperature to the human body. As a result, the semantic segmentation may classify these noisy pixels belong to human. To overcome this drawback, a post-processing stage containing three phases is introduced.

The first phase is the connectivity-filter to remove thermal noises that have a similar human temperature, for example, noises that generated by the human body or a warm object that have similar temperature with the humans such as a warm cup of coffee. The connectivity-filter is based on morphological operations. Specifically, the 8-connected algorithm [38] finds each connected component in the mask generated by the semantic segmentation network. The methodology behind this algorithm is to cluster each object based on the connectivity of its values. Each value in the mask mentioned above belongs to the same object if it has the same value (0 or 1) and is connected along the diagonal, horizontal, or vertical direction. Any connected component which is less than or equal to 30 pixels is considered to be a thermal noise and to be removed. This size is calculated based on finding the minimum human-size acquired using the TSA from a maximum distance of 9 m. Next, the second image processing technique used is the Flood-Fill algorithm [39] to fill holes in human existence in the mask obtained from the last step. These holes may appear as a result of thick clothing, which reduces the temperature acquired for the human body by using the TSA. This decreases in the human temperature values can be seen as background pixels by the semantic segmentation network.

Since the colourmap is a colourful representation of the scene temperatures, human at far long distances from each other lead to colour the distant human presence to background colours. Therefore, the second phase of the post-processing

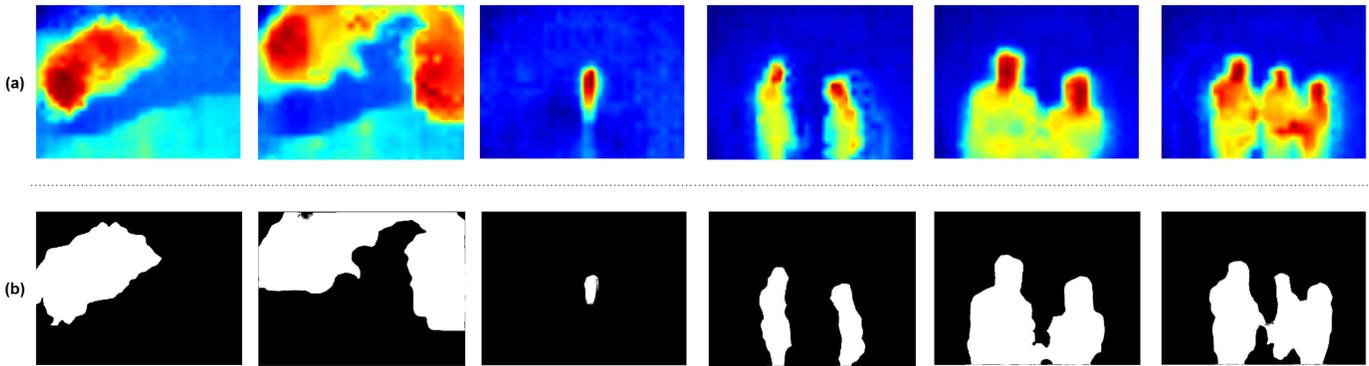


Fig. 4: Illustrative results of the proposed framework, (a) the thermal images after applying the pre-processing techniques, (b) the human presence locations after using semantic segmentation, connectivity filter, and the extra human validation techniques.

stage is to repeat the previous steps starting from obtaining a new RGB image of the interpolated heat-map without the locations of segmented human presence and adding the new segmented human, if found to the previously segmented mask. This extra-human presence validation phase repeats until the scene turned to an empty occupancy.

Fig 4 shows a few examples of applying the pre-processing, semantic segmentation, extra-human validation, and the connectivity filter to locate the human presence in the thermal images obtained by the TSA. Figs 4 (a) shows the pre-processed heat-maps in different sensor location, human pose, human-human distance, and sensor-human distance, where Figs 4 (b) shows the corresponding human presence in these thermal images after applying the semantic segmentation, extra-human validation, and the connectivity filter. As mentioned earlier, the encoder-decoder network for the semantic segment the human heat-map were only trained to detect the human presence with only one person from a vertical position, and these illustrative examples are all unseen data for the network.

At this stage, the human presence in the heat-maps converted to RGB images has been determined. However, human temperature values are lost because they are converted to RGB values to focus on pixel intensity rather than the temperature value. Therefore, the third phase of the post-processing stage is to restore human temperatures through multiplying the masks obtained using the semantic segmentation, connectivity filter, and the extra human validation, which shown in 4 (b), by the interpolated heat-maps to count the number of people as described in the following section.

D. Estimating the Occupancy Using Machine Learning Approach

The final stage of the proposed methodology is to count the number of people using the TSA. If the sum of the computed mask from the semantic segmentation and the post-processing techniques is zero, it means no human presence in the thermal scene. Otherwise, two different machine learning approaches were evaluated to estimate the occupancy. In particular, a classification model using Adaptive Boosting (AdaBoost) [40] and a regression model using shallow neural network [41] have been developed to count the number of people after

segmenting the human presence from the thermal images. The primary difference between these two approaches is that classification deals with the problem of predicting a discrete class label, where the output of the regression is a continuous quantity.

1) *Classification for Occupancy Estimation* : Boosting is a type of classification algorithms, which seeks to boost the accuracy of a weak learner (a classifier that performs poor in a given classification problem) by having multi-learners and then fuse them to a strong learner. The weak learners in AdaBoost are decision trees with a single split, called decision stump. The weights of the observations in the first iteration are equal. The classification model in AdaBoost is then improved by training the decision trees sequentially. Each of these weak learners tries to correct its predecessor by focusing on the incorrect observations in the previous iterations by updating weights. Therefore, misclassified objects in the previous iteration carry more weights than the correct classified observations to enable the weak learner to classify the hard samples in the training dataset.

In this paper, an extension of the AdaBoost algorithm to a multi-class problem called AdaBoost.M2 described in Algorithm 1 is used as a holistic classification approach. The e -th training set for this algorithm includes the segmented human heat-maps x obtained after applying the pre-processing, semantic segmentation, and the post-processing techniques, where y represents the class label (the number of people belongs to the set Y). The distribution $D_t(i, y)$ is maintained over the training set E and updated sequentially in each iteration c based on the output of that iteration. As mentioned earlier, misclassified training samples carries more weights than those correctly observed in the next iteration. By doing so, the update rule guarantees the upper bounds on training and generalisation error rates.

2) *Regression for Occupancy Estimation*: The second step to estimate the occupancy is a regression through estimating the relationships between the segmented heat-maps and the number of people in the scene using an artificial neural network. In particular, a shallow neural network with only one hidden layer with sigmoid neurons and one output layer is used to determine the number of people in the scene. The input of

Algorithm 1 The AdaBoost.M2 algorithm to classify each thermal scene into a class label, which represents the number of people in each scene.

Input: 1) Series of E of training samples $\{(x_1, y_1), \dots, (x_e, y_e)\}$ with labels $y_e \in Y = \{1, \dots, j\}$
 2) D represents the distribution over the E samples
 3) Weak learning algorithm **DecisionTree**
 4) Counter C for the number of iterations
 1: **Initialize:** The weight vector: $w_{i,y}^1 = D(i)/(J-1)$, where $i = 1, \dots, E, y \in Y - \{y_i\}$.
 2: **for** $c = 1, 2, \dots, C$ **do**
 3: $q_c(i, y) = \frac{w_{i,y}^c}{\sum_{y \neq y_i} w_{i,y}^c}$
 4: $D_c(i, y) = \frac{W_i}{\sum_{i=1}^E W_i} (y \neq y_i)$
 5: **Call DecisionTree.** \triangleright Given the distribution D , and label weighting function q_c ; return a hypothesis $G_c : X \times Y \rightarrow [0, 1]$
 6: $\epsilon_c = \frac{1}{2} \sum_{i=1}^E D_c(i, y) \left(1 - g_c(x_i, y_i) + \sum_{y \neq y_i} q_c(i, y) g_c(x_i, y) \right)$ \triangleright Calculate the pseudo-loss of g_c .
 7: $\beta_c = \epsilon_c / (1 - \epsilon_c)$
 8: $w_{i,y}^{c+1} = w_{i,y}^c \beta_c^{\frac{1}{2}(1+g_c(x_i, y_i) - g_c(x_i, y))}$ \triangleright Update the new weights vector, for $i = 1, \dots, E, y \in Y - \{y_i\}$
 9: **end for**
Output: $g_f(x) = \arg \max_{y \in Y} \sum_{c=1}^C \log \frac{1}{\beta_c} g_c(x, y)$

the network is the segmented heat-maps, and the output is the number of people. The network is trained using the Levenberg-Marquard backpropagation algorithm [42]. This algorithm aims to minimise the sum of the squares of deviations $S(\beta)$ of a set of pair $n(x_i, \hat{y}_i)$ of input heat-maps x and the number of people \hat{y} by finding the parameters β of the model output $f(x, \beta)$.

$$\hat{\beta} \in \operatorname{argmin}_{\beta} S(\beta) \equiv \operatorname{argmin}_{\beta} \sum_{i=1}^n [\hat{y}_i - f(x_i, \beta)]^2 \quad (6)$$

The training of the network terminates when an increase in the mean square error of the validation dataset is detected. In this network, in contrast to classification, the result in the regression is a continuous value. Therefore, the output \hat{y} is rounded to the nearest decimal point as our goal is to estimate the discrete number of people.

IV. EXPERIMENTS

To evaluate the performance of the proposed methodology, experiments were conducted with two different configurations of sensor locations. Correspondingly, two kinds of data were collected in different indoor environments and sensor locations.

At the first stage, the data was obtained while placing the sensor on a vertical position, as shown in Fig 5(a). Within this stage, two subsets of data were collected. The first subset consists of 47 thermal scenes in which only one person moves in the sensor field of view up to 9 meters long. The thermal objects in this subset are then labelled as either human or background objects. The second subset is used to analyse the effect of human distance on sensor performance. It is obtained in a human presence at distances of 0.5 m to 9 m away from the sensor. This subset is collected every 0.5 m, and its size is 325. The third subset collected with one, two, and three

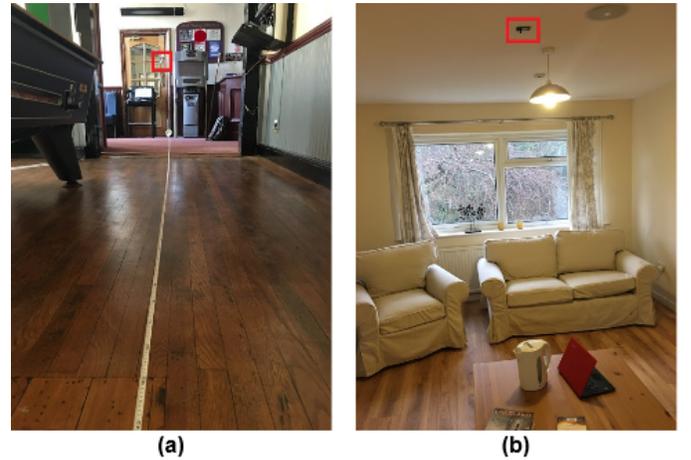


Fig. 5: Data collection stages from two different indoor environments, (a) the sensor is placed on the wall, (b) the sensor is on the ceiling.

different occupants moving in the sensor field of view to assess the performance of the occupancy estimation system. The size of this subset is 214.

The second stage aims to assess the adaptability of the proposed framework to work in a different indoor environment and sensor location. In this stage, the sensor is placed on the ceiling of the room, as shown in Fig 5(b). The dataset was collected in four different scenarios: one, two, three, and four occupants were moving in the scene. The size of this dataset is 203. In addition to the above dataset, 128 thermal scenes were collected from two empty-human environments and sensor locations with thermal noises such as hot kettle, laptop, and heater when turned on to evaluate the proposed framework ability to detect the empty occupancy environment. In total, 917 thermal scenes collected to conclude the results

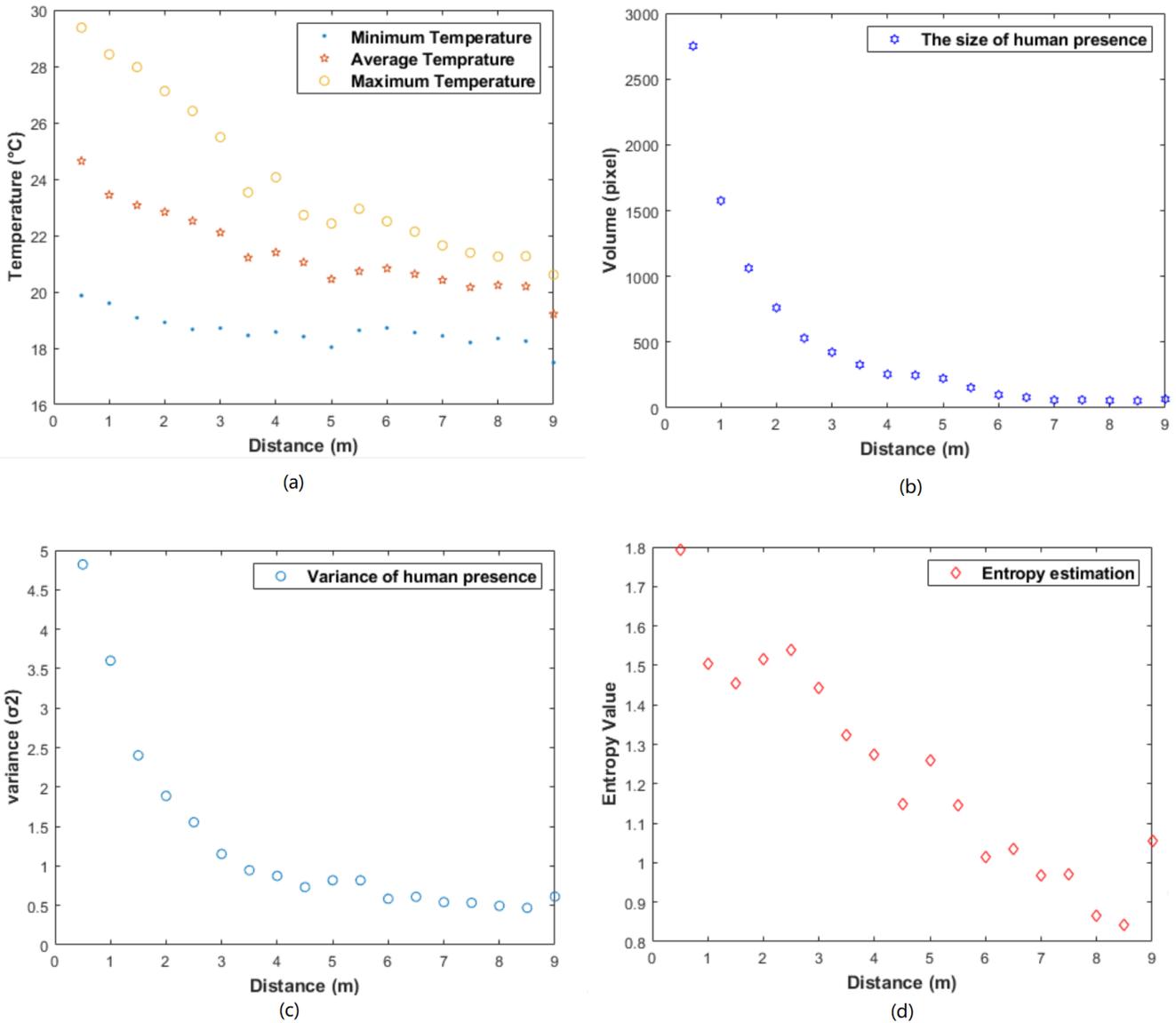


Fig. 6: The effect of distance on human presence, (a) the minimum, maximum, and average temperatures, (b) the size of the human presence in the thermal scene, (c) the variance in human temperature, (d) the estimate of the entropy.

of this paper.

A. Human-Sensor Analysis

To assess the impact of the distance on the human presence using TSA, a subset of the described dataset above is used. The dataset has different human heat-maps every 0.5m and up to 9m in length as described earlier, which makes a total of 18 distance steps. Further, an average human heat-map at every 0.5m is computed. This average human heat-map for each distance step aims to avoid biased analysis of a random selection of a human heat-map.

Fig 6(a) shows the effect of the human-sensor distance on the value of the acquired temperature. Specifically, the minimum, maximum, and average temperature of a human presence at different distances are shown. It can be seen the acquired human temperatures decrease when a person moves

away from the sensor.

As shown in Fig 6(b), and Fig 6(c) the sizes and the temperature variances of human presence vary depends on the relative location of the human subject to the sensor. One observation can be drawn from these figures; there is a significant decrease in the size of the human presence and temperature variance in the first 1.5m. The decrease in the size of the human presence and the variance continues beyond 1.5m, but with a smaller interval.

The common pattern of Fig 6(a), Fig 6(b), and Fig 6(c) is that they all have relatively stable values starting from a distance of 6m meters and beyond. However, this stability does not exist in the entropy estimation of human presence at different distances, as shown in Fig 6(d). Besides, it can be identified that a linear relationship between the entropy metric for the human presence and the distance for every 1m.

Based on these results, it can be concluded that the TSA and all the metrics calculated based on the thermal images are quite sensitive to the distance. Furthermore, the calculated statistical metrics, in particular, the entropy point estimate can be used to determine the human distance from the sensor using a suitable function approximator (e.g. our shallow ANN).

B. Occupancy Estimation Experimental Results

The first experiment was to use the collected empty occupancy dataset and evaluate the performance of the proposed framework to detect the empty human environment before proceeding to the classification or regression models as described in Section III-D. The proposed framework was able to detect the empty occupancy with 100% accuracy. This performance validates the proposed pre-processing, semantic segmentation, and post-processing techniques in segmenting only the human presence from the thermal scene.

The second experiment was to examine the performance of the classification approach using AdaBoost. The used dataset obtained from the vertical position. In this validity, the dataset is divided into 70% for model training and 30% for testing. The performance of this classification model in occupancy estimation is 98.43% achieved accuracy.

The third experiment is to validate the performance of the classification model with a different machine learning approach. In this experiment, the regression approach based on a shallow neural network is used. The same data used to train and test the classification model, is also used to train and test the shallow network. This dataset is divided into 70% for training, validating and testing the shallow neural network during the network training stage, and 30% for testing the performance of the trained shallow neural network to count the number of people in the thermal scene. The performance of the regression model in occupancy estimation is 93.75%.

The results of the above experiments show that the classification model has better accuracy than the regression model. The assumed reason for the lower accuracy in the regression was due to the uncertainty for some of the regression outputs. Hence, as mentioned above, the output of the regression is a continuous numerical value. Since this paper is concerned with counting a discrete number of people, the output of the regression model is rounded to the nearest decimal number. Uncertainty occurs when the output has one half (e.g. 1.5, 2.5, etc.).

V. DISCUSSION

This experiment has two main aims. The first aim is to validate the adaptability feature of the proposed framework for operating in a different indoor environment and sensor location. Therefore, the sensor is installed on the ceiling of the room in a different home. Hence, the trained deep convolutional encoder-decoder network from a vertical sensor location is used to segment the human presence from the overhead thermal scenes. The second aim of this experiment is to validate the possibility to parametrise the proposed framework to predict more people in the scene. Therefore, the dataset collected for this experiment contains more people

compared to the previous experiment described in Section IV-B.

The configuration of this experiment dataset is divided in the same way as the vertical sensor dataset used in the previous experiment is divided with the same classification and regression approaches. Regarding the classification approach using the AdaBoost algorithm, the system achieves an accuracy of 100% in estimating the occupancy from 1 to 4 different occupants moving in the thermal scene. However, the system achieves 58.33% accuracy using a shallow neural network to estimate the occupancy.

The classification approach using the AdaBoost algorithm shows a high-performance in estimating the occupancy in different sensor location and environment using the proposed pre-processing, human segmentation, and post-processing techniques. However, the regression approach shows a lower performance in estimating the occupancy from the overhead thermal scenes. This was due to the high uncertainty of the output of the regression model in the overhead thermal scenes compared to vertical-based scenes. On the other hand, the regression approach is a promising approach to estimate the occupancy in an unsupervised learning problem, in which the number of people is greater than the number used to train the model.

Furthermore, the proposed approach of this paper has been compared with the state-of-the-art approaches. Table I shows a comparison of the proposed approach and other notable works in terms of the experimental setup, adaptability of the sensor placement, occupancy estimation technique, and experimental results.

VI. CONCLUSION AND FUTURE WORK

Thermal sensor array (TSA) is a privacy-friendly and low-resolution sensing technique. This paper has proposed a novel framework which is adaptive to TSA location for human segmentation and occupancy estimation. The proposed framework is evaluated in different indoor environments, sensor locations, the number of occupants, and machine learning approaches. Furthermore, an investigation in the impact of human distance on the TSA is provided. From the obtained results, it can be concluded that the use of a deep convolutional encoder-decoder network with the pre-processing and post-processing techniques that consider the characteristics of the low-resolution TSA is an accurate approach for human segmentation. Besides, the adaptive boosting classification algorithm provides accurate results to estimate the occupancy in the proposed framework.

This research confirms that human-centred applications using TSA should focus on segmenting human presence rather than human shape detection. This is because the intra-class variations in the human presence using the TSA is relatively high compared to the normal camera with respect to the human-sensor distance and sensor location. Future work could be conducted to evaluate the accuracy of the entropy point estimate and other statistical measurements in predicting the distance between the human and the sensor location. Also, to assess the reliability of this low-cost TSA to detect the

TABLE I: A comparison of the experimental setup, sensor placement, occupancy estimation method and results of the proposed system with the state-of-the-art.

Paper/Method	Ref.	Sensor	Adaptive Sensor Placement	Sensor Location	Estimation Method	Accuracy
Beltran et al.	[20]	TSA, PIR	Non-adaptive	Ceiling	K-Nearest Neighbors	NA
Gomez et al.	[22]	TSA	Non-adaptive	Wall	Convolutional Neural Network	53.7%
Tyndall et al.	[10]	TSA, PIR	Non-adaptive	Ceiling	K* algorithm	82.56%
Metwaly et al.	[25]	TSA	Non-adaptive	Ceiling	Feedforward Neural Network	98.90%
Proposed in this paper		TSA	Adaptive	Wall	AdaBoost.M2	98.43%
	Wall			Shallow neural network	93.75%	
	Ceiling			AdaBoost.M2	100%	
	Ceiling			Shallow neural network	58.33%	

abnormal human temperature in the human presence using reliable medical thermometers. Besides, future work could also be directed to maximise the occupancy estimation coverage area through a multi-sensor fusion approach.

ACKNOWLEDGMENT

A.N. is supported by Nottingham Trent University through a fully-funded Scholarship Scheme.

REFERENCES

- [1] U. S. G. A. Office, *Long term care: aging baby boom generation will increase demand and burden on federal and state budgets*. US General Accounting Office, 2002.
- [2] S. Quine and S. Morrell, "Fear of loss of independence and nursing home admission in older australians," *Health & social care in the community*, vol. 15, no. 3, pp. 212–220, 2007.
- [3] J. Han, S. Pan, M. K. Sinha, H. Y. Noh, P. Zhang, and P. Tague, "Smart home occupant identification via sensor fusion across on-object devices," *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 3-4, p. 23, 2018.
- [4] A. Benmansour, A. Bouchachia, and M. Feham, "Multioccupant activity recognition in pervasive smart home environments," *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, p. 34, 2016.
- [5] A. Toleikyte, L. Kranzl, and A. Müller, "Cost curves of energy efficiency investments in buildings—methodologies and a case study of lithuania," *Energy Policy*, vol. 115, pp. 148–157, 2018.
- [6] C. Jiang and A. Mita, "Automatic spatial attribute and travel pattern generation for simulating living spaces for elderly individuals living alone," *Building and Environment*, p. 106776, 2020.
- [7] A. Naser, A. Lotfi, J. Zhong, and J. He, "Heat-map based occupancy estimation using adaptive boosting," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2020, (accepted).
- [8] H. Saha, A. R. Florita, G. P. Henze, and S. Sarkar, "Occupancy sensing in buildings: A review of data analytics approaches," *Energy and Buildings*, vol. 188, pp. 278–285, 2019.
- [9] Z. Chen, C. Jiang, and L. Xie, "Building occupancy estimation and detection: A review," *Energy and Buildings*, vol. 169, pp. 260–270, 2018.
- [10] A. Tyndall, R. Cardell-Oliver, and A. Keating, "Occupancy estimation using a low-pixel count thermal imager," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3784–3791, 2016.
- [11] A. Howedi, A. Lotfi, and A. Pourabdollah, "Exploring entropy measurements to identify multi-occupancy in activities of daily living," *Entropy*, vol. 21, no. 4, p. 416, 2019.
- [12] A. R. Pratama, A. Lazovik, and M. Aiello, "Office multi-occupancy detection using ble beacons and power meters," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2019, pp. 0440–0448.
- [13] A. S. Crandall and D. J. Cook, "Coping with multiple residents in a smart environment," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, no. 4, pp. 323–334, 2009.
- [14] Q. Li, W. Huangfu, F. Farha, T. Zhu, S. Yang, L. Chen, and H. Ning, "Multi-resident type recognition based on ambient sensors activity," *Future Generation Computer Systems*, 2020.
- [15] J. Yin, M. Fang, G. Mokhtari, and Q. Zhang, "Multi-resident location tracking in smart home through non-wearable unobtrusive sensors," in *International Conference on Smart Homes and Health Telematics*. Springer, 2016, pp. 3–13.
- [16] Z. Chen, C. Jiang, M. K. Masood, Y. C. Soh, M. Wu, and X. Li, "Deep learning for building occupancy estimation using environmental sensors," in *Deep Learning: Algorithms and Applications*. Springer, 2020, pp. 335–357.
- [17] S. Mamidi, R. Maheswaran, and Y. Chang, "Smart sensing, estimation, and prediction for efficient building energy management," in *Multi-agent Smart Computing Workshop*, 2011.
- [18] K. P. Lam, M. Höynck, B. Dong, B. Andrews, Y.-S. Chiou, R. Zhang, D. Benitez, J. Choi *et al.*, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," *IBPSA Building Simulation*, vol. 145, pp. 1452–1459, 2009.
- [19] T. Wang and D. Cook, "smrt: Multi-resident tracking in smart homes with sensor vectorization." *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [20] A. Beltran, V. L. Erickson, and A. E. Cerpa, "Thermosense: Occupancy thermal based sensing for hvac control," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.
- [21] Y. Yuan, X. Li, Z. Liu, and X. Guan, "Occupancy estimation in buildings based on infrared array sensors detection," *IEEE Sensors Journal*, vol. 20, no. 2, pp. 1043–1053, 2019.
- [22] A. Gomez, F. Conti, and L. Benini, "Thermal image-based cnn's for ultra-low power people recognition," in *Proceedings of the 15th ACM International Conference on Computing Frontiers*, 2018, pp. 326–331.
- [23] H. Mohammadmoradi, S. Munir, O. Gnawali, and C. Shelton, "Measuring people-flow through doorways using easy-to-install ir array sensors," in *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2017, pp. 35–43.
- [24] M. Cokbas, P. Ishwar, and J. Konrad, "Low-resolution overhead thermal triprwire for occupancy estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 88–89.
- [25] A. Metwaly, J. P. Queralta, V. K. Sarker, T. N. Gia, O. Nasir, and T. Westerlund, "Edge computing with embedded ai: Thermal image analysis for occupancy estimation in intelligent buildings," in *Proceedings of the INTelligent Embedded Systems Architectures and Applications Workshop 2019*, 2019, pp. 1–6.
- [26] L. Walmsley-Eyre and R. Cardell-Oliver, "Hierarchical classification of low resolution thermal images for occupancy estimation," in *2017 IEEE 42nd conference on local computer networks workshops (LCN Workshops)*. IEEE, 2017, pp. 9–17.
- [27] G. Cerutti, R. Prasad, and E. Farella, "Convolutional neural network on embedded platform for people presence detection in low resolution thermal images," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7610–7614.
- [28] A. Naser, A. Lotfi, J. Zhong, and J. He, "Human activity of daily living recognition in presence of an animal pet using thermal sensor array," in *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2020, pp. 1–6.
- [29] Z. Liu, M. Yang, Y. Yuan, and K. Y. Kan, "Fall detection and personnel tracking system using infrared array sensors," *IEEE Sensors Journal*, 2020.
- [30] L. Tao, T. Volonakis, B. Tan, Z. Zhang, Y. Jing, and M. Smith, "3d convolutional neural network for home monitoring using low resolution thermal-sensor array," in *3rd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2019)*, 2019, pp. 1–6.
- [31] Z. Chen, Y. Wang, and H. Liu, "Unobtrusive sensor-based occupancy facing direction detection and tracking using advanced machine learning algorithms," *IEEE Sensors Journal*, vol. 18, no. 15, pp. 6360–6368, 2018.

- [32] S. Savazzi, V. Rampa, S. Kianoush, A. Minora, and L. Costa, "Occupancy pattern recognition with infrared array sensors: A bayesian approach to multi-body tracking," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4479–4483.
- [33] N. Gu, B. Yang, and T. Zhang, "Dynamic fuzzy background removal for indoor human target perception based on thermopile array sensor," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 67–76, 2019.
- [34] Far infrared thermal sensor array. Accessed: 2019-12-19. [Online]. Available: <https://www.melexis.com/en/product/MLX90640/Far-Infrared-Thermal-Sensor-Array>
- [35] F. Riquelme, C. Espinoza, T. Rodenas, J.-G. Minonzo, and C. Taramasco, "ehomesensors dataset: An infrared thermal sensor dataset for automatic fall detection research," *Sensors*, vol. 19, no. 20, p. 4565, 2019.
- [36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [38] H. Heijmans, "Introduction to connected operators," *Nonlinear filters for image processing*, pp. 207–235, 1999.
- [39] Y. Lu, J. Huang, and P. Zheng, "A cfd-dem study of bubble dynamics in fluidized bed using flood fill method," *Chemical Engineering Journal*, vol. 274, pp. 123–131, 2015.
- [40] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [41] M. Bianchini and F. Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 8, pp. 1553–1565, 2014.
- [42] S. Sapna, A. TAMILARASI, M. P. Kumar *et al.*, "Backpropagation learning algorithm based on levenberg marquardt algorithm," *Comp Sci Inform Technol (CS and IT)*, vol. 2, pp. 393–398, 2012.



Abdallah Naser obtained his BSc in Computer Engineering from Cyprus International University with High-Honor and his MSc in Information Security and Biometrics with Distinction from the University of Kent. He is currently pursuing his PhD degree at Nottingham Trent University, UK. Abdallah granted a patent for inventing a new way to secure online accounts using concepts of pattern recognition from the Palestinian patent office. Besides, he was one of the winning teams in K-Hackathon competition organised by Middle

East Technical University for developing a classification IoT system to detect car accidents in addition to other academic awards. His current research interests focus on utilising privacy-preserving sensors for occupancy monitoring through image processing and machine learning approaches.



Ahmad Lotfi (M'96-SM'08) received his BSc. and MTech. in control systems from Isfahan University of Technology, Iran and Indian Institute of Technology Delhi, India respectively. He received his PhD degree in Learning Fuzzy Systems from University of Queensland, Australia in 1995. He is currently a Professor of Computational Intelligence at Nottingham Trent University, where he is leading the research group in Computational Intelligence and Applications. His research focuses on the identification

of progressive changes in behaviour of elderly people suffering from Dementia or any other cognitive impairments. Accurate identification of progressive changes through utilisation of unobtrusive sensor network or robotics platform will enable carers (formal and informal) to intervene when deemed necessary. Research collaboration is established with world-leading researchers. He has worked in collaboration with many healthcare commercial organisations and end-users including Tunstall Healthcare Group and Nottingham Adult Care. He has received external funding from Innovate UK, EU and industrial companies to support his research. He has authored and co-authored over 180 scientific papers in the area of computational intelligence, internet of things, abnormal behaviour recognition and ambient intelligence in highly prestigious journals and international conferences. He has been invited as an Expert Evaluator and Panel Member for many European and International Research Programmes.



Junpei Zhong (S'10-M'15-SM'20) is an independent research fellow at Nottingham Trent University, UK. He graduated from South China University of Technology with a Bachelor in Automation. He obtained his MPhil degree at the Hong Kong Polytechnic University, Hong Kong, and Ph.D. degree at the University of Hamburg, Germany. Before joining NTU, he has been a researcher at the University of Hertfordshire (UK), University of Plymouth (UK), Waseda University (Japan) and the National Institute of Advanced

Industrial Science and Technology (AIST, Japan). He has participated in several EU and Japanese projects centering around machine learning, assistive technologies, and cognitive robots. His current research focuses on cognitive assistive robots and their usage in wellbeing and elderly healthcare. He is the organizer of the IEEE Symposium on Domestic Robotics 2019 (part of the Symposium Series of Computational Intelligence 2019) in Xiamen, China. He also organized special sessions and workshops in ICIRA 2019, IEEE-RAS Humanoids 2018, IEEE ICDL-EpiRob 2018, and guest editors of several journals such as IEEE Transaction on Cognitive and Developmental Systems, Interaction Studies and Complexity.