IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 10, 2021, accepted May 31, 2021, date of publication June 4, 2021, date of current version June 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3086431

Device-Free Localization Using Privacy-Preserving Infrared Signatures Acquired From Thermopiles and Machine Learning

NATHANIEL FAULKNER[1,2], FAKHRUL ALAM[1,2], (Senior Member, IEEE), MATHEW LEGG[1], (Member, IEEE), AND SERGE DEMIDENKO[1,2], (Fellow, IEEE)

[1]Department of Mechanical and Electrical Engineering, School of Food and Advanced Technology, Massey University, Auckland 0632, New Zealand
[2]School of Engineering and Technology, Sunway University, Bandar Sunway 47500, Malaysia

Corresponding author: Fakhrul Alam (f.alam@massey.ac.nz)

The work of Nathaniel Faulkner was supported in part by the Massey University Doctoral Scholarship.

ABSTRACT The development of an accurate passive localization system utilizing thermopile sensing and artificial intelligence is discussed in this paper. Several machine learning techniques are explored to create robust angular and radius coordinate models for a localization target with respect to thermopile sensors. These models are leveraged to develop a reconfigurable passive localization system that can use a varying number of thermopiles without the need for retraining. The proposed robust system achieves high localization accuracy (with the median error between 0.13 m and 0.2 m) while being trained using a single human subject and tested against multiple other subjects. It is shown that the proposed system does not experience any significant performance deterioration when localizing a subject at different ambient temperatures or with different configurations of the thermopile sensors placement.

INDEX TERMS Device-free localization (DFL), human sensing, indoor positioning system (IPS), infrared sensing, machine learning, passive localization, thermopile.

I. INTRODUCTION

Smart cities [1] and smart homes [2] are radically changing how we live by offering, among other things, location based services [3] and ambient assisted living [4] requiring reliable positioning systems. Two recent decades have seen intensive research activities associated with the development of Indoor Positioning System (IPS) solutions [5].

IPS can be of active and passive types. Active or device-based solutions use a network of static nodes (often termed as anchors) to localize a transceiver carried by a human target. Given the immense popularity of mobile phones, many solutions propose to locate individuals by tracking their phones. These techniques leverage the large number of onboard sensors (e.g. camera, Inertial Measurement Unit (IMU), light-sensors) and communication capabilities (cellular, Wi-Fi, Bluetooth) of the phones [6]. The passive or Device-Free Localization (DFL) systems [7] do not require the tracked entity to carry a transceiver. Passive positioning can be achieved by using regular camera vision techniques. However, there is a privacy issue that has to be considered here: people are normally quite reluctant to have such imaging devices, particularly in private areas of their residences. Camera-based techniques are also impacted by the illumination conditions.

Various alternative sensing techniques, based on the use of the Radio Frequency (RF) Received Signal Strength Indicator [8], Wi-Fi Channel State Information [9], visible light [10], [11], and electric field [12], [13] were proposed for DFL. Localization using pressure-sensitive [14] and capacitive [15] floors were also investigated. There were reports on techniques that applied single-pixel cameras [16], ultrasonic [17], and acoustic [18] sensing. Footstep induced vibrations captured by seismic sensors were also proposed for the localization [19]. Whilst considerable progress has been achieved in the DFL-associated research, the area is still of significant and on-going interest amongst researchers aiming to improve existing techniques and develop new solutions.

Human subjects can be detected from their Infrared (IR) emission. In most indoor surroundings, a person having a higher temperature than the environment can be distinguished from the background. Two popular devices used for

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Verticale.

IR localization are Passive IR (PIR) and Thermopile sensors. PIR sensors are commonly used as motion detectors in security systems. However, PIR-based techniques require a relatively large number of sensors. They may also need significant sensor modifications (e.g., [20]), making commercial off-the-shelf motion sensors unusable. Besides, they are inherently incapable of localizing a stationary target as PIR sensors require relative motion between them and the target. Rotating sensors [21] or shutters [22] could overcome that issue. Regrettably, such sensors are more complex, expensive and are characterized by increased power consumption.

A. LOCALIZATION USING THERMOPILE SENSORS

Low-resolution thermopile sensors (e.g., AMG8833 GridEYE[1]) are effectively thermal cameras that can detect both stationary and moving targets. At the same time, due to their lower image resolution, such sensors do not compromise the privacy of subjects. Figures 1(a) and 1(b) illustrate the difference between an image acquired from a standard camera and a typical output of a thermopile sensor, taken at its maximum 8×8 pixel resolution. As well as being privacy-preserving, thermopiles are invariant to changes in illumination.

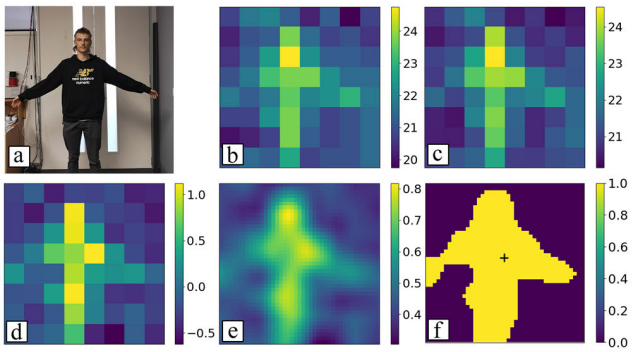

FIGURE 1. The sensor data pre-processing and parametric model process used to find the angle of the subject. Photo (a) shows a standard smartphone camera image of the subject taken simultaneously from the same position as that obtained by the thermopile sensor. Plot (b) shows the raw input from the sensor. The remaining plots illustrate (c) the Gaussian denoising application, (d) SVD background removal and data normalization, (e) interpolation, and (f) thresholding and averaging of pixel positions above the threshold to find the center of the largest blob.

Shetty *et al.* proposed tracking a subject using the foreground regions from the thermopile images [23]. Unfortunately, the authors did not report any accuracy values. Using a similar method, Kuki *et al.* were able to obtain an accuracy ranging between 0.15 m and 0.35 m (depending upon the activity) in a 2.56 m^2 area using a 4×4 pixel sensor [24]. The same authors then extended their work to achieve multi-person detection [25]. Qu *et al.* performed multi-target localization using a ceiling-mounted sensor [26]. They were able to distinguish between subjects even for crossover events. They also investigated the sensor lens distortion and performed distortion correction. Ng *et al.* were able to locate subjects with an accuracy of approximately 0.5 m in a 12.5 m^2 area using five sensors [27]. Kowalski *et al.* [28] were able to localize a subject to a 0.5 m grid square with a 73% probability in a 3.75 m^2 area. Whilst the accuracy of that setup was lower than results in some other reported works, the authors extended their sensors to have a 180-degree Field-of-View (FOV) by collocating three GridEye sensors directed 60 degrees from each other whilst also covering a much larger area. Tariq *et al.* used a low-resolution 16-pixel thermopile sensor with a variety of neural networks to achieve 0.096 m Root Mean Squared Error (RMSE) in a 9 m^2 space [29]. Narayana *et al.* [30] were able to locate a subject with a median 0.22 m accuracy within a 72 m^2 area using a higher resolution 32×24-pixel sensor. It should be noted that their proposed system required an additional calibrated PIR sensor for depth estimation.

Singh and Aksanli [31] compared the application of various Machine Learning (ML) classifiers for the detection and activity recognition of multiple human subjects using thermopile sensors. Similarly, Tateno *et al.* [32] and Tao *et al.* [33] used deep learning networks for fall detection and activity recognition respectively by utilizing ceiling-mounted sensors. Gochoo *et al.* [34] used a deep learning network to classify 26 separate yoga poses.

B. KEY CONCEPT AND CONTRIBUTION

The model-based localization techniques proposed in the literature (e.g., [23], [27]) appear to be unable to accurately capture the complex relationship between IR data and the relative position of the target with respect to the sensors. The reported ML techniques (e.g., [28], [29]) largely adopt the fingerprinting approach. They employ a single dataset, collected with one test subject, that is split for training, validation and testing. Therefore, it is difficult to ascertain whether these systems generalize well to different environments or subjects.

This paper proposes a new approach for target (i.e., a human subject) localization based on training models of thermopile sensors with the application of ML techniques. These models provide accurate estimation of an Angular Coordinate (AC) and Radius Coordinate (RC) of the target with respect to the center of the sensor, which are the direction and range to the target. Therefore, in the proposed approach, the subject can be localized by using just a single sensor (Fig. 2(a)). If the subject is in FOV of at least two sensors of known positions, two ACs can be estimated and used for positioning in a manner that is similar to the Angle of Arrival (AoA) [35] method (Fig. 2(b)). Similarly, if there are three or more sensors, one can use the lateration technique [35] to find the target position, using the distances between the subject and the sensors (Fig. 2(c)).

In the proposed approach, the sensor models for AC and RC need to be trained just once for one sensor. They can be then transferred to other sensors without measurable compromise in terms of localization accuracy. This leads to a robust and reconfigurable indoor positioning system that does

[1]https://na.industrial.panasonic.com/products/sensors/sensors-automotive-industrial-applications/lineup/grid-eye-infrared-array-sensor

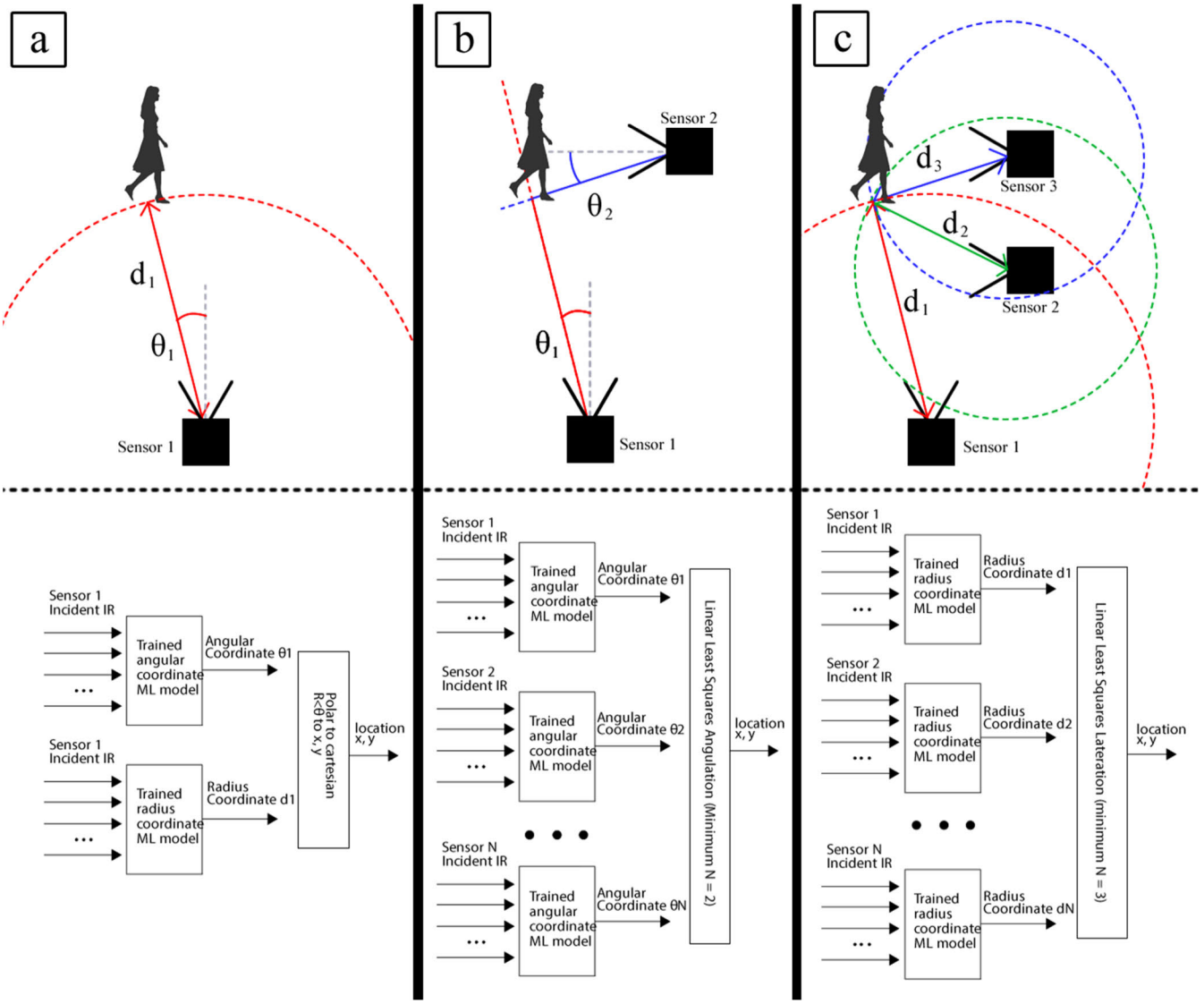

FIGURE 2. Three methods of positioning a subject where the sensor positions and orientations are known a priori. Diagram (a) shows the positioning using a combination of a single AC and RC/range. Diagram (b) shows the angulation using a minimum of two ACs. Diagram (c) shows the lateration using a minimum of three ranges.

not need to be retrained when deployed outside the training environment. The performance of the proposed system was tested with several different subjects. Each subject walked arbitrary paths for several minutes. The system was able to localize a subject, which it was not trained upon, with a median error of less than 0.2 m. The results show that the proposed approach is largely invariant to the subjects, system configuration, and deployment environments.

The remainder of the paper is structured as follows. Section II discusses data acquisition, ground truth estimation, and data preprocessing methods. Section III discusses the training, tuning, and evaluation of the various ML models used for the estimation of the angular and radius coordinates of a subject. Section IV demonstrates how these models can be used for positioning, and it also investigates the positioning performance of the system. Section V concludes the paper and discusses the limitations and future work to address these.

II. SYSTEM DEVELOPMENT

A. DATA ACQUISITION

The thermopile sensor used in this work is the Grid-EYE AMG8833. An interface to connect the AMG8833 to a computer was designed and constructed. It uses the STM32f103[2] microcontroller and a USB to serial adapter (Fig. 3). An arbitrary number of sensors may then be connected to a computer via USB cables whilst using a simple script for logging incoming sensor data, a corresponding device ID, and the timestamp of the data to a text file.

Several different datasets were recorded. Firstly, a 1.8 m tall male subject walked around a test area for approximately one hour. Two sensors were affixed to the walls at a height of 1.4 m (chosen to be above the height of most furniture) as

[2]https://www.st.com/content/st_com/en/products/microcontrollers-microprocessors/stm32-32-bit-arm-cortex-mcus/stm32-mainstream-mcus/stm32f1-series/stm32f103/stm32f103cb.html

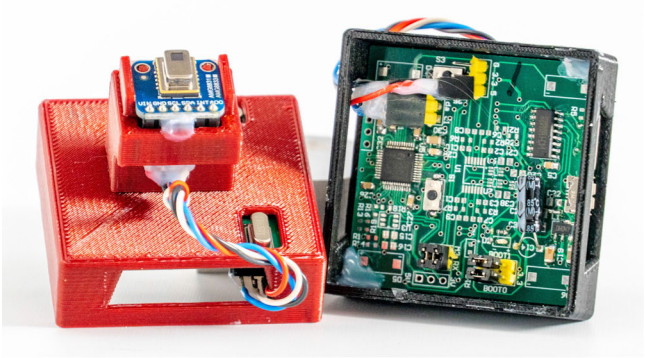

FIGURE 3. The thermopile sensor unit, housing both the thermopile module and the microcontroller used to communicate with a computer.

seen in Fig. 4(a). The ambient temperature was measured to be at 24 °C. This dataset is henceforth referred to as *Dataset 1*. The second dataset (*Dataset 2*) was taken at a later date in the same test area, using five different subjects, henceforth known as *Subjects 1-5*. The subjects were between 1.65 and 1.85 m tall (one female and four males). *Subject 1* was the subject used to collect *Dataset 1*. Each subject (including *Subject 1*) walked around the test area (having the same layout as the one used for the *Dataset 1* collection) for approximately 5-7 minutes each. The ambient temperature was measured to be at 22 °C. The third dataset (*Dataset 3*) was taken at another date, in a different room, with a three-sensor setup as seen in Fig. 4(b). It only featured *Subject 1* walking around the test area for approximately 10 minutes. The ambient temperature was at 26°C. The final dataset (*Dataset 4*) was also taken in this room, on another date, with *Subject 1* moving within the test area for approximately 7 minutes. The positions of the three sensors (see Fig. 4(c)) were different from those chosen for *Dataset 3*. The ambient temperature was again at 22 °C.

These multiple datasets were taken for specific purposes. *Dataset 1* was used to train various sensor models discussed later in this paper. *Datasets 2-4* were employed to investigate the generalizability of the proposed approach with respect to different environments or configurations and with different subjects from whence the system was trained.

B. GROUND TRUTH ESTIMATION

Accurate ground truth is very important when designing and evaluating a positioning system. In order to train and test a robust model, a large amount of labelled data is required. It is possible to mark out predefined paths and have a subject walking whilst following them at a set pace. However, such an arrangement is not ideal as it requires a high level of concentration from the subject. Besides, it can potentially force the subject into an unnatural gait. Ideally, a ground truth system should accurately track subjects as they naturally walk within the testing area. For this reason, the HTC Vive[3] was used as the ground truth tool. In previous works, it was found to be accurate to within several mm for extended time periods [15], [36]. The "tracking puck" needs to be kept within the line of sight of the "lighthouses". Therefore, the puck was attached to the subject's head, to approximate the subject's position in two dimensions.

[3] https://developer.vive.com/eu/vive-tracker-for-developer/

C. DATA PRE-PROCESSING

Data received from the thermopile sensors should be preprocessed to make them resistant to changes in ambient conditions. After that, the data can be used to train or test various ML models.

1) GAUSSIAN DENOISING

The temperature data produced by the thermopiles are noisy and as such, a single pixel fluctuates between frames randomly. This could cause the misdetection of subjects. To address it, each pixel is taken as a single time-series element and a one-dimensional Gaussian kernel is then applied along with the time-series data for each pixel (Fig. 1(c)). Each 8×8 frame is flattened into a single 64×1 vector:

$$\boldsymbol{f} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{64} \end{bmatrix}. \tag{1}$$

Such vectors form columns of a matrix:

$$\mathrm{A} = \begin{bmatrix} \boldsymbol{f_1} & \boldsymbol{f_2} & \cdots & \boldsymbol{f_N} \end{bmatrix}, \tag{2}$$

which is a time-series of length N samples with each row representing a single pixel (p_n) with respect to time. Each row of A is convolved with a 1-dimensional Gaussian kernel G. A kernel with a sigma value of 3 was empirically chosen with the following 5 samples defining the function:

$$G = \begin{bmatrix} 0.1784 & 0.2104 & 0.2223 & 0.2104 & 0.1784 \end{bmatrix} \tag{3}$$

2) BACKGROUND REMOVAL

The background temperature of a room is prone to change over time. However, such temperature changes are highly correlated between pixels in an empty scene. Moreover, the difference between two given background pixels at the same point in time appears to stay constant over time. This is illustrated in the first panel of Fig. 5. Each vertical slice represents a single frame from the thermopile. The first set of approximately 1800 samples in the figure corresponds to the room with no subject being present for approximately 3 minutes. The fluctuations in the background with time as the room temperature changes can be seen by the vertical strata in the data, whereas the constant offset between pixels can be seen in the horizontal stratification. Singular Value Decomposition (SVD) was employed for dimension reduction (such an approach has been used in computer vision for separating the background and foreground in videos [37]). SVD factorizes [38] the M × N matrix **A** (in this case M is 64 and N is the number of samples in the dataset):

$$\mathrm{A} = ULV^T. \tag{4}$$

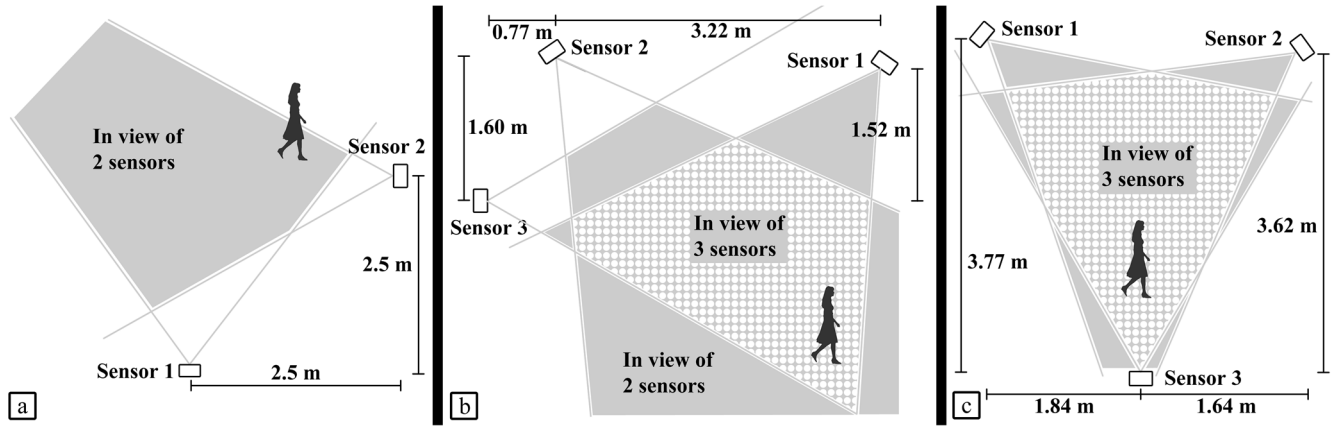

FIGURE 4. Room layouts for *Dataset 1* and *Dataset 2* (a), *Dataset 3* (b), and *Dataset 4* (c).

Here, the columns of U are the left singular vectors and the columns of V are the right singular vectors; $\mathbf{L}$ contains the singular values of $\mathbf{A}$ in a diagonal matrix, arranged in descending order; and T represents the transpose operation. The matrix $\mathbf{A}$ can be reassembled by multiplying the matrices together as given in Equation (4). However, it is possible to modify values of $\mathbf{L}$ before reassembling, to various effects. The singular values effectively represent how strongly each singular vector contributes to the matrix. The background data are highly correlated across the dataset, both in time and across the frame. Therefore, a good representation of the background can be found by reassembling the data using only the first singular value (the most dominant dimension) and zeroing out the others (see the bottom panel of Fig. 5). The foreground can therefore be found by doing the opposite – zeroing the first singular value and reassembling the matrix, as shown in the middle panel of Fig. 5.

3) DATA NORMALIZATION

After the background subtraction, there is still some variation in mean and standard deviation between different datasets. Furthermore, it is often advantageous to have input data for machine learning ranging between 0 and 1. Both Min-Max scaling and standardization were used in different circumstances. In a permanent, real-world deployment of the system, this would be done using a predetermined number of previous samples (e.g., several minutes' worth). When the system is first powered on, it would require an initial self-calibration period until a sufficient number of samples is captured.

III. SENSOR MODELS

The authors propose to create transferable sensor models that are trained once and then can be used for any sensor of the same make (e.g., the GridEye sensors). Locations of the sensors could be set arbitrarily. The same model could be used for all sensors regardless of their positions in a room without any retraining. In essence, the sensors are calibrated to produce the angular and/or radius coordinates of a target in two dimensions.

Multiple sensor models were trained and validated on *Dataset 1*, which was the largest one with 32,000 data points (frames). It was randomly split into 80-10-10 training-validation-test segments. The split was the same for each trained model (i.e., the same segments of data were used for training, validation, and test for each model). Each model (including the outlined below parametric model that did not require training data) was then tested against the data from five subjects taken at a different time (*Dataset 2*) as well as the test split from *Dataset 1*. This was done to investigate the generalizability of the models between subjects.

A. ANGULAR COORDINATE MODELS

These models take the temperature data from the sensor (after they were preprocessed as outlined in Section II D) as an input. They then output the angular coordinates of the subject with respect to the sensor. The ground truth angle and distance are computed using the subject's ground truth position and the sensor position (see θ_1 and d_1 respectively in Fig. 2(a)). A parametric method was used, as well as several ML methods (such as Multi-Layer Perceptron (MLP), random forest, Weighted K-Nearest Neighbor (WKNN)) to develop the AC models. The advantage of the parametric model is that it does not require any prior training. At the same time, the studied ML methods require a one-off training of the model that then ideally could be generalized to any subject or room.

1) PARAMETRIC MODEL

The devised parametric model utilizes an approach similar to the thermopile positioning [23] and capacitive floor footprints detection reported in [15]. After the temperature data are pre-processed (as described above in Section II D), it is reshaped to an 8×8 frame. Then it undergoes bicubic interpolation to a 55×55 matrix (see Fig. 1(e)) with each element of the matrix being referred to as a single pixel. Binary thresholding is then applied to select only the foreground objects. After that, the connected component analysis is employed for blob detection. If the number of blobs is more

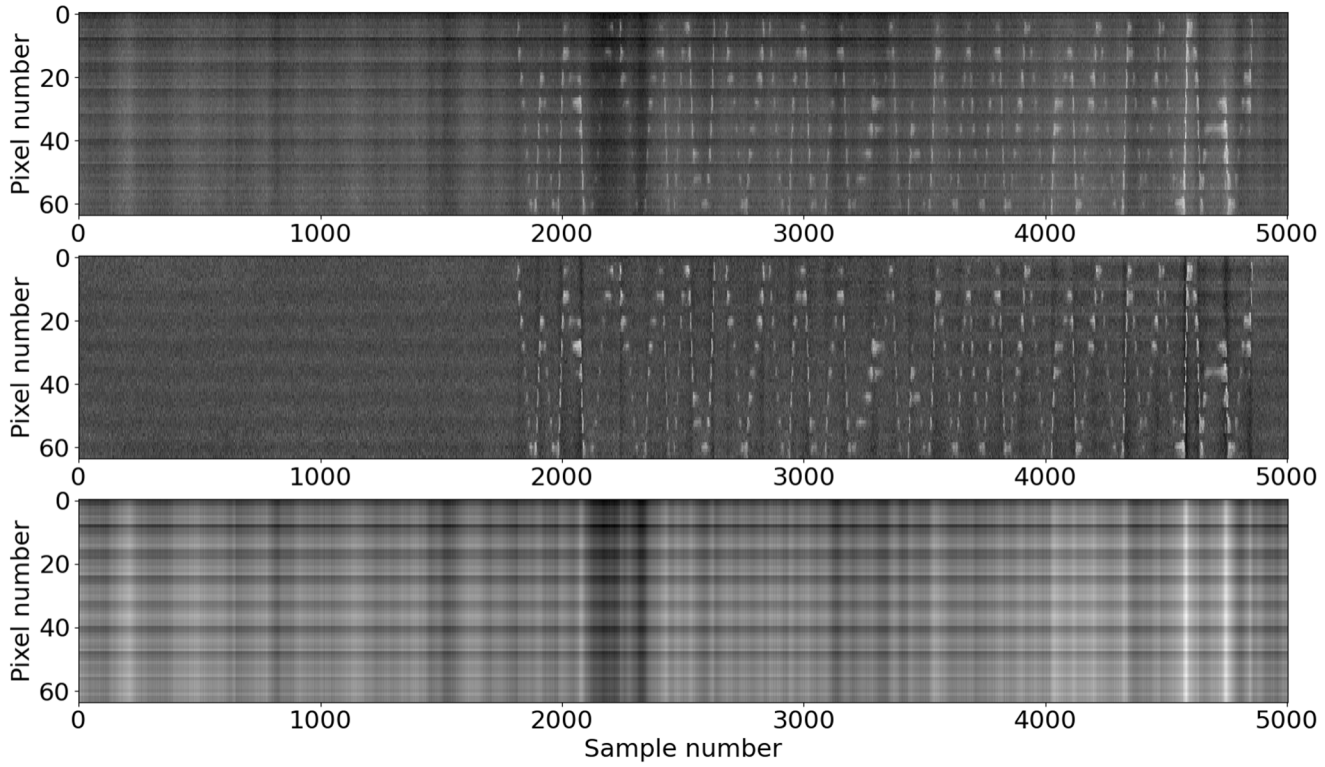

FIGURE 5. Plots demonstrating how the thermopile temperature matrix A (see Equation 2) is processed to remove background thermal effects. Each column is a single flattened frame, f from the sensor, with the horizontal axis representing time. The first approximately 1800 samples show an empty room, with the rest showing a subject moving around. The top panel is before background removal, the middle panel is after background removal and the bottom panel the removed background.

than one, the largest of them is assumed to be the subject. The center of the blob is found by taking the mean x and y positions of all the pixels in the blob. Fig. 1(f) illustrates this process. The position of the blob is then converted into an angular form:

$$\theta = \left(\frac{x}{FW} - 0.5\right) \times FOV. \tag{5}$$

Here θ is the AC of the subject with respect to the sensor, x is the position of the center of the blob pertaining to the frame (in pixels), FW is the frame width in pixels (55 pixels after the interpolation), and FOV is the width of the horizontal field of view of the sensor (60 degrees for GridEye). A zero value of θ indicates that a subject is on a line perpendicular to the sensor (i.e., in the middle of the field of view). Negative values of θ correspond to the left-hand side while positive - to the right-hand side plane.

2) MULTI-LAYER PERCEPTRON MODEL

An MLP model was trained with 64×1 input vectors of pre-processed temperature data from the sensor while the outputs were ACs of the subject with respect to the sensor. The input layer of the MLP had 64 perceptions fully connected to the first hidden layer. The output layer was a single perceptron, fully connected to the final hidden layer. A grid search within a wide range of hyperparameters was used to tune the MLP. Each candidate model, defined by a unique combination of hyperparameters, was trained for 1000 epochs and tested against the validation data after each epoch. Early stopping was used to avoid overtraining by observing the validation RMSE curve over these 1000 epochs. The final models for each candidate model were then sorted by validation RMSE.

TABLE 1. Final hyperparameter values selected for each AC model after hyperparameter tuning.

Model	Hyperparameter	Value
MLP	Number of hidden layers	2
	Hidden layer size	500
	Hidden layer activation	ReLU
	Output layer activation	Sigmoid
Random Forest	Number of estimators	500
	Minimum sample per leaf node	1
	Minimum samples to split leaf	2
WKNN	K-value	2
	Distance metric	Euclidean

It was found that 2 hidden layers of 500 perceptrons with ReLU activation [39] on the hidden layers and sigmoid activation on the output layer gave the best performance (see Table 1). Several other parameter combinations gave similar performance results whilst enabling trading of the performance (RMSE) for a simpler model if needed. Larger models take significantly longer time and more computing resources to both train and run. Whilst training is a one-off event and can be generalized to multiple environments or

subjects, running the network has potential processing considerations. This is because the models could be run on the sensors themselves as opposed to on a PC. Such sensor-based solutions are resource-constrained. Thus, a simpler model is preferred as it would run faster and with lower power consumption.

3) RANDOM FOREST MODEL

A random forest regressor was trained in a manner that was similar to the MLP case by searching through a range of parameters to find optimal hyperparameters. It was observed that limiting the maximum tree depth or maximum leaf nodes had very little impact on the accuracy of the models. The best model had 500 estimators, a minimum of 2 samples to split a leaf node and a minimum of 1 sample per leaf node (see Table 1). However, the RMSE difference between 100 and 500 estimators was less than 1%, thus suggesting an opportunity for utilizing a simpler regressor.

4) WEIGHTED K-NEAREST-NEIGHBOR MODEL

WKNN regressor was optimized for the number of neighbors (K) and the distance metrics. It was found that a K-value of 2 with either Euclidean or Canberra distance metric provided the best performance.

B. PERFORMANCE OF ANGULAR COORDINATE MODELS

The performance of each model for estimating the AC for *Dataset 2* can be seen in Fig. 6. The parametric model was significantly outperformed by the three ML models. Also, there were no significant differences between these three machine learning models. At the same time, MLP might be somewhat preferable over WKNN in cases where a large database is needed. The model was processed on a PC with the raw data arriving from the device. However, it might be preferred in the future to move the model onto the microcontroller where storage concerns could preclude the use of a large WKNN database. The results showed that while the accuracy of the AC estimation was decreased compared to the test set split from *Dataset 1*, it was reasonably subject-invariant. *Subject 1* was used for training the model. The obtained results were on par with those for the other four subjects (whilst worse than the training set). It suggested that the main difference was in the environments (the training set was taken on a different day, with the temperature of the room being approximately 2°C lower).

C. RADIUS COORDINATE OR RANGE MODELS

The range models estimate the distance between the subject and the sensor. In a similar manner as for the AC models, a flattened 8×8 frame from the sensor was inputted as a 64×1 vector. The output of the model was the range. The range can be taken as the radius coordinate and then it can be combined with the angular coordinate to perform single-sensor based positioning (Fig. 2(a)). Also, the distances from multiple sensors can potentially be used for the lateration (Fig. 2(c)).

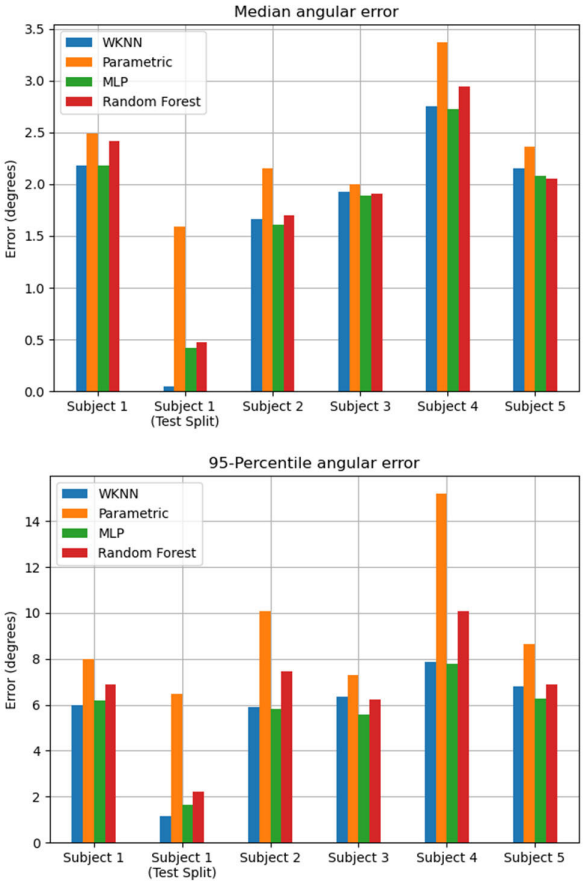

FIGURE 6. Median and 95-percentile angular coordinate error for the four angular coordinate models. Each model was tested on all five subjects (*Dataset 2*), plus the 10% test split from the training dataset (*Dataset 1*).

TABLE 2. Final hyperparameter values selected for each range model after hyperparameter tuning.

Model	Hyperparameter	Value
MLP	Number of hidden layers	3
	Hidden layer size	500
	Hidden layer activation	ReLU
	Output layer activation	Sigmoid
Random Forest	Number of estimators	100
	Minimum sample per leaf node	1
	Minimum samples to split leaf	2
WKNN	K-value	2
	Distance metric	Canberra

Three regressors (MLP, random forest and WKNN) were trained. The hyperparameters are listed in Table 2.

D. PERFORMANCE OF RANGE MODELS

The three different range estimation models were compared for *Dataset 2*, i.e., for five subjects (Fig. 7). The MLP outperformed the other two models quite significantly. It was more robust to variations of the environment and different subjects than the other methods. Interestingly, the WKNN and random forest methods struggled most with *Subject 1* upon whose data the models were trained albeit with data collected on a different day.

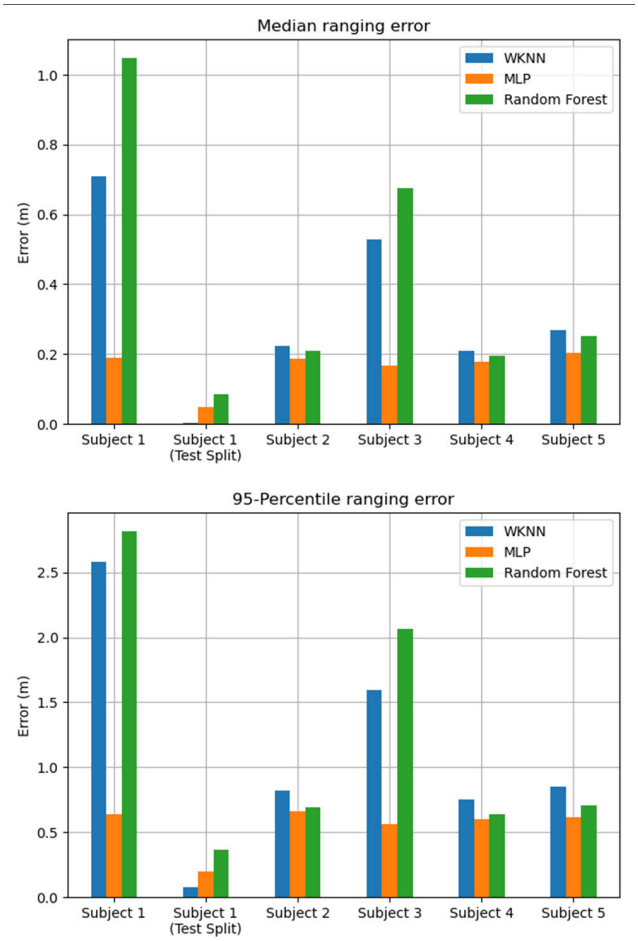

FIGURE 7. Median and 95-percentile radius coordinate error for the three range models. Each model was tested on all five subjects (*Dataset 2*), plus the 10% test split from the training dataset (*Dataset 1*).

IV. POSITIONING METHODS

Several methods of positioning a subject can now be developed using the sensor models proposed in Section III. The use of the AC and range with a single sensor or just the AC-based model on multiple sensors allows for the ease of configuring the system. Sensors can be set up at suitable locations in a room. Additional sensors can be incorporated for extended coverage or higher accuracy without the need for retraining. This provides a significant advantage over the multiple sensor fingerprint-based method that relies on the sensor number and the geometry remaining consistent. For such fingerprint-based methods, retraining would be required if sensors are to be spaced at different distances, or additional sensors need to be added.

A. MODEL-BASED POSITIONING

1) SINGLE-SENSOR BASED POSITIONING USING THE AC AND RANGE

Positioning can be performed using a single sensor using AC and range data. Two MLP models are employed for subject positioning, with one of them outputting the AC whilst the other, the RC - the distance between the subject and sensor. Simple geometry is then used to calculate the position of the subject relative to the sensor, see Fig 2(a). It should be noted that once the range is estimated from three or more sensors, it is possible to perform the lateration-based localization. However, the reported research did not pursue that approach as it would require extra sensors while preliminary results did not show noticeable performance benefits.

2) MULTIPLE SENSOR BASED POSITIONING USING AC

This approach uses the positions of multiple sensors and the angular coordinates of the subject with regard to the sensors. ACs are estimated using the MLP model (outlined in Section III A). The position of the subject is found similarly to a standard AoA technique. Fig. 2(b) shows an example where two sensors are used. If more than two sensors are employed, the system is over defined and linear least squared estimation can be used [35]. The problem can be formulated as

$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{q} = \boldsymbol{b} \tag{6}$$

where:

$$\boldsymbol{A} = \begin{bmatrix} \sin(\theta_1) & -\cos(\theta_1) \\ \sin(\theta_2) & -\cos(\theta_2) \\ \vdots & \vdots \\ \sin(\theta_n) & -\cos(\theta_n) \end{bmatrix}, \tag{7}$$

$$\boldsymbol{b} = \begin{bmatrix} \sin(\theta_1)x_1 - \cos(\theta_1)y_1 \\ \sin(\theta_2)x_2 - \cos(\theta_2)y_2 \\ \vdots \\ \sin(\theta_n)x_n - \cos(\theta_n)y_n \end{bmatrix}, \tag{8}$$

and $\boldsymbol{q}$ is a measure of the noise. The estimate of $\boldsymbol{x}$, the 2×1 position vector is:

$$\hat{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} \tag{9}$$

B. FINGERPRINT-BASED POSITIONING

The proposed model-based positioning techniques were benchmarked against fingerprint-based techniques. Fingerprinting is commonly used for ML-based positioning as reported in the literature. A single-sensor based fingerprint positioning used an MLP with its input being a flattened array of the pixels from a single sensor. The output presents the x and y coordinates of the subject relative to the sensor. Training, validation, and testing were done following the process described in Section III A. A dual-sensor fingerprint-based positioning was also implemented where an MLP was trained to take a flattened array of 128 pixels (64 from each sensor) to estimate the position of the subject.

C. POSITIONING RESULTS EVALUATION

The four different positioning methods were evaluated using each of the five subjects (corresponding to *Dataset 2*). It can be seen from Fig. 8 that two-sensor-based positioning provided higher accuracy than the single-sensor ones. The proposed dual-sensor AC-based positioning was the most accurate. However, the field of view of the dual-sensor

models (being the intersection of the individual FOVs) was smaller than the total area (Fig. 4(a)). The single sensor models offered a larger coverage, at the cost of lower accuracy.

One solution could be to use the dual-sensor configuration where there is coverage and utilize the single-sensor-based solution only where there is coverage by a single sensor. Another solution could be to use a higher density of sensors or employ sensors with a wider FOV (e.g., [28]).

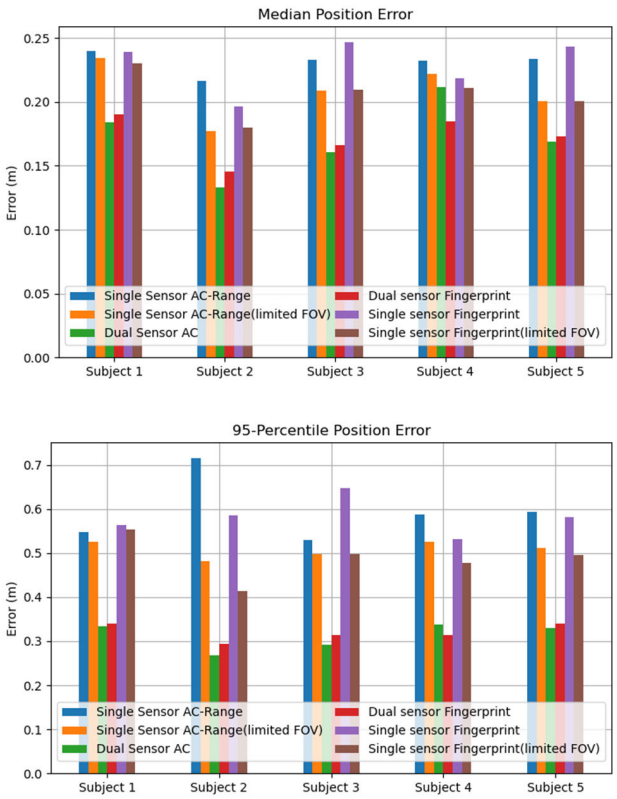

FIGURE 8. Median and 95-percentile position error for the four different positioning methods, plus the two single sensor methods being limited to the same FOV as the dual sensor models (the intersection of the two FOVs).

In order to make a fair comparison, the single-sensor based positioning algorithms were also run a second time whilst using only the data that corresponded to the combined FOVs of both the sensors in the dual-sensor configuration. While there was a modest improvement for the single-sensor systems, the dual sensor configurations were still more accurate.

It was observed that there was not much variation in position errors between the subjects. As each subject walked about randomly, they entered and exited the FOV of the sensors. Time intervals between a subject entering and exiting sensor FOVs were saved as for individual paths. For each subject, that equated to between 25 to 35 paths of varying lengths and trajectories. Some examples of paths can be seen in Fig 9.

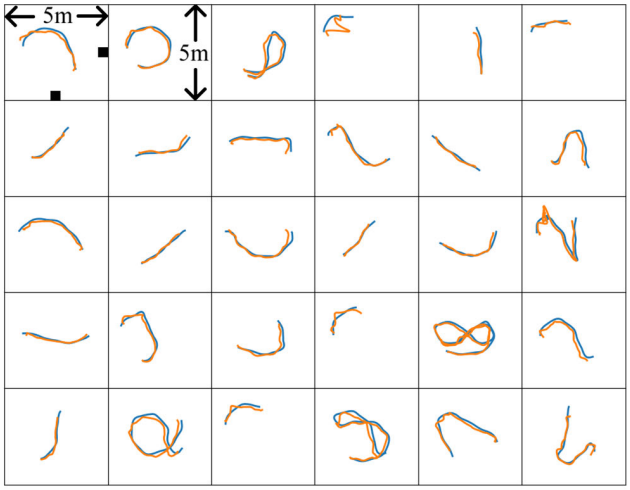

FIGURE 9. The paths walked by *Subject 1* in *Dataset 2*. The HTC Vive ground truth is shown in blue, along with the estimated positions using the dual camera AC-based method in orange. The black squares in the top left panel represent the sensor positions (shown only for the first panel).

D. PERFORMANCE IN DIFFERENT CONFIGURATIONS

A three-sensor system layout (Fig. 4(b)) was implemented to carry out the experiment (corresponding to the case of *Dataset 3* outlined in Section II B). The AC-based position evaluations were performed using all combinations of sensor pairs (1-2), (2-3), (3-1) as well as for the three-sensors case. The AC MLP model used was trained on *Dataset 1* that was acquired at an earlier date and in a different room. Fig. 10 and Table 3 show the localization results. When localization was carried out with the pairs of sensors, the obtained positioning accuracy was similar for all of them (the median varied between 0.13 m and 0.14 m; the 95-percentile varied between 0.27 m and 0.34 m for each of the three pairs). This was on par with what was observed for *Dataset 2* even though the relative sensors' positionings were markedly different. It clearly demonstrates the robust nature of the AC-based localization system. As seen in Fig. 11, the positioning accuracy could be improved by using the ACs from all three sensors.

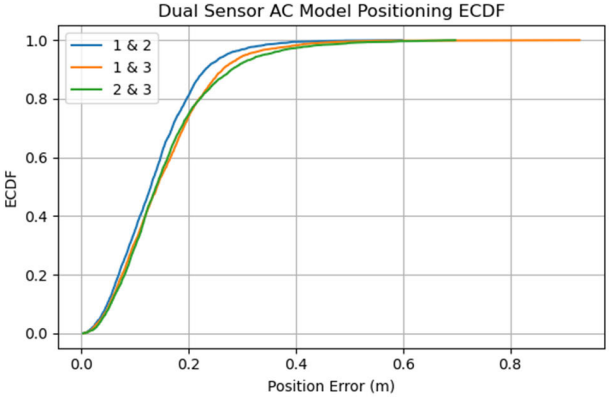

FIGURE 10. Dual sensor AC model positioning accuracy for pairs of sensors for *Dataset 3*.

TABLE 3. Median and 95-percentile errors for the various positioning models used on Dataset 3.

Model	Sensor id	Median error (m)	95-Percentile error (m)
Three-sensor AC model	1, 2 & 3	0.11	0.23
Dual-sensor AC model	1 & 2	0.13	0.27
	2 & 3	0.14	0.34
	3 & 1	0.14	0.31
Dual-sensor fingerprint	1 & 2	0.22	0.41
	2 & 3	0.35	0.98
	3 & 1	0.95	2.24
Single-sensor fingerprint	1	0.19	0.45
	2	0.17	0.39
	3	0.16	0.52
Single-sensor AC/range	1	0.17	0.40
	2	0.18	0.51
	3	0.22	0.56

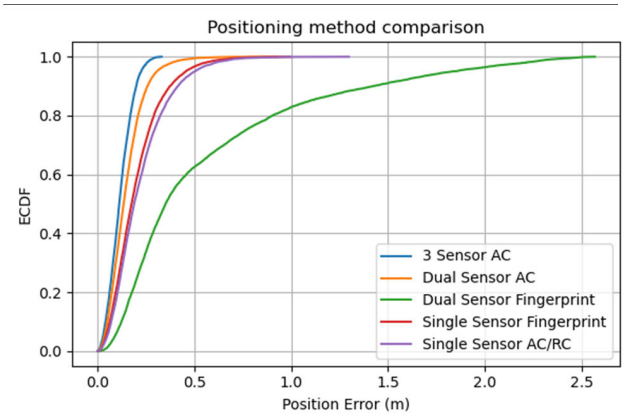

FIGURE 11. Positioning accuracy for the different positioning models on *Dataset 3*. For the dual sensor models, the errors were combined from each of the three pairwise combinations. For the single sensor models the errors were combined from the three individual sensors.

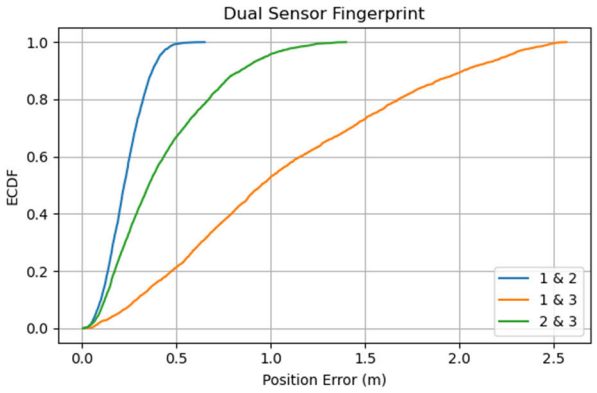

FIGURE 12. Dual sensor Fingerprint-based positioning accuracy for pairs of sensors for *Dataset 3*.

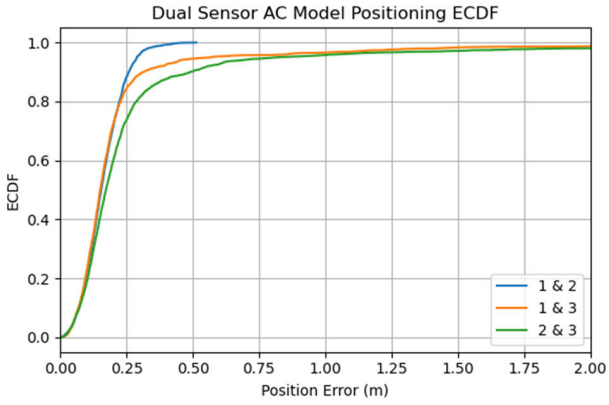

FIGURE 13. Dual sensor AC model positioning accuracy for pairs of sensors for *Dataset 4*. Note the graph has been truncated at 2 m along the x axis for clarity due to the very large worse case errors for 1-3 and 2-3 (8.5 m and 97.5 m respectively).

Such flexibility is not readily available with the other positioning techniques.

The performance of the single-sensor based positioning methods (AC plus range and fingerprinting) were not impacted by changes in the configuration. The accuracy was consistently close between *Dataset 2* and *Dataset 3* while less accurate than when employing the multiple-sensor based methods with AC estimation. In contrast, the dual sensor fingerprinting method experienced significant accuracy degradation. Closer inspection revealed that the accuracy of the (1-2) pair was nearer to that achieved with *Dataset 2* (see Fig. 12). The two sensors of the (1-2) pair were placed in a relatively similar configuration to those employed for acquiring *Dataset 2*. However, the accuracy levels achieved with (2-3) and, especially, (3-1) sensor pairs were considerably poorer due to the sensor positions being significantly different from the *Dataset 2* acquisition case.

While the AC-based method is shown to be robust and flexible, the selection of sensor locations still needs to be done judiciously. This is evident from the AC-based positioning performance shown in Fig. 13 for *Dataset 4*. The sensor layout is given in Fig 4(c) whilst the dataset details are outlined in Section II B. The accuracy for the sensor pair (1-2) appears to be similar to what was observed for *Dataset 2* and *Dataset 3*. However, the accuracy achieved by the sensor pairs (1-3) and (2-3) shows significant degradation. In all previous cases, the sensors were out of FOV of each other, whereas in the configuration under discussion, *Sensor 3* was within the fields of view of both *Sensor 1* and *Sensor 2*. Thus, if a subject stands directly between *Sensor 3* and one of the other sensors, the two angular coordinates would produce near-parallel lines. Therefore, even a small error in the AC estimations can cause a very large error in the position estimation. This is only a problem when the subject is in the FOV of two sensors, the sensors are within the FOV of each other, and the subject is directly between them. This can be mitigated through continuous tracking (as long as the target does not move along the direct path from one sensor to the other). However, the sensors could be positioned to ensure that such a scenario is unlikely to happen any often. Also,

since the single-sensor based positioning is not impacted, the system can switch to a single sensor operation mode in such a scenario.

V. CONCLUSION AND FUTURE WORK

ML model-based systems showed great promise in performing accurate localization using single and multiple thermopile sensors. Multiple-sensor based positioning was shown to be more accurate than the single-sensor based one. However, the single-sensor based positioning offered an important advantage of larger area coverage.

The ML regressors were trained with one human subject and tested with other subjects as well as in different environments as opposed to only training and testing on the same dataset. The model-based techniques generalized well.

The fingerprinting-based positioning also appeared to be able to cope with a change of subjects and environments when only a single sensor was used. However, the fingerprinting-based positioning with multiple sensors was essentially limited by the configuration (e.g., the number of sensors and their relative positions) that was used for training.

The most apparent limitation of the proposed system is that it is only capable of localizing a single subject. This could be addressed in the future by applying a stacked approach. The authors were able to train an accurate classifier (over 95%) to detect the presence of a subject in a frame. This can be further extended to count the number of subjects. Several different models can then be trained for a varying number of subjects. With multiple sensors, it should be possible to track the subjects without losing their identity. In the case of crossover events where one subject is occluded by another, it is unlikely that subjects are occluded from both sensors. It may also be possible to assign a short-term identity to a person based on their thermal profile [26] whilst they are within the FOV. The impact of changing the heights of the sensors have also not been investigated and can be a topic of further studies.

Thermopile sensors rely on the subject emitting IR, which is influenced by the clothes the subject is wearing. For example, while wearing a very heavy coat, a subject may not be visible to the sensor except at very close proximity. The experiments were undertaken in standard office wear that is appropriate for the ambient temperature. An exploratory investigation was undertaken where *Subject 1* wore a thick winter jacket. Positioning nearer the sensors appeared to be relatively unaffected. However, the performance was degraded at further distances as the jacket reduced the effective range of the sensor. It should be noted that the subject was much warmer than comfortable and would not have worn such apparel in a climate-controlled room. This would be an interesting area for a future investigation.

The performance of additional ML techniques (e.g., recurrent neural networks) and the impact of hyperparameters on ML techniques have been also identified for a future study.

ACKNOWLEDGMENT

Nathaniel Faulkner and Fakhrul Alam would like to thank Sunway University for granting them the Visiting Research Fellow (2020) and Adjunct Professor (2021–2022) appointments respectively.

REFERENCES

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[2] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," *J. Cleaner Prod.*, vol. 140, pp. 1454–1464, Jan. 2017.

[3] I. A. Junglas and R. T. Watson, "Location-based services," *Commun. ACM*, vol. 51, no. 3, pp. 65–69, 2008.

[4] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 579–590, May 2013.

[5] F. Khelifi, A. Bradai, A. Benslimane, P. Rawat, and M. Atri, "A survey of localization systems in Internet of Things," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 761–785, Jun. 2019.

[6] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2017.

[7] F. Alam, N. Faulkner, and B. Parr, "Device-free localization: A review of non-RF techniques for unobtrusive indoor positioning," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4228–4249, Mar. 2021.

[8] D. Konings, F. Alam, F. Noble, and E. M.-K. Lai, "Device-free localization systems utilizing wireless RSSI: A comparative practical investigation," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2747–2757, Apr. 2019.

[9] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5217–5230, Jun. 2018.

[10] N. Faulkner, F. Alam, M. Legg, and S. Demidenko, "Watchers on the wall: Passive visible light-based positioning and tracking with embedded light-sensors on the wall," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2522–2532, May 2020.

[11] D. Konings, N. Faulkner, F. Alam, E. M.-K. Lai, and S. Demidenko, "FieldLight: Device-free indoor human localization using passive visible light positioning and artificial potential fields," *IEEE Sensors J.*, vol. 20, no. 2, pp. 1054–1066, Jan. 2020.

[12] A. R. Akhmareh, M. Lazarescu, O. B. Tariq, and L. Lavagno, "A tagless indoor localization system based on capacitive sensing technology," *Sensors*, vol. 16, no. 9, p. 1448, Sep. 2016.

[13] X. Tang and S. Mandal, "Indoor occupancy awareness and localization using passive electric field sensing," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 11, pp. 4535–4549, Nov. 2019.

[14] M. Andries, O. Simonin, and F. Charpillet, "Localization of humans, objects, and robots interacting on load-sensing floors," *IEEE Sensors J.*, vol. 16, no. 4, pp. 1026–1037, Feb. 2016.

[15] N. Faulkner, B. Parr, F. Alam, M. Legg, and S. Demidenko, "CapLoc: Capacitive sensing floor for device-free localization and fall detection," *IEEE Access*, vol. 8, pp. 187353–187364, 2020.

[16] J. Zhao, N. Frumkin, P. Ishwar, and J. Konrad, "CNN-based indoor occupant localization via active scene illumination," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2636–2640.

[17] E. A. Wan and A. S. Paul, "A tag-free solution to unobtrusive indoor tracking using wall-mounted ultrasonic transducers," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Sep. 2010, pp. 1–10.

[18] Y. Guo and M. Hazas, "Localising speech, footsteps and other sounds using resource-constrained devices," in *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2011, pp. 330–341.

[19] M. Mirshekari, S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh, "Occupant localization using footstep-induced structural vibration," *Mech. Syst. Signal Process.*, vol. 112, pp. 77–97, Nov. 2018.

[20] S. Narayana, R. V. Prasad, V. S. Rao, T. V. Prabhakar, S. S. Kowshik, and M. S. Iyer, "PIR sensors: Characterization and novel localization technique," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2015, pp. 142–153.

[21] Y. Li, D. Li, Y. Cheng, G. Liu, J. Niu, and L. Su, "A novel human tracking and localization system based on pyroelectric infrared sensors: Demonstration abstract," in *Proc. 15th Int. Conf. Inf. Process. Sensor Netw.*, 2016, pp. 1–2.
[22] L. Wu, Y. Wang, and H. Liu, "Occupancy detection and localization by monitoring nonlinear energy flow of a shuttered passive infrared sensor," *IEEE Sensors J.*, vol. 18, no. 21, pp. 8656–8666, Nov. 2018.
[23] A. D. Shetty, Disha, B. Shubha, and K. Suryanarayana, "Detection and tracking of a human using the infrared thermopile array sensor—Grid-eye," in *Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, Jul. 2017, pp. 1490–1495.
[24] M. Kuki, H. Nakajima, N. Tsuchiya, and Y. Hata, "Human movement trajectory recording for home alone by thermopile array sensor," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 2042–2047.
[25] M. Kuki, H. Nakajima, N. Tsuchiya, K. Kuramoto, S. Kobashi, and Y. Hata, "Mining multi human locations using thermopile array sensors," in *Proc. IEEE 43rd Int. Symp. Multiple-Valued Log.*, May 2013, pp. 59–64.
[26] D. Qu, B. Yang, and N. Gu, "Indoor multiple human targets localization and tracking using thermopile sensor," *Infr. Phys. Technol.*, vol. 97, pp. 349–359, Mar. 2019.
[27] H. M. Ng, "Human localization and activity detection using thermopile sensors," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2013, pp. 337-338.
[28] C. Kowalski, K. Blohm, S. Weiss, M. Pfingsthorn, P. Gliesche, and A. Hein, "Multi low-resolution infrared sensor setup for privacy-preserving unobtrusive indoor localization," in *Proc. 5th Int. Conf. Inf. Commun. Technol. Ageing Well e-Health (ICTAWE)*, Heraklion, Greece, vol. 1, 2019, pp. 183–188.
[29] O. B. Tariq, M. T. Lazarescu, and L. Lavagno, "Neural networks for indoor person tracking with infrared sensors," *IEEE Sensors Lett.*, vol. 5, no. 1, pp. 1–4, Jan. 2021.
[30] S. Narayana, V. Rao, R. V. Prasad, A. K. Kanthila, K. Managundi, L. Mottola, and T. V. Prabhakar, "LOCI: Privacy-aware, device-free, low-power localization of multiple persons using IR sensors," in *Proc. 19th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2020, pp. 121–132.
[31] S. Singh and B. Aksanli, "Non-intrusive presence detection and position tracking for multiple people using low-resolution thermal sensors," *J. Sensor Actuator Netw.*, vol. 8, no. 3, p. 40, Jul. 2019.
[32] S. Tateno, F. Meng, R. Qian, and Y. Hachiya, "Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor," *Sensors*, vol. 20, no. 20, p. 5957, Oct. 2020.
[33] L. Tao, T. Volonakis, B. Tan, Z. Zhang, and Y. Jing, "3D convolutional neural network for home monitoring using low resolution thermal-sensor array," in *Proc. 3rd IET Int. Conf. Technol. Act. Assist. Living (TechAAL)*, London, U.K., 2019, pp. 1–6.
[34] M. Gochoo, T.-H. Tan, S.-C. Huang, T. Batjargal, J.-W. Hsieh, F. S. Alnajjar, and Y.-F. Chen, "Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7192–7200, Aug. 2019.
[35] R. Zekavat and R. M. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*. Hoboken, NJ, USA: Wiley, 2011.
[36] A. H. A. Bakar, T. Glass, H. Y. Tee, F. Alam, and M. Legg, "Accurate visible light positioning using multiple-photodiode receiver and machine learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
[37] D. Chetverikov, S. Fazekas, and M. Haindl, "Dynamic texture as foreground and background," *Mach. Vis. Appl.*, vol. 22, no. 5, pp. 741–750, Sep. 2011.
[38] S. J. Prince, *Computer Vision: Models, Learning, and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[39] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375

NATHANIEL FAULKNER received the B.E. degree (Hons.) in electronics and computer engineering from Massey University, New Zealand, in 2016, where he is currently pursuing the Ph.D. degree. Since 2020, he has been appointed as a Visiting Research Fellow with the School of Engineering and Technology, Sunway University, Malaysia. His research interests include indoor positioning, embedded systems design, and the Internet of Things.

FAKHRUL ALAM (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical and electronic engineering from BUET, Bangladesh, and the M.S. and Ph.D. degrees in electrical engineering from Virginia Tech, USA. He is currently an Associate Professor with the Department of Mechanical and Electrical Engineering, School of Food and Advanced Technology, Massey University, New Zealand. He has been appointed as an Adjunct Professor with the School of Engineering and Technology, Sunway University, Malaysia, since 2021. His research interests include indoor localization, 5G and visible light communication, the IoT, and wireless sensor networks.

MATHEW LEGG (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in physics from The University of Auckland, New Zealand. He is currently a Senior Lecturer with the Department of Mechanical and Electrical Engineering, School of Food and Advanced Technology, Massey University. His research interests include the development of acoustic/ultrasonic measurement systems and techniques for acoustic imaging, non-destructive testing, and remote sensing.

SERGE DEMIDENKO (Fellow, IEEE) graduated in computer engineering from the Belarusian State University of Informatics and Radioelectronics and received the Ph.D. degree from the Institute of Engineering Cybernetics, Belarusian Academy of Sciences. He is currently a Professor and the Dean of the School of Engineering and Technology, Sunway University, Malaysia. He is also associated with the Department of Mechanical and Electrical Engineering, School Food and Advanced Technology, Massey University, New Zealand. His research interests include electronic design and test, signal processing, instrumentation, and measurements. He is also fellow of IET and U.K. Chartered Engineer.