Indoor human detection based on Thermal Array Sensor data and adaptive background estimation

Anna A. Trofimova[1], Andrea Masciadri[1], Fabio Veronese[1], Fabio Salice[1]

[1]Department of Electronics, Information and Bioengineering Politecnico di Milano, via Anzani 42, 22100, Como, ITALY
Email: anna.trofimova91@gmail.com, andrea.masciadri@polimi.it, fabio.veronese@polimi.it, fabio.salice@polimi.it

Abstract

Low Resolution Thermal Array Sensors are widely used in several applications in indoor environments. In particular, one of these cheap, small and unobtrusive sensor provides a low-resolution thermal image of the environment and, unlike cameras, it is capable to detect human heat emission even in dark rooms. The obtained thermal data can be used to monitor older seniors while they are performing daily activities at home, to detect critical situations such as falls. Most of the studies in Activity Recognition using Thermal Array Sensors requires Human Detection techniques to recognize humans passing in the sensor field of view. This paper aims to improve the accuracy of the algorithms used so far by considering the temperature environment variation. This method leverages an adaptive background estimation and a noise removal technique based on Kalman Filter. In order to properly validate the system, a novel installation of a single sensor has been implemented in a Smart Environment: the obtained results show an improvement in human detection accuracy with respect to the state of the art, especially in case of disturbed environments.

Keywords

Thermal Array Sensor, Indoor Human Detection, Smart Environment

1. Introduction

Facilitate older seniors independent living has become an important issue of the current research in the field of Assistive Technologies fostered by worldwide governments. Indeed, as reported in the technical report of the United Nations [1], the over-60 world population in the next 15 years is projected to grow by 56 percent, from 901 millions to more than 1.4 billions. The ageing process is particularly evident in Europe and in Northern America, where in 2015, more than one out of five people was over 60, and it is growing rapidly in other regions as well. Thus, Bierhoff et Al. [2], for example, report Health-Care Systems among the products and services based on Smart Home technologies. In particular, since their aim is to detect critical conditions or predict them on early stages and alert caregivers, Emergency Treatment services are crucial for older adults living in smart homes. This kind of systems usually rely on a network of sensors to

unobtrusively monitor the life of a person, providing feedback to his/her beloved [3]. A concrete example of a widely demanded feature by the families of the elderly people who live alone is a Fall Detector, indeed falls are nowadays considered as the most frequent hazard for older seniors and they may endanger the physical and psychological health of a person, hindering independent living [4]. For this reasons, monitoring elderly people at home makes them feel safer and help their relatives to be more confident, knowing the well-being of their beloved.

Low Resolution Thermal Array Sensors (LR-TASs) are very suitable in a home environment for substantial reasons. Thanks to their low-resolution, these sensors provide useful data without invading the privacy of the dweller as it could happen using cameras or microphones. Furthermore, these devices are small, cheap, easy to be installed in a normal room, and they can work even in absence of light.

LR-TASs are composed of $m \times n$ infrared sensing elements, acquiring the temperature of a two dimensional area. In this paper we refer to experiments conducted using the Grid-Eye [5] sensor developed by Panasonic. This device is an 8-by-8 LR-TAS with sampling rate of 10 samples/s, a temperature range from -20°C to 80°C with 0.25°C resolution, field of view of 60°, and maximum distance declared to detect humans of 5m. Moreover, Grid-Eye sensor comes with an on-board thermistor that provides the environment temperature from -20°C to 80°C with 0.0625°C resolution. It communicates through I^2C interface to a wireless station which sends all the data to a central processing unit for the analysis.

2. Previous works

Human localization has several applications in Smart Home Environments, e.g., surveillance, health monitoring, and energy management. In particular, LR-TAS has been used to accomplish different tasks.

The goal of the work proposed by Sixsmith et Al. [6] is to detect falls of the older seniors through SIMBAD: Smart Inactivity Monitor using Array-Based Detectors. This system relies on two parallel modules to raise alarms. First of all it analyzes target motion to detect falls characteristic dynamics. Then, it monitors target inactivity and compares it with a map of acceptable inactivity periods in different locations in the field of view. This system has been tested in laboratory simulating predefined fall scenarios, reporting limited results in true positive rate without any false positive. This research includes also results related to a trial lasting two months in a single occupant house. After a training period where experts tuned the system parameters according to the output, still an unacceptably high false-alarm rate emerged.

Erickson et Al. [7] use a Thermal Array Sensor Network in order to measure occupancy of a building. The provided information is used to control heating, cooling, ventilation, and lighting of the building to optimize energy usage. The method proposed by the authors consists in removing the background to detect the pixels of the matrix that refer to humans, followed by an analysis of the connected components using K-Nearest Neighbors classifier, to estimate the number of people. Unfortunately, no significant results have been reported.

In Basu et Al. [8] the authors present a method to estimate number of people and the direction of their motion from LR-TASs data. A Support Vector Machine has been used to classify connected components and local peak counts, estimating the number of persons with 80% accuracy. Finally, they inferred the direction of a subject motion across a set of scenes using cross-correlation analysis.

Mashiyama et Al. [9] report a system for Activity Recognition using LR-TASs. The proposed method aims to detect five activities - No event, Stopping, Walking, Sitting, and Falling - in three steps: human body detection, feature extraction and classification. Considered the high performances of this method in classifying activities as reported by its authors, we have decided to further analyze and reimplement it in order to compare results obtained in a field trial test.

For sake of clarity and completeness let's report the crucial passages of the work proposed by Mashiyama et Al. with the notation used in the rest of this publication. Given an instant of time t, the frame $I(t)$ represents the set of $T_{i,j}$ measurements taken by a LR-TAS sensor at time t, each one related to the corresponding (i, j) pixel. Fixing a time windows τ, the variance $v_{i,j}(t)$ of each pixel with $t \geq \tau$ is computed as it follows:
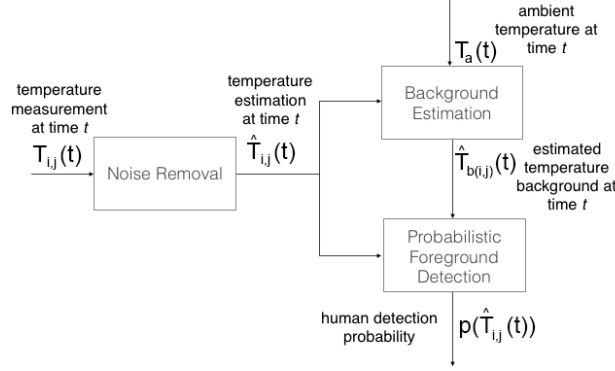
Figure 1: Algorithmic flow of the proposed Human Detection method

$$v_{i,j}(t) = \frac{1}{\tau} \sum_{k=t-(\tau-1)}^{t} \left(T_{i,j}(t) - \overline{T_{i,j}(k)} \right)^2, \text{where } \overline{T_{i,j}(t)} = \frac{1}{\tau} \sum_{k=t-(\tau-1)}^{t} T_{i,j}(t). \tag{1}$$

If the obtained variance $v_{i,j}(t)$ exceeds a given threshold V_{th}, a moving person (Walking, Sitting, or Falling) is detected in the current frame. Conversely, if no movement has been detected, the discrimination between a Stopping person or No event is done according to the difference (T_{diff}) between a person temperature T_p and the background temperature T_b. Given n_{temp} as the number of pixel covered by a standing person, the average of the first n_{temp} pixels of a frame ordered by descending temperature gives T_p. Similarly, the average of the remaining pixels gives T_b. Finally, only if T_{diff} exceeds a given threshold T_{th} a standing person is revealed. The authors tested their system in a test bed experiment, reporting particularly good accuracy results in classifying the mentioned activities, specially considering just the detection phase, excluding the activity classification method.

Most of the work done in activity recognition and human detection using LR-TAS report experimental data obtained by tests performed in a controlled environment. However, as highlighted by Sixsmith et Al. [6], there are some limitations in this approach that have to be considered when building an effective indoor monitoring system. First of all, the positioning of the sensors must take into account the geometry of the environment and its contents to ensure that the vision of the sensor is not obstructed. Indeed, a system to detect falls would be useless if it would not guarantee its effectiveness over the entire walkable area of the house. Moreover, another factor requiring a deep study is the noise management: radiators, appliances, heaters or sunlight reflections have to be considered in the model of the system. Our work is mainly focused in improving human detection's performances handling noisy data.

3. Human detection

The following method aims at retrieving a probability estimation of the presence of at least one person in the LR-TAS field of view. The main steps of the algorithm are summarized in the flow presented in Figure 1: noise removal, background estimation and probabilistic foreground detection.

3.1   Noise removal

LR-TAS (Low Reslution - Termal Array Sensor) raw temperature data are characterized by the presence of noise perturbing the desired measured signal. These type of sensors usually denote low accuracy on a single measurement: Grid-Eye sensor, for example, report the value within Typ. $2.5°C$ [5]. In order to remove the effect of the noise on the measured signal, we will consider the indoor environment as a dynamic system influenced by the external temperature, air conditioning systems, and the presence of people and appliances

in general. Reducing the noise components from the measured signal makes the human detection system more robust to external variations. As already mentioned, Grid Eye is a matrix of 8×8 sensing elements measuring the temperature of a certain region of space. To get the temperature distribution in the space, the temperature evolution process on a single region will be taken as an independent dynamic system and, hence, the measurement made by a single sensing element will be filtered independently from other measurements.

3.1.1 Kalman Filtering

Consider a dynamic system $\mathcal{S}$ represented as follows:

$$\mathcal{S} : \begin{cases} x(t) = F \cdot x(t-1) + \xi(t) \\ y(t) = s(t) + \eta(t) = H \cdot x(t) + \eta(t) \end{cases}, \tag{2}$$

where $s(t)$ is the variable to be estimated, $y(t)$ is the value obtained measuring $s(t)$ which is affected by the measurement noise term $\eta(t)$, $x(t)$ is the state variable at time t, ξ models the process noise, F is the system matrix and H is the measurement matrix. The proposed noise removal technique is based on Kalman Filter (KF) [10] and it is composed of two phases: extrapolation and correction. During the extrapolation phase the filter receives a prediction of the system state $\tilde{x}(t)$ for the current step t using the system state estimate $\hat{x}(t-1)$ made on the previous step. During the correction phase the state prediction $\tilde{x}(t)$ is adjusted by current measurements $y(t)$ to obtain the corrected estimate $\hat{x}(t)$. The state prediction is expressed as:

$$\tilde{x}(t) = F \cdot \hat{x}(t-1), \tag{3}$$

while the state estimate is represented as:

$$\hat{x}(t) = \tilde{x}(t) + K(t) \cdot [y(t) - H \cdot \tilde{x}(t)], \tag{4}$$

where $K(t)$ is the Kalman Gain [10] at time t.

3.1.2 LR-TAS data filtering

In order to obtain the expected value of the measured temperature, separating it from the noise component, we applied the Kalman-Filtering technique described in the previous paragraph. Let the state variable $x(t)$ be represented by $T_{i,j}(t)$: the average temperature of the objects placed in the field of view of the sensing element in position (i, j) at time t. Similarly, $y(t)$ is the measure of $T_{i,j}(t)$ as acquired by the sensing element in position (i, j) at time t. Finally, in order to get the prediction on the state as described in Eq. (3), the system matricies have to be set as follows:

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } H = \begin{bmatrix} 1 & 0 \end{bmatrix}. \tag{5}$$

Thus, from Eq. (3):

$$\tilde{T}_{i,j}(t) = F \begin{bmatrix} \hat{T}_{i,j}(t-1) \\ \Delta\hat{T}_{i,j}(t-1) \end{bmatrix} = \begin{bmatrix} \hat{T}_{i,j}(t-1) + \Delta\hat{T}_{i,j}(t-1) \\ \Delta\hat{T}_{i,j}(t-1) \end{bmatrix}, \tag{6}$$

while the variable to be estimated $s(t)$ can be derived from Eq. (2):

$$s(t) = H\hat{x}(t) = \hat{T}_{i,j}(t) \tag{7}$$

The result of the application of KF to measurements collected by a single sensing element is shown in Figure 2: data have been collected in a perturbed environment with appliances in the sensor field of view and activating the air conditioning system. The influence of these factors on the ambient temperature is evident analyzing the room temperature collected by the on-board sensor thermistor (Fig. 3).
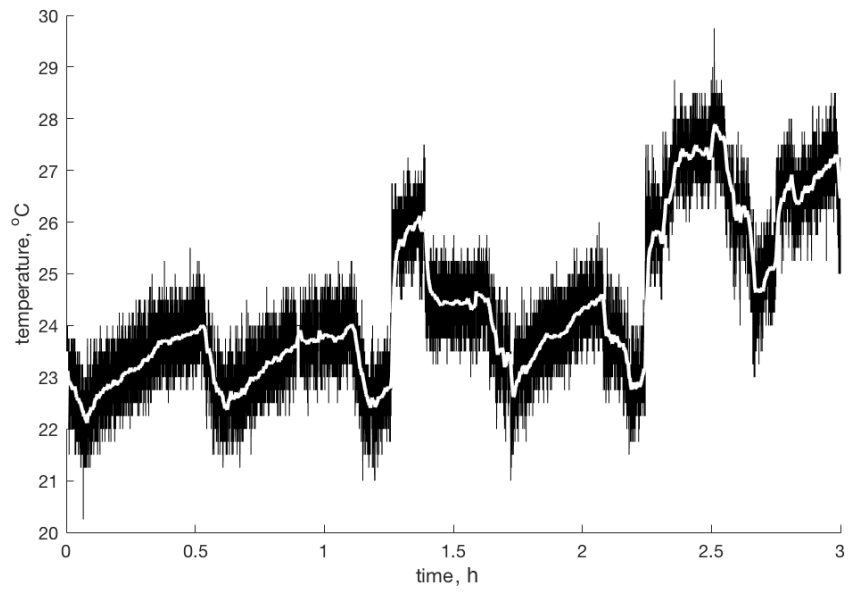
Figure 2: Result of application of Kalman Filter to the temperature measurements done by a single cell sensor: original signal (black) and filtered (white).
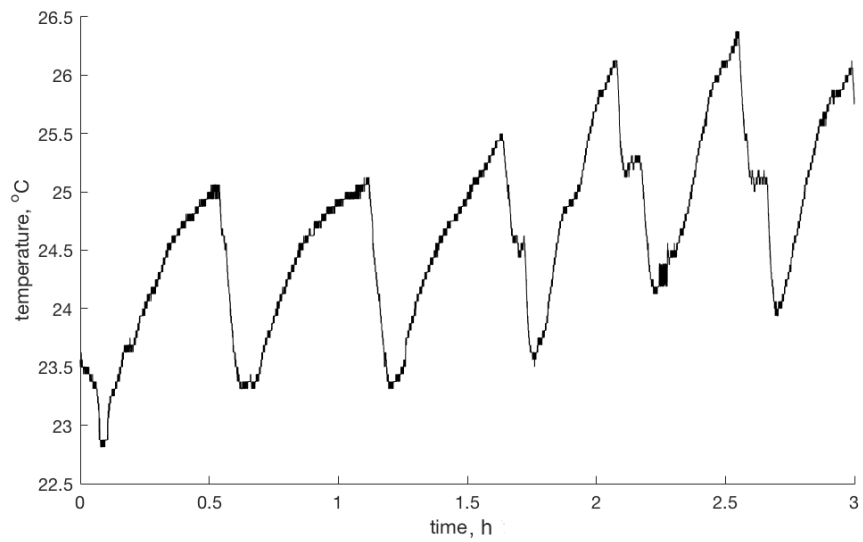

Figure 3: Ambient temperature measurements done by the on-board sensor thermistor. The oscillation is caused by the air conditioning system.
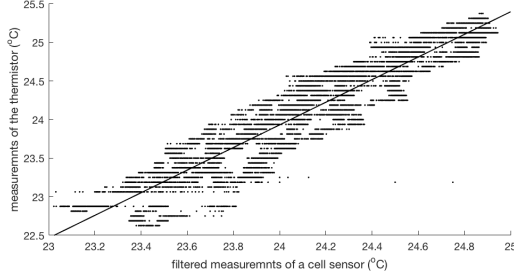
Figure 4: Dependence between ambient temperature and filtered sensor cell measurements.

3.2. Background Estimation

The fundamental assumption to discriminate humans from the background is that the human temperature distribution has to differ from the ambient temperature distribution. In this condition, the human recognition task converges to the analysis of the difference between the current measurements of the sensor cells and the corresponding values of estimated temperature background. Nevertheless, the temperature background estimation should adapt to the environmental condition changes that can be relatively rapid.

Assuming that the thermistor measurement of the ambient temperature is almost not affected by the presence of humans, this information can be used as a reference to detect changes in the environmental conditions. Thus, the dependence between the background temperature $T_{b(i,j)}(t)$ of the sensing element in position (i,j) and the ambient temperature $T_a(t)$ is a function $f(T_a(t)) = T_{b(i,j)}(t)$, it is possible to compute $T_{b(i,j)}(t)$ from $T_a(t)$. The analysis of the sensor cells and thermistor measurements shows a linear dependence (Fig. 4): the average correlation coefficient between $T_a(t)$ and $T_{b(i,j)}(t)$ is 0.62, and between $T_a(t)$ and $T_{\hat{b(i,j)}}(t)$ is 0.9. Hence, the function can be expressed as:

$$T_{b(i,j)}(t) = f(T_a(t)) = \begin{bmatrix} 1 & T_a(t) \end{bmatrix} \cdot \beta. \tag{8}$$

In order to compute β that globally minimized the least square errors, it is necessary to collect a great number of samples. In the final implementation of the proposed method, to make the learning period shorter and let the background estimation algorithm work on-line, an approximation of β is used [11]. Setting:

$$A = \begin{bmatrix} 1 & T_a(t-1) \\ 1 & T_a(t-2) \\ \vdots & \vdots \\ 1 & T_a(t-\tau) \end{bmatrix} \text{ and } B = \begin{bmatrix} T_{i,j}(t-1) \\ T_{i,j}(t-t) \\ \vdots \\ T_{i,j}(t-\tau) \end{bmatrix}, \tag{9}$$

β as been computed as it follows:

$$\hat{\beta}(t) = A^{\ddagger}B, \tag{10}$$

where τ is a time window and $\ddagger$ means pseudo inverse operation.

Thus, the estimated background temperature is computed as it follows:

$$T_{\hat{b(i,j)}}(t) = \begin{bmatrix} 1 & T_a(t) \end{bmatrix} \cdot \hat{\beta}, \tag{11}$$

while the residual squares are given by:

$$R(t) = (T_{\hat{b(i,j)}}(t) - T_{i,j}(t))^2. \tag{12}$$

Finally, since the estimation of $\hat{\beta}$ should involve only the background related measurements, Eq. (10) is extended using the analysis of the residual squares:
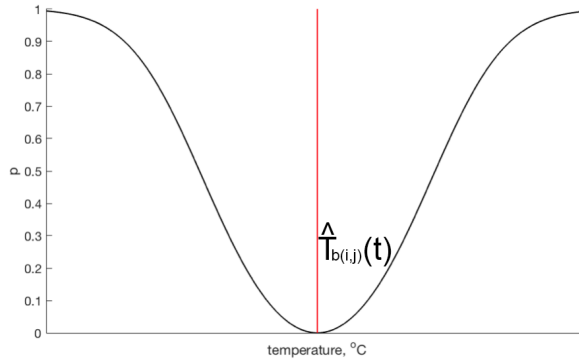
6

Figure 5: Modeled probability function for foreground detection.

$$\hat{\beta}(t) = \begin{cases} \hat{\beta}(t-1), & \text{if } R(t) \geq R_{th} \\ A^{\ddagger}B & \text{otherwise} \end{cases} \tag{13}$$

3.3. Probabilistic Foreground Detection

In order to provide as much information as possible in uncertain situations, the proposed method also computes for every cell the probability of human detection in every instant. For this reason, we modeled the probability function $q(T_{i,j}(t))$, describing whether the measurement belongs to the background temperature distribution T_b, as a logistic function:

$$q(T_{i,j}(t)) = \frac{2}{1 + e^{k \cdot R(t)}}, \tag{14}$$

where k is the steepness of the function and the probability $p(T_{i,j}(t))$ that that measurement does not belong to the background distribution is (Fig. 5):

$$p(T_{i,j}(t)) = 1 - q(T_{i,j}(t)). \tag{15}$$

4. Installation and Results

The proposed method aims to improve the accuracy of Human Detection algorithms using LR-TAS data in noisy environments. For this reason, the environment has been perturbed during the experiment using air-conditioning system, appliances, and exposing the sensor to sunlight reflection.

4.1. Sensor mount

In the literature, there are two opinions on the placement of LR-TAS for human detection: the wall [12], [13] and the ceiling [9]. After carrying out several experiments, it was found that the installation on the wall has several drawbacks:

- Under real condition, furniture and other objects can obstruct the view of the sensor.

- The movement of a human - coming closer and further to the installation point - influences the amount of pixels representing him/her.

- Since the sensor tends to average the value of the temperature in the observed space, human movement also affects the temperature distribution.
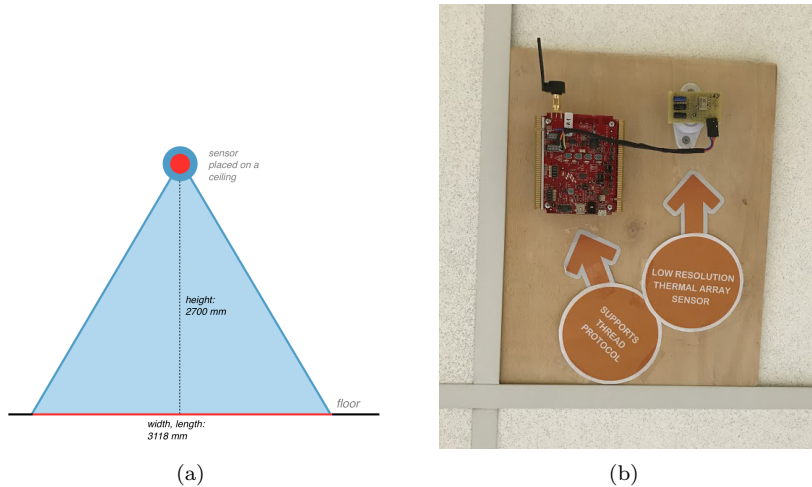
Figure 6: LR-TAS: ceiling installation.

For these reasons, the sensor was mounted on the ceiling at the height of 2.7m with a resultant detection area on the floor of approximately $9m^2$ (Fig. 6). Streaming data have been transmitted and stored on a central device. The comunnication has been implemented using Thread protocol [14]: an innovative solution designed by the Thread Group for IoT applications. This protocol allows to easily create an IPv6, meshed, robust and secure network of sensors.

4.2. Experiments

We have collected several datasets for a total duration of 4 days. During this period people were asked to perform everyday activities passing and staying under the sensor. The experiment data have been manually annotated to validate the proposed algorithm: every pixel is labeled as "1" if it represents human and "0" otherwise.

4.3. Results

The proposed method has been tested on the retrieved datasets. Figure 7(a) shows the performance of the temperature background estimation for a single pixel sensor. Comparing it with Figure 7(b), it is visible that where the filtered measurements are distant from the estimated background, the probability of human detection is very high.

In order to compare the obtained results with the method that, in our knowledge, reports the best accuracy value in the literature, we implemented also the human detection algorithm proposed by Mashiyama et Al. [9]: Figure 7(c) shows T_{diff} computed using the method proposed by Mashiyama. Unfortunately, authors do not suggest any method to compute the threshold T_{th} and it is very difficult to manually tune it.
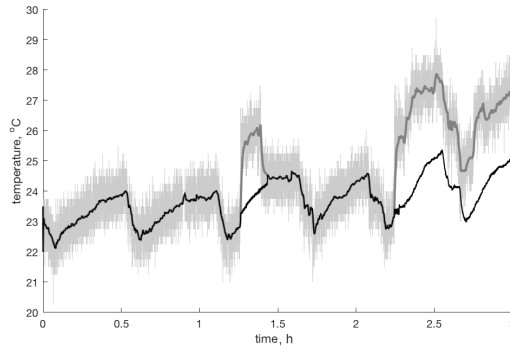
To compare the results of the work of two algorithms, the measurement is said to represent human when a frame contains at least one measurement whose probability $p(m(t)) > p_{th}$. The parameters used in both of the algorithms are provided in Table 1, while the obtained results are presented in Table 2.
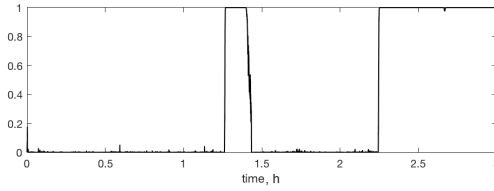
5. Conclusions

We have presented a novel technique to detect humans in indoor environments using Low Resolution Thermal Array Sensor. This approach consider the temperature variation in the room due to external dynamics and noise. A Kalman Filter has been used to filter the noise on the temperature measurements while a background

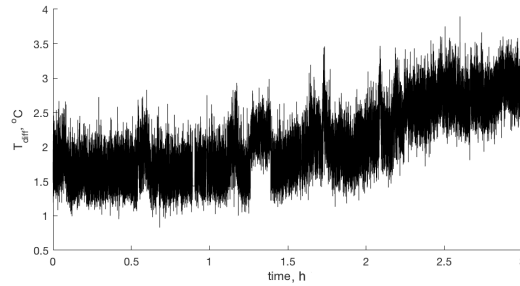| Proposed method		Mashiyama et Al. method	
τ	10	τ	10
R_{th}	0.01	n_{temp}	3
k	1.5	T_{th}	2.5
p_{th}	0.9	V_{th}	0.6

Table 1: Parameters used in the proposed method and Mashiyama et Al. method

(a) Background estimation results: original measurements (light gray), filtered measurements (gray), estimated background(black)

(b) The probability of Human Detection computed with the proposed method

(c) T_{diff} as computed with the method by Mashiyama et Al.

Figure 7: Outcome of different steps computed over 3 hours sample data.

	Proposed method	Mashiyama et Al. method
TP	0.9988	0.3964
TN	0.9408	0.9950
FP	0.0592	0.0050
FN	0.0012	0.6036
precision	0.9441	0.9876
recall	0.9988	0.3964
ACCURACY	0.97	0.70

Table 2: Comparison between the proposed method and Mashiyama et Al. method in detecting humans using LR-TAS in a noisy environment.

estimation technique aims to separate the background from humans.

Final results show an improvement in the Human Detection accuracy compared to the state of the art when performing a field trial in a real environment passing from 70% to 97%.

The mentioned results have been collected using a single sensor installation, however a multisensor system needs to be implemented in order to setting up a real scenario in a smart environment. This extension, which requires to handle technical e theoretical problems - from placing the sensors to retrieving one overall model with the global state of dwellers and environment - will be part of our future work.

Acknowledgments

References

[1] U. Nations, "World population ageing 2015," tech. rep., Department of Economic and Social Affairs, Population Division, New York, 2015.

[2] I. Bierhoff, A. van Berlo, J. Abascal, B. Allen, A. Civit, K. Fellbaum, E. Kemppainen, N. Bitterman, D. Freitas, and K. Kristiansson, "3. smart home environment," 2007.

[3] S. Mangano, H. Saidinejad, F. Veronese, S. Comai, M. Matteucci, and F. Salice, "Bridge: Mutual reassurance for autonomous and independent living," Intelligent Systems, IEEE, vol. 30, no. 4, pp. 31–38, 2015.

[4] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, and J. Lundy, "Fall detection-principles and methods," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1663–1666, IEEE, 2007.

[5] "Panasonic - grid-eye: infrared array sensor." https://na.industrial.panasonic.com/products/sensors/sensors-automotive-industrial-applications/grid-eye-infrared-array-sensor, 2016.

[6] A. Sixsmith and N. Johnson, "A smart sensor to detect the falls of the elderly," IEEE Pervasive Computing, vol. 3, no. 2, pp. 42–47, 2004.

[7] V. L. Erickson, A. Beltran, D. A. Winkler, N. P. Esfahani, J. R. Lusby, and A. E. Cerpa, "Thermosense: thermal array sensor networks in building management," in Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, p. 87, ACM, 2013.

[8] C. Basu and A. Rowe, "Tracking motion and proxemics using thermal-sensor array," arXiv preprint arXiv:1511.08166, 2015.

[9] S. Mashiyama, J. Hong, and T. Ohtsuki, "Activity recognition using low resolution infrared array sensor," in 2015 IEEE International Conference on Communications (ICC), pp. 495–500, IEEE, 2015.

[10] R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.

[11] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning, vol. 1. NY Springer, 2013.

[12] L. I. L. Gonzalez, M. Troost, and O. Amft, "Using a thermopile matrix sensor to recognize energy-related activities in offices," Procedia Computer Science, vol. 19, pp. 678–685, 2013.

[13] Y. Jeong, K. Yoon, and K. Joung, "Probabilistic method to determine human subjects for low-resolution thermal imaging sensor," in Sensors Applications Symposium (SAS), 2014 IEEE, pp. 97–102, IEEE, 2014.

[14] "Thread group - iot designed network protocol." https://www.threadgroup.org/, 2017.