

Article

Energy Savings in Buildings Based on Image Depth Sensors for Human Activity Recognition

Omar Mata ¹, Juana Isabel Méndez ¹, Pedro Ponce ^{1,*}, Therese Pepper ², Alan Meier ³ and Arturo Molina ^{1,*}¹ Institute of Advanced Materials for Sustainable Manufacturing, Tecnológico de Monterrey, Monterrey 64849, Mexico² Institute for Energy and Environment, University of California, Berkeley, CA 94720, USA³ Energy and Efficiency Institute, University of California, Davis, CA 95616, USA

* Correspondence: pedro.ponce@tec.mx (P.P.); armolina@tec.mx (A.M.)

Abstract: A smart city is a city that binds together technology, society, and government to enable the existence of a smart economy, smart mobility, smart environment, smart living, smart people, and smart governance in order to reduce the environmental impact of cities and improve life quality. The first step to achieve a fully connected smart city is to start with smaller modules such as smart homes and smart buildings with energy management systems. Buildings are responsible for a third of the total energy consumption; moreover, heating, ventilation, and air conditioning (HVAC) systems account for more than half of the residential energy consumption in the United States. Even though connected thermostats are widely available, they are not used as intended since most people do not have the expertise to control this device to reduce energy consumption. It is commonly set according to their thermal comfort needs; therefore, unnecessary energy consumption is often caused by wasteful behaviors and the estimated energy saving is not reached. Most studies in the thermal comfort domain to date have relied on simple activity diaries to estimate metabolic rate and fixed values of clothing parameters for strategies to set the connected thermostat's setpoints because of the difficulty in tracking those variables. Therefore, this paper proposes a strategy to save energy by dynamically changing the setpoint of a connected thermostat by human activity recognition based on computer vision preserving the occupant's thermal comfort. With the use of a depth sensor in conjunction with an RGB (Red–Green–Blue) camera, a methodology is proposed to eliminate the most common challenges in computer vision: background clutter, partial occlusion, changes in scale, viewpoint, lighting, and appearance on human detection. Moreover, a Recurrent Neural Network (RNN) is implemented for human activity recognition (HAR) because of its data's sequential characteristics, in combination with physiological parameters identification to estimate a dynamic metabolic rate. Finally, a strategy for dynamic setpoints based on the metabolic rate, predicted mean vote (PMV) parameter and the air temperature is simulated using EnergyPlus™ to evaluate the energy consumption in comparison with the expected energy consumption with fixed value setpoints. This work contributes with a strategy to reduce energy consumption up to 15% in buildings with connected thermostats from the successful implementation of the proposed method.

Citation: Mata, O.; Méndez, J.I.; Ponce, P.; Pepper, T.; Meier, A.; Molina, A. Energy Savings in Buildings Based on Image Depth Sensors for Human Activity Recognition. *Energies* **2023**, *16*, 1078. <https://doi.org/10.3390/en16031078>

Academic Editors: Edmundas Kazimieras Zavadskas and Daniele Milone

Received: 7 November 2022

Revised: 28 November 2022

Accepted: 12 December 2022

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: energy savings; recurrent neural networks; metabolic rate; thermal comfort; depth sensor; computer vision

1. Introduction

The International Energy Agency (IEA) estimated that buildings have become the third largest energy consumer in the world [1]. Generally, energy usage in buildings is expended on lighting, electrical equipment, and heating, ventilation, and air conditioning (HVAC) systems. HVAC systems, which play an important role in ensuring occupant comfort, are among the largest energy consumers in buildings with up to 60% of the total

energy consumption in households [1]. Performance enhancements to traditional HVAC systems therefore offer an exciting opportunity for significant reductions in energy consumption.

Several studies show that almost 50% [2] of USA's energy usage in buildings is utilized for indoor climate conditioning. Moreover, the worldwide energy consumption by HVAC equipment also shows considerable high values. In Europe, they represent around 40% of energy consumption [3]. In China, about 20% of the total energy usage in the year 2004 is reported with a constant annual increase [4]. In the Middle East, more than 65% of energy consumption is reported for use of cooling systems [5] and in Mexico, the cooling system makes up to 44% of the total energy consumption [6].

The increase in building energy consumption is highly affected by building design, change of occupant comfort standard, building operation, maintenance, and HVAC system design. Especially important has been the intensification of energy consumption in HVAC systems, which has now become almost essential in parallel to the spread in the demand for thermal comfort, considered a luxury not long ago. All those aspects should be conceived with energy consumption and occupant comfort in mind.

Thermal comfort is all about human satisfaction with the thermal environment. The design and calculation of air conditioning systems to control the thermal environment in a way that also achieves an acceptable standard of air quality inside a building should comply with the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standard 55-2017 [7]. This standard acknowledges the two main research areas in thermal comfort: thermal physiology and human behavior.

As for the first area, it includes two common indicators called predicted mean vote (PMV) and the predicted percentage of dissatisfied (PPD) known as the Fanger's model [2,3]. The PMV predicts the mean value of the votes on the seven-point thermal sensation scales:

- +3 Hot;
- +2 Warm;
- +1 Slight warm;
- 0 Neutral;
- −1 Slight cool;
- −2 Cool;
- −3 Cold.

On the other hand, the PPD represents the prediction of the percentage of occupants that feel uncomfortable and its value ranges from 0% to 100%. Thus, an acceptable range of thermal comfort goes from slight warm to slight cool for 20% of dissatisfied for the residential sector [2]. The ASHRAE-55 calls this the PMV-PPD model described by the Fanger's equation, considered to be the milestone of the development of a thermal model [8]:

$$M - W = C + R + E + (C_{res} + E_{res}) + S \quad (1)$$

where:

M: the metabolic rate;

W: mechanical work is done;

C: convective heat loss from the clothed body;

R: radiative heat loss from the clothed body;

E: evaporative heat loss from the clothed body;

C_{res} : convective heat loss from respiration;

E_{res} : evaporative heat loss from respiration;

S: the rate at which heat is stored in the body tissues.

The ASHRAE-55 standard uses empirical tables with common activities and their respective met units (metabolic rate) and clothing insulations for different garments in clo

units. However, the use of fixed value tables has some limitations in real time analysis and only works for statistical analysis or estimations.

The second area is related to a hypothesis in which the perception of thermal comfort is related to outdoor weather conditions, and it is based on adaptive opportunities of occupants to control their own comfort [1,4,9]. This thermal adaptation was defined by three categories [10]: behavioral adjustment which includes personal modifications such as removing garments, or doing physical activity, and external modifications such as opening a window or changing the air conditioner; physiological adaptation referring to the acclimatization or even genetic adaptation; psychological adaptation refers to the expectations due to past experiences [10].

Both PMV and the adaptive models are aggregate models, which means they are designed to predict the average thermal comfort of large populations and ultimately present limitations in real case scenarios [11]. As many studies have shown [12–15], the measurement of thermal comfort in office buildings is limited by the subjectivity and high dependency on the six mandatory parameters used for heating and air conditioning setpoint controls. Four variables are related to the environment:

- Air temperature;
- Air speed;
- Humidity;
- Radiant temperature.

Two variables are related to the occupant:

- Metabolic rate;
- Clothing insulation.

Furthermore, the work in [16] highlights the difficulty and the cost of obtaining some of these variables. For example, mean radiant temperature and air speed are two of the environmental variables that are not typically monitored as they require expensive instruments for measuring [16]. Moreover, the two personal variables metabolic rate and clothing insulation are said to be impossible to collect in real time [17]; hence, the process is simplified with the assumed values or fixed set of data collected from laboratory or field measurements, which ultimately causes erroneous estimations [14].

Hence, the contribution of this paper is focused on a methodology for an on-line estimation of the metabolic rate of a single occupant to improve simulations of energy consumption in smart homes with an HVAC system, as there is still not a way to measure the two variables of thermal comfort related to the occupant: metabolic rate and clothing insulation [17]. The metabolic rate estimation is based on Human Activity Recognition with RGB-D data using a skeleton-based model over a 3D representation with a recurrent neural network as the classification method. The RGB-D data are intended to reduce privacy issues in comparison with the RGB data, as the 3D skeleton model is used to reduce data used for the classification method compared with a pixel's image data. The recognized activity is paired with a metabolic rate value that is used as an input variable for the human-centered approach of the adaptive thermal comfort on a simulation for comparing energy savings between setpoints with fixed values and adaptable setpoints.

This paper is organized as follows: Section 2 shows the literature overview for human activity recognition. Section 3 describes the materials and methods used for the proposal. Section 4 shows the results of the proposal. Finally, Section 5 discusses and presents the improvements from implementing depth sensors for activity recognition and its impact on energy consumption analysis in HVAC systems.

2. Literature Overview

Thermostats that control HVAC systems are employed in about 85% of households; thus, they represent an opportunity for saving energy at home. Initial approaches for saving energy through connected thermostats are presented with gamification techniques [18,19], data analysis [20], behavior analysis [21], and usability of interfaces [22–24].

However, a first approach using the adaptive model to measure the differences between increasing or decreasing the thermostat setpoint depending on the season was analyzed. Reductions and thermal comfort were achieved. The research suggested more information about the householder and how, for instance, clothing insulation and metabolic rates affect thermal comfort. Hence, in [25,26], the authors proposed to measure thermal comfort in smart homes through dynamic clothing insulation with cameras; the activities were inferred depending on the position of the householder and the clothing insulation of twelve homes. They found that energy reductions are achievable and that tailored strategies were required as not all the homes achieved thermal comfort, and there were homes where the comfort was not met and the energy consumption increased. Therefore, in [27,28], they proposed using cameras to measure the *clo* value dynamically through a Convolutional Neural Network (CNN) model classification and obtain the thermal comfort range of a household in Concord, California. These approaches considered only the *clo* value and assumed a metabolic rate of 1.0. Furthermore, in [25,26], the authors pointed out the need to measure the activities as well to obtain dynamic thermal comfort models instead of conventional models.

Human Activity Recognition (HAR) requires a series of physical actions that construct one physical activity, where a physical action is defined as any bodily movement produced by skeletal muscles and the activity itself is the sequence of those produced movements [29]. HAR research focuses on types of activities that humans perform within a time interval; thus, it is based on sensors and/or video data analysis. Moreover, the two types of sensors found in HAR are: wearables and ambient sensors [30].

Wearables sensors are attached to the person's body to measure a given attribute such as electrocardiogram (ECG), location, temperature, motion, electroencephalogram (EEG), etc. [31–34]. All the data of these sensors may be sent to another device for processing; regularly this device can be a computer, an embedded system, or a smartphone. Moreover, the smartphone itself can be used as a wearable sensor because of all the technological progress made on them [35–37]. The main disadvantages of wearables are they require batteries, thus charging them may be annoying for the user; some data may be synchronized manually because of no communication between them; and finally, the user may feel them intrusive so they may not wear them at all [38].

On the other hand, ambient sensors are not intrusive and may be connected directly to a source of power. Video cameras today are low-cost devices to obtain the necessary data over time. A sequence of images is directly used to make human activity recognition and they need to be processed by a computer. The disadvantage of the camera is that there may be privacy issues; therefore, a strong acceptance of this technology may be needed by the user. Other ambient sensors (also known as binary sensors) such as motion detectors, pressure sensors, contact switches, etc., can be an effective way to track human activity [39,40].

The field of HAR has become an important research area due to its increasing number of applications and therefore, recent surveys about this field offer a precise description of state-of-the-art methods, publicly available databases, and actual research challenges [30,41–44]. The vision-based HAR research is divided based on data type; the most common is Red, Green, and Blue (RGB) data from a normal camera (CCTV, webcam, etc.) and the Red, Green, Blue and Depth (RGB-D) data that incorporate depth information. The RGB data have achieved lower accuracy compared to the RGB-D data but the configuration complexity, high cost, and the need for big datasets of the RGB-D data are the reasons why RGB data is mainly used [45].

A key component for vision-based HAR is human body modeling. The three most common types are: skeleton-based model, contour-based model, and volume-based model [46]. The skeleton-based model represents a set of joint locations following a human body skeletal structure; this model is visually identified as a stick-figure. The contour-based model contains the contour information of the human body, and it is often represented with rectangles of a person silhouette. Lastly, the volume-based model is

represented in 3D by geometric shapes (cylinders, cones, cubes) or meshes that resemble a human body [47].

Depending on the data type and the human body model used for HAR, a 2D or 3D representation of the data can be selected to work with. Regarding the data type, the RGB only offers information in 2D while the RGB-D can work with 2D and 3D data representation. On the other hand, regarding human body modeling the skeleton-based model can be used with 2D and 3D data, the contour-based model with 2D data and the volume-based-model with 3D data [47].

Finally, the most used methods of classification for human activities can be divided in two: conventional neural networks and other machine learning methods [41]. Machine learning methods such as K-nearest neighbors (KNN), Support Vector Machines (SVM), decision trees, and Hidden Markov Models (HMM) are mainly used with wearables sensors and some ambient sensors; while neural networks methods such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) are used with vision-based sensors [41,43].

Figure 1 depicts a diagram with the previously described classification of vision-based HAR-related works based on: data representation, data type, body modeling and methods for classification. Moreover, the figure shows the characteristics this paper focuses on.

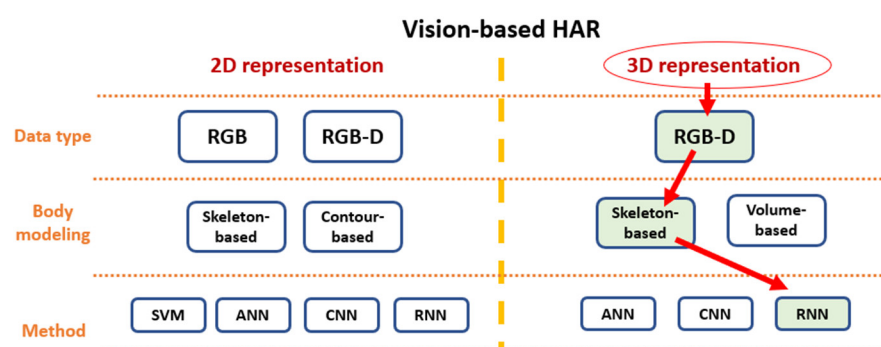


Figure 1. Vision-based HAR components classification based on the literature overview.

3. Materials and Methods

Figure 2 presents the general methodology proposed for obtaining MET values depending on the activity detected on-line. A MET is a ratio of the working metabolic rate relative to the resting metabolic rate, where one MET is the energy spent sitting at rest. First, the data preprocessing consists of obtaining the 3D-axis joints information of a human skeleton shape detected by a depth vision sensor. This 3D data is then transformed to change orientation and size in a new reference 3D plane. Next, the transformed 3D data is input in a Recurrent Neural Network with a classification layer that detects the activity performed by a human; this network needs to be trained with a custom-created database. Finally, a MET value is obtained depending on the activity detected. On the other hand, a simulation of energy usage on a house with an HVAC adaptive setpoint based on the MET value over a whole week is made to get an estimated energy saving that can be achieved with the proposed methodology.

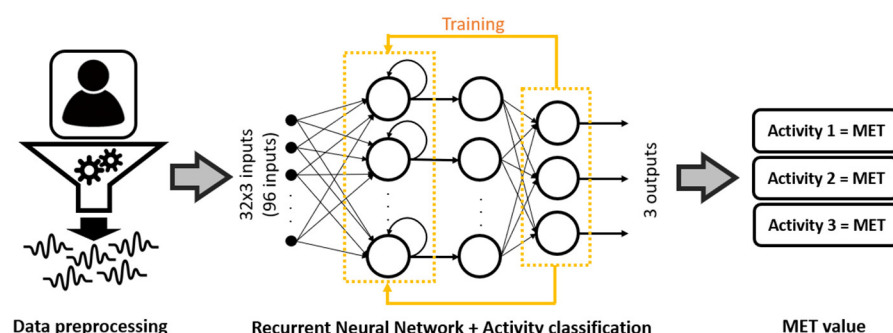


Figure 2. General methodology for estimating an on-line MET value with an RNN.

3.1. Data Preprocessing

The data preprocessing consists of five steps that take a combination of image and depth information into signals that can be classified to recognize different activities. Figure 3 depicts the five steps of the proposed methodology. Each step is described next.



Figure 3. Proposed five steps for image preprocessing.

The first step is to extract the joints in a 3D coordinate system of a skeleton model using the Azure Kinect Body Tracking SDK (Software Development Kit). This uses the depth sensor's information built in the Kinect Azure and an ANN to track multiple human bodies at the same time. Then, the preprocessing is made with the 3D coordinate information of each joint for the objective to make the data invariant to different orientations towards the camera.

The skeleton model consists of 32 joints (Figure 4a) over a 3D frame of reference that depends on the actual position of the camera as shown in Figure 4. The 3D coordinate system is represented as metric $[X, Y, Z]$ coordinate triplets with units in millimeters. The origin $[0, 0, 0]$ is located at the focal point of the camera with the orientation such that the positive X-axis points right, the positive Y-axis points forward, and the positive Z-axis points up.

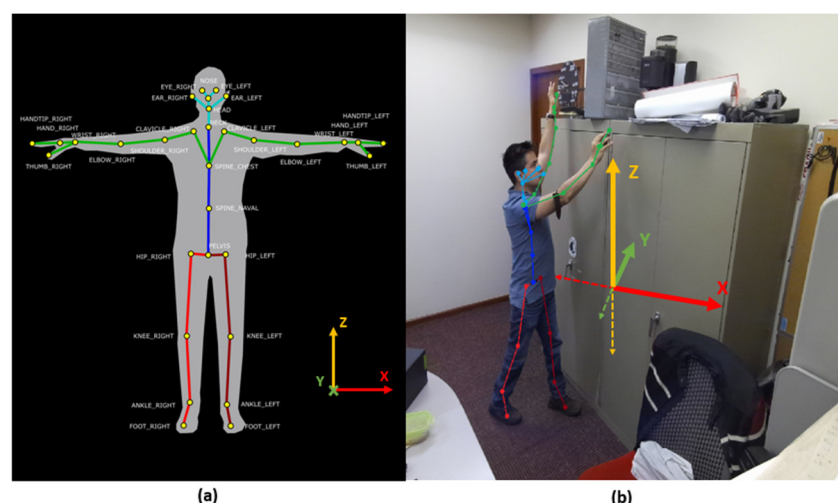


Figure 4. (a) Skeleton model joints; (b) skeleton model detected over the RGB image with its 3D coordinate reference.

As the joint position and orientation are estimates relative to the global depth sensor's frame of reference, the second step of the preprocessing is to translate the skeleton joints to a new reference centered on the origin of the coordinate plane to eliminate distance variability of the subject. Figure 5 depicts the original joints data on a side view (plane YZ) and a top view (plane XY) with the translated data, which now is centered on the joint corresponding to the pelvis on the origin (0, 0, 0). In the same Figure 5, it can be noticed that the 3D model is inclined in reference to plane XY, although the real position of the body is in perpendicular position to the floor (standing). This is due to the actual position of the camera/depth sensor in reference to the floor. Hence, the next step is to correct the pitch and roll rotations due to the position of the camera/depth sensor to make the activity recognition independent of where the camera/depth sensor is located. Figure 6 depicts the roll and pitch angles referenced to the camera and how it can affect the perspective view of the person.

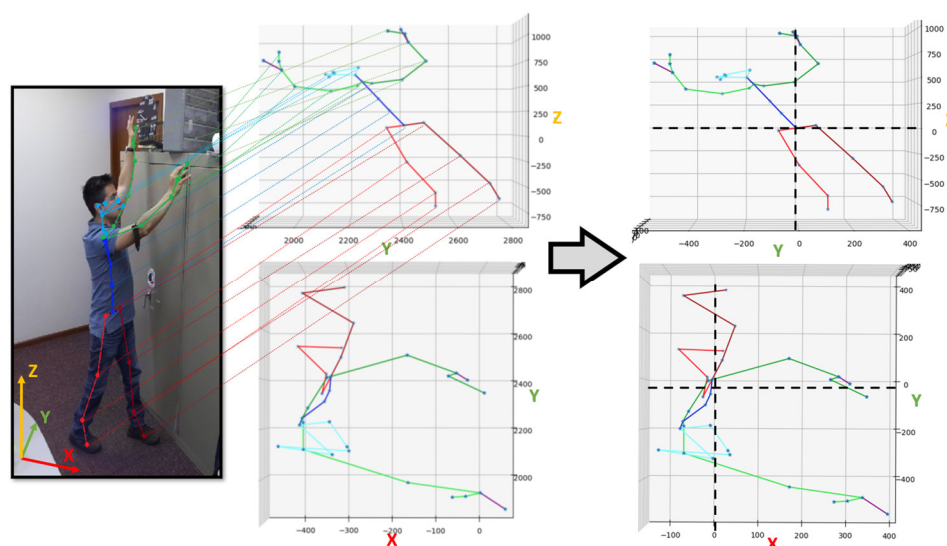


Figure 5. Step 2: Joints translation to the origin of the coordinate plane (0, 0, 0) on the pelvis joint.

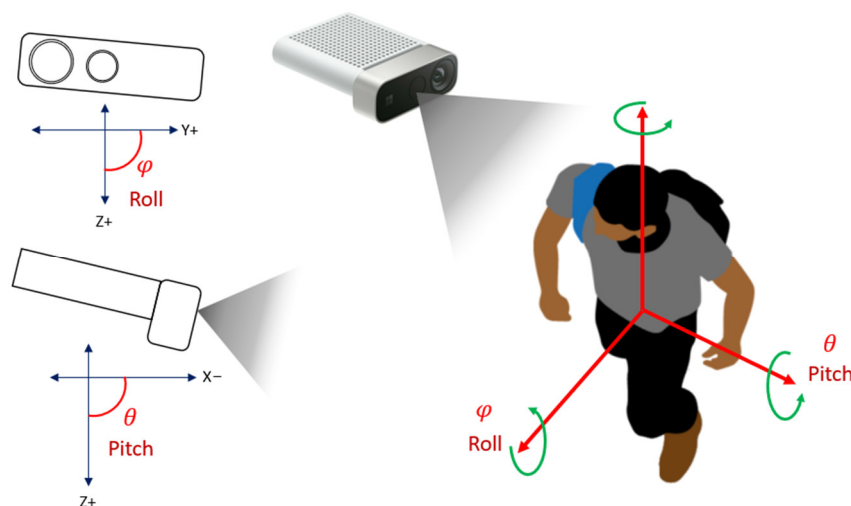


Figure 6. Pitch and roll angles from the camera and human body perspective correction.

The correction of the position of the camera/depth sensor uses the Inertial Measurement Unit's (IMU) accelerometer of the device with three axes to determine the angle individually of each axis. The reference position of the device is taken with the X- and Y-

axes in the plane of horizon with 0 g field and the Z-axis orthogonal to the horizon with 1 g field in rest.

To perform a rotation in Euclidean space, a rotation matrix is used to transform a vector such as the earth's gravitational field vector g . In a 3D coordinate system, the rotations of the X-, Y- and Z-axes in a counterclockwise direction when looking towards the origin are represented by the matrices in Equations (2)–(4) [48]:

$$R_x(\varphi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & \sin(\varphi) \\ 0 & -\sin(\varphi) & \cos(\varphi) \end{bmatrix}, \quad (2)$$

$$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad (3)$$

$$R_z(\phi) = \begin{bmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

Any rotation can be given as a composition of rotations about three axes (Euler's rotation theorem), thus the matrix \mathbf{R} with the order roll, pitch, and yaw and with the effect of the earth's gravitational field of 1 g initially aligned downwards along the Z-axis is shown in Equation (5) and solved in (6) and (7):

$$\mathbf{R}_{xyz} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = R_x(\varphi)R_y(\theta)R_z(\phi) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (5)$$

$$\begin{aligned} \mathbf{R}_{xyz} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ \cos \phi \sin \theta \sin \varphi - \cos \varphi \sin \phi & \cos \varphi \cos \phi + \sin \theta \sin \varphi \sin \phi & \cos \theta \sin \varphi \\ \cos \varphi \cos \phi \sin \theta + \sin \varphi \sin \phi & \cos \varphi \sin \theta \sin \phi - \cos \phi \sin \varphi & \cos \theta \cos \varphi \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned} \quad (6)$$

$$\mathbf{R}_{xyz} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin \theta \\ \cos \theta \sin \varphi \\ \cos \theta \cos \varphi \end{bmatrix} \quad (7)$$

Rewriting Equation (7) relating the normalized accelerometer reading A to the rotation angles gives Equation (8):

$$\begin{bmatrix} -\sin \theta \\ \cos \theta \sin \varphi \\ \cos \theta \cos \varphi \end{bmatrix} = \frac{1}{\sqrt{A_x^2 + A_y^2 + A_z^2}} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} \quad (8)$$

Thus, the roll and pitch angles equations can be obtained solving Equation (8) with the normalized accelerometer reading as shown in Equations (9) and (10) [49]:

$$\theta = \tan^{-1} \left(\frac{A_x}{\sqrt{A_y^2 + A_z^2}} \right), \quad (9)$$

$$\varphi = \tan^{-1} \left(\frac{A_y}{\sqrt{A_x^2 + A_z^2}} \right), \quad (10)$$

where:

A_x —normalized accelerometer reading in X-axis;

A_y —normalized accelerometer reading in Y-axis;

A_z —normalized accelerometer reading in Z-axis.

Figure 7 shows the result of step 3 that corrects the human skeleton position by applying pitch and roll rotations. The rotation is applied to the skeleton with Euler angle transformations [50] with the parameters previously calculated from the accelerometer. This process is only made once at starting the activity recognition as it is assumed the depth sensor is in a fixed position.

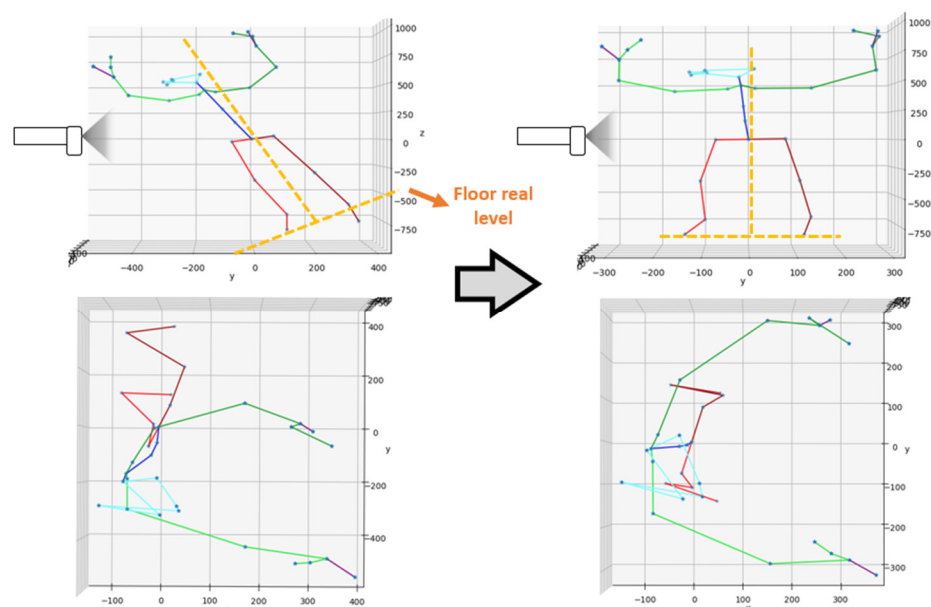


Figure 7. Step 3: Pitch and roll angles correction.

Step 4 is to make a yaw rotation of the skeleton (through the Z-axis) to make it always face front to the depth sensor (anatomically anterior position) as Figure 4a. Figure 8 depicts the implemented rotation of the skeleton from a side view (YZ plane) and the top view (plane XY). This process takes the relative position of the left and right clavicle joints forming a vector that should be parallel to the X-axis and the nose-head joint vector pointing negative into the Y-axis. In this way the invariance in position of the skeleton is achieved to be always in the same reference. The last step of the transformations is to eliminate the variance of heights.

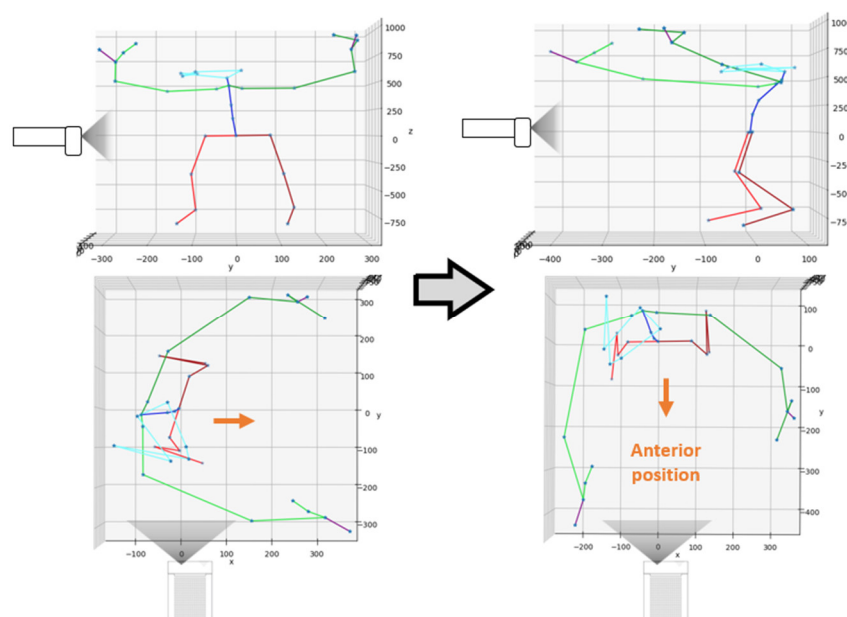


Figure 8. Step 4: Yaw rotation on skeleton to face on $-Y$ direction (anatomically anterior position).

As the reference values of the joints are in millimeters, the subjects' measures add a variable to the classification process that should be removed. Figure 9 depicts the result of step 5, the normalization of all values from -1 to 1 , applying Equation (11) to each joint's data.

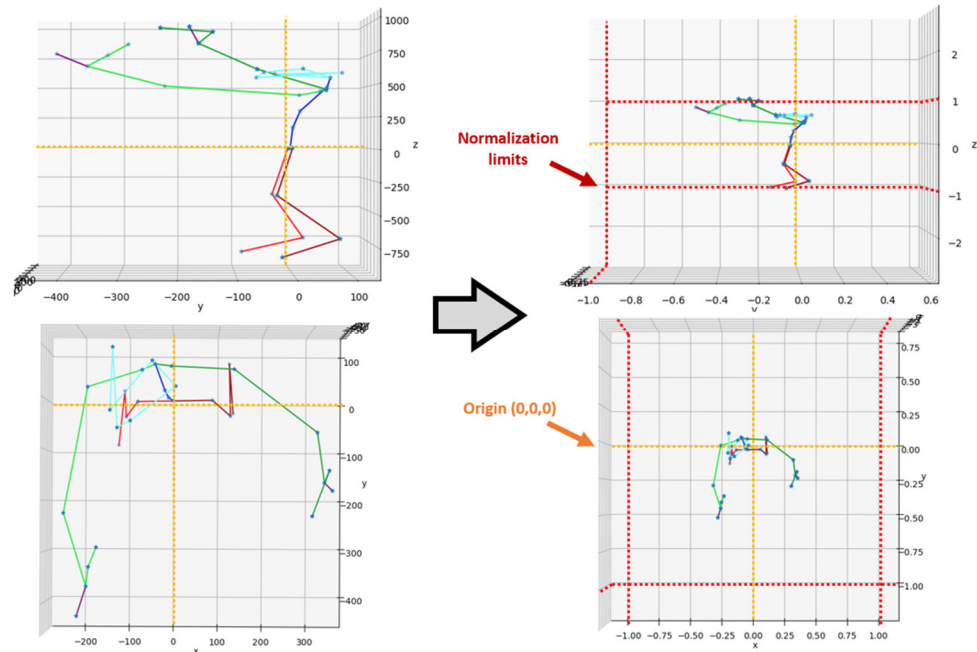


Figure 9. Step 5: Normalization of skeleton joint values between -1 and $+1$ (scaling).

$$x' = 2 \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (11)$$

Finally, this process is repeated for each frame of the captured data as a set of frames will be required to detect the activity. The time series generated with the data of each of the three axes of each joint is used as one feature for the Recurrent Neural Network. Figure 10 depicts the data that make up the input for the RNN, the sequential data of each axis for every joint.

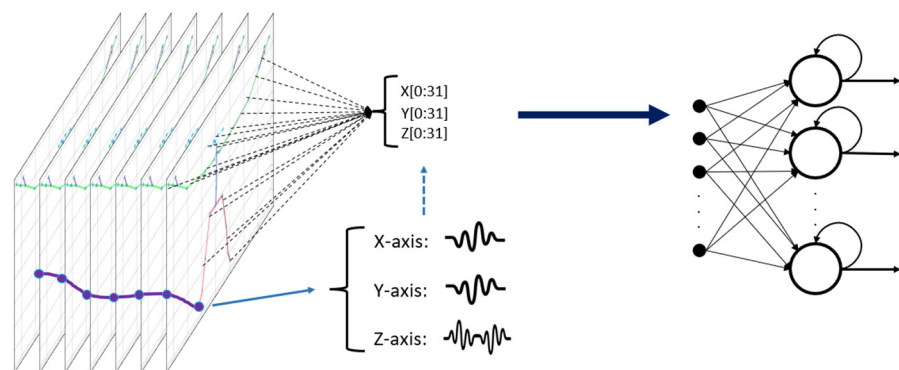


Figure 10. Frame sequence representing time series data for three axes of each joint.

3.2. RNN + Activity Classification

Recurrent Neural Networks (RNN) are typically used to solve time series analysis problems, hence the use of this type of network in the Human Activity Recognition problem.

Figure 11 depicts a representation of an RNN where X_t is some input in the form of a vector representing a time series, h_t is the output hidden state vector, and the blue line is the loop representing that the output is fed back as an input in the network. Unrolling the basic representation of the RNN, it is clear that the loop allows information to be passed from one step of the network to the next, where t represents the number of observations in time. Therefore, an RNN consists of a function F dependent on the past state vector and the current input feature which outputs the current hidden state vector h_t , as stated in Equation (12):

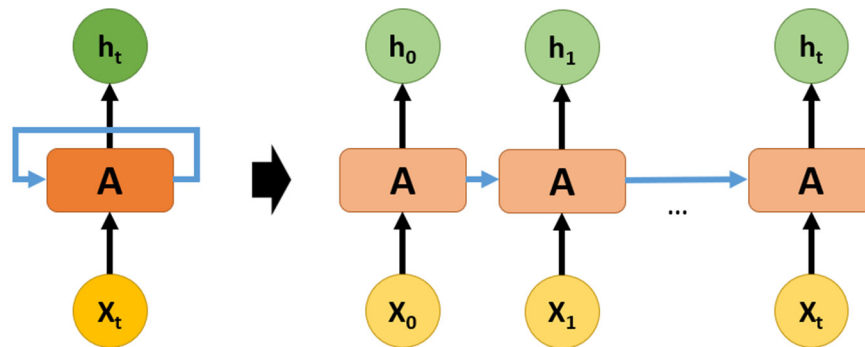


Figure 11. Recurrent Neural Network unrolled equivalent.

$$h_t = F(X_t, h_{t-1}) \quad (12)$$

However, the RNN is highly susceptible to the vanishing gradient problem because the hidden layer of one observation is used to train the hidden layer of the next observation, meaning that the cost function of the network is calculated for each observation [51]. Therefore, the cost function calculated at a deep layer will be used to change the weights of neurons at the shallow layers; because of the multiplicative nature of the backpropagation algorithm, the gradients calculated at those deep layers either have too small or too large of an impact on the weights of neurons in the shallow layers [51]. This effect is depicted in Equation (13), where the gradient on the current state vector h_c from the past state vector h_p is the product of gradients for all intermediate state vectors:

$$\frac{\partial h_c}{\partial h_p} = \prod_{r=0}^{c-p-1} \frac{\partial}{\partial h_{c-r-1}} F(X_{c-r}, h_{c-r-1}) \quad (13)$$

There are many techniques to try to solve the vanishing gradient problem [52,53], but the most important is a specific type of network called Long Short-Term Memory Networks (LSTMs). The LSTM solves the problem by setting the weight initialization to 1 but also adding new components to the traditional RNN architecture: forget gate, input gate, cell state, and output gate. Figure 12 depicts the difference between a normal RNN (Figure 13a) and an LSTM architecture (Figure 12b).

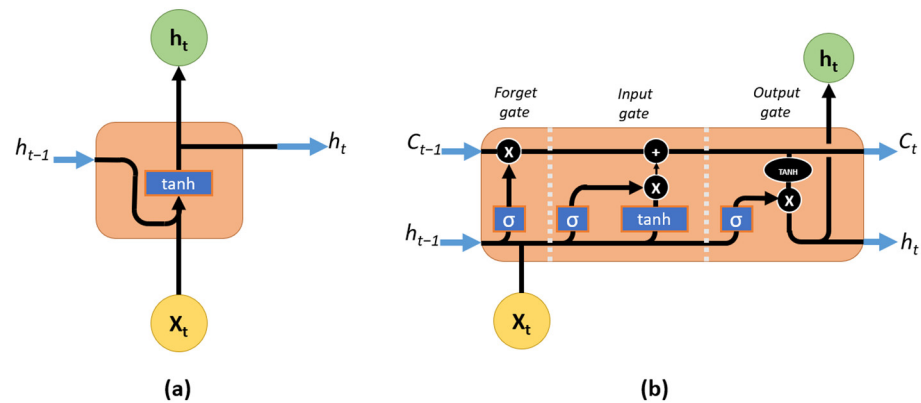


Figure 12. Comparison between RNN and LSTM. (a) RNN diagram and (b) LSTM diagram.

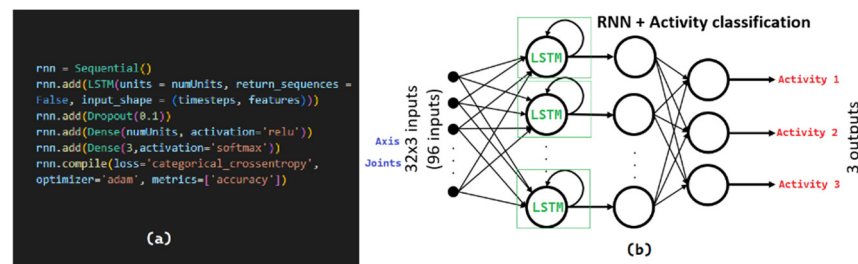


Figure 13. Implemented classification network. (a) Python code; (b) architecture of neural network.

With the forget layer, operated by a sigmoid function, the magnitude of the gradient of LSTM does not decrease, thereby avoiding the gradient problem [54]. The output of the forget layer is between 0 and 1 for each value in the cell state, where a 0 represents to completely forget the value while a 1 represents to totally keep the state as shown in Equation (14):

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (14)$$

The next part handles what information is stored in the cell state by including the input gate layer with a sigmoid function and a tanh function that creates the vector for new \tilde{C}_t candidate values [54]. Therefore, the updated cell state is described by Equation (15):

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (15)$$

Finally, the output gate is a filtered version of the cell state evaluated by a sigmoid function that decides what parts of the cell states are used [54]. Hence, the output h_t is described in Equation (16):

$$h_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) * \tanh(C_t) \quad (16)$$

As for the specific architecture of the LSTM used in this paper, the model is defined as a sequential Keras model with a single hidden layer and a dropout layer of 10% with the goal of reducing overfitting of the model to the training data. A dense, fully connected layer is implemented to interpret the features extracted by the LSTM and its final output layer implements a softmax function to classify the three activities: raising hands, sitting, and walking. The inputs of the network consist in 96 data that represent 3 axis values for each of the 32 joints. Figure 13 depicts the architecture described and the Python code for implementation.

3.3. Study Case: Metabolic Rate Dynamic Analysis Applied on Thermostats

A dataset of daily activities for periods of 15 min during a week's time was obtained from the RNN classification. Thus, 672 observations were obtained with 3 different activities. Then, two energy simulations were performed during the extremely hot week for a household located in Concord, California. The first simulation was the baseline that considered the building, electric loads, and occupation schedules presented in [55] with a fixed value of metabolic rate. This home has two conditioning zones: bedroom two and living room zones. This paper analyzed the living room. The cooling setpoint was 24.4 °C, and the heating setpoint was 21.7 °C, the same initial values considered in [27]. As the extremely hot week was during the summer period, the clo value considered was 0.5 [7]. The second simulation considered the three different activities in the dining and living zone. The energy model was simulated using LadybugTools V1.5.0 software plugin for Grasshopper by Ladybug Tools LLC, USA [56,57].

Then, a strategy to save energy considering thermal comfort was proposed to be compared to the first two simulations. This strategy consisted of increasing or decreasing the cooling and heating setpoints by 1 °C [58,59] or even turning off the thermostat, depending on the following considerations:

1. The difference between outdoor temperature and operative temperature. As the operative temperature tends to match the outdoor temperature, we will call heating tendency when the outdoor temperature is higher than the operative temperature and cooling tendency otherwise.
2. The thermal sensation scale evaluation with the PMV equation [60]. If the thermal sensation at a particular moment is negative, the occupant feeling tends to be cool while a positive value means the occupant feeling tends to be hotter.
3. Four rules are obtained with the combination of the two previous considerations. If the natural tendency is heat and the occupant sensation is negative, the AC is turned off, but if the occupant sensation is positive then the setpoints decrease by 1 °C. Moreover, if the natural tendency is cooling and the occupant sensation is negative, the setpoint is increased by 1 °; on the contrary, if the occupant sensation is positive, then the AC is turned off.

This strategy is evaluated with two more simulations, the first one using the previously described baseline and the second considering the same three activities' recognition of the last simulations. Finally, both results are compared for energy consumption and total comfort state.

Moreover, those activities were converted into W per person because the energy simulation requires that measure. Table 1 depicts these activities, the metabolic rate, and the W/person. The W/person was calculated by multiplying 58.1 W/m² equal to 1 met, and 1.8m² is equal to the skin surface of an average individual of 1.70 m in height and 68 kg [61]. Table 1 depicts these activities, the metabolic rate, and the W/person.

Table 1. Activities considered for the energy simulation.

Activity	Met	W/Person
Desk work	1.8	188
Standing/walking	2.5	261
Cleaning light	2.3	241
Out	0	0

Finally, a comparison between the base model and the dynamic activities model was performed. This comparison included the differences between the total hours of thermal comfort and the total kWh HVAC consumption.

4. Results

This section presents the results of two simulated processes, first the activity recognition depicted in Figure 2 and then the energy saving simulation with the dynamic setpoint for HVAC systems. First, the activity recognition results are shown with the use of an RNN and how a small dataset of activities was created to train the neural network. Then, the evaluation of a simulated model of a house with an HVAC system to obtain an analysis of energy savings between a model with a fixed setpoint and one with an adaptive setpoint is presented.

4.1. Activity Recognition

To show the capability of an RNN to classify activities of daily living (ADL) with the proposed methodology, a small dataset of three activities (sitting, walking, raising arms) was created as most of the available datasets are vision-based (images) or sensor-based as reviewed by [62,63].

The total data gathered for training included 201,600 values as shown in Table 2. This corresponds to 40, 50 and 50 repetitions of each of the three activities to train: sitting, walking, and raising arms. Each repetition consists of 15 timesteps at 2.5 frames per second; and each observation has 96 values corresponding to the x-axis, y-axis, and z-axis values for 32 joints of a 3D skeleton human model. The activities were performed by four different subjects indistinctly with parameters shown in Table 3.

Table 2. Total data for training the RNN.

Activity	Repetitions	Timesteps per Action	Joints per Observation	Axes per Joint	Recorded Data
Sitting	40	15	32	3	57,600
Walking	50	15	32	3	72,000
Raising arms	50	15	32	3	72,000
Total					201,600

Table 3. Training subject's physiology information.

Subject	Gender	Age	Height [cm]	Weight [kg]
Person A	Male	26	1.67	68
Person B	Male	25	1.71	72
Person C	Female	29	1.60	69
Person D	Male	24	1.81	61

Of the total data gathered, the values of five joints were discarded: nose, eye left, eye right, ear left, and ear right as they are not necessary since they do not provide relevant information for the detection of the activity.

Figure 14 depicts the office plan where the training and test data were gathered and the four different positions where the device was located. For the training data, the camera/depth sensor was placed on position 3, while for the testing data the device was placed on the four positions marked to evaluate if the proposed methodology can deal with different view perspectives for classifying the activity. Table 4 shows the position characteristics for each location referenced to the camera/depth sensor. Figure 15 shows the position 2 (a) and position 3 (b) different view perspectives for the testing data.

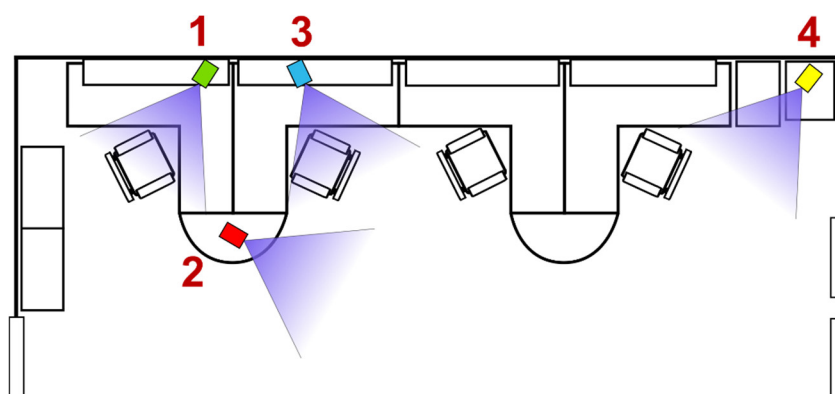


Figure 14. Depth sensor locations blueprint for testing and training inside an office, each number indicates where the depth sensor was located

Table 4. Camera/depth sensor location information.

Location	Height (cm)	Pitch [°]	Roll [°]
1	173	18.9	−1.8
2	87	−2.4	1.2
3	182	20.2	2.8
4	121	6.3	1.3



Figure 15. (a) Camera/depth sensor position 2 view; (b) camera/depth sensor position 3 view.

Moreover, different levels of ambient lightning were used for the testing data. For measuring the light, precision light sensor 1127 was used. Three levels of lightning for each different position of the camera/depth sensor were tested: fully illuminated (513 lux), partially illuminated (235 lux) and dark (4 lux).

The data recorded for evaluating the model in which 15, 17 and 16 repetitions for sitting, walking, and raising arms respectively were recorded are shown in Table 5. The data were recorded in different camera/depth sensor positions (Figure 14), different lighting, with partial occlusions and with three different subjects (Table 6) as depicted in Figure 16.

Table 5. Total data for testing the RNN.

Activity	Repetitions	Timesteps per Action	Joints per Observation	Axis per Joint	Recorded Data
Sitting	15	15	32	3	21,600
Walking	17	15	32	3	24,480
Raising arms	16	15	32	3	23,040
Total					69,120

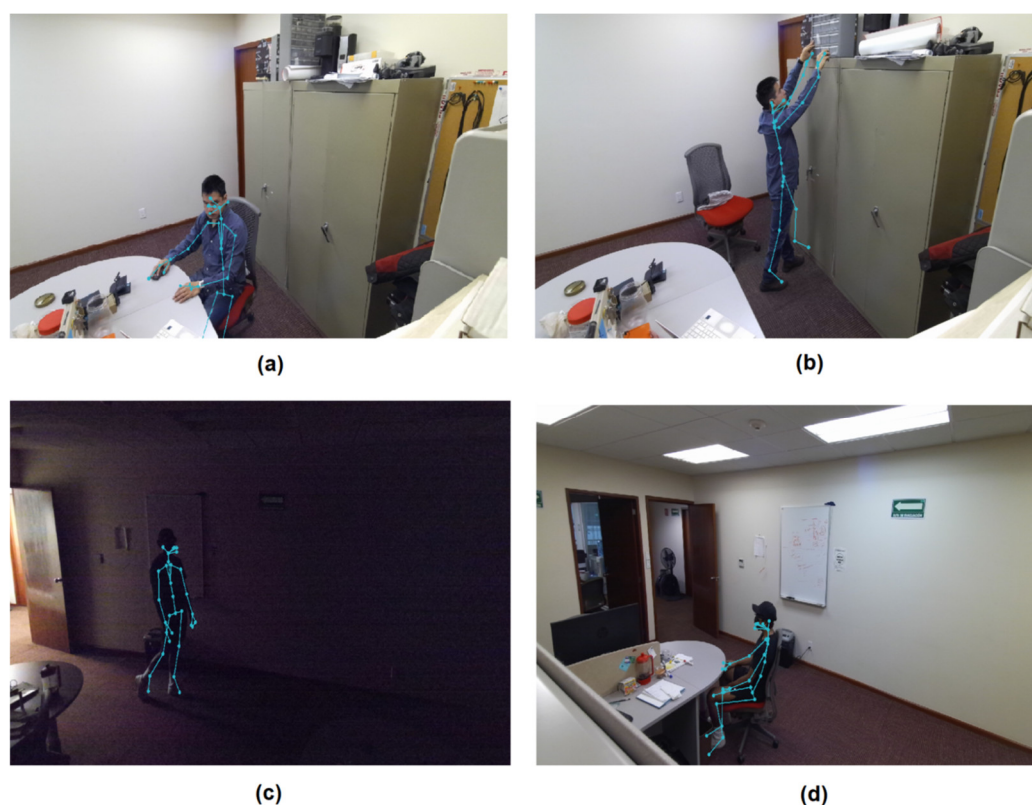


Figure 16. (a) Partial occlusion for sitting on subject 1; (b) raising hands in different positions from the trained data; (c) different lighting; (d) sitting position for subject 2.

Table 6. Test subject's physiology information.

Subject	Gender	Age	Height [m]	Weight [kg]
Person A	Male	22	1.77	72
Person B	Male	44	1.72	80
Person C	Male	34	1.64	65

Because of the stochastic nature of neural networks, different models will result when training with the same data configuration. Therefore, the evaluation of the RNN model was repeated multiple times for a specific number of epochs to be trained and then changed to compare the results. Table 7 summarizes the mean and standard deviation of the performance of the model for 5, 10, 15 and 20 epochs. The mean gives the average accuracy of the model on the dataset, whereas the standard deviation gives the average variance of the accuracy from the mean.

Table 7. Performance of the RNN model.

Epochs	Mean	Standard Deviation
5	93.33%	6.166
10	94.79%	3.125
15	90.21%	6.458
20	91.88%	4.792

After observing the results, the best values correspond to the model for 10 epochs of training. In addition, Figure 17 depicts a confusion matrix showing the performance of the model with the test data.

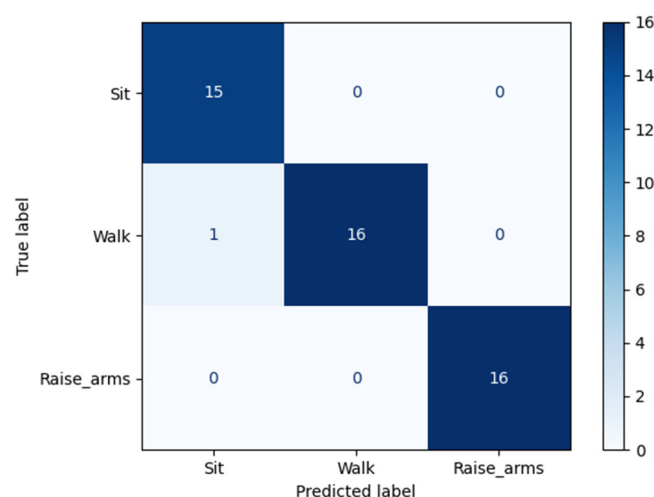


Figure 17. Confusion matrix results.

The results obtained with the trained model show a very high accuracy and validates the methodology proposed for activity recognition where instead of doing image classification we only use 81 signals over 15 timesteps that represent the movement in three axes for the skeleton joints of a human model.

4.2. Energy Savings Simulation

An experiment consisting of four simulations is proposed to evaluate the power consumption; these were made using LadybugTools V1.5.0 by Ladybug Tools LLC, USA. The experiment first consisted of two simulations comparing the estimated energy consumption of a living room with an HVAC system, as described in Section 3.3, for a constant met value set to 1.1 and with variable met values, as described in Table 1, emulating the process of activity recognition, as described in Section 3.1. The simulation is configured to evaluate the parameters every 15 min over a period of 24 h for 7 days (a complete week) but only the time between 7:00 to 21:00 was considered for the results as it is the busiest time for that specific room. The results given by the simulations are:

- Condition: Value between -3 and $+3$ representing the PMV index within the thermal sensation scale.
- Comfort: Binary value that evaluates whether the occupant is comfortable (1) or not (0) with the current environmental and occupant-related variables according to the adaptive thermal comfort model.
- Energy: Energy consumed in kWh.

The results for the first simulation with constant met values and the second simulation with variable met values are listed in Table 8.

Table 8. Results of first and second simulation.

(No Energy Saving Strategy)	Met = 1.1	Variable Met
Condition (average)	−0.7347	0.3367
Comfort (sum)	104	72
Energy (sum)	7.597	7.54

The ideal average of the condition should be 0 as it would represent that for every period of 15 min, the thermal sensation is “normal”. More positive values would represent that the thermal sensation is “hotter” and more negative values would represent a “colder” sensation. For a constant met value, the general sensation would be slightly cold; as for the variable met simulation, the sensation is almost normal with a little tendency to be a bit hot.

The result of the sum of comfort values represents how many periods of fifteen minutes the occupant felt comfortable according to the adaptive thermal comfort model. The higher the value the more comfortable the occupant is. It can be observed that the simulation with constant met values has a higher value.

The sum of energy consumed in kWh is the third observable result. For both simulations the consumption is almost the same with a difference of 0.0567 kWh.

The second part of the experiment consists of two more simulations. This time the proposed strategy for saving energy described in Section 3.3 was applied to the setpoint limits of the thermostat and the other parameters remain the same as for the first two simulations. The results for these new simulations are shown in Table 9.

Table 9. Results of third and fourth simulation.

(With Energy Saving Strategy)	Met = 1.1	Variable Met
Condition (average)	−0.6581	0.3648
Comfort (sum)	134	61
Energy (sum)	5.0557	6.5234

The condition result shows that using a constant met value, the average sensation for the whole week is colder than having a variable met. In comparison with the previous simulations, for the constant met the condition improved as it got closer to zero.

The sum of comfort for a constant met doubled for the variable met. As for comparing with the first simulations, the comfort increased for a constant met but decreased for variable met.

The sum of energy consumption is almost 1.5 kWh less for the constant met simulation than for the variable met simulation. However, in comparison with the first two simulations, both decreased at least 15% with the energy saving strategy proposed.

5. Discussion

This paper focuses on three main aspects to propose a strategy to try to reduce energy consumed by a HVAC system in a building without compromising the thermal comfort of the occupant. The first one considers a dynamic met value that can change according to the activity carried out by the occupant in the calculations of comfort. Moreover, the activity must be detected on-line to let the thermal comfort models update as the occupant-related variables change. Therefore, the way to go is a vision-based system, as deep learning techniques have significantly progressed [46] and offer less intrusive sensing.

The second aspect is using a depth sensor-based system to recognize human activities of daily living to avoid the main challenges an RGB-based classification system faces. With the presented methodology that uses a skeleton model with 3D data of 32 joints to make a classification using a simple LSTM network, it is shown that the recognition of activities can be achieved with high accuracy and with less data for training in comparison with similar public available datasets [46]. Moreover, the manipulation of 3D information allowed the recognition without affecting the position in which the camera was placed, the orientation of the occupant with respect to the camera when carrying out the activity or even the physiological differences of the occupants, as could be demonstrated in the tests carried out and obtaining a high level of accuracy.

The last aspect is the strategy to save energy by increasing or decreasing by 1 °C the heating and cooling setpoints of a connected thermostat. The proposed strategy showed in the simulations that the comfort level for a constant met value is higher than the one for a variable met value, showing that actual models are not giving a real perspective of the occupant's comfort as they are estimating higher values of comfort when in real-life scenarios depending on the activity of the occupant, the comfort values should be lower. The results also showed that the energy consumption decreased by 33% compared to the simulations with constant met value and 14.2% comparing with the variable met values

simulations. As the variable met simulation offers more realistic information it is important to notice that the 14.2% of energy saving comes with a decrease of 11 points in comfort, meaning that in eleven time slots of 15 min of the whole week the occupant felt not comfortable; that is, 165 min less than without using the energy saving strategy. A 14.2% of energy saving for a 1.63% decrease in comfort can be considered an acceptable strategy; moreover, the decrease in comfort can be improved by introducing the capacity of changing occupant's clothes in future work.

The implementation of an on-line estimation of metabolic rate on a connected thermostat opens the possibility to implement energy saving strategies that currently are limited to just the information obtained by the environment sensors allocated in the thermostat. The simulation presented in this paper shows a strategy that reaches 14% of energy saving compared to a strategy that does not include the on-line metabolic rate information, showing the importance of adding the information of all thermal comfort parameters. Furthermore, incorporating a vision-based sensing system allows not just to incorporate the metabolic rate information to the thermal comfort analysis but also the clothing insulation of a person to increase even more the thermal comfort estimations.

6. Conclusions

In this paper, a preprocessing methodology for using 3D data from a depth sensor was proposed. By using the preprocessed data, the classification algorithm that used an LSTM neural network was able to effectively classify three different common activities of daily life to later assign them a MET value. The activity recognition process validates the ability to identify the MET values on-line inside a smart home or smart building. Moreover, the simulation results for energy savings with a variable MET value as part of the comfort model reduced the energy consumption by 14% without significantly affecting the comfort of the occupants. Therefore, it can be concluded that the inclusion of the on-line metabolic rate information offers a more accurate picture of the thermal comfort analysis to propose energy saving strategies based on HVAC systems. Moreover, the proposed strategy showed positive results for saving energy and can be improved by including clothing detection based on the same vision system.

As this paper only considered three activities and a fixed set of rules for the energy saving strategy, our future work can include the increment of the activity database and the investigation of a reinforcement learning (RL) algorithm to improve the energy saving strategy. This strategy could learn to maximize comfort and minimize energy consumption by modifying the connected thermostat setpoint.

Author Contributions: Writing—original draft, O.M., J.I.M. and P.P.; Writing—review and editing, T.P., A.M. (Alan Meier) and A.M. (Arturo Molina) All authors have read and agreed to the published version of the manuscript. All authors will be informed about each step of manuscript processing including submission, revision, revision reminder, etc., via emails from our system or the assigned Assistant Editor. All authors have read and agreed to the published version of the manuscript.

Funding: This research project is supported by Tecnológico de Monterrey and CITRIS under the collaboration ITESM-CITRIS Smart thermostat, deep learning, and gamification project (<https://citriss-uc.org/2019-itesm-seed-funding/> (accessed on 20 September 2022)). Agreement: TECNOLÓGICO DE MONTERREY–CITRIS 2019.

Acknowledgments: The authors would like to acknowledge the financial and the technical support of Tecnológico de Monterrey in the production of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ADL	Activities of Daily Living
AHSRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
ECG	Electrocardiogram
EEG	Electroencephalogram
HAR	Human Activity Recognition
HMM	Hidden Markov Model
HVAC	Heating, Ventilation, and Air Conditioning
IEA	International Energy Agency
KNN	K-nearest neighbors
kWh	kilowatt-hour
LSTM	Long Short-Term Memory Networks
MET	A ratio of the working metabolic rate relative to the resting metabolic rate
PPD	Predicted Percentage of Dissatisfied
PMV	Predicted Mean Vote
RNN	Recurrent Neural Network
RGB	Red, Green, and Blue
RGB-D	Red, Green, Blue and Depth
SDK	Software Development Kit
SVM	Support Vector Machines

References

1. Transition to Sustainable Buildings—Analysis. (n.d.). IEA. Available online: <https://www.iea.org/reports/transition-to-sustainable-buildings> (accessed on 16 August 2022).
2. Poel, B.; van Cruchten, G.; Balaras, C.A. Energy performance assessment of existing dwellings. *Energy Build.* **2007**, *39*, 393–403. <https://doi.org/10.1016/j.enbuild.2006.08.008>.
3. Balaras, C.A.; Grossman, G.; Henning, H.-M.; Infante Ferreira, C.A.; Podesser, E.; Wang, L.; Wiemken, E. Solar air conditioning in Europe—An overview. *Renew. Sustain. Energy Rev.* **2007**, *11*, 299–314. <https://doi.org/10.1016/j.rser.2005.02.003>.
4. Yao, Y.; Chen, J. Global optimization of a central air-conditioning system using decomposition–coordination method. *Energy Build.* **2010**, *42*, 570–583. <https://doi.org/10.1016/j.enbuild.2009.10.027>.
5. El-Dessouky, H.; Ettouney, H.; Al-Zeefari, A. Performance analysis of two-stage evaporative coolers. *Chem. Eng. J.* **2004**, *102*, 255–266. <https://doi.org/10.1016/j.cej.2004.01.036>.
6. Oropeza-Perez, I.; Petzold-Rodriguez, A.H. Analysis of the Energy Use in the Mexican Residential Sector by Using Two Approaches Regarding the Behavior of the Occupants. *Appl. Sci.* **2018**, *8*, 2136. <https://doi.org/10.3390/app8112136>.
7. ASHRAE. (2017). Standard 55-2017 -- Thermal Environmental Conditions for Human Occupancy. *American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc.*
8. Fanger, P.O. Thermal Comfort: analysis and applications in environmental engineering. **1970**. Available online: <https://www.cabdirect.org/cabdirect/abstract/19722700268> (accessed on 8 August 2022).
9. Van Hoof, J.; Mazej, M.; Hensen, J.L.M. Thermal comfort: Research and practice. *Front. Biosci.-Landmark* **2010**, *15*, 765–788. <https://doi.org/10.2741/3645>.
10. De Dear, R.; Brager, G.S. Developing an Adaptive Model of Thermal Comfort and Preference. 1998. Available online: <https://escholarship.org/uc/item/4qq2p9c6> (accessed on 9 August 2022).
11. Arakawa Martins, L.; Soebarto, V.; Williamson, T. A systematic review of personal thermal comfort models. *Build. Environ.* **2022**, *207*, 108502. <https://doi.org/10.1016/j.buildenv.2021.108502>.
12. Li, D.; Menassa, C.C.; Kamat, V.R. Personalized human comfort in indoor building environments under diverse conditioning modes. *Build. Environ.* **2017**, *126*, 304–317. <https://doi.org/10.1016/j.buildenv.2017.10.004>.
13. Karmann, C.; Schiavon, S.; Arens, E. Percentage of Commercial Buildings Showing at least 80% Occupant Satisfied with their Thermal Comfort. 2018. Available online: <https://escholarship.org/uc/item/89m0z34x> (accessed on 20 August 2022).
14. Aryal, A.; Becerik-Gerber, B. Energy consequences of Comfort-driven temperature setpoints in office buildings. *Energy Build.* **2018**, *177*, 33–46. <https://doi.org/10.1016/j.enbuild.2018.08.013>.
15. Huizenga, C.; Abbaszadeh, S.; Zagreus, L.; Arens, E.A. Air Quality and Thermal Comfort in Office Buildings: Results of a Large Indoor Environmental Quality Survey. 2006. Available online: <https://escholarship.org/uc/item/7897g2f8> (accessed on 5 September 2022).
16. Kim, J.; Schiavon, S.; Brager, G. Personal comfort models—A new paradigm in thermal comfort for occupant-centric environmental control. *Build. Environ.* **2018**, *132*, 114–124. <https://doi.org/10.1016/j.buildenv.2018.01.023>.

17. D'Ambrosio Alfano, F.R.; Palella, B.I.; Riccio, G. The role of measurement accuracy on the thermal environment assessment by means of PMV index. *Build. Environ.* **2011**, *46*, 1361–1369. <https://doi.org/10.1016/j.buildenv.2011.01.001>.
18. Ponce, P.; Meier, A.; Mendez, J.; Pfeffer, T.; Molina, A.; Mata, O. Tailored Gamification and Serious Game Framework Based on Fuzzy Logic for Saving Energy in Smart Thermostats. *J. Clean. Prod.* **2020**, *262*, 121167. <https://doi.org/10.1016/j.jclepro.2020.121167>.
19. Avila, M.; Méndez, J.I.; Ponce, P.; Pfeffer, T.; Meier, A.; Molina, A. Energy Management System Based on a Gamified Application for Households. *Energies* **2021**, *14*, 3445. <https://doi.org/10.3390/en14123445>.
20. Meier, A.; Ueno, T.; Pritoni, M. Using Data from Connected Thermostats to Track Large Power Outages in the United States. *Appl. Energy* **2019**, *256*, 113940. <https://doi.org/10.1016/j.apenergy.2019.113940>.
21. Pfeffer, T.; Pritoni, M.; Meier, A.; Aragon, C.; Perry, D. How People Use Thermostats in Homes: A Review. *Build. Environ.* **2011**, *46*, 2529–2541. <https://doi.org/10.1016/j.buildenv.2011.06.002>.
22. Pfeffer, T.; Perry, D.; Pritoni, M.; Aragon, C.; Meier, A. Facilitating Energy Savings with Programmable Thermostats: Evaluation and Guidelines for the Thermostat User Interface. *Ergonomics* **2013**, *56*, 463–479. <https://doi.org/10.1080/00140139.2012.718370>.
23. Ponce, P.; Pfeffer, T.; Molina, A. Usability Perceptions and Beliefs about Smart Thermostats by Chi-Square Test, Signal Detection Theory, and Fuzzy Detection Theory in Regions of Mexico. *Front. Energy* **2018**, *13*, 522–538. <https://doi.org/10.1007/s11708-018-0562-2>.
24. Ponce, P.; Pfeffer, T.; Molina, A. Framework for Evaluating Usability Problems: A Case Study Low-Cost Interfaces for Thermostats. *Int. J. Interact. Des. Manuf.* **2018**, *12*, 439–448. <https://doi.org/10.1007/s12008-017-0392-1>.
25. Méndez, J.I.; Medina, A.; Ponce, P.; Pfeffer, T.; Meier, A.; Molina, A. A real-time adaptive thermal comfort model for sustainable energy in interactive smart homes: Part I. In *Proceedings of the International Conference on Smart Multimedia, Marseille, France, 25–27 August 2022*; Springer: Berlin/Heidelberg, Germany, 2022.
26. Medina, A.; Méndez, J.I.; Ponce, P.; Pfeffer, T.; Meier, A.; Molina, A. A real-time adaptive thermal comfort model for sustainable energy in interactive smart homes: Part II. In *Proceedings of the International Conference on Smart Multimedia, Marseille, France, 25–27 August 2022*; Springer: Berlin/Heidelberg, Germany, 2022.
27. Medina, A.; Méndez, J.I.; Ponce, P.; Pfeffer, T.; Meier, A.; Molina, A. Using Deep Learning in Real-Time for Clothing Classification with Connected Thermostats. *Energies* **2022**, *15*, 1811. <https://doi.org/10.3390/en15051811>.
28. Medina, A.; Méndez, J.I.; Ponce, P.; Pfeffer, T.; Molina, A. Embedded Real-Time Clothing Classifier Using One-Stage Methods for Saving Energy in Thermostats. *Energies* **2022**, *15*, 6117. <https://doi.org/10.3390/en15176117>.
29. Abdullah, M.; Negara, A.; Sayeed, S.; Choi, D.-J.; Sonai, K. Classification Algorithms in Human Activity Recognition using Smartphones. *International Journal of Biomedical and Biological Engineering* **2012**, *6*, 8.
30. Dang, L.M.; Min, K.; Wang, H.; Piran, M.J.; Lee, C.H.; Moon, H. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognit.* **2020**, *108*, 107561.
31. Iglesias, J.; Cano, J.; Bernardos, A.M.; Casar, J.R. A ubiquitous activity-monitor to prevent sedentariness. In *Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, Seattle, WA, USA, 21–25 March 2011; pp. 319–321. <https://doi.org/10.1109/PERCOMW.2011.5766894>.
32. Choujaa, D.; Dulay, N. TRACME: Temporal activity recognition using mobile phone data. In *Proceedings of the 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, Shanghai, China, 17–20 December 2008; Volume 1, pp. 119–126. <https://doi.org/10.1109/EUC.2008.33>.
33. Parkka, J.; Ermes, M.; Korpipaa, P.; Mantyjarvi, J.; Peltola, J.; Korhonen, I. Activity classification using realistic data from wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 119–128. <https://doi.org/10.1109/TITB.2005.856863>.
34. Jatoba, L.C.; Grossmann, U.; Kunze, C.; Ottenbacher, J.; Stork, W. Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. In *Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, Canada, 20–24 August 2008, pp. 5250–5253. <https://doi.org/10.1109/IEMBS.2008.4650398>.
35. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explor. Newsl.* **2011**, *12*, 74–82. <https://doi.org/10.1145/1964897.1964918>.
36. Bayat, A.; Pomplun, M.; Tran, D.A. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Comput. Sci.* **2014**, *34*, 450–457. <https://doi.org/10.1016/j.procs.2014.07.009>.
37. Marron, J.J.; Labrador, M.A.; Menendez-Valle, A.; Fernandez-Lanvin, D.; Gonzalez-Rodriguez, M. Multi sensor system for pedestrian tracking and activity recognition in indoor environments. *Int. J. Ad. Hoc. Ubiquitous Comput.* **2016**, *23*, 3–23. <https://doi.org/10.1504/IJAHUC.2016.078480>.
38. Lara, O.D.; Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **2012**, *15*, 1192–1209.
39. Wilson, D.H.; Atkeson, C. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In *Pervasive Computing*; Gellersen, H.W., Want, R., Schmidt, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 62–79. https://doi.org/10.1007/11428572_5.
40. Ordóñez, Fco. J.; de Toledo, P.; Sanchis, A. Activity Recognition Using Hybrid Generative/Discriminative Models on Home Environments Using Binary Sensors. *Sensors* **2013**, *13*, 5460–5477. <https://doi.org/10.3390/s130505460>.
41. Jobanputra, C.; Bavishi, J.; Doshi, N. Human activity recognition: A survey. *Procedia Comput. Sci.* **2019**, *155*, 698–703.
42. Poppe, R. A survey on vision-based human action recognition. *Image Vis. Comput.* **2010**, *28*, 976–990.

43. Bux, A.; Angelov, P.; Habib, Z. Vision based human activity recognition: A review. *Adv. Comput. Intell. Syst.* **2017**, *513*, 341–371.
44. Yang, B.; Li, X.; Hou, Y.; Meier, A.; Cheng, X.; Choi, J.H.; Wang, F.; Wang, H.; Wagner, A.; Yan, D.; et al. Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-a review. *Energy Build.* **2020**, *224*, 110261.
45. Oyedotun, O.K.; Khashman, A. Deep learning in vision-based static hand gesture recognition. *Neural Comput. Appl.* **2017**, *28*, 3941–3951.
46. Liu, Z.; Zhu, J.; Bu, J.; Chen, C. A survey of human pose estimation: The body parts parsing based methods. *J. Vis. Commun. Image Represent.* **2015**, *32*, 10–19.
47. Chen, Y.; Tian, Y.; He, M. Monocular human pose estimation: A survey of deep learning-based methods. *Comput. Vis. Image Underst.* **2020**, *192*, 102897. <https://doi.org/10.1016/j.cviu.2019.102897>.
48. Goldstein, H. *Classical Mechanics Addison-Wesley Series in Physics*, 2nd ed.; Addison-Wesley: Boston, MA, USA, 1980.
49. Pedley, M. Tilt Sensing Using a Three-Axis Accelerometer. Freescale Semiconductor; AN3461. 2013. Available online: <https://www.nxp.com/docs/en/application-note/AN3461.pdf> (accessed on 3 September 2022).
50. Pio, R. Euler angle transformations. *IEEE Trans. Autom. Control* **1966**, *11*, 707–715. <https://doi.org/10.1109/TAC.1966.1098430>.
51. Hochreiter, S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **1998**, *6*, 107–116.
52. Arjovsky, M.; Shah, A.; Bengio, Y. Unitary evolution recurrent neural networks. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1120–1128.
53. Schmidhuber, J.; Hochreiter, S. Long short-term memory. *Neural. Comput.* **1997**, *9*, 1735–1780.
54. Takeuchi, D.; Yatabe, K.; Koizumi, Y.; Oikawa, Y.; Harada, N. Real-time speech enhancement using equilibrated RNN. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 851–855.
55. Méndez, J.I.; Pfeffer, T.; Ponce, P.; Meier, A.; Molina, A. Empowering Saving Energy at Home through Serious Games on Thermostat Interfaces. *Energy Build.* **2022**, *263*, 112026. <https://doi.org/10.1016/j.enbuild.2022.112026>.
56. Ladybug Tools Ladybug Tools. Home Page. Available online: <https://www.ladybug.tools/> (accessed on 2 May 2021).
57. Associates, R.M. Rhinoceros 3D. Available online: <https://www.rhino3d.com/es/> (accessed on 10 July 2022).
58. Hoyt, T.; Arens, E.; Zhang, H. Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings. *Building and Environment.* **2014**, *88*, 89–96. <https://doi.org/10.1016/j.buildenv.2014.09.010>.
59. Cai, M.; Ramdasalli, S.; Pipattanasomporn, M.; Rahman, S.; Malekpour, A.; Kothandaraman, S.R. Impact of HVAC set point adjustment on energy savings and peak load reductions in buildings. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018; pp. 1–6. <https://doi.org/10.1109/ISC2.2018.8656738>.
60. Ekici, C. Measurement Uncertainty Budget of the PMV Thermal Comfort Equation. *Int. J. Thermophys.* **2016**, *37*, 48. <https://doi.org/10.1007/s10765-015-2011-3>.
61. U.S. Department of Energy. *EnergyPlus™ Version 9.5.0 Documentation: Input Output Reference*; U.S. Department of Energy: Washington, DC, USA, 2021.
62. Sharma, V.; Gupta, M.; Pandey, A.K.; Mishra, D.; Kumar, A. A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets. *Appl. Artif. Intell.* **2022**, *36*, 2093705. <https://doi.org/10.1080/08839514.2022.2093705>.
63. Ariza-Colpas, P.P.; Vicario, E.; Oviedo-Carrascal, A.I.; Butt Aziz, S.; Piñeres-Melo, M.A.; Quintero-Linero, A.; Patara, F. Human Activity Recognition Data Analysis: History, Evolutions, and New Trends. *Sensors* **2022**, *22*, 3401. <https://doi.org/10.3390/s22093401>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.