

ORIGINAL ARTICLE OPEN ACCESS

# Skeleton-Based Posture Recognition for Home Care From Virtual Unmanned Aerial Vehicle

Andrés Bustamante<sup>1</sup> | Lidia M. Belmonte<sup>1,2</sup> | António Pereira<sup>3,4</sup> | Rafael Morales<sup>1,2</sup> | Antonio Fernández-Caballero<sup>1,5,6</sup>

<sup>1</sup>Instituto de Investigación en Informática de Albacete, Universidad de Castilla-La Mancha, Albacete, Spain | <sup>2</sup>Departamento de Ingeniería Eléctrica, Electrónica, Automática y Comunicaciones, Universidad de Castilla-La Mancha, Albacete, Spain | <sup>3</sup>Computer Science and Communications Research Centre, Polytechnic Institute of Leiria, Leiria, Portugal | <sup>4</sup>Institute of New Technologies—Leiria Office, INOV INESC INOVAÇÃO, Leiria, Portugal | <sup>5</sup>Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete, Spain | <sup>6</sup>Biomedical Research Networking Centre in Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

**Correspondence:** Antonio Fernández-Caballero ([antonio.fdez@uclm.es](mailto:antonio.fdez@uclm.es))**Received:** 12 December 2024 | **Revised:** 9 June 2025 | **Accepted:** 17 July 2025

**Funding:** This work was supported by the grant 2022-GRIN-34436 funded by Universidad de Castilla-La Mancha and by “ERDF A way of making Europe,” by Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) (through Red [225RT0169]), by Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III, MICIU, and by Junta de Comunidades de Castilla-La Mancha/European Science Foundation (grant number SBPLY/24/180225/000225). This work was also partially supported by National Funds through the Portuguese funding agency, FCT-Fundação para a Ciência e a Tecnologia, within project UIDB/04524/2020 and the project PTDC/CCI-COM/2416/2021, and by Portuguese National funds through FITEC-Programa Interface, with reference CIT “INOV-INESC Inovação-Financiamento Base.”

**Keywords:** heuristics | home care | posture recognition | skeleton | unmanned aerial vehicle | virtual reality

## ABSTRACT

This article presents a novel approach for real-time posture recognition in monitoring scenarios, utilising a virtual camera simulated on a UAV within virtual environments. Leveraging the MediaPipe Pose library, key points of the body skeleton are extracted, focusing on a subset of 8 key points for computational efficiency. Through the integration of heuristic algorithms based on physical proportions of the human body, the proposed methodology provides accurate estimations of three distinct postures: lying, standing, and sitting. This heuristic-based approach offers a computationally efficient alternative to traditional machine learning and deep learning methods, ensuring real-time performance and scalability. The efficiency of the framework is demonstrated through experiments that show its potential applications in various fields, including healthcare, virtual reality, and human-computer interaction. This approach achieved an average precision of 98.08% for virtual images. Success rates were 100%, 95.8%, and 98.9% for standing, sitting, and lying postures, respectively. Furthermore, the original classification model, which was tuned for virtual images, was tested on real images without any alteration to the parameter values. Its good performance demonstrates its potential for generalisation and application in diverse environments. Overall, this work contributes to the advancement of posture recognition technology, offering a versatile and accessible solution for posture analysis in dynamic monitoring environments.

## 1 | Introduction

Human posture recognition has been a focus of research in computer vision and artificial intelligence for the past few decades (Kaur et al. 2024). Posture recognition involves analysing the original information from the target human or

humans, captured by a sensor device or camera, through a series of algorithms to obtain posture (Dentamaro et al. 2024; Hassan et al. 2024; Jiang et al. 2023; Mim et al. 2023). Body posture recognition is particularly beneficial in facilities that care for the elderly, who are more prone to accidents due to disability or age-related deterioration (Carretero et al. 2007).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Expert Systems* published by John Wiley & Sons Ltd.

They also make work easier by keeping facilities under constant surveillance without or with rather complex assistive robots (Sanchez-Comas et al. 2020).

Monitoring the elderly is essential for optimal and rapid home assistance. Home care by unmanned aerial vehicles (UAVs) is becoming a hot topic due to their ability to monitor dependent people indoors (Belmonte, García, Morales, et al. 2021; Martínez et al. 2021; Belmonte, García, Segura, et al. 2021; Belmonte, Morales, and Fernández-Caballero 2019) and provide a warning or alarm as soon as possible based on the relevant events detected. Unlike static vision systems or other ground robots, a UAV equipped with an on-board camera can overcome potential barriers to the home, such as stairs, as well as access remote points, avoid dead angles, position itself in front of the person, and accompany the person in real time (Belmonte, García, Segura, et al. 2021; Wan et al. 2021). In addition, indoor UAVs offer innovative approaches to improve safety and mobility through improved trajectory planning (Belmonte, Morales, García, et al. 2019).

Virtual reality (VR) is a complementary technology that is useful for focusing on human factors in the development and deployment of home care UAVs. It is necessary to test different technical solutions before implementing them in real homes. VR technology allows testing in controlled environments before building a physical device. In fact, VR has traditionally been used to simulate environments due to its ability to reduce time and cost (Seth et al. 2011), eliminate risk to people and equipment (Roberts et al. 2015; Li et al. 2018), and train (Bric et al. 2016). This is particularly true for robotics (Zhang et al. 2018; Rohmer et al. 2013), as it allows rapid prototyping and algorithm development, as well as system verification, among other benefits. Although current UAV technology makes them safer by allowing them to move outdoors automatically detecting and avoiding obstacles, indoor experiments are better done in VR for safety reasons (Belmonte, García, Segura, et al. 2021; Belmonte, García, Morales, et al. 2021).

Traditional posture recognition methods use classical image segmentation algorithms to segment an image or video. We believe that posture estimation based on heuristic, probabilistic, and trigonometric analysis is mathematically simpler to implement and less computationally expensive than using more advanced algorithms or classifiers (Newell et al. 2016; Belagiannis and Zisserman 2017). Heuristics are renowned for their simplicity and speed. They offer quick-to-implement solutions that provide immediate results. Furthermore, heuristics do not require large datasets or substantial computational power, unlike machine learning models. Another advantage is that heuristics are easy to understand because they are explainable. However, their simplicity may result in solutions that are not always the most accurate or optimal.

In this context, skeletons have become a widely used data modality for posture recognition, where skeletal data are used to construct high-level features that characterise 3D configurations of postures (Liu et al. 2020). In fact, the extraction of human skeletal landmarks from two-dimensional (2D) or depth images through posture estimation is a common basis for vision-based posture recognition. Thus, there are several ways to detect

the human skeleton. One way is to use dedicated devices such as the Microsoft Kinect (Moon et al. 2016; Schwarz et al. 2012), which use depth sensors and cameras that, together with its own algorithm, can generate a virtual skeleton of the person(s) in its field of view. The alternative to not using depth devices is to use colour images to recognise human bodies (Rudol and Doherty 2008) and to estimate the actions a person is performing (Martínez et al. 2021; Penmetsa et al. 2014). This second option, which is the one chosen in this study, allows the use of software libraries that are not specifically designed for a particular device but can be integrated into a multitude of vision sensors. This also means that there is no dependence on a specific price range or limitation in terms of the environmental conditions.

The aim of this research is to estimate a person's posture using virtual colour images captured by a virtual UAV camera. Considering the important issue of caring for the aged adult, the postures considered in this article are when the person is *standing*, *sitting*, or *lying*. This last posture could, for example, indicate a fall, which is very important for people who live alone. The contributions of this work to the state of the art are the following:

1. Novel approach using MediaPipe and Unity. This work presents a novel methodology for posture recognition using MediaPipe and Unity. Using landmark data from MediaPipe, coupled with geometric figure generation in Unity, this study pioneers a unique fusion of these technologies for posture classification.
2. Heuristic-based posture inference. Unlike existing approaches that rely heavily on machine learning and deep learning algorithms, this article proposes a heuristic-based method for posture detection. By analysing the variations in width, height and aspect ratio of geometric figures representing different body segments, this study offers an alternative, computationally efficient solution for posture recognition.

This article is structured as follows. Section 2 presents the posture detection solution developed, detailing the set of heuristic rules. The results of the proposal are presented and discussed in Section 3, where a comparison is made with other related work. Lastly, Section 4 offers the most relevant conclusions of this study.

## 2 | Materials and Methods

### 2.1 | General Description

As described before, the aim of this work is to detect three postures in virtual colour images as captured by a virtual home care UAV. The three postures, *Standing*, *Sitting*, and *Lying*, are shown in virtual environments such as the one displayed in Figure 1.

The 3D environments were created using the Unity game engine (<https://www.unity.com>). Unity is an engine that facilitates the creation of virtual environments for experimentation and testing. Among a wide range of possibilities, some research with UAVs (Unity Simulation UAV) (Belmonte, García, Segura,

et al. 2021; Hentati et al. 2018) was developed using this graphical engine.

According to Figure 2, Unity was used to create virtual environments and Python to process images sent from Unity via the Message Queuing Telemetry Transport (MQTT) protocol (<https://mqtt.org/>). The MQTT Unity-to-Python transmission method offers several advantages for VR applications, including multi-threading, communication, processing, value return flow, and real-time posture detection (Bustamante et al. 2022). The scenario includes the interior of a flat where an elderly person lives. There is also a UAV containing the camera and the texture renderer, which is used to capture the image of the scene associated with the camera and send it via MQTT to be processed. In addition, as shown in Figure 2, posture estimation is based on selected skeletal landmarks to obtain the three postures sought. The following sections describe the different steps of the methodology used, including Google's MediaPipe Pose (<https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md>) library for image processing from Python, as well as the heuristics used to estimate the postures of the human avatar based on a percentage value.



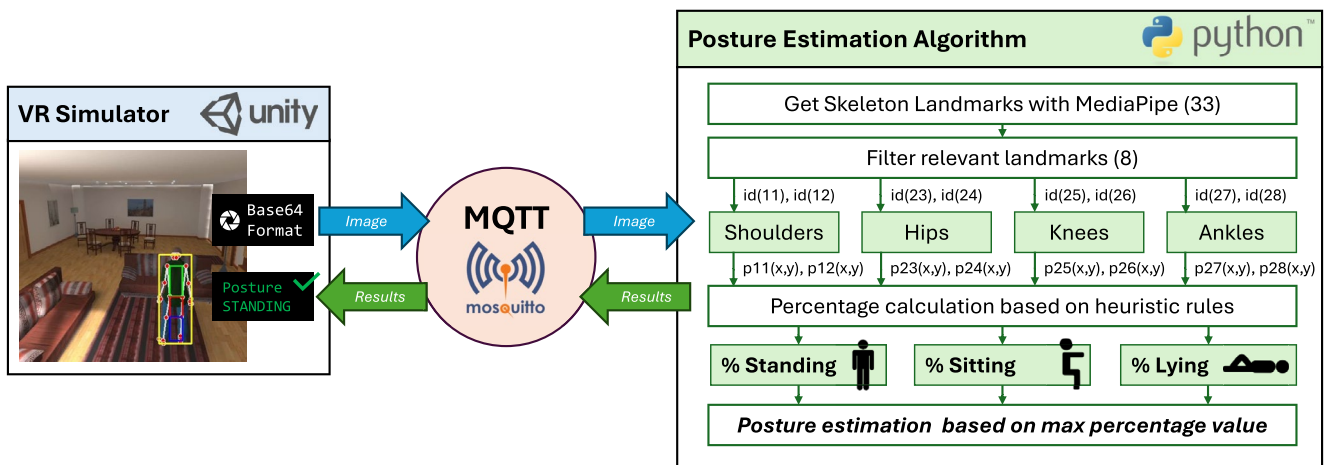
**FIGURE 1** | The three postures that are to be identified (Standing, Sitting, and Lying).

## 2.2 | Body Skeleton Detection

Image processing is the basis for posture recognition and can directly affect the extraction of feature points and the result of posture classification, thus impacting the recognition rate of posture. The usual main tasks in the processing stage are noise reduction, human skeleton keypoint detection, scaling, grey level normalisation, and image segmentation. Body skeleton keypoint detection mainly detects keypoint information such as human joints (Li et al. 2024) and facial features (Zhang et al. 2024). The output is the skeletal feature of the human body, which is the primary part of posture recognition and behaviour analysis, mainly used for segmentation and alignment.

As image processing for posture estimation is highly dependent on image quality, distance to the object, viewing angle, and lighting, this work has used the MediaPipe Pose solution to take advantage of precise robust bipedal skeleton detection and real-time performance on mobile phones and PCs through CPU-based inference (Kim et al. 2023). This ensures validation and estimation of postures with the lowest possible margin of error. The image processing with MediaPipe is done by an echo program in Python that is independent of the echo simulator in Unity. First, the simulator is programmed to take screenshots of the Render Texture component in the UAV's camera every 0.6 s. The captured image is then streamed in base64 format over the MQTT protocol to the Python program. Python receives the image and processes it using OpenCV (<https://opencv.org>). OpenCV decodes the base64 image and creates an object ready to be processed with MediaPipe.

As mentioned before, MediaPipe Pose will provide the required keypoints in a robust manner. In fact, MediaPipe Pose is a machine learning solution for high-fidelity posture tracking. It provides solutions for different types of applications, such as face detection (Lugaresi et al. 2019), hand bone detection and tracking (Halder and Tayade 2021), and any type of tracking related to a human body and its parts. In this work, we use the MediaPipe algorithm to obtain the 33 reference points that make up the entire human skeleton. However, we only use eight landmarks. Eight of them are used to generate evaluation parameters and angles. These in turn estimate the avatar's current posture using

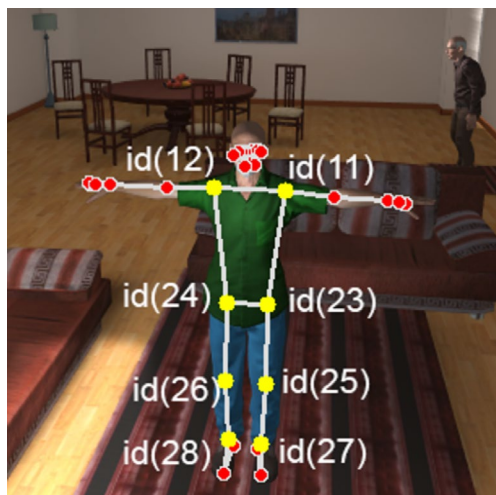


**FIGURE 2** | Block diagram of the implemented solution.

heuristics that provide posture percentages. The landmarks used are shown as yellow dots in the human avatar in Figure 3.

Figure 4 shows recognition examples for the three different postures studied. When keypoints are detected, the maximum and minimum  $x$  and  $y$  values of the keypoints are used to define a rectangular bounding box around the body (yellow rectangle). Note that MediaPipe processes and generates a humanoid skeleton of the first person detected in an image, no matter how many objects or obstacles there are. This is an advantage, as other algorithms may generate false positives (Dalal et al. 2006), which would require additional processing to optimise the results and minimise errors (Pietraszek 2004). The performance of the MediaPipe Pose solution is remarkable and no longer requires complex systems or multiple cameras to achieve similar results (Puwein et al. 2015).

For the eight selected body keypoints (corresponding to the left and right shoulders, hips, knees, and ankles), their  $(x, y)$  coordinates are extracted from the landmarks to obtain three bounding boxes covering the upper body (green rectangle), middle body (red rectangle) and lower body (blue rectangle) based on the height and width of the person (see Figure 4).



**FIGURE 3** | Landmarks highlighting the joints used in this proposal: Left shoulder = id (11); right shoulder = id (12); left hip = id(23); right hip = id (24); left knee = id (25); right knee = id (26); left ankle = id (27); right ankle = id (28).

## 2.3 | Body Posture Detection

Traditional machine learning-based recognition methods mainly describe and infer human posture based on human body models and extract image posture features through algorithms that have high requirements on feature representation and spatial position relationship of keypoints. In general, our approach represents a novel application of a few body landmarks and heuristics for posture detection. The potential future applications of this approach are numerous, including monitoring physical activity levels, detecting falls, or assessing sleep quality.

Once the bounding boxes have been generated in the previous stage, we use the height and width of the four generated bounding boxes as basic parameters to obtain the posture. The aspect of the four rectangles varies in different ways, so heuristics based on the differences in widths and heights will provide fast and accurate results for posture classification. It is important to note that these heuristics use predetermined thresholds and conditions based on the body characteristics of *Standing*, *Sitting*, and *Lying* poses. More specifically, the heuristics are based on the well-studied body proportions of Leonardo da Vinci's Vitruvian Man.

Finally, the percentage of a person in the *Standing*, *Sitting*, and *Lying* poses is obtained. To do this, the values obtained by the heuristics are normalised, and the winning pose is determined based on the highest normalised value. This process provides a hierarchical analysis of the poses and helps to identify the most likely posture given the input body measurements. The percentages of each posture are then plotted using the Plotly library (see Figure 5).

### 2.3.1 | Parameters for Heuristics

To estimate the avatar's posture, we use different heuristics for the three postures. Although the logic within these heuristics is different, they use the same parameters, mentioned above and described in Table 1.

In addition, aspect ratios need to be calculated.

$$\text{Aspect ratio of upper body: } AR_{ub} = \frac{w_{ub}}{h_{ub}} \quad (1)$$



**FIGURE 4** | Examples of bounding boxes for the three postures studied (*Standing*, *Sitting*, and *Lying*).

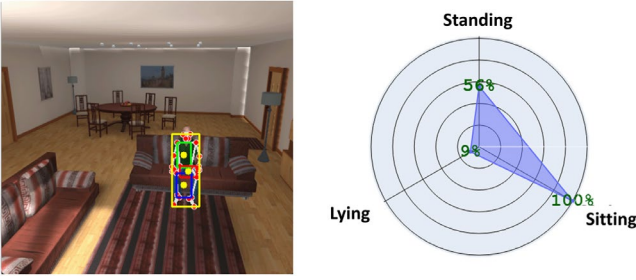


FIGURE 5 | Posture percentages graph.

TABLE 1 | Basic parameters used to detect posture.

Feature	Nomenclature
Height of upper body bounding box	$h_{ub}$
Height of middle body bounding box	$h_{mb}$
Height of lower body bounding box	$h_{lb}$
Width of upper body bounding box	$w_{ub}$
Width of middle body bounding box	$w_{mb}$
Width of lower body bounding box	$w_{lb}$
Centre of the upper body bounding box	$c_{ub}(x, y)$
Centre of the middle body bounding box	$c_{mb}(x, y)$
Centre of the lower body bounding box	$c_{lb}(x, y)$

$$\text{Aspect ratio of middle body: } AR_{mb} = \frac{w_{mb}}{h_{mb}} \quad (2)$$

$$\text{Aspect ratio of lower body: } AR_{lb} = \frac{w_{lb}}{h_{lb}} \quad (3)$$

$$\text{Aspect ratio of whole body: } AR_b = \frac{w_{ub} + w_{mb} + w_{lb}}{h_{ub} + h_{mb} + h_{lb}} \quad (4)$$

The angle between the three centre points,  $c_{ub}(x_1, y_1)$ ,  $c_{ub}(x_2, y_2)$  and  $c_{ub}(x_3, y_3)$  is also calculated as

$$\theta_b = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) - \arctan\left(\frac{y_3 - y_1}{x_3 - x_1}\right) \quad (5)$$

It is also necessary to obtain the differences in height between the upper and middle body, and the middle and lower body (in %). This makes it possible not to depend on the distance of the avatar in the input image. If the avatar is closer to the UAV camera, the bounding boxes will be larger and vice versa. However, the percentages do not vary with distance and therefore more accurate heuristics can be used.

$$\text{Height difference between upper and middle body: } \Delta h_{ub,mb} = (|h_{ub} - h_{mb}| \times 100) / \max(h_{ub}, h_{mb}) \quad (6)$$

$$\text{Height difference between middle and lower body: } \Delta h_{mb,lb} = (|h_{mb} - h_{lb}| \times 100) / \max(h_{mb}, h_{lb}) \quad (7)$$

The values provided by the heuristics for the *Standing*, *Sitting* and *Lying* postures are called  $v^{sta}$ ,  $v^{sit}$  and  $v^{ly}$ , respectively. The heuristics have been obtained through a process of experimentation. An increment constant  $\Delta v$  will be added for each rule that is fulfilled for the *Standing*, *Sitting* and *Lying* postures.

### 2.3.2 | Heuristics for the *Standing* Pose

This section describes the estimation of a person's *Standing* position based on the following heuristics. First, a full body aspect ratio ( $AR_b$ ) check is performed. If the height of the bounding box is much greater than its width, it is assumed that the person in avatar form is *Standing*. In this case,  $v^{sta}$  is given a value that cannot be reached by heuristics.

$$v^{sta} = \begin{cases} 10 \times \Delta v, & \text{if } AR_b \leq AR_b^{sta} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Thus, the rest of the heuristics are applied only if  $v^{sta} = 0$ . In the first place, we have the heuristics that relate to the aspect ratio, the height, and the width.

$$v^{sta} = \begin{cases} v^{sta} + \Delta v, & \text{if } AR_{b_{min}}^{sta} < AR_b \leq AR_{b_{max}}^{sta} \\ v^{sta} + \Delta v, & \text{if } AR_{ub_{min}}^{sta} < AR_{ub} \leq AR_{ub_{max}}^{sta} \\ v^{sta} + \Delta v, & \text{if } AR_{mb_{min}}^{sta} < AR_{mb} \leq AR_{mb_{max}}^{sta} \\ v^{sta} + \Delta v, & \text{if } h_{ub} > h_{mb}/2 \\ v^{sta} + \Delta v, & \text{if } h_{ub} > h_{lb} \\ v^{sta} + \Delta v, & \text{if } h_{mb} \geq h_{lb} \\ v^{sta} + 2 \times \Delta v, & \text{if } h_{mb} - h_{lb} < 6 \\ v^{sta} + \Delta v, & \text{if } w_{ub} < w_{mb}/2 \\ v^{sta} + \Delta v, & \text{if } w_{ub} < w_{lb} \\ v^{sta} + \Delta v, & \text{if } w_{mb} < w_{lb} \end{cases} \quad (9)$$

Now, the angle  $\theta_b$  is explored as

$$v^{sta} = v^{sta} + \Delta v, \text{ if } \theta_{b_{min}}^{sta} \leq \theta_b \leq \theta_{b_{max}}^{sta} \quad (10)$$

Then comes the heuristics based on the percentage of difference in height.

$$v^{sta} = \begin{cases} v^{sta} + 2 \times \Delta v, & \text{if } \Delta h_{mb,lb}^{sta} < \Delta h_{mb,lb} < \Delta h_{mb,lb_{max}}^{sta} \\ v^{sta} + 1.5 \times \Delta v, & \text{if } \Delta h_{ub,mb} \leq \Delta h_{ub,mb_{min}}^{sta} \end{cases} \quad (11)$$

Finally,  $v^{sta}$  is reinforced if the previous posture was also *Standing*.

$$v^{sta} = v^{sta} + 0.25 \times \Delta v, \text{ if PreviousPose} = \textit{Standing} \quad (12)$$

### 2.3.3 | Heuristics for the *Sitting* Pose

This section describes how to estimate a person's *Sitting* posture using the following heuristics. First, the full body aspect ratio ( $AR_b$ ) is checked. If the height of the bounding box is much smaller than its width, it is assumed that the avatar has a high

probability of being *Sitting*. In this case,  $v^{sit}$  is given a value that is not easily reached by heuristics.

$$v^{sit} = \begin{cases} 5 \times \Delta v, & \text{if } AR_b \geq AR_b^{sta} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

The rest of the heuristics are only applied if  $v^{sit} = 0$ . The explanation of these heuristics is similar to that of the *Standing* heuristics.

$$v^{sit} = \begin{cases} v^{sit} + \Delta v, & \text{if } AR_{b_{min}}^{sit} < AR_b \leq AR_{b_{max}}^{sit} \\ v^{sit} + \Delta v, & \text{if } AR_{ub_{min}}^{sit} < AR_{ub} \leq AR_{ub_{max}}^{sit} \\ v^{sit} + \Delta v, & \text{if } AR_{mb_{min}}^{sit} < AR_{mb} \leq AR_{mb_{max}}^{sit} \\ v^{sit} + \Delta v, & \text{if } h_{ub} > h_{mb} \\ v^{sit} + \Delta v, & \text{if } h_{ub} > h_{lb} \\ v^{sit} + \Delta v, & \text{if } h_{mb} \leq h_{lb} \\ v^{sit} + \Delta v, & \text{if } w_{ub} < w_{mb} \\ v^{sit} + \Delta v, & \text{if } w_{ub} < w_{lb} \\ v^{sit} + \Delta v, & \text{if } w_{mb} > w_{lb} \end{cases} \quad (14)$$

$$v^{sit} = v^{sit} + \Delta v, \text{ if } \theta_{b_{min}}^{sit} \leq \theta_b \leq \theta_{b_{max}}^{sit} \quad (15)$$

$$v^{sit} = \begin{cases} v^{sit} + \Delta v, & \text{if } \Delta h_{mb,lb_{min}}^{sit} < \Delta h_{mb,lb} < \Delta h_{mb,lb_{max}}^{sit} \\ v^{sit} + 1.5 \times \Delta v, & \text{if } \Delta h_{ub,mb_{min}}^{sit} < \Delta h_{ub,mb} \leq \Delta h_{ub,mb_{max}}^{sit} \end{cases} \quad (16)$$

$$v^{sit} = v^{sit} + 0.25 \times \Delta v, \text{ if PreviousPose} = \textit{Sitting} \quad (17)$$

### 2.3.4 | Heuristics for the *Lying* Pose

This section describes how to estimate a person's *Lying* posture using the following heuristics. First, the full body aspect ratio ( $AR_b$ ) is checked. If the height of the bounding box is much smaller than its width, it is assumed that the avatar has a high probability of *Lying* (as in the case of *Sitting*). In this way  $v^{lyi}$  is given a value that is not easily reached by the heuristics.

$$v^{lyi} = \begin{cases} 5 \times \Delta v, & \text{if } AR_b \geq AR_b^{sta} \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

The rest of the heuristics are only applied if  $v^{lyi} = 0$ . Again, the explanation of these heuristics is similar to that of the *Standing* heuristics.

$$v^{lyi} = \begin{cases} v^{lyi} + \Delta v, & \text{if } AR_{b_{min}}^{lyi} < AR_b \leq AR_{b_{max}}^{lyi} \\ v^{lyi} + \Delta v, & \text{if } AR_{mb_{min}}^{lyi} < AR_{mb} \leq AR_{mb_{max}}^{lyi} \\ v^{lyi} + 3 \times \Delta v, & \text{if } (h_{ub} > h_{mb})(h_{ub} < h_{lb})(h_{mb} < h_{lb}) \\ v^{lyi} + 3 \times \Delta v, & \text{if } (w_{ub} < w_{mb})(w_{ub} < w_{lb})(w_{mb} > w_{lb}) \end{cases} \quad (19)$$

$$v^{lyi} = \begin{cases} v^{lyi} + 2.5 \times \Delta v, & \text{if } \Delta h_{mb,lb} \geq \Delta h_{mb,lb_{min}}^{lyi} \\ v^{lyi} + 1.5 \times \Delta v, & \text{if } \Delta h_{ub,mb} > \Delta h_{ub,mb_{min}}^{lyi} \end{cases} \quad (20)$$

$$v^{lyi} = v^{lyi} + 0.25 \times \Delta v, \text{ if PreviousPose} = \textit{Lying} \quad (21)$$

### 2.3.5 | Normalisation and Posture Classification

The final stage of posture detection involves normalising the values obtained in the previous section in relation to the heuristics for each type of posture. Normalisation guarantees that at least one posture reaches a probability 100%.

$$\begin{cases} \overline{v^{sta}} = \frac{v^{sta} \times 100}{\max(v^{sta}, v^{sit}, v^{lyi})} \\ \overline{v^{sit}} = \frac{v^{sit} \times 100}{\max(v^{sta}, v^{sit}, v^{lyi})} \\ \overline{v^{lyi}} = \frac{v^{lyi} \times 100}{\max(v^{sta}, v^{sit}, v^{lyi})} \end{cases} \quad (22)$$

Finally, the winning posture is the one with the highest normalised score.

$$\text{CurrentPose} = \begin{cases} \textit{Standing}, & \text{if } (\overline{v^{sta}} > \overline{v^{sit}})(\overline{v^{sta}} > \overline{v^{lyi}}) \\ \textit{Sitting}, & \text{if } (\overline{v^{sit}} > \overline{v^{sta}})(\overline{v^{sit}} > \overline{v^{lyi}}) \\ \textit{Lying}, & \text{if } (\overline{v^{lyi}} > \overline{v^{sta}})(\overline{v^{lyi}} > \overline{v^{sit}}) \end{cases} \quad (23)$$

$$\text{PreviousPose} = \text{CurrentPose} \quad (24)$$

## 3 | Results

### 3.1 | Data

To evaluate the proposed heuristic model for detecting the three postures—*Standing*, *Sitting*, and *Lying*—three distinct video sequences were generated, each corresponding to one posture and recorded at a frame rate of five frames per second. During these sequences, the spatial positioning and orientation of the animated characters were periodically adjusted. A total of 1010 images were analysed, distributed as follows: 395 images for *Standing*, 432 for *Sitting*, and 183 for *Lying*. The sequences were processed using Python and MediaPipe, and the probability of each posture was calculated based on the described heuristics.

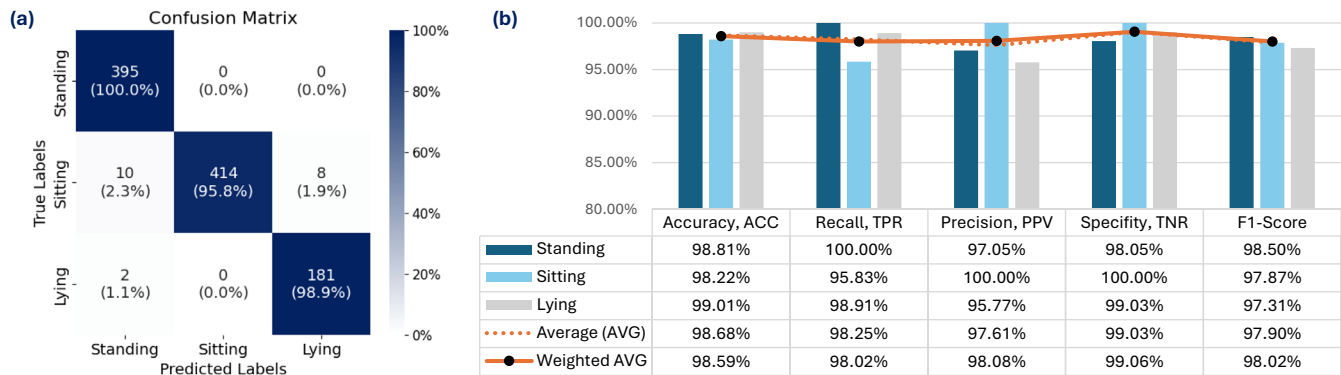
Due to the high reliability of MediaPipe in posture monitoring (Kumar et al. 2025), it was used to determine the optimal parameters and values needed to detect the studied body postures, utilising the properties of the aspect ratio. This approach guarantees adaptability across different environments. The variable values used in the heuristics were obtained through experimentation and are detailed in Table 2. The findings are discussed below.

### 3.2 | Results Per Posture

A confusion matrix and a table summarising key metrics were created to analyse the performance of the classification model based on the heuristic rules described above (see Figure 6). These metrics include accuracy (ACC), recall (TPR), precision

**TABLE 2** | Values obtained for the variables of the heuristics.

Standing		Sitting		Lying	
$AR_b^{sta} = 0.2$		$AR_b^{sta} = 1.0$			
$AR_{b_{min}}^{sta} = 0.08$	$AR_{b_{max}}^{sta} = 0.36$	$AR_{b_{min}}^{sit} = 0.17$	$AR_{b_{max}}^{sit} = 0.48$		
$AR_{ub_{min}}^{sta} = 0.2$	$AR_{ub_{max}}^{sta} = 0.7$	$AR_{ub_{min}}^{sit} = 0.9$	$AR_{ub_{max}}^{sit} = 5.0$	$AR_{ub_{min}}^{lyi} = 2.6$	$AR_{ub_{max}}^{lyi} = 48.0$
$\theta_{b_{min}}^{sta} = -3$	$\theta_{b_{min}}^{sta} = 2$	$\theta_{b_{min}}^{sit} = -13$	$\theta_{b_{min}}^{sta} = 8$		
$\Delta h_{mb,lb_{min}}^{sta} = 12$	$\Delta h_{mb,lb_{max}}^{sta} = 30$	$\Delta h_{mb,lb_{min}}^{sit} = 30$	$\Delta h_{mb,lb_{max}}^{sit} = 71$	$\Delta h_{mb,lb_{min}}^{lyi} = 71$	
	$\Delta h_{ub,mb_{max}}^{sta} = 40$	$\Delta h_{ub,mb_{min}}^{sit} = 40$	$\Delta h_{ub,mb_{max}}^{sit} = 76$	$\Delta h_{ub,mb_{min}}^{lyi} = 76$	

**FIGURE 6** | Results for virtual images: (a) Confusion matrix and (b) metrics per posture (*Standing*, *Sitting*, and *Lying*).

(PPV), specificity (TNR) and the F1-score, representing the harmonic mean of precision and recall. Each parameter was calculated individually for each class (posture) and globally using the arithmetic and weighted averages, taking into account the number of images in each class.

The following conclusions can be drawn from analysing the confusion matrix. The *Standing* posture was recognised with 100% success, with no confusion errors involving the *Sitting* or *Lying* postures. Classification of the *Sitting* posture also demonstrated a high success rate of 95.8%, though a few misclassifications occurred. Specifically, the heuristic confused *Sitting* with *Standing* (2.3%) and *Lying* (1.9%). Finally, the *Lying* posture achieved a success rate of 98.9%, with minor confusion involving the *Standing* posture (1.1%). Some misclassifications arose due to the avatar's movements, including frequent rotations and limb crossings. Consequently, MediaPipe landmarks generated a slightly distorted human skeleton, causing the heuristic to misidentify the character or assign an incorrect posture.

In terms of model metrics, the results demonstrate exceptional performance, with all weighted average values surpassing 98%. Specifically, the model achieves an accuracy (ACC) of 98.59%, a precision (PPV) of 98.08%, and a specificity (TNR) of 99.04%, ensuring high classification reliability. Accuracy measures the correctness of the overall classification, precision denotes the proportion of correctly identified positive cases, and specificity evaluates the model's ability to correctly classify negative instances while minimising false positives.

As shown in Table 3 and discussed in Section 4, our approach aligns with previous studies on skeleton-based posture recognition, whether heuristic, machine learning, or deep learning techniques are employed.

### 3.3 | Generalisation to Real-World Data

Following experimentation with virtual images, an important issue in computer vision is the potential for generalisation to real-world images. Testing the heuristic-based approach in real-world scenarios helps evaluate its efficiency. Experiments have been conducted with human participants to validate the effectiveness and robustness of the proposed approach beyond a controlled VR setting (see Figure 7).

Figure 8 shows the results of applying the same heuristic approach and variable values to three different real-world video sequences, each representing a specific posture type and processed at 20–30 frames per second. This figure presents the corresponding confusion matrix and performance metrics for the assessment.

A total of 2741 images were processed: 555 for the *Standing* posture, 863 for *Sitting* and 1323 for *Lying*. As in the virtual scenario, the person (in the same posture) continuously moves and rotates throughout the video sequences, thus changing their position and orientation relative to the camera.

The confusion matrix and metrics suggest that the classification model performs well when generalising to real-world images,

**TABLE 3** | Performance comparison with other skeleton-based approaches.

	<b>Paper</b>	<b>Classification method</b>	<b>Sensor</b>	<b>Precision (%)</b>
Posture		Deep learning methods		
	Galvão et al. (2021)	Spatial-temporal graph convolution network (ST-GCN)	RGB	98.62
	Zhang et al. (2022)	Convolutional neural network (CNN)	RGB	98.50
	Cao et al. (2024)	Lookup-based convolutional neural network (LCNN)	RGB-D	95.31
Falling	Zahan et al. (2023)	Graph convolutional network (GCN)	RGB	88.71
		Machine learning methods		
	Ramirez et al. (2022)	Random forest (RF), Support vector machine (SVM), Multilayer perceptron (MLP), K-nearest neighbour (KNN)	RGB	99.81
	Alaoui et al. (2019)	Decision tree (DT), Random forest (RF), Support vector machine (SVM), K-nearest neighbour (KNN)	RGB	98.50
Abnormal sitting		Heuristic methods		
	Chen et al. (2020)	Heuristics	RGB	97.00
		Deep learning methods		
	Li et al. (2023)	Graph convolutional network (GCN)	RGB	95.61
Standing, sitting and lying		Deep learning methods		
	Yadav, Tiwari, et al. (2022)	Convolutional long short-term memory (ConvLSTM)	RGB-D	98.67
	Yadav, Luthra, et al. (2022)	Convolutional neural network (CNN) and Gated recurrent unit (GRU)	RGB	93.94
Ours	Guerra et al. (2023)	Bidirectional long short-term memory (BLSTM), Bidirectional gated recurrent unit (BGRU)	RGB-D	84.00
		Heuristic methods		
		Heuristics (virtual images)	RGB	98.08
		Heuristics (real-world images)	RGB	91.40

with all weighted average values exceeding 91%. The *Lying* posture is classified most accurately, followed by *Standing*, while *Sitting* has the lowest classification accuracy. Evaluating the model on real-world images yields slightly less effective results than on virtual images, as shown in Figure 9. However, it should be noted that the parameter values presented in Table 2 were optimised for the virtual environment. Let us highlight that the same values were applied in this test. Fine-tuning these parameters for real-world conditions would likely yield even better results, bringing performance closer to that achieved in the virtual environment, in which all experimentation was carried out to tune parameters and determine heuristic rules.

### 3.4 | Discussion

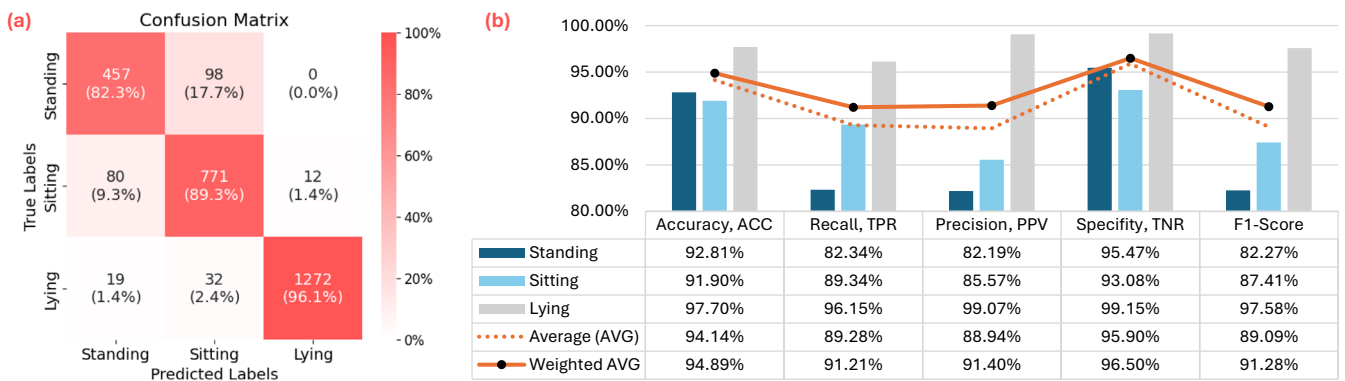
The work presented in this article is based on the detection of three different postures in colour images with a skeleton-based

approach. It is now mandatory to compare our results with those of similar studies dedicated to monitoring elderly people for home care. Consequently, our study has been compared with approaches related to the detection of typical standing, sitting, lying, and/or falling postures. Of course, any comparison can be criticised and improved. Indeed, considerable work has been conducted on posture detection based on skeletonisation of the human body. Here, our aim is to showcase some of the latest advances in this field.

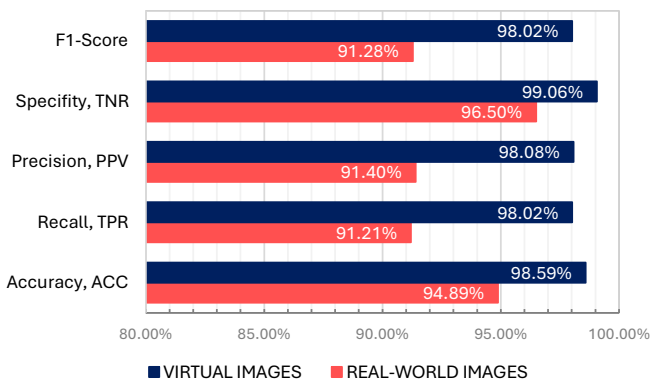
Table 3 shows the performance of methods dealing with one, two, or three of the proposed postures. The table shows the performance of approaches that use colour (RGB), as is the case here, and depth sensors (RGB-D). It also presents results for techniques employing machine learning, deep learning, or heuristics for classification. As can be seen, our heuristic-based method achieves success rates that are very close to, or even better than, those of other studies that use machine or deep learning methods.



**FIGURE 7** | Examples of bounding boxes for the three postures studied (*Standing*, *Sitting*, and *Lying*) in real-world images.



**FIGURE 8** | Results for real-world images: (a) Confusion matrix and (b) metrics per posture (*Standing*, *Sitting*, and *Lying*).



**FIGURE 9** | Comparison of weighted average metrics for virtual and real-world images.

When examining the falling posture first, it is evident that all methods demonstrate excellent classification results. The more traditional machine learning-based methods offer accuracy that is on a par with that of deep learning-based methods. The results shown for machine learning-based methods are those obtained using the most successful of these methods, that is, random forest (RF) (Ramirez et al. 2022; Alaoui et al. 2019). This is not surprising, given that RF has traditionally produced excellent

posture detection results by combining the outputs of multiple decision trees to obtain a single result. In the case of deep learning-based methods, two of these works are inspired by convolutional neural networks (CNNs) (Zhang et al. 2022; Cao et al. 2024), while two others are based on recurrent neural networks (RNNs) (Galvão et al. 2021; Zahan et al. 2023). CNNs are typically used to solve problems involving spatial data, such as images, whereas RNNs are better suited to analysing temporal and sequential data, such as text or video. Accordingly, the first two approaches classified images, while the latter two used videos. Finally, the heuristic-based work (Chen et al. 2020) achieves excellent precision, demonstrating that simple, non-learning-based methods are still a valid approach to action detection and classification.

Another article that detects abnormal sitting posture also uses an RNN. The results are comparable to those of the previous case. It would have been expected that the results for the detection of three postures (standing, sitting and lying) would be worse than for a single posture. However, Table 3 shows that this is not the case. Here, we only present papers based on deep learning methods. Noteworthy are two papers (Yadav, Luthra, et al. 2022; Yadav, Tiwari, et al. 2022) that use a ConvLSTM (a type of recurrent neural network for spatio-temporal prediction

with convolutional structures in the input-to-state and state-to-state transitions) and a CNN and an RNN, respectively. Another work (Guerra et al. 2023) has worse results, although the authors also work with CNNs and RNNs, in this case, bilateral ones. As can be seen, our work has achieved a very competitive level of precision. Thus, our heuristic-based approach offers a computationally efficient alternative to traditional machine learning and non-traditional deep learning methods.

On the other hand, the following limitations have been identified in the application of the proposed heuristic-based approach:

1. Limitations of MediaPipe for posture recognition. MediaPipe is a useful tool for detecting the posture of the first person it detects and provides information on joint and limb positions. We have identified the optimal parameters (and their corresponding values) for detecting various body postures using MediaPipe. However, the system encounters significant limitations when handling multiple individuals within the frame (e.g., a caretaker and an elderly person).
2. False positives may appear. In VR testing, false positives appeared, specifically when the avatar was lying down with bent knees, creating the impression that it was sitting. This type of false positive could be corrected in future work by analysing additional parameters to minimise the margin of error.

## 4 | CONCLUSIONS

This article introduced a novel proposal for the estimation of three postures during the monitoring of a person for the purpose of home care. A heuristic-based approach offered a computationally efficient alternative to traditional machine learning and deep learning methods, ensuring real-time performance and scalability.

The approach used the virtual camera of a UAV in virtual home environments. The keypoints of the body skeleton were obtained using the MediaPipe Pose library. In this work, only eight of the 33 points provided by this library were used. MediaPipe proved to be an optimal library in terms of performance for generating and monitoring a person. A series of heuristics based on the physical proportions of the body were then used to provide percentage values for each posture, thus providing accurate estimates of three different postures: standing, sitting, and lying. The highest value was the one that determined the avatar's current posture.

The main conclusions have to do with the following trends in computer vision, and more concretely in posture recognition:

- Real-time posture recognition. This research contributes to real-time posture recognition systems. By integrating MediaPipe and Unity, the proposed framework enables the recognition of lying, standing, and sitting postures in real-time scenarios, facilitating applications in healthcare, gaming, and human-computer interaction.
- Cross-platform compatibility. The use of Unity as a development platform ensures cross-platform compatibility,

allowing the proposed posture recognition system to be deployed seamlessly across different devices and operating systems. This aspect increases the accessibility and applicability of the developed solution.

- Low computational overhead. By using heuristic computations on geometric figures derived from MediaPipe landmarks, this approach minimises computational overhead compared to traditional machine learning-based methods. This feature makes the proposed system suitable for resource-constrained environments such as mobile devices or embedded systems.
- Customisation and adaptability. The modular design of the proposed system in Unity allows for easy customisation and adaptation to different user scenarios and environments. Researchers and developers will be able to extend the functionality of the system by integrating additional features or refining existing algorithms, fostering continued innovation in posture recognition technology.
- Potential for extension to other applications. Beyond posture recognition, the methodology presented in this article has potential for application in related areas such as gesture recognition, human motion analysis, and virtual reality interactions. This versatility opens up avenues for future research and development in diverse areas.

Lastly, as we tested under the virtual reality paradigm, there were experiments that gave us surprising results, such as when we tried a posture that broke the laws of physics but gave positive results.

Our future work will focus on making progress in the following areas:

1. Expanding testing to more complex VR scenarios and conducting additional pipeline tests in more complex and realistic indoor scenarios involving furniture, partial occlusions, and dynamic interactions, to better mimic real-world conditions.
2. Addressing the false positives by refining the heuristic rules to reduce the margin of error, particularly for postures prone to misclassification, and by considering the incorporation of additional parameters or features to improve accuracy.
3. Incorporating user feedback through studying the possibility of including feedback from potential end users, such as caregivers or individuals in need of home care, to assess the system's practical usability and acceptability.
4. Analysing the viability of integrating machine learning components to complement the heuristic-based approach and exploring other applications of the proposed technology.
5. Expanding the study of posture detection when multiple individuals appear in the scene, using other pose estimation systems, such as OpenPose (<https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/?ref=blog.roboflow.com>) or YOLO (You Only Look Once) networks (Chen et al. 2022; Dai and Liu 2023; Bustamante et al. 2024).

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Alaoui, A. Y., S. El Fkihi, and R. O. H. Thami. 2019. "Fall Detection for Elderly People Using the Variation of Key Points of Human Skeleton." *IEEE Access* 7: 154786–154795. <https://doi.org/10.1109/ACCESS.2019.2946522>.
- Belagiannis, V., and A. Zisserman. 2017. "Recurrent Human Pose Estimation." In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (fg 2017)*, 468–475. IEEE. <https://doi.org/10.1109/FG.2017.64>.
- Belmonte, L. M., A. S. García, R. Morales, J. L. de la Vara, F. López de la Rosa, and A. Fernández-Caballero. 2021. "Feeling of Safety and Comfort Towards a Socially Assistive Unmanned Aerial Vehicle That Monitors People in a Virtual Home." *Sensors* 21, no. 3: 908. <https://doi.org/10.3390/s21030908>.
- Belmonte, L. M., A. S. García, E. Segura, P. Novais, R. Morales, and A. Fernández-Caballero. 2021. "Virtual Reality Simulation of a Quadrotor to Monitor Dependent People at Home." *IEEE Transactions on Emerging Topics in Computing* 9, no. 3: 1301–1315. <https://doi.org/10.1109/TETC.2020.3000352>.
- Belmonte, L. M., R. Morales, and A. Fernández-Caballero. 2019. "Computer Vision in Autonomous Unmanned Aerial Vehicles—A Systematic Mapping Study." *Applied Sciences* 9, no. 15: 3196. <https://doi.org/10.3390/app9153196>.
- Belmonte, L. M., R. Morales, A. S. García, E. Segura, P. Novais, and A. Fernández-Caballero. 2019. "Trajectory Planning of a Quadrotor to Monitor Dependent People." In *Understanding the Brain Function and Emotions*, 212–221. Springer International Publishing. [https://doi.org/10.1007/978-3-030-19591-5\\_22](https://doi.org/10.1007/978-3-030-19591-5_22).
- Bric, J. D., D. C. Lumbard, M. J. Frelich, and J. C. Gould. 2016. "Current State of Virtual Reality Simulation in Robotic Surgery Training: A Review." *Surgical Endoscopy* 30, no. 6: 2169–2178. <https://doi.org/10.1007/s00464-015-4517-y>.
- Bustamante, A., L. Belmonte, R. Morales, A. Pereira, and A. Fernández-Caballero. 2022. "Video Processing From a Virtual Unmanned Aerial Vehicle: Comparing Two Approaches to Using Opencv in Unity." *Applied Sciences* 12: 5958. <https://doi.org/10.3390/app12125958>.
- Bustamante, A., L. Belmonte, R. Morales, A. Pereira, and A. Fernández-Caballero. 2024. "Bridging the Appearance Domain Gap in Elderly Posture Recognition With Yolov9." *Applied Sciences* 14: 9695. <https://doi.org/10.3390/app14219695>.
- Cao, Y., M. Guo, J. Sun, X. Chen, and J. Qiu. 2024. "Fall Detection Based on LCNn and Fusion Model of Weights Using Human Skeleton and Optical Flow." *Signal, Image and Video Processing* 18: 833–841. <https://doi.org/10.1007/s11760-023-02776-9>.
- Carretero, S., J. Garcés, and F. Ródenas. 2007. "Evaluation of the Home Help Service and its Impact on the Informal Caregiver's burden of Dependent Elders." *International Journal of Geriatric Psychiatry* 22, no. 8: 738–749. <https://doi.org/10.1002/gps.1733>.
- Chen, T., Z. Ding, and B. Li. 2022. "Elderly Fall Detection Based on Improved Yolov5s Network." *IEEE Access* 10: 91273–91282. <https://doi.org/10.1109/ACCESS.2022.3202293>.
- Chen, W., Z. Jiang, H. Guo, and X. Ni. 2020. "Fall Detection Based on Key Points of Human-Skeleton Using OpenPose." *Symmetry* 12, no. 5: 744. <https://doi.org/10.3390/sym12050744>.
- Dai, Y., and W. Liu. 2023. "Gl-yolo-lite: A Novel Lightweight Fallen Person Detection Model." *Entropy* 25, no. 4: 587. <https://doi.org/10.3390/e25040587>.
- Dalal, N., B. Triggs, and C. Schmid. 2006. "Human Detection Using Oriented Histograms of Flow and Appearance." In *Computer Vision – ECCV 2006*, 428–441. Springer. [https://doi.org/10.1007/11744047\\_33](https://doi.org/10.1007/11744047_33).
- Dentamaro, V., V. Gattulli, D. Impedovo, and F. Manca. 2024. "Human Activity Recognition With Smartphone-Integrated Sensors: A Survey." *Expert Systems With Applications* 246: 123143. <https://doi.org/10.1016/j.eswa.2024.123143>.
- Galvão, Y. M., L. Portela, J. Ferreira, P. Barros, O. A. de Araújo Fagundes, and B. J. T. Fernandes. 2021. "A Framework for Anomaly identification Applied on Fall Detection." *IEEE Access* 9: 77264–77274. <https://doi.org/10.1109/ACCESS.2021.3083064>.
- Guerra, B. M. V., S. Ramat, G. Beltrami, and M. Schmid. 2023. "Recurrent Network Solutions for Human Posture Recognition Based on Kinect Skeletal data." *Sensors* 23, no. 11: 5260. <https://doi.org/10.3390/s23115260>.
- Halder, A., and A. Tayade. 2021. "Real-Time Vernacular Sign Language Recognition Using MediaPipe and Machine Learning." *International Journal of Research Publication and Reviews* 2: 9–17.
- Hassan, E., M. S. Hossain, S. Elmuogy, A. Ghoneim, K. AlMutib, and A. Saber. 2024. "Ultimate Pose Estimation: A Comparative Study." *Expert Systems* 41, no. 9: e13586. <https://doi.org/10.1111/exsy.13586>.
- Hentati, A. I., L. Krichen, M. Fourati, and L. C. Fourati. 2018. "Simulation Tools, Environments and Frameworks for UAV Systems Performance Analysis." In *2018 14th International Wireless Communications & Mobile Computing Conference*, 1495–1500. IEEE. <https://doi.org/10.1109/IWCMC.2018.8450505>.
- Jiang, X., Z. Hu, S. Wang, and Y. Zhang. 2023. "A Survey on Artificial Intelligence in Posture Recognition." *Computer Modeling in Engineering and Sciences* 137, no. 1: 35–82. <https://doi.org/10.32604/cmescs.2023.027676>.
- Kaur, H., V. Rani, and M. Kumar. 2024. "Human Activity Recognition: A Comprehensive Review." *Expert Systems* 41, no. 11: e13680. <https://doi.org/10.1111/exsy.13680>.
- Kim, J.-W., J.-Y. Choi, E.-J. Ha, and J.-H. Choi. 2023. "Human Pose Estimation Using Mediapipe Pose and Optimization Method Based on a Humanoid Model." *Applied Sciences* 13: 2700. <https://doi.org/10.3390/app13042700>.
- Kumar, S., S. Madesh, B. Nikitha, and Maheshwari. 2025. "Intelligent Posture Monitoring System With Real Time Notifications Using Mediapipe and Opencv." In *Computing Technologies for Sustainable Development. IRCCTSD 2024. Communications in Computer and Information Science*, edited by P. D. Sivakumar, R. Ramachandran, C. Pasupathi, and P. Balakrishnan, 145–155. Springer.
- Li, L., G. Yang, Y. Li, D. Zhu, and L. He. 2023. "Abnormal Sitting Posture Recognition Based on Multi-Scale Spatiotemporal Features of Skeleton Graph." *Engineering Applications of Artificial Intelligence* 123: 106374. <https://doi.org/10.1016/j.engappai.2023.106374>.
- Li, M., H. Hu, J. Xiong, X. Zhao, and H. Yan. 2024. "TSwin-Pose: Enhanced Monocular 3D Human Pose Estimation With Jointflow." *Expert Systems With Applications* 249: 123545. <https://doi.org/10.1016/j.eswa.2024.123545>.
- Li, X., W. Yi, H.-L. Chi, X. Wang, and A. P. Chan. 2018. "A Critical Review of Virtual and Augmented Reality (VR/AR) Applications in Construction Safety." *Automation in Construction* 86: 150–162. <https://doi.org/10.1016/j.autcon.2017.11.003>.

- Liu, J., Y. Wang, Y. Liu, S. Xiang, and C. Pan. 2020. "3D PostureNet: A Unified Framework for Skeleton-Based Posture Recognition." *Pattern Recognition Letters* 140, no. 8: 143–149. <https://doi.org/10.1016/j.patrec.2020.09.029>.
- Lugaresi, C., J. Tang, H. Nash, et al. 2019. "MediaPipe: A Framework for Building Perception Pipelines."
- Martínez, A., L. M. Belmonte, A. S. García, A. Fernández-Caballero, and R. Morales. 2021. "Facial Emotion Recognition From an Unmanned Flying Social Robot for Home Care of Dependent People." *Electronics* 10, no. 7: 868. <https://doi.org/10.3390/electronics10070868>.
- Mim, T. R., M. Amatullah, S. Afreen, et al. 2023. "GRU-INC: An Inception-Attention Based Approach Using GRU for Human Activity Recognition." *Expert Systems With Applications* 216: 119419. <https://doi.org/10.1016/j.eswa.2022.119419>.
- Moon, S., Y. Park, D. W. Ko, and I. H. Suh. 2016. "Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering." *International Journal of Advanced Robotic Systems* 13, no. 2: 65. <https://doi.org/10.5772/62415>.
- Newell, A., K. Yang, and J. Deng. 2016. "Stacked Hourglass Networks for Human Pose Estimation." In *Computer Vision – ECCV 2016*, 483–499. Springer International Publishing. [https://doi.org/10.1007/978-3-319-46484-8\\_29](https://doi.org/10.1007/978-3-319-46484-8_29).
- Penmetsa, S., F. Minhuj, A. Singh, and S. Omkar. 2014. "Autonomous UAV for Suspicious Action Detection Using Pictorial Human Pose Estimation and Classification." *Electronic Letters on Computer Vision and Image Analysis* 13, no. 1: 18. <https://doi.org/10.5565/rev/elcvia.582>.
- Pietraszek, T. 2004. "Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection." In *Lecture Notes in Computer Science*, 102–124. Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-30143-1\\_6](https://doi.org/10.1007/978-3-540-30143-1_6).
- Puwein, J., L. Ballan, R. Ziegler, and M. Pollefeys. 2015. "Joint Camera Pose Estimation and 3D Human Pose Estimation in a Multi-Camera Setup." In *Computer Vision – ACCV 2014*, 473–487. Springer International Publishing. [https://doi.org/10.1007/978-3-319-16808-1\\_32](https://doi.org/10.1007/978-3-319-16808-1_32).
- Ramirez, H., S. A. Velastin, P. Aguayo, E. Fabregas, and G. Farias. 2022. "Human Activity Recognition by Sequences of Skeleton Features." *Sensors (Basel)* 22, no. 11: 3991. <https://doi.org/10.3390/s22113991>.
- Roberts, D. J., A. S. Garcia, J. Dodiya, R. Wolff, A. J. Fairchild, and T. Fernando. 2015. "Collaborative Telepresence Workspaces for Space Operation and Science." In *2015 IEEE Virtual Reality*, 275–276. IEEE.
- Rohmer, E., S. P. Singh, and M. Freese. 2013. "V-REP: A Versatile and Scalable Robot Simulation Framework." In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1321–1326. IEEE.
- Rudol, P., and P. Doherty. 2008. "Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery." In *2008 IEEE Aerospace Conference*, 1–8. IEEE.
- Sanchez-Comas, A., K. Synnes, and J. Hallberg. 2020. "Hardware for Recognition of Human Activities: A Review of Smart Home and AAL Related Technologies." *Sensors (Basel)* 20, no. 15: 4227. <https://doi.org/10.3390/s20154227>.
- Schwarz, L. A., A. Mkhitarian, D. Mateus, and N. Navab. 2012. "Human Skeleton Tracking From Depth Data Using Geodesic Distances and Optical Flow." *Image and Vision Computing* 30, no. 3: 217–226. <https://doi.org/10.1016/j.imavis.2011.12.001>.
- Seth, A., J. M. Vance, and J. H. Oliver. 2011. "Virtual Reality for Assembly Methods Prototyping: A Review." *Virtual Reality* 15, no. 1: 5–20. <https://doi.org/10.1007/s10055-009-0153-y>.
- Wan, J., M. Li, M. O'Grady, X. Gu, A. A. H. Alawlaqi, and G. O'Hare. 2021. "Time-Bounded Activity Recognition for Ambient Assisted Living." *IEEE Transactions on Emerging Topics in Computing* 9, no. 1: 471–483. <https://doi.org/10.1109/TETC.2018.2870047>.
- Yadav, S. K., A. Luthra, K. Tiwari, H. M. Pandey, and S. A. Akbar. 2022. "ARFDNet: An Efficient Activity Recognition & Fall Detection System Using Latent Feature Pooling." *Knowledge-Based Systems* 239: 107948. <https://doi.org/10.1016/j.knosys.2021.107948>.
- Yadav, S. K., K. Tiwari, H. M. Pandey, and S. A. Akbar. 2022. "Skeleton-Based Human Activity Recognition Using ConvLSTM and Guided Feature Learning." *Soft Computing* 26: 877–890. <https://doi.org/10.1007/s00500-021-06238-7>.
- Zahan, S., G. M. Hassan, and A. Mian. 2023. "SDFA: Structure-Aware Discriminative Feature Aggregation for Efficient Human Fall Detection in Video." *IEEE Transactions on Industrial Informatics* 19, no. 8: 8713–8721. <https://doi.org/10.1109/TII.2022.3221208>.
- Zhang, X., G. Foderaro, C. Henriquez, and S. Ferrari. 2018. "A Scalable Weight-Free Learning Algorithm for Regulatory Control of Cell activity in Spiking Neuronal Networks." *International Journal of Neural Systems* 28, no. 2: 1750015. <https://doi.org/10.1142/S0129065717500150>.
- Zhang, Y., X. Zheng, W. Liang, S. Zhang, and X. Yuan. 2022. "Visual Surveillance for Human Fall Detection in Healthcare IoT." *IEEE Multimedia* 29, no. 1: 36–46. <https://doi.org/10.1109/MMUL.2022.3155768>.
- Zhang, Z., H. Wang, Q. You, L. Chen, and H. Ning. 2024. "A Novel Temporal Adaptive Fuzzy Neural Network for Facial Feature Based Fatigue Assessment." *Expert Systems With Applications* 252: 124124. <https://doi.org/10.1016/j.eswa.2024.124124>.