






## Human Activity Recognition via Feature Extraction and Artificial Intelligence Techniques: A Review

### Reconocimiento de actividades humanas por medio de extracción de características y técnicas de inteligencia artificial: una revisión

José Camilo Eraso-Guerrero <sup>1</sup>, Elena Muñoz-España <sup>2</sup>, Mariela Muñoz-Añasco <sup>3</sup>

Fecha de Recepción: 03 de enero de 2022

Fecha de Aceptación: 04 de julio de 2022

**Cómo citar:** Eraso-Guerrero, J.C. Muñoz-España, E. y Muñoz-Añasco, M. (2022). Human Activity Recognition via Feature Extraction and Artificial Intelligence Techniques: A Review. *Tecnura*, 26(74), 213-236. <https://doi.org/10.14483/22487638.17413>

### Abstract

**Context:** In recent years, the recognition of human activities has become an area of constant exploration in different fields. This article presents a literature review focused on the different types of human activities and information acquisition devices for the recognition of activities. It also delves into elderly fall detection via computer vision using feature extraction methods and artificial intelligence techniques.

**Methodology:** This manuscript was elaborated following the criteria of the document review and analysis methodology (RAD), dividing the research process into the heuristics and hermeneutics of the information sources. Finally, 102 research works were referenced, which made it possible to provide information on current state of the recognition of human activities.

**Results:** The analysis of the proposed techniques for the recognition of human activities shows the importance of efficient fall detection. Although it is true that, at present, positive results are obtained with the techniques described in this article, their study environments are controlled, which does not contribute to the real advancement of research.

**Conclusions:** It would be of great impact to present the results of studies in environments similar to reality, which is why it is essential to focus research on the development of databases with real falls of adults or in uncontrolled environments.

**Keywords:** human activity recognition, fall detection, type of activities, feature extraction, convolutional neural networks

<sup>1</sup>Electronics engineer, Master's candidate in Automatics. Universidad del Cauca. Popayán, Colombia.

Email: [joseeraso@unicauca.edu.co](mailto:joseeraso@unicauca.edu.co)

<sup>2</sup>Electronics and Telecommunications engineer, specialist in Industrial Informatics, specialist in Telematic Networks and Services, Master's in Electronics and Telecommunications. Professor at Universidad del Cauca. Popayán, Colombia.

Email: [elenam@unicauca.edu.co](mailto:elenam@unicauca.edu.co)

<sup>3</sup>Industrial engineer, Master's in Business Administration, Master's in Automatics, PhD in Automatics, Robotics, and Industrial Informatics. Professor at Universidad del Cauca. Popayán.

Email: [mamunoz@unicauca.edu.co](mailto:mamunoz@unicauca.edu.co)

## Resumen

**Contexto:** En los últimos años, el reconocimiento de actividades humanas se ha convertido en un área de constante exploración en diferentes campos. Este artículo presenta una revisión de la literatura enfocada en diferentes tipos de actividades humanas y dispositivos de adquisición de información para el reconocimiento de actividades, y profundiza en la detección de caídas de personas de tercera edad por medio de visión computacional, utilizando métodos de extracción de características y técnicas de inteligencia artificial.

**Metodología:** Este manuscrito se elaboró con criterios de la metodología de revisión y análisis documental (RAD), dividiendo el proceso de investigación en heurística y hermenéutica de las fuentes de información. Finalmente, se referenciaron 102 investigaciones que permitieron dar a conocer la actualidad del reconocimiento de actividades humanas.

**Resultados:** El análisis de las técnicas propuestas para el reconocimiento de actividades humanas muestra la importancia de la detección eficiente de caídas. Si bien es cierto en la actualidad se obtienen resultados positivos con las técnicas descritas en este artículo, sus entornos de estudio son controlados, lo cual no contribuye al verdadero avance de las investigaciones.

**Conclusiones:** Sería de gran impacto presentar resultados de estudios en entornos semejantes a la realidad, por lo que es primordial centrar el trabajo de investigación en la elaboración de bases de datos con caídas reales de personas adultas o en entornos no controlados.

**Palabras clave:** reconocimiento de la actividad humana, detección de caídas, tipos de actividades, extracción de características, redes neuronales convolucionales

## Table of Contents

	Page
<b>Introduction</b>	215
<b>Classification of human activities</b>	216
<b>Information acquisition methods</b>	217
<b>Feature extraction</b>	218
Global feature extraction . . . . .	218
Local feature extraction . . . . .	220
Depth-based feature extraction . . . . .	221
Convolutional neural networks . . . . .	222
<b>Conclusions</b>	224
<b>Acknowledgments</b>	225
<b>References</b>	225

## INTRODUCTION

Human activity recognition (HAR) aims to model user behavior and automatically identify the tasks they perform by observing and analyzing human behavior (Saini *et al.*, 2018, Uzunovic *et al.*, 2018, Brophy *et al.*, 2018), which results in the recognition of people's activities, identities, personalities, and psychological state (Vrigkas *et al.* (2015)).

In recent years, HAR has become an area of constant exploration in different fields; its applications are a current research subject, as it helps automate processes and activities that may go unnoticed by the human eye or may constitute tedious tasks. For Lohit *et al.* (2018), a human pose transmits the configuration of the body parts and implicit predictive information on people's subsequent movement, dynamic information that may be utilized in various applications. By reviewing the literature, it can be found that HAR exhibits a growing demand in the fields of entertainment (Lawrence *et al.*, 2010, Han *et al.*, 2013, Akhavian & Behzadan, 2016); video surveillance systems (Ryoo, 2011, Preis *et al.*, 2012, Liu *et al.*, 2014, Ben Mabrouk & Zagrouba, 2018, Cosar *et al.*, 2017, J.-W. Hsieh *et al.*, 2014); emergency rescue and emergency robotics (Durrant-Whyte *et al.*, 2012); smart cities, sports performance, military applications, medical monitoring for caring of the elderly, and diverse health care (Banos *et al.*, 2012, Chen *et al.*, 2012, Avci *et al.*, 2010, Kim *et al.*, 2010, Sazonov *et al.*, 2011, Ismail *et al.*, 2015, Rafferty *et al.*, 2017); among others (Elbasiony & Gomaa.). The common factor of research on HAR is the set of problems under study, which involve recognizing a specific activity such as the weather, object protection, and lighting conditions, among others. Moreover, an activity may vary from one person to another (Y. Yang *et al.*, 2019, Kahani *et al.*, 2019). Thus, it is essential to find different ways to optimize the recognition of human activities.

Modern and efficient methods for healthcare are now being proposed, such as the use of blockchain and the Internet of Things (Pava *et al.*, 2021). However, this review prioritizes the works and advances on HAR, specifically elderly fall detection since, according to the World Health Organization (WHO), the proportion of the planet's population over 60 years will double from 11 to 22% between 2000 and 2050 (2015). In huge numbers, this age group will grow from 605 million to 2 billion in the course of half a century. As a consequence, caring for and monitoring the health of the elderly will become an essential and daunting task. Li *et al.* (2018) state that approximately 58% of elderly over 80 years old have passed away after a severe fall due to physical trauma, mild traumatic brain injury, hip fracture, among others, which may discourage this population from working out. A sedentary lifestyle in the elderly is another problem that entails other health consequences, such as obesity and cardiovascular diseases (Suto & Oniga, 2019).

As workout and movement are vital for the elderly, monitoring and recognizing daily activities is essential to provide them with proper healthcare. According to Li *et al.* (2018), the automatic detection of falls or movements that may affect health can significantly reduce the consequences of the incident. It may also allow tracking and reporting anomalies from patterns of normal daily behaviors by adults

with high risk of falling.

This document presents a review of the state of the art regarding 1) the classification of human activities, 2) the methods for HAR information acquisition and 3) the methods that have been used for feature extraction from videos and images in order to recognize the activities of the elderly.

This research was conducted following the criteria for review methodology and document analysis (Barbosa-Chacón *et al.*, 2013), dividing the research process into the heuristics and hermeneutics of the different information sources.

## CLASSIFICATION OF HUMAN ACTIVITIES

Human activity recognition is a current research topic due to its various applications in the entertainment industry, video surveillance, healthcare, robotics, smart cities, sports performance, and military applications. Therefore, the main objective of this work is to review the field of HAR with a focus on elderly care.

Human activities are classified depending on their complexity and duration. For Hassan *et al.* (2018), activities are divided into short-term activities and simple and complex tasks: firstly, short-term activities such as the transition between sitting and standing up; secondly, basic activities such as walking and reading; and complex activities, which involve scenarios where there is an interaction with objects or people. On the other hand, Vrigkas *et al.* (2015) propose another mechanism to classify activities which also considers their complexity. A summary of this classification is shown in Table 1.

**Table 1.** Classification of human activities

Classification	Description
<i>Gestures</i>	Primitive movements of a person's body parts, which correspond to a particular action.
<i>Atomic actions</i>	A person's movements that are part of more complex activities.
<i>Human-object interaction or human-human interaction</i>	Human activities that involve two or more people or objects.
<i>Activities in group</i>	Activities carried out by groups of people.
<i>Behaviors</i>	Physical activities associated with feelings, personality, and the psychological state of an individual.
<i>Events</i>	High-level activities that describe social actions between individuals and indicate a person's social roles.

Source: Vrigkas *et al.* (2015)

Research related to recognizing activities carried out by the elderly is focused on identifying short-term and basic-specific activities. An example of this is the work carried out by [Khan \*et al.\* \(2011\)](#), which aimed to detect six specific activities (forward falls, backward falls, chest pain, faints, vomiting, and headaches). On the other hand, [X. Ma \*et al.\* \(2014\)](#) attempted to recognize other six activities (people falling, flexing, sitting, squatting, or lying down). In turn, [Amiri \*et al.\* \(2014\)](#) increased the number of activities to be recognized (a person cleaning a table, drinking a drink, taking or dropping an object, reading, sitting, standing up, writing, using a phone, and falling), which also expanded the difficulty and vagueness of the system due to occlusion issues and the similarity between actions ([Yu \*et al.\*, \(2013\)](#)).

## INFORMATION ACQUISITION METHODS

The first step to recognize a determined human activity is obtaining information for subsequent processing. This process may be carried out in different ways. The first method is based on environmental sensors, such as pressure, acoustic, electromyography, and different sensors that may be integrated and distributed around the environment where the identification of different activities is required ([L. Yang \*et al.\* \(2016\)](#)). The use of different sensors may entail high costs and could be an intrusive method. Aspects such as the arrangement and generation of different types of sensors should also be taken into account, especially in underdeveloped territories, as discussed by [Nivia-Vargas & Jaramillo-Jaramillo \(2018\)](#).

The second method also receives information through portable sensors such as contact sensors, gyroscopes, and accelerometers ([Rosati \*et al.\*, 2018](#)). Using methods based on sensors has a specific set of difficulties when recognizing elderly activities. According to [Khan \*et al.\* \(2013\)](#), the elderly often forget to wear portable sensors, and their use in different parts of the body causes frustration since it limits their movement.

On the other hand, ([Kwolek & Kepski \(2014\)](#)) argue that the vast majority of elderly people do not enjoy using sensors, as they generate excessive false alarms. Some daily activities are wrongly detected as falls, which may also frustrate users.

This article delves into the third information acquisition method: incorporating computer vision using cameras, depth sensors, and image processing techniques. According to [Yu \*et al.\*, 2013](#) and [Amiri \*et al.\* \(2014\)](#), this is a non-intrusive method that can extract a large amount of information in comparison with portable sensor methods. Furthermore, this method is not easily affected by noise in the environment. On that premise, [Panahi & Ghods \(2018\)](#) highlight the technological progress of extracting images from video using RGB (red, green, and blue) cameras or using depth map images to determine the different distances of objects or people. [L. Yang \*et al.\* \(2016\)](#) divide the vision-based method into three categories: methods using standard RGB cameras, 3D-based methods using multiple cameras, and 3D-based methods using depth cameras.

The vision-based method also has its limitations, which include a lack of privacy, as it implies having a camera in the environment at all times. Moreover, [Concone et al. \(2019\)](#) criticize its computational cost, since this method may rarely run in real-time, and they highlight the fact that the performance of the method strongly depends on the position of the cameras.

## FEATURE EXTRACTION

Although HAR has been a continuous topic of research in the last decade, there are still different aspects that hinder the accurate recognition of elderly activities. With the computer vision method, this includes features such as the weather, object protection or occlusion, lighting conditions, the similarity between some activities, clutter in the background of the image, privacy problems, and other specific difficulties that may cause false detections. For this reason, it is vital to study the different methods in order to optimize the recognition of these activities, especially regarding fall detection.

Some studies ([Yu et al., 2013](#), [Goudelis et al., 2015](#)) argue that the most essential step for successful activity recognition is to select a method for feature extraction from an image or a video. Different methods have been proposed whose purpose is to effectively distinguish non-intentional actions such as falls from other daily activities. This review of the state of the art focuses on the classification of feature extraction methods presented by [S. Zhang et al. \(2017\)](#), which classifies them in terms of their approach: local characteristics, global characteristics, and depth-based representation. Also, the current method based on convolutional neural networks is attached.

For [Das Dawn & Shaikh \(2016\)](#), the shape or edge of an object are relevant data that can be used to determine local characteristics. However, global information involves flow description or movement in a video.

### Global feature extraction

This method allows extracting global descriptions from videos and images, which, according to [S. Zhang et al. \(2017\)](#), allows localizing the human subject and isolating them from the background, using subtraction methods to acquire their silhouette and shape. Other global representation methods are 3D space-time volumes, which monitor a human being's silhouette for a determined period of time. There is also the Fourier Transform method, which is based on monitoring the frequency of a silhouette for activity recognition.

Various other studies use global feature extraction to recognize human actions, especially for elderly fall detection. In general, research in this field takes advantage of the silhouette of the human body to reach its objective.

Elderly fall detection is the main objective of several works ([Khan et al. \(2011\)](#), [Khan et al. \(2013\)](#), [Yu et al., 2012](#), [Yu et al., 2013](#), [Foroughi et al., 2008](#)), which focus on extracting the human silhouette for subsequent processing. These studies have several differences. [Khan et al. \(2011\)](#) use the human silhouet-



te to extract information from the elderly using R-transform, invariant scale, rotation characteristics, and Kernel discriminant analysis (KDA) as they attempt to detect human falls while considering the different distances of people in front of the camera. On the other hand, the works by [Yu et al., 2012](#), [Yu et al., 2013](#) have several common factors: both techniques detect falls in the elderly, extract the adult's silhouette, and calculate the human figure's center of mass. Nonetheless, [Yu et al., 2013](#) employ the method presented by [Rougier et al. \(2007\)](#) to extract and delimit people's silhouettes via ellipse features and look for the structural characteristics and shape of human actions, locating its centroid as a fall detection method. Meanwhile, [Yu et al., 2012](#) calculate the centroid of the human silhouette and identify the person's orientation. To this effect, at least two cameras are needed, both of them synchronized in order to minimize occlusion. Finally, ([Foroughi et al., 2008](#)) use the human silhouette as captured from videos or images to identify histograms of its segmented projection, analyzing temporary changes in an elderly person's head in order to recognize a possible fall.

As these systems were implemented in different contexts, they reported different performances. However, the study areas of these works were controlled environments such as small apartments with multiple cameras, few lighting changes, and high computational costs. For instance, [Auvinet \(n.d\)](#) used various cameras to extract 3D images, aiming to detect and analyze the volume in the elderly silhouette's vertical space, activating a falling alarm when the volume distribution was abnormally close to the floor. For an extended period, this method reached a recognition effectiveness of 99,7%, albeit using eight simultaneous cameras, having a high computational cost with regard to synchronization and performance, and making the system challenging to implement on a daily basis.

On the other hand, ([Nguyen et al., 2016](#)) aimed to recognize indoor human actions tested in different environments with natural lighting and different shadows, as well as involving diverse daily activities, which caused several failures. Nonetheless, as their method is based on the use of a single RGB camera, it is easy to implement and entails a low computational cost. Falls in the elderly are detected by analyzing movement orientation and magnitude, changes in the human shape, and movement in the image's histogram. The authors suggest using additional techniques in future research, which includes detecting the head and the inactivity zone.

Optical flow is a global extraction technique used to extract and describe silhouettes on moving or dynamic backgrounds. [Efros et al. \(2003\)](#) used this method to recognize actions performed by soccer and tennis players and ballet dancers in TV broadcasts. The authors suggest applying this technique to extract the dynamic background, focusing only on the sportsmen's silhouette.

Despite the fact that systems using global feature extraction have performed well in controlled environments, [S. Zhang et al. \(2017\)](#) have exposed the difficulties of these systems given their noise sensitivity and viewpoint changes. Furthermore, authors such as ([Goudelis et al., 2015](#)) have indicated that methods based on silhouettes and figures lack solidity and generalization, as they depend on an accurate extraction of the human silhouette and the different geometric transformations, which may be distorted by the distance and position of the subject.

## Local feature extraction

S. Zhang *et al.* (2017) explain that this method focuses on specific local patches determined by interest point detectors or dense sampling, which densely cover the content of a video or an image. The first interest point detector was proposed by Harris & Stephens (1988) and is known for being an excellent corner detector, giving rise to further research works such as the one by Laptev & Lindeberg (2003), who proposed 3D space-time interest points (STIP). The latter would become the main interest point detectors and inspire even further research (Chakraborty *et al.*, 2012, Laptev, 2005, Nguyen *et al.*, 2015), which aimed to optimize these techniques.

According to Das Dawn & Shaikh (2016), STIP is an essential technique for robust interest points extraction from a video or image in the space-time domain, such as a corner point or an isolated point where the intensity is maximum or minimum –even endpoints of lines and curves.

Amiri *et al.* (2014) focused on simulating a smart home environment using two cameras and a Kinect sensor placed between them. Local feature extraction with space-time techniques was implemented using the Hariss3D algorithm as a feature detector and STIP as a feature descriptor. The system's main difficulties are occlusion problems and clutter in the background, since tracking the human body is a challenging and an error-prone task. The capacity of the Kinect sensor to recognize skeletal information only for objects in the range of 1,2 to 3,5 m may also have caused recognition problems. On the other hand, Berlin & John (2016) used Harris's corner point detectors differently, including the histogram form of the diverse images in order to recognize different activities performed in two sets with controlled environments. The results showed 95 and 88 % recognition rates for Set1 and Set2, respectively.

Conversely, Venkatesha & Turk (2010) attempted online human activity recognition, that is to say, without storing any video. The systems immediately learned the actions in the scene and classified them, considering the shape of human actions. They also used interest point extraction techniques while analyzing the histogram of the image in order to identify the action performed. This method showed a recognition effectiveness of 87 % in non-complex actions. Meanwhile, Peng *et al.* (2016) obtained a similar recognition rate, albeit combining local space-time characteristics and the construction of a visual dictionary, proposing a hybrid super vector.

Zhu *et al.* (2011) presented another technique based on recognizing an action through feature coding of local 3D space-time gradients within the framework of scattered code. By doing so, each space-time characteristic is transformed into a linear combination of some 'atoms' in a dictionary trained to detect local movement and appearance features. This method provides an increase in scale invariance, achieving the recognition of some basic activities. Considering the above studies and that proposed by H.-B. Zhang *et al.* (2019), local feature extraction does not require pre-processing activities such as background segmentation or human detection. It also offers scale invariance and rotation and is stable under lighting changes and more resistant to occlusion than global feature extraction.



S. Zhang *et al.* (2017) highlighted the fact that, although these detectors achieve satisfactory results in HAR, they have a significant deficiency: the calculation of stable interest points is often inadequate, as “discriminative” and “correct” interest points are difficult to identify. Similarly, H.-B. Zhang *et al.* (2019) faced some difficulties with the current local feature extraction method, as it is easily affected by changes in camera view, background movement, and camera movement.

## Depth-based feature extraction

The development of depth sensors such as the Microsoft Kinect (Shotton *et al.*, 2011) has allowed higher access to depth maps and the real-time position of skeletal joints, thus contributing to HAR via computer vision.

Various studies (X. Ma *et al.* (2014), Planinc & Kampel, 2013, Nizam *et al.* (2017), Mastorakis & Makris (2014), Yao *et al.*, 2017, Jalal *et al.* (2012)) have used the Kinect sensor as an information acquisition instrument and employed its depth images for HAR. The difference lies in the characteristics that each researcher wanted to extract. For example, X. Ma *et al.* (2014) conducted a complex study aiming to recognize six human actions (people falling, bending, sitting down, squatting, walking, and lying down) while combining global extraction techniques from depth images and analyzing changes in the human shape in short periods of time. On the other hand, (Nizam *et al.* (2017)) focused on extracting the elderly center of mass and added the angle between the human body and the floor plan. If this data is below specific thresholds, then a fall is detected.

On the other hand, Kwolek & Kepski (2014) complemented the use of depth images and calculated the distance from the human center of mass to the ground using an accelerometer for elderly fall detection. In this approach, if the acceleration exceeds a threshold value, it means that the person is in motion. At that moment, the depth sensor begins to extract information in order to detect a possible fall. However, the process requires calibrating the cameras and accelerometers, which increases its computational cost. Nizam *et al.* (2017) also used a Kinect sensor to study the speed and position of a person. Thus, if a high speed in a short time is detected, it is assumed that a fall has occurred. The fall is confirmed or discarded by analyzing the position of the body. This system has an average precision of 93,94 %.

Mastorakis & Makris (2014) attempted elderly fall detection by using a Kinect sensor to extract the environment’s 3D image, aiming to obtain a 3D bounding box surrounding the older person. Here, when the bounding box changes its width, height, and depth, the speed is analyzed. When the speed is higher than a certain threshold, it is considered that a fall has occurred. In turn, Yao *et al.*, 2017 used depth images to extract information such as the movement of the human torso, the 3D positions of the central hip joint, the central shoulder joint, and the height of a person’s centroid. With this method, a fall can be identified when the rates of the aforementioned characteristics reach threshold values. Despite this robust method, using only a Kinect makes the system dependent on the distance at which the sensor is working.

Unlike the aforementioned studies, [Jalal et al. \(2012\)](#) did not calculate the distance to the ground of any part of the human body. Instead, they combined the extraction of some global features such as depth data in order to recognize the elderly's daily activities. To this effect, R-transform was used to extract depth silhouettes of elderly body parts, and a hidden Markov model was subsequently used to train and recognize daily household activities. The results showed an average recognition rate of 96,55%.

According to [X. Ma et al. \(2014\)](#), light is not a problem when extracting silhouettes, since the Kinect sensor uses infrared light. This is very advantageous, as the sensor can also recognize human silhouettes in the dark and extract information from the human skeleton for HAR ([Yong Du et al., 2015](#)). This technique has been widely applied in different studies ([Keceli & Burak Can, 2013](#), [Pazhoumand-Dar et al., 2015](#), [X. Yang & Tian, 2014](#), [Hbali et al., 2018](#)). However, occlusion represents a problem with this approach, as recognition can be affected if the human body is occluded by any object. Therefore, several studies ([Ni et al., 2013](#), [Jalal et al., 2017](#), [Liu & Shao, 2013](#)) have merged spatial-temporal features with RGB cameras and depth data to reduce the occlusion problem. Data merging makes the processing volume larger, which increases feature dimensions. These factors increase the computational complexity of the algorithm for activity recognition.

## Convolutional neural networks

Finally, the current state of the art highlights the growing importance and impact of using convolutional neural networks (CNN) for HAR, as well as their classification and optimization in recent years. Different authors have adopted the use of CNN as a recognition method. For instance, [Hsieh & Jeng \(2018\)](#) applied a feedback CNN of optical flow to video transmission incorporating point estimation histograms, the limit of the object in motion, and limits of the subject in order to detect falls. Moreover, [Yan et al., 2018](#) proposed a novel model of dynamic skeletons called Spatial Temporal Graph Convolutional Networks. This model automatically learns the spatial and temporal patterns of data, which allows for a higher generalization capacity. [Xu et al. \(2020\)](#) also based their research on mapping the human skeleton to predict falls using OPENPOSE, thus obtaining a skeletal map and transforming it into a dataset to then feed the CNN. Other studies involving CNNs are based on the movement of a person. [Wang et al. \(2015\)](#) extracted the trajectory in a determined scenario while attempting to recognize and classify different activities. [Núñez-Marcos et al. \(2017\)](#) used optical flow images as the neural network's input, followed by a training phase to detect falls. Similarly, [Espinosa et al. \(2019\)](#) incorporates optical flow to a CNN that not only learns static information. CNNs have also been used in studies that incorporate depth maps from a Kinect sensor for fall detection ([Rah-nemoonfar & Alkittawi, 2018](#), [Adhikari et al. \(2017\)](#)). [Adhikari et al. \(2017\)](#) concludes that combining RGB image background subtraction and depth images with CNNs provides a possible solution to monitor falls based on indoor videos.

**Table 2.** Databases of human falls

Data base	Videos	Data provided	Environment	Population	Types of falls
URFD (Kwolek & Kepski (2014))	70 videos, 30 falls, and 40 daily activities	RGB depth images, images, and accelerometer signals	Indoors	Adult people	People falling while standing and sitting on a chair
LE2I (Charfi <i>et al.</i> , 2013)	191 videos: falls and daily activities	RGB images	Realistic indoors, home environments, and offices with variable lighting, occlusion, and cluttered and textured background	Adult people	Falls when walking, stumbling, and falling from chairs
CMDFALL (COMVIS-PTIT, n.d.)	600 videos with 20 human actions including falls	RGB images, depth images, and accelerometer signals	Home simulation indoors	30 men and 20 women between 21 and 40 years old	Falling backwards and forwards, to the left, and to the right
FALL-UP (Martínez-Villaseñor <i>et al.</i> , 2019)	361 videos including falls and daily activities	RGB images, accelerometer signals and different indoor images	Sensors	17 adults between 18 and 24 years old	Different falls
Multiple Cameras Fall Dataset (Auvinet (n.d))	192 videos including falls and daily activities	RGB images	Realistic and indoor home environments with occlusion, disorder, texture, variable lighting, and movement in the background	Adults	Backward falls, forward falls
UCF101 (Soomro <i>et al.</i> , 2012)	13.000 clips, 27 hours of video data with 101 human actions	RGB images	Controlled and realistic environments, moving cameras, and cluttered background	Adult people	Different falls

Source: Authors

Lu *et al.* (2017) used a three-dimensional convolutional neural network (3D-CNN) to extract the spatial characteristics of 2D images. It also incorporated video motion information to detect falls, thus reducing the failures caused by image noise, lighting variations, and occlusion. Similarly, C. Ma *et al.* (2019) incorporated a 3D-CNN, albeit hiding the facial regions optically perceivable in the video capture phase, thus helping to protect privacy while using surveillance cameras. Khraief *et al.* (2020) used a CNN with its own characteristics. In this study, the authors created a multi-stream CNN –a CNN with multiple flows. That is to say, four CNNs fed by the same images but extracting different features from them (color, texture, depth, shape, movement). Finally, they concatenated the four CNNs in order to obtain a unique classification of activities for fall detection.

A different CNN-based method to detect falls was proposed by Sreenidhi (2020), in which feature extraction from images of people falling was carried out. The CNN employed facial recognition because the author manifests that human expression when falling is highly distinguishable.

When working with CNNs, a large amount of data is required for training, which may be disadvantageous. For that reason, some authors (Cai *et al.*, 2019, Khraief *et al.* (2020), Li *et al.* (2018)) have used networks based on pre-trained architectures such as AlexNet, VGG16 Krizhevsky *et al.*, 2012, and ResNet He *et al.* (2016).

According to El Kaid *et al.* (2019), although the application of CNNs in activity recognition is successful, it has been done in very restricted environments. None of these networks are flexible enough to work well outside their domain. In this vein, the studies by Debard *et al.* (2016) and Fan *et al.* (2017) are concerned with the functioning of algorithms for detecting human actions, considering real falls, global and local feature extraction, and feature extraction through CNN.

Accordingly, the vast majority of studies focus on HAR using short video data segments captured in artificial environments, optimal conditions, and simulated falls by actors. Thereupon, Debard *et al.* (2016) selected algorithms with a good percentage of activity recognition when used with databases created in controlled or acted scenarios in order to implement them in real environments and falls of real adults. The authors concluded that said algorithms did not have the same efficiency, since they do not consider image quality, overexposure problems, occlusion, and changes in lighting conditions, thus demonstrating that not all the specifications for a robust system in real-world situations are met. Modern clustering techniques, such as the one proposed by Contreras-Contreras *et al.* (2022) could be applied to such databases.

## CONCLUSIONS

This work reviewed the progress made on human activity recognition with an emphasis on elderly falls, showing different devices and techniques to acquire data for subsequent processing and recognition. Furthermore, the main feature extraction methods used in the literature to detect human falls were presented.

Despite the fact that the vast majority of the proposed techniques have a high fall detection percentage and perform well in controlled environments, methods such as global feature extraction are highly sensitive to noise, occlusion, and viewpoint changes.

Similarly, local feature extraction shows a high deficiency when calculating correct interest points. Moreover, interest points are affected by changes in camera view, movement of the background, and camera movement. In addition, the main issue with extracting depth-based features is occlusion, as it may affect human skeleton extraction. Current applications of convolutional neural networks have been successful. However, they are placed in controlled and restricted environments, and their performance is not very good outside their domain.

This work demonstrates the importance of an efficient fall detection method, as well as the great potential of this research area going forward. Although good results are obtained by using the different techniques proposed by the authors mentioned in this paper, the environments where these techniques have been used are controlled, unrealistic, or use simulated falls, which does not contribute to the real advancement of this field. Therefore, presenting the results of these studies in environments similar to reality would have a positive impact, which is why studies that focus on elaborating databases with real falls of adults in non-controlled environments become essential.

## ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to Universidad del Cauca for supporting this research process.

## REFERENCES

- [Adhikari *et al.* (2017)] Adhikari, K., Bouchachia, H., & Nait-Charif, H. (2017, May 8-12). *Activity recognition for indoor fall detection using convolutional neural network* [Conference presentation]. 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). Nagoya, Japan. <https://doi.org/10.23919/MVA.2017.7986795> ↑Ver página 222
- [Akhavian & Behzadan, 2016] Akhavian, R., & Behzadan, A. H. (2016). Smartphone-based construction workers' activity recognition and classification. *Automation in Construction*, 71(Part 2), 198-209. <https://doi.org/10.1016/j.autcon.2016.08.015> ↑Ver página 215
- [Amiri *et al.* (2014)] Amiri, S. M., Pourazad, M. T., Nasiopoulos, P., & Leung, V. C. M. (2014). Improved human action recognition in a smart home environment setting. *IRBM*, 35(6), 321-328. <https://doi.org/10.1016/j.irbm.2014.10.005> ↑Ver página 217, 220

- [Auvinet (n.d)] Auvinet, E., Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (n.d.). *Multiple cameras fall dataset*. <http://www.iro.umontreal.ca/~labimage/Dataset/> ↑Ver página 219, 223
- [Auvinet *et al.*, 2011] Auvinet, E., Multon, F., Saint-Arnaud, A., Rousseau, J., & Meunier, J. (2011). Fall detection with multiple cameras: An occlusion-resistant method based on 3-D silhouette vertical distribution. *IEEE Transactions on Information Technology in Biomedicine*, 15(2), 290-300. <https://doi.org/10.1109/TITB.2010.2087385> ↑Ver página
- [Avci *et al.*, 2010] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., & Havinga, P. (2010, February 22-25). *Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey* [Conference presentation]. 23th International Conference on Architecture of Computing Systems, Hannover, Germany. <https://ieeexplore.ieee.org/document/5759000> ↑Ver página 215
- [Banos *et al.*, 2012] Banos, O., Damas, M., Pomares, H., Prieto, A., & Rojas, I. (2012). Daily living activity recognition based on statistical feature quality group selection. *Expert Systems with Applications*, 39(9), 8013-8021. <https://doi.org/10.1016/j.eswa.2012.01.164> ↑Ver página 215
- [Barbosa-Chacón *et al.*, 2013] Barbosa-Chacón, J. W., Barbosa-Herrera, J. C., & Rodríguez-Villabona, M. (2013). Revision y análisis documental para estado del arte: una propuesta metodológica desde el contexto de la sistematización de experiencias educativas. *Scielo Analytics*, 27, 83-105. [https://doi.org/10.1016/S0187-358X\(13\)72555-3](https://doi.org/10.1016/S0187-358X(13)72555-3) ↑Ver página 216
- [Ben Mabrouk & Zagrouba, 2018] Ben Mabrouk, A., & Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 91, 480-491. <https://doi.org/10.1016/j.eswa.2017.09.029> ↑Ver página 215
- [Berlin & John (2016)] Berlin, S. J., & John, M. (2016, October 24-27). *Human interaction recognition through deep learning network* [Conference presentation]. 2016 IEEE International Carnahan Conference on Security Technology (ICCST), Orlando, FL, USA. <https://doi.org/10.1109/CCST.2016.7815695> ↑Ver página 220
- [Brophy *et al.*, 2018] Brophy, E., Domínguez-Veiga, J. J., Wang, Z., & Ward, T. E. (2018, June 21-22). *A machine vision approach to human activity recognition using photoplethysmograph sensor data* [Conference presentation]. 2018 29th Irish Signals and Systems Conference (ISSC), Belfast, UK. <https://doi.org/10.1109/ISSC.2018.8585372> ↑Ver página 215
- [Cai *et al.*, 2019] Cai, X., Liu, X., Li, S., & Han, G. (2019, October 16-19). *Fall detection based on colorization coded MHI combining with convolutional neural network* [Conference presentation]. 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China. <https://doi.org/10.1109/ICCT46805.2019.8947223> ↑Ver página 224



- [Chakraborty *et al.*, 2012] Chakraborty, B., Holte, M. B., Moeslund, T. B., and González, J. (2012). Selective spatio-temporal interest points. *Computer Vision and Image Understanding*, 116(3), 396-410. <https://doi.org/10.1016/j.cviu.2011.09.010> ↑Ver página 220
- [Charfi *et al.*, 2013] Charfi, I., Miteran, J., Dubois, J., Atri, M., & Tourki, R. (2013). Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. *Journal of Electronic Imaging*, 22(4), 041106. <https://doi.org/10.1117/1.JEI.22.4.041106> ↑Ver página 223
- [Chen *et al.*, 2012] Chen, L., Nugent, C. D., & Wang, H. (2012). A knowledge-driven approach to activity recognition in smart homes. *IEEE Transactions on Knowledge and Data Engineering*, 24(6), 961-974. <https://doi.org/10.1109/TKDE.2011.51> ↑Ver página 215
- [COMVIS-PTIT, n.d.)] Computer Vision Department of the MICA International Research Institute & Posts & Telecommunications Institute of Technology (COMVIS-PTIT) (n.d.). Continuous multimodal multi-view dataset of human fall (CMD FALL). <https://www.mica.edu.vn/perso/Tran-Thi-Thanh-Hai/CMD FALL.html> ↑Ver página 223
- [Concone *et al.* (2019)] Concone, F., Re, G. Lo, & Morana, M. (2019). A fog-based application for human activity recognition using personal smart devices. *ACM Transactions on Internet Technology*, 19(2), 1-20. <https://doi.org/10.1145/3266142> ↑Ver página 218
- [Contreras-Contreras *et al.* (2022)] Contreras-Contreras, G. F., Medina-Delgado, B., Acevedo-Jaimes, B. R., & Guevara-Ibarra, D. (2022). Metodología de desarrollo de técnicas de agrupamiento de datos usando aprendizaje automático. *Tecnura*, 26(72), 42-58. <https://doi.org/10.14483/22487638.17246> ↑Ver página 224
- [Cosar *et al.*, 2017] Cosar, S., Donatiello, G., Bogorny, V., Garate, C., Alvares, L. O., & Bremond, F. (2017). Toward abnormal trajectory and event detection in video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3), 683-695. <https://doi.org/10.1109/TCSVT.2016.2589859> ↑Ver página 215
- [Das Dawn & Shaikh (2016)] Das Dawn, D., & Shaikh, S. H. (2016). A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector. *The Visual Computer*, 32(3), 289-306. <https://doi.org/10.1007/s00371-015-1066-2> ↑Ver página 218, 220
- [Debard *et al.* (2016)] Debard, G., Mertens, M., Deschodt, M., Vlaeyen, E., Devriendt, E., Dejaeger, E., Milisen, K., Tournoy, J., Croonenborghs, T., Goedemé, T. Tuytelaars, T., & Vanrumste, B. (2016). Camera-based fall detection using real-world versus simulated data: How far are we from the solution? *Journal of Ambient Intelligence and Smart Environments*, 8(2) 149-168. <https://doi.org/10.3233/AIS-160369> ↑Ver página 224

- [Durrant-Whyte *et al.*, 2012] Durrant-Whyte, H., Roy, N., & Abbeel, P. (2012). *Robotics: Science and Systems VII*. MIT Press. ↑Ver página 215
- [Efros *et al.* (2003)] Efros, Berg, Mori, & Malik. (2003, October 13-16). *Recognizing action at a distance* [Conference presentation]. 9th IEEE International Conference on Computer Vision, Nice, France. <https://doi.org/10.1109/ICCV.2003.1238420> ↑Ver página 219
- [El Kaid *et al.* (2019)] El Kaid, A., Baïna, K., & Baïna, J. (2019). Reduce false positive alerts for elderly person fall video-detection algorithm by convolutional neural network model. *Procedia Computer Science*, 148, 2-11. <https://doi.org/10.1016/j.procs.2019.01.004> ↑Ver página 224
- [Elbasiony & Gomaa,] Elbasiony, R., & Gomaa, W. (2020). A survey on human activity recognition based on temporal signals of portable inertial sensors. In A. E. Hassanien, A. T. Azar, T. Gaber, R. Bhatnagar, & M. F. Tolba (Eds.), *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)* (pp. 734-745). Springer. [https://doi.org/10.1007/978-3-030-14118-9\\_72](https://doi.org/10.1007/978-3-030-14118-9_72) ↑Ver página 215
- [Espinosa *et al.* (2019)] Espinosa, R., Ponce, H., Gutiérrez, S., Martínez-Villaseñor, L., Brieva, J., & Moya-Albor, E. (2019). A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset. *Computers in Biology and Medicine*, 115, 103520. <https://doi.org/10.1016/j.combiomed.2019.103520> ↑Ver página 222
- [Fan *et al.* (2017)] Fan, Y., Levine, M. D., Wen, G., & Qiu, S. (2017). A deep neural network for real-time detection of falling humans in naturally occurring scenes. *Neurocomputing*, 260, 43-58. <https://doi.org/10.1016/j.neucom.2017.02.082> ↑Ver página 224
- [Foroughi *et al.*, 2008] Foroughi, H., Aski, B. S., & Pourreza, H. (2008). *Intelligent video surveillance for monitoring fall detection of elderly in home environments* [Conference presentation]. 2008 11th International Conference on Computer and Information Technology, Khulna, Bangladesh. <https://doi.org/10.1109/ICCITECHN.2008.4803020> ↑Ver página 218, 219
- [Goudelis *et al.*, 2015] Goudelis, G., Tsatiris, G., Karpouzis, K., & Kollias, S. (2015). Fall detection using history triple features. In ACM (Eds.), *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '15* (art. 81). ACM Press. <https://doi.org/10.1145/2769493.2769562> ↑Ver página 218, 219
- [Han *et al.*, 2013] Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with Microsoft Kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5), 1318-1334. <https://doi.org/10.1109/TCYB.2013.2265378> ↑Ver página 215

- [Harris & Stephens (1988)] Harris, C., & Stephens, M. (1988). A combined edge and corner detector. In C. J. Taylor (Ed.), *Proceedings of the Alvey Vision Conference* (pp. 23.1-23.6). Alvey Vision Club. [↑Ver página 220](#)
- [Hassan *et al.* (2018)] Hassan, M. M., Uddin, M. Z., Mohamed, A., & Almogren, A. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, 81, 303-313. <https://doi.org/10.1016/j.future.2017.11.029> [↑Ver página 216](#)
- [Hbali *et al.*, 2018] Hbali, Y., Hbali, S., Ballihi, L., & Sadgal, M. (2018). Skeleton-based human activity recognition for elderly monitoring systems. *IET Computer Vision*, 12(1), 16-26. <https://doi.org/10.1049/iet-cvi.2017.0062> [↑Ver página 222](#)
- [He *et al.* (2016)] He, K., Zhang, X., Ren, S., & Sun, J. (2016, June 27-30). Deep residual learning for image recognition [Conference presentation]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. <https://doi.org/10.1109/CVPR.2016.90> [↑Ver página 224](#)
- [J.-W. Hsieh *et al.*, 2014] Hsieh, J.-W., Chuang, C.-H., Alghyaline, S., Chiang, H.-F., & Chiang, C.-H. (2014). Abnormal scene change detection from a moving camera using bags of patches and spiderweb map. *IEEE Sensors Journal*, 15(5), 2866-2881. <https://doi.org/10.1109/JSEN.2014.2381257> [↑Ver página 215](#)
- [Hsieh & Jeng (2018)] Hsieh, Y.-Z., & Jeng, Y.-L. (2018). Development of home intelligent fall detection iot system based on feedback optical flow convolutional neural network. *IEEE Access*, 6, 6048-6057. <https://doi.org/10.1109/ACCESS.2017.2771389> [↑Ver página 222](#)
- [Ismail *et al.*, 2015] Ismail, S. J., Rahman, M. A. A., Mazlan, S. A., & Zamzuri, H. (2015, October 18-20). *Human gesture recognition using a low cost stereo vision in rehab activities* [Conference presentation]. 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS), Langkawi, Malaysia. <https://doi.org/10.1109/IRIS.2015.7451615> [↑Ver página 215](#)
- [Jalal *et al.*, 2017] Jalal, A., Kim, Y.-H., Kim, Y.-J., Kamal, S., & Kim, D. (2017). Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognition*, 61, 295-308. <https://doi.org/10.1016/j.patcog.2016.08.003> [↑Ver página 222](#)
- [Jalal *et al.* (2012)] Jalal, A., Uddin, M. Z., Kim, J. T., & Kim, T.-S. (2012). Recognition of human home activities via depth silhouettes and  $\mathfrak{R}$  transformation for smart homes. *Indoor and Built Environment*, 21(1), 184-190. <https://doi.org/10.1177/1420326X11423163> [↑Ver página 221, 222](#)
- [Kahani *et a.*, 2019] Kahani, R., Talebpour, A., & Mahmoudi-Aznaveh, A. (2019). A correlation based feature representation for first-person activity recognition. *Multimedia Tools and Applications*, 78, 21673-21694. <https://doi.org/10.1007/s11042-019-7429-3> [↑Ver página 215](#)

- [Keceli & Burak Can, 2013] Keceli, A. S., & Burak Can, A. (2013, April 24-26). *Recognition of human actions by using depth information* [Conference presentation]. 2013 21st Signal Processing and Communications Applications Conference (SIU), Haspolat, Turkey. <https://doi.org/10.1109/SIU.2013.6531211> ↑Ver página 222
- [Khan *et al.* (2011)] Khan, Z. A., & Sohn, W. (2011). Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care. *IEEE Transactions on Consumer Electronics*, 57(4), 1843-1850. <https://doi.org/10.1109/TCE.2011.6131162> ↑Ver página 217, 218
- [Khan *et al.* (2013)] Khan, Z. A., & Sohn, W. (2013). A hierarchical abnormal human activity recognition system based on R-transform and kernel discriminant analysis for elderly health care. *Computing*, 95(2), 109-127. <https://doi.org/10.1007/s00607-012-0216-x> ↑Ver página 217, 218
- [Khraief *et al.*, 2019] Khraief, C., Benzarti, F., & Amiri, H. (2019). Convolutional Neural network based on dynamic motion and shape variations for elderly fall detection. *International Journal of Machine Learning and Computing*, 9(6), 814-820. <https://doi.org/10.18178/ijmlc.2019.9.6.878> ↑Ver página
- [Khraief *et al.* (2020)] Khraief, C., Benzarti, F., & Amiri, H. (2020). Elderly fall detection based on multi-stream deep convolutional networks. *Multimedia Tools and Applications*, 79, 19537-19560. <https://doi.org/10.1007/s11042-020-08812-x> ↑Ver página 224
- [Kim *et al.*, 2010] Kim, E., Helal, S., & Cook, D. (2010). Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, 9(1), 48-53. <https://doi.org/10.1109/MPRV.2010.7> ↑Ver página 215
- [Krizhevsky *et al.*, 2012] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). *ImageNet classification with deep convolutional neural networks* [Conference presentation]. 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf> ↑Ver página 224
- [Kwolek & Kepski (2014)] Kwolek, B., & Kepski, M. (2014). Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine*, 117(3), 489-501. <https://doi.org/10.1016/j.cmpb.2014.09.005> ↑Ver página 217, 221, 223
- [Laptev & Lindeberg (2003)] Laptev, I., & Lindeberg, T. (2003, October 13-16). *Space-time interest points* [Conference presentation]. Ninth IEEE International Conference on Computer Vision, Nice, France. <https://doi.org/10.1109/ICCV.2003.1238378> ↑Ver página 220

- [Laptev, 2005] Laptev, I. (2005). On space-time interest points. *International Journal of Computer Vision*, 64, 107-123. <https://doi.org/10.1007/s11263-005-1838-7> ↑Ver página 220
- [Lawrence *et al.*, 2010] Lawrence, E., Sax, C., Navarro, K. F., & Qiao, M. (2010, February 10-16). *Interactive games to improve quality of life for the elderly: Towards integration into a WSN monitoring system* [Conference presentation]. 2010 Second International Conference on EHealth, Telemedicine, and Social Medicine, Saint Marteen, Netherlands Antilles. <https://doi.org/10.1109/eTELEMED.2010.21> ↑Ver página 215
- [Li *et al.* (2018)] Li, H., Shrestha, A., Fioranelli, F., Kernec, J. Le, & Heidari, H. (2018, October 28-31). *Hierarchical classification on multimodal sensing for human activity recognition and fall detection* [Conference presentation]. 2018 IEEE SENSORS, New Delhi, India. <https://doi.org/10.1109/ICSENS.2018.8589797> ↑Ver página 215, 224
- [Li *et al.*, 2017] Li, X., Pang, T., Liu, W., & Wang, T. (2017, October 14-16). *Fall detection for elderly person care using convolutional neural networks* [Conference presentation]. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China. <https://doi.org/10.1109/CISP-BMEI.2017.8302004> ↑Ver página
- [Liu & Shao, 2013] Liu, L., & Shao, L. (2013). Learning discriminative representations from RGB-D video data. In F. Rossi (Ed.), *IJCAI '13: Proceedings of the Twenty-Third international joint conference on Artificial Intelligence* (pp. 1493-1500). ACM <https://dl.acm.org/doi/10.5555/2540128.2540343> ↑Ver página 222
- [Liu *et al.*, 2014] Liu, Y., Li, X., & Jia, L. (2014, June 29 - July 4). *Abnormal crowd behavior detection based on optical flow and dynamic threshold* [Conference presentation]. 11th World Congress on Intelligent Control and Automation, Shenyang, China. <https://doi.org/10.1109/WCICA.2014.7053189> ↑Ver página 215
- [Lohit *et al.* (2018)] Lohit, S., Bansal, A., Shroff, N., Pillai, J., Turaga, P., & Chellappa, R. (2018, June 18-22). *Predicting dynamical evolution of human activities from a single image* [Conference presentation]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA. <https://doi.org/10.1109/CVPRW.2018.00079> ↑Ver página 215
- [Lu *et al.* (2017)] Lu, N., Ren, X., Song, J., & Wu, Y. (2017, August 20-23). *Visual guided deep learning scheme for fall detection* [Conference presentation]. 2017 13th IEEE Conference on Automation Science and Engineering (CASE), Xi'an, China. <https://doi.org/10.1109/COASE.2017.8256202> ↑Ver página 222
- [C. Ma *et al.* (2019)] Ma, C., Shimada, A., Uchiyama, H., Nagahara, H., & Taniguchi, R. (2019). Fall detection using optical level anonymous image sensing system. *Optics & Laser Technology*, 110, 44-61. <https://doi.org/10.1016/j.optlastec.2018.07.013> ↑Ver página 224



- [X. Ma *et al.* (2014)] Ma, X., Wang, H., Xue, B., Zhou, M., Ji, B., & Li, Y. (2014). Depth-based human fall detection via shape features and improved extreme learning machine. *IEEE Journal of Biomedical and Health Informatics*, 18(6), 1915-1922. <https://doi.org/10.1109/JBHI.2014.2304357> ↑Ver página 217, 221, 222
- [Martínez-Villaseñor *et al.*, 2019] Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., & Peñafort-Asturiano, C. (2019). UP-Fall detection dataset: A multimodal approach. *Sensors*, 19(9), 1988. <https://doi.org/10.3390/s19091988> ↑Ver página 223
- [Mastorakis & Makris (2014)] Mastorakis, G., & Makris, D. (2014). Fall detection system using Kinect's infrared sensor. *Journal of Real-Time Image Processing*, 9(4), 635-646. <https://doi.org/10.1007/s11554-012-0246-9> ↑Ver página 221
- [Nguyen *et al.*, 2015] Nguyen, T. V., Song, Z., & Yan, S. (2015). STAP: Spatial-Temporal Attention-Aware Pooling for action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(1), 77-86. <https://doi.org/10.1109/TCSVT.2014.2333151> ↑Ver página 220
- [Nguyen *et al.*, 2016] Nguyen, V. A., Le, T. H., & Nguyen, T. T. (2016). Single camera based fall detection using motion and human shape features. In ACM (Eds.), *Proceedings of the Seventh Symposium on Information and Communication Technology - SoICT '16*. (pp. 339-344) ACM Press. <https://doi.org/10.1145/3011077.3011103> ↑Ver página 219
- [Ni *et al.*, 2013] Ni, B., Pei, Y., Moulin, P., & Yan, S. (2013). Multilevel depth and image fusion for human activity detection. *IEEE Transactions on Cybernetics*, 43(5), 1383-1394. <https://doi.org/10.1109/TCYB.2013.2276433> ↑Ver página 222
- [Nivia-Vargas & Jaramillo-Jaramillo (2018)] Nivia-Vargas, A. M., & Jaramillo-Jaramillo, I. (2018). La industria de sensores en Colombia. *Tecnura*, 22(57), 44-54. <https://doi.org/10.14483/22487638.13518> ↑Ver página 217
- [Nizam *et al.* (2017)] Nizam, Y., Mohd, M. N. H., & Jamil, M. M. A. (2017). Human fall detection from depth images using position and velocity of subject. *Procedia Computer Science*, 105, 131-137. <https://doi.org/10.1016/j.procs.2017.01.191> ↑Ver página 221
- [Núñez-Marcos *et al.* (2017)] Núñez-Marcos, A., Azkune, G., & Arganda-Carreras, I. (2017). Vision-based fall detection with convolutional neural networks. *Wireless Communications and Mobile Computing*, 2017, 9474806. <https://doi.org/10.1155/2017/9474806> ↑Ver página 222
- [OMS, 2015] OMS (WHO) (2015). *Datos interesantes acerca del envejecimiento*. <http://www.who.int/ageing/about/facts/es/> ↑Ver página



- [Panahi & Ghods (2018)] Panahi, L., & Ghods, V. (2018). Human fall detection using machine vision techniques on RGB-D images. *Biomedical Signal Processing and Control*, 44, 146-153. <https://doi.org/10.1016/j.bspc.2018.04.014> ↑Ver página 217
- [Pava *et al.*, 2021] Pava, R., Pérez-Castillo, J. N., & Niño-Vásquez, L. F. (2021). Perspectiva para el uso del modelo P6 de atención en salud bajo un escenario soportado en IoT y blockchain. *Tecnura*, 25(67), 112-130. <https://doi.org/10.14483/22487638.16159> ↑Ver página 215
- [Pazhoumand-Dar *et al.*, 2015] Pazhoumand-Dar, H., Lam, C.-P., & Masek, M. (2015). Joint movement similarities for robust 3D action recognition using skeletal data. *Journal of Visual Communication and Image Representation*, 30, 10-21. <https://doi.org/10.1016/j.jvcir.2015.03.002> ↑Ver página 222
- [Peng *et al.* (2016)] Peng, X., Wang, L., Wang, X., & Qiao, Y. (2016). Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Computer Vision and Image Understanding*, 150, 109-125. <https://doi.org/10.1016/j.cviu.2016.03.013> ↑Ver página 220
- [Planinc & Kampel, 2013] Planinc, R., & Kampel, M. (2013). Introducing the use of depth data for fall detection. *Personal and Ubiquitous Computing*, 17(6), 1063-1072. <https://doi.org/10.1007/s00779-012-0552-z> ↑Ver página 221
- [Preis *et al.*, 2012] Preis, J., Kessel, M., Werner, M., & Linnhoff-Popien, C. (2012). *Gait Recognition with Kinect*. [https://www.researchgate.net/publication/239862819\\_Gait\\_Recognition\\_with\\_Kinect/citations](https://www.researchgate.net/publication/239862819_Gait_Recognition_with_Kinect/citations) ↑Ver página 215
- [Rafferty *et al.*, 2017] Rafferty, J., Nugent, C. D., Liu, J., & Chen, L. (2017). From activity recognition to intention recognition for assisted living within smart homes. *IEEE Transactions on Human-Machine Systems*, 47(3), 368-379. <https://doi.org/10.1109/THMS.2016.2641388> ↑Ver página 215
- [Rahnemoonfar & Alkittawi, 2018] Rahnemoonfar, M., & Alkittawi, H. (2018, December 10-13). *Spatio-temporal convolutional neural network for elderly fall detection in depth video cameras* [Conference presentation]. 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA. <https://doi.org/10.1109/BigData.2018.8622342> ↑Ver página 222
- [Rosati *et al.*, 2018] Rosati, S., Balestra, G., & Knaflitz, M. (2018). Comparison of different sets of features for human activity recognition by wearable sensors. *Sensors*, 18(12), 4189. <https://doi.org/10.3390/s18124189> ↑Ver página 217
- [Rougier *et al.* (2007)] Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2007, May 21-23). *Fall detection from human shape and motion history using video surveillance* [Conference presentation]. 21st

- International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), Niagara Falls, ON, Canada. <https://doi.org/10.1109/AINAW.2007.181> ↑Ver página 219
- [Ryoo, 2011] Ryoo, M. S. (2011, November 6-13). *Human activity prediction: Early recognition of ongoing activities from streaming videos* [Conference presentation]. 2011 International Conference on Computer Vision, Barcelona, Spain. <https://doi.org/10.1109/ICCV.2011.6126349> ↑Ver página 215
- [Saini *et al.*, 2018] Saini, R., Kumar, P., Roy, P. P., & Dogra, D. P. (2018). A novel framework of continuous human-activity recognition using Kinect. *Neurocomputing*, 311, 99-111. <https://doi.org/10.1016/j.neucom.2018.05.042> ↑Ver página 215
- [Sazonov *et al.*, 2011] Sazonov, E., Metcalfe, K., Lopez-Meyer, P., & Tiffany, S. (2011, November 28 - December 1). *RF hand gesture sensor for monitoring of cigarette smoking* [Conference presentation]. 2011 Fifth International Conference on Sensing Technology, Palmerson North, New Zealand. <https://doi.org/10.1109/ICSensT.2011.6137014> ↑Ver página 215
- [Shotton *et al.*, 2011] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011, June 20-25). *Real-time human pose recognition in parts from single depth images* [Conference presentation]. CVPR 2011, Colorado Springs, CO, USA. <https://doi.org/10.1109/CVPR.2011.5995316> ↑Ver página 221
- [Soomro *et al.*, 2012] Soomro, K., Roshan, A., & Shah, M. (2012). UCF101: A Dataset of 101 human actions classes from videos in the wild. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1212.0402> ↑Ver página 223
- [Sreenidhi (2020)] Sreenidhi, I. (2020). Real-time human fall detection and emotion recognition using embedded device and deep learning. *International Journal of Emerging Trends in Engineering Research*, 8(3), 780-786. <https://doi.org/10.30534/ijeter/2020/28832020> ↑Ver página 224
- [Suto & Oniga, 2019] Suto, J., & Oniga, S. (2019). Efficiency investigation from shallow to deep neural network techniques in human activity recognition. *Cognitive Systems Research*, 54, 37-49. <https://doi.org/10.1016/j.cogsys.2018.11.009> ↑Ver página 215
- [Uzunovic *et al.*, 2018] Uzunovic, T., Golubovic, E., Tucakovic, Z., Acikmese, Y., & Sabanovic, A. (2018, October 21-23). *Task-based control and human activity recognition for human-robot collaboration* [Conference presentation]. IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington DC, USA. <https://doi.org/10.1109/IECON.2018.8591206> ↑Ver página 215

- [Venkatesha & Turk (2010)] Venkatesha, S., & Turk, M. (2010, August 23-26). *Human activity recognition using local shape descriptors* [Conference presentation]. 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey. <https://doi.org/10.1109/ICPR.2010.902> ↑Ver página 220
- [Vrigkas et al. (2015)] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2, 28. <https://doi.org/10.3389/frobt.2015.00028> ↑Ver página 215, 216
- [Wang et al. (2015)] Wang, L., Qiao, Y., & Tang, X. (2015, June 7-12). *Action recognition with trajectory-pooled deep-convolutional descriptors* [Conference presentation]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA. ↑Ver página 222
- [Xu et al. (2020)] Xu, Q., Huang, G., Yu, M., & Guo, Y. (2020). Fall prediction based on key points of human bones. *Physica A: Statistical Mechanics and Its Applications*, 540, 123205. <https://doi.org/10.1016/j.physa.2019.123205> ↑Ver página 222
- [Yan et al., 2018] Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *Computer Vision and Pattern Recognition*, 32(1), 12328. <https://doi.org/10.1609/aaai.v32i1.12328> ↑Ver página 222
- [L. Yang et al. (2016)] Yang, L., Ren, Y., & Zhang, W. (2016). 3D depth image analysis for indoor fall detection of elderly people. *Digital Communications and Networks*, 2(1), 24-34. <https://doi.org/10.1016/j.dcan.2015.12.001> ↑Ver página 217
- [X. Yang & Tian, 2014] Yang, X., & Tian, Y. (2014). Effective 3D action recognition using EigenJoints. *Journal of Visual Communication and Image Representation*, 25(1), 2-11. <https://doi.org/10.1016/j.jvcir.2013.03.001> ↑Ver página 222
- [Y. Yang et al., 2019] Yang, Y., Hou, C., Lang, Y., Guan, D., Huang, D., & Xu, J. (2019). Open-set human activity recognition based on micro-Doppler signatures. *Pattern Recognition*, 85, 60-69. <https://doi.org/10.1016/j.patcog.2018.07.030> ↑Ver página 215
- [Yao et al., 2017] Yao, L., Min, W., & Lu, K. (2017). A new approach to fall detection based on the human torso motion model. *Applied Sciences*, 7(10), 993. <https://doi.org/10.3390/app7100993> ↑Ver página 221
- [Yong Du et al., 2015] Yong Du, Wang, W., & Wang, L. (2015, June 7-12). *Hierarchical recurrent neural network for skeleton based action recognition* [Conference presentation]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA. <https://doi.org/10.1109/CVPR.2015.7298714> ↑Ver página 222

- [Yu *et al.*, 2012] Yu, M., Naqvi, S. M., Rhuma, A., & Chambers, J. (2012). One class boundary method classifiers for application in a video-based fall detection system. *IET Computer Vision*, 6(2), 90-100. <https://doi.org/10.1049/iet-cvi.2011.0046> ↑Ver página 218, 219
- [Yu *et al.*, 2013] Yu, M., Yu, Y., Rhuma, A., Naqvi, S. M. R., Wang, L., & Chambers, J. A. (2013). An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment. *IEEE Journal of Biomedical and Health Informatics*, 17(6), 1002-1014. <https://doi.org/10.1109/JBHI.2013.2274479> ↑Ver página 217, 218, 219
- [H.-B. Zhang *et al.* (2019)] Zhang, H.-B., Zhang, Y.-X., Zhong, B., Lei, Q., Yang, L., Du, J.-X., & Chen, D.-S. (2019). A comprehensive survey of vision-based human action recognition methods. *Sensors*, 19(5), 1005. <https://doi.org/10.3390/s19051005> ↑Ver página 220, 221
- [S. Zhang *et al.* (2017)] Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, 2017, 3090343. <https://doi.org/10.1155/2017/3090343> ↑Ver página 218, 219, 220
- [Zhu *et al.* (2011)] Zhu, Y., Zhao, X., Fu, Y., & Liu, Y. (2011). Sparse coding on local spatial-temporal volumes for human action recognition. In R. Kimmel, R. Klette, & A. Sugimoto (Eds.), *Computer Vision - ACCV 2010* (pp. 660-671). Springer. [https://doi.org/10.1007/978-3-642-19309-5\\_51](https://doi.org/10.1007/978-3-642-19309-5_51) ↑Ver página 220

