Human Body Posture Recognition Approaches: A Review

Mohammed A. Ali[1], Abir J. Hussain[2], Ahmed T. Sadiq[1]

[1]Department of Computer Sciences, University of Technology,
Baghdad, Iraq
[2]School of Computer Sciences and Mathematics, Liverpool John Moores University,
Liverpool, England

Abstract—**Human body posture recognition has become the focus of many researchers in recent years. Recognition of body posture is used in various applications, including surveillance, security, and health monitoring. However, these systems that determine the body's posture through video clips, images, or data from sensors have many challenges when used in the real world. This paper provides an important review of how most essential hardware technologies are used in posture recognition systems. These systems capture and collect datasets through accelerometer sensors or computer vision. In addition, this paper presents a comparison study with state-of-the-art in terms of accuracy. We also present the advantages and limitations of each system and suggest promising future ideas that can increase the efficiency of the existing posture recognition system. Finally, the most common datasets applied in these systems are described in detail. It aims to be a resource to help choose one of the methods in recognizing the posture of the human body and the techniques that suit each method. It analyzes more than 80 papers between 2015 and 2020.**

Index Terms—**Acceleration based, Computer vision, Health monitoring, Human body posture recognition, Security.**

I. Introduction

Human body posture recognition (HBPR) is an essential topic in modern technology and focuses on many researchers in computer science and engineering (Patel, Bhatt, and Patel, 2017). Artificial intelligence algorithms are used in these technologies to recognize the position of the human body. In computer vision and electronic devices, such as sensors and smartphones used in many applications (Hameed, Alwan, and Ateia, 2020), the human body's posture is critical. Due to the rapid advancement of image processing and other technologies, HBPR is used in a range of applications and tests, including computer-based intelligent video surveillance and pattern recognition (Lo Presti and La Cascia, 2016; Hsiao, et al., 2018; Zhang, Wu, and Wang, 2020; Wan, et al., 2020).

ARO-The Scientific Journal of Koya University
Vol. X, No.1 (2022), Article ID: ARO.10930, 10 pages
DOI: 10.14500/aro.10930
Received: 26 January 2022; Accepted: 17 May 2022
Regular research paper: Published: 13 June 2022
Corresponding author's e-mail: cs.19.21@grad.uotechnology.edu.iq.iq

In recent years, deep learning technology has been used to achieve successful results in a variety of tasks (Ghafoor, et al., 2021), including image object recognition (Zhao, et al., 2019), classification (Rawat and Wang, 2017), and semantic segmentation (Sun and Wang, 2018), and so on. The use of deep learning technology in HBPR has resulted in rapid growth. The advances were made on engineered networks with excellent estimation capacity, rich datasets (Colleges, et al., 2014; Joo, et al., 2019; Mehta, et al., 2018) for feeding networks, and a more realistic exploration of body models (Kanazawa, et al., no date). Even though there is some current interest in HBPR, there is still a need to summarize the most recent deep learning-based research.

HBPR systems usually use both supervised learning and unsupervised learning. Unsupervised systems rely on rules, while supervised systems use algorithms trained in advance with unique datasets (Gupta, et al., 2020).

The review paper classified HBPR systems into two types: Acceleration-based and vision-based methods (Hsiao, et al., 2018; Abedi, et al., 2020). Fig. 1 shows classification of data capture hardware for human body posture. Posture recognition systems perform sequential data collection from wearable devices or cameras (Zhang and Callaghan, 2020). Acceleration-based methods are divided into two hardware types: Wearable sensors and smartphone devices. The sensors require users to wear multiple accelerometers to collect data (Gupta, et al., 2020; Mo, et al., 2016). Some methods rely on the input data from smartphone sensors such as gyroscopes and accelerometers. These devices recognize the positions of the human body using built-in or stand-alone sensors (Gupta, et al., 2020; Khokhlov, et al., 2018; Voicu, et al., 2019). Many designs, from fitness bands with very limited smartphones to larger implementation platforms, including desktop computers (Khokhlov, et al., 2018). In the final step, they are categorized using the model estimation of the main points of the body through the neural network (Gupta, et al., 2020). Collecting data in the vision-based method utilize one or more conventional cameras and specialized cameras for depth images (Mirza and Al-Talabani, 2021). The vision-based system has the benefit of not requiring users to wear any sensors, but its performance is heavily influenced by visual angle, lighting conditions, and other environmental variables. On the other hand, the accelerometer-based

http://dx.doi.org/10.14500/aro.10930

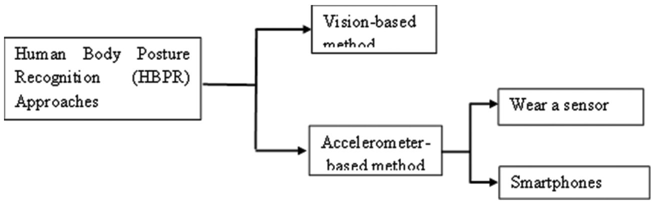

Fig. 1. Classification Hardware to capture and collect dataset.

method requires users to wear a sensor, but it eliminates any external interference. Most cell phones now have a built-in accelerometer, allowing users to construct an accelerometer-based posture system on the mobile platform without the use of extra equipment (Chen and Xue, 2016).

Several reviews have been published during the past decade to summarize the related work on human pose estimation (HPE). For instance, Liu, et al., 2015, a study of HPE based on body parts parsing from numerous input sources, multiple view and single view. Zhang, et al., 2016, discuss the challenges of estimating human pose and include a data-driven analysis of recent approaches to estimating human pose, including depth-based approaches and conventional image-based methods. This research focuses on methods that use RGB and depth image data. Another aspect is presented in the review (Lo Presti and La Cascia, 2016). It highlighted the 3D skeleton-based action classification methods and the new research in this field. Sarafianos, et al., 2016 another paper showed a review of 3D HPE approaches using various inputs (e.g., multiview or monocular, film, or single image). Gong, et al., 2016, this paper provided a review of traditional HPE approaches focused on monocular vision with a few deep learning-based methods. Patel et al., 2017, propose a review discussing how various steps of the human body pose are used by showing various methods used for each step of the system. (Wu, Sharma and Blumenstein, 2017) The three types of datasets were shown, and various state-of-the-art deep learning-based strategies for human behavior detection were proposed (Ben Mabrouk and Zagrouba, 2018). Another paper reviewed the two key measures that make up a video monitoring scheme, behavior representation and behavior simulation. Wang, et al., 2018, this article showed a survey of four types of red green blue-depth (RGB-D) motion recognition: RGB-D based, skeleton based, depth based, and RGB based. Dang, et al., 2019, this review illustrated methods for predicting human pose based on deep learning and their methodologies. Zhang, et al., 2019, covered the latest developments of hand-designed action features in RGB and depth details, modern deep learning-based action feature representation methods, products in human-object interaction recognition, and the emerging technologies hot topic of action detection methods. Chen, Tian, and He, 2020, this paper reviewed four types of monocular HPE analysis and human pose datasets focused on deep learning. The review of Gupta, et al., 2020, concentrated on several recent academic articles on multiple behavior recognition methods, wearable devices, smartphone sensors, and vision-based tools used to recognize.

The main contributions of this review paper are as follows:

- A study of devices used in HBPR systems also identifies each device's limitations and advantages
- Review of essential methods used in HBPR systems
- Classification of a dataset used in HBPR systems into several categories, depending on the hardware used.

The rest of this paper is organized as follows. Section II contains a literature review and compares procedures with practices and their precision. Section III includes the most used datasets and a short description of the information they contain. A discussion of the research regarding its benefits and drawbacks is included in Section IV. Finally, the conclusions and future directions of promising studies that could improve the effectiveness of HBPR appear in Section V.

II. LITERATURE REVIEW

This paper discusses many methods to discern the human body's pose, whether in cameras, sensors, or smartphones.

A. Vision-based Methods

HBPR has become a common research subject in computer vision (Ding, et al., 2020, Abdulhussein and Raheem, 2020). Posture information is helpful for tasks such as activity detection and content extraction. The method of inferring the locations of 2D or 3D human body parts from still photographs or videos is known as human pose recognition (HPR) (Dang, et al., 2019; Gupta, et al., 2020). Traditional and depth cameras are used to track human body identification in various settings, including sports stadiums, indoor and outdoor events, shopping centers, educational facilities, hospitals, and highways (Zhang and Callaghan, 2020). Estimating the human experience is a challenging problem in computer vision, but it has many real-world applications (Dang, et al., 2019).

From a single image or video, 2D HPR calculates the positions of human joints. HPR 2D algorithms traditionally focus on handcraft feature extraction and sophisticated body models to obtain local representations and global positional structures before deep learning can have a significant effect on vision-based human posture appreciation (Zhang, Wu, and Wang, 2020; Yan, Coenen and Zhang, 2016; Zhou and Zhang, 2020). Single individual posture estimation and multipersonal posture estimation are two types of contemporary 2D human postural estimation approaches based on deep learning (Chen, Tian, and He, 2020; Zhang, Wu, and Wang, 2020).

Color and depth images, as well as human skeletal records, are all readily accessible to researchers. Several location recognition algorithms have been suggested using skeletal data derived from the Kinect (Rahmani and Bennamoun, 2017; Liu, et al., 2020). These algorithms reduce the effect of poor lighting. They also remove the need for pre-processing, such as segmentation and object detection in complex backgrounds, allowing for improved image quality (Ding, et al., 2020).

A photograph or other input form can be used to generate a 3D model of human body joints, known as 3D human pose estimation. Commercial technologies like Kinect with a depth sensor, Vicon with an optical sensor, and the Captury with multiple cameras have been used for 3D body pose estimation, although these systems only work in extremely limited contexts or require unique markers on the human body to perform (Wu, Sharma and Blumenstein, 2017; Abdulhussein and Raheem, 2020). The three depth sensor-based human labor identification forms are osteoid data, depth image, and methods based on depth and skeleton (Rahmani and Bennamoun, 2017).

B. Acceleration-based Methods

Smartphones are the most helpful tool in our daily lives, and modern technology helps them consistently meet their user's needs and wishes. Smartphone designers continually add new features to the devices to make these systems more usable and functional. Sensors enhance mobile capabilities and play an essential role in environmental awareness. As a result, most smartphones come equipped with a range of sensors to capture a wealth of information about everyday activities. Sensors collect data from body movements and then identify behaviors (Ding, et al., 2020; Gupta, et al., 2020). Accelerometers and gyroscopes are the most common (Kareem, Ali and Jasim, 2020). The accelerometer sensor monitors changes in motion, whereas the gyroscope is used to track the object's orientation. Most commercial smartphones contain at least two, a microsystem or a micro-electro-mechanical system (MEMS; Nandy, Saha and Chowdhury, 2020; Khokhlov, et al., 2018). MEMS-based accelerometers are more practical than mechanical accelerometers. People currently wear or carry these smartphones and wearables with built-in MEMS sensors. These sensors allow the use of many fascinating lifestyle tracking applications. Since smartphones include sensors (gyroscope, accelerometer, compass, etc.) as well as networking capabilities (such as Wi-Fi and Bluetooth), the data gathered from the sensors may be sent to a remote server for analysis and classification (Nandy, Saha and Chowdhury, 2020). Google Fit, Samsung Fitness, and Noom Walk are activity recognition applications for Android OS mobile devices. Likewise, Human Activity Tracker and Pacer are activity recognition apps for iOS devices that gather data from built-in sensors for health-care purposes (Khokhlov, et al., 2018; Wan, et al., 2020; Nandy, Saha and Chowdhury, 2020; Voicu, et al., 2019) support vector machine, k-nearest neighbors algorithm (k-NN), Bootstrap aggregating (Bagging), and Adaptive Boosting (Ada Boost) are some of the human action recognition innovations used by smartphones in recent studies (Ding, et al., 2020; Gupta, et al., 2020).

On the other hand, wearable technology collects data from sensors connected to the subject, used for continuous monitoring. Since human activity includes various physical motions, the recognition of human activity requires data from several sensors mounted on various areas of the person's body (Gupta, et al., 2020). Furthermore, the use of different sensors is now standard as a plug-and-play option. For example, sensors, such as temperature, tone, light, and potentiometers, may be mounted directly on the boards of popular Arduino or Raspberry Pi computers (Vecchio, Mulas, and Cola, 2017). Furthermore, the number of premium sensor products available is constantly increasing, enabling consumers to track various conditions such as air quality, carbon monoxide, ambient pressure, capacity, humidity, gas leakage, and hydrogen (Kamiŝalić, et al., 2018). Several articles have been published on multiple human activity detection technologies using body sensors or wearable sensors in several environments, including senior centers, detox centers, and systems for identifying mental illnesses (Hsiao, et al., 2018; Wu, Sharma, and Blumenstein, 2017). Motion sensors, such as acceleration, video, proximity, and other wearable sensors, are examples of body sensors. These sensors commonly track driving, sitting, biking, standing, group meetings, reading, and other body movements.

The design of wearable sensors must consider user compatibility. Activity tracking sensors must be light and comfortable. Multiple datasets use activity tracking sensors. Human activities can be classified using statistics after function extraction and modeling, and a machine learning algorithm is implemented. Activity identification requires translating low-level sensor data to higher-level abstractions (Sun and Wang, 2018). Vector machine-based classification, neural network-based classification, and pattern mating-based classification are the most common algorithms (Neili, et al., 2017; Gouiaa and Meunier, 2017; Hassan, et al., 2018).

Table I is a comparison of previous reviews with this review. The comparison was based on the topics covered and studied.

Methods of collecting and capturing image data or other types impact many factors in choosing to build a human body posture recognition system and choosing algorithms and techniques in classification. Table II analyzes the accuracy of these algorithms depending on the type of data used.

III. DATASET

The dataset has a fundamental role in Human Body Posture Recognition. HBPR, whether through computer vision-based systems or acceleration-based systems using deep learning, requires a large dataset. The dataset size is critical, but the learning improvement can be costly for a large dataset. Each method of determining the body's activities has a specific dataset that depends on the application and the devices used to capture the input. Table III presents the most common datasets available to users that are carefully collected to be trainable and testable.

HMDB51 and UCF 101 are the most often utilized RGB datasets for evaluating proposed methods. These two datasets have been used in virtually all studies of current deep learning-based techniques to test algorithm efficacy. However, the RGBD and skeleton datasets have seen less use in deep learning-based algorithms than the RGB dataset. The fact that these datasets are small scale is one of the primary reasons. Deep learning-based techniques for depth and skeleton data are becoming major research subjects, thanks to the emergence of large scale and complex RGBD and skeleton datasets, such as the NTU RGB+D dataset (Zhang, et al., 2019).

This review paper will divide the dataset according to the type of camera, the samples were collected, RGB images,

 http://dx.doi.org/10.14500/aro.10930

and depth images collected by depth cameras.

A. RGB Dataset

1. Writing posture dataset (WPD) (Cao, et al., 2017): The WPD was gathered in a laboratory setting with a calibrated binocular camera, and RGB video was captured for each subject. It contains 113,400 pictures taken from 30 people of various genders and heights. For each frame, skeleton data and a position category are supplied
2. MPII (Andriluka, et al., 2014): The dataset is from YouTube videos. It includes 410 human activities, and each image is provided with an activity label
3. Southeast University SEU: This information was initially compiled by Zhao, et al., 2012. Each video in the collection was captured using a side-mounted Logitech C905 CCD camera in daylight circumstances. The dataset was created with the help of 10 male and 10 female drivers. Each film was shot in natural light during the day
4. HMDB51 (Zhao, et al., 2012): The dataset contains many real videos culled from various sources, including films and web videos. There are 6849 video clips in the collection, divided into 51 activity types. Again, they were gathered from various sources (public databases, movies such as YouTube, Google videos, and Prelinger archive)
5. Hollywood2 (Marszałek, Laptev and Schmid, 2009): The dataset consists of 3669 video clips and about 20 of 1 h of the film, with 12 classes of human activities and ten scenery classes. The collection contains video snippets from 69 films. It was suggested that realistic and complex environments be provided (cluttered background, multiple persons.etc.)
6. Olympic (Niebles, Chen and Fei-Fei, 2010): The Olympic Sports dataset includes videos of athletes participating in various sports. It collected all its video sequences from YouTube and used Amazon Mechanical Turk to annotate its class labels
7. UCF101: It was produced by the center for research in Computer Vision, University of Central Florida, the USA, in 2012 (Soomro, Zamir and Shah, 2012). It is an expansion of the UCF50 dataset (Reddy and Shah, 2013), which comprises 50 activity types. It is made of 13,320 videos of 101 realistic action categories taken from YouTube. It delivers the most significant variation in activities and realistic situations (viewpoint, illumination conditions.etc.) (Fig. 2a).

Table I
Summary and Comparison of the Related Review of Human Body Posture Recognition

No.	Survey, review, and reference	Topic/main focus
	A survey of human pose estimation: The body parts parsing based methods (Liu, et al., 2015)	The recent advances in vision-based human pose estimation
	A survey on human pose estimation (Zhang, et al., 2016)	Discuss the difficulties of human pose estimation and give a data-driven overview of recent approaches
	3D skeleton-based human action classification: A survey (Lo Presti and La Cascia, 2016)	Highlights the motives and challenges of this relatively new research field
	3D human pose estimation: A review of the literature and analysis of covariates (Sarafianos, et al., 2016)	Various inputs (e.g., multiview or monocular, film, or single image)
	Human pose estimation from monocular images: A comprehensive survey (Gong, et al., 2016)	Monocular vision with a few deep learning-based methods
	Human body posture recognition – A survey (Patel, Bhatt and Patel, 2017)	A different phase of human body posture
	Recent advances in video-based human action recognition using deep learning: A review (Wu, Sharma and Blumenstein, 2017)	Three types of datasets, various state-of-the-art deep learning-based strategies
	Abnormal behavior recognition for intelligent video surveillance systems: A review (Ben Mabrouk and Zagrouba, 2018).	Two key measures that make up a video monitoring scheme, behavior representation, and behavior simulation
	RGB-D-based human motion recognition with deep learning: A survey (Wang, et al., 2018)	There are four types of RGB-D motion recognition: RGB-D based, skeleton based, depth based, and RGB based
	Deep learning-based 2D human pose estimation: A survey (Dang, et al., 2019)	Methods for predicting human pose based on deep learning and the methodologies used are discussed
	A comprehensive survey of vision-based human action recognition methods (Zhang, et al., 2019)	Review of human action recognition methods and provide a comprehensive overview of recent approaches in human action recognition research
	Monocular human pose estimation: A survey of deep learning-based methods ours (Chen, Tian and He, 2020)	Four classes of deep learning-based monocular HPE analysis and human pose datasets
	A survey on human activity recognition and classification (Gupta, et al., 2020)	Various methods of activity recognition
	Our review	Two famous techniques are capturing and collecting a dataset of human body posters and analyzing modern classification methods

B. Depth and Skeleton Dataset

1. MSR-Action3D (Li, Zhang and Liu, 2010): A depth sequence action dataset is recorded by a depth camera. Wanqing Li, a researcher at Microsoft Research Redmond, developed it. It contains 567 depth map sequences of 10 people doing 20 different action types twice or 3 times (Fig. 2b)
2. UTKinect-Action (Xia, Chen and Aggarwal, 2012): The UT-Kinect dataset is a depth sequence action recognition dataset. There are 10 different sorts of actions. RGB, depth, and skeleton joint positions were captured in three channels
3. Northwestern-UCLA (Wang, et al., 2014): The multiview 3D event dataset comprises RGB, depth, and human skeletal data simultaneously recorded by three Kinect cameras. There are 10 activity categories in this dataset. A group of 10 performers carries out each action. This dataset includes data from several perspectives
4. NTU RGB+D datasets (Shahroudy, et al., 2016): NTU RGB+D is a large-scale dataset recognizing RGB-D human actions. It includes 56,880 samples from 40 topics representing 60 action classes. There are three types of activities, 40 everyday acts, nine health-related actions, and

Table II
Systematic Analysis of Human Body Posture Recognition Algorithm Based on Prior Research

Paper	Feature	Device	Type image	Method	Dataset	Accuracy (%)
Ding, et al., 2020	Skeleton	Kinect	RGB-D image	Bagging approach and RIPPER rule learning algorithm	MSR-Action3D	94.5
					Microsoft MSRC-12	97.6
					UTKinect- Action	98.1
					Baduanjin posture (local dataset)	99.6
El Amine Elforaici, et al., 2018	3D skeleton	Kinect2	RGB-D image	(CNNs) + SVM classification	Local dataset	
					RGB	93.3
					Depth	95.7
Hsiao, et al., 2018	Spatial features or body part features.	FSR sensors and infrared array sensors	Pressure sensor data+infrared sensor data	Fuzzy c-means clustering algorithm	Local dataset	88
Zhang, Wu and Wang, 2020	Spatiotemporal information	RGB camera	RGB images	Improved two-branch multistage convolutional neural network (M-CNN)	Postures of fall	98.70
Wan, et al., 2020	-	Smartphone	Acceleration data	Proposed CNN model	UCI	92
					Pamap2 datasets	91
Chen and Xue, 2016	-	Android phone	Acceleration data	CNN modified	31,688 samples from 100 subjects	93.8
Zhang, Yan and Li, 2018	3D body skeletons	RGB camera	RGB images	Multistage CNN architecture	ARM	94.9
					Back	93.9
					Legs	94.6
Huang, et al., 2019	-	Ultra-wide band	Acceleration data	Least square estimation (LSE) method and the improved extended Kalman filtering (iEKF) algorithm	-	-
Kale, et al., 2018	-	Wirelessly (Wi-Fi) acquiring	Sensor data	Artificial neural networks	Dataset of 44,800 samples	97.58
Ren, et al., 2020	-	Kinect	RGB-D image	Hybrid fuzzy logic and machine learning approach	A small dataset containing 19,800	97.1
Liu, et al., 2020	Gaussian voxel feature	RGB camera	RGB images	3D CNN	WPD	97.77
					BPD	98.16
Quan, et al., 2019	Forward kinematics model of the human skeleton	ASUS Xtion PRO	RGB-D image	Unsupervised learning algorithms, such as GNG and PSO	No dataset used	-
Neili, et al., 2017	Skeleton data	-	RGB-D	ConvNets and SVM classifier	CAD60	99.66
Gouiaa and Meunier, 2017	Shadow images	Camera and two infrared light sources	Synthetic data	CNN	Dataset captured in laboratory	99
Liang and Hu, 2020	Coordinate threshold of joint points	Kinect V1	RGB image	Deep neural network ResNet	MPII	91
Yan, Coenen and Zhang, 2016	Hand position	RGB camera	RGB image	CNN	SEU dataset	99.47
					Driving-Posture-atNight dataset	99.3
					Driving-Posture-inReal dataset	95.77
Zhou and Zhang, 2020	Appearance, audio, and skeleton	RGB camera	RGB image	SVM	HMDB51	85
					UCF	81
					Hollywood2	74
					Olympic	76
Jalal, et al., 2020	Extract full-body human silhouette (energy features) and key points context-aware features		RGB	Ray optimization+K-Ary tree hashing	UCF50	80.9
					Hmdb51	82.1
					Olympic	90.83.
Liu, Liu and Chen, 2017	Sequence-based transform	Kinect V1	RGB+D	CNN	Northwestern-	92.61
					UCLA	73.8
					NTU RGB+D datasets	87.21
Liu, *et al*., 2019	Skeleton posture	Kinect V2.0	RGB+D	Posture-CNN model	1620 pose images (including six postures)	99.01

(Contd...)

Table II
(Continued)

Paper	Feature	Device	Type image	Method	Dataset	Accuracy (%)
Rahmani, Mian and Shah, 2018	Dense trajectories from videos+Dense trajectories from mocap sequences	-	Synthetic data	Robust non-linear knowledge transfer model	INRIA Xmas Motion Acquisition Sequences	74.1
					N-UCLA Multiview dataset	78.1
					UWA3D Multiview Activity II Dataset	67.4
Kamel, et al., 2019	-		RGB-D	CNN	MSRACTION3D DATASET	94.51
					UTD-MHAD DATASET	88.14
Wang and Liu, 2020	Direction cosine method	Kinect depth sensor	RGB-D	BP neural network		Reach more than 90
Xiao, Cui and Li, 2020	Three-level spatial pyramid feature extraction	-	RGB	Dual aggregation deep network and CXQDA distance metric	UCF50	62
					HMDB51	66
					Hollywood2	92
					Olympic	89
Zhao and Obonyo, 2020	-	Wearable Inertial Measurement Units (IMUs) sensors	-	CNN+LSTM	Datasets collected from four workers on construction sites	-
Zhou, et al., 2018	Spatiotemporal information	-	RGB	Two-stream MiCT-Net	UCF101	94.7
					HMDB51	70.5
Santhoshkumar and Kalaiselvi Geetha, 2019	-	-	RGB	FDCNN	(University of York) and GEMEP dataset	95.4
Cao, et al., 2017	-	-	RGB	Greedy parsing algorithm	MPII	79.7
					COCO	84.9
Wang, et al., 2016	Upper and lower body ratio of the human silhouette	Kinect	Depth images	LVQ neural network	-	97
Kołodziej, et al., 2019	Statistical parameters	Smartphone,	Acceleration signals	Quadratic discriminant analysis (QDA)	-	90–95
Kim and Lee, 2020	-	-	RGB	Proposed a lightweight stacked hourglass network	MPII	90.8
					LSP	91.7
Saini, et al., 2019	3D skeleton	Kinect	-	Bidirectional long short-term memory neural network (BLSTM-NN)	KARD dataset	96.67
					CAD-60 dataset	79.58
Ning, Zhang and He, 2018	-	-	-	Stacked hourglass design and inception-ResNet module	MPII	91.2
					LSP	93.9
Lei et al. (Zhao and Chen, 2020)	Time-domain features+Frequency domain features	Inertial sensor	Sensor- data	SVM	-	96
Zhang, et al., 2017	-	Smartphone, RFID		A hierarchical algorithm with backward reasoning, proposal (multifusion)	-	85.7
Nandy, Saha and Chowdhury, 2020	-	Smartphone	-	Ensemble of classifiers	-	96
Voicu, et al., 2019	-	Smartphone	-	Neural network	-	93

SVM: Support vector machine

11 joint actions.

IV. Discussion

This review examined different techniques for recognizing body position and analyzed many papers, results, and algorithms used for each technique. In addition, we mentioned the details of the dataset that is used with each of the techniques. Despite significant advances in body location recognition, there are limitations to the techniques that make them ideal only for specific applications.

We mention and analyze the limitations and benefits in this section. The advantages of vision-based systems are that they are easy to use and do not require cooperation from those watched. Furthermore, they provide the capacity to understand dynamic systems and high-level operations (Wu, Sharma, and Blumenstein, 2017). The disadvantages of vision-based systems include the camera position, the background clutter, and limited coverage. Furthermore, there is a vast range of additional criteria, necessitating the production of

Fig. 2: The samples from dataset: (a) UCF101, RGB image dataset, (b) MSR-Action3D, depth image dataset.

Table III
The Popular Datasets of Human Body Posture Recognition. the (✓) Mark Represents Finding a Feature in the Dataset, but the (X) Mark Means it is Not Found

Dataset name	Color	Depth	Skeleton	Samples	Classes
MSR-Action3D (Li, Zhang and Liu, 2010)	X	✓	✓	3224	20
UTKinect-Action (Xia, Chen and Aggarwal, 2012)	✓	X	✓	3795	10
writing posture dataset (WPD) (Cao, et al., 2017)	✓	X	✓	113,400	15
MPII (Andriluka, et al., 2014)	✓	X	X	24,984	20
Southeast University SEU (Zhao, et al., 2012)	✓	X	X	24,210	4
HMDB51 (Zhao, et al., 2012)	✓	X	X	6849	51
Hollywood2 (Marszałek, Laptev and Schmid, 2009)	✓	X	X	3669	12
Olympic (Niebles, Chen and Fei-Fei, 2010)	✓	X	X	783	16
UCF101 (Reddy and Shah, 2013)	✓	X	X	13,320	101
Northwestern-UCLA (Wang, et al., 2014)	X	✓	✓	1494	10
NTU RGB+D datasets (Shahroudy, et al., 2016)	X	✓	✓	56,880	60

costly training videos (Rahmani and Bennamoun, 2017). In addition, some elements must be handcrafted to solve complex problems. As a result, the key disadvantage of these methods is that they are problem dependent, making them difficult to execute in the real world, despite their high success in action detection (Wu, Sharma, and Blumenstein, 2017). Video editing is computationally effective. The Kinect is less effective outdoors, where a certain amount of solar IR and ferromagnetic radiation can create substantial noise and even wash out the scene produced by it.

Furthermore, the Kinect consumes more energy, has a lower resolution, and is not as quickly and inexpensively available as regular cameras (Zhang, Yan, and Li, 2018). The Kinect also fails to satisfy the criterion of monitoring different individuals. All cameras, including Kinect, have the drawback of invading the subject's privacy, which has led to debates about health surveillance applications for the elderly (Huang, et al., 2019).

One advantage of acceleration-based systems is that they achieve high emotion recognition accuracy (Huang, et al., 2019). Furthermore, they have high real-time efficiency and accuracy (Wang and Liu, 2020). However, one limitation of these systems is that they are restricted to the classification activities for which they were developed (Ravi, et al., 2016). Furthermore, they can burden the subject, jeopardizing the interactive experience (Ding, et al., 2020). Sensors also have the drawback of requiring multiple data processing steps, and they can involve the invasion of the subject's personal space (Huang, et al., 2019). Furthermore, acceleration sensors cannot obtain static information, so they have difficulty identifying locations, shapes, and other objects (Huang, et al., 2019). Furthermore, this method necessitates users to wear sensors on their bodies, which may be inconvenient or cumbersome (Ren, et al., 2020). As a result, interest in wearable sensors is low, and the growth of the technology is slow (Wang and Liu, 2020). Furthermore, wearable sensors reduce user interface while increasing hardware costs (Yan, Coenen, and Zhang, 2016). Finally, since they are battery powered, they must be maintained daily to continue working (Saini, et al., 2019).

Despite its challenges, vision-based technology is often used because cameras of this type are available in many places indoors and outdoors. However, recent developments have reduced the limitations. Furthermore, sensor-based technologies have recently entered many applications and have specificity. However, their main limitations are high price, low data accuracy, and failure to diagnose static objects. Therefore, the application of sensors is limited, unlike the vision-based technology, despite its shortcomings.

V. Conclusion and Future Work

Human body posture recognition has been extensively studied in recent years, and it plays an essential role in many applications. The methods used to recognize the poses and activities that humans perform are explored in this review paper. Vision-based techniques are discussed, as well as smartphone sensors and wearable sensors. Furthermore, it presents the most common datasets used to describe human body posture. Color or depth color cameras are used to capture data in vision-based systems. Alternately, sensors mounted on the human body while assessing the body's position through a two-dimensional or three-dimensional image, smartphone devices, and wearable sensors obtain data. Each strategy has its own set of benefits and drawbacks.

Because of their precision in the wearing phase, the sensors have a restricted application. Despite its drawbacks, vision-based technology can be advanced by researching and developing modern cameras. It is possible to use sensors on patients, the elderly, and athletes due to the lack of limitations in wearing those instruments. Classifying the data used in this technology are mainly artificial intelligence methods and deep learning. The primary advantage of deep learning is that it removes the need to extract features manually.

Furthermore, deep learning has shown reliable results. As a result, video monitoring and other activities may benefit from vision-based technologies. Deep learning is superior to other approaches to this technology, and it is possible to combine several deep learning algorithms for improved performance.

References

Abdulhussein, A.A. and Raheem, F.A., 2020. Hand gesture recognition of static letters American sign language (ASL) using deep learning. *Engineering and Technology Journal*, 38(6A), pp.926-937.

Abedi, W.M.S., Ibraheem-Nadher, D. and Sadiq, A.T., 2020. Modified deep learning method for body postures recognition. *International Journal of Advanced Science and Technology*, 29, 3830-3841.

Andriluka, M., Pishchulin, L., Gehler, P. and Schiele, B., 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp.3686-3693.

Ben Mabrouk, A. and Zagrouba, E., 2018. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 91, pp.480-491.

Cao, Z., Simon, T., Wei, S.E. and Sheikh, Y., 2017. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In: *Proceedings 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, pp.1302-1310.

Chen, Y. and Xue, Y., 2016. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. In: *Proceedings 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*. IEEE, pp.1488-1492.

Chen, Y., Tian, Y. and He, M., 2020. Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*, 192, p.102897.

Colleges, Gujarat Technological University, 2014. *Microsoft COCO*. Eccv, (June). Colleges, Gujarat Technological University, pp.740-755.

Dang, Q., Yin, J., Wang, B. and Zheng, W., 2019. Deep learning based 2D human pose estimation: A survey. *Tsinghua Science and Technology*, 24, pp. 663-676.

Ding, W., Hu, B., Liu, H., Wang, X. and Huang, X., 2020. Human posture recognition based on multiple features and rule learning. *International Journal of Machine Learning and Cybernetics*, 11(11), pp.2529-2540.

El Amine Elforaici, M., Chaaraoui, I., Bouachir, W., Ouakrim Y. and Mezghani, N., 2018. Posture Recognition Using an rgb-d Camera: Exploring 3d Body Modeling and Deep Learning Approaches. In: *2018 IEEE Life Sciences Conference, LSC*. IEEE, pp.69-72.

Ghafoor, K.J., Rawf, K.M., Abdulrahman, A.O. and Taher, S.H., 2021. Kurdish Dialect Recognition using 1D CNN. *Aro-the Scientific Journal of Koya University*, 9(2), pp.10-14.

Gong, W., Zhang, X., Gonzàlez, J., Sobral, A., Bouwmans, T., Tu, C. and Zahzah, E., 2016. Human pose estimation from monocular images: A comprehensive survey. *Sensors*, 16(12), p.1966.

Gouiaa, R. and Meunier, J., 2017. Learning cast shadow appearance for human posture recognition. *Pattern Recognition Letters*, 97, pp.54-60.

Gupta, A., Gupta, K., Gupta, K. and Gupta, K., 2020. A Survey on Human Activity Recognition and Classification. In: *2020 International Conference on Communication and Signal Processing* (*ICCSP*). IEEE, pp.915-919.

Hameed, F.S., Alwan, H.M. and Ateia, Q.A., 2020. Pose estimation of objects using digital image processing for pick-and-place applications of robotic arms. *Engineering and Technology Journal*, 38(5A), pp.707-718.

Hassan, M.M., Huda, S., Uddin, M.Z., Almogren, A. and Alrubaian, M., 2018. Human activity recognition from body sensor data using deep learning. *Journal of Medical Systems*, 42(6), p.99.

Hsiao, R.S., Chen, T.X., Bitew, M.A., Kao, C.H. and Li, T.Y., 2018. Sleeping posture recognition using fuzzy c-means algorithm. *BioMedical Engineering Online*, 17(S2), p.157.

Huang, X., Wang, F., Zhang, J., Hu, Z. and Jin, J., 2019. A posture recognition method based on indoor positioning technology. *Sensors*, 19(6), p.1464.

Jalal, A., Akhtar, I. and Kim, K., 2020. Human posture estimation and sustainable events classification via Pseudo-2D stick model and K-ary tree hashing. *Sustainability* (*Switzerland*), 12(23), pp.1-24.

Joo, H., Simon, T., Li, X., Liu, H., Tan, L., Gui, L., Banerjee, S., Godisart, T., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S. and Sheikh, Y., 2019. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1), pp.190-204.

Kale, H., Mandke, P., Mahajan, H. and Deshpande, V., 2018. Human Posture Recognition using Artificial Neural Networks. In: *Proceedings of the 8th International Advance Computing Conference, IACC 2018*. IEEE, pp.272-278.

Kamel, A., Sheng, B., Yang, P., Li, P., Ruimin, S. and Feng, D.D., 2019. Deep convolutional neural networks for human action recognition using depth maps and postures. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(9), pp.1806-1819.

Kamiŝalić, A., Fister, I., Turkanović, M. and Karakatič, S., 2018. Sensors and functionalities of non-invasive wrist-wearable devices: A review. *Sensors* (*Switzerland*), 18(6), p.1714.

Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J., 2018. End-to-end recovery of human shape and pose. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp.7122-7131.

Kareem, A.A., Ali, W.H. and Jasim, M.H., 2020. Design and implementation of a wireless system to locate a user in indoor environments. *Engineering and Technology Journal*, 38(11A), pp.1640-1651.

Khokhlov, I., Reznik, L., Cappos, J. and Bhaskar, R., 2018. Design of Activity Recognition Systems with Wearable Sensors. In: *2018 IEEE Sensors Applications Symposium* (*SAS*). IEEE, pp.1-6.

Kim, S.T. and Lee, H.J., 2020. Lightweight stacked hourglass network for human pose estimation. *Applied Sciences*, 10(18), p.6497.

Kołodziej, M., Majkowski, A., Tarnowski, P., Rak, R.J., Gebert, D. and Sawicki, D., 2019. Registration and analysis of acceleration data to recognize physical activity. *Journal of Healthcare Engineering*, 2019, pp.9497151.

Li, W., Zhang, Z. and Liu, Z., 2010. Action Recognition Based on a Bag of 3D Points. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, pp.9-14.

Liang, M. and Hu, Y., 2020. Application of Human Body Posture Recognition Technology in Robot Platform for Nursing Empty-Nesters. In: *2020 6th International Conference on Control, Automation and Robotics, ICCAR 2020*. IEEE, pp.91-95.

Liu, G., Lin, L., Zhou, W., Zhang, R., Yin, H., Chen, J. and Guo, H., 2019. A posture recognition method applied to smart product service. In: *Procedia CIRP*, pp.425-428.

Liu, J., Wang, Y., Liu, Y., Xiang, S. and Pan, C., 2020. 3D PostureNet: A unified framework for skeleton-based posture recognition. *Pattern Recognition Letters*, 140, pp.143-149.

Liu, M., Liu, H. and Chen, C., 2017. Enhanced skeleton visualization for view invariant human action recognition. *Pattern Recognition*, 68, pp.346-362.

Liu, Z., Zhu, J., Bu, J. and Chen, C., 2015. A survey of human pose estimation: The body parts parsing based methods. *Journal of Visual Communication and Image Representation*, 32, pp.10-19.

Lo Presti, L. and La Cascia, M., 2016. 3D skeleton-based human action classification: A survey. *Pattern Recognition*, 53, pp.130-147.

Marszałek, M., Laptev, I. and Schmid, C., 2009. *Marszałek-cvpr09-Action in Context*, pp.2929-2936.

Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W. and Theobalt, C., 2018. Monocular 3D Human Pose Estimation in the Wild Using Improved CNN Supervision. In: *Proceedings 2017 International Conference on 3D Vision, 3DV 2017*, pp.506-516.

Mirza, S.F. and Al-Talabani, A.K., 2021. Efficient kinect sensor-based kurdish sign language recognition using echo system network. *ARO The Scientific Journal of Koya University*, 9(2), pp.1-9.

Mo, L., Li, F., Zhu, Y. and Huang, A., 2016. Human Physical Activity Recognition Based on Computer Vision with Deep Learning Model. In: *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE, pp.1-6.

Nandy, A., Saha, J. and Chowdhury, C., 2020. Novel features for intensive human activity recognition based on wearable and smartphone sensors. *Microsystem Technologies*, 26(6), pp.1889-1903.

Neili, S., Gazzah, S., El Yacoubi, M.A. and Ben Amara, N.E., 2017. Human Posture Recognition Approach Based on ConvNets and SVM Classifier. In: *Proceedings 3rd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2017*. IEEE, pp.1-6.

Niebles, J.C., Chen, C.W. and Fei-Fei, L., 2010. Modeling temporal structure of decomposable motion segments for activity classification. *Lecture Notes in Computer Science* (*Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), *6312 LNCS* (*PART 2*), pp.392-405.

Ning, G., Zhang, Z. and He, Z., 2018. Knowledge-guided deep fractal neural networks for human pose estimation. *IEEE Transactions on Multimedia*, 20(5), pp.1246-1259.

Patel, P., Bhatt, B. and Patel, B., 2017. Human Body Posture Recognition a Survey. In: *2017 International Conference on Innovative Mechanisms for Industry Applications* (*ICIMIA*). IEEE, pp.473-477.

Quan, W., Woo, J., Toda, Y. and Kubota, N., 2019. Human posture recognition for estimation of human body condition. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 23(3), pp.519-527.

Rahmani, H. and Bennamoun, M., 2017. Learning action recognition model from depth and skeleton videos. In: *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, pp.5833-5842.

Rahmani, H., Mian, A. and Shah, M., 2018. Learning a deep model for human action recognition from novel viewpoints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3), pp.667-681.

Ravi, D., Wong, C., Lo, B. and Yang, G.Z., 2016. Deep Learning for Human Activity Recognition: A Resource Efficient Implementation on Low-power Devices. In: *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks* (*BSN*). IEEE, pp.71-76.

Rawat, W. and Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9), pp.2352-2449.

Reddy, K.K. and Shah, M., 2013. Recognizing 50 human action categories of web videos. *Machine Vision and Applications*, 24(5), pp.971-981.

Ren, W., Ma, O., Ji, H. and Liu, X., 2020. Human posture recognition using a hybrid of fuzzy logic and machine learning approaches. *IEEE Access*, 8, pp.135628-135639.

Saini, R., Kumar, P., Kaur, B., Roy, P.P., Dogra, D.P. and Santosh, K.C., 2019. Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare. *International Journal of Machine Learning and Cybernetics*, 10(9), pp.2529-2540.

Santhoshkumar, R. and Geetha, M.K., 2019. Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks. In: *Procedia Computer Science*, pp.158-165.

Sarafianos, N., Boteanu, B., Ionescu, B. and Kakadiaris, I.A., 2016. 3D human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, 152, pp.1-20.

Shahroudy, A., Liu, J., Ng, T.T. and Wang, G., 2016. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*. pp.1010-1019.

Soomro, K., Zamir, A.R. and Shah, M., 2012. *UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild*. *arXiv*, 2012, 12120402.

Sun, W. and Wang, R., 2018. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, 15(3), pp.474-478.

Vecchio, A., Mulas, F. and Cola, G., 2017. Posture recognition using the interdistances between wearable devices. *IEEE Sensors Letters*, 1(4), pp.1-4.

Voicu, R.A., Dobre, C., Bajenaru, L. and Ciobanu, R.L., 2019. Human physical activity recognition using smartphone sensors. *Sensors*, 19(3), p.458.

Wan, S., Qi, L., Xu, X., Tong, C. and Gu, Z., 2020. Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25(2), pp.743-755.

Wang, J. and Liu, X.H., 2020. Human Posture Recognition Method Based on Skeleton Vector with Depth Sensor. In: *IOP Conference Series: Materials Science and Engineering*, p.012035.

Wang, J., Nie, X., Xia, Y., Wu, Y. and Zhu, S.C., 2014. Cross-view Action Modeling, Learning, and Recognition. In*: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.2649-2656.

Wang, P., Li, W., Ogunbona, P., Wan, J. and Escalera, S., 2018. RGB-D-based human motion recognition with deep learning: A survey. *Computer Vision and Image Understanding*, 171, pp.118-139.

Wang, W.J., Chang, J.W., Haung, S.F. and Wang, R.J., 2016. Human posture recognition based on images captured by the kinect sensor. *International Journal of Advanced Robotic Systems*, 13(2), p.54.

Wu, D., Sharma, N. and Blumenstein, M., 2017. Recent Advances in Video-based Human Action Recognition Using Deep Learning: A Review. In: *2017 International Joint Conference on Neural Networks* (*IJCNN*). IEEE, pp.2865-2872.

Xia, L., Chen, C. and Aggarwal, J., 2012. *View Invariant Human Action Recognition Using Histograms of 3D Joints The University of Texas at Austin. CVPR 2012 HAU3D Workshop*, pp.20-27.

Xiao, J., Cui, X. and Li, F., 2020. Human action recognition based on convolutional neural network and spatial pyramid representation. *Journal of Visual Communication and Image Representation*, 71, p.102722.

Yan, C., Coenen, F. and Zhang, B., 2016. Driving posture recognition by convolutional neural networks. *IET Computer Vision*, 10(2), pp.103-114.

Zhang, H., Yan, X. and Li, H., 2018. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, 94, pp.1-10.

Zhang, H.B., Lei, Q., Zhong, B.N., Du, J.X. and Peng, J.L., 2016. A survey on human pose estimation. *Intelligent Automation and Soft Computing*, 22(3), pp.483-489.

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), p.1005.

Zhang, J., Wu, C. and Wang, Y., 2020. Human fall detection based on body posture spatio-temporal evolution. *Sensors* (*Switzerland*), 20(3), p.946.

Zhang, S. and Callaghan, V., 2020. Real-time human posture recognition using an adaptive hybrid classifier. *International Journal of Machine Learning and Cybernetics*, 12(2), pp.489-499.

Zhang, S., McCullagh, P., Zheng, H. and Nugent, C., 2017. Situation awareness inferred from posture transition and location: Derived from smartphone and smart home sensors. *IEEE Transactions on Human-Machine Systems*, 47(6), pp.814-821.

Zhao, C.H., Zhang, B.L., He, J. and Lian, J., 2012. Recognition of driving postures by contourlet transform and random forests. *IET Intelligent Transport Systems*, 6(2), pp.161-168.

Zhao, J. and Obonyo, E., 2020. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. *Advanced Engineering Informatics*, 46, p.101177.

Zhao, L. and Chen, W., 2020. Detection and recognition of human body posture in motion based on sensor technology. *IEEJ Transactions on Electrical and Electronic Engineering*, 15(5), pp.766-770.

Zhao, Z.Q., Zheng, P., Xu, S.T. and Wu, X., 2019. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11), pp.3212-3232.

Zhou, E. and Zhang, H., 2020. Human action recognition toward massive-scale sport sceneries based on deep multi-model feature fusion. *Signal Processing: Image Communication*, 84, p.115802.

Zhou, Y., Sun, X., Zha, Z.J. and Zeng, W., 2018. MiCT: Mixed 3D/2D Convolutional tube for human action recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp.449-458.