People Tracking in a Smart Room Using Kinect Sensor

Heba A. Salim
Collage Of Engineering
University of Basrah, Iraq
pgs2343@uobasrah.edu.iq

Musaab Alaziz
Collage Of Engineering
University of Basrah, Iraq
mosab.adil@uobasrah.edu.iq

Turki Y. Abdalla
Collage Of Engineering
University of Basrah, Iraq
turki.abdalla@uobasrah.edu.iq

2021 International Conference on Communication & Information Technology (ICICT) | 978-1-6654-3914-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICT52195.2021.9567897

***Abstract*—— In this paper, a new method for people tracking in a smart room using a Kinect sensor is proposed. The approach is based on the skeleton data with the (X, Y, Z) coordinate values of each joint in the human body which is provided by the Kinect sensor. For data classification, the Support Vector Machine (SVM) technique is used. To achieve this goal 14 movement classes are defined. Experiments were conducted on 12 subjects each one performs 14 movements in each experiment, the training dataset is created manually by capturing the subject movements during all experiments. The result of these we get after training the SVM model shows that the average accuracy is 90.2%.**

Keywords—Kinect Sensor, Motion tracking, Kinect skeleton, Support Vector Machine (SVM)

I. INTRODUCTION

People tracking is an important feature for smart environment analysis. Many research studies have been carried, and most rely on having a hardware device or sensors attached to the human's body to track them inside the room, or sensors attached to the furniture in that room [23][24]. In the project reported in this paper, a Kinect sensor is used to track people without the need for any additional hardware. The context knowledge gained from people's positions can assist us in predicting what people expect from a smart environment [1]. In nearly 9 years, much effort has been expended in examining human motion using the Kinect sensor. The Kinect sensor, which was released in 2010 [16], is capable of capturing not only color information but also depth and motion depth information. Because of its low cost and free SDK (Software Development Kit), the Kinect is becoming more popular [9]. The Microsoft Kinect for Windows SDK can track and capture data from a user's skeleton at 30 frames per second. Each tracked skeleton is constructed from the three-dimensional coordinates of twenty joints [5]. Initially, in this work, for data classification, the Support Vector Machine (SVM) technique is used, which provided good results [2], and the gait skeleton information is used to recognize a person's action [17], and this data is easily extracted from the 3D skeletal joint coordinates provided by the Kinect sensor, the skeletal data for each person consists of 20 joints, as shown in Fig. 1. Kinect V1 [3] provides these data. The novelty of this work lies in the way the method used to calculate the joint distance, the angle between joints, and the distance between the subject and the Kinect. This provides us with the features that will be used to improve classification accuracy [3].

The rest of the paper is structured as follows. In Section II, the hardware system and the components of the Kinect sensor are described; in Section III, we describe how the Kinect sensor recognizes human posture; and in Section IV, all experiments and scenarios are presented. Section V describes the Support Vector Machine (SVM) classifier and displays the final results. Finally, in Section VI, concluding remarks are provided.

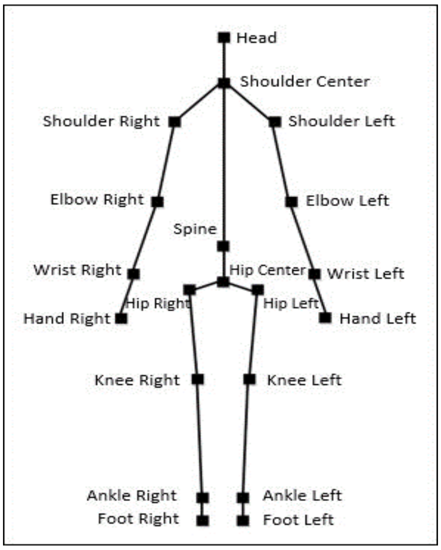


Fig. 1. 20 Joints Skeleton [4].

II. KINECT SENSOR

It is a motion-sensing device that was created for the Xbox 360 gaming console. Microsoft's Kinect software, which is embedded in the device, allows for human gesture recognition [4]. The Kinect sensor incorporates a wide range of advanced sensing hardware. It is equipped with a depth sensor, a color camera, and a four-microphone array, allowing it to capture full-body 3D motion, facial recognition, and voice recognition [6]. As illustrated in Fig. 2. the Kinect Sensor is made of:

1) A motorized Tilt:

The Kinect sensor has a tilt motor that could be used to tilt the camera and it increases the possible interaction space of the camera by +27 and -27 degrees [4].

Fig. 2. The Kinect Sensor.

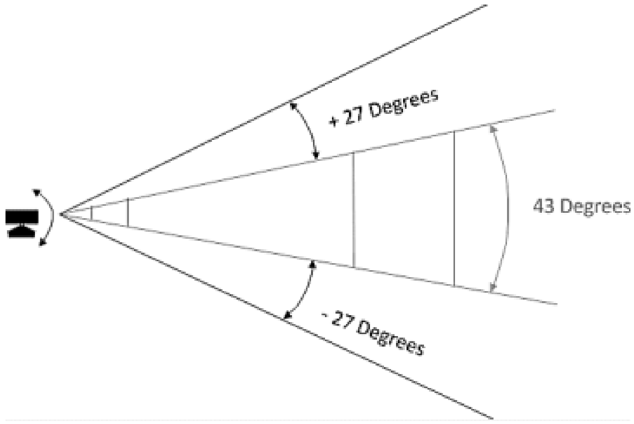


Fig. 3. Kinect Interaction Space[7].

1) An RGB color camera:

Identifies the red, green, and blue color components, as well as the body type and facial features. It has a 640x480 pixel resolution and a frame rate of 30 frames per second. This aids in both facial and body recognition [4].

2) A depth sensor:

A monochrome CMOS sensor and an infrared projector aid in the creation of 3D imagery throughout the room. It also calculates the distance between each point on the human's body by sending out invisible near-infrared light and measuring its "time of flight" after it reflects off the objects [4].

3) A microphone:

Is a set of four microphones that can separate the player's voice from other background noises, allowing them to use their voices as an additional control feature [4].

III. HUMAN POSTURE RECOGNITION BASED ON KINECT SENSOR

A. Feature extraction and data processing

With the skeleton tracked by the Kinect; first, extract the joint position, each joint has three values of the (X, Y, Z) coordinates [9]. From these values, many features can be derived such as joint angles, the distance between joints, the distance between each joint from the Kinect sensor, and the distance from the Kinect to the human skeleton. The total number of these features will be 46 features as shown in Fig.4 and TABLE II. From these features, 14 types of body positions are specified that the person will perform while present in the room as shown in TABLE I. These features and classes will be used to build the dataset that will be used to train the classification model (SVM). After training, the average accuracy is determined.

B. Human position dataset building

For data collection a dataset capture tool created in C# using Visual Studio 2017, the output window displays the Kinect skeleton and the (X, Y, Z) coordinate values, and the values of the features and body positions. To collect data for the dataset, the Kinect sensor sets at 170 cm height and the angle of the Kinect is -22 degrees. Neon lights are used in the testing room (which do not affect the results of recording skeleton data because Kinect uses an infrared camera, which works very well in the dark).

TABLE I. THE SELECTED CLASSES AND THEIR DESCRIPTION

Class Number	**Detailed description**
1.	Standing in the room
2.	Prayer position (sitting on the floor)
3.	Sitting on the floor with stretched legs
4.	Sitting on the floor with crossed legs
5.	Lying the floor
6.	Lying the floor with one leg raised
7.	Sitting on the bed
8.	Standing on the bed
9.	Standing on the bed with stretched legs
10.	Lying the bed with one leg raised
11.	Lying the bed
12.	Sitting on the chair
13.	Sitting on the chair with crossed legs (The right leg on the left leg)
14.	Sitting on the chair with crossed legs (The left leg on the right leg)

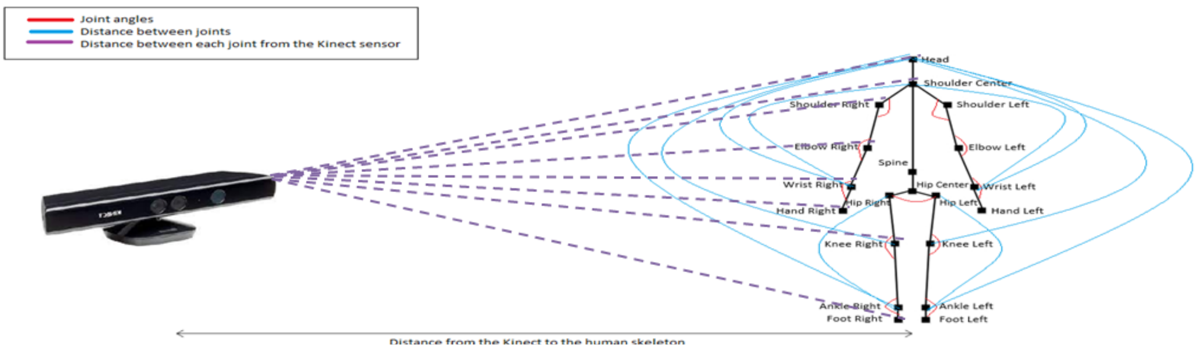


Fig. 4. Kinect sensor with the human skeleton and the measured features, the red arcs represent the joint angles, the blue arcs represent the distance between joints, and the purple dashed lines represent the distance between each joint and the Kinect sensor.

TABLE II. THE 46 FEATURE AND THEIR NAMES

Feature name			
a. Distance from the Kinect to the human skeleton	*b. Joint angles*	*c. Distance between joints*	*d. Distance between each joint from the Kinect sensor*
	1. Shoulder Left Angle	1. Head and Wrist Right	1. Head Distance
	2. Elbow Left Angle	2. Shoulder Center and Wrist Right	2. Shoulder Center Distance
	3. Wrist Left Angle	3. Head and Knee Right	3. Shoulder Right Distance
	4. Shoulder Right Angle	4. Head and Wrist Left	4. Shoulder Left Distance
	5. Elbow Right Angle	5. Head and Ankle Right	5. Elbow Right Distance
	6. Wrist Right Angle	6. Shoulder Center and Wrist Left	6. Spine Distance
	7. Hip Left Angle	7. Hip Right and Ankle Right	7. Hip Center Distance
	8. Knee Left Angle	8. Head and Knee Left	8. Hip Right Distance
	9. Ankle Left Angle	9. Hip Left and Ankle Left	9. Hip Left Distance
	10. Ankle Left Angle	10. Head and Ankle Left	10. Knee Right Distance
	11. Knee Right Angle		11. Knee Left Distance
	12. Ankle Right Angle		12. Elbow Left Distance
	13. Spine Angle		13. Wrist Right Distance
	14. Shoulder Center Angle		14. Wrist Left Distance
	15. Hip Center Angle		15. Hand Right Distance
			16. Hand Left Distance
			17. Ankle Right Distance
			18. Ankle Left Distance
			19. Foot Right Distance
			20. Foot Left Distance

C. Features mathematical calculations

1) Measure the distance from the Kinect to the human skeleton:

Every single joint has 3 values: X, Y, and Z. It is projected in a Cartesian coordinate system. The (0, 0, 0) point is the position of the sensor. The distance between the player and the device is represented by a mathematical vector.

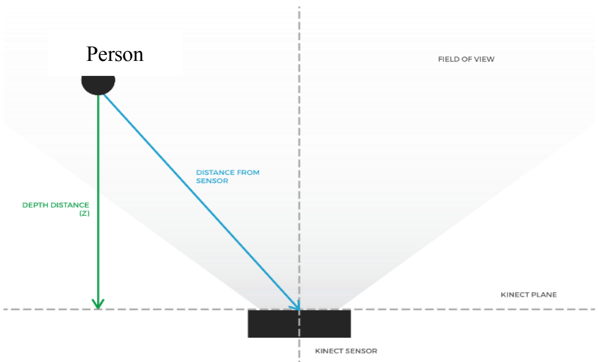


Fig. 5. The distance between the human skeleton and the sensor is represented by a mathematical vector (drawn in blue). The Position is a set of X, Y, and Z coordinates in the 3D space. The "Z" value is the distance between the player and the plane.

The length of a vector is given by the formula below:

$$D1 = \sqrt{X^2 + Y^2 + Z^2} \qquad (1)$$

2) Measure the distance between joints:

To find the distance the joints should be specified first, let the first joint be $J1$ and the second joint $J1$, the distance equation will be as follows:

$$D2 = \sqrt{(J1.X - J2.X)^2 + (J1.Y - J2.Y)^2 + (J1.Z - J2.Z)^2} \qquad (2)$$

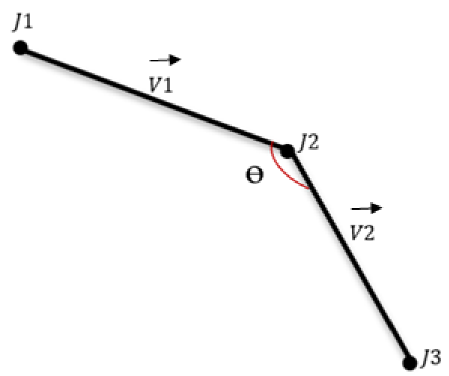
3)Measure the distance between each joint and the Kinect :

The same approach that used to measure the distance between the skeleton and Kinect is used here, but the joint type is specified, as shown in the equation below:

$$D3 = \sqrt{(Joint_{type}.X)^2 + (Joint_{type}.Y)^2 + (Joint_{type}.Z)^2} \qquad (3)$$

4)Measure the Angle between joints :

The vectors V1 and V2 are used o measure the angle between two joints, as shown in Fig. 6. , and the equation will be :



$$\theta = cos^{-1} \frac{\overrightarrow{V1}.\overrightarrow{V2}}{\left\|\overrightarrow{V1}.\overrightarrow{V2}\right\|} \qquad (4)$$

Fig. 6. Measure the distance between joints, and joint angle

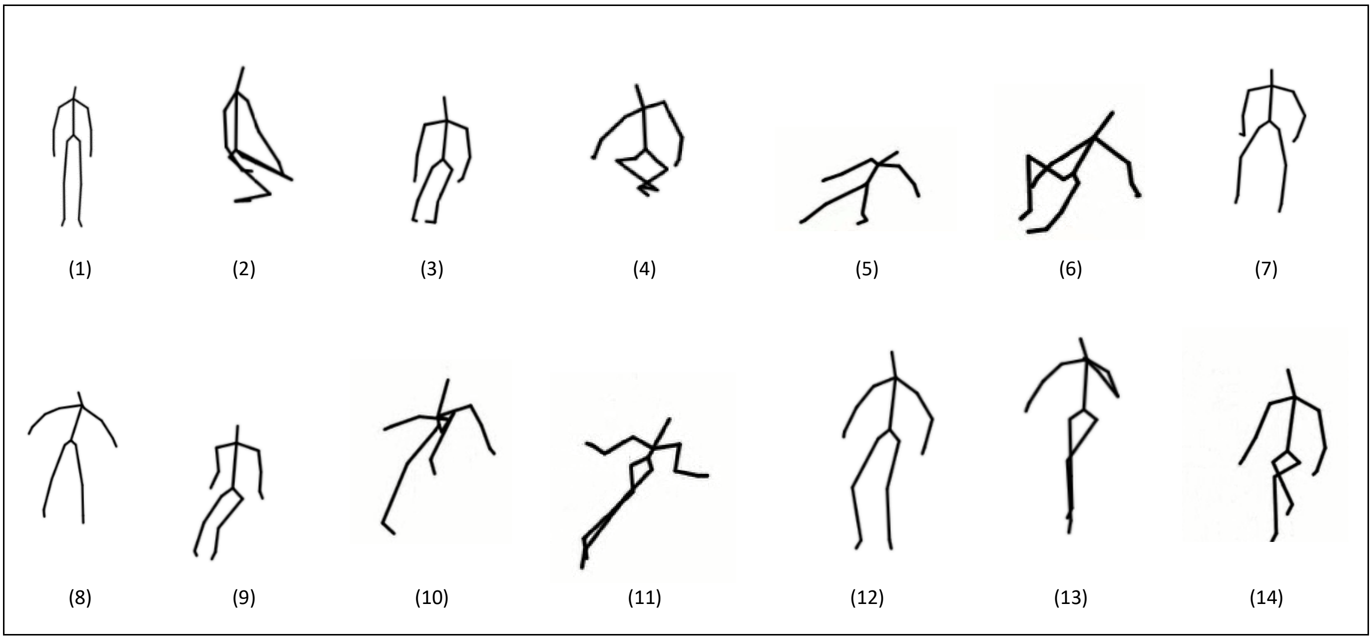


Fig. 7. Body position types, extracted from the experiments.

IV. EXPERIMENTS AND SCENARIOS

We recorded the 14 positions of 12 people for our experiments (9 men and 3 women). We record 5 times for each position for about 10 to 20 seconds. There have been 1120 experiments in total. The following scenarios are assigned to the subject.:

3) Scenario for standing position:
The subject stands straight and relaxed with his or her arms at his or her sides. His/her eyes are drawn to Kinect (see Fig. 7 (1)) . The distance between the subject and the Kinect is 148 cm in the first and second recording times, and 176 cm in the other three recording times.

4) Scenario for prayer position:
In all experiments, the subject sits on the ground in a prayer position with his/her right side in front of the Kinect (see Fig. 7 (2)), and the distance between the subject and the Kinect is 150 cm.

5) Scenario for sitting on the floor with stretched legs position:
In all experiments, the subject sits on the ground in front of the Kinect with his/her legs straightened (see Fig. 7 (3)), and the distance between the subject and the Kinect is 226 cm.

6) Scenario for sitting on the floor with crossed legs position:
In all experiments, the subject sits on the ground in front of the Kinect with his/her legs crossed (see Fig. 7 (4)), and the distance between the subject and the Kinect is 220 cm.

7) Scenario for lying on the floor position:
The subject's body lying on the ground in a vertical direction from the viewpoint of the Kinect (see Fig. 7 (5)), in all experiments, the distance between the subject and the Kinect is 126 cm.

8) Scenario for lying on the floor with one leg raised position:
The subject's body lying on the ground with one leg raised in a vertical direction from the viewpoint of the Kinect (see Fig. 7 (6)), In all experiments, the distance between the subject and the Kinect is 126 cm.

9) Scenario for sitting on the bed position:
The subject sits on the bed, with his/her hands are on the bed, and keeps his/her back straight (see Fig. 7 (7)), In all experiments, the distance between the subject and the Kinect is 186 cm.

10) Scenario for standing on the bed position:
The subject stands straight on the bed with his/her arms relaxed His/her gaze is drawn to Kinect (see Fig. 7 (8)). In all experiments, the distance between the subject and the Kinect is 226 cm.

11) Scenario for sitting on the bed with stretched legs position:
The subject sits on the bed in front of the Kinect with his/her legs are straightened (see Fig. 7 (9)), the distance between the subject and Kinect is 212 cm in all experiments.

12) Scenario for lying on the bed with one leg raised position:
The subject's body lying on the bed with one leg raised in a vertical direction from the viewpoint of the Kinect (see Fig. 7 (10)), the distance between the subject and Kinect is 230 cm in all experiments.

13) Scenario for lying on the bed position:
The subject's body lying on the bed in a vertical direction from the viewpoint of the Kinect (see Fig. 7 (11)), In all experiments, the distance between the subject and the Kinect is 230 cm.

14) Scenario for sitting on the chair position:
The subject sits on the chair, with his/her hands are on the chair sides, with his/her feet touches the floor, and keeps his/her back straight(see Fig. 7 (12)) , In all experiments, the distance between the subject and the Kinect is 221 cm.

15) Scenario for sitting on the chair with the right leg on the left leg position:
The subject sits on the chair, with his/her hands are on the chair sides, and his/her right leg on the left leg, and keeps his/her back straight (see Fig. 7 (13)), In all experiments, the distance between the subject and the Kinect is 221 cm.

16) Scenario for sitting on the chair with the left leg on the right leg position:
The subject sits on the chair, with his/her hands are on the chair sides, and his/her left leg on the right leg, and keeps his/her back straight (see Fig. 7 (14)), In all experiments, the distance between the subject and the Kinect is 221 cm.

V. CLASSIFICATION: SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are cutting-edge large-margin classifiers that have recently gained popularity in visual pattern recognition and other applications [13][14]. SVM can obtain decision-making rules and achieve low error for independent test sets, allowing it to efficiently solve learning problems. Also, it can ensure higher performance in a variety of practical applications and the accuracy of long-term predictions. For our work, the SVM model is trained using the dataset collected after completing all experiments. First, create an SVM model and train it using our dataset. To achieve good performance the SVM classifier's radial basis function (RBF) kernel [10], [15], is used. The data is divided using cross-validation, randomly selecting 50% of the data for training and 50% for testing, and repeating this process for 10 iterations, at each iteration, the accuracy is determined, and then determine the average accuracy, the average accuracy was 90.2%. From the confusion matrix shown in Fig. 9. The average value of the Recall and precision is determined for each class; their values are shown in TABLE III.

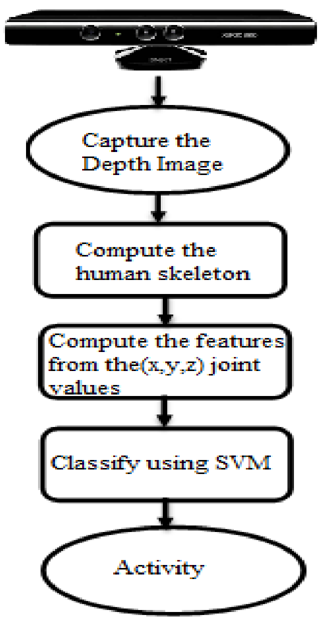


Fig. 8. Main steps of activity classification using the skeleton information and SVM classifier

TABLE III. THE 14 CLASSES AND THEIR RECALL AND PRECISION VALUES FOR THE SVM MODEL

Class Number	Class Name	Recall %	Precision %
1.	Standing in the room	0.932891	0.848343
2.	Prayer position (sitting on the floor)	0.887235	0.934926
3.	Sitting on the floor with stretched legs	0.932529	0.918025
4.	Sitting on the floor with crossed legs	0.939682	0.964340
5.	Lying the floor	0.899020	0.920667
6.	Lying the floor with one leg raised	0.869527	0.869803
7.	Sitting on the bed	0.932194	0.950073
8.	Standing on the bed	0.924460	0.936955
9.	Standing on the bed with stretched legs	0.927179	0.940658
10.	Lying the bed with one leg raised	0.849523	0.854580
11.	Lying the bed	0.845845	0.837072
12.	Sitting on the chair	0.917644	0.935132
13.	Sitting on the chair with crossed legs (The right leg on the left leg)	0.864786	0.911648
14.	Sitting on the chair with crossed legs (The left leg on the right leg)	0.900427	0.833511

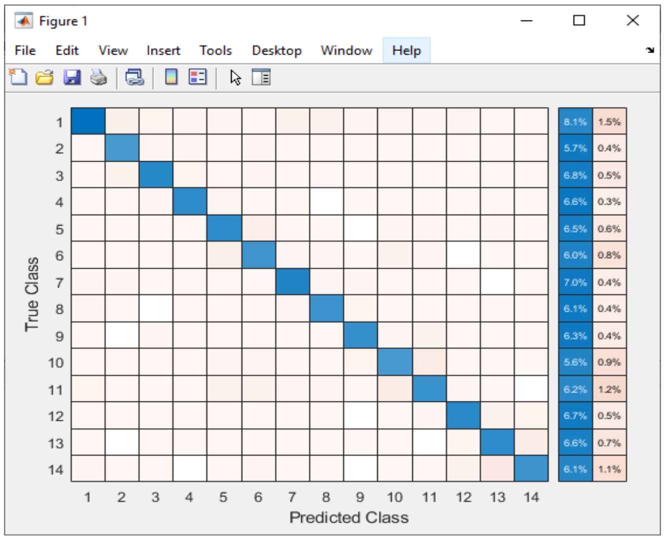


Fig. 9. Confusion matrix of 14 classes for the SVM model "Captured from MATLAB R2019b"

VI. CONCLUSION

In this paper, a method for tracking people and recognizing actions in a smart room is proposed, by using the skeleton provided by the Kinect sensor. We ended up running 1120 experiments using 46 different features extracted from the human skeleton and tracked by the Kinect sensor. For classification, the SVM classifier is used, an average accuracy of 90.2% is get. The result obtained shows that the skeleton allows us to classify the 14 body positions well. In the future, we plan to experiment with different classification methods and compare the resulting classification accuracy of the SVM accuracy, also plan to add more experiments to improve the accuracy.

VII. RELATED WORK

There have been lots of works proposed for people tracking and action recognition, many of them define only the basics body positions such as standing and sitting, also they used few numbers of features and experiments. While in our work the number of body positions is 14 with 46 features for each position and the total number of experiments is 1120. In this section, we describe some of these related works. S. Majumder and N. Kehtarnavaz, et al. [9] have proposed an algorithm for identifying basic human postures from still images that were created using C++ and the open pose library, an algorithm for identifying basic human postures from still images were created. Using this method, two postures, sitting and standing, were classified. Only human skeleton information is available. Wei, Qiao, and Lee et al. [5] a KSCC algorithm for calibrating the Kinect skeleton coordinate for remote physical training applications was proposed. Rahman and Gavrilova (2017) et al. [3] present a method for identifying people by using sensor-based gait data. The goal of this project is to identify a person using Kinect 3D skeletal joint gait data. The gait cycle of each individual is detected, and features are trained using a KNN classifier. Ben Tamou, Ballihi, and Aboutajdine et al. [11] proposed a novel approach to human action recognition based on depth camera-extracted skeleton joints The 3D coordinates of skeleton joints are subtracted in their method. Bhattacharya, Czejdo, and Perez et al. [12] In the context of aircraft marshaling, they discussed machine learning techniques for gesture classification. The characteristics distinguish their research. They used the joint coordinates data stream from the Kinect sensor as the feature describing the moving human body in video data. They chose the most accurate one using machine learning techniques (SVM, linear kernel).

REFERENCES

[1] J. Zhao, G. Zhang, L. Tian, and Y. Q. Chen, "REAL-TIME HUMAN DETECTION WITH DEPTH CAMERA VIA A PHYSICAL RADIUS-DEPTH DETECTOR AND A CNN DESCRIPTOR School of Computer Science , Shanghai Key Laboratory of Intelligent Information Processing , Fudan University , China," no. July, pp. 1536–1541, 2017.

[2] S. Maudsley-Barton, J. McPhee, A. Bukowski, D. Leightley, and M. H. Yap, "A comparative study of the clinical use of motion analysis from kinect skeleton data," in *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, 2017, vol. 2017-Janua, pp. 2808–2813, doi: 10.1109/SMC.2017.8123052.

[3] M. W. Rahman and M. L. Gavrilova, "Kinect gait skeletal joint feature-based person identification," in *Proceedings of 2017 IEEE 16th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2017*, 2017, pp. 423–430, doi: 10.1109/ICCI-CC.2017.8109783.

[4] A. Jana, "Kinect for Windows SDK Programming Guide," in *PACT Publishing*, 2012, pp. 1–366.

[5] T. Wei, Y. Qiao, and B. Lee, "Kinect skeleton coordinate calibration for remote physical training," in *MMEDIA - International Conferences on Advances in Multimedia*, 2014, no. c, pp. 66–71.

[6] A. Shahroudy, J. Liu, T. T. Ng, and G. Wang, "NTU RGB+D: A large scale dataset for 3D human activity analysis," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 1010–1019, 2016, doi: 10.1109/CVPR.2016.115.

[7] Microsoft, "Kinect Sensor," no. November, pp. 1371–1372, 2012, doi: 10.13140/2.1.1068.5124.

[8] MathWorks: https://www.mathworks.com/help/imaq/data-streams-returned-by-the-kinect.html , March 2021

[9] S. Ghazal and U. S. Khan, "Human posture classification using skeleton information," *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.*, vol. 2018-January, pp. 1–4, 2018, doi: 10.1109/ICOMET.2018.8346407.

[10] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors (Basel).*, vol. 18, no. 1, pp. 1–20, 2017, doi: 10.3390/s18010018.

[11] A. Ben Tamou, L. Ballihi, and D. Aboutajdine, "Automatic learning of articulated skeletons based on mean of 3D joints for efficient action recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 4, pp. 1–19, 2017, doi: 10.1142/S0218001417500082.

[12] S. Bhattacharya, B. Czejdo, and N. Perez, "Gesture Classification with Machine Learning using Kinect Sensor Data," no. December, pp. 348–351, 2016, doi: 10.1109/EAIT.2012.6407958.

[13] M. Awad and R. Khanna, "Efficient learning machines: Theories, concepts, and applications for engineers and system designers," *Effic. Learn. Mach. Theor. Concepts, Appl. Eng. Syst. Des.*, no. January, pp. 1–248, 2015, doi: 10.1007/978-1-4302-5990-9.

[14] M. G. A. Komang, M. N. Surya, and A. N. Ratna, "Human activity recognition using skeleton data and support vector machine," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, pp. 0–8, 2019, doi: 10.1088/1742-6596/1192/1/012044.

[15] B. Seddik, S. Gazzah, and N. Essoukri Ben Amara, "Human-action recognition using a multi-layered fusion scheme of Kinect modalities," *IET Comput. Vis.*, vol. 11, no. 7, pp. 530–540, 2017, doi: 10.1049/iet-cvi.2016.0326.

[16] R. Lun and W. Zhao, *A Survey of Applications and Human Motion Recognition with Microsoft Kinect*, vol. 29, no. 5. 2015.

[17] G. Roy, A. Bhuiya, A. Mukherjee, and S. Bhaumik, "Kinect Camera Based Gait Data Recording and Analysis for Assistive Robotics-An Alternative to Goniometer Based Measurement Technique," *Procedia Comput. Sci.*, vol. 133, pp. 763–771, 2018, doi: 10.1016/j.procs.2018.07.121.

[18] T. T. H. Tran, T. L. Le, and J. Morel, "An analysis on human fall detection using skeleton from Microsoft kinect," *2014 IEEE 5th Int. Conf. Commun. Electron. IEEE ICCE 2014*, pp. 484–489, 2014, doi: 10.1109/CCE.2014.6916752.

[19] S. Majumder and N. Kehtarnavaz, "Vision and Inertial Sensing Fusion for Human Action Recognition: A Review," *IEEE Sens. J.*, vol. 21, no. 3, pp. 2454–2467, 2021, doi: 10.1109/JSEN.2020.3022326.

[20] V. R. Reddy and T. Chattopadhyay, "Human activity recognition from Kinect captured data using stick model," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8511 LNCS, no. PART 2, pp. 305–315, 2014, doi: 10.1007/978-3-319-07230-2_30.

[21] M. Alaziz, Z. Jia, R. Howard, X. Lin, and Y. Zhang, "MotionTree: A Tree-Based In-Bed Body Motion Classification System Using Load-Cells," *Proc. - 2017 IEEE 2nd Int. Conf. Connect. Heal. Appl. Syst. Eng. Technol. CHASE 2017*, pp. 127–136, 2017, doi: 10.1109/CHASE.2017.71.

[22] A. Elm, "Evaluation of Body Position Measurement and Analysis using Kinect," p. 56, 2014.

[23] S. Fallmann and L. Chen, "Computational sleep behavior analysis: A survey," *IEEE Access*, vol. 7, pp. 142421–142440, 2019, doi: 10.1109/ACCESS.2019.2944801.

[24] J. Neumann, J. R. Casas, D. Macho, and J. R. Hidalgo, " Integration of audiovisual sensors and technologies in a smart room," *Pers. Ubiquitous Comput.*, vol. 13, no. 1, pp. 15–23, 2009, doi: 10.1007/s00779-007-0172-1.