IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 7.918 ⁙ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12117

AI System for Human Gesture Recognition and Control

Dhruva L[1], Vishesh S[2]

BE (CSE), Dr. AIT, Bangalore, India[1]

BE, MBA, Bangalore, India[2]

Abstract: Our paper introduces a gesture human action recognition system to monitor the various human actions and aids in the creation of the Human Robot Interaction (HRI) interface. The depth image of the person is initially acquired using the Kinect sensor 2.0, while the colour image is initially obtained using the RGB Sensor. The Histogram of Oriented Gradient (HOG) Features is calculated by combining the data from the image and the colour image into a single image. Finally, these traits are transformed into robot-recognizable actions. This system is capable of identifying a wide range of human behaviours, including walking, waving the left and right hands, bowing, and holding the left and right hands. Neural Network (NN) classifiers are being used to recognise the different gestures. In essence, Haar cascade classifiers are used to examine if a picture matches both positively and negatively. The actions taken by the human are what determine the robot's entire movement. The person standing in front of the Kinect sensor is completely responsible for controlling the device. This can be utilised as a real-time application and put into action as a human replacement where necessary.

Keywords: Kinect sensor 2.0, Human Robot Interaction (HRI) Interface, Histogram of Oriented Gradient (HOG), Neural Network (NN), human action recognition system, RGB Sensor, real time application and Haar cascade classifiers.

I. INTRODUCTION

The assessment of human posture has long been a problem in computer vision, having applications in Human PC Connection, movement capture, and action recognition. The high component of the research area, the vast array of opportunities available, and the requirements, such as refraining from body part infiltration and rejecting unsustainable ideas, all contribute to the emergence of complexity. Different challenges include various bodily limits, jumbled foundation, and brightness alterations. In this study, an attempt is made to reconstruct the human body using stick skeletons using RGB and depth data obtained from Microsoft Kinect sensors. Segment two gives concise outline for the Kinect sensor and segment four depicts a calculation with brief diagrams for different advances.

II. SYSTEM OVERVIEW

The computer software helps us to trace and supervise activities of human and interactive feedback for health care and other applications. Here, we present system overview of design and implementation. The system has of major elements such as: Human activity tracking and recognition, detection, real time feedback, interactive feedback and data collection program. Human Activity Tracking and Recognition: Human Activity recognition is the major element in our system. We have a laptop or a computer that connects to Kinect sensor. A motion sensing input device which is a Kinect senor has three major parts: Colour video camera: This camera aids in facial recognition and other detection features by detecting three colour components red, green and blue. Depth sensor: An IR projector and monochrome CMOS sensor works together to see the room in 3D regardless of lighting conditions. Multi array microphone: This is an array of 4 microphones that can separate the voices of users from the noise in the room. The outcome of detection is carried in real time with user. Interactive Feedback: Here users are allowed to register with the Kinect and inform the user if any inappropriate or wrong movement is detected.

The registration is a process that recognizes and traces a consented worker. The real time activities of human are recognized for proper movements. In this project we use Microsoft Kinect to explore its use on robot and to examine robotic sensor. Basically, Kinect based robot was designed for applications in robotic fields such as in education, medical, entertainment etc. So in this method we measure a human body. A human skeleton model has seven body parts involving eight points. We use anthropometric data which is the study of measurements and proportions of human body from a given data source book to estimate the proportions or size of the body parts. Here we use visual studio software and Arduino IDE software for coding purpose. Visual studio software is used for coding in transmitter part while the Arduino IDE software is used for coding in the receiver part. This project gives a brief overview of latest research on robotics

DOI: 10.17148/IJARCCE.2023.12117

using Kinect. Till now machines used were either automated or remote controlled. Were these can be remotely controlled by RF, IR, IOT (using Bluetooth and WI-FI and RF module). In this project, the machines are controlled virtually using a Natural User Interface (NUI) console Kinect. Here we are controlling a four-wheel robot action using NRF wireless communication. Here the transmitter section can be considered as the user part and receiver section can be considered as an application part. In the transmitter section we have Kinect sensor, pc or laptop, microcontroller and a NRF transmitter. The main heart of the project is the Kinect sensor. Here the Kinect sensor is connected to the laptop.

When the user sits in front of the sensor and performs various actions it is recorded by the sensor. The laptop is then connected to the microcontroller which is Arduino UNO. The various actions performed by the human are tracked by the Kinect sensor, and then the Kinect sensor calculates the difference between the threshold values and the tracked values. Arduino UNO is used for analysing the data and to transfer the data to the NRF. NRF is a wireless transceivers module. Coming to the receiver section we have NRF receiver, microcontroller, motor driver and DC motors.

The NRF receiver transfer the data received by the NRF transmitter to the Arduino UNO microcontroller, which is again transferred to the motor driver circuit. This motor driver circuit controls a set of two dc motors simultaneously in any direction. These dc motors will run accordingly as per the instructions given by the Arduino microcontroller which is coded by the Arduino IDE. In the Arduino IDE the actions are converted to the binary code, and these binary codes are executed to the actions by the robotic car. The microcontroller used here is the Arduino UNO, which is coded by the Arduino IDE for converting the actions performed by the user in front of the Kinect sensor to the binary code which helps the robot car to follow the actions. NRF module is a transceivers module which is used in both the transmitter and the receiver section for data transmission.

The range of the NRF is between 10 -20 meters.Kinect sensor and its software development kit (SDK), human machine interface of personal computer has achieved a new level where the users directly interact with the human body movements. This new form of HMI has quickly spread to various dimensions including the education, medical care, entertainments etc. Motor driver circuit is used for controlling a set of two DC motors simultaneously in any direction. Then these DC motors will run accordingly as per the instructions given. The language used for coding is C++, VB (visual basic), Python, JavaScript, and many more languages. We use visual studio in the transmitter section to track the various actions of human or user performed in front of the Kinect sensor, and then helps the Kinect sensor to calculate the difference between the threshold values and the tracked values.

Arduino IDE is a cross-platform application. The language used for coding is C and C++. We use Arduino IDE in the receiver section. In the Arduino IDE the various actions performed by the human or user in front of the Kinect sensor are converted into binary code. Then these binary codes are implemented to the actions by the robotic car. Figure 1 shows the system overview of our assignment. [1]

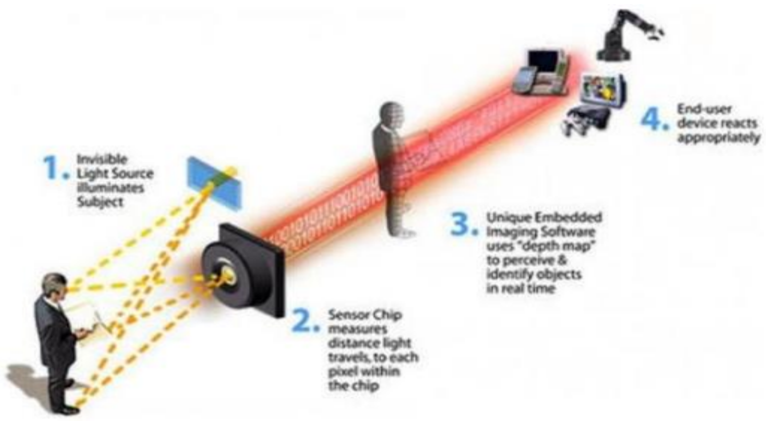

Figure 1 shows the system overview of our assignment.

III. SYSTEM FLOWCHART

As shown in the Flowchart diagram, initially the Kinect sensor captures an image of the person standing in front of the Kinect and these actions performed by the humans are converted to commands with the help of the visual studio. The Kinect sensor can capture the image in 3D by considering in all the three dimensions including the xy-yz-zx plane. Using the RGB Sensor the colour image of the person is obtained. This colour image is converted into the depth image and the

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023
DOI: 10.17148/IJARCCE.2023.12117

stick skeletal model of the image is obtained. The joint angle calculation is done by the sensor and the related valid angle is been calculated. The angles for each action is been fixed as a threshold value and these values gets changed from one person to other. Every time when a person stands in front of the sensor threshold value keeps getting changed and these changes depends on person's height weight and body posture and his current positions. For every action there is a fixed threshold value when a person stands in front of the sensor the value nearest to the corresponding threshold value the corresponding actions get executed. After the image is captured image processing steps are being carried out to remove the unwanted noise present in the background of the image. In the Foreground segmentation, noise is removed using the morphological operations like erosion and dilation. Later the small blob is removed to obtain a Proper Image.

This helps us to focus only on the required part of the image. Skin colour segmentation is the next step performed to the foreground segmented image. For this, we will take the RGB image as the input and convert to HSV colour space and the pixel value will range between the HSV_{max} and HSV_{min} with values from 255 to 0. We do Blob analysis to get skin mask for the obtained image. The next step will be the arm fitting, in this step we make use of the Euclidean transform algorithm. In this we consider the pivot points and we will perform the angular search for these pivot points (shoulder for elbow and elbow for wrist points).Direction with the maximum EDT values are considered and end point skin value is considered as 255. If there is no final point then the value with the second largest EDT values is considered.

These data in form of the codes are transferred with the help of the NRF (Narrow radio frequency).NRF is been used in both the transmitter and as well as in the receiver side. It can transmit and receive the data simultaneously. These data are sent to the Arduino UNO towards the receiver end. In the receiver end, using the Arduino IDE Software finally, the commands are converted to machine level codes. The robot will enact same actions performed by the humans with the help of machine codes. Data is been obtained from the anthropometric source book to analyze the size of the body parts. [2] Width set is considered as twice the face width, shoulder points are halfway between the face and rectangle base lines with shoulders, arm's length is considered as 1.12 times the face width of the object, and these analyses are based on the anthropometric data. Figure 2 shows the system flowchart. Figure 3 shows the stick skeletal model. Figure 4 shows the data analysis. Figure 5 shows segmented depth image and figure 6 shows the skin segmented image. Figure 7 shows the angular search for pivot points.[3]

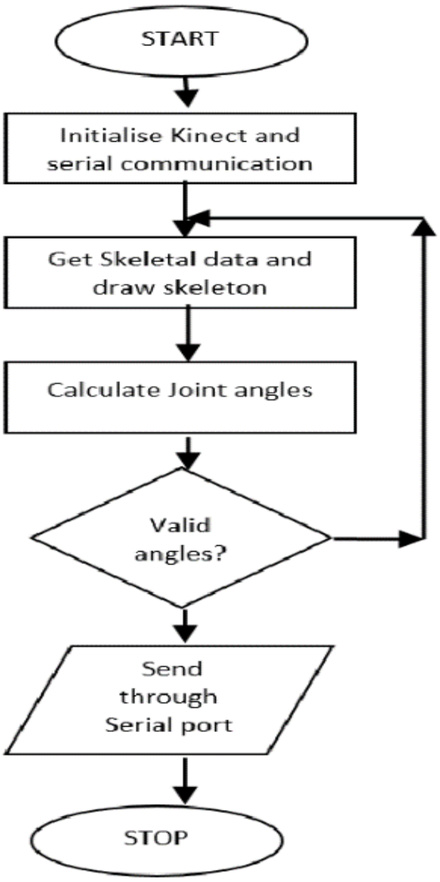

Figure 2 shows the system flowchart.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁂ Impact Factor 7.918 ⁂ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12117

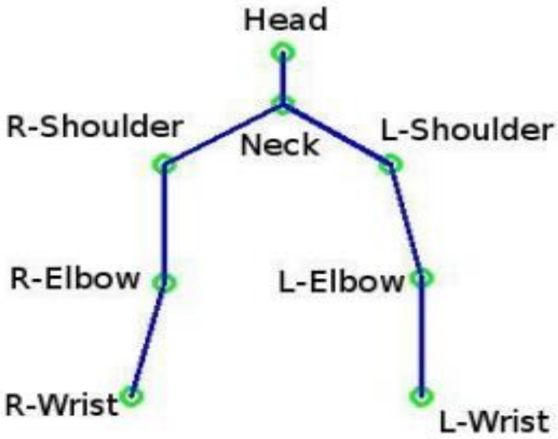

Figure 3 shows the stick skeletal model.

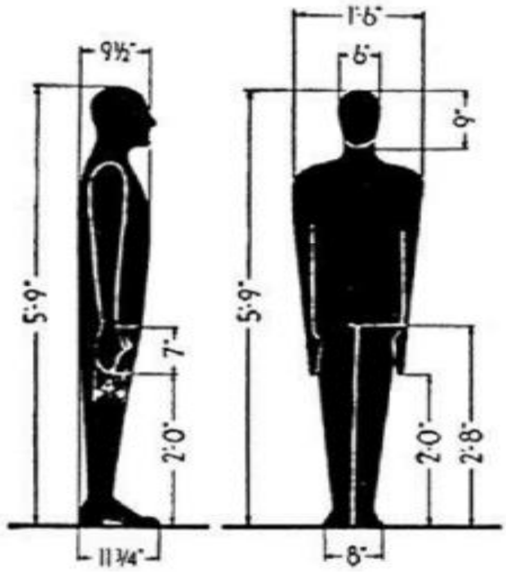

Figure 4 shows data analysis.

Figure 5 shows foreground segmented depth image.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.918 ⋇ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12117

Figure 6 shows the skin segmented image.

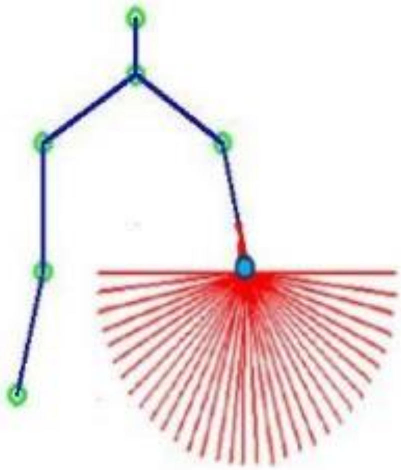

Figure 7 shows the angular search for pivot points.

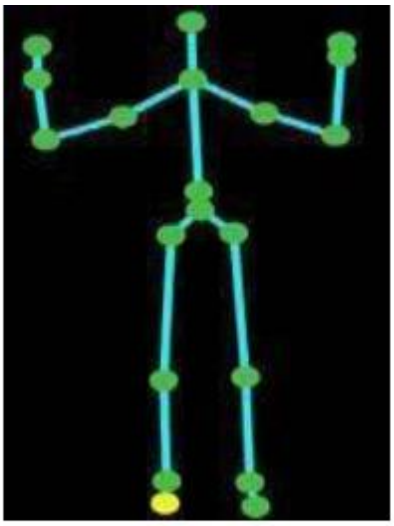

Figure 8 shows the gestures for the robot to move forward.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ※ Impact Factor 7.918 ※ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12117

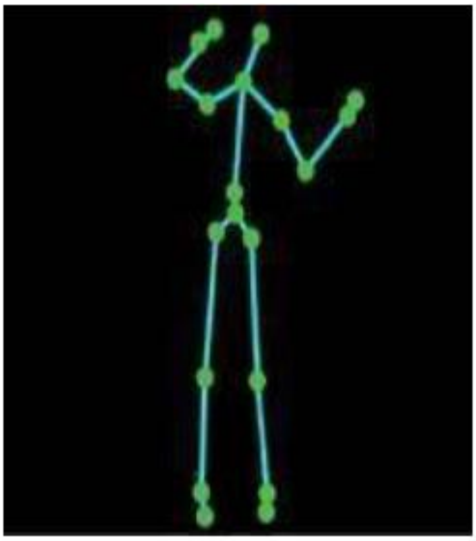

Figure 9 shows the gesture to turn right.

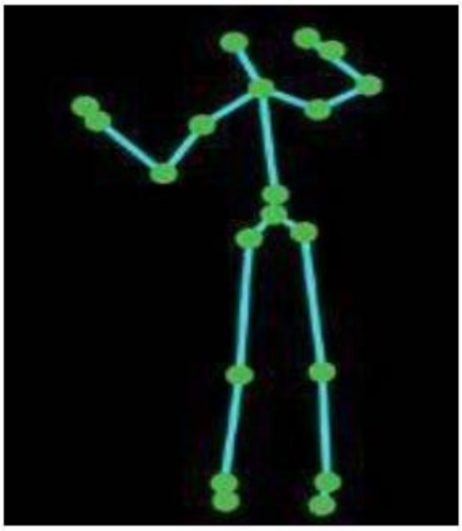

Figure 10 shows the gesture to turn left.

IV. RESULTS

The arudino microcontroller is placed inside the robot. And the robot will receive the control signals from pc. To simulate the robots actions are given by the person standing in front of the Kinect sensor. [4] The wheel robot is of 75cm which is capable of moving forward, backward, right, left, and also rotates in clockwise and anticlockwise direction. The results are shown by human body hand skeleton positions without using any devices like remote robot can be controlled.

The speed is of 100centimeters per second. In the experiment eight kind of human actions can be performed. The algorithm was inserted in Microsoft Kinect libraries. The result is implemented on skeleton body. This algorithm works well on the images with less occlusion or meteorology. The distance transform map becomes a bit unreliable because of arm occlusions. In our paper, we have considered only the upper body and based on the upper body actions the robot enacts. The detection is done for each frame separately.

The depth image captured by Kinect is 512*424 pixels and colour images are of 1920*1080 pixels. The system normalizes videos is of 20 frames. If the sequence contained less than 20 frames and then the frames are copied and inserted. Figure 8, 9 and 10 shows the gestures for the robot to move forward, right and left respectively.

ISO 3297:2007 Certified ⵈ Impact Factor 7.918 ⵈ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12117

V. CONCLUSIONS AND FUTURE WORK

The main goal of this paper is to achieve the human action track recognition of the system. We have been implemented the distance transform algorithm to analyse the parameters for the human stick skeletal model only to the upper body. We included the obstacle avoidance algorithm to detect the obstacle in the image. The system can detect the eight different actions completely performed by the human successfully.

The average recognition rates were 88.7%, 74.3% and 51.2%.The data transfer speed achieved by the system is 2 Mbps. In the paper, we have performed the experiments by considering only the upper body, In the future days, we aim to implement the idea to the entire Human body system. We can work on to decrease the number of delays and to improve the overall efficiency of the system and give importance to achieve the complete accuracy of the system. Data speed of the system should be increased. This system can recognize the eight different actions, as a future work we intend to extend many more actions. We have achieved the action command system as a future work we can aim to achieve the voice command control system. [5]

REFERENCES

[1]. Jungong Han; Ling Shao; Dong Xu; Shotton, J., "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review," in Cybernetics, IEEE Transactions on, vol.43, no.5, pp.1318-1334, Oct. 2013 DOI: 10.1109/TCYB.2013.2265378.

[2]. P. Baerlocher and R. Boilic, Inverse Kinematics Techniques for the Interactive Posture Control of Articulated Figures, PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2001.

[3]. Machida, E.; Meifen Cao; Murao, T.; Hashimoto, H., "Human motion tracking of mobile robot with Kinect 3D sensor," in SICE Annual Conference (SICE), 2012 Proceedings of, vol., no.,pp.2207- 2211,20-23 Aug.2012.

[4]. Xu, D. and Acosta, A. (2005) An Analysis of the Inverse Kinematics for a 5- DOF Manipulator. International Journal of Automation and Computing, 2, 114-124. http://dx.doi.org/10.1007/s11633- 005- 0114-1.

[5]. EvangelosKalogerakis, Aaron Hertzmann, and Karan Singh. Learning 3d mesh segmenta-tion and labeling. ACM Trans. Graph., 29(4), 2010.