e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

VIRTUAL YOGA SYSTEM USING KINECT SENSOR

Praful Wagh[*1], Sachin Patil[*2], Rushikesh Shrivastav[*3], Shubham Bachhav[*4]

[*1,2,3,4]Students, Department Of Computer Engineering, Loknete Gopinathji Munde Institute Of Engineering Education & Research, Nashik, Maharashtra, India.

ABSTRACT

Physical Yoga has become a big part of life for patients suffering from orthopedic injuries. But as time goes by, patients tend to get tired and demotivated due to repetitive and tedious exercises the recognition of poses is a field of investigation that takes incredible significance for oneself preparing in different sports. Due to its body tracking and depth sensor, Kinect provides a low-cost option for recognising Yoga positions. We propose an interactive system for perceiving a few postures for learning Yoga that will be characterised by a level of difficulty and coordinated with command voices to imagine the guidelines and pictures about the stances to be executed in this research. It also allows a therapist to customize exercises according to the specific needs and challenges of individual patients. The Microsoft Kinect sensor recognises skeleton and joint positions and provides three-dimensional information about the user's body. This enables an immersive and natural interaction between the user and the system. We use the Microsoft Kinect sensor to detect and track the movements of user and speech out message form system.

Keyword: Kinect Sensor, Motion Capture, Virtual Physical Therapy, Serious Game, Rehabilitation, Interface Video Game, Interactive Technologies, Movement Recognition.

I. INTRODUCTION

Physical therapy are optimal when assessment, monitoring, adherence to the therapy program and patient engagement can be achieved. Different procedures are involved in standard physical therapy and rehabilitation practise. These processes are usually intensive, time consuming, dependent on the expertise of the therapist and implies the collaboration of the patient who is usually asked to perform the therapy multiple times at home with no supervision. Injury or disease to the bone, articular, or soft tissue structures of the joints can result in discomfort, as well as a reduction in mobility, strength, stability, and functional use of the upper and lower extremities body traditional methods of rehabilitation is quite tedious and boring. As exercises are to be done for a long period of time, patients tends to loose interest over time because of monotonous exercises routines. Also, traditional physiotherapy involve use of goniometer which is a very old device used to measure degrees of flexion and extension of joints but it is inaccurate.

The Microsoft Kinect Sensor has a number of advanced sensing components. It has a depth sensor, a colour camera, and a four-microphone array for full-body 3D motion capture, facial recognition, and voice recognition, among other features. A Kinect-based system can help patients do rehabilitation exercises correctly, promote patient accountability, and allow clinicians to rectify exercise faults. All these features allow to develop an integrated Human Computer Interaction (HCI) system in which happens as an open-ended dialog between the user and the computer. This system also provides some command voices so that the user can interact with the program such as the possibility of changing different Yoga poses instructions for a correct position. The System will prepare yoga program for the person. User should begin the system and kinect sensing element to perform the exercise. The voice instruction and, as a result, the action taken by the user will be the device's input. The kinect sensing element can locate the figure's twenty body points and relay the information to the system. The system is capable of calculating both RGB and depth images. The system may then calculate the movement and compare it to the rule-based instruction. If a user performs well, the system will reward them and direct them to the next step instruction to perform next yoga poses. Also, we can make weekly report of yoga exercise. Microsoft has announced new sensors that can establish watershed in RGB-D cameras for use with gaming consoles. There are three types of sensors: RGB, audio, and depth. Both of them perform important responsibilities, such as detecting movement, allowing users to play games using their bodies as a controller, and identifying player sounds. This Microsoft Kinect development also aided other computer vision applications like robotics and action identification.

In Kinect sensor the arrangement of the infrared (IR) projector, the color camera, and the IR camera. The IR projector is paired with the IR camera, which is a monochrome complementary metal oxide semiconductor (CMOS) sensor, to form the depth sensor. The depth-sensing technology is licensed from the Israeli company Prime Sense. Although the particular technology isn't revealed, it is based on the notion of structured light. The IR projector is a series of IR dots created by an infrared laser passing through a diffraction grating. Gesture recognition system has been divided into five different phases. They are data acquisition, segmentation and pre-processing, feature extraction and finally the recognition. Data acquisition involves capturing the image using Microsoft Kinect. The image needs to be segmented in order to remove the unwanted background details. The third phase of this process is the image processing which is further subdivided into a number of steps which include noise removal, edge detection, contour detection, and normalization of the image to add the accuracy of the detection. The features are then extracted from the segmented and pre-processed image for recognition. Finally, the image is recognized as a meaningful gesture based on the gesture analysis.

II. LITERATURE SURVEY

[1] The ability to recognise gestures is critical for human-machine interaction. In this study, we suggest a method for employing a Kinect depth camera to identify human motions. The camera sees the topic in the front plane and creates a depth image of it in the plane facing the camera. This depth image is then utilised to remove the background before generating the subject's depth profile. Furthermore, the difference between succeeding frames determines the subject's motion profile, which is utilised to recognise motions.

[2] Based on new learnable triangulation approaches that incorporate 3D information from many 2D views, we describe two unique solutions for multi-view 3D human posture estimation. A basic differentiable algebraic triangulation with the addition of confidence weights computed from the input images is the initial (baseline) solution. The second approach is based on a unique volumetric aggregation method based on intermediate 2D backbone feature maps. The combined volume is subsequently adjusted with 3D convolutions, resulting in final 3D joint heatmaps and the ability to model a human position beforehand. Importantly, both algorithms are end-to-end differentiable, allowing us to optimise the goal metric directly. On the Human3.6M dataset, we show that the solutions are transferable across datasets and that they significantly improve the multi-view state of the art.

[3] We show how a multi-camera system may be used to train fine-grained detectors for keypoints that are prone to occlusion, such as hand joints. This method is known as multiview bootstrapping: first, a keypoint detector is employed to generate noisy labels in various hand views. The outliers are then triangulated in 3D using multiview geometry. Finally, to strengthen the detector, the reprojected triangulations are used as fresh labelled training data. We repeat the process, each time generating more labelled data. The minimal number of views required to obtain goal true and false positive rates for a given detector is calculated mathematically. A hand keypoint detector is trained using this way.

[4] Deep High-Resolution Representation Learning for Human Pose Estimation is an official pytorch implementation. The human pose estimation problem is the topic of this research, with a focus on learning trustworthy high-resolution representations. A high-to-low resolution network produces low-resolution representations, and most available approaches recover high-resolution representations from these low-resolution representations. Our suggested network, on the other hand, keeps high-resolution representations throughout the entire process. We begin with a high-resolution subnetwork as the first stage, then add high-to-low resolution subnetworks one by one to construct additional stages, and then join the multi-resolution subnetworks in parallel. We perform many multi-scale fusions so that each of the high-to-low resolution representations receives information from other parallel representations on a regular basis, resulting in rich high-resolution representations.

[5] We offer the first real-time method for capturing a human's whole global 3D skeletal posture using a single RGB camera in a stable, temporally consistent manner. A new convolutional neural network (CNN)-based pose regressor is combined with kinematic skeleton fitting in our method. Our new fully-convolutional pose formulation simultaneously regresses 2D and 3D joint positions in real time and does not require tightly clipped input frames. On the basis of a coherent kinematic skeleton, a real-time kinematic skeleton fitting approach employs CNN output to provide temporally robust 3D global posture reconstructions. Our method is

the first monocular RGB method that can be used in real-time applications like 3D character control; previously, the only monocular methods for such applications required specialised RGB-D cameras. Our results are qualitatively equivalent to, and in some cases superior than, those obtained using monocular RGB-D techniques like the Kinect. We show, however, that our approach is more widely applicable than RGB-D solutions, as it works for outdoor situations, community recordings, and low-quality commodity RGB cameras.

III. MOTIVATION

Following the current market trend, this project intends to take Exercise expertise a step further by offering a totally new method to the use of read work, as there are no comparable programmes for Windows-based PCs. Specifically, the gestures for the virtual workout will be supplied to the Purposed system. The integration of the Kinect sensing element and thus the management over diversion application, which distinguishes it from other similar virtual applications, is the most important goal of the appliance, which distinguishes it from other similar virtual applications; that is, the lack of need to hold or possibly touch a keyboard, mouse, or controller device to play the sport. This application model can contain a lot of pre-programmed motions, but the actual diversion setting will be able to replicate your movements like in motion detecting games. However, you will gain knowledge of motion sensing games as a result of this. For explicit virtual workout, this programme might be provided a gesture set. As a result, the consumer may only require shopping for the Kinect sensing device and not the gaming console.

IV. PROBLEM DEFINITION

For years, the traditional medical science method to physiotherapy has been in use, and it has been highly laborious for the instructor and dull for the patient. Issues such as Patients must adhere to their workout programmes on a consistent basis for effective elbow rehabilitation. Traditional workout programmes are repetitious, unpleasant, and time-consuming. The patient loses interest and becomes fatigued when using obsolete tools like goniometers. To address this issue, a novel motion detection approach based on the Kinect Sensor can be applied to physiotherapy in a way that patients would find intriguing and engaging. As a result of this notion, a Virtual Yoga System based on the Kinect Sensor might be created to provide new and creative ways to recover, making treatment more fun and thus increasing motivation. Physiotherapists should focus on improving their skills in this area. As technology advances in the health industry, it is becoming a part of therapy and treatment alternatives. Furthermore, it decreases workload by effectively utilising physiotherapy time while still providing therapy.

V. EXISTING SYSTEM

Another group of professors directs the scholars into positions and allows them to feel them to gain a better comprehension. In general, the majority of possibilities for persons who are blind or have low vision to interact in yoga have required contact with a yoga teacher who has the necessary information and knowledge to accommodate them. Multiple sets of CDs were made available in homes for practising yoga, so that any busy person might do so without feeling obligated to do so. Exergames, or workout video games, have been a popular way to increase exercise participation in recent years. Exergames will provide fitness activities and serve as a springboard to a variety of more advanced workouts. However, many people are unable to play these games due to physical limitations.

VI. PROPOSED SYSTEM

To execute the workout, the user must first turn on the equipment and the kinect sensor device. The voice instruction and, as a result, the action taken by the user will be the device's input. The kinect sensing element can locate the figure's twenty body points and relay the information to the system. The system is capable of calculating both RGB and depth images. The system may then calculate the movement and compare it to the rule-based instruction. If the user performs well, the system can congratulate them and provide them instructions on how to execute the next exercise.

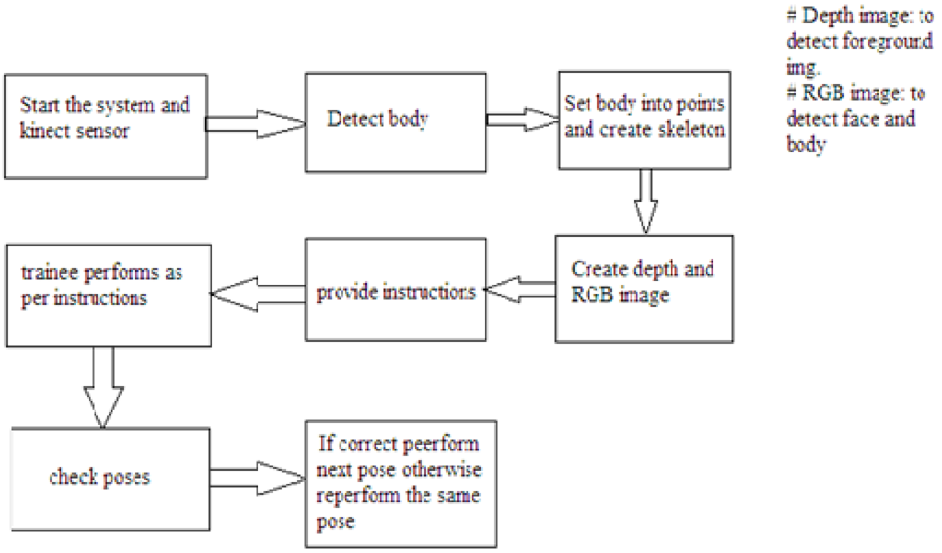
Start the system and kinect sensor
Detect body
Set body into points and create skeleton
# Depth image: to detect foreground img.
# RGB image: to detect face and body
Create depth and RGB image
provide instructions
trainee performs as per instructions
check poses
If correct peerform next pose otherwise reperform the same pose

Fig 1: Workflow Diagram

VII. SYSTEM IMPLEMENTATION

The configuration of the infrared (IR) projector and the colour camera in the Kinect sensor, and the IR camera. The depth sensor is made up of an infrared projector and an infrared camera, both of which are monochrome complementary metal oxide semiconductors. (CMOS) sensor. The depth-sensing technology is licensed from the Israeli company Prime Sense. Despite the fact that the particular technology isn't revealed, it is based on the structured light principle. The IR projector is a series of IR dots created by an infrared laser passing through a diffraction grating. Gesture recognition system has been divided into five different phases. They are data acquisition, segmentation and pre-processing, feature extraction and finally the recognition. Data acquisition involves capturing the image using Microsoft Kinect. The image needs to be segmented in order to remove the unwanted background details. The third phase of this process is the image processing which is further subdivided into a number of steps which include noise removal, edge detection, contour detection, and normalization the image to add the accuracy of the detection. The features are then extracted from the segmented and pre-processed image for recognition. Finally, the image is recognized as a meaningful gesture based on the gesture analysis.

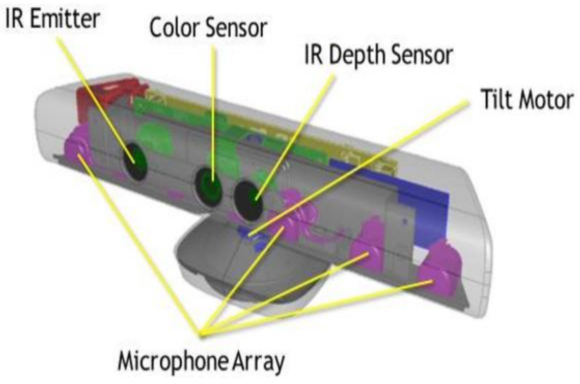

Fig 2: Kinect Sensor Diagram

The Kinect sensor bar is a Microsoft RGB-D camera that assists in the development of gesture recognition systems by recording depth information and turning it into 3D coordinates. The kinect sensor is shown in Figure 2. The kinect sensor consists of the following parts:-

1. Infrared Sensor

The IR sensor, also known as an IR emitter, emits infrared rays on the individual who is executing the required gesture that the system must recognise. Even in low-light situations, the IR sensor aids in capturing the subject. It's compatible with the depth camera.

2. Depth Camera

The depth camera, in conjunction with the IR sensor, aids in the capture of depth images of the subject, such as the joint configuration data of a person making a certain gesture. The depth camera captured a sample image.

3. RGB Camera

The RGB camera of the Kinect sensor is a normal camera, just like the other ones. The RGB camera has a 1280x960 resolution which is capable of storing three-channel data, thus making it possible to capture a color image.

4. Multi-Array Microphone

A multi-array microphone contains four microphones for capturing sound. The four microphones make it possible to record audio as well as find the location of the sound source and the direction of the audio wave.

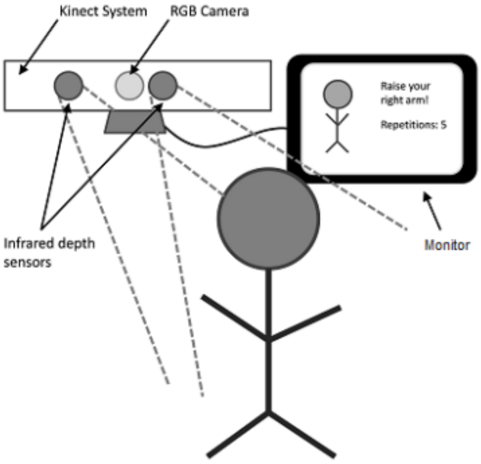

Fig 3: Overview of intended Setup

1. MODULES

A module is a software or hardware item that is separate from the rest of the system. The three aspects of our proposed system are as follows:-

1. Skeleton Detection and Tracking

2. Recognition of gestures

3. Recognize Yoga Posture

4. Text to Speech (TTS)

1. Skeleton Detection and Tracking

We can detect the skeleton and track the joint points of the human body in this module for yoga posture. We can easily identify the pixels that represent the players using the raw depth data returned by the Kinect sensor. Skeleton tracking doesn't just track the joints by reading the player's data; it also tracks the entire body's movement.

2. Recognition of Gestures

The skeleton structures were successfully discovered in this module thanks to the clever advancement of some gadgets that employed the kinect sensor to assess depth. These kinect sensors have been used to observe human body movements, and they can give sufficient accuracy while tracking the entire body in real-time mode with low latency.

3. Recognize Yoga Posture

In this module, we will execute yoga postures and then go on to the next one. We will also calculate yoga pose estimation and classify it.

4. Text to Audio

This module uses a text-to-speech method to deliver the message. It's a multi-language voice synthesizer that turns digital text into natural-sounding speech. It can be easily integrated into any embedded system thanks to its simple command-based interface.

VIII. FUTURE SCOPE

This system has a lot of potential in the future. While it clearly attempts to increase people's motivation to exercise and become in shape, it may be developed and modified in such a way that it produces a personalized training model for them, tailored to their lifestyle and needs. The device is expected to provide precision that is comparable to that of a personal trainer. The possibilities for future research are endless, and they promise to improve aged care patients' motivation to accurately report their yoga sessions. At no expense, the elderly can have simple access to the most engaging and interactive yoga system practises.

IX. CONCLUSION

With the advancement of technology, our daily lives have gotten more easy, and people are losing more and more time and desire for physical activity. As a result, several technologies and apps will be developed to make exercising more convenient and accessible anywhere. This new accessible Exercises will allow any igroup of individuals to access exercise while in reception, which will benefit both their physical and mental health. This new accessible Exercises will allow any group of individuals to access exercise while in reception, which will benefit both their physical and mental health. The Kinect continues to progress, update, and evolve into a very viable tool for joint injury rehabilitation. As a result, Kinect has found a solution to this difficulty. Furthermore, in order to improve the user interface, we are considering adding more elements in the future, such as an award mechanism and other plugins.

X. REFERENCES

[1] K. Aberman, P. Li, D. Lischinski, O. Sorkine-Hornung, D. Cohen-Or, and B. Chen. Skeleton-aware networks for deep motion retargeting. ACM Transactions on Graphics (TOG), 39(4):62, 2020.

[2] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. OpenPose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

[3] E. Aksan, M. Kaufmann, and O. Hilliges. Structured prediction helps 3d human motion modelling. In Proceedings of the IEEE International Conference on Computer Vision, pages 7144–7153, 2019.

[4] A. Haque, B. Peng, Z. Luo, A. Alahi, S. Yeung, and L. FeiFei. Towards viewpoint invariant 3d human pose estimation. In Proc. ECCV, pages 160–177. Springer, 2016.

[5] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov. Learnable triangulation of human pose. In Proc. ICCV, pages 7718–7727, 2019.

[6] Pavlovic, V.I. and Sharma, R. and Huang, T.S., ”Vision-Based Gesture Recognition: A Review”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997.

[7] J. Romero, D. Tzionas, and M. J. Black. Embodied hands: Modeling and capturing hands and bodies together. ACM Transactions on Graphics (ToG), 36(6):245, 2017.

[8] T. Simon, H. Joo, I. Matthews, and Y. Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In Proc. CVPR, 2017.

[9] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In Proc. CVPR, pages 5693–5703, 2019.