International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

CAPTURING THE MOTION OF A PERSON USING COMPUTER VISION

[1]Md Toufeeq Ahmed , [2] MekaPoojaReddy, [3] Surabhi NavyaSri

[1] Assistant Professor, Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women, Hyderabad, Telangana, India.

[1] toufeeq002@gmail.com

[2,3]Students, Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women, Hyderabad, Telangana, India.

[2] poojareddie09@gmail.com , [3] navya16surabhi@gmail.com

Abstract

3D motion capture is a technology that captures the movements of a person or object in three dimensions. It involves attaching sensors or markers to the body or object and using specialized software to track their movement and translate it into a digital model. This technology is commonly used in the film and gaming industries to create realistic animations and special effects, as well as in sports and biomechanics research to analyze and improve athletic performance. In addition, 3D motion capture can be used in virtual reality and augmented reality applications, allowing users to interact with digital environments in a more natural and immersive way. The use of 3D motion capture has become increasingly widespread in recent years, with advances in hardware and software making it more accessible and easier to use. Now that machine learning and artificial intelligence are growing in popularity, we are researching the use of artificial Intelligence and machine learning in the field of performance capture for the entertainment industry.

The current approaches are typically too pricey for the general people. We are focusing on combining artificial intelligence with third-party technologies (like Unity) to provide outstanding motion capture results. Input costs will decrease as a result, and innovation will rise. In this project, we will talk about developing a 3D motion capturing model using OpenCV and replicating the motion in Unity.

With the popularity of King Kong, Pirates of the Caribbean 2, Avatar, and other films. The virtual Characters in these films have become popular and well loved by audiences. The creation of these virtual characters is different from traditional 3D animation but is based on real character movements and expressions. An overview of several mainstream motion capture systems in the field of motion capture is presented, and the application of motion capture technology in film and animation is explained in detail. The current motion capture technology is mainly based on complex human markers and sensors, which are costly, while deep- learning- based human pose estimation is becoming a new option. However, most existing methods are based on a single person or picture estimation, and there are many challenges for video multi person estimation. The experimental results show that a simple design of the human motion capture system is achieved.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

I INTRODUCTION

The expansion of technological impact in the entertainment sector has shown to be effective. Motion Capture (MoCap) is the process of using cameras and specifically built suits to capture the movement of objects or people. Its impact on the entertainment business has been significant. Motion capture has been popular in movies and video games since the 1970s. In comparison to previous procedures, it has been complimented for its reduced latency, replication of complex moves, and data production. The demand for specific software and hardware, as well as the expense of relevant inputs, has been widely criticized. The goal of this project is to drastically lower the cost of MoCap. The app eliminates the need for sophisticated lenses and motion capture suits. With the help of Python's. OpenCV, scripting and Unity, we were able to complete the job. The user uploads video to the app, which is then pre-processed with OpenCV & Unity, which aids in the creation of animated clips of the user. These procedures demonstrate the capacity to capture performance in a coherent way.

With the popularity of King Kong, Pirates of the Caribbean 2, Avatar, and other films, the virtual characters in these films have become popular and well loved by audiences. The creation of these virtual characters is different from traditional 3D animation but is based on real character movements and expressions. An overview of several mainstream motion capture systems in the field of motion capture is presented, and the application of motion capture technology in film and animation is explained in detail. The current motion capture technology is mainly based on complex human markers and sensors, which are costly, while deep-learning- based human pose estimation is becoming a new option

However, most existing methods are based on a single person or picture estimation, and there are many challenges for video multi person estimation.The experimental results show that a simple design of the human motion capture system is achieved.Motion capture technology is more mature and common in the film and television industry. After capturing the motion data of professional actors, doing specific processing, and thenbinding with the character model in the film and television works, we can get 3D virtual animation .The currently used pose capture system is mainly divided into two categories: sensor capture and optical capture. The former is more mature, characterized by fast transmission speed and more accurate pose data; the disadvantage is the higher cost, and wearable devices are less convenient to use. In contrast, optical capture is the opposite, and there are two types of optical capture: unmarked and marked.

The object of this Project is a marker less capture system, where a common 2D image or video is used as input to capture the human body's joint point data using target detection and feature extraction. Although it is not yet widely used due to its unstable performance, its advantages such as ease of use, flexibility, and low cost should not be overlooked.In recent years, numerous scholars at home and abroad have proposed considerable convolutional neural network models and other auxiliary methods for human pose estimation, covering single to multi person, 2D to 3D, and picture to video. However, human pose is a complex nonlinear model, and environmental noise, occlusion, and spatial depth ambiguity are the main hindrances to this task. If the input object is video data, it is also a difficult task to output a high frame rate and smooth and stable pose. Most of the existing methods are based on image-based 3D pose estimation,or for single-person videos. The actual motion capture application objects are many times facing multi personscenes, and the characters must have contact with the virtual physical space, so we propose a 3D multi person estimation model from the video to meet the practical needs. There are two general types of 3D pose estimation: one is regressive regressing the 3Dcoordinates of the nodes directly from 2D data, which require the data to be 3D

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

labeled, which is often difficult to obtain and the other is the lifting type where the two-dimensional pose is first obtained and then a mapping method is trained to lift it to the three- dimensional space on top of the two- dimensional one

II LITERATURE SURVEY

The proposed system first recognizes the body in an input image or the video, then the model extracts image/video coordinates. The Unity software helps to create a model using basic scripting of python & C#. The external program allows us to generate an animated rendition of our model, which is efficient enough in comparison to Mo Cap suits.3D motion capture, also known as motion capture or mocap, is a technology that capturesthe movements of a person or object in three dimensions. It is often used to create realistic animations and special effects in film, television, and video games, as well as in sports andbiomechanics research to analyze and improve athletic performance. To capture motion, sensors or markers are attached to the body or object. These markers can be small reflective balls, LED lights, or other types of sensors. The markers are then tracked by specialized cameras or other sensors that are placed around the capture area. As the person or object moves, the markers are tracked in real-time, and the data is fed into software that translates the movement into a digital model. The digital model can then be used to createa3D animation or to analyze the movement of the person or object. In the film and gaming industries, 3D motion capture is often used to create realistic character animations and special effects. In sports and biomechanics, it can be used to analyze the movements of athletes and identify areas for improvement. 3D motion capture can also be used in virtual reality and augmented reality applications, allowing users to interact with digital environment sin a more natural and immersive way. In these applications, users wear specialized headsets or other devices that track their movements and translate them into the digital environment. Overall, 3D motion capture is a powerful and versatile technologythat is used in a variety of industries to create realistic animations, analyze movement, and enhance the way we interact with digital environments.In comparison to previous procedures, it has been complimented for its reduced latency, replication of complex moves, and data production. The demand for specific software andhardware, as well as the expense of relevant inputs, has been widely criticized. The goal of this project is to drastically lower the cost of MoCap. The app eliminates the need for sophisticated lenses and motion capture suits. With the help of Python's OpenCV, scripting and Unity, we were able to complete the job. OpenCV (Open Source Computer Vision) is a free and open- source library of computer vision and machine learning algorithms designed to help developers build applications that can analyze, understand, and manipulate visual data It was originally developed by Intel in 1999 and has since become one of the most widely used libraries in the field of computer vision, with a large community of contributors and users around the world.OpenCV provides a wide range of tools and functions for image and video processing, including feature detection, object recognition, and image stitching. It also offers-machine learning algorithms for tasks such as classification, cluster and regression. OpenCV is written in C++ and has interfaces for Python, Java, and MATLAB, making it easy to use in a variety of programming languages and environments. One of the key strengths of OpenCV is its ability to handle a wide range of visual data, including images, videos, and depth maps. It can be used for tasks such as image classification, object detection, face recognition, and augmented reality. Unity is a cross-platform game engine and development platform that is often used in conjunction with 3D motion capture. Unity provides a range of tools and features for creating 3Danimations, including the ability to import and manipulate 3D models and animations, create custom materials and sharers, and build interactive environments. When using Unity for 3D motion capture, developers can

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

import motion capture data captured using specialized hardware and software into the Unity editor. This data can then be used to drive the animation of 3D models or characters in the Unity scene. Unity also provides a range of tools and features for smoothing and refining motion capture data, such as the ability to re target animations to different character rigs and the ability to blend between different animations to create smooth transitions. In addition to its animation tools, Unity also provides a range of features for building interactive 3D environments, including support for physics, lighting, and audio. This makes it an ideal platform for creating immersive 3D experiences that utilize motion capture data, such as interactive games, simulations, and virtual reality applications.

The user uploads video to the app, which is then per-processed with OpenCV& Unity, which aids in the creation of animated clips of the user. These procedures demonstrate the capture performance in a coherent way.we need to combine the model with the captured motion data to achieve matching with the model, so as to drive the movement of the model. Finally, the model is matched with the captured data, and the model can follow the captured motion data to move

III PROBLEM DEFINITION

Object detection is usually achieved by object detectors or background subtraction. An object detector is often a classifier that scans the image by a sliding window and labels each sub image defined by the window as either object or background. Generally, the classifier is built by offline learning on separate datasets or by online learning initialized with a manually labeled frame at the start of a video. Alternatively, background subtraction compares images with a background model and detects the changes as objects. It usually assumes that no object appears in images when building the background model. Such requirements of training examples for object or background modeling actually limit the applicability of above-mentioned methods in automated video analysis Another category of object detection methods that can avoid training phases are motion- based methods which only use motion information to separate objects from the background. Given a sequence of images in which foreground objects are present and moving differently from the background, can we separate the objects from the background automatically. The goal is to take the image sequence as input and directly output a mask sequence. The most natural way for motion-based object detection is to classify pixels according to motion patterns, which is usually named motion segmentation. These approaches achieve both segmentation and optical flow computation accurately and they can work in the presence of large camera motion. However, they assume rigid motion or smooth motion in respective regions, which is not generally true in practice. In practice, the foreground motion can be very complicated with no rigid shape changes. Also, the background may be complex, including illumination changes and varying textures such as waving trees and sea waves. The video includes an operating escalator, but it should be regarded as background for human tracking purpose. An alternative motion-based approach is background estimation. Different from background subtraction, it estimates a background model directly from the testing sequence. Generally, it tries to seek temporal intervals inside which the pixel intensity is unchanged and uses image data from such intervals for background estimation. However, this approach also relies on the assumption of static background. Hence, it is difficult to handle the scenarios with complex background or moving cameras.

A novel algorithm is proposed for moving object detection which falls into the category of motion based methods. It solves the challenges mentioned above in a unified framework named Detecting Contiguous Outliers in the LOw-rank Representation (DECOLOR). We assume that the underlying background images are linearly correlated.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Thus, the matrix composed of vectorized video frames can be approximated by a low-rank matrix, and the moving objects

can be detected as outliers in this low-rank representation. Formulating the problem as outlier detection allows us to get rid of many assumptions on the behavior of foreground. The low-rank representation of background makes it flexible to accommodate the global variations in the background. Moreover, DECOLOR performs object detection and background estimation simultaneously without training sequences. The main contributions can be summarized as follows

A new formulation is proposed in which outlier detection in the low-rank representation in which the outlier support and the low-rank matrix are estimated simultaneously. We establish the link between our model and other relevant models in the framework of Robust Principal Component Analysis (RPCA). Differently from other formulations of RPCA, we model the outlier support explicitly. We demonstrate that, although the energy is no convex, DECOLOR achieves better accuracy in terms of both object detection and background estimation compared against the state-of-the-art algorithm of RPCA .

In other models of RPCA, no prior knowledge on the spatial distribution of outliers has been considered. In real videos, the foreground objects usually are small clusters. Thus, contiguous regions should be preferred to be detected. Since the outlier support is modeled explicitly in our formulation, we can naturally incorporate such contiguity prior using Markov Random Fields (MRFs). Use a parametric motion model to compensate for camera motion. The compensation of camera motion is integrated into our unified framework and computed in a batch manner for all frames during segmentation and background estimation. Background subtraction is a widely used for detecting moving objects. The ultimate goal is to“subtract” the background pixels in a scene leaving only the foreground objects of interest. If one has a model of how the background pixels behave the “subtraction” process is very simple Background subtraction usually consists of three attributes besides the basic structure of the background model, background initialization background maintenance (updating the background model to account) and foreground/background pixel classification.

IV SYSTEM IMPLEMENTATION

In the area of moving object detection a technique robust to background dynamics using background subtraction with adaptive pixel-wise background model update is described. A foreground-background pixel classification method using adaptive thresholding is presented. Another technique that is robust to sudden illumination changes using an illumination model and a statistical test is presented. We also propose a fast implementation of the method using an extension of integral images. a novel and simple method for moving object detection. The method is based on background subtraction. The first step, referred to as background model initialization,

is to construct a background model using the initial frames. The step is based on temporal frame differencing because the background model is not available at the start of a sequence. A typical example is when a sequence starts with a moving foreground object, part of the background model will be covered and hence the background model will not be available. Once the initial background model is constructed, with each subsequent frame, we detect if there is any sudden illumination change. If there is no illumination change, simple background subtraction can be used to find the foreground pixels. The threshold used for this process is determined adaptively according to the current frame. After thresholding the difference of the pixel intensities between the background model and the current frame, a foreground object mask is generated. In the presence of sudden illumination change, we use the illumination effect, which is expressed as a ratio of pixel intensities. We determine if the pixel is a foreground or

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

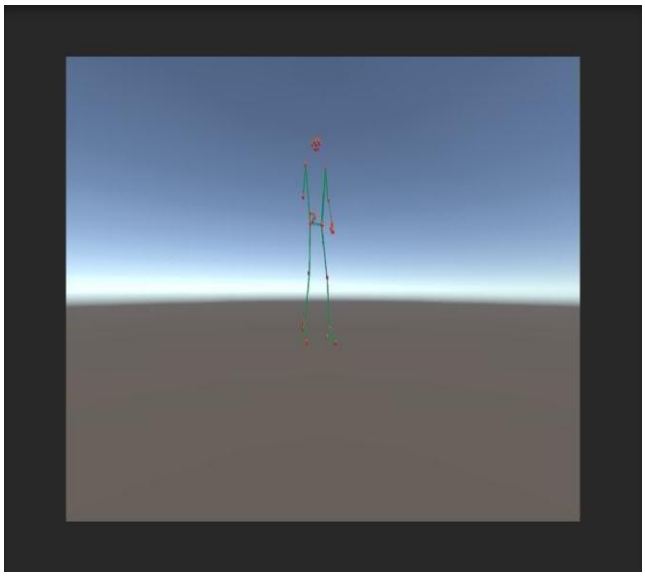
background pixel by the statistics of the illumination effect of its neighboring pixels. A foreground object mask is generated according to this statistic. The process of obtaining the foreground mask is referred to as foreground-background pixel classification. After obtaining the foreground object mask, the background model should be updated because the scene dynamics are always evolving. We propose a technique that updates the background model according to two factors: how long a pixel has been a background pixel, and how large the value of the pixel is in the difference frame. The first challenge is that it is not a convex problem, because of the no convexity of the low rank constraint and the group sparsity constraint. Furthermore, we also need to simultaneously recover matrix B and F, which is generally a Chicken-and-Egg problem. In our framework, alternating optimization and greedy methods are employed to solve this problem.

The first focus on the fixed rank problem (i.e., rank equals to 3), and then will discuss how to deal with the more general constraint of rank ≤ 3. 4 is divided into two sub problems with unknown B or F, and solved by using two steps iteratively: To initialize this optimization framework, we simply choose Binit = φ, and Finit = 0. Greedy methods are used to solve both sub problems. Then the αk rows with largest values are preserved, while the rest rows are set to zero. This is the estimated F in the first step. In the second step, φ is computed as per newly-updated F.. Singular value decomposition (SVD) is applied on φ. Then three eigenvectors with largest eigenvalues are used to reconstruct

B. Two steps are alternatively employed until a stable solution of ˆB is found. The greedy method of solving Eq. 5 discovers exact αk number of foreground trajectories, which may not be the real foreground number. On the contrary, B can be always well estimated, since a subset of unknown number of background trajectories is able to have a good estimation of background subspace. Since the whole framework is based on greedy algorithms, it does not guarantee a global minimum. In our experiments, however, it is able to generate reliable and stable results. The above-mentioned method solves the fixed rank problem, but the rank value in the background problem usually cannot be pre- determined. To handle this undetermined rank issue, we propose a multiple rank iteration method. Then the fixed rank optimization procedure is performed on each specific rank starting from 1 to 3. The output of the current fixed rank procedure is fed to the next rank as its initialization. We obtain the final result B(3) and F(3) in the rank-3 iteration. Given a data matrix of K × 2L with K trajectories over L frames, the major calculation is O(KL2) for SVD on each iteration. Convergence of each fixed rank problem is achieved iterations on average. The overall time complexity is O(KL2). To explain why our framework works for the general rank problem, we discuss two examples. First, if the rank of B is 3 (i.e., moving cameras), then this framework discovers an optimal solution in the third iteration, i.e., using rank-3 model. The reason is that the first two iterations, i.e. the rank- 1 and rank-2 models, cannot find the correct solution as they are using the wrong rank constraints. Second, if the rank of the matrix is 2 (i.e., stationary cameras), then this framework obtains stable solution in the second iteration. This solution will not be affected in the rank-3 iteration. The reason is that the greedy method is used. When selecting the eigenvectors with three largest eigenvalues, one of them is simply flat zero. Thus B does not change, and the solution is the same in this iteration. Note that low rank problems can also be solved using convex relaxation on the constraint problem [2]. However, our greedy method on unconstrained problem is better than convex relaxation in this application. Convex relaxation is not able to make use of the specific rank value constraint (≤ 3 in our case). The convex relaxation uses λ to implicitly constrain the rank level, which is hard to constrain a matrix to be lower than a specific rank value. First demonstrate that our approach

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

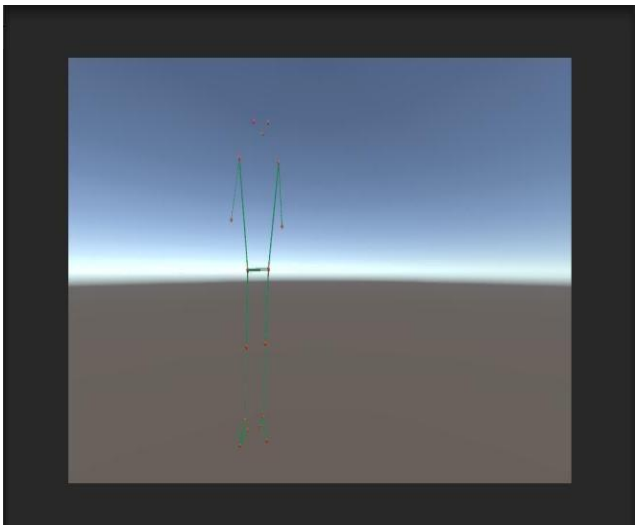
handles both stationary cameras and moving cameras automatically in a unified framework, by using the General Rank constraint (GR) instead of the Fixed Rank constraint (FR). Here use two videos to show the difference. One is "VHand" from a moving camera (rank(B) = 3), and the



other is "truck" captured by stationary camera (rank(B) = 2). The use distribution of L2 norms of estimated foreground trajectories to show how well background and foreground is separated in our model. For a good separation result, F should be well estimated. Thus large for foreground trajectories and small for background ones. In other words, it distribution has an obvious difference between the foreground region and the background region. The FR model also finds a good solution, since rank-3 perfectly fits the FR model. However, the FR constraint fails

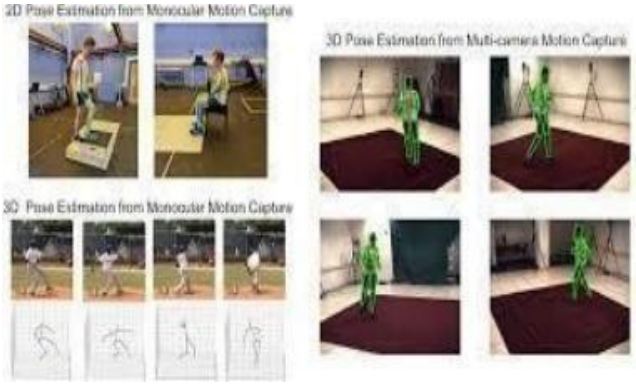
when the rank of B is 2, where the distribution of between B and F are mixed together. On the other hand, GR−2 can handle this well, since the data perfectly fits the constraint. On GR−3 stage, it uses the result from GR−2 as the initialization, thus the result on GR−3 still holds. The figure shows that the distribution of _ ˆ Fi_2 from the two parts has been clearly separated in the third column of the bottom row. This experiment demonstrates that the GR model can handle more situations than the FR model. Since in real applications it is hard to know the specific rank value in advance, the GR model provides a more flexible way to find the right solution. The algorithm is implemented in MATLAB. All experiments are run on a desktop PC with a 3.4 GHz Intel i7 CPU and 3 GB RAM. Since the graph cut is operated for each frame separately, as discussed in Section 3.3.2, the dominant cost comes from the computation of SVD in each iteration. The CPU times of DECOLOR for sequences in Figs are 26.2,13.3, 14.1, 11.4, and 14.4 seconds, while those of PCP are26.8, 38.0,

15.7, 39.1, and 21.9 seconds, respectively. All results are obtained with a convergence precision of 10_4.The



memory costs of DECOLOR and PCP are almost the same since both of them need to compute SVD. The peak values of memory used in DECOLOR for sequences in Figures are around 65 MB and 210 MB, respectively.

V RESULTS



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

VI CONCLUSION

3D motion capture using Open CV is a powerful and widely used technology that allows for the real-time tracking and analysis of the three-dimensional movements of objects or people. Open CV is a free and open-source library of computer vision and machine learning algorithms that provides a number of tools and techniques for 3D motion capture, including feature extraction, object tracking, and deep learning-based approaches. This technology has a wide range of applications in industries such as film and gaming, sports and biomechanics research, and virtual and augmented reality. Overall, 3D motion capture using OpenCV is a valuable tool for creating realistic animations and special effects, analyzing and improving athletic performance, and enabling natural and immersive interactions with digital environments. We can see that the 3D simulation that we have created has accurately replicated the movements of a person in the video. This shows us the scope of this model and simulation software like Unity being used in a variety of fieldssuch as the film industry, game development industry and so on. Although animation is still in its infancy in terms of performance, motion capture technology has been around for a while. Even if there are now better tools for working with data, technical proficiency is still necessary. As a result of the rise in performance capture projects, more people are regularly using motion capture It was challenging to include all aspects of motion capture in a single essay, especially the ones that dealt with software. Motion capture has been expensive for the computer animation industry, but it is now widely acknowledged as a valuable tool in the visual effects arsenal, and it will remain so for a very long time.

The application of modern motion capture technology in the production of film and television animation. With the continuous development and improvement of motion capture technology, motion capture technology will surely get more and more important applications in our daily lives. The emergence of optical motion capture systems has greatly reduced the cost of filming and animation production and has made film and animation pictures realistic. The motion capture model in this article can capture the actions of multiple people in a non restricted environment. After the human body pose is corrected, it basically conforms to the real human motion. It works well in three-dimensional virtual characters.

The cameras differentiate a human from a background, and then identify he position of a number of features or joints, such as shoulders, knees, elbows and hands. Some systems can also track hands or specific gestures, though this is not true of all skeletal tracking systems. Once those joints are identified, the software connects them into a humanoid skeleton and tracks their position real time. This data can then be used to drive interactive displays, games, VR or AR experiences or any number of other unique integration, such as displaying your ‘shadow’ projected on the side of a real car. Using depth cameras of any kind allows the skeletal tracking system to disambiguate between overlapping or occluded objects or limbs, as well as making the system more robust to different lighting conditions than a solely 2D camera-based algorithm would. There are a number of skeletal tracking solutions today that support Intel® Real Sense™ depth cameras, including the recently launched cubemos Skeleton Tracking SDK

REFERENCES

1) International Journal of Research in Engineering and Science (IJRES) ISSN(Online): 2320-9364, ISSN (Print): 2320-9356 Volume 9 Issue 7 ǁ 2021 ǁ PP.

2) Yating Wei, Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production, Volume 2022 Hindawi journals .

3) International Journal of Research in Engineering and Science (IJRES) ISSN(Online): 2320-9364, ISSN (Print): 2320-9356, Volume 9 Issue 7 ǁ 2021 ǁ PP.

4) Müller, M., Röder, T., & Clausen, M. (2018). Motion Capture and Recognition Using Computer Vision. In

Artificial Intelligence and Computer Vision (pp. 227-242). Springer, Cham.

5) Loper, M., Mahmood, N., Romero, J., & Black, M. J. (2014). SMPL: A skinned multi- person linear model. ACM Transactions on Graphics (TOG), 34(6), 248.

6) Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: a review. Computer vision and image understanding, 73(3), 428-440.

7) Seo, Y. D., Han, S. S., & Ko, H. S. (2016). A real-time motion capture system using multiple Kinect v2 sensors for sport training. Journal of Sports Science and Medicine, 15(4), 638.

AUTHORS

Authors:

MekaPoojaReddy B. Tech scholar, Department of Electronics And Communication Engineering, Bhoj Reddy Engineering College for Women.Santhosh Nagar Cross Roads,Vinay Nagar, Saidabad, Hyderabad, Telangana-500059. An enthusiastic learnerwith highly motivational and leadership skills. Always willing to innovate new things,whichcanimprovetechnology.

Mr. Md Toufeeq Ahmed is working as Assistant Professor in Electronics and Communication Engineering department of Bhoj Reddy Engineering College for Women, Hyderabad. He completed his B.E. (ECE) and M.E. (Digital Systems)from Osmania University, Hyderabad. He has more than 6 years of teaching experience. His research interests include Analog IC design, Analog and Mixed signal IC design and Data Converters for lowpowe Rvlsi applications

Surabhi Navya Sri B.Tech Scholar, Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women. Santhosh Nagar Cross Roads,Vinay Nagar, Saidabad, Hyderabad, Telangana-500059. A highly skilled, talented and knowledgeable candidate with extensive knowledge in the field of electronics and eagerto learn innovative things which can develop the existing technology