3D HUMAN POSE ESTIMATION USING MACHINE LEARNING

Pradakshinaa P[1],Harini S[2], Dhivyamanohar C[3], Kiruthika V R[4]

[1] Student, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India
[2] Student, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India
[3] Student, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India
[4] AssistantProfessor, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India

ABSTRACT

Human pose estimation is a fundamental task in computer vision and artificial intelligence that involves the estimation of the spatial configuration of a human body in an image or video. Accurate pose estimation is crucial for a wide range of applications, including human-computer interaction, augmented reality, virtual reality, biomechanics, and action recognition. While 2D pose estimation can provide valuable information about the pose in image space, 3D human pose estimation aims to recover the three-dimensional positions of body joints, offering a more complete and informative representation of human movement. We propose a method that uses a convolutional neural network (CNN) to estimate human pose by analyzing the projection of the depth and ridge data, which represent local maxima in a distance transform map. To fully utilize the 3D information of depth points, we propose a method to project the depth and ridge data in various directions. The proposed projection method reduces the loss of 3D information, stack data can avoid joint drift, and CNN improves localization accuracy. Separate humans from the background using depth data and extract highlight data from human silhouettes. Project depth and elevation data to XY, XZ, and ZY planes. ResNet-101 accepts 6 rendered images and uses heatmaps to generate 2D heatmaps and offsets. Create 2D key points for each plane using the soft-argmax operation. Obtain detailed 3D joint positions using fully connected layers. In experiments on SMMC-10, EVAL, and ITOP datasets, the proposed method achieved improved pose estimation accuracy. The proposed method can eliminate the loss of 3D information and displacement of joint positions that may occur during human pose estimation.

Keyword - *Human pose Estimation, Spatial Configurations, three-dimensional space, RGB images*

1. INTRODUCTION

Human pose estimation is the task of finding the parameters of a human body model, such as the length and orientation of body parts (head, trunk, limbs, etc.) that fit an input image. Fast depth imaging devices can extract rich information from depth images, thereby simplifying human pose estimation. An approach to detect human joints from observed input images using a pre-trained body part detector. Used a joint detector based on geodesic features to locate body joints in depth data. A head-torso detector based on rabid candidates and a pattern matching algorithm for each limb, but required that the upper body and face be visible without obstruction. Human pose recognition approach that predicts the representation of mid-body parts for human pose estimation, but this prediction usually requires expensive training steps and large human pose space. We needed a large number of training examples to cover. Used regression forests to directly identify co-occurrences from the votes of each pixel, but the votes were modeled, which required more complex training. Described a method to convert depth data into an average representation without background subtraction. Although most of these detection approaches do not completely miss body parts, they can only detect visible parts, which suffers from occlusion problems and significantly reduces the accuracy of human pose estimation. A productive approach to find human joints by fitting a predefined human body model to the observed input images. An iterative closest point (ICP) method for human pose estimation and human body tracking, but it cannot be used due to computational complexity.

2. SCOPE

The motivation to work on 3D human pose estimation using machine learning comes from its transformative potential in different fields. In fields such as computer vision and robotics, accurate 3D gesture estimation enables robots and machines to better understand and interact with humans, facilitating safer and more intuitive human-machine collaboration. In the entertainment industry, it paves the way for more immersive virtual reality experiences and realistic character animations, increasing user interaction and realism. Additionally, in sports and healthcare, 3D pose estimation can aid injury prevention, rehabilitation, and performance analysis by providing detailed insight into human movement. Additionally, applications can be extended to security and surveillance to enhance the tracking and identification of individuals in crowded or complex environments. Finally, advances in 3D human pose estimation using machine learning will revolutionize the way we interact with technology, entertainment and education, and monitor and improve human well-being in various applications.

3. LITERATURE REVIEW

The literature survey led for this study investigates existing works and ongoing exploration led in the field of "3D Human Posture Assessment utilizing AI ". This part plans to give an exhaustive outline of the cutting edge strategies and systems, recognize holes in the flow research, and establish the groundwork for the proposed arrangement. In this part, we arrange and talk about existing chips at "3D Human Pose estimation using ML". Each word is given its comparison segment number.

3.1 MONOCULAR 3D HUMAN POSE ESTIMATION

Monocular 3d Human pose estimation in the wild using improved CNN supervision" presents a significant advancement in monocular 3D human pose estimation using Convolutional Neural Networks (CNNs). The authors introduced a two-stage framework that leverages both 2D and 3D information, achieving state-of-the-art results. The method is sensitive to occlusions and may not perform well in crowded scenes. Identified Gap: Addressing occlusion handling and improving robustness in complex environments.

3.2 MASK R-CNN

Mask R-CNN addresses the problem of instance segmentation, which involves not only detecting objects in an image but also segmenting them at the pixel level.The primary contribution of Mask R-CNN is the integration of instance segmentation into the existing object detection framework. It achieves state-of-the-art results on instance segmentation tasks while maintaining competitive object detection performance.

3.3 VOXELPOSE TO GAUGE 3D STANCES

VoxelPose to gauge 3D stances of numerous individuals from different camera sees. VoxelPose straightforwardly works in the 3D space and, consequently, tries not to settle on mistaken choices in every camera view. In particular, they initially anticipate 2D posture heatmaps for all perspectives (the stage one), then twist the heatmaps to a typical 3D space and develop a component volume took care of into a Cuboid Proposition Organization (CPN) to confine all individuals examples (the stage two), and, at last, build a better grained highlight volume and gauge a 3D posture for every proposition (the last stage).

4. METHODOLOGY

This project will mainly concentrate on creating and assessing a machine learning-based 3D human pose estimation model. Our project scope encompasses tasks such as gathering data, preparing the data, developing the model, training it, and evaluating its performance. It's important to recognize that the field of 3D pose estimation is extensive, and our project may not encompass all its complexities.

4.1 Convolutional Neural Network

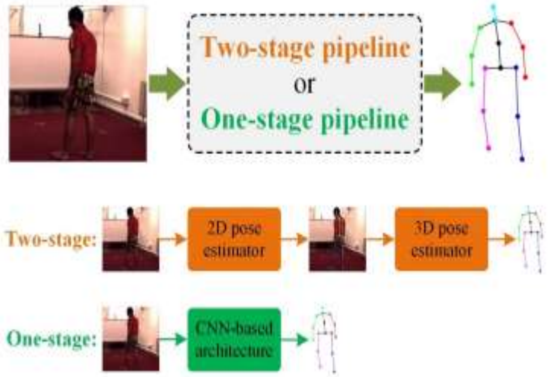
Convolutional neural networks, a kind of artificial neural networks, have completely changed the way that image processing and computer vision are researched. It is widely utilized in many different applications, including face identification, image segmentation, and object recognition. The biological framework of the human brain's visual cortex serves as the foundation for CNN's architectural design. Among the layers that make up this system are the input layer, convolutional layer, pooling layer, and fully connected layer. The raw data from the input layer, which is a picture, is subjected to a number of filters in the convolutional layer. A variety of elements in the image, including edges, corners, and forms, are detected using the filters. The pooling layer is used to lower the output's dimensionality and improve its computational efficiency after the convolution layer’s output. The classification process is then carried out using the identified features and the completely linked layer. Convolutional filters are used in CNN's operating system to extract information from an image. The input image is compressed with the filters, which are micro matrices, to create a feature map. A specific feature's activation at a certain spot in the image is represented by the feature map. The best combination of filters to utilize to identify different features in an image are learned by the CNN during the training phase.

5. PROPOSED WORK

The purpose of 3D human pose estimation using machine learning is to accurately and automatically determine the 3D positions and orientations of joints and parts of the human body from 2D images and video frames. This technology is especially valuable in fields as diverse as computer vision, robotics, and human-computer interaction, as it allows machines to understand and interpret human body movements as well as gesture recognition, virtual reality, and biomechanics. Can be used for applications such as analysis. The steps involved in 3D human pose estimation usually include several key steps. First, the process begins with data collection, where a dataset of images or video frames containing people is collected. Next, a preprocessing step may be required to normalize and enrich the input data. Then, feature extraction techniques are used to identify important parts of the human body, such as joint locations. A machine learning model, often a deep neural network such as a convolutional neural network (CNN) or a recurrent neural network (RNN), is trained on this data to learn the relationship between a two-dimensional state and the corresponding three-dimensional state. Once the model is trained, it is applied to new images or video frames to predict the 3D pose of a human. Post-processing techniques can be used to adjust and smooth the estimated state.

5.1 SINGLE-PERSON HUMAN POSE ESTIMATION

Single-person 3D pose estimation is divided into 2-step and 1-step categories. The method is as shown in Figure 3.1a. The two-step method consists of obtaining 2D joint positions using a 2D key point detection model. Convert 2D key points to 3D key points using deep learning techniques. Such an approach in the first stage, suffers from the ambiguity of depth inherent in the second key stage. A problem that many works aim to solve. One-step method means three-dimensional regression obtains detailed positions directly from static images. These methods require a lot of training data 3D annotation is available, but manual annotation is expensive and tedious. A standard method for single and multi-person 3D pose estimation. From the input 2D image and the predicted 3D human pose in (a) are obtained from the sample and GT.3.6M human dataset.



(a) Single-person 3D pose estimation.

FIGURE 3.1 Single-person 3D estimation

5.2 DIRECT REGRESSION

This method directly maps the joints and joints of the body. Features of the human body model. If the model is familiar, If you predict the 17 important points of a certain person, the result will be a 2.17 vector containing X and Y coordinates. All the expected signs are shown in Figure 3.2 below.

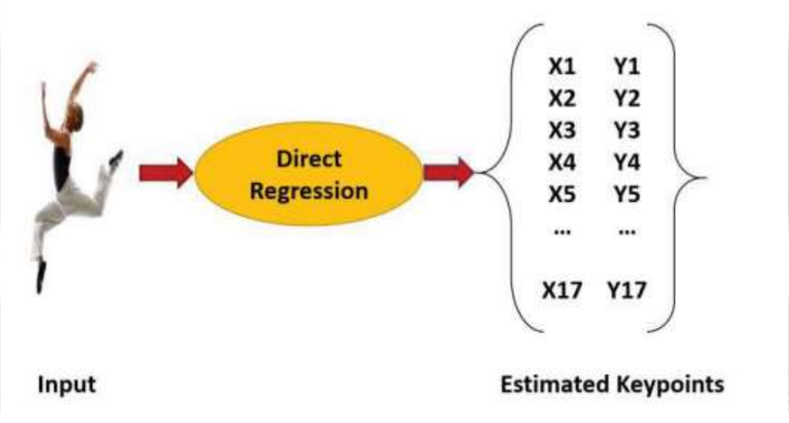


FIGURE 3.2 Direct Regression

5.3 HEATMAP REGRESSION

Heat map regression is widely used for 2D humans position estimation and grouping of key locations: for Example - hands, face, body. With heat map in frameworks, pixel values are usually used as follows: Probability that the corresponding pixels are mapped milestones in this framework. By adopting this technique, It is easy to practice and can achieve pixel-level accuracy.

6. ADVANTAGES

This technique leverages the capabilities of deep learning models to accurately predict the three-dimensional positions of human body joints and keypoints.3D human pose estimation using machine learning offers the promise of highly accurate and adaptable results in diverse real-world scenarios. While benefiting from end-to-end learning, temporal insights, and scalability, this approach demands substantial data, computational resources, and consideration of ethical and anatomical challenges. As technology advances, the potential for applications in fields like sports, healthcare, and animation remains compelling, underscoring the significance of a thoughtful and context-aware approach to implementation.

7. CONCLUSION

The success of our 3D pose estimation model can be attributed to several important factors. The variety and size of the dataset played an important role in training a robust model. Larger and more diverse data sets allow the model to better generalize to different situations and conditions. The combination of CNN and recurrent layers allows the model to capture both spatial and temporal information, improving pose estimation accuracy. The computational efficiency of the model is a major advantage. It opens up opportunities for real-time applications in areas such as gaming and augmented reality. Continuous optimization of model inference speed is beneficial.

6. REFERENCES

[1]. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

[2]. https://journals.ametsoc.org/view/journals/bams/87/9/bams-87-9-1195.xml

[3].https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2004GL022045

[4].https://link.springer.com/chapter/10.1007/978-3-031-06767-9_23