

Using a 3d graphics engine, create a robust real-time action or pose recognition training model.

¹ Mr. Pradeep Bedi, Graphic Era Hill University

Received: 29 December 2021 / Accepted: 29 January 2022

© The Author(s) 2022

Abstract:

Human pose recognition is a crucial one. The 3D images are divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model free and model based. In model free the estimation is done in two stages namely: single stage and 2D to 3D. In this single stage using direct prediction as well as body structure constraints are used. The assumptions which are made are the joints are named as parts. When the parts are connecting it is known as pair or limb. Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suits small amount of tasks. To accommodate more tasks mask R-CNN is used. The parameters used for the position are x, y and z. The orientation parameters are q₀, q_x, q_y and q_z. The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used. This contains 256 nodes including information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude. The rendering of the objects are carried out using the blender rendering software. The 3D model is presented to a blender and the synthetic dataset was matched. Then the prediction was made in three categories namely bounding box, location and rotation prediction. The other method is top-down, bottom-up and real time. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the parts are identified then the pose. But in bottom up approach it is vice-versa. The input is first fed into the transformation, projection and rotation. Then the transformed image is fed into the Key Point Estimation (KPE). This produces a confidence map using the gestures and then it is fed into the Transformation Parameter Estimation (TPE). The EP-PP transformation takes place. For the wide angle images are trwide-angleained well before and the EP-PP transformation takes place. The learning and prediction was carried out using the radio frequency tracker. If the movement is fast then it is difficult to track it. In such scenario, it is using convolutional neural network based compensation. Then it is fed into the 3D graphics software for the projection.

Keywords— Key Point Estimation, R-CNN, ResNet, DRN

1. Introduction

This invention relates to designing the pose recognition training model using 3d graphics engine in real time. Human pose recognition is a crucial one. The 3D images are divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model free and model based. In model free the estimation is done in two stages namely: single stage and 2D to 3D. In this single stage using direct prediction as well as body structure constraints are used. The assumptions which are made are the joints are named as parts. When the parts are connecting it is known as pair or limb. Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suits small amount of tasks. To accommodate more tasks mask R-CNN is used. The parameters used for the position are x, y and z. The orientation parameters are q₀, q_x, q_y and q_z. The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor.

Then the multilayer perceptron is used. This contains 256 nodes including the information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude. The rendering of the objects are carried out using the blender rendering software. The 3D model is presented to a blender and the synthetic dataset was matched. Then the prediction was made in three categories namely bounding box, location and rotation prediction. The other method is top-down, bottom-up and real time. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the parts are identified then the pose. But in bottom up approach it is vice-versa. The input is first fed into the transformation, projection and rotation. Then the transformed image is fed into the Key Point Estimation (KPE). This produces a confidence map using the gestures and then it is fed into the Transformation Parameter Estimation (TPE). The EP-PP transformation takes place. For the wide

angle images are trained well before and the EP-PP transformation takes place. The learning and prediction was carried out using the radio frequency tracker. If the movement is fast then it is difficult to track it. In such scenario, it is using convolutional neural network based compensation. Then it is fed into the 3D graphics software for the projection.

2. Materials and Methods

Jimmy Wu et al., suggest the object pose estimation using the pose interpreter network. Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suits small amount of tasks. This limitation was remedy by mask R-CNN that can handle multiple instances. The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used. This contains 256 nodes including the information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude. The rendering of the objects are carried out using the blender rendering software.

Yucheng Chen et al., suggest the 3D human pose estimation. It is divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model free and model based. In model free the estimation is done in two stages namely: single stage and 2D to 3D. In this single stage using direct prediction as well as body structure constraints are used. In model based three prediction techniques are used namely SMPL based, Kinematic model based, and probabilistic model. The other method is top-down, bottom-up and real time.

Pau Bramon Mora suggests the 3D pose regression of real time object with the synthetic data. The 3D model is presented to a blender and the synthetic dataset was matched. Then it was fed to Deep Neural Network (DNN). Then the prediction was made in three categories namely bounding box, location and rotation prediction.

Shi Chen and Kazuyuki Demachi suggest human pose estimation using transformation parameters. In this method, the input is first fed into the transformation, projection and rotation. Then the transformed image is fed into the Key Point Estimation (KPE). This produces a confidence map using the gestures and then it is fed into the Transformation Parameter Estimation (TPE).

Stefan Hinterstoisser et al., suggest pose estimation of texture less 3D objects. Here LINEMOD is used for object detection. Pose estimation was done by Kinect. This approach let to do the feature redundancy. To improve the corrected based on LINEMOD.

Mingliang Fu et al., suggest pose tracking in real time for the texture less object. The learning and prediction was carried out using the radio frequency tracker. If the movement is fast the it is difficult to track it. For which it is using convolutional neural network based compensation.

Sudharshan Chandra Babu suggests human pose estimation. The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used.

Bharath Raj suggests the human pose estimation using deep learning methods. The assumptions which are made are the joints are named as parts. When the parts are connecting it is known as pair or limb. Then it is connected. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the parts are identified then the pose. But in bottom up approach it is vice-versa.

Grégory Rogez et al., suggest pose estimation. It is made possible using a convolutional Neural Network. In this single stage using direct prediction as well as body structure constraints are used.

Albert Cleetus suggests human pose estimation in animals and games. In this method the 2D images are converted into 3D using the annotations.

Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suits small amount of tasks. This limitation was remedy by mask R-CNN that can handle multiple instances. The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used. This contains 256 nodes including the information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude. The rendering of the objects are carried out using the blender rendering software. Then it is fed into the 3D graphics software for the projection.

The 3D images are divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model free and model based. In model free the estimation is done in two stages namely: single stage and 2D to 3D. In this single stage using direct prediction as well as body structure constraints are used. In model based three prediction techniques are used namely SMPL based, Kinematic model based, and probabilistic model. The other method is top-down, bottom-up and real time. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the

parts are identified then the pose. But in bottom-up approach it is vice-versa.

The 3D model is presented to a blender and the synthetic dataset was matched. Then it was fed to Deep Neural Network (DNN). Then the prediction was made in three categories namely bounding box, location and rotation prediction.

Human pose estimation using various transformation parameters. In this method, the input is first fed into the transformation, projection and rotation. Then the transformed image is fed into the Key Point Estimation (KPE). This produces a confidence map using the gestures and then it is fed into the Transformation Parameter Estimation (TPE).

The assumptions which are made are the joints are named as parts. When the parts are connecting it is known as pair or limb. Then it is connected. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the parts are identified then the pose. But in bottom up approach it is vice-versa.

The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used.

Detailed Description of the Invention:

Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suits small amount of tasks. This limitation was remedy by mask R-CNN that can handle multiple instances. The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used. This contains 256 nodes including the information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude. The rendering of the objects are carried out using the blender rendering software. Then it is fed into the 3D graphics software for the projection.

The 3D images are divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model free and model based. In model free the estimation is done in two stages namely: single stage and 2D to 3D. In this single stage using direct prediction as well as body structure constraints are used. In model based three prediction techniques are used namely SMPL based, Kinematic model based, and probabilistic model. The other method is top-down, bottom-up and real time. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the

parts are identified then the pose. But in bottom-up approach it is vice-versa.

The 3D model is presented to a blender and the synthetic dataset was matched. Then it was fed to Deep Neural Network (DNN). Then the prediction was made in three categories namely bounding box, location and rotation prediction.

Human pose estimation has various transformation parameters as in figure 1. In this method, the input is first fed into the transformation, projection and rotation. Then the transformed image is fed into the Key Point Estimation (KPE). This produces a confidence map using the gestures and then it is fed into the Transformation Parameter Estimation (TPE). The EP-PP transformation takes place as in figure 2. For the wide angle images as in figure 3 the images are trained well before and the EP-PP transformation takes place.

The assumptions which are made are the joints are named as parts. When the parts are connecting it is known as pair or limb. Then it is connected. For the multi person pose estimation, Top down as well as bottom up approach is made. In top down approach the parts are identified then the pose. But in bottom up approach it is vice-versa.

The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used.

The learning and prediction was carried out using the radio frequency tracker. If the movement is fast then it is difficult to track it. In such scenario, it is using convolutional neural network based compensation.

Drawings

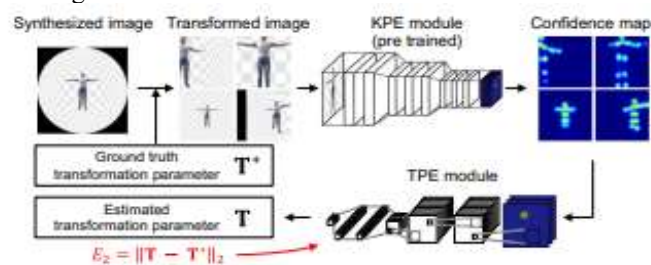


Figure 1: Overall Architecture

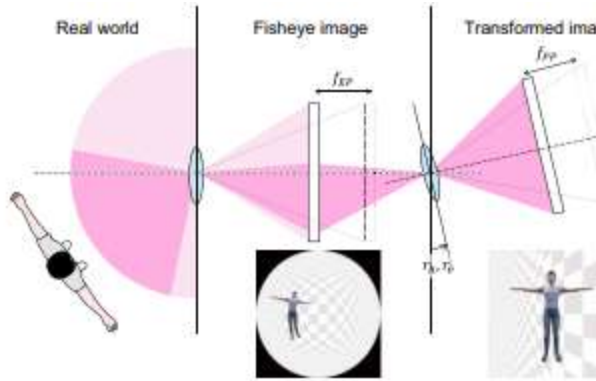


Figure 2: EP-PP transformation

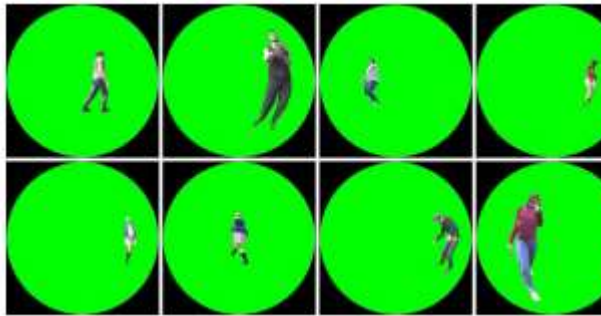


Figure 3: Wide angle images

3. Conclusion

The 3D images are divided into two categories namely 3D single and 3D multiple. The 3D single is further classified into 3D model-free and model-based. In model-free the estimation is done in two stages namely: single-stage and 2D to 3D. In this single-stage using direct prediction as well as body, structure constraints are used. Using Dilated Residual Network (DRN) the semantic segmentation was done. This will suit small amount of tasks. To accommodate more tasks mask R-CNN is used. The parameters used for the position are x , y and z . The orientation parameters are q_0 , q_x , q_y and q_z . The rotation was given by two parameters w and Θ . The features are extracted using ResNet feature extractor. Then the multilayer perceptron is used. This contains 256 nodes including the information about position and orientation. The 20 output layer from the orientation and the 15 output layer from the position are normalized to form a unit magnitude.

References

- [1] [n. d.]. MSCOCO keypoint detection evaluation metric. <https://cocodataset.org/#keypoints-eval> (2016).
- [2] M. Andriluka, U. Iqbal, E. Ensafutdinov, L. Pishchulin, A. Milan, J. Gall, and Schiele B. 2018. PoseTrack: A Benchmark for Human Pose Estimation and Tracking. In CVPR.

- [3] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2014. 2d human pose estimation: New benchmark and state of the art analysis. In CVPR.
- [4] Federico Angelini, Zeyu Fu, Yang Long, Ling Shao, and Syed Mohsen Naqvi. 2018. Actionxpose: A novel 2d multi-view pose-based algorithm for real-time human action recognition. arXiv preprint arXiv:1810.12126 (2018).
- [5] Anurag Arnab, Carl Doersch, and Andrew Zisserman. 2019. Exploiting Temporal Context for 3D Human Pose Estimation in the Wild. In CVPR.
- [6] Bruno Artacho and Andreas Savakis. 2020. UniPose: Unified Human Pose Estimation in Single Images and Videos. In CVPR.
- [7] Vasileios Belagiannis and Andrew Zisserman. 2017. Recurrent human pose estimation. In FG.
- [8] Abdallah Benzine, Florian Chabot, Bertrand Luvion, Quoc Cuong Pham, and Catherine Achard. 2020. PandaNet: Anchor-Based Single-Shot Multi-Person 3D Pose Estimation. In CVPR.
- [9] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. 2016. Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image. In ECCV.
- [10] Adrian Bulat and Georgios Tzimiropoulos. 2016. Human pose estimation via convolutional part heatmap regression. In ECCV.