Assignment 10

Name: Gaurav Ganesh Bhogale

Referral ID: SIRSS2101

POSE ESTIMATION USING COMPUTER VISION

Q1. What is human pose estimation?

Human pose estimation is the process of estimating the configuration of the body (pose) from a single, typically monocular, image. The main component of human pose estimation is the modelling of the human body. There are three of the most used types of human body models: skeleton-based model, contour-based, and volume-based.

Q2. What is 2D human pose estimation?

2D human pose estimation is used to estimate the 2D position or spatial location of human body keypoints from visuals such as images and videos. Traditional 2D human pose estimation methods use different hand-crafted feature extraction techniques for the individual body parts.

Early computer vision works described the human body as a stick figure to obtain global pose structures. However, modern deep learning based approaches have achieved major breakthroughs by improving the performance significantly for both single-person and multi-person pose estimation. Some popular 2D human pose estimation methods include OpenPose, CPN, AlphaPose, and HRNet.

Q3. What is 3D human pose estimation?

3D Human Pose Estimation is used to predict the locations of body joints in 3D space. Besides the 3D pose, some methods also recover 3D human mesh from images or videos. This field has attracted much interest in recent years since it is used to provide extensive 3D structure information related to the human body. It can be applied to various applications, such as 3D animation industries, virtual or augmented reality, and 3D action prediction. 3D human pose estimation can be performed on monocular images or videos (normal camera feeds).

Using multiple viewpoints or additional sensors (IMU or LiDAR), 3D pose estimation can be applied with information fusion techniques, which is a very challenging task. While 2D human datasets can be easily obtained, collecting accurate 3D pose image annotation is time-consuming, and manual labelling is not practical and expensive. Therefore, although 3D pose estimation has made significant advancements in recent years, especially due to the progress made in 2D human pose estimation, there are still several challenges to overcome: Model generalization, robustness to occlusion, and computation efficiency.

A popular library that uses neural networks for real-time 3D human pose estimation, even for multi-person use cases, is OpenPose.

Q4. Explain Human Body Modelling.

In human pose estimation, the location of human body parts is used to build a human body representation (such as a body skeleton pose) from visual input data. Therefore, human body modelling is an important aspect of human pose estimation. It is used to represent features and key points extracted from visual input data. Typically, a model-based approach is used to describe and infer human body poses and render 2D or 3D poses.

Most methods use an N-joints rigid kinematic model where a human body is represented as an entity with joints and limbs, containing body kinematic structure and body shape information.

There are three types of models for human body modelling:

- **Kinematic Model**, also called skeleton-based model, is used for 2D pose estimation as well as 3D pose estimation. This flexible and intuitive human body model includes a set of joint positions and limb orientations to represent the human body structure. Therefore, skeleton pose estimation models are used to capture the relations between different body parts. However, kinematic models are limited in representing texture or shape information.

- **Planar Model**, or contour-based model, that is used for 2D pose estimation. The planar models are used to represent the appearance and shape of a human body. Usually, body parts are represented by multiple rectangles approximating the human body contours. A popular example is the Active Shape Model (ASM) that is used to capture the full human body graph and the silhouette deformations using principal component analysis.

- **Volumetric model**, which is used for 3D pose estimation. There exist multiple popular 3D human body models used for deep learning-based 3D human pose estimation for recovering 3D human mesh. For example, GHUM & GHUML(ite), are fully trainable end-to-end deep learning pipelines trained on a high-resolution dataset of full-body scans of over 60'000 human configurations to model statistical and articulated 3D human body shape and pose.

Q5. How does pose estimation works?

Pose estimation utilizes pose and orientation to predict and track the location of a person or object. Accordingly, pose estimation allows programs to estimate spatial positions ("poses") of a body in an image or video. In general, most pose estimators are 2 steps frameworks that detect human bounding boxes and then estimate the pose within each box.

Pose estimation operates by finding key points of a person or object. Taking a person, for example, the key points would be joints like the elbow, knees, wrists, etc. There are two types of pose estimation: multi-pose and single pose. Single pose estimation is used to estimate the poses of a single object in a given scene, while multi-pose estimation is used when detecting poses for multiple objects.

Q6. Explain Pose Estimation with deep learning.

deep learning has been shown to outperform classical computer vision methods in various tasks, including image segmentation or object detection. Therefore, deep learning techniques brought significant advances and performance gains in pose estimation tasks.

Next, we will list and review the popular pose estimation methods.

The Most popular Pose Estimation methods

- Method #1: OpenPose

- Method #2: High-Resolution Net (HRNet)

- Method #3: DeepCut

- Method #4: Regional Multi-Person Pose Estimation (AlphaPose)

- Method #5: Deep Pose

- Method #6: PoseNet

- Method #7: Dense Pose

Q7. Applications of Pose Estimation:

Training Robots

Typical use cases of pose estimators is in the application of making robots learn certain crafts. In place of manually programming robots to follow trajectories, robots can be made to learn actions and movements by following the tutor's posture look or appearance**.**

Motion Tracking for Consoles

Other applications of pose estimation are in-game applications, where human subjects auto-generate and inject poses into the game environment for an interactive gaming experience. For instance, Microsoft's Kinect used 3D pose estimation (using IR sensor data) to track the motion of the human players and to use it to render the actions of the characters virtually into the gaming environment.