ASSIGNMENT-10

K Manoj
Referral Id:**SIRSS2309**

POSE ESTIMATION USING COMPUTER VISION

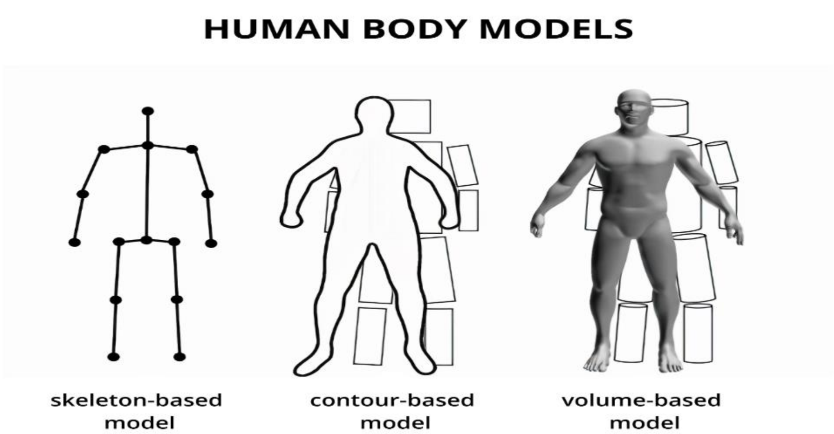
What is Human Pose Estimation?

Human pose estimation is a computer vision-based technology that detects and analyzes human posture. The main component of human pose estimation is the modeling of the human body. There are three of the most used types of human body models: skeleton-based model, contour-based, and volume-based.

Skeleton-based model consists of a set of joints (keypoints) like ankles, knees, shoulders, elbows, wrists, and limb orientations comprising the skeletal structure of a human body. This model is used both in 2D and 3D human pose estimation techniques because of its flexibility.

Contour-based model consists of the contour and rough width of the body torso and limbs, where body parts are presented with boundaries and rectangles of a person's silhouette.

Volume-based model consists of 3D human body shapes and poses represented by volume-based models with geometric meshes and shapes,



What is OpenCV?

Computer vision is a process by which we can understand how the images and videos are stored and manipulated, also it helps in the process of retrieving data from either images or videos. Computer Vision is part of Artificial Intelligence. Computer-Vision plays a major role in Autonomous cars, Object detections, robotics, object tracking, etc.

OpenCV is an open-source library mainly used for computer vision, image processing, and machine learning. It gives better output for real-time data, with the help of OpenCV, we can process images and videos so that the implemented algorithm can be able to identify objects such as cars, traffic signals, number plates, etc., and faces, or even handwriting of a human. With the help of other data analysis libraries, OpenCV is capable of processing the images and videos according to one's desire.

Multi-Person Pose Estimation

Multi-Person pose estimation is more difficult than the single person case as the location and the number of people in an image are unknown. Typically, we can tackle the above issue using one of two approaches:

- The simple approach is to incorporate a person detector first, followed by estimating the parts and then calculating the pose for each person. This method is known as the **top-down** approach.
- Another approach is to detect all parts in the image (i.e. parts of every person), followed by associating/grouping parts belonging to distinct persons. This method is known as the **bottom-up** approach.

Categories of pose estimation:

When working with people, these keypoints represent major joints like elbows, knees, wrists, etc. This is referred to as human pose estimation.

Humans fall into a particular category of objects that are flexible. By bending our arms or legs, keypoints will be in different positions relative to others. Most inanimate objects are rigid. For instance, the corners of a brick are always the same distance apart regardless of the brick’s orientation. Predicting the position of these objects is known as rigid pose estimation.

There’s also a key distinction to be made between 2D and 3D pose estimation. 2D pose estimation simply estimates the location of keypoints in 2D space relative to an image or video frame. The model estimates an X and Y coordinate for each keypoint. 3D pose estimation works to transform an object in a 2D image into a 3D object by adding a z-dimension to the prediction.

3D pose estimation allows us to predict the actual spatial positioning of a depicted person or object. As you might expect, 3D pose estimation is a more challenging problem for machine learners, given the complexity required in creating datasets and algorithms that take into account a variety of factors – such as an image’s or video’s background scene, lighting conditions, and more.

Finally, there is a distinction between detecting one or multiple objects in an image or video. These two approaches can be referred to as single and multi pose estimation, and are largely self-explanatory: Single pose estimation approaches detect and track one person or object, while multi pose estimation approaches detect and track multiple people or objects.

Human Body Model:

Every pose estimation algorithm agrees upon a body model beforehand. It allows the algorithm to formalize the problem of human pose estimation into that of estimating the body model parameters. Most algorithms use a simple N-joint rigid kinematic skeleton model (N is typically between 13 to 30) as the final output. Formally, kinematic models can be represented as a graph, where each vertex V represents a joint. The edges E can encode constraints or prior beliefs about the structure of the body model.

Such a model suffices for most applications. However, for many other applications such as character animation, a more elaborate model may be needed. Some techniques have considered a highly detailed mesh models, representing the whole body with a point cloud.

Another rather primitive body model that was used in earlier Pose Estimation pipelines is a shape-based body model. In shape-based models, human body parts are approximated using geometric shapes like rectangles, cylinders, conics etc.

Feature Extraction

Feature extraction in Machine Leaning refers to the creation of derived values from raw data (such as an image or video in our case), that can be used as input to a learning algorithm. Features can either be explicit or implicit. Explicit features include conventional Computer Vision based features like Histogram of Oriented Gradients (HoG) and Scale Invariant Feature Transform (SIFT). These features are calculated explicitly before feeding the input to the following learning algorithm. Implicit features refers to deep learning based feature maps like outputs from complex Deep Convolutional Neural Networks (CNNs). These feature maps are never created explicitly, but are a part of a complete pipeline trained end-to-end.

How does Pose Estimation work?

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.

Human pose estimation on the popular MS COCO Dataset can detect 17 different keypoints (classes). Each keypoint is annotated with three numbers (x,y,v), where x and y mark the coordinates, and v indicates if the keypoint is visible.

Pose Estimation with Deep Learning:

With the rapid development of deep learning solutions in recent years, 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.

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

Use Cases and Applications of Pose Estimation:

Human pose estimation has been utilized in a wide range of applications, including human-computer interaction, motion analysis, augmented reality, and robotics.

Most Popular Pose Estimation Applications:

- Application 1: Human Activity Estimation
- Application 2: Motion Transfer and Augmented Reality
- Application 3: Training Robots
- Application 4: Motion Tracking for Consoles
- Application 5: Human Fall Detection

Augmented Reality and Virtual Reality

As of today, pose estimation interfaced with augmented and virtual reality applications gives users a better online experience. For instance, users can virtually learn how to play games like tennis via virtual tutors who are pose represented.

More so, pose estimators can also be interfaced with augmented reality-based applications. For example, The United States Army experiments with augmented reality programs to be used in combat. These programs aim to help soldiers distinguish between enemies and friendly troops, as well as improve night vision.

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.

Outlook and Future Trends

New technologies and methods make it possible to decrease the size of AI models, making pose estimation algorithms less “heavy” and much more efficient. This is the basis for the real-world implementation of human pose detection.

As a result, it becomes possible to deploy pose estimation algorithms to edge devices and perform on-device machine learning (Edge AI). Edge Inference makes the technology scalable, more robust for mission-critical applications (offline capability), and private (no visuals need to be sent to the cloud). An example of a lightweight pose estimation model for Edge ML is Lightweight OpenPose.

Platforms such as Viso Suite make it possible to quickly use and scale pose estimation in real-world applications by deploying pose estimation models to edge devices that are connected to cameras.

What’s Next?

Pose estimation is a fascinating aspect of computer vision that can be applied in multiple fields, including technology, healthcare, business, and others. Aside from its prominence in modeling human characters via Deep Neural Networks that learn various key points, it is also used for security and surveillance systems.

Other popular applications of compute vision include image classification, image segmentation, face detection, or object detection.