# **TASK 10**

# **REFERRAL ID:SIRSS2159**

# **FULL NAME : SHEREEN ANTHONY**

**ESTIMATION USING COMPUTER VISION**

**Pose estimation is a computer vision technique that is used to predict the configuration of the body(POSE) from an image. The reason for its importance is the abundance of applications that can benefit from technology.**

What is open cv?

Opencv is an open source library which is very useful for computer vision applications such as video analysis, CCTV footage analysis and image analysis. OpenCV is written by C++ and has more than 2,500 optimized algorithms. When we create applications for computer vision that we don’t want to build from scratch we can use this library to start focusing on real world problems. There are many companies using this library today such as Google, Amazon, Microsoft and Toyota. Many researchers and developers contribute. We can easily install it in any OS like Windows, Ubuntu and MacOS.

Uses of open cv

OpenCV is the huge open-source library for **the computer vision, machine learning, and image processing** and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human.

**What is pose estimation?**

Human pose estimation and tracking is a computer vision task that includes detecting, associating, and tracking semantic key points. Examples of semantic keypoints are “right shoulders,” “left knees,” or the “left brake lights of vehicles.”

The performance of semantic keypoint tracking in live video footage requires high computational resources what has been limiting the accuracy of pose estimation. With the latest advances, new applications with real-time requirements become possible, such as self-driving cars and last-mile delivery robots.

Today, the most powerful image processing models are based on convolutional neural networks (CNNs). Hence, state-of-the-art methods are typically based on designing the CNN architecture tailored particularly for human pose inference.

##### bottom-up vs. Top-down methods

All approaches for pose estimation can be grouped into bottom-up and top-down methods.

* **Bottom-up methods** estimate each body joint first and then group them to form a unique pose. Bottom-up methods were pioneered with DeepCut (a method we will cover later in more detail).
* **Top-down methods** run a person detector first and estimate body joints within the detected bounding boxes.

##### The importance of Pose Estimation

In traditional [object detection](https://viso.ai/deep-learning/object-detection/), people are only perceived as a bounding box (a square). By performing pose detection and pose tracking, computers can develop an understanding of human body language. However, conventional pose tracking methods are neither fast enough nor robust enough to occlusions to be viable.

High-performing real-time pose detection and tracking will drive some of the biggest trends in computer vision. For example, tracking the human pose in real-time will enable computers to develop a finer-grained and more natural understanding of human behavior.

This will have a big impact on various fields, for example, in autonomous driving. Today, the majority of self-driving car accidents are [caused by “robotic” driving](https://rmi.org/safe-self-driving-cars/), where the self-driving vehicle conducts an allowed but unexpected stop, and a human driver crashes into the self-driving car. With real-time human pose detection and tracking, the computers are able to understand and predict pedestrian behavior much better – allowing more natural driving.

### What is Human Pose Estimation?

Human pose estimation aims at predicting the poses of human body parts and joints in images or videos. Since pose motions are often driven by some specific human actions, knowing the body pose of a human is critical for action recognition.

##### 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](https://viso.ai/deep-learning/openpose/), CPN, AlphaPose, and HRNet (we will cover them and others later in this article).

##### 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](https://viso.ai/computer-vision/image-annotation/) is time-consuming, and manual labeling 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](https://viso.ai/deep-learning/openpose/).

##### Human body modeling

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 modeling is an important aspect of human pose estimation. It is used to represent features and keypoints 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 modeling:

* **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](https://www.researchgate.net/publication/349703947_Deep_Active_Shape_Model_for_Face_Alignment_and_Pose_Estimation) (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](https://github.com/google-research/google-research/tree/master/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. It can be used to infer

##### Main challenges

Human pose estimation is a challenging task as the body’s appearance joins changes dynamically due to diverse forms of clothes, arbitrary occlusion, occlusions due to the viewing angle, and background contexts. Pose estimation needs to be robust to challenging real-world variations such as are lighting and weather.

Therefore, it is challenging for image processing models to identify the fine-grained joint coordinates. It is especially difficult to track small and barely visible joints.

### Head pose estimation

Estimating the head pose of a person is a popular computer vision problem. Head pose estimation has multiple applications such as aiding in gaze estimation, modeling attention, fitting 3D models to video and performing face alignment.

Traditionally head pose is computed with the use of keypoints from the target face and by solving the 2D to 3D correspondence problem with a mean human head model.

The ability to recover the 3D pose of the head is a by-product of keypoint-based facial expression analysis that is based on the extraction of 2D facial keypoints with deep learning methods. Those methods are robust to occlusions and extreme pose changes.

### 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**](https://viso.ai/computer-vision/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.

**Challenges**

There is a huge variety of challenging difficulties concerning pose estimation.

First of which is a clear image, more often than not the images available are not clear, making it difficult to identify the subjects.

Secondly, all the parts of the body that we use for the key points should be visible. In cases of low-resolution images, this task becomes difficult. Also, in the case of multiple people present in the image, this becomes an even more difficult task.

In the case of Multi pose estimation, there is the problem of overlapping. Usually, we can see multiple people, and the key points may overlap.

With the advent of smartphones and their cameras, there is another problem with the images being of varied sizes. The rescaling or resizing is another challenge where we ought to maintain the details while making sure that the model can precisely extract the relevant features.

**Use cases and applications**

   Human activity and movement:

Tracking the variations in the pose of a person over a period can also be used for activity, gesture, and gait recognition. Some applications could be as below:

-         Application to identify if a person follows the exercise regime properly

-         Application to identify the health of a person – In case the person collapses inform somebody close or paramedics

-         Application to identify body language and flag suspicious individuals to nearby authorities - used at airports

B.     Augmented reality

-         CGI Application to track the human pose variations to render graphical animations – e.g: Thanos

C.     Animation & gaming

-         Identify and track movements in Gaming

D.     Robotics

-         Robots can be taught to mimic human poses, activities by tracking and following human instructor demonstrations, instead of manually programming robots