Computer vision is a field of study that enables computers to interpret and understand the visual world. It seeks to automate tasks that the human visual system can do.

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

Feature extraction is a type of dimensionality reduction. Feature extraction can be performed on texts as part of NLP or on images for computer vision tasks.

2D vision systems, which only capture data in two dimensions, 3D vision systems can also collect height/depth data from the surface of the object to accurately digitalize the object's shape and present it in a digital, three-dimensional format.

2D vision, the target object image acquired for processing is a flat, two-dimensional view. The 2D image does not provide any height information at all – there's X and Y data, but no Z-axis depth of field data.

The target object image in a 3D vision is no longer just a flat image. It's a three-dimensional point cloud with accurate coordinates that shows the exact location of every pixel in space.

In computer vision, CNN stands for Convolutional Neural Network. It is a class of deep, feed-forward artificial neural networks most commonly applied to analyzing visual imagery.

It enables a computer to understand and interpret the image or visual data.

CNN utilizes spatial correlations which exist with the input data. Each concurrent layer of the neural network connects some input neurons.

Computer vision helps the autonomous vehicle to plan its route and make informed decisions in real time, by continuously analysing the surroundings.

1. Obstacle Detection and Avoiding collision : Computer vision algorithms can identify and track various obstacles, stationary & moving, such as vehicles, pedestrians, road debris and plan appropriate path to avoid collisions.
2. Recognize Traffic Sign: Computer vision algorithms can identify and interpret traffic signs. Autonomous vehicles utilize this information to adjust their speed and respond accordingly. As the vehicle approaches a stop sign, the computer vision system recognizes the sign's shape and colour.
3. Lane Detection: autonomous vehicles can detect & track lane markings on the road due to computer vision. This allows the vehicle to maintain its position within the lane ensuring it stays within its designated path, even on curved roads.

Challenges, such as Viewpoint Variation, variable lighting conditions, complex backgrounds, noisy/cluttered input, data quality pose a problem for real-world deployment object detection using computer vision..

Thus while AI can detect an object, it might happen it might not fully understand the full context under which that object is being viewed. Thus, leading to inaccurate interpretation.

Non-maximum suppression (NMS) is a post-processing technique of object detection which eliminate duplicate detections and select the most relevant bounding boxes that correspond to the detected objects. It is used to remove incorrect detection results.

By applying non-maximum suppression, object detection algorithms can produce more accurate and reliable results by reducing duplicate detections and improving localization of objects within an image.

Image segmentation is a crucial task in computer vision, where the goal is to divide an image into different meaningful and distinguishable regions or objects. It divide an image into different parts or segments to identify the objects and boundaries within it. It is a fundamental task in various applications such as object recognition, tracking, and detection, medical imaging, and robotics.

Optical flow is a technique used to describe image motion. It is usually applied to a series of images that have a small time step between them, for example, video frames. The uses of optical flow is mainly in the field of Object Tracking. The optical flow can be used as an estimation of object velocity and position of object in the next frame.

R-CNN, or Region-based Convolutional Neural Network, is a method for object detection that involves using a CNN to classify object proposals, or regions of interest (ROIs), within an image. R-CNN is that it can handle many object classes, but it is computationally expensive, requiring the CNN to be run on each ROI individually.

Fast R-CNN is an improvement over R-CNN that addresses the computational inefficiency of the original method. It achieves this by using a single CNN to process the entire image and generate the ROIs rather than running the CNN on each ROI individually. Fast R-CNN significantly reduces the computational cost but is still expensive as it requires a separate classifier for each object class.

Mask R-CNN is an extension of Faster R-CNN and works by adding a branch for predicting an object mask (Region of Interest) in parallel with the existing branch for bounding box recognition.  
It is very efficient and adds only a small overhead to Faster R-CNN.