Computer Vision Interview Questions

1. What is Computer Vision?

Computer Vision uses images and videos to understand a real-world scene. Just like Humans use eyes for capturing light, receptors in the brain for accessing it, and the visual cortex for processing it. Similarly, a computer understands images, videos, or a real-world scenario through machine learning algorithms and AI self-learning programming.

2. What are machine learning algorithms available in OpenCV?

An OpenCV is open for all and free cross-platform where you get a library of real-time computer vision programming functions. It is developed by Intel and is mostly written in the C++ programming language. A JavaScript version is also available as OpenCV.js which is built for web platforms.

Some machine learning libraries available in OpenCV are:

Artificial Neural Networks, Random Forest, Support Vector Machine, Decision Tree Learning, Convolution Neural Networks, Boosting and Gradient Boosting Trees, Expectation-Maximization algorithm, Naive Bayes classifier, K-nearest neighboring algorithm.

3. How many types of image filters in OpenCV?

Image filters used in OpenCV are:

- Bilateral Filter
- Blur Filter
- Box Filter
- Dilate
- Build Pyramid
- Erode
- Filter2D
- Gaussian Blur
- Deriv
- Gabor Kernels
- Laplacian
- Median Blur.

4. How many types of image filters in OpenCV?

The face recognition algorithm is basically the computer application that is used for tracking, detecting, identifying, or verifying the human faces simply from the image or the video that has been captured using the digital camera.

Some popular but evolving algorithms are:

- PCA- Principal Component Analysis
- LBPH- Local Binary Pattern Histograms
- k-NN (nearest neighbors) algorithm
- Eigen's faces
- Fisher faces
- SIFT- Scale Invariant Feature Transform
- SURF- Speed Up Robust Features

5. What Do You Mean by Color Model?

A Color Model is a coordinate system and a subset of visible colors. With "Color Model" we create a whole range of colors from a limited set of primary colors like RGB (Red Green Blue). Color Models are of two types: Additive and Subtractive.

6. What is Dynamic Range?

Dynamic Range is a ratio of small and large values that is assumed by a certain quantity. It is used in signals, photography, sounds, and light. From a photographic point of view, it is a ratio of minimum and maximum measuring light intensity or the lightest and darkest regions also called color contrast,

7. Define Digital Image?

A digital image is an image that is comprised of the elements of the picture, they are also admitted as pixels. Each pixel is with the finite and the discrete numbers of the numerical representation which belong to its intensity and the gray level which is considered as its output from the functions of the two-dimensions that is feed by the input of spatial coordinates those are denoted by the x-axis and y-axis.

8. What Is Meant by Mach Band Effect?

- Mach Band Effect is an optical illusion. It emphasizes the differentiation between edges of the somewhat varying shades of grey when they reach each other.
- The extreme left side is dark grey and it converts into the lighter shades as they move to the right side of the plate.

9. What is Sampling and Quantization? Explain with an example why the inputs in computer vision problems can get huge. Provide a solution to overcome this challenge.

We use sampling and Quantization to convert analog images to digital images. An image has two things.

1. Coordinates

Digitizing of coordinates is called Sampling. That is, converting the coordinates of the analog images to the digital images.

2. Intensity/Amplitude

Digitizing of Amplitude or Intensity is called Quantization. That is converting the Amplitude or Intensity of an analog image to a digital image.

10. Explain with an example why the inputs in computer vision problems can get huge. Provide a solution to overcome this challenge.

Consider a 500x500 pixel RGB image fed to a fully connected neural network for which the first hidden layer has just 1000 hidden units. For this image, the number of input features will be 500*500*3=750,000, i.e. the input vector will be 750,000 dimensional. The weight matrix at the first hidden layer will therefore be a 1000x750,000-dimensional matrix which is huge in size for both computations as well as storage. We can use convolution operation, which is the basis of convolutional neural networks, in order to address this challenge.

11. Explain with an example why the inputs in computer vision problems can get huge. Provide a solution to overcome this challenge.

- The earlier layers of the neural network detect simple features of an image, such as edges or corners.
- As we go deeper into the neural network, the features become increasingly complex, detecting shapes and patterns.
- The later layers of the neural network are capable of detecting complex patterns such as complete objects.

12. How do you address the issue of the edge pixels being used less than the central pixels during convolutional operation?

In order to address the issue of the filter or kernel extracting information from the edge pixels less in comparison to the central pixel, we can use padding. Padding is essentially adding one or more additional rows or columns of pixels along the boundary of the image. The padding forms the new edge pixels of the image and therefore results in insufficient extraction of information from the original edge pixels. Padding provides the added advantage of preventing the shrinking of an image as a result of the convolution operations.

13. How many parameters are to be learned in the pooling layers?

No parameters are to be learned in the pooling layers. In general, the pooling layer has a set of hyperparameters describing the filter size and the stride length, which are set and work as a fixed computation.

14. Explain why mirroring, random cropping, and shearing are some techniques that can help in a computer learning problem.

There is often limited data available to solve computer vision problems, which just isn't enough to train the neural networks. Techniques like mirroring, random cropping, and shearing can help augment the existing dataset and create more training data from the existing data, thereby ameliorating the issue of limited training data.

15. Mention a method that can be used to evaluate an object localization model. How does it work?

Intersection over Union (also known as IoU) is a commonly used method to evaluate the performance of the object localization model. The overlap between the ground truth bounding box and the predicted bounding box is checked, and the ratio of the intersection of the areas to the Union of the areas is calculated. If this ratio called IoU is found greater than some threshold (usually set to 0.5 or higher), the prediction of the model is considered correct.