Module 1

1. **What is computer vision? Explain with an example:**

Computer vision is a field of artificial intelligence and computer science that focuses on enabling computers to interpret, understand, and process visual information from the world, much like humans do with their eyes. It involves developing algorithms and systems that can extract meaningful information from images or video data. Computer vision has various applications, including image and video analysis, object recognition, and scene understanding.

Example: One common application of computer vision is facial recognition. Facial recognition software can identify and verify individuals by analyzing features such as the arrangement of eyes, nose, and mouth in a given image or video stream. It's used in security systems, unlocking smartphones, and even tagging friends in photos on social media.

2. **How is light considered as an electromagnetic wave?**

Light is considered an electromagnetic wave because it exhibits both electric and magnetic properties. According to electromagnetic theory, light is a result of oscillating electric and magnetic fields propagating through space. James Clerk Maxwell's equations describe the behavior of electromagnetic waves, and they show that changing electric fields create magnetic fields and vice versa. Light consists of oscillating electric and magnetic fields that are perpendicular to each other and propagate through space as waves, carrying energy and information.

3. **Mention any 10 applications of computer vision:**

Computer vision has numerous applications, and here are ten examples:

- Facial recognition
- Autonomous vehicles (self-driving cars)
- Medical image analysis (e.g., detecting tumors in medical scans)
- Object detection and tracking in surveillance systems
- Augmented reality (overlaying digital information on the real world)
- Handwriting recognition
- Quality control in manufacturing
- Gesture recognition (e.g., for gaming or human-computer interaction)
- Content-based image retrieval (searching for images based on their content)
- Agricultural robotics (e.g., identifying and managing crops and pests)

4. **Write a short note on the history of computer vision:**

The history of computer vision can be summarized as follows:

- **1950s-1960s:** Early work in computer vision involved simple tasks like edge detection and pattern recognition. The field was primarily experimental, and computers had limited processing power.
- **1970s:** This decade saw the development of foundational techniques like the Hough transform for shape analysis, and the development of early image processing algorithms for feature extraction.
- **1980s:** The 1980s brought advances in 3D vision, object recognition, and stereo vision. Prominent algorithms like the RANSAC algorithm for robust fitting were introduced.
- **1990s:** Researchers focused on understanding complex scenes, motion analysis, and object tracking. The era also saw the use of neural networks and the development of the Viola-Jones object detection framework.
- **2000s:** The 2000s brought significant progress in image classification and recognition, driven in part by large datasets and deep learning methods. The ImageNet dataset and convolutional neural networks (CNNs) became key milestones.

5. Explain with examples of computer vision algorithms in the 1970s:

In the 1970s, computer vision was in its infancy, with a focus on fundamental techniques for image analysis. Two notable developments are:

- **The Hough Transform:** The Hough Transform, initially developed in the late 1960s but widely utilized in the 1970s, was a breakthrough in pattern recognition. It allowed for the detection of shapes and patterns within images. One of its primary applications was line detection. This algorithm could identify lines in images, even if they were broken or incomplete. It played a critical role in early computer vision systems used for tasks like extracting road lanes from aerial imagery.
- **SIFT (Scale-Invariant Feature Transform):** While SIFT was primarily developed in the 1990s and 2000s, it builds upon earlier work from the 1970s. SIFT is a feature extraction technique that can detect and describe distinctive local features within images. These features are invariant to changes in scale, rotation, and illumination, making them invaluable for object recognition and image matching. SIFT laid the foundation for key computer vision applications in subsequent decades.

6. Explain with examples of computer vision algorithms in the 1980s:

The 1980s marked a significant shift towards 3D vision and advanced object recognition techniques. Here are some examples:

- **RANSAC (Random Sample Consensus):** RANSAC, though introduced in the late 1980s, became a fundamental algorithm in computer vision. It is used for robust model fitting, especially in the presence of outliers. RANSAC iteratively estimates model parameters from a subset of data points and is widely used in tasks like fitting lines or planes to 3D point clouds. This robust estimation is essential in applications such as 3D reconstruction and structure from motion.

7. Explain with examples of computer vision algorithms in the 1990s:

In the 1990s, computer vision expanded its focus towards complex scene understanding and motion analysis. Here are some notable examples:

- **Optical Flow Algorithms:** Optical flow techniques emerged in the 1990s as a way to estimate the motion of objects in a scene. These algorithms analyze the movement of pixels between consecutive frames in video sequences, allowing for the tracking of object motion. Optical flow has applications in video analysis, such as tracking moving objects and understanding the dynamics of scenes, making it valuable in fields like surveillance, robotics, and sports analytics.
- **Viola-Jones Object Detection Framework:** The Viola-Jones framework, introduced in the late 1990s, was a significant advancement in object detection. This framework employs Haar-like features and a cascade of classifiers to perform real-time face detection, laying the foundation for broader object detection applications. It became a cornerstone in applications like facial recognition, image-based biometrics, and video surveillance.

8. Explain with examples of computer vision algorithms in the 2000s:

The 2000s were a transformative period for computer vision, largely due to the advent of deep learning and the availability of large datasets. Here are some notable examples:

- **Convolutional Neural Networks (CNNs):** CNNs gained prominence in image processing and computer vision tasks during the 2000s. They are deep learning models designed for feature extraction and image classification. The ImageNet Large Scale Visual Recognition Challenge in 2012 demonstrated the power of CNNs in image classification, leading to their widespread adoption. Applications expanded to object recognition, scene understanding, and even image generation.
- **ImageNet Dataset:** The ImageNet dataset, introduced in 2009, played a pivotal role in the advancement of computer vision. It contains millions of labeled images covering a wide range of object categories. ImageNet facilitated the training of deep learning models, especially CNNs, and allowed them to achieve state-of-the-art performance in image classification and object recognition tasks. This dataset revolutionized the field and contributed to breakthroughs in areas like image understanding and automated image tagging.

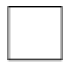
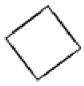
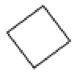
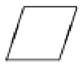
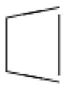
Module 2

Explain the basic set of 2D planar transformations:

In 2D computer graphics and geometry, planar transformations are used to manipulate and transform 2D objects. The basic set of 2D planar transformations includes:

- **Translation (T):** This moves an object by a specified distance in the x and y directions.
- **Rotation (R):** It rotates an object by a given angle around a fixed point.
- **Scaling (S):** This enlarges or shrinks an object in the x and y directions.
- **Shearing (Sh):** Shearing skews an object by displacing points along one axis.
- **Reflection (Ref):** It flips or mirrors an object over a specified axis.

Describe the hierarchy of 2D coordinate transformations with a note on degrees of freedom and what they preserve:

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} I & t \end{bmatrix}_{2\times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} R & t \end{bmatrix}_{2\times 3}$	3	lengths	
similarity	$\begin{bmatrix} sR & t \end{bmatrix}_{2\times 3}$	4	angles	
affine	$\begin{bmatrix} A \end{bmatrix}_{2\times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{H} \end{bmatrix}_{3\times 3}$	8	straight lines	

**********Refer Textbook for note page number 42**********

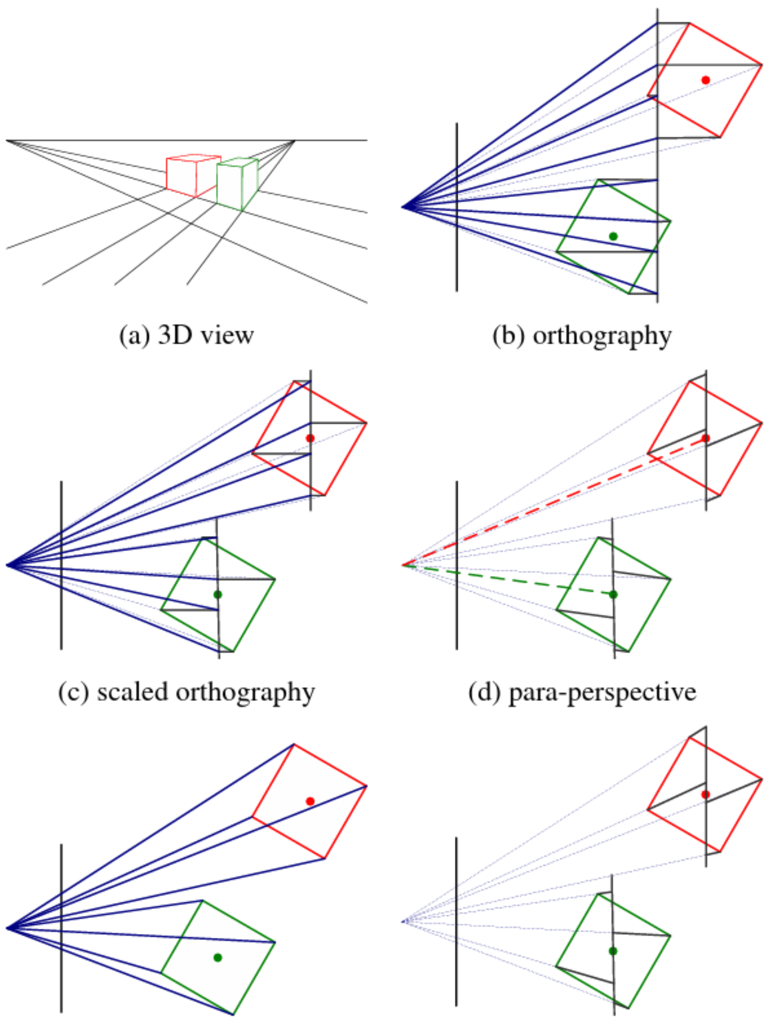
What are quaternions? Explain any use of them with examples:

Quaternions are a mathematical concept used to represent rotations in 3D space. They are an extension of complex numbers and have four components: one real part and three imaginary parts. Quaternions are particularly useful in computer graphics, animation, and robotics to avoid the problems associated with gimbal lock (a phenomenon in Euler angles) and to perform smooth and efficient 3D rotations.

Example: In 3D computer graphics, quaternions can be used to smoothly interpolate between two orientations. For instance, in 3D animation, you can use quaternions to smoothly animate a character's rotation from one pose to another, avoiding the jerky motion that can occur when using other rotation representations.

With neat diagrams, illustrate 5 types of 3D to 2D projection models:

- **Orthographic Projection:** In orthographic projection, all parallel lines remain parallel in the projection. It's often used for technical and architectural drawings.
- **Perspective Projection**: In perspective projection, objects farther from the viewer appear smaller. It's used in 3D rendering to create a sense of depth and realism.
- **Oblique Projection**: Oblique projection is a combination of orthographic and perspective projection. It's used to show objects with one face parallel to the viewer.
- **Isometric Projection:** In isometric projection, all three axes are equally foreshortened. This creates a 3D appearance but without the distortion of perspective projection.
- **Axonometric Projection:** Axonometric projection is a general term for non-perspective 3D projections like isometric and dimetric, where angles between axes are preserved.



(a) 3D view (b) orthography

(c) scaled orthography (d) para-perspective

(e) perspective (f) object-centered

Elucidate the difference between calibration matrix and camera matrix:

The calibration matrix and camera matrix are both matrices that are used in computer vision to map between 3D world coordinates and 2D image coordinates. However, there is a subtle difference between the two matrices.

The calibration matrix is a 3x3 matrix that represents the intrinsic parameters of the camera. The intrinsic parameters are the properties of the camera that are independent of the pose of the camera, such as the focal length, principal point, and skew coefficient.

The camera matrix is a 3x4 matrix that represents the intrinsic and extrinsic parameters of the camera. The extrinsic parameters are the properties of the camera that depend on the pose of the camera, such as the position and orientation of the camera.

The camera matrix can be obtained by multiplying the calibration matrix by a 4x4 pose matrix. The pose matrix represents the position and orientation of the camera relative to the world coordinate system.

In mathematical terms, the camera matrix is defined as follows:

P = K[R|t]

where:

- P is the 3x4 camera matrix
- K is the 3x3 calibration matrix
- R is the 3x3 rotation matrix
- t is the 3x1 translation vector

The camera matrix can be used to project 3D world coordinates into 2D image coordinates as follows:

x = P X

What is meant by normalized device coordinates and where are they used:

Normalized Device Coordinates (NDC) are a coordinate system used in computer graphics and rendering. NDC space is a device-independent, normalized space in which the visible region of a scene is defined. In NDC, the viewing volume is a cube with boundaries ranging from -1 to 1 in each dimension.

NDC coordinates are used for a final stage of transformation in the graphics pipeline before mapping to screen coordinates. It allows for abstraction from the actual display resolution and aspect ratio. The conversion from world or camera coordinates to NDC coordinates involves dividing by the viewing volume dimensions.

Explain 3 types of lens distortions:

- **Radial Distortion:** Radial distortion occurs when points in an image are displaced from their true locations due to the lens not being perfectly geometrically accurate. There are two types of radial distortion: barrel distortion (outward bulging) and pincushion distortion (inward pinching).
- **Tangential Distortion:** Tangential distortion occurs when the lens and the image plane are not perfectly parallel. It leads to a bending or tilting of straight lines in an image.
- **Chromatic Aberration:** Chromatic aberration is the phenomenon where different wavelengths (colors) of light are focused at slightly different points, resulting in color fringing. This occurs due to the variation in the refractive index of a lens with respect to the wavelength of light.

Explain BRDF with an example:

BRDF stands for Bidirectional Reflectance Distribution Function. It describes how light is reflected at a surface point in different directions. The BRDF of a material defines how it scatters light, which is crucial in computer graphics for rendering realistic surfaces.

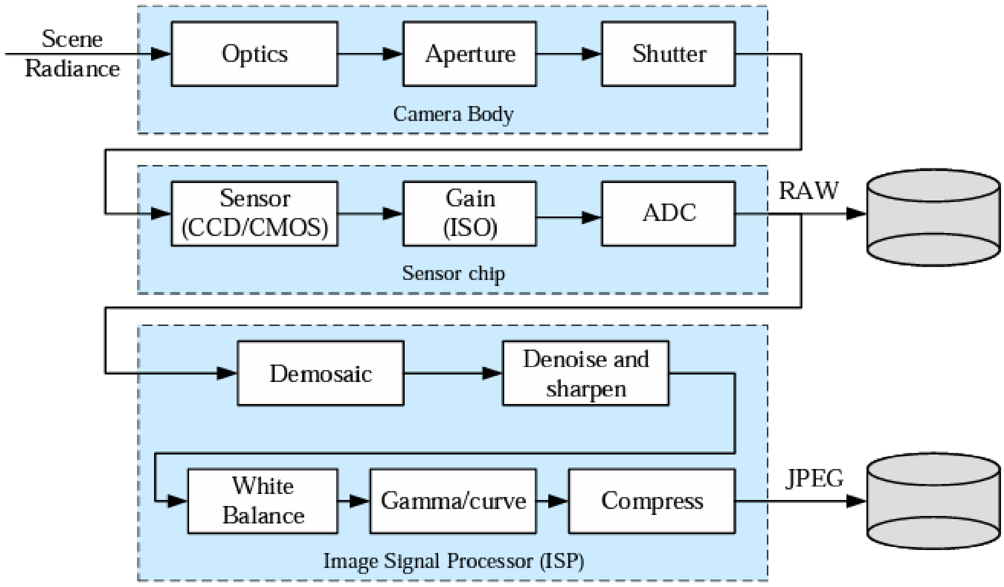
Example: Consider a metallic surface with a high specular BRDF. When light hits this surface, it reflects a significant portion of the incident light in a mirror-like fashion, creating a shiny and reflective appearance. The BRDF parameters determine the intensity and direction of this reflection, impacting how the material appears in a rendered image.

What is meant by a Lambertian surface? How is it connected with BRDF:

A Lambertian surface is an idealized model used in computer graphics and computer vision. It is perfectly diffuse, meaning it scatters light uniformly in all directions, regardless of the viewing or lighting angle. The reflectance of a Lambertian surface is constant for all incoming and outgoing directions.

The connection between a Lambertian surface and BRDF is that a Lambertian material has a specific BRDF, which is constant and equal to the albedo (reflectance) of the material divided by π. This means that it doesn't depend on the lighting or viewing direction. Lambertian surfaces are often used as a simple and efficient approximation for modeling diffuse reflection in computer graphics.

explain the image sensing pipeline in a digital camera:



1. **Lens and Aperture:** The process begins with the camera's lens. The lens's primary function is to focus incoming light onto the camera's image sensor. It plays a crucial role in determining the image's clarity, depth of field, and exposure. The aperture, which is part of the lens, controls the amount of light that enters the camera. It does so by adjusting the size of the aperture's diaphragm, which is represented as an f-number (e.g., f/2.8, f/8). A smaller f-number means a wider aperture, allowing more light to pass through.

2. **Image Sensor:** Once the light passes through the lens, it reaches the image sensor. Most digital cameras use either a Complementary Metal-Oxide-Semiconductor (CMOS) sensor or a Charge-Coupled Device (CCD) sensor. These sensors consist of an array of millions of individual photosites (pixels). Each photosite is responsible for capturing and converting light into an electrical charge.

3. **Photodiodes and Sensitivity:** Within each pixel on the image sensor, there is a photodiode. Photodiodes are light-sensitive devices that generate an electrical charge when exposed to light. The amount of charge generated depends on the intensity and duration of the light hitting the photodiode. Some photosites are more sensitive to red, green, or blue light, allowing the sensor to capture color information.

4. **Analog-to-Digital Conversion (ADC):** After the image sensor collects the charge from each photosite, the analog signal is converted into a digital signal by an Analog-to-Digital Converter (ADC). This process quantizes the electrical charges into discrete values, which are represented as digital pixel values. The bit depth of the ADC determines the number of colors or shades of gray that can be represented.

5. **Signal Processing:** Once the image data is in digital format, various signal processing operations are applied to improve the image quality. These processes may include white balance correction, noise reduction, sharpening, and color correction. These adjustments are made to enhance the visual quality of the image.

6. **Storage:** The processed digital image is then stored in the camera's memory, typically on a memory card. The image file format and compression settings affect the size of the stored image.

7. **Display:** Some digital cameras have a built-in display, such as an LCD screen, that allows users to preview and review captured images. The display provides a quick way to assess the quality of the image and make adjustments if needed.

8. **Data Transfer:** Images can be transferred from the camera to a computer or other devices for further processing and sharing. This can be done using various methods, including USB, Wi-Fi, or Bluetooth.

Give examples of color spaces. Write the equations for converting from one color space to another?

Tb page no 90—91

Write a short note on image compression:

Image compression is the process of reducing the size of digital images while preserving the essential visual information. There are two main types of image compression:

- **Lossless Compression:** Lossless compression reduces file size without any loss of image quality. It's suitable for images where every detail must be preserved, such as medical images or line drawings. Common lossless compression formats include PNG and GIF.

- **Lossy Compression:** Lossy compression achieves higher compression ratios by discarding some image data. This results in some loss of quality, but it's imperceptible in many cases. It is commonly used for photographs and web images. JPEG is a popular lossy compression format.

*****************Refer textbook and write for full marks*********************