MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

MDS6401 / Segment 03

PRINCIPLES OF 3D VISION AND DEPTH PERCEPTION

Table of Contents

1. Understanding Stereo Vision and Image Blurring in Photography 4
2. The Pinhole Camera: Basic Principles and Image Formation 6
3. Depth of Field and Aperture Size in Stereo Vision 7
4. Coordinate Frames and Perspective Projection in Imaging 9
5. Projection Vectors and Pixel Mapping in Imaging 11
6. Homogeneous Coordinates and Transformation in Imaging 13
7. Extrinsic Parameters in Camera Calibration 15
8. From 2D to 3D: Understanding Depth Reconstruction Using Stereo Vision 16
9. Key Concepts in Stereo Vision and Epipolar Geometry 18
    9.1 Epipolar Geometry and 3D Reconstruction in Stereo Vision 21
10. Stereo Matching: Finding Corresponding Points in Stereo Vision 22
11. Image Similarity Measures and Stereo Matching Techniques 25
12. Summary 27

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Introduction

Stereo vision is crucial for perceiving depth in images. It is achieved by capturing two images from slightly different perspectives. This technique, along with understanding the impact of focal length and aperture on image sharpness, is fundamental in photography and imaging systems. By converting 3D scenes into 2D images and back into 3D views, we can enhance visualisation and depth perception, vital in fields like computer vision, robotics, and photography. We will delve into the fascinating world of stereo vision and its critical role in perceiving image depth. Alongside stereo vision, we will explore the impact of focal length and aperture size on image sharpness, which are crucial concepts in photography and imaging systems. Additionally, we will learn about the basic principles of the pinhole camera, coordinate transformations in imaging, and advanced techniques in stereo-matching. These topics will provide a comprehensive understanding of how 3D scenes are converted into 2D images and back into 3D views, enhancing our ability to visualise and analyse depth information in various applications.

Learning Objectives

At the end of this topic, you will be able to:

- Explain the principles of stereo vision and depth perception
- Describe the effects of focal length and aperture on image sharpness
- Explore the basics of the pinhole camera and its role in image formation
- Outline the transformation between coordinate frames in imaging
- List the methods of stereo matching and their practical applications

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

1. Understanding Stereo Vision and Image Blurring in Photography

In stereo vision, the objective is to perceive depth in images, typically achieved through binocular vision. When you capture an image using a standard camera, you obtain a 2D representation of the scene. However, binoculars or stereoscopic cameras capture two images from slightly different perspectives, allowing the brain to perceive depth. The critical parameters here are the distance between the observation points (known as the baseline) and the focal length of the lenses.

The Concept of Blurring and Focal Length

When capturing a 3D scene, it is crucial to convert this into a 2D image while preserving depth information and then back into a 3D view for enhanced visualisation. This intricate transformation process involves understanding how light rays interact with the capturing medium.

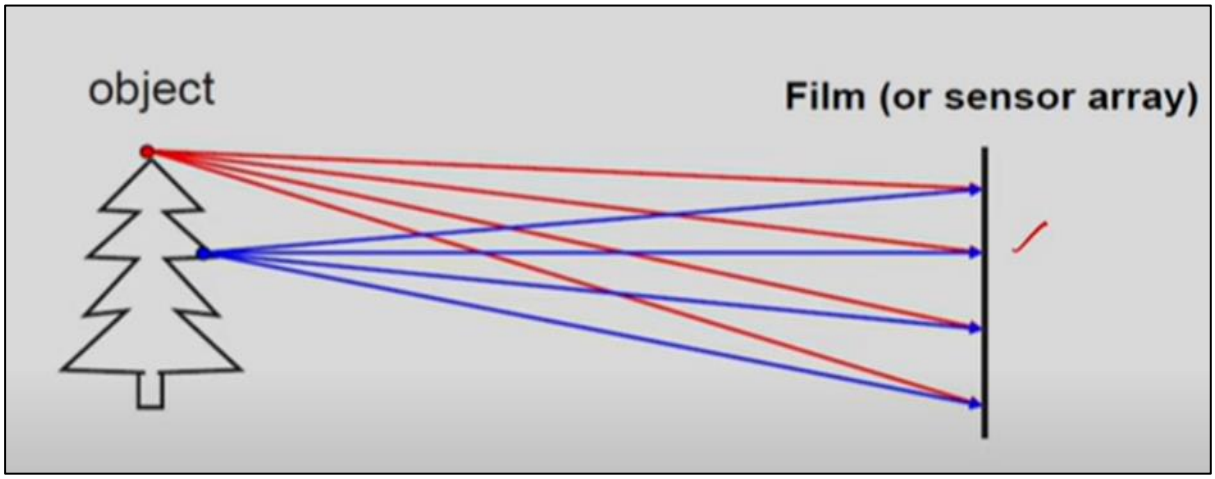

Figure 1: Illustration of Light Rays and Image Blurring on a Film or Sensor Array

Figure 1 illustrates that photons from an object (the tree) travel in straight lines and hit the film or sensor array. Due to their varying angles and positions, these photons interfere with each other, causing a blurred image. This blurring occurs because light from different object parts hits the sensor at various points, leading to overlapping and interference.

We must focus only on the light rays originating from a specific point on the object to mitigate this blurring effect.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

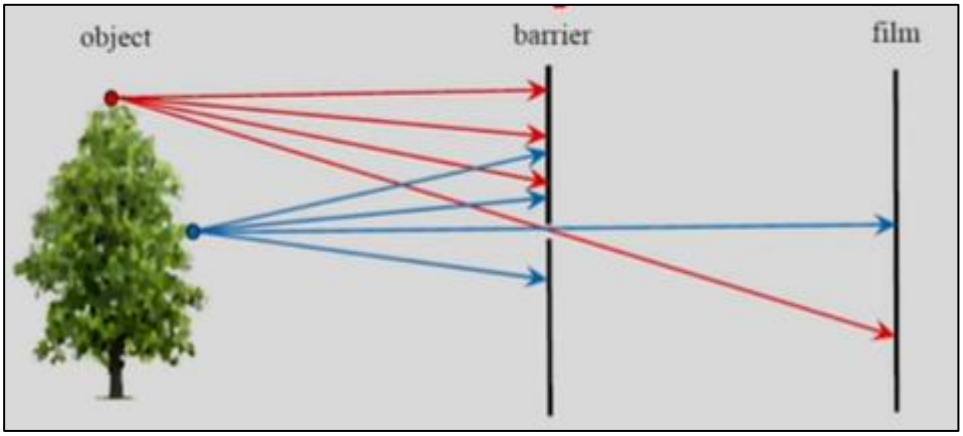

Figure 2: Effect of Barrier (Aperture) on Light Ray Interference and Image Sharpness

One effective method to reduce blurring is introducing a barrier (aperture), as depicted in Figure 2. The barrier selectively allows light rays from a particular point on the object to pass through while blocking the rest. This selective allowance reduces interference from unwanted light rays, enhancing the focus on the desired object.

In photography, this barrier is known as the aperture. The aperture size determines how much light is allowed to reach the sensor, thereby controlling the focus and sharpness of the image.

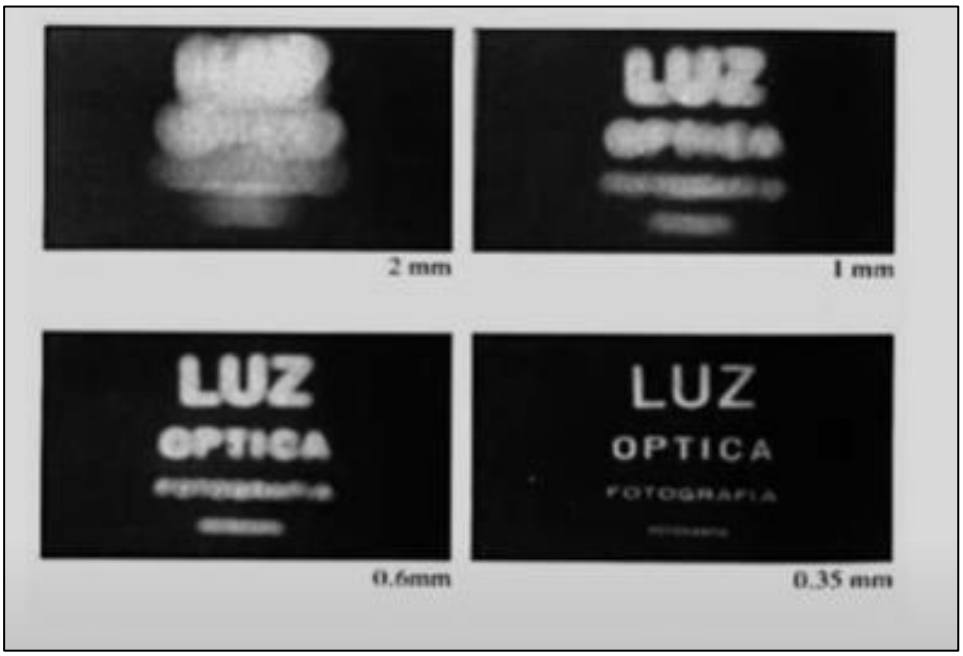

Figure 3: Impact of Aperture Size on Image Sharpness

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Figure 3 demonstrates the impact of varying aperture sizes on image sharpness. The image becomes progressively sharper as the aperture size decreases (from 2mm to 0.35mm). A smaller aperture reduces blurring by limiting the light rays that can interfere with the primary light rays from the object.

However, it is essential to note that the aperture size cannot be reduced indefinitely. An excessively small aperture would block most light rays, making it impossible to form a clear image. Additionally, a very small aperture can introduce diffraction effects, further degrading the image quality.

2. The Pinhole Camera: Basic Principles and Image Formation

The pinhole camera is one of the simplest and most fundamental image-capture tools. This device operates on basic principles of optics that are often taught in high school science classes. Figure 4 illustrates the key components and processes involved in a pinhole camera's functioning.

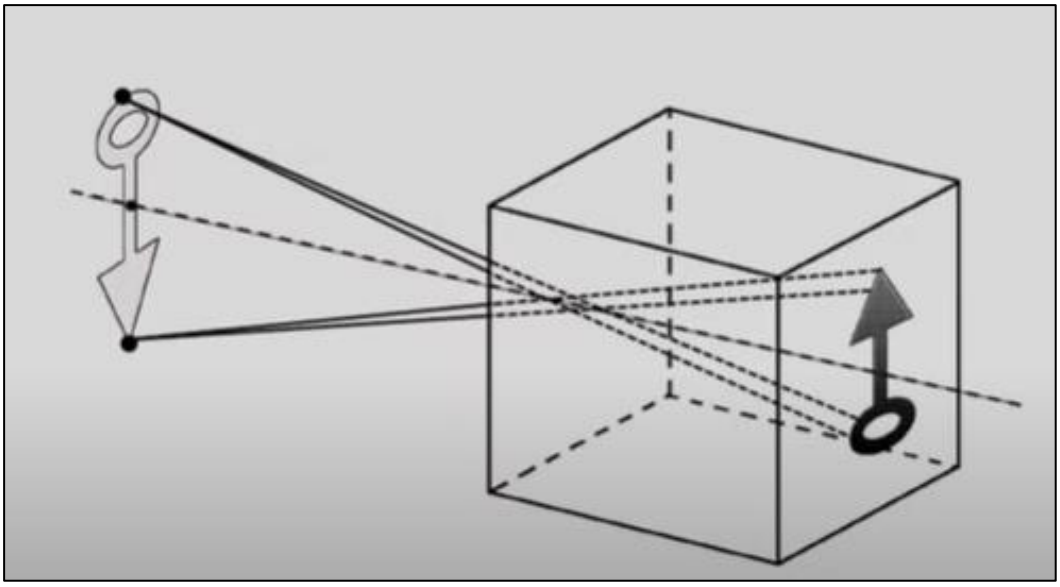

Figure 4: Principle of Image Formation in a Pinhole Camera

The object (in this case, represented by an arrow and circle) is situated in front of a box with a small aperture, known as the pinhole. Light rays from the object pass through this pinhole

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

and converge to form an inverted image on the opposite side of the box, referred to as the image plane.

Key Components and Concepts

- **Object Plane:** This is where the object being captured is located.
- **Pinhole:** A small aperture that allows light rays to pass through. It acts as a barrier that limits the light, ensuring that only light rays from specific angles reach the image plane.
- **Image Plane:** The plane where the inverted image of the object is formed. This occurs because the light rays cross each other as they pass through the pinhole.

Image Formation

The light rays originating from different points on the object travel in straight lines and pass through the pinhole, forming an inverted image on the image plane. This is due to the crossing of light paths. For instance, light from the top of the object travels downward through the pinhole and hits the lower part of the image plane, while light from the bottom travels upward to the top.

Human Eye and Image Reversal

A similar process occurs in the human eye, where the lens focuses light onto the retina, creating an inverted image. However, the brain processes this image, reversing it to appear upright, allowing us to perceive the world correctly.

3. Depth of Field and Aperture Size in Stereo Vision

The primary challenge in stereo vision is accurately identifying the depth of an image or field. Stereo vision involves capturing two slightly different perspectives of the same scene to create a sense of depth. However, understanding the role of aperture size is crucial for achieving the desired focus and clarity in images.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

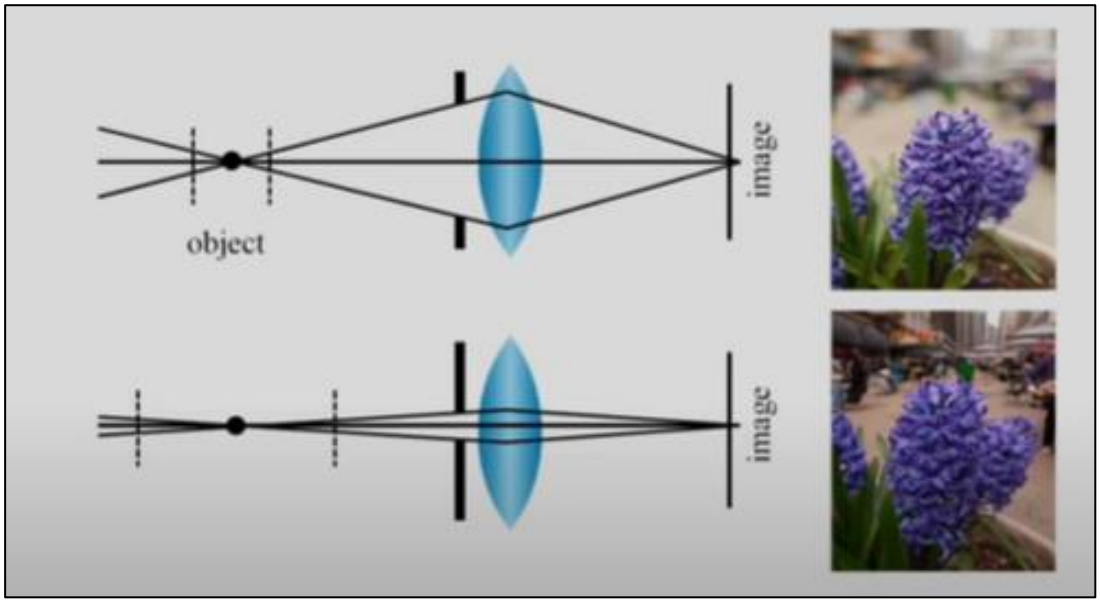

Figure 5: Effect of Aperture Size on Depth of Field in Photography

Figure 5 demonstrates the relationship between aperture size and depth of field. Two scenarios are illustrated: one with a larger aperture and one with a smaller one, each affecting the focus and background blur differently.

Key Concepts

- Object: The subject being captured.
- Lens: The optical component that focuses light rays.
- Image Plane: The plane where the image is formed.

Impact of Aperture Size

The aperture size plays a significant role in determining the depth of field in an image:

- Large Aperture (Top Image): When the aperture is large, more light rays from the object can pass through the lens. This results in a shallow depth of field, meaning the object is in sharp focus while the background appears blurred. This effect is ideal for isolating the subject from its surroundings.
- Small Aperture (Bottom Image): A smaller aperture restricts the number of light rays entering the lens. This increases the field depth, bringing the object and its background into clearer focus. This effect is suitable for scenes with important subjects and background details.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Practical Implications

In practical photography, adjusting the aperture size allows photographers to control the focus and background blur. Using a larger aperture, they can create a shallow depth of field, highlighting a specific subject while blurring out distracting backgrounds. With a smaller aperture, they can achieve a greater depth of field, capturing more details in both the foreground and background.

4. Coordinate Frames and Perspective Projection in Imaging

Understanding the transformation between different coordinate frames is crucial in imaging and computer vision. While we often work within Cartesian or Euclidean coordinate systems, we need to consider world coordinate frames, camera coordinate frames, and image planes for imaging purposes.

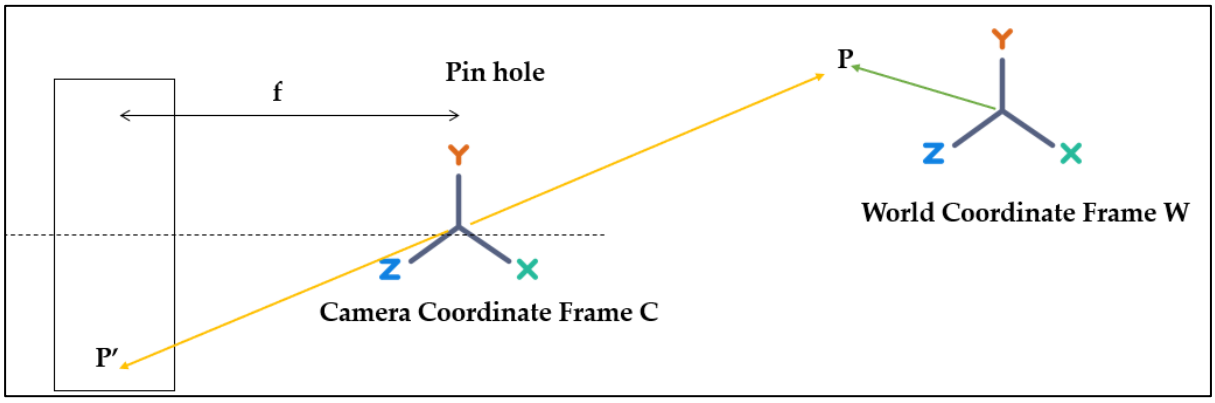

Figure 6: Transformation from World Coordinate Frame to Camera Coordinate Frame in Perspective Projection

Figure 6 illustrates the relationship between the world coordinate frame (W), the camera coordinate frame (C), and the image plane. The object P exists in the world coordinate frame, and its position is defined by a vector from the origin to P. This vector encompasses all three coordinates (X, Y, Z).

The camera coordinate frame is centred at the pinhole of the camera, where the reference point lies. The vector from the camera to the object P is the projection vector. Once the image is captured by the camera, the object P is projected onto the image plane at the point P'.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

This transformation process can be broken down into three main stages:

1. **World Coordinate Frame to Camera Coordinate Frame:** This is a 3D-to-3D transformation in which the object's coordinates in the world frame are converted to the camera frame.
2. **Camera Coordinate Frame to Image Plane:** This is a 3D to 2D transformation, where the 3D coordinates are projected onto a 2D image plane, resulting in the loss of depth information (Z coordinate).

Transformation Process

The transformation from the world coordinate frame to the image plane involves several steps:

1. World Coordinate Frame to Camera Coordinate Frame (3D to 3D):
   - Let P= $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be the object coordinates in the world frame.
   - Apply the coordinate transformation to get the coordinates in the camera frame, $C = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$
2. Camera Coordinate Frame to Image Plane (3D to 2D):
   - The perspective projection transforms the 3D coordinates C to 2D coordinates I = $\begin{pmatrix} x \\ y \end{pmatrix}$ on the image plane.

Loss of Depth Information

During the transformation from the camera coordinate frame to the image plane, the depth information (Z coordinate) is lost. The image plane can only represent 2D coordinates (X, Y). As a result, the final image projection lacks the depth information in the 3D world coordinate frame.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

5. Projection Vectors and Pixel Mapping in Imaging

Understanding the transition between different coordinate systems, including world coordinate frames, camera coordinate frames, and image planes, is essential in imaging and computer vision. This understanding facilitates the accurate projection of 3D objects into 2D images, enabling effective image analysis and processing.

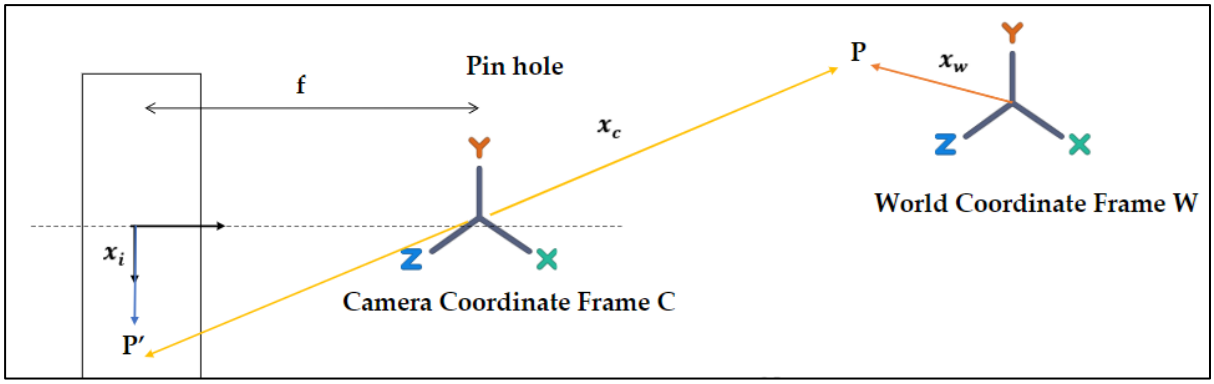

Figure 7: Projection of World Coordinate to Camera Coordinate and Image Plane

Figure 7 illustrates the relationship between the world coordinate frame (W), the camera coordinate frame (C), and the image plane. The object P exists in the world coordinate frame, and its position is defined by a vector from the origin to P. This vector encompasses all three coordinates (X, Y, Z).

The camera coordinate frame is centred at the pinhole of the camera, where the reference point lies. The vector from the camera to the object P is the projection vector x_c. Once the image is captured by the camera, the object P is projected onto the image plane at point P'.

The projection from the camera coordinate frame to the image plane involves the transformation of 3D coordinates to 2D coordinates:

$$x_i = T \cdot x_c$$

$$y_i = T \cdot y_c$$

The equations demonstrate how the coordinates, x_i and y_i in the image plane are derived from the coordinates x_c and y_c in the camera coordinate frame using the transformation matrix T.

$$u = m_i \cdot T \cdot x_c$$

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

$$v = m_j \cdot T \cdot y_c$$

The above equations provide a detailed view of the pixel mapping process. It shows how the coordinates in the image plane are further transformed into pixel coordinates on the image sensor. Here, m_i and m_j are the pixel densities along the x and y axes, respectively.

Figure 8 highlights the importance of considering the origin's position in the image plane.

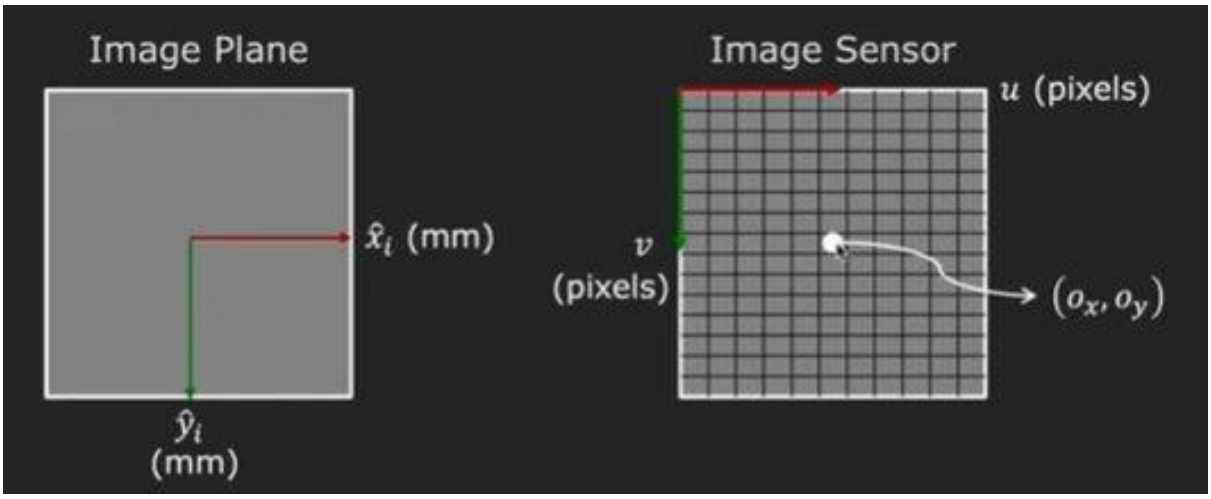

Figure 8: Mapping Image Plane Coordinates to Pixel Coordinates on an Image Sensor

If the origin is not at the centre but at the top-left corner, an offset needs to be added to the transformation equations:

$$u = m_i \cdot T \cdot x_c + O_x$$

$$v = m_i \cdot T \cdot y_c + O_x$$

The above equations consolidate the above concepts, showing how the camera's intrinsic parameters (such as pixel density and offsets) are used to calibrate the camera and accurately map the coordinates from the camera frame to the image plane and then to the pixel grid.

Importance of Intrinsic Parameters

Intrinsic parameters are critical for camera calibration. These parameters, which include pixel density, focal length, and offsets, are inherent to the camera's design and calibration. Knowing these parameters allows for accurately converting 3D to 2D-pixel coordinates, which is essential for precise image capturing and processing.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

6. Homogeneous Coordinates and Transformation in Imaging

Nonlinear Systems and Linearisation

In image processing and computer vision, we often encounter complex nonlinear systems with which to work directly. These systems need to be linearised to simplify computations and facilitate easier manipulation. Additionally, working with matrices is a preferred approach due to their computational efficiency.

Homogeneous Coordinate System

One of the most effective methods to achieve linearisation and ease of matrix manipulation is using homogeneous coordinates. This system introduces an additional fictitious coordinate, enabling the transformation of nonlinear equations into linear ones.

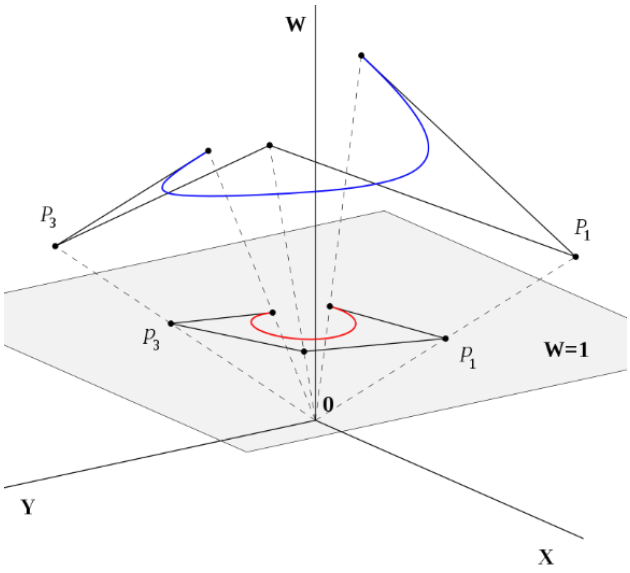

Figure 9: Transformation of 2D Coordinates to Homogeneous Coordinates

Figure 9 demonstrates a typical transformation matrix A that combines rotation R and translation T. This matrix converts coordinates from one frame to another, from a world coordinate frame to a camera coordinate frame.

In homogeneous coordinates, the transformation matrix becomes:

$$A = [R\ T]$$

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

This matrix enables the transformation by considering rotation and translation, making working with different coordinate frames easier.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\,\theta & -sin\,\theta & 0 \\ 0 & sin\theta & cos\,\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} \rightarrow \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

Euclidean homogeneous

The above equation shows the process of converting Euclidean coordinates to homogeneous coordinates. This conversion introduces an additional coordinate, making the system linear and suitable for matrix operations.

The homogeneous representation of a 2D point u=(u,v) as a 3D point is u'=(u',v',w'). The additional coordinate w' is a fictitious component that facilitates linearisation as below:

$$u \equiv \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{w}u \\ \tilde{w}v \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \tilde{u}$$

Advantages of Homogeneous Coordinates

- Linearisation: Homogeneous coordinates help convert nonlinear systems into linear ones, simplifying the computation and manipulation of complex transformations.
- Matrix Representation: Using matrices for transformations allows for efficient computational processes, as matrix operations are well-supported in various mathematical and programming frameworks.
- Additional Dimension: Introducing a fictitious coordinate adds a new dimension to the system, providing additional flexibility in representing and transforming coordinates.

Application in Robotics

In robotics, homogeneous coordinates are particularly useful for normalising transformations and scaling. For example, a robot with stereo vision systems can use homogeneous coordinates to convert 2D image points to 3D points, facilitating tasks like path planning and navigation.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

By representing transformations as matrices, robots can efficiently compute the best paths using algorithms like A* and Dijkstra based on the 3D views captured by their cameras. The additional dimension in homogeneous coordinates ensures that the transformations remain accurate and consistent.

7. Extrinsic Parameters in Camera Calibration

We must delve into homogeneous coordinates and extrinsic parameters to understand how 2D images are transformed into 3D representations. Homogeneous coordinates allow for the linearisation of transformations, making it easier to work with 3D points and their projections onto 2D planes.

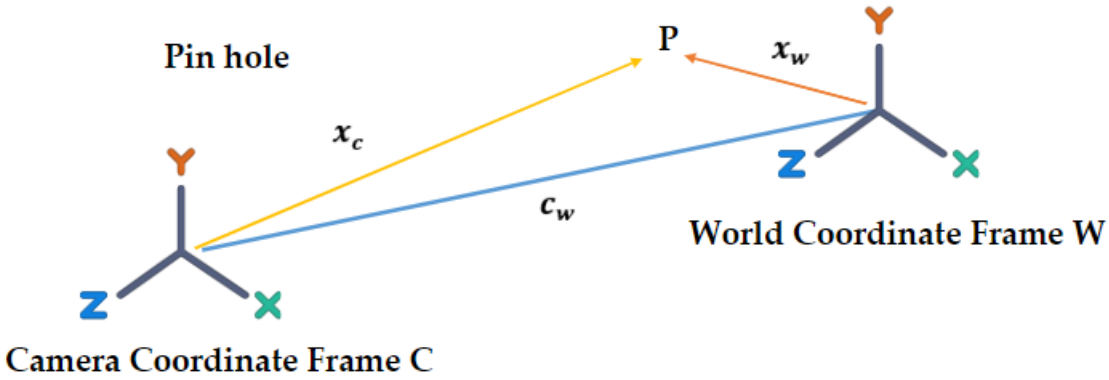

Figure 10: Transformation from World Coordinate Frame to Camera Coordinate Frame

Figure 10 illustrates the relationship between the camera coordinate frame (C) and the world coordinate frame (W). The point P in the world coordinate frame has coordinates x_w. The camera coordinate frame is centred at the pinhole, with coordinates x_c representing the same point in the camera's view.

The transformation between these coordinate frames can be expressed as:

$$x_c = R(x_w - c_w) = Rx_w + t$$

R is the rotation matrix, and t is the translation vector. This equation transforms the world coordinates x_w to camera coordinates x_c. The above equation describes how the coordinates in the world frame are transformed into the camera frame, incorporating rotation and translation.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Homogeneous Coordinates

Converting from 2D to 3D involves adding an extra dimension, represented as W in the homogeneous coordinate system. This extra coordinate facilitates the linearisation of transformations and enables matrix operations for computational efficiency.

When an image is represented in 2D, it has X and Y coordinates. By converting it to a homogeneous coordinate system, we add one more dimension, W, along the Z direction. This conversion allows for a more comprehensive representation of the point in 3D space.

Extrinsic Parameters

Extrinsic parameters define the position and orientation of the camera in the world coordinate system. These parameters include:

1. Rotation Matrix (R): Describes the camera's orientation relative to the world coordinate frame.
2. Translation Vector (t): Describes the camera's position relative to the world coordinate frame.

Given these extrinsic parameters, the position of a point P in the world coordinate frame can be transformed into the camera coordinate frame using the equation:

$$x_c = R(x_w - c_w)$$

The equation can be expressed in a simplified form:

$$x_c = Rx_w + t$$

8. From 2D to 3D: Understanding Depth Reconstruction Using Stereo Vision

In image processing and computer vision, transforming the 3D coordinates of an object to a 2D image plane is a well-understood process using transformation matrices. However, the reverse—reconstructing 3D coordinates from 2D images—presents a more complex challenge. Stereo vision solves this problem by utilising multiple images captured from different perspectives.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

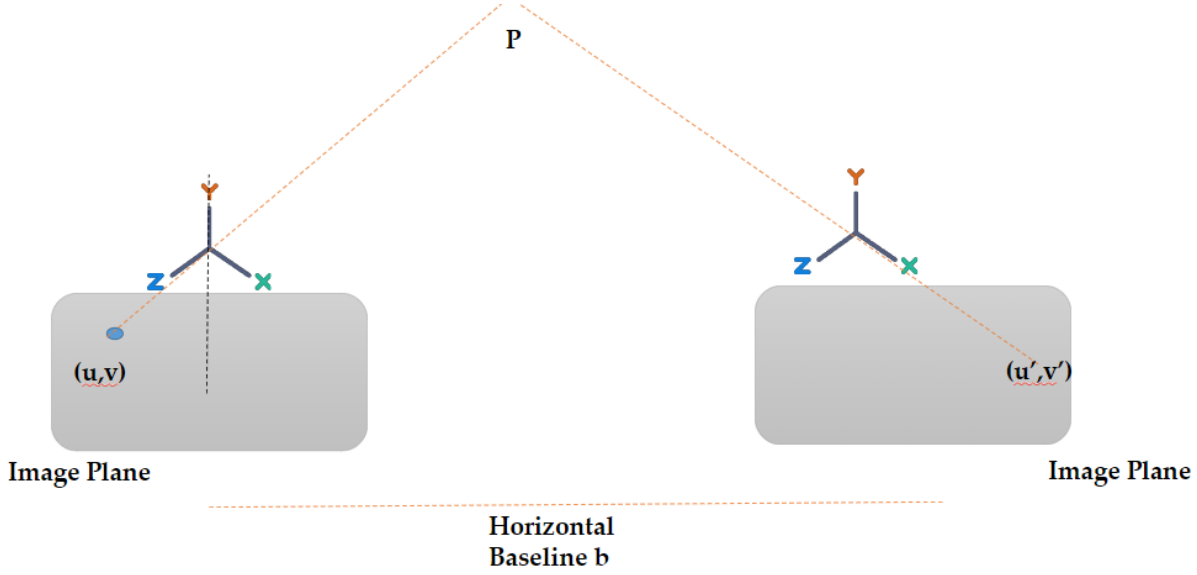

Figure 11: Stereo Vision for Depth Reconstruction: Horizontal Baseline and Disparity

Figure 11 illustrates the concept of stereo vision for depth reconstruction. It shows two cameras positioned with a horizontal baseline between them, capturing images of the same object from different angles.

1. **Camera Coordinate Frames:** The left and right cameras have their own coordinate frames. The coordinates (u, v) represent the projection of the object P on the left camera's image plane, while (u', v') represent the projection on the right camera's image plane.
2. **Horizontal Baseline (b):** The distance between the two cameras' centres is the horizontal baseline, which is crucial for calculating the depth of information.
3. **Projection and Disparity:** The object P is projected onto the image planes of both cameras. The slight difference in position between the projections (u, v) and (u', v') is referred to as disparity. This disparity is essential for determining the depth of the object.

Depth Reconstruction Process

To reconstruct the 3D coordinates of an object from 2D projections, we utilise the following steps:

1. **Projection Identification:** Identify the projections (u, v) and (u', v') of the object P on the left and right image planes, respectively.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

2. **Line of Sight Intersection:** Each camera captures a line of sight extending from its centre through the projection point on the image plane. These lines of sight intersect at the object's actual position in the 3D world.
3. **Calculating Depth:** By using the disparity (difference in projection positions) and the known baseline distance, we can calculate the depth (z-coordinate) of the object. The depth calculation is based on the principle that the point of intersection of the lines of sight from both cameras is where the object lies.

Disparity and Depth

Disparity is the object's position difference between the two image planes. It is directly related to the depth of the object:

$$Depth\ (z) = \frac{b \cdot f}{d}$$

where:

b is the horizontal baseline

f is the focal length of the cameras

d is the disparity

A larger disparity indicates a closer object, while a smaller disparity indicates a farther object.

9. Key Concepts in Stereo Vision and Epipolar Geometry

To understand how we recover the 3D structure of a scene using multiple images, we need to delve into several fundamental concepts: intrinsic parameters, extrinsic parameters, perspective projection, and stereo vision. These concepts are crucial for accurate 3D reconstruction and depth estimation.

Intrinsic and Extrinsic Parameters

- **Intrinsic Parameters**

Intrinsic parameters are inherent to the camera and define its internal characteristics. These include the focal length, principal point, and pixel density.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

These parameters are necessary for converting 3D coordinates in the camera coordinate frame to 2D coordinates on the image plane.

- **Extrinsic Parameters**

Extrinsic parameters, including rotation and translation matrices, define the camera's position and orientation in the world coordinate system.

They transform coordinates from the world coordinate frame to the camera coordinate frame.

- **Perspective Projection**

Perspective projection is projecting 3D points onto a 2D image plane. This transformation involves both intrinsic and extrinsic parameters.

The projection converts the spatial arrangement of objects in the 3D scene into a 2D image, maintaining the perspective view of the scene.

- **Stereo Vision**

Stereo vision involves using two or more images from different viewpoints to reconstruct a scene's 3D structure. This technique mimics binocular vision in humans, where each eye captures a slightly different view, allowing the brain to perceive depth.

In computer vision, stereo vision can be achieved using multiple cameras or a single camera moving between positions.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

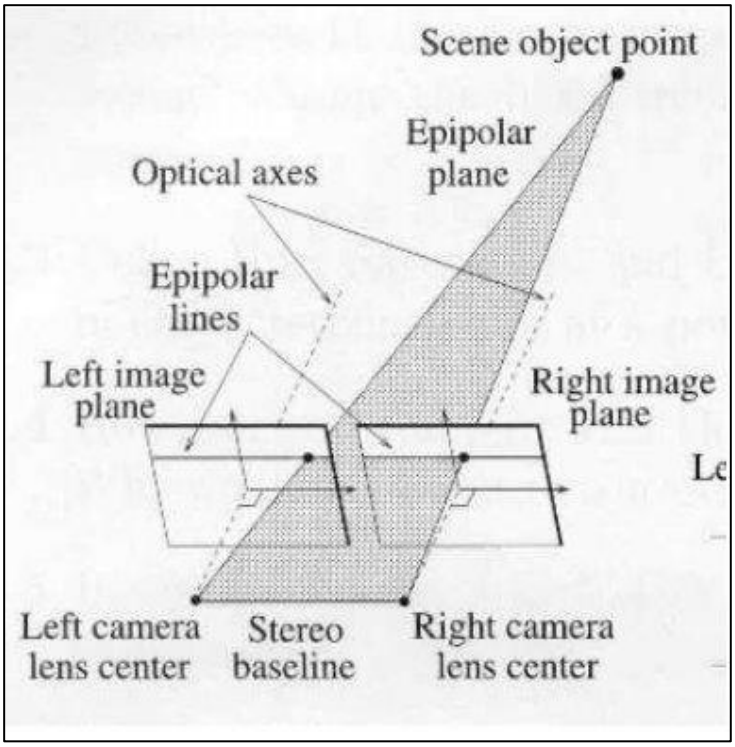

Figure 12: Epipolar Geometry in Stereo Vision

Figure 12 illustrates the concept of epipolar geometry in stereo vision. Key components include the epipolar plane, epipolar lines, and the cameras' projection centres.

1. **Optical Axes:** The central lines of sight for each camera.
2. **Epipolar Plane:** The plane passes through the projection centres of both cameras and the scene object point.
3. **Epipolar Lines:** The intersection of the epipolar plane with the image planes of both cameras.
4. **Stereo Baseline:** The distance between the centres of projection of the two cameras.
5. **Image Planes:** The 2D planes where the projections of the 3D scene are captured.

Epipolar Geometry

- Epipolar geometry is the study of the geometric relationship between two views of a scene. Understanding how points in one image correspond to those in another is crucial.
- Epipolar Plane: The plane that passes through the scene object point and the optical centres of the two cameras.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

- Epipolar Lines: The lines where the epipolar plane intersects the image planes of the cameras. These lines are essential for finding corresponding points in stereo images.

Conjugate Pair and Disparity

- Conjugate Pair: Refers to corresponding points in the left and right images that represent the same scene point.
- Disparity: The difference in the position of corresponding points in the left and right images. Disparity is used to calculate the depth of the scene point.

9.1 Epipolar Geometry and 3D Reconstruction in Stereo Vision

Epipolar geometry is fundamental in computer vision (CV) and photogrammetry, especially in stereo vision and 3D reconstruction. It involves using epipolar lines and planes to understand the depth and spatial relationships between multiple images taken from different viewpoints.

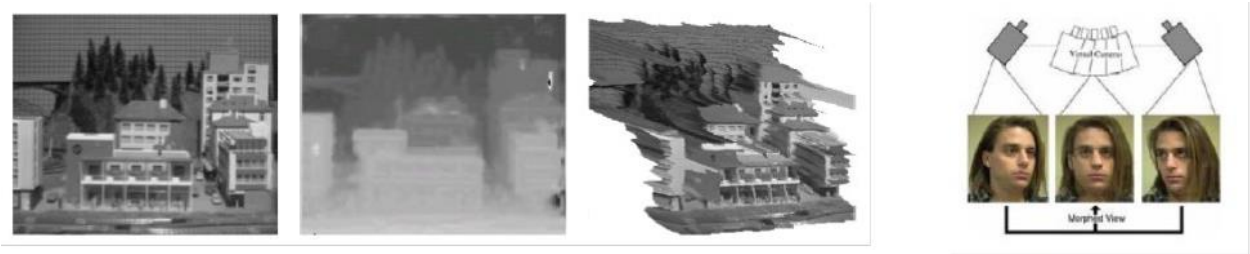

Figure 13: 3D Reconstruction Using Depth Maps and Multiple Camera Views

Figure 13 illustrates the sequence of the process of 3D reconstruction using stereo vision:

1. Original Image (Left): The first image is the actual captured image of a scene.
2. Depth Map (Centre): The second image represents the depth map, which shows the depth information (third coordinate) for each point in the scene. This map indicates how far each point is from the camera.
3. 3D Reconstruction (Right): The third image is the 3D reconstruction of the scene using the depth information. This reconstructed image provides a three-dimensional view, showing the spatial relationships and relative distances between objects in the scene.
4. Stereo Vision Setup (Far Right): The diagram illustrates the use of multiple cameras to capture different viewpoints of the same scene. This setup is essential for obtaining the depth information required for 3D reconstruction.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

How Depth Perception Works

Depth perception in human vision is based on the differences in the appearance of objects between the left and right eyes. This disparity allows the brain to perceive depth and spatial relationships. Similarly, stereo vision systems use multiple cameras to capture images from different perspectives in computer vision.

Epipolar Lines and Planes

- Epipolar Line: The epipolar line is a key concept in stereo vision. It is the line of intersection between the epipolar and image planes. Each point in one image corresponds to an epipolar line in the other image.
- Epipolar Plane: This plane passes through the centres of projection of the two cameras and the point in the scene. It helps determine the corresponding points in the left and right images.

Practical Applications

- 3D Reconstruction: Using the depth information and the principles of epipolar geometry, we can reconstruct a 3D scene from 2D images. The depth map provides the necessary information to understand the spatial arrangement of objects.
- Advanced Photography: Techniques like those used in high-quality videography, such as the multi-camera setups in Netflix productions, leverage stereo vision to capture detailed depth information and create immersive visual experiences.
- Augmented Reality and Robotics: Epipolar geometry is used in augmented reality (AR) and robotics to understand and navigate the environment accurately.

10. Stereo Matching: Finding Corresponding Points in Stereo Vision

Stereo-matching is a critical process in stereo vision. It aims to identify corresponding points in two or more images captured from different viewpoints. This process is essential for reconstructing the 3D structure of a scene by determining the depth of information from the disparity between corresponding points.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

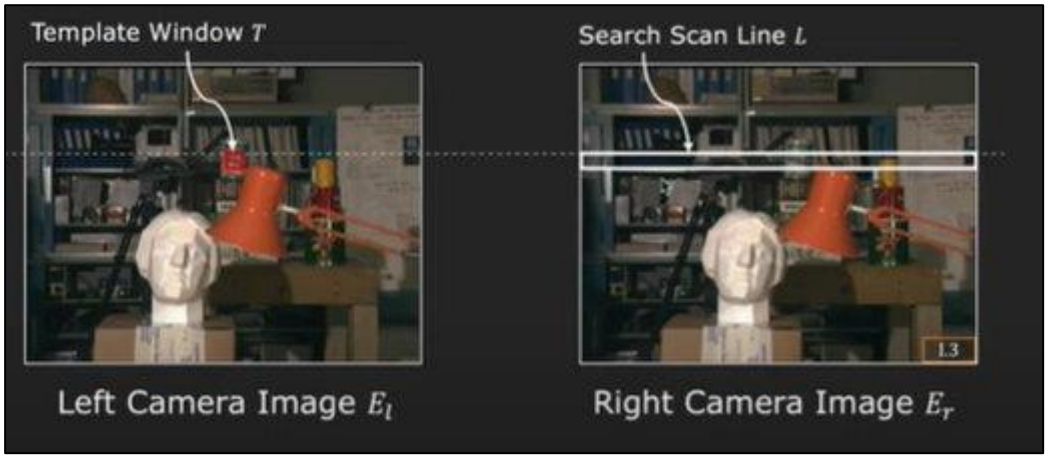

Figure 14: Stereo Matching Using Epipolar Geometry: Template and Search Scan Line

Figure 14 demonstrates the concept of stereo matching using epipolar geometry:

1. **Left Camera Image (E_l):** The left side shows an image captured by the left camera, with a red box highlighting a specific point of interest.
2. **Right Camera Image (E_r):** The right side shows an image captured by the right camera, with a white box indicating the search area for the corresponding point of the red box in the left image.
3. **Template Window (T):** The red box in the left camera image represents a template window to locate the corresponding point in the right camera image.
4. **Search Scan Line (L):** The white box in the right camera image represents the epipolar line along which the corresponding point is expected to lie.

Epipolar Constraint and Epipolar Lines

The epipolar constraint is a key principle in stereo vision that reduces the search space for finding corresponding points. According to this constraint, the corresponding point in one image must lie on the epipolar line in the other image. This reduces the two-dimensional search problem to a one-dimensional search along the epipolar line.

- **Epipolar Line:** The intersection of the epipolar plane with the image plane. It represents the possible locations of the corresponding point in the other image.
- **Epipolar Plane:** The plane formed by the 3D scene point and the optical centres of the two cameras.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Stereo Matching Process

1. Identify the Template Window (T): Select a window around the point of interest in the left camera image.
2. Search Along the Epipolar Line (L): Use the epipolar constraint to limit the search area for the corresponding point to the epipolar line in the right camera image.
3. Template Matching: Compare the template window with windows along the epipolar line in the right image using similarity measures such as the Jaccard index or cosine similarity.
4. Determine Corresponding Point: Identify the window in the right image that best matches the template window from the left image. This matching point represents the corresponding coordinates in the right image.

Key Concepts

1. Disparity: The difference in coordinates of the corresponding points in the left and right images. The disparity is used to calculate depth, with greater disparity indicating closer objects and smaller disparity indicating farther objects.
2. Depth and Disparity Relationship: Depth is inversely proportional to disparity. As the disparity increases, the depth decreases, and vice versa.
3. Baseline: The distance between the two cameras' projection centres. The disparity is directly proportional to the baseline; a larger baseline increases disparity, making depth estimation more accurate.

Practical Application

Stereo matching is widely used in various applications, including:

- 3D Reconstruction: Recovering the 3D structure of a scene from multiple 2D images.
- Robotics: Enabling robots to perceive depth and navigate their environment.
- Augmented Reality: Integrating virtual objects into real-world scenes with accurate depth perception.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

11. Image Similarity Measures and Stereo Matching Techniques

Various similarity measures are employed to match corresponding points in stereo images accurately. These measures help quantify the differences or similarities between image patches, enabling effective stereo matching. The primary similarity measures include the Sum of Absolute Differences (SAD), the Sum of Squared Differences (SSD), and Zero-mean Normalised Cross-Correlation (ZNCC).

1. Sum of Absolute Differences (SAD):
   - Formula: $s = \sum_{(u,v)\in I} |I_1[u,v] - I_2[u,v]|$
   - Toolbox Function: 'sad ( )'
   - Description: Measures the absolute differences between pixel values of two image patches. It is simple and fast but may not handle noise well.

2. Sum of Squared Differences (SSD):
   - Formula: s = $\sum_{(u,v)\in I} (I_1[u,v] - I_2[u,v])^2$
   - Toolbox Function: 'ssd ( )'
   - Description: Computes the squared differences between pixel values, providing a more robust measure against noise than SAD.

3. Zero-mean Normalised Cross-Correlation (ZNCC):
   - Formula:

$$s = \frac{\sum_{(u,v)\in I} I_1[u,v] \cdot I_2[u,v]}{\sqrt{\sum_{(u,v)\in I} I_1^2[u,v] \cdot \sum_{(u,v)\in I} I_2^2[u,v]}}$$

   - Toolbox Function: 'zncc ( )'
   - Description: Normalises the image patches before computing the cross-correlation, making it invariant to linear changes in illumination.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

Figure 15: Comparative Analysis of Stereo Matching Techniques for Depth Estimation

Figure 15 showcases the application of these similarity measures in stereo matching:

1. Left Image: The initial image captured by the left camera.
2. Right Image: The initial image captured by the right camera.
3. Ground Truth: The reference depth map obtained through precise measurements.
4. SSD (Window size=21): This depth map was obtained using SSD with a fixed window size of 21. It provides basic matching but may lack accuracy in varying depth regions.
5. SSD—Adaptive Window: This is an improved method in which the window size adapts based on the local image characteristics, enhancing the matching accuracy.
6. State of the Art: The latest stereo-matching technique closely matches the ground truth and significantly improves depth estimation.

Practical Considerations in Stereo Matching

- Epipolar Geometry: Utilised to constrain the search for corresponding points to the epipolar line, reducing the computational complexity.
- Disparity and Depth Relationship: Depth is inversely proportional to disparity, directly proportional to the baseline distance between the cameras.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION

- Challenges in Stereo Matching: Issues such as occlusions, textureless regions, and image distortions can affect the accuracy of stereo matching. Adaptive methods and advanced algorithms aim to mitigate these challenges.

12. Summary

In this topic, we discussed:

- Stereo Vision: Captures depth using binocular vision, utilising two images from different perspectives.
- Image Blurring: Caused by varying angles and positions of light rays hitting the sensor.
- Aperture Size: Controls light entry and impacts image sharpness; smaller apertures reduce blurring but can introduce diffraction.
- Pinhole Camera: Demonstrates basic image formation principles using a small aperture to project inverted images.
- Coordinate Transformation: Essential for converting 3D world coordinates to 2D image coordinates, involving perspective projection.
- Homogeneous Coordinates: Simplify transformations and calculations in imaging by introducing an additional fictitious coordinate.
- Depth Reconstruction: Achieved using stereo vision, where the disparity between images from different viewpoints determines depth.
- Epipolar Geometry: Reduces search space in stereo matching by constraining corresponding points to the epipolar line.
- Stereo Matching Techniques: Various similarity measures like SAD, SSD, and ZNCC help identify corresponding points in stereo images.
- Practical Applications: These include 3D reconstruction, robotics, and augmented reality, which enhance depth perception and image analysis.

©COPYRIGHT 2023 (VER. 1.0), ALL RIGHTS RESERVED. MANIPAL ACADEMY OF HIGHER EDUCATION