Indira Gandhi National Open University

MCS-230
Digital Image Processing and Computer Vision

Block

3

COMPUTER VISION-I

Unit 8	Introduction to computer Vision	211
Unit 9	Single Camera Models	224
Unit 10	Multiple Cameras	236

PROGRAMME DESIGN COMMITTEE

Prof. (Retd.) S.K. Gupta , IIT, Delhi
Prof. Ela Kumar, IGDTUW, Delhi
Prof. T.V. Vijay Kumar JNU, New Delhi
Prof. Gayatri Dhingra, GVMITM, Sonipat
Mr. Milind Mahajan,. Impressico Business Solutions, New Delhi
Sh. Shashi Bhushan Sharma, Associate Professor, SOCIS, IGNOU
Sh. Akshay Kumar, Associate Professor, SOCIS, IGNOU
Dr. P. Venkata Suresh, Associate Professor, SOCIS, IGNOU
Dr. V.V. Subrahmanyam, Associate Professor, SOCIS, IGNOU
Sh. M.P. Mishra, Assistant Professor, SOCIS, IGNOU
Dr. Sudhansh Sharma, Assistant Professor, SOCIS, IGNOU

COURSE DESIGN COMMITTEE

Prof. T.V. Vijay Kumar, JNU, New Delhi
Prof. S.Balasundaram, JNU, New Delhi
Prof. D.P. Vidyarthi, JNU, New Delhi
Prof. Anjana Gosain, USICT, GGSIPU New Delhi
Dr. Ayesha Choudhary, JNU, New Delhi
Sh. Shashi Bhushan Sharma, Associate Professor, SOCIS, IGNOU
Sh. Akshay Kumar, Associate Professor, SOCIS, IGNOU
Dr. P. Venkata Suresh, Associate Professor, SOCIS, IGNOU
Dr. V.V. Subrahmanyam, Associate Professor, SOCIS, IGNOU
Sh. M.P. Mishra, Assistant Professor, SOCIS, IGNOU
Dr. Sudhansh Sharma, Assistant Professor, SOCIS, IGNOU

SOCIS FACULTY

Prof. P. Venkata Suresh, Director, SOCIS, IGNOU
Prof. Sandeep Singh Rawat, SOCIS, IGNOU
Dr. Akshay Kumar, Associate Professor, SOCIS, IGNOU
Dr. Sudhansh Sharma, Assistant Professor, SOCIS, IGNOU
Prof. V.V. Subrahmanyam, SOCIS, IGNOU
Prof. Divakar Yadav, SOCIS, IGNOU, New Delhi
Dr. M.P. Mishra, Associate Professor, SOCIS, IGNOU

COURSE PREPARATION TEAM

This Course is adapted from MMTE-003–Digital Image Processing and Pattern Recognition of School of Sciences, IGNOU

Content Editor

Prof S. Balasundaram
School of Computers and Systems Sciences
JNU, New Delhi

Course Writer

Unit 8, Unit 9 and Unit 10
Dr. Ayesha Choudhary
School of Computers and Systems Sciences
JNU, New Delhi

COURSE COORDINATOR

Dr.Sudhansh Sharma,
Assistant Professor,
School of Computers and Information Sciences, IGNOU.

PRINT PRODUCTION

Sh Sanjay Aggarwal
Assistant Registrar, MPDD, IGNOU, New Delhi

June, 2023

©Indira Gandhi National Open University, 2023

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110068.
Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by MPDD, IGNOU.
Laser Typesetter: Tessa Media & Computers, C-206, Shaheen Bagh, Jamia Nagar, New Delhi-110025

BLOCK 3 INTRODUCTION

This Block relates to the coverage of the various topics, relevant from the point of view of computer vision, which includes the introduction to the actual meaning of computer vision, along with various camera models and the transformations involved in computer vision. The block also coves the concepts related to single camera and multiple camera environments. The unit wise distribution of content is given below:

Unit 8, includes the introduction to the actual meaning of computer vision, along with various camera models and the transformations involved in computer vision

Unit 9, includes the concepts related to single camera model environment viz. perspective projection, homography, camera calibration, and affine motion models.

Finally, Unit 10 involves the various concepts relevant to multiple camera environment like stereo vision, point correspondence, epipolar geometry and optical flow.

ignou
THE PEOPLE'S
UNIVERSITY

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 8 INTRODUCTION TO COMPUTER VISION

Structure		Page No.
8.1	Introduction	211
8.2	Objectives	211
8.3	Introduction to Computer Vision	212
8.4	Camera Models	212
8.5	Projections	214
8.6	Transformations	215
8.7	Summary	222
8.8	Solutions / Answers	223
8.8	References	223

8.1 INTRODUCTION

The purpose of this chapter is to introduce the subject of computer vision. Till now, we have been going through the concepts of image processing. In the previous units, we have learnt about digital images, and the various algorithms for processing of digital images. We would now like to introduce the subject of computer vision and its various concepts. We know that a camera is modelled on the human vision, and we require two eyes to be able to see in the three dimensional world. As we read further, we shall first discuss the camera models, and the various transformations that occur for imaging to occur.

In Section 8.2, we shall discuss the objectives of the unit.In Section 8.3, we shall introduce the subject of computer vision. We then discuss the various camera models in Section 8.4 and transformations in Section 8.5. Finally, we summarise the discussion in Section 8.6 and Section 8.7, we discuss some problems and obtain their solutions.

8.2 OBJECTIVES

The objective of this unit is to introduce the subject of computer vision. “A picture is worth a thousand words” holds true asonecan see that an image of any scene has a lot of detailed information. Further, it is easy to note thata color image has more information than a grey-scale image. In today’s world, camera technology has become very cheap and ubiquitous Cameras are the instruments through which we capture an image. The mathematics behind the camera model helps us to understand computer vision. Therefore, we need to understand the various camera models. As it has beendiscussed earlier in digital image processing, that a digital image is a matrix of non-negative integers, therefore, any transformation applied on a matrix can be applied on a digital image.

In this unit, we shall study various geometric transformations. We shall finally summarise the unit.

8.3 INTRODUCTION TO COMPUTER VISION

Computer vision is the field of study that endeavours to develop techniques to help machines understand the content of images and videos captured by single and/or multiple cameras. It seems to be a trivial problem for human beings to understand and recognize contents of images and videos once seen, however, images and videos have a lot of content and it is not easy for a machine to focus on the relevant portions of an image, understand the content, recognize familiar objects and faces and "tell the story". However, for machines to carry out intelligent tasks by "seeing" its surroundings, current computer vision and machine learning techniques need to be learnt, understood and further developed. Applications of computer vision exist in every sphere of life, from medical image understanding to autonomous vehicles and robotics. Therefore, it is an important field of study.

8.4 CAMERA MODELS

A camera is a basic tool in computer vision. It is essential to record an image of the scene around us to be able to analyse the scene. Therefore, it is important to learn the model of the camera. In general, a camera is a physical device that captures an image by allowing light to pass through a small hole (the aperture) onto a light sensitive surface. The lens in a camera focus the light entering through the aperture onto the imaging surface or the light-sensitive surface. Also, a shutter mechanism controls the amount of time for which the photosensitivesurface is exposed to light.

A camera therefore, maps the 3D world onto a 2D plane (the image). There are two major classes of camera models: camera models with a finite center and camera models with center at infinity. We shall focus our attention to the camera model with a finite center and discuss the simplest camera model: the pin-hole camera model.

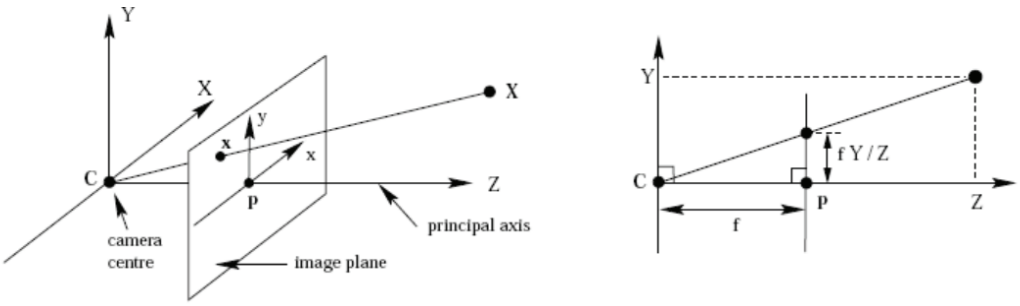


Figure 1: The pin-hole camera geometry. Figure taken from [1]

8.4.1 The Pin-Hole Camera Model

In a pin-hole camera, a barrier with a small hole (pin-hole) is placed between the 3D object and the 2D image plane (light-sensitive plane). Each point on the 3D object emits/ reflects multiple light rays out of which only one or very few of these light rays fall on the image plane by passing through the pin-hole. Thereby, one can see that there exists a mapping between the points of the 3D object and the 2D plane forming an image and such a mapping will be a projection where the pin-hole is called the centre of projection.

Formulation of the pin-hole camera model

Assume that the centreof projection is the origin of the 3D Euclidean coordinate system. Let $Z = f$be the image plane or the focal plane. Then, the centre of projection ('*C*') is called the camera centre or optical centre and *f* is the focal length. The *principal axis or the principal ray* is the line from the camera centre that is perpendicular to the image plane. The point of intersection of the principal axis and the image plane is known as the *principal point*. The *principal plane/projection plane* is the plane passing through the principal point parallel to the image plane, i.e., the image plane becomes the projection plane.

Under the pin-hole camera projection, a 3D point $\mathbf{X} = (X, Y, Z)^T$ is mapped to a 2D point $\mathbf{x} = (x, y)^T$ on the image plane where, the line joining the camera centre Cto the 3D point **X** intersects the image plane at **x**.

From Figure 1, it can be seen that by rules of similar triangles,

$$\frac{X}{Z} = \frac{x}{f} \text{and} \frac{Y}{Z} = \frac{y}{f} \Rightarrow x = \frac{fX}{Z} \text{and} y = \frac{fY}{Z}$$

This impliesthatthe point $\mathbf{X} = (X, Y, Z)^T \in \mathbf{R}^3$ is mapped to the point $\boldsymbol{x} = (fX/Z, fY/Z)^T \in \mathbf{R}^2$ on the image plane. Therefore, a pin-hole camera does a perspective projection of a 3D scene onto a 2D plane.

Note that it is not possible to write the above perspective projection transformation in matrix form. To overcome this problem, we introduce homogeneous coordinates system.

8.4.2 Homogeneous Coordinates

A coordinate system where every point having three coordinates, $(x, y, w)^T$, is called a homogeneous coordinatesystem provided $w \neq 0$ in which for points $\mathbf{x}_1 = (x_1, y_1, w_1)^T$ and $\mathbf{x}_2 = (x_2, y_2, w_2)^T$, we have $\mathbf{x}_1 = \mathbf{x}_2 \Leftrightarrow \frac{x_1}{w_1} = \frac{x_2}{w_2}$ and $\frac{y_1}{w_1} = \frac{y_2}{w_2}$ will be satisfied. Clearly, the points of a homogeneous coordinate system belong to the 2- dimensional space, $\Re^3 \setminus (0,0,0)^T$.

A two dimensional point in the Euclidean coordinate system can be represented as a point in the homogeneous coordinate system and vice-versa.

In fact, for a 2D point $(x, y)^T$ its corresponding point in the homogeneous coordinate systemistaken as $(x, y,1)^T$. Similarly, for a given a point $(x, y, w)^T$ and $w \neq 0$ in homogeneous coordinates, its corresponding 2D point is obtained as: Since

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} x/w \\ y/w \\ w/w \end{pmatrix} = \begin{pmatrix} x/w \\ y/w \\ 1 \end{pmatrix}$$

and therefore its corresponding 2D point becomes (x/w, y/w).

Likewise, as a point in the homogeneous coordinate system having four coordinates $(X, Y, Z, W)^T$ such that $W \neq 0$, the following additional condition holds: For $\mathbf{X}_1 = (X_1, Y_1, Z_1, W_1)^T$, $\mathbf{X}_2 = (X_2, Y_2, Z_2, W_2)^T$ from the homogenous coordinate systemwe have $(X_1, Y_1, Z_1, W_1)^T = (X_2, Y_2, Z_2, W_2)^T \Leftrightarrow \frac{X_1}{W_1} = \frac{X_2}{W_2}, \frac{Y_1}{W_1} = \frac{Y_2}{W_2}$ and $\frac{Z_1}{W_1} = \frac{Z_2}{W_2}$.

Clearly, these points belong to $\Re^4 \setminus (0,0,0,0)^T$.

As in the 2D case, a 3D point (X, Y, Z) will be associated with its corresponding homogeneous coordinates point (X, Y, Z,1). Conversely, any point (X, Y, Z, W) from the homogeneous coordinate system can be associated to its corresponding 3D point in the Euclidean spaceto be (X/W, Y/W, Z/W).

An important point to remember is that all scalar multiples of a homogeneous vector represent the same point.

8.5 PROJECTIONS

(i) Perspective projection

Consider the pin-hole camera model as a perspective projection transformation discussed in 8.4. Using matrix transformation in homogeneous coordinates

$$\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} fX \\ fY \\ fZ \\ Z \end{pmatrix}$$

the projection point in the Euclidean space can be obtained

$$x = f\frac{X}{Z},\ y = f\frac{Y}{Z}$$

This way the perspective projection of a 3D point via the center of projection (the pin-hole) onto a point on the image plane $Z = f$ can be obtained in matrix form.

(ii) Orthographic projection

Orthographic projection is the projection of a 3D object onto a plane by a set of parallel rays that are orthogonal tothe image plane, i.e., it is a parallel projection. In this projection, the center of projection is taken at infinity.

For any point $(X, Y, Z)^T$ from the object, its orthographic projection on the image plane $Z = f$ is given by

$x = X, y = Y$

The important properties of the orthographic projection are that parallel lines remain parallel and the size of the object does not change.

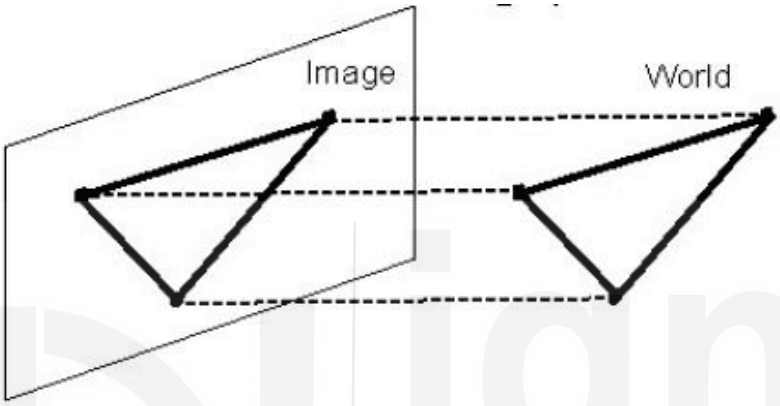


Figure 2. Orthographic projection is shown in the above image. (Image source: Internet)

(iii) Weak perspective projection

A perspective projection is a non-linear transformation, however, under certain circumstances, we can approximate it by a linear transformation, that is, as a scaled orthographic projection. The important conditions for these are that the object lies close to the optical axis and that the dimension of the object are small compared to its average distance $\overline{Z}$ from the camera.

Equations of the weak perspective projection are:

$$x = \frac{fX}{z} = \frac{fX}{\overline{z}}; y = \frac{fY}{z} = \frac{fY}{\overline{z}}$$

in which each point is scaled by $\frac{f}{\overline{z}}$

8.6 TRANSFORMATIONS

Geometric transformations play an important role in Computer Vision. In this section, we discuss the important transformations that we shall require in the future in this course.

8.6.1 Euclidean Transformations

The most important property of Euclidean transformation is that it preserves lengths and angle measures. They are the most commonly used transformations consisting of translation and rotation.

(i) Translation

When a point is moved from one location to another along straight line paths, it is known as a translation. In 2D, let $(x, y)^T$ be any point and t_x and t_y denote the translations along *x*- and y- directions respectively. Then, the new coordinates of the point $(x', y')^T$ are given by

$x' = x + t_x$ *and* $y' = y + t_y$

Or,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x + t_x \\ y + t_y \end{pmatrix}$$

In homogeneous coordinates, translation is given as a matrix vector product as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

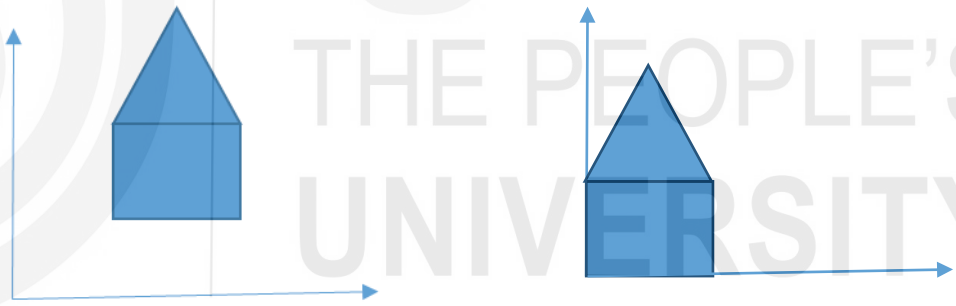


Fig (a) An object (b) Object translated to the origin

For example: If there is a rectangle with coordinates (2, 2), (2, 6), (5,2) and (5, 6) and it is to be translated to the origin. Then the translation vector will be

$$\begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix}$$

Therefore, $\begin{bmatrix} x_1' \\ y_1' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} xx_i \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix}$,

Similarly, $\begin{bmatrix} x_2' \\ y_2' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{4} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} x_3' \\ y_3' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{3} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} x_4' \\ y_4' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{3} \\ \mathbf{4} \\ \mathbf{1} \end{bmatrix}$

The graphical representation is given below:

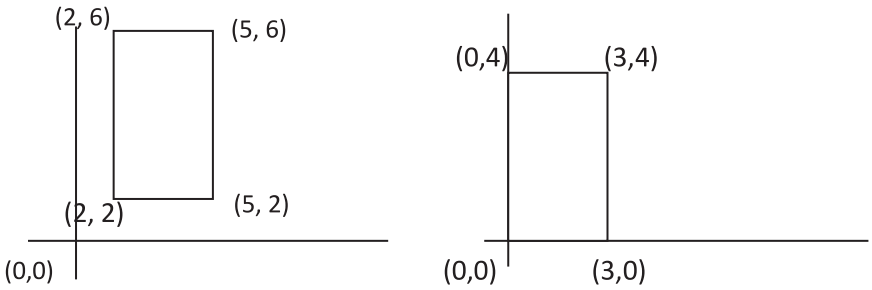


Fig: a) Square before Translation **b) Square after Translation**

(ii) Rotation

A rotation is specified by an angle θ, and the pivot-point,$(x, y)^T$ about which the object is to be rotated. A two-dimensional rotation applied to an object re-positions it along a circular path in the 2D-plane. A positive value of θ definescounter-clockwise rotation about the pivot point while a negative value of θ defines clockwise rotation. Rotation about the originis given as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

where, R = $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is known as the rotation matrix.

In homogeneous coordinates, the rotation about the origin with respect to angle θ is given as

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

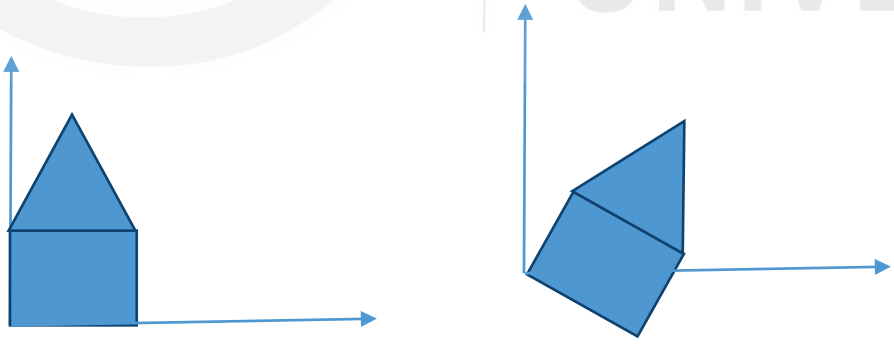


Fig. (c) Object at the origin **(d) Object rotated at the origin**

Example 2: Perform a 45^0 rotation of a triangle ABC with coordinates A: (0,0),B: (1,1),C: (5,2) about the origin.

Solution: We can represent the given triangle, in matrix form, using homogeneous coordinates of the vertices:

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 5 & 2 & 1 \end{bmatrix}$$

The matrix of rotation is: $R_\theta = R_{45^\circ} = \begin{bmatrix} \cos 45 & \sin 45 & 1 \\ -\sin 45 & \cos 45 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

So the new coordinates A'B'C' of the rotated triangle ABC can be found as:

$$\begin{bmatrix} A' \\ B' \\ C' \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix} . R_{45^\circ} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \sqrt{2} & 1 \\ 3\sqrt{2}/2 & 7\sqrt{2}/2 & 1 \end{bmatrix}$$

Thus A'=(0,0), B'=(0,√2), C'=(3√2/2,7√2/2)

The following *Figure (a)* shows the original, triangle ABC and *Figure (b)* shows the triangle ABC after the rotation.

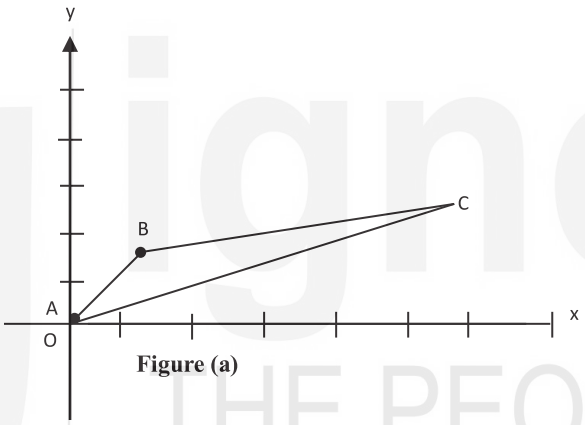


Figure (a)

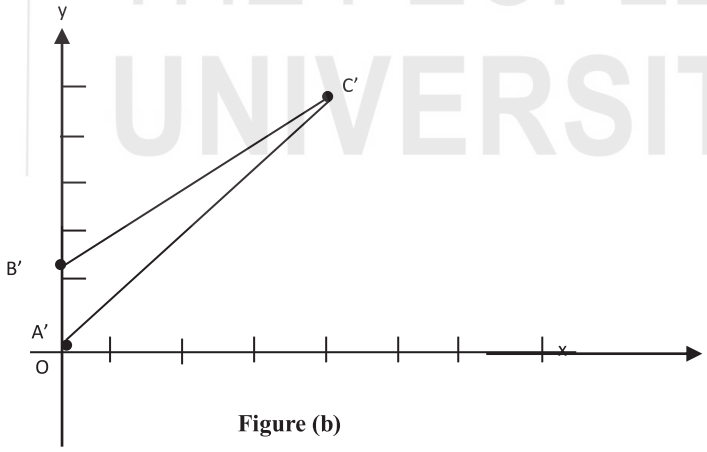


Figure (b)

A **rigid body transformation** are combinations of rotation and translations. A general rigid body transformation can be represented using the homogeneous coordinates as

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

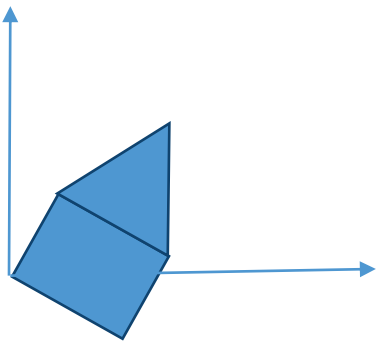


Fig (e) Object at origin after rotation

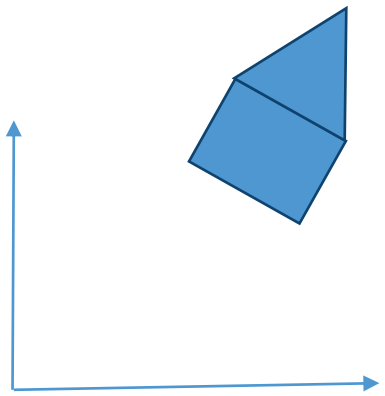


Fig (f) Object rotated and translated from initial positioning Fig (e)

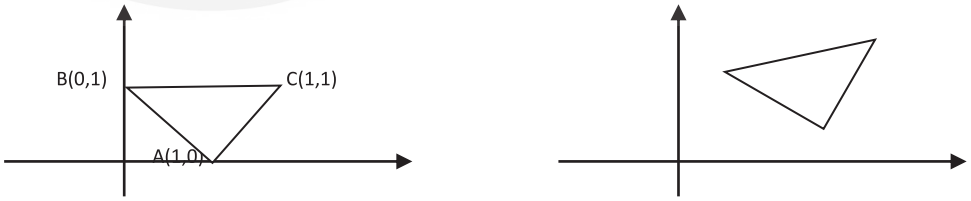
The rigid body transformations preserve angles between vectors and length of vectors therefore parallel lines remain parallel after a rigid body transformation.

Example: Consider the triangle (1,0), (0,1), (1,1). To translate it by $(t_x, t_y) = (1, 0)$, and rotation by $\theta = 30°$

$$\begin{pmatrix} x_1' & x_2' & x_3' \\ y_1' & y_2' & y_3' \\ 1 & 1 & 1 \end{pmatrix} = \begin{bmatrix} \cos 30 & -\sin 30 & 1 \\ \sin 30 & \cos 30 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} =$$

$$= \begin{bmatrix} 1+\cos 30 & 1-\sin 30 & \cos 30-\sin 30+1 \\ \sin 30 & \cos 30 & \sin 30+\cos 30 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1.866 & 0.5 & 1.366 \\ 0.5 & 0.866 & 1.366 \\ 1 & 1 & 1 \end{bmatrix}$$



8.6.2 Affine Transformations

Euclidean transformations do not change the shape of the object. Lines transform to lines, planes to planes and circles to circles. However, the lengths and angles are preserved by Euclidean transformation. Affine transformations are an extension of the Euclidean transformations which do not preserve lengths and angles. That is, under an affine transformation, circle may transform to an ellipse, however, a line will transform to a line. The important affine transformations are scaling and shear. Moreover,

translation and rotation also are affine transformations, since affine transformations are an extension of the Euclidean transformations.

(i) Scaling

By scaling, the dimensions of an object are either compressed or expanded. A scaling transformation is carried out by multiplying the coordinate values of eachvertex$(x, y)^T$by scaling factors Sxand Sy, in the x– and y- directions respectively, to produce the transformed coordinates

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Example: Consider the triangle ABC with coordinates A =(0,0), B = (1,0)and C = (1,1). If the scaling factor along X-axis, $S_x = 4$ and along Y-axis, $S_y = 1$, then calculate the new coordinates of the sheared triangle ABC.

In matrix form, the shear transformation will be

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Then,

$$A' = \begin{pmatrix} x_2' \\ y_2' \end{pmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$B' = \begin{pmatrix} x_1' \\ y_1' \end{pmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

$$C' = \begin{pmatrix} x_3' \\ y_3' \end{pmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

Therefore, the scaled triangle, $A'B'C'$ will have the coordinates, $A' = (0,0), B' = (4,0)$ *and* $C' = (4,1)$

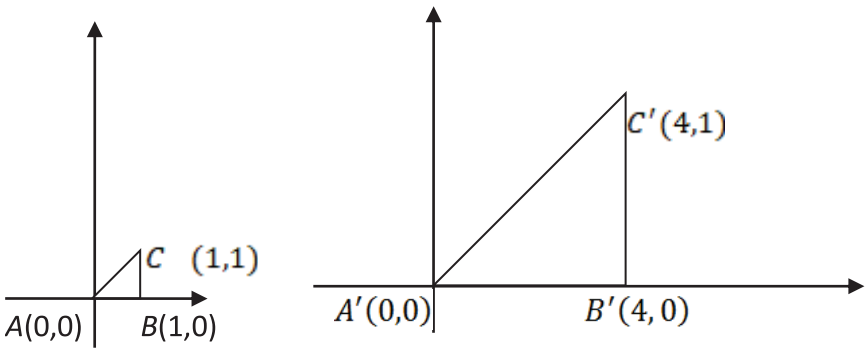


Fig (a) The original triangle ABC **Fig(b) The Scaled triangle $A'B'C'$**

(ii) Shear

A transformation that slants the shape of an object is called the shear transformation. Shearing transformation can also be carried out in both X, Y directions or only one the directions. The new coordinates after shearing in X direction are given by:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & sh_x \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and shear in Y-direction is given by:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ sh_y & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Here., sh_x and sh_y are the shearing factors along the X- and Y-directions respectively and are given as inputs.

Then, in matrix form,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & sh_x \\ sh_y & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Example:

Consider the triangle ABC with coordinates A = (0,0), B = (1,0) and C = (1,1). If the shearing factor along X-axis, $sh_x = 4$ and along Y-axis, $sh_y = 1$, then calculate the new coordinates of the sheared triangle ABC.

In matrix form, the shear transformation will be

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Then,

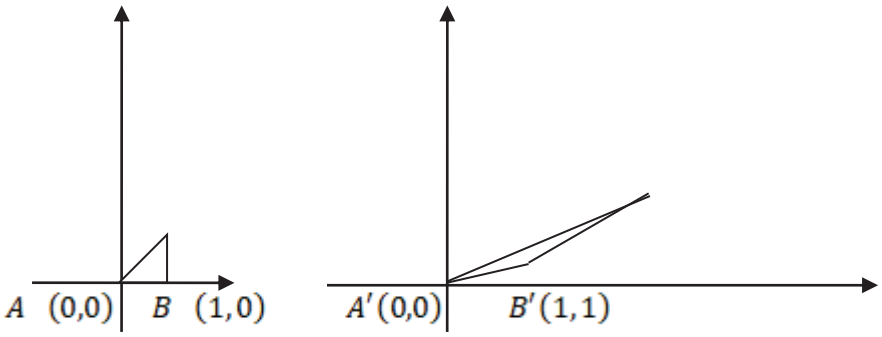
$$A' = \begin{pmatrix} x_2' \\ y_2' \end{pmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$B' = \begin{pmatrix} x_1' \\ y_1' \end{pmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$C' = \begin{pmatrix} x_3' \\ y_3' \end{pmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

Therefore, the sheared triangle, $A'B'C'$ will have the coordinates $A' = (0,0), B' = (1,1), and\ C' = (5,2)$

-



General Affine Transformation

Using the homogeneous coordinates, a general affine transformation is defined of the form:

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{pmatrix}$$

where the sub-matrix $\begin{pmatrix} a & b \\ d & e \end{pmatrix}$ need not be a rotation matrix. An affine transformation preserves parallelism but does not preserve angles.

8.6.3 Projective Transformations

Projective transformations are the most 'general' forms of linear transformations. It does not preserve angles, distances, therefore, parallelism also. Therefore, under the projective transformation, parallel lines do not remain parallel. The only thing that it preserves is collinearity of points and therefore straight lines remain straight.

A projective transformation in the most generic form in homogeneous coordinate system is given as:

$$P = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix}$$

8.7 SUMMARY

In this unit, we have studied an introduction to computer vision, the basic pin-hole camera model. Homogeneous coordinates were introduced, which helps in defining the projection transformations in terms of matrices. We have discussed perspective projection, orthographic projection and weak perspective projection. Geometric transformations are important for computer vision and we have discussed Euclidean, Affine and Projective transformations and their representation in homogeneous coordinates.

8.8 QUESTIONS AND SOLUTIONS

Q1. Which projection does the pin-hole camera represent?

Ans. 1 Pin-hole camera represents the perspective projection, such that it represents a mapping between the points on the 3D object and its image formed on the 2D plane

Q 2. What are the rigid body transformations?

Ans.2 The rigid body transformations are translation and rotation. Translation in movement along a line, while rotation moves a point by an angle around a pivot point.

Q 3. What are the three classes of transformations?

Ans 3. The three classes of transformations are: Projective, Affine and Euclidean transformations. The projective transformation preserves only collinearity, affine transformation preserves parallelism, while Euclidean transformation preserves lengths and angles.

8.9 REFERENCES

[1] Richard Hartley and Andrew Zisserman "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press, April 2004.

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 9 SINGLE CAMERA

Structure **Page No.**

9.1 Introduction 224
9.2 Objectives 224
9.3 Camera Models 224
9.4 Perspective Projection 226
9.5 Homography 229
9.6 Camera Calibration 231
9.7 Affine Motion Models 234
9.8 Summary 235
9.9 Solutions / Answers 235

9.1 INTRODUCTION

In this unit, we shall study the various aspects of computer vision related to single camera. This is important since in many applications, we may have data from only a single camera to work with. In the previous unit, we discussed about the pin-hole camera model, the perspective projection and homogeneous coordinates. We saw how homogeneous coordinates helped in representing the perspective projection as a matrix. In this unit, we shall discuss the pin-hole camera model and perspective projection in more detail and also discuss the camera matrix. We shall then discuss the camera calibration process for a single camera, and the affine motion models. Finally, we shall summarize this unit in Section 9.8.

9.2 OBJECTIVES

The objectives of this unit are as follows:

- To learn about the camera model in detail,
- To understand the concept of camera matrix,
- To understand the process of camera calibration
- To understand the affine motion model.
- To give an overview of the aspects of computer vision related to a single camera and the estimation of 3D parameters from a single camera.

9.3 CAMERA MODELS

As discussed in Unit- 8, a pin-hole camera does a perspective projection of a 3D scene onto a 2D plane.

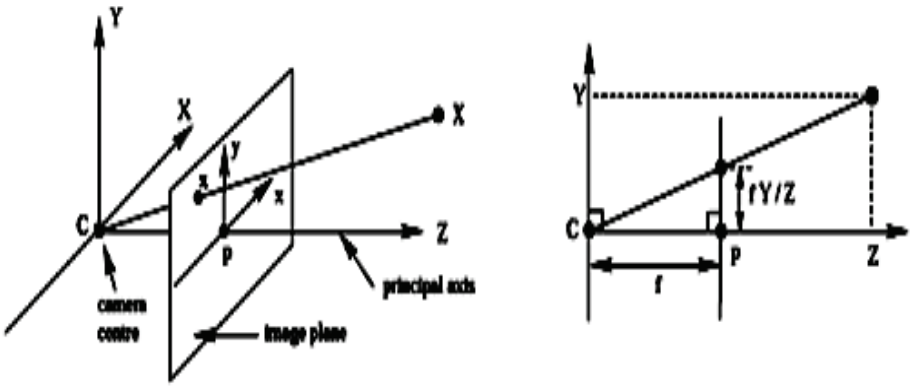


Figure-1: The pin-hole camera model [1]

Under the pin-hole camera projection, a 3D point **X** = (X, Y, Z) is mapped(∼) to a 2D point **x** = *(x, y)*on the image plane where, the line joining the camera centre*C* to the 3D point **X** intersects the image plane at **x**. Using the homogeneous coordinates, we represent **X** = (X, Y, Z, 1) and **x** = *(x, y,1)* and therefore, the pin-hole camera model can be represented as

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (1)$$

In this simplified version, we have assumed that the origin of the world coordinate system is at the centre of projection (pin-hole) and that *f* is the focal length of the camera. The focal length is the distance between the camera centre (pin-hole/ centre of projection) to the image plane. To assume that the image is not inverted, it is assumed that the image plane is in front of the pin-hole, although physically the image plane is present behind the pin-hole.

Therefore, we see that in the pin-hole camera model a perspective projection occurs. The image coordinates are related to the world coordinates as given by Equation 1under this perspective projection where the camera centre is at the origin of the world coordinate system. However, in real-life it is not always possible to keep the camera (pin-hole/ centre of lens) at the origin of the world coordinate system. To bring the object into the camera's view, we may need to move the camera away from the origin. In Unit-8, we studied about transformations. In real-world, we shall have to carry out a sequence of translations and rotations of the camera to bring the object of interest in its view. In the next section, we shall discuss two important concepts: (a) intrinsic parameters of a camera and, (b) extrinsic parameters of a camera.

Before we discuss these parameters, we have to understand that both the pixel coordinate system and the world coordinate systems are related by the following physical parameters: (a) size of pixels (b) position of principal

point (c) focal length of the lens and (d) position and orientation of the camera.

The internal or intrinsic parameters of the camera define the relation between the pixel coordinates of a point on the image with the corresponding camera coordinates that exist in the camera reference frame.

The external/ extrinsic parameters of the camera are the parameters that define the location and orientation of the camera coordinate frame with respect to a known world coordinate frame.

Questions to check your progress:

1. What is a pinhole camera?
2. What does f, the focal length represent in a pin-hole camera model?

9.4 PERSPECTIVE PROJECTION

A camera projects the 3D world onto a 2D image plane.It is a perspective projection and is represented by a 3x4 matrix such that the left 3x3 submatrix is non-singular. We shall now see how the camera matrix encodes the camera parameters.

9.4.1 External/Extrinsic Parameters of a Camera

The external camera parameters define the relation between the known world coordinate frame and the unknown camera coordinate frame. The two coordinate systems are related by a rotation and translation. Therefore, determining the external parameters of the camera implies:

(a) finding the translation vector between the relative positions of the origins of the camera coordinate frame and the world coordinate frame.

(b) Finding the elements of the rotation matrix to align the corresponding axes of the two coordinate frames.

Therefore, the extrinsic camera parameters help us in finding the relation between the coordinates of a 3D point in the world coordinate system with its coordinates in the camera coordinate system.
Let R be the rotation matrix,

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

and T be the translation vector,

$$T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

Then, for a 3D point A, whose world coordinates be $A_w = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}$ and camera coordinates are $A_c = \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$,

$$A_c = R(A_w - T)$$

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_w - T_x \\ Y_w - T_y \\ Z_w - T_z \end{bmatrix} \quad -(2)$$

$$X_C = R_1^T(A_w - T)$$
$$Y_C = R_2^T(A_w - T)$$
$$Z_C = R_3^T(A_w - T)$$

where R_i^T corresponds to the *i*-th row of the rotation matrix

Equation (2) gives the relation between the coordinates of a 3D point A, in the camera coordinate system and the world coordinate system using the extrinsic camera parameters.

Question for review of progress:

1. Define and describe the external parameters of a camera
2. What does the rotation matrix in the camera external parameters represent?
3. What does the translation vector in the camera external parameters represent?

9.4.2 Internal/Intrinsic Parameters of a Camera

Equation 1 assumes that the coordinates in the image plane are measured with the principal point as the origin. As we discussed in Unit-8, the *principal axis or the principal ray* is the line from the camera centre that is perpendicular to the image plane. Therefore, the *principal point* is the point of intersection of the principal axis and the image plane. In general, the pin-hole camera model is a geometrical model, implying that all points and focal length are measured in millimetres or centimetres or meter. However, the camera coordinates are measured in pixel distance or height and width of pixels (sampling distance).

Therefore, the internal camera parameters characterize the following:

a) The perspective projection (focal length)
b) The relation between pixel size and image plane coordinates
c) The geometric distortions introduced by the optics

The relation between the camera coordinates and the image plane coordinates is given by the perspective projection:

$$x = f\frac{X_c}{Z_c} = f\frac{R_1^T(A_w - T)}{R_3^T(A_w - T)}; y = f\frac{Y_c}{Z_c} = f\frac{R_2^T(A_w - T)}{R_3^T(A_w - T)} \quad (3)$$

9.4.2.1 Relation between image plane coordinates and pixels

Let u_x and u_y be the coordinates of the principal point in pixels and s_x, s_y are the sizes of pixels in the horizontal and vertical directions in millimeters. Therefore,

$x = -(x_{im} - u_x)s_x \Rightarrow x_{im} = -x/s_x + u_x$;

$y = -(y_{im} - u_y)s_y \Rightarrow y_{im} = -y/s_y + u_{y\text{-}(4)}$ - (4)

where, (x,y) are the pixel coordinates of the image of the 3D points. Therefore,

$$\begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix} = \begin{bmatrix} -1/s_x & 0 & u_x \\ 0 & -1/s_y & u_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad - (5)$$

Hence, we can relate the pixel coordinates with the world coordinates from Equations (3) and (4) as

$$x_{im} = \left(-f/s_x\right)\left(\frac{R_1^T(A_w - T)}{R_3^T(A_w - T)}\right) + u_x$$

$$y_{im} = \left(-f/s_y\right)\left(\frac{R_2^T(A_w - T)}{R_3^T(A_w - T)}\right) + u_y$$

Therefore, the matrix of intrinsic parameters, also known as the camera calibration matrix is

$$K = \begin{bmatrix} -f/s_x & 0 & u_x \\ 0 & -f/s_y & u_y \\ 0 & 0 & 1 \end{bmatrix}$$

And the matrix representing the external or extrinsic parameters of the camera is

$$[R|T] = \left[\begin{array}{ccc|c} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{array}\right]$$

In homogeneous coordinates,

$$\begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix} = \begin{bmatrix} -f/s_x & 0 & u_x \\ 0 & -f/s_y & u_y \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{array}\right] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad -(6)$$

P = K[R|T] is called the camera projection matrix, which is a 3x4 matrix, where K represents the intrinsic parameters and [R|T] represent the extrinsic parameters of the camera.

Questions for review of progress:

1. What do the internal parameters of a camera represent?
2. What role do the coordinates of the principal point play?

9.5 HOMOGRAPHY

A planar homography is a transformation that relates two planes. Therefore, a homography relates two images of the same scene, i.e. homography maps an image from one view to another. Given points in one image and the homography matrix, H, we can find the corresponding points in the other image.

A transformation or a mapping $h: P^2 \rightarrow P^2$ is a homography if and only if there exist a non-singular 3x3 matrix H such that for any point in $x \in P^2$, then

$$x' = h(x) = Hx$$

An important condition for a transformation to be a homography is that it maps collinear points to collinear points, that is, if x_1, x_2 and x_3 lie on a line then, $h(x_1)$, $h(x_2)$ and $h(x_3)$ also lie on a line.

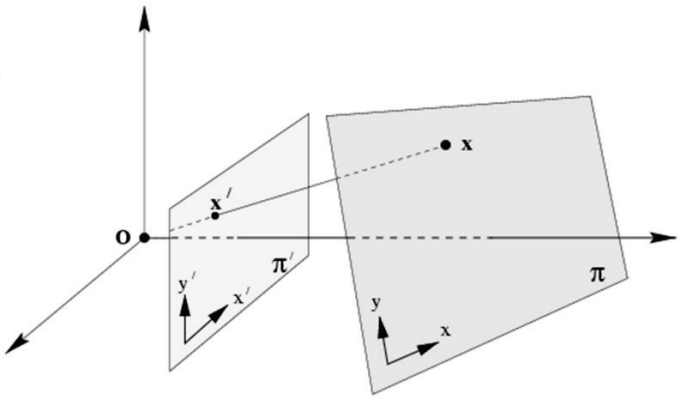


Fig. 2. The projection maps points on one plane to points on another plane. These points are related by a homography, H, such that $x' = Hx$. (Fig. taken from [1])

9.5.1 Homography Estimation: Direct Linear Transformation Algorithm

We can estimate the homography between two images of the same scene under the assumption that the point correspondences are given and that there is planar motion of the camera.

We assume that $x_i' \overset{i}{\leftrightarrow} x_i$ are the given set of point correspondences. We need to compute the 3x3 homography matrix H.

For each point correspondence, we have,

$$\begin{bmatrix} x_i' \\ y_i' \\ w_i' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix}$$

Therefore, one pair of correspondences will give two independent equations. There are 8 unknowns, since H has 9 entries but is defined up to scale. In this case, we can divide each entry of H by h_{33}, and fix the last entry as 1.

Therefore, H has 8 degrees of freedom. Each point correspondence gives 2 independent equations, therefore, to find the 8 unknowns we shall need at least 4 point correspondences.

However, in practice, the point correspondences may not be free of noise, therefore a larger number of point correspondences are used to estimate the best solution.

Direct Linear Transformation Algorithm

Each point correspondence gives rise to two independent equations.

$$\frac{x_i'}{w_i'} = \frac{x_i h_{11} + y_i h_{12} + w_i h_{13}}{x_i h_{31} + y_i h_{32} + w_i h_{33}}$$

$$\frac{y_i'}{w_i'} = \frac{x_i h_{21} + y_i h_{22} + w_i h_{23}}{x_i h_{31} + y_i h_{32} + w_i h_{33}}$$

Therefore, if we simplify, we get the equation $A_i\, h = 0$ where,

$$A_i = \begin{bmatrix} w_i' x_i & w_i' y_i & w_i' w_i & 0 & 0 & 0 & -x_i' x_i & -x_i' y_i & -x_i' w_i \\ 0 & 0 & 0 & w_i' x_i & w_i' y_i & w_i' w_i & -y_i' x_i & -y_i' y_i & -y_i' w_i \end{bmatrix}$$

$$h = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix}$$

And by stacking all the equations of this form, we will get the system of equation **Ah = 0.** To obtain the solution of the system of equations Ah = 0, we obtain the Singular Value Decomposition (SVD)[2] of **A**. If SVD(A) = UDV^T where, D is the diagonal matrix with singular values arranged in descending order, then the solution, **h** is the last column of V, since for the system of equation **Ah=0**, the solution **h** is the smallest singular vector which corresponds to the smallest singular value.

Planar homography finds a wide range of applications such as in removing perspective distortion, creating panoramas, 3D reconstruction, etc.

For example:

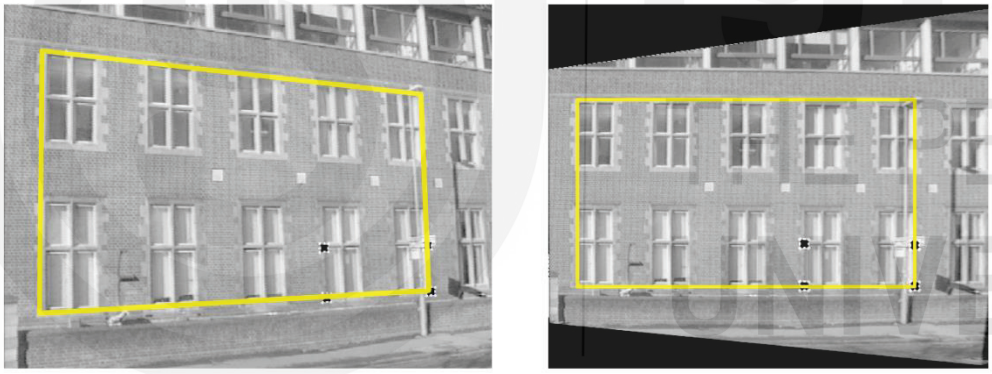


Fig. 3. The figure shows that 4 point correspondences suffice to remove the perspective distortion from a planar building façade . Fig. taken from [1]

Questions for review of progress:

1. Planar homography helps in relating the points on one plane with the corresponding points on another plane. True or False.
2. Write out the complete equations to find the system Ah = 0.

9.6 CAMERA CALIBRATION

The aim of camera calibration is to estimate the intrinsic and extrinsic parameters of a camera. To compute the intrinsic and extrinsic camera parameters, we need to know a set of correspondences between the world points (X, Y, Z) and their projections on the image (x,y). Therefore, the first step is to establish the set of correspondences between the world points and

their projections on the image plane. To do so, in general, images of a known calibration object are used. The calibration object has a known 3D geometry and location in space. Moreover, it generates image points which can be accurately located.

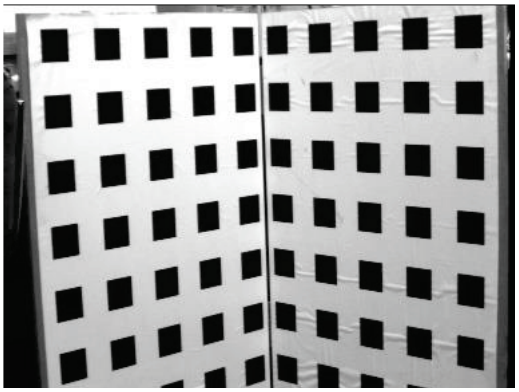


Fig 4. Image of a calibration object. (Image taken from [1])

If we consider the object in Fig. 4, then it can be seen that there are equal spaced black squares on a white background, such that the corner points are clearly visible and can be extracted easily. Also, if we assume a point on this object to be the origin of the world coordinate system, and define the 3 axes, then we can find the coordinate of each corner point in 3D. Moreover, the projection of these corner points can be found in the image and therefore, we can establish a set of correspondences between 3D points and their corresponding image points by using a known calibration object.

We discuss the linear solution to compute the estimate of the camera matrix P given n> 6, world to image corresponding points $\{X_i \leftrightarrow x_i\}$.Gold Standard algorithm is the algorithm for estimating the camera matrix from accurate world to image corresponding points.

Step 1: **Normalization**. In this step, we compute a similarity transformation T to normalise the 2D image points. To carry out normalization, the points should be translated such that their centroid should be at the origin and scaled in such a manner that the root mean squared distance from the origin is √2.Similarly, compute a similarity transformation U to normalise the 3D world points. The 3D points should also be normalised such that centroid of the points is translated to the origin and root mean squared distance from the origin is √3. (This works well for the case when the variation in depth of the points is less such as in a calibration object).

Step 2: **Direct Linear Transform (DLT):**
Assuming that the camera matrix is represented by

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13}m_{14} \\ m_{21} & m_{22} & m_{23}m_{24} \\ m_{31} & m_{32} & m_{33}m_{34} \end{bmatrix}$$

For each corresponding point, we have

$x_i = M\ X_i$

Therefore,

$$\begin{bmatrix} x_h \\ y_h \\ w \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$x = \frac{x_h}{w} = \frac{m_{11}X_w + m_{12}Y_w + m_{13}Z_w + m_{14}}{m_{31}X_w + m_{32}Y_w + m_{33}Z_w + m_{34}}$$

$$y = \frac{y_h}{w} = \frac{m_{21}X_w + m_{22}Y_w + m_{23}Z_w + m_{24}}{m_{31}X_w + m_{32}Y_w + m_{33}Z_w + m_{34}}$$

Therefore, each point correspondence gives us two independent equations. Therefore,

$$A_w = \begin{bmatrix} wX_w & wY_w & wZ_w & w & 0 & 0 & 0 & 0 & -x_hX_w & -x_hY_w & -x_hZ_w & -x_h \\ 0 & 0 & 0 & 0 & wX_w & wY_w & wZ_w & w & -y_hX_w & -y_hY_w & -x_hZ_w & -y_h \end{bmatrix}$$

and,

$$\text{m} = \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{bmatrix}$$

We can write by stacking A_w for each corresponding point, we generate the 2nx12 matrix **A**, such that **Am = 0,** where **m** is the vector that contains the entries of the matrix M. The solution of **Am = 0**, subject to $\|\boldsymbol{m}\| = \mathbf{1}$ is then obtained from the unit singular vector of A corresponding to the smallest singular value of A. We use Singular Value Decomposition (SVD) to find the solution as in the case of homography computation. This gives us the linear solution for M, which is used as an initial estimate of M.

Step 3: The measurement errors need to be reduced. Therefore, we minimize the geometric error $\sum_i d(x_i, MX_i)^2$ over M using the linear estimate as the starting point and an iterative algorithm such as Levenberg-Marquardt.

Step 4: De-normalization: Finally, the camera matrix for the original (unnormalized) coordinates is obtained as

$$M' = T^{-1}MU$$

Therefore, M' is the camera matrix. Using QR-decomposition[2], the camera matrix can be decomposed as $M' = M_{int}M_{ext}$, where, M_{int} consists of the internal parameters of the camera and is an upper-triangular matrix and M_{ext} is the matrix of external parameters of the camera.

Questions for review:

1. What are the minimum number of point correspondences between 3D points and 2D points required for camera calibration?
2. Write down the equations needed to set up the system Am = 0.

9.7 AFFINE MOTION MODEL

We discussed the Affine transformation in Unit 8. To recall, an affine transformation is a geometric transformation that preserves lines and parallelism.

In general, an affine projection is a combination of a linear mapping + translation in homogeneous coordinates. That is,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} = AX + \boldsymbol{b}$$

where, $\boldsymbol{b}$ denotes the projection of the world origin.

We now define an affine camera.

9.7.1 Affine Camera

An affine camera is the camera whose projection matrix M has the last row of the form (0,0,0,1). An important property of the affine camera is that it maps the points at infinity to points at infinity. Therefore, an affine camera is a camera at infinity, implying that the camera center lies on the plane at infinity. The affine camera preserves parallelism.

As we calibrate a projective camera, similarly we can estimate an affine camera also. An affine camera matrix is a 3x4 projection matrix with the last row given as (0,0,0,1).

In general, the affine motion model is used for approximating the flow pattern of the camera motion in a video.

9.8 SUMMARY

In this unit, we have discussed the single camera geometry, including the camera model, perspective projection, the camera parameters, homography, camera calibration and the affine motion model. The camera parameters are of two types, the internal camera parameters and the external camera parameters which together form the camera matrix. We also discussed how image to image homography can be computed with the knowledge of sufficient number of corresponding points. We also discussed that if sufficient number of world to image corresponding points are known then the camera can be calibrated and the projection matrix can be estimated. This can be done easily with a calibration object. We also discussed the affine transformations, the affine camera, and the affine motion model.

9.9 REFERENCES

[1] Richard Hartley and Andrew Zisserman "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press, 2004.

[2] Gene H. Golub and Charles F. Van Loan, "Matrix Computations", Third Edition, The John Hopkins University Press, 1996.

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 10 MULTIPLE CAMERA

Structure	Page No.
10.1 Introduction	236
10.2 Objectives	236
10.3 Stereo Vision	237
10.4 Point correspondences	237
10.5 Epipolar Geometry	238
10.6 Motion: Optical Flow	240
10.7 Summary	242
10.8 Solutions / Answers	242

10.1 INTRODUCTION

In the previous unit, we learnt about the camera models and how an image is created. We also learnt that there is loss of depth when a picture is taken by a single camera. In this unit, we shall see how we can recover the depth of a scene when there are multiple cameras. In this course, we shall study about the case when there are two cameras, that is also known as Stereo Vision. There are more limitations of a single camera system. Apart from loss of depth information, a camera is a sensor that has a fixed view volume. This implies that there is a range of scene that a single camera can see. Therefore, if the object of interest moves out of the view volume, it can no longer be observed by the camera. This may cause issues in applications like security and surveillance in wide areas. Moreover, as we saw in the study of the camera model, we can only see what lies between the camera lens and the image plane, implying that if the object of interest lies behind another object, then occlusion occurs and we are not able to view the object of interest. However, if there was another camera that could view the object of interest from another view, then we will be able to observe the object of interest despite occlusion from one view. Therefore, multiple camera systems have many advantages over single camera systems. However, this also leads to questions such as how many cameras are enough, where should the cameras be placed and how much overlap should the cameras have between their views. The answers to these questions are dependent on factors such as the cost of equipment, the type of cameras used and the application of the camera system.

10.2 OBJECTIVES

The objectives of this unit are to :

- Learn about the stereo vision system.
- Discuss the concept of point correspondences and epipolar geometry.
- Discuss the concept of motion and optical flow that allows a computer vision system to find the movement pattern in a video.

10.3 STEREO VISION

A stereo vision system consists of two cameras such that both the cameras can capture the same scene, however with some disparity between the views. One stereo vision system that we can easily relate to are the two eyes that we have.

The process of using two images of the scene captured by a stereo camera system to extract 3D information is known as stereo vision. Since the stereo pair, or the images taken by the stereo system, enable us to get the 3D information of the scene, therefore, it finds wide application in autonomous navigation of robots, autonomous cars, virtual and augmented reality, etc.

In stereo vision, the 3D information is obtained by estimating the relative depth of the points in the scene. The corresponding points in the stereo pair are matched and a disparity map is created which helps in estimating the depth of the points. We shall first understand the concept of corresponding points.

10.4 POINT CORRESPONDENCES

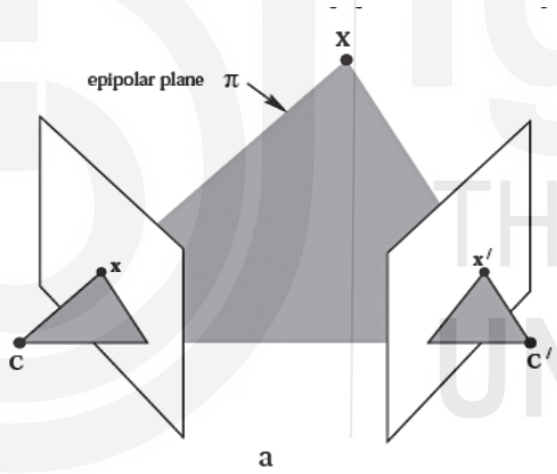


Fig10.1 The figure shows the concept of a stereo pair. C and C' are the two camera centers and x and x' are the images of the 3D point X in the corresponding image planes. x and x' are said to be corresponding points. Fig taken from [1]

As shown in Figure 10.1, the stereo pair consist of two cameras that are at a certain distance apart. The line joining the camera centers is known as the baseline. The view volume of the two cameras is such that they view a common area in the scene. A 3D point that lies in the common view volume of the two cameras will be imaged by both the cameras. Therefore, as shown in Figure 10.1, the 3D point X is visible to both the cameras and therefore, it has an image x in Camera 1 (with camera center C) and image x' in Camera 2 (with camera center C'). Therefore, x and x' are called corresponding points.

Given a point in one image, we can find it corresponding point in the other image because of epipolar geometry which we shall study next.

10.5 EPIPOLAR GEOMETRY

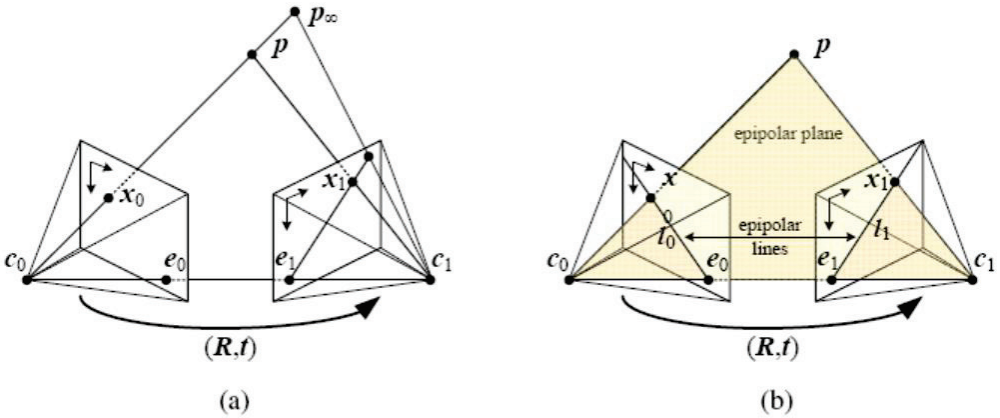


Figure 10.2. (a) ,(b) Epipolar geometry (Image taken from [2])

Epipolar geometry describes the geometric relation between two views of the same scene captured by a stereo vision system. In Figure 10.2 (a), it can be seen that the two cameras (C_0 and C_1) are related by a rotation (R) and translation (t), implying that the relative pose of the cameras are an inherent part of epipolar geometry. The 3D point $p$is imaged in both the image planes I_0 and I_1. In I_0, the image of p is denoted by x_0 and in I_1, the image of p is denoted by x_1. It can also be seen that the two camera centres, C_0 and C_1 and the 3D point p are coplanar. This plane Π is known as the Epipolar plane. Moreover, the ray joining C_0 and $x_0$extends towards infinity p_∞ and is intersected by the ray joining C_1 and x_1 at the 3D point p. Therefore, we can see that in the case of stereo vision, it is possible to find the 3D point if it is imaged by both the cameras and the point correspondences are known.

The line joining the two camera centres C_0 and C_1 is known as the **baseline**. The baseline intersects the two image planes at e_0 and e_1 respectively. The point of intersection of the baseline with an image plane is known as an **epipole**. Therefore, e_0 is the left epipole and e_1 is the right epipole. The left epipole e_0 is the image of the camera center C_0 in the left image plane I_0 while, the right epipole e_1 is the image of the camera centreC_1 in the right image planeI_1. The epipolar line is the intersection of the epipolar plane with the image plane. An important point to be noted is that every epipolar line passes through the epipole, therefore, the point of intersection of all the epipolar lines is the epipole. Another point to be noted is that the epipolar

line $\boldsymbol{l_1}$ in I_1 is the image of the back projected ray joining the camera centre $\boldsymbol{C_0}$ with the 3D point $\boldsymbol{p}$, while the epipolar lime $\boldsymbol{l_0}$ in I_0 is the image of the back projected ray, that is the ray from the camera centre $\boldsymbol{C_1}$ with the 3D point $\boldsymbol{p}$.

Therefore, given $\boldsymbol{x_0}$, the corresponding point $\boldsymbol{x_1}$ in the second image is constrained to lie on the corresponding epipolar line. The Fundamental matrix, $\boldsymbol{F}$, represents the epipolar geometry algebraically. An important point to be noted is that the fundamental matrix defines the relation between corresponding points as given by Equation 10.1.

$$\boldsymbol{x_1^T F\, x_0 = 0} \qquad (10.1)$$

More precisely, the fundamental matrix $\boldsymbol{F}$ maps a given point $\boldsymbol{x_0}$ from the first image to the epipolar line $\boldsymbol{l_1}$ in the second image, $\boldsymbol{l_1 = F x_0}$ and since the corresponding point $\boldsymbol{x_1}$ of $\boldsymbol{x_0}$ in the second image lies on $\boldsymbol{l_1}$ implies Equation 10.1 is satisfied.

Since $\boldsymbol{x_1^T l_1 = 0}$, **and** $\boldsymbol{l_1 = F x_0}$ therefore, $\boldsymbol{x_1^T F\, x_0 = 0}$.

$\boldsymbol{F}$ is a matrix of size 3x3 with rank 2 and therefore, if we have at least 8 correspondences, then $\boldsymbol{F}$ can be computed from the point correspondences.

An important property of the fundamental matrix is that $\boldsymbol{l_1 = F x_0}$ and $\boldsymbol{l_0 = F^T x_1}$. Moreover, since the epipole lies on the epipolar lines, therefore, $\boldsymbol{F e_0 = 0}$ and $\boldsymbol{e_1^T F = 0}$.

Solving for the Fundamental Matrix

Given point correspondences, x_i and x_i' we can setup a system of equations using Equation (1) to solve for F.

$x_i'^T F\, x_i = 0$ implies

$$[x_i' \quad y_i' \quad 1] \begin{pmatrix} f_{10} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \qquad (10.2)$$

which gives Equation 10.3 on solving:

$$x_i' f_{10} x_i + x_i' f_{12} y_i + x_i' f_{13} + y_i' f_{21} x_i + y_i' f_{22} y_i + y_i' f_{23} + f_{31} x_i + f_{32} y_i + f_{33} = 0 \qquad (10.3)$$

If we consider *m*such correspondences, then we can solve the system of equation for the 9 unknowns in F.

$$\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx'_m & x_my'_m & x_m & y_mx'_m & y_my'_m & y_m & x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0 \quad (10.4)$$

This system requires atleast 8 points to solve since F has 8 degrees of freedom. Therefore, this is also known as the 8-point algorithm.

We use Singular Value Decomposition (SVD) to solve the system of equations in Equation 10.4.If the cameras are calibrated and the matrices $K$and$K'$of internal parameters are known then, we can use the fundamental matrix, F, to compute the essential matrix, E, given by Equation 10.5.

$$E = K'FK \quad (10.5)$$

The Essential matrix, E, is used to determine the camera pose i.e. the positioning and alignment of camera.

10.6 MOTION

In our daily lives, we perceive, understand and predict motion rather easily. Motion perception is very strong in the human vision system. Motion or perception of motion during the imaging process can be caused by some of the following reasons:

1. The camera is static but the object is moving.
2. The camera is moving but the object is static.
3. Both the camera and objects are moving
4. The camera is static but light source and objects are moving.

Motion perception plays an important role in applications of computer vision such as activity recognition, surveillance, 3D reconstruction, and many others. Therefore, it is important to estimate the motion from images.

10.6.1 Optical Flow

Estimating motion of the pixels in a sequence of images or videos has a very large number of applications. Optical flow is used to compute where the motion information between sequence of images. It helps to determine a dense point to point correspondence between pixels of an image at time t

with the pixels in the image at time t+1. Therefore, optical flow is said to compute the motion in a video or sequence of images.

Optical flow is based on the assumption that across consecutive frames, the pixel intensities do not rapidly change. It also assumes that neighbouring pixels have similar pattern of motion. Most often, it also assumes that the luminance remains constant. We assume that *f(x,y,t)* is a pixel in the image taken at time *t,* moves to a point *(x+δx, y+δy)* at the time *t+δt,* that is, to *f(x+δx, y+δy, t+δt)*. Since they are the same point, and therefore, the above assumption can be written mathematically as

$$f(x+\delta x, y+\delta y, t+\delta t) = f(x,y,t) \qquad (10.6)$$

Equation (10.6) forms the basis of the 2D motion constraint equation and holds true given that *δx, δy, δt* are small. Taking into consideration the first order Taylor series expansion about $f(x,y,t)$ in Equation (10.6), we get

$$f(x+\delta x, y+\delta y, t+\delta t) = f(x,y,t) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial t}\delta t + (\textit{Higher order terms}) - (10.7)$$

Assuming that the higher order terms are very small, we ignore them. Then, from Equations (3) and (4), we get:

$$\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial t}\delta t = 0$$

or, $$\frac{\partial f}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial f}{\partial t} = 0$$

or, $$\frac{\partial f}{\partial x}v_x + \frac{\partial f}{\partial y}v_y + \frac{\partial f}{\partial t} = 0 \qquad (10.8)$$

where, $v_x = \frac{\delta x}{\delta t}$, $v_y = \frac{\delta y}{\delta t}$ are the image velocities or optical flow at *(x,y)* at time *t.*

$\frac{\partial f}{\partial x} = f_x; \frac{\partial f}{\partial y} = f_y; \frac{\partial f}{\partial t} = f_t$ are the image intensity derivatives at *(x, y)*.

Then, from Equation (10.8), we get

$$(f_x, f_y).(v_x, v_y) = -f_t \qquad (10.9)$$

Equation (10.9) is called the 2D Motion Constraint Equation.

There are various methods for solving for optical flow. The Lucas and Kanade optical flow algorithm [3] and Horn and Shunck [4]optical flow methods are two of the most popular methods for estimating the optical flow.

10.7 SUMMARY

In this unit we learned about various concepts related to Multiple camera models required for computer vision. The concept of stereo vision was discussed and the concept was extended to point correspondence and Epipolar geometry; finally the unit was completed with the motion oriented concepts which includes optical flow.

10.8 QUESTIONS AND SOLUTIONS

Q1. What is Stereo Vision ?
Sol. Ref. 10.3

Q2. Describe the concept of Point Correspondance.
Sol. Ref. 10.4

Q3. Discuss the term Epipolar Geometry.
Sol. Ref. 10.5

Q4. Explain Optical flow, in context of motion perception in computer vision.
Sol. Ref. 10.6

10.9 REFERENCES

[1] Hartley and Zisserman "Multiple View Geometry, Oxford Press
[2] Computer Vision: Algorithms and Applications by Richard Szeliski
[3] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", DARPA Image Understanding Workshop, 1981, pp 121-130. (also IJCAI'81, pp674-679)
[4] B. K. P. Horn and B. G. Shunck "Determining Optical Flow", Artificial Intelligence 17, 1981, pp. 185- 204.