

# Uncalibrated Stereo

Shree K. Nayar

Monograph: FPCV-4-2

Module: Reconstruction II

Series: First Principles of Computer Vision

Computer Science, Columbia University

April 2025

[FPCV Channel](#)

[FPCV Website](#)

Consider the following scenario. You and your friend are standing in front of a monument, and each of you takes a photo of it. You have different cameras and you have no knowledge of how your cameras were positioned and oriented with respect to each other. It turns out that, if you know the internal parameters of the two cameras, you can compute the translation and rotation of one camera with respect to the other. Once that is done, you have a calibrated stereo system and hence can compute the three-dimensional structure of the monument. This approach is called uncalibrated stereo.

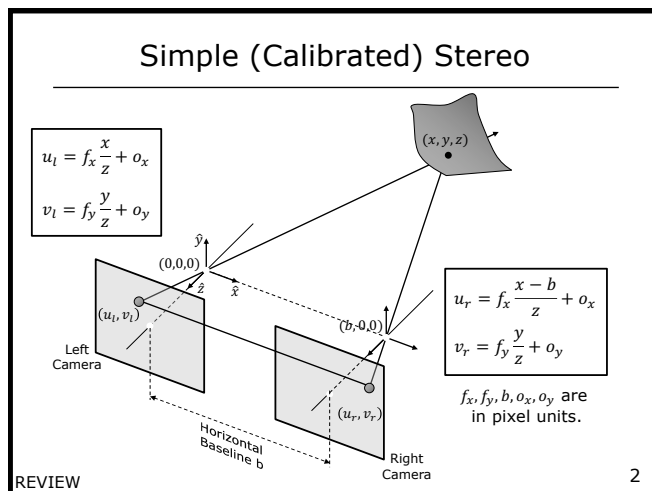
## Uncalibrated Stereo

Shree K. Nayar  
Columbia University

Topic: Uncalibrated Stereo, Module: Reconstruction II  
First Principles of Computer Vision

1

### Simple (Calibrated) Stereo



### Depth and Disparity

Solving for  $(x, y, z)$ :

$$x = \frac{b(u_l - o_x)}{(u_l - u_r)} \quad y = \frac{bf_x(v_l - o_y)}{f_y(u_l - u_r)} \quad z = \frac{bf_x}{(u_l - u_r)}$$

where  $(u_l - u_r)$  is called the Disparity.

REVIEW

3

First, we let's briefly review calibrated stereo, or simple stereo. In simple stereo, we have a left camera and an identical right camera that is displaced with respect to the left one along the  $x$ -axis by a distance  $b$ , called the baseline. From images taken with these two cameras, we can compute the 3D structure of the scene.

For a point  $(u, v)$  in the left image, we first find the corresponding point in the right image using template matching. We now have two outgoing rays and the intersection of these rays is where the physical scene point corresponding to the matched image points lies. If we know the baseline  $b$  and the internal parameters  $f_x, f_y, o_x$ , and  $o_y$ , we can compute the coordinates  $x, y$ , and  $z$  of the scene point using the perspective projection equations for the left and right cameras. The term  $(u_l - u_r)$  in the expressions for  $x, y$ , and  $z$  is called disparity.

We will now explore the more ambitious problem of uncalibrated stereo. The goal of uncalibrated stereo is to recover the 3D structure of a scene from two arbitrary views of it. We will assume that the internal parameters of both cameras are known. In the case of modern digital cameras, these parameters are often included in the meta-tag of each captured image. However, we do not know the rotation and translation of one camera with respect to the other.

In order to solve this problem, we need to formulate a geometric relationship between the two cameras. This relationship can be concisely described using the concept of epipolar geometry. Epipolar geometry relates points in the left and right images through a single 3x3 matrix called the fundamental matrix. We develop a method for computing the fundamental matrix from a small number of corresponding points in the left and right images. Then, we use the fundamental matrix to find the rotation and translation of one camera with respect to the other. At this point, the stereo system is fully calibrated.

Then, to recover the 3D structure of the scene, we will need to find dense correspondences between the two images. Ideally, for every point in the left image, we want to find the corresponding point in the right image. We will show that, for each point in the left image, finding correspondence reduces to a 1D search in the right image. Once we have all the correspondences, we compute depth using a least-squares estimation process. Finally, we will describe how stereo vision is exploited by different animals, and explore how it works in the case of humans.

Let us examine the problem of uncalibrated stereo in more detail. Assume that we have a monument and two photos (referred to as left and right views) are captured using two different cameras. We want to compute the 3D structure of the scene from these two images. We will assume that we know the internal parameters of the two cameras, i.e., the focal lengths  $f_x, f_y$  and the location  $o_x, o_y$  of the principal point of each camera are known. This is not an unrealistic assumption since we can either calibrate the camera using the method described in the previous lecture, or retrieve the

## Uncalibrated Stereo

Method to estimate 3D structure of a static scene from two arbitrary views.

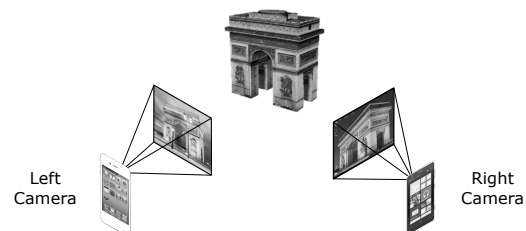
Topics:

- (1) Problem of Uncalibrated Stereo
- (2) Epipolar Geometry
- (3) Estimating Fundamental Matrix
- (4) Finding Dense Correspondences
- (5) Computing Depth
- (6) Stereopsis: Stereo in Nature

4

## Uncalibrated Stereo

Compute 3D structure of static scene from two arbitrary views



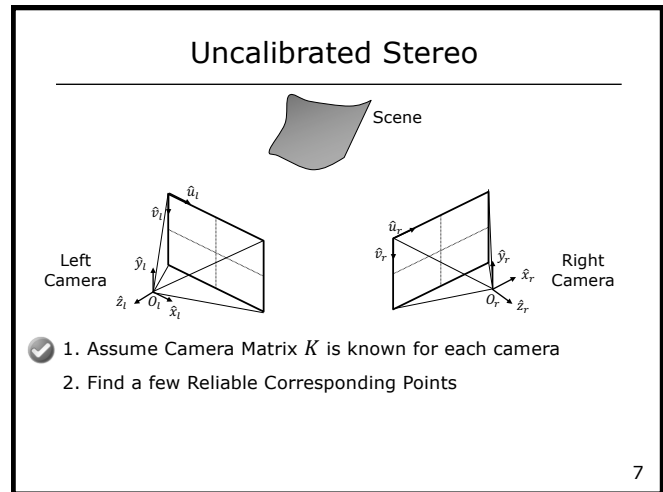
Intrinsics ( $f_x, f_y, o_x, o_y$ ) are known for both views/cameras.

Extrinsics (relative position/orientation of cameras) are unknown.

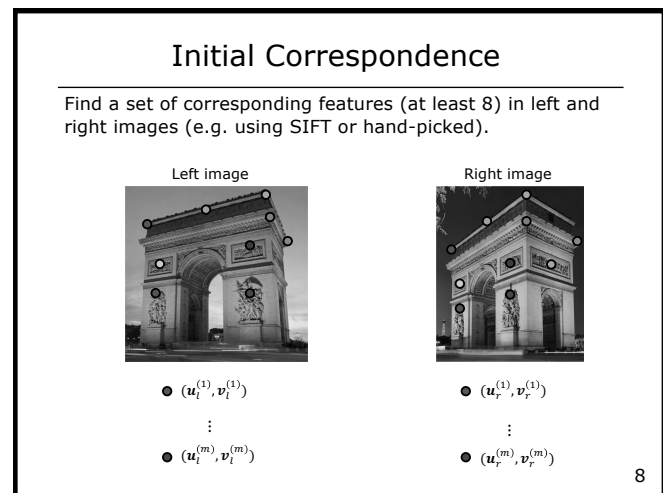
6

internal parameters from the meta-tag embedded in each image. To calibrate our stereo system, we want to find its unknown extrinsic parameters, which are the relative position and orientation of one camera with respect to the other camera.

Here we show a 3D scene imaged by the two cameras. The left camera has a 3D coordinate frame  $O_l$ , and the right camera has its frame  $O_r$ . It should be noted that we do not know the relationship between  $O_l$  and  $O_r$ . All we have are two images where each pixel is defined by its image coordinates  $u$  and  $v$ . Once again, we are assuming that the camera matrix  $K$  (made of intrinsic parameters) of each camera is known.



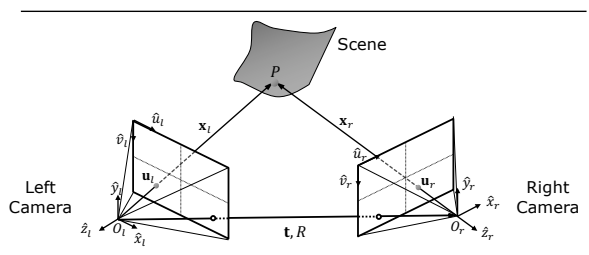
To find the relationship between the 3D frames  $O_l$  and  $O_r$  of the two cameras, we first need to find a small number of reliable correspondences between the two images. For this, we can apply SIFT and chose the most robust matches. For the calibration method we will soon describe, we need a minimum of eight corresponding pairs of image points.





After finding the initial corresponding pairs, we will find the rotation  $R$  and translation  $\mathbf{t}$  of each camera with respect to the other. At this point, our stereo system is calibrated. The next step is to find dense correspondences between the two images. Since  $R$  and  $\mathbf{t}$  are known, each correspondence can be found using a 1D search. Finally, with the dense correspondences, we can compute the 3D structure of scene by triangulation.

### Uncalibrated Stereo



1. Assume Camera Matrix  $K$  is known for each camera
2. Find a few Reliable Corresponding Points
3. Find Relative Camera Position  $\mathbf{t}$  and Orientation  $R$
4. Find Dense Correspondence
5. Compute Depth using Triangulation

9

The method for finding rotation and translation between the two cameras from images taken by them is based on a concept called epipolar geometry, which we will take a closer look at now.

## Epipolar Geometry

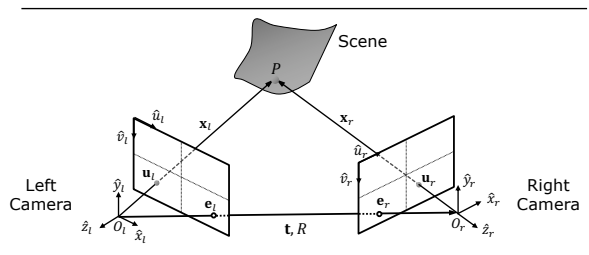
Shree K. Nayar  
Columbia University

Topic: Uncalibrated Stereo, Module: Reconstruction II  
First Principles of Computer Vision

10

Shown here again are the left and right cameras. The projection of the center of the left camera  $O_l$  onto the right camera image and the projection of the center of the right camera  $O_r$  onto the left camera image are indicated by the points  $\mathbf{e}_l$  and  $\mathbf{e}_r$ . These two points are called the epipoles of the stereo system. Any given stereo system has a unique pair of epipoles,  $\mathbf{e}_l$  and  $\mathbf{e}_r$ . Note that while they happen to lie within the right and left images in our figure, in general, the epipoles could lie outside the images as well.

### Epipolar Geometry: Epipoles



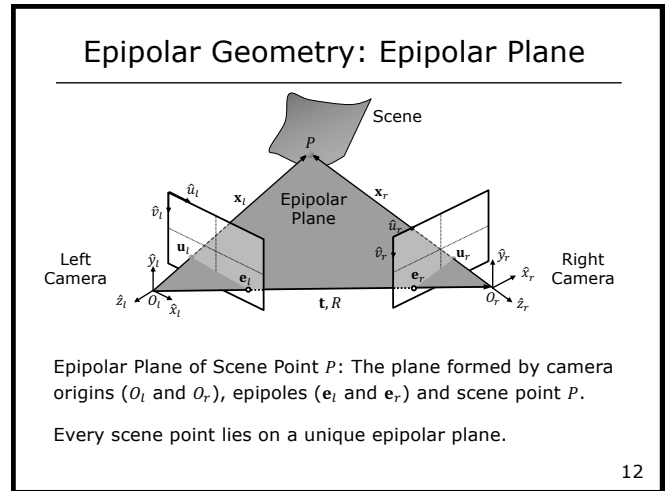
Epipole: Image point of origin/pinhole of one camera as viewed by the other camera.

$\mathbf{e}_l$  and  $\mathbf{e}_r$  are the epipoles.

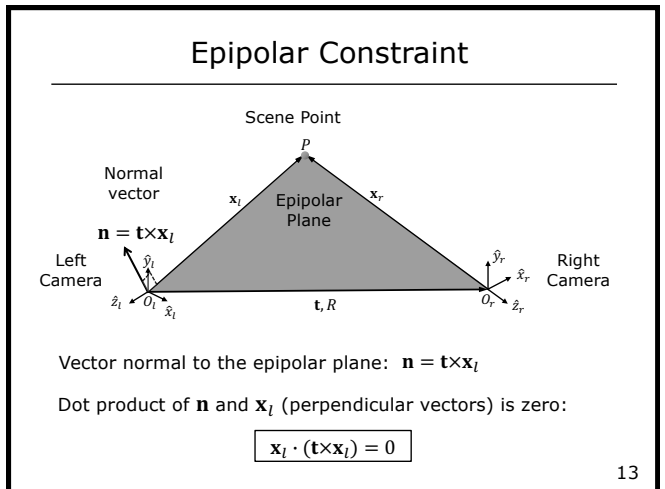
$\mathbf{e}_l$  and  $\mathbf{e}_r$  are unique for a given stereo pair.

11

Consider a plane formed by a scene point  $P$ ,  $O_l$ , and  $O_r$ . This plane also includes the epipoles  $e_l$  and  $e_r$ . This is called the epipolar plane corresponding to the scene point  $P$ . Each point in the scene has a unique epipolar plane. We will use the epipolar plane to setup a constraint that includes the external parameters  $\mathbf{t}$  and  $R$  that are of interest to us.



Let  $\mathbf{n}$  denote a vector that is normal to the epipolar plane. We can calculate  $\mathbf{n}$  as the cross product of the unknown translation vector  $\mathbf{t}$  and the vector  $\mathbf{x}_l$  that corresponds to the point  $P$  in the left camera's frame. Since the normal vector must be perpendicular to  $\mathbf{x}_l$ , the dot product of  $\mathbf{n}$  and  $\mathbf{x}_l$  equals 0. This gives us the epipolar constraint:  $\mathbf{x}_l \cdot (\mathbf{t} \times \mathbf{x}_l) = 0$ . We now want this constraint to include both the translation  $\mathbf{t}$  and the rotation  $R$ .



### Epipolar Constraint

Writing the epipolar constraint in matrix form:

$$\mathbf{x}_l \cdot (\mathbf{t} \times \mathbf{x}_l) = 0$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{bmatrix} = 0 \quad \text{Cross-product definition}$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0 \quad \text{Matrix-vector form}$$

$T_x$

$\mathbf{t}_{3 \times 1}$ : Position of Right Camera in Left Camera's Frame  
 $R_{3 \times 3}$ : Orientation of Left Camera in Right Camera's Frame

$$\mathbf{x}_l = R \mathbf{x}_r + \mathbf{t} \quad \text{2}$$

14

### Epipolar Constraint

Substituting into the epipolar constraint gives:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \left( \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \right) = 0 \quad \text{3}$$

$\mathbf{t} \times \mathbf{t} = 0$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Essential Matrix  $E$

$$E = T_x R$$

[Longuet-Higgins 1981] 15

Shown here is the epipolar constraint in matrix form. The cross-product can be expanded and written as the product of a translation matrix  $T_x$  a  $\mathbf{x}_l$  [1]. We know that  $\mathbf{t}$  is the position of the right camera in the left camera's frame and  $R$  is the orientation of the left camera in the right camera's frame. We can therefore relate the 3D coordinates of point  $P$  in the left camera frame to its 3D coordinates in the right camera frame as  $\mathbf{x}_l = R\mathbf{x}_r + \mathbf{t}$ , which can be written in matrix form as [2].

We can now substitute the right side of expression [2] for  $\mathbf{x}_l$  in the epipolar constraint [1], to get equation [3]. Note that the last term in [3]—the product of the translation matrix and the translation vector—equals  $\mathbf{t} \times \mathbf{t}$ , which is 0. Thus, we are left with the product of two 3x3 matrices, which we will define as a new matrix called the essential matrix  $E$ . Introduced by Louquet-Higgins in 1981, the essential matrix is the product of the translation and rotation matrices and relates  $\mathbf{x}_l$  and  $\mathbf{x}_r$ .

An interesting property of the essential matrix  $E$  is that it can be decomposed into the translation matrix  $T_x$  and the rotation matrix  $R$ . Notice that  $T_x$  is a skew symmetric matrix, which means that any element  $a_{ij}$  equals  $-a_{ji}$ . We also know from previous lectures that the rotation matrix  $R$  is an orthonormal matrix. It turns out that the product of a skew symmetric matrix and an orthonormal matrix can be decomposed into the two component matrices using singular value decomposition. This is what makes the essential matrix special—if we can compute it, we can find the translation  $\mathbf{t}$  and the rotation  $R$ .

### Essential Matrix $E$ : Decomposition

$$E = T_x R$$

$$\begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Given that  $T_x$  is a Skew-Symmetric matrix ( $a_{ij} = -a_{ji}$ ) and  $R$  is an Orthonormal matrix, it is possible to "decouple"  $T_x$  and  $R$  from their product using "Singular Value Decomposition".

Take Away: If  $E$  is known, we can calculate  $\mathbf{t}$  and  $R$ .

MATH PRIMER

16

How do we then compute the essential matrix? The epipolar constraint relates  $\mathbf{x}_l$  and  $\mathbf{x}_r$  which are 3D coordinates of the same scene point. We unfortunately do not know the 3D locations of scene points as this is what we are ultimately interested in finding. However, we do know the projection of a scene point onto the images, i.e, we know  $(u_l, v_l)$  and  $(u_r, v_r)$ . So, our goal is to recast our epipolar constraint in terms of these image coordinates.

### How do we find $E$ ?

Relates 3D position  $(x_l, y_l, z_l)$  of scene point w.r.t left camera to its 3D position  $(x_r, y_r, z_r)$  w.r.t. right camera

$$\mathbf{x}_l^T E \mathbf{x}_r = 0$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

3D position in left camera coordinates
3x3 Essential Matrix
3D position in right camera coordinates

Unfortunately, we don't have  $\mathbf{x}_l$  and  $\mathbf{x}_r$ .

But we do know corresponding points in image coordinates.

17

## Incorporating the Image Coordinates

Perspective projection equations for left camera:

$$\begin{aligned} u_l &= f_x^{(l)} \frac{x_l}{z_l} + o_x^{(l)} & v_l &= f_y^{(l)} \frac{y_l}{z_l} + o_y^{(l)} \\ z_l u_l &= f_x^{(l)} x_l + z_l o_x^{(l)} & z_l v_l &= f_y^{(l)} y_l + z_l o_y^{(l)} \end{aligned}$$

Representing in matrix form:

$$\boxed{1} \quad z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} z_l u_l \\ z_l v_l \\ z_l \end{bmatrix} = \begin{bmatrix} f_x^{(l)} x_l + z_l o_x^{(l)} \\ f_y^{(l)} y_l + z_l o_y^{(l)} \\ z_l \end{bmatrix} = \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} \\ 0 & f_y^{(l)} & o_y^{(l)} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}$$

Known  
Camera Matrix  $K_l$

REVIEW

18

## Incorporating the Image Coordinates

Left camera

$$z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} \\ 0 & f_y^{(l)} & o_y^{(l)} \\ 0 & 0 & 1 \end{bmatrix}}_{K_l} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}$$

Right camera

$$z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} \\ 0 & f_y^{(r)} & o_y^{(r)} \\ 0 & 0 & 1 \end{bmatrix}}_{K_r} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix}$$

$$\mathbf{x}_l^T = [u_l \quad v_l \quad 1] z_l K_l^{-1^T}$$

$$\mathbf{x}_r = K_r^{-1} z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix}$$

19

Shown on the left are the perspective projection equations for the left camera, which give us expressions for  $u_l$  and  $v_l$ . We multiply both sides by  $z_l$ , and using homogeneous coordinates, we can write  $z_l u_l$ ,  $z_l v_l$ , and  $z_l$  in matrix form  $\boxed{1}$ . This matrix, in turn, can be written as the product of the camera matrix  $K_l$  and the 3D coordinates  $(x_l, y_l, z_l)$  of the scene point in the left camera  $\boxed{2}$ . Since we know the internal parameters of the two cameras,  $K_l$  is known to us. We can obtain a similar equation for the right camera. These two equations can be rewritten to get the expressions shown on the right for  $\mathbf{x}_l^T$  and  $\mathbf{x}_r$ .

## Incorporating the Image Coordinates

Epipolar constraint:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Rewriting in terms of image coordinates:

$$[u_l \quad v_l \quad 1] \cancel{z_l} K_l^{-1^T} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} K_r^{-1} \cancel{z_r} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \quad \boxed{1}$$

$$\begin{aligned} z_l &\neq 0 \\ z_r &\neq 0 \end{aligned}$$

20

## Incorporating the Image Coordinates

Epipolar constraint:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Rewriting in terms of image coordinates:

$$[u_l \quad v_l \quad 1] K_l^{-1^T} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} K_r^{-1} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \quad \boxed{2}$$

21

Substituting our expressions for  $\mathbf{x}_l^T$  and  $\mathbf{x}_r$  into the epipolar constraint in slide 17, we get expression  $\boxed{1}$ . In  $\boxed{1}$  we have the known image coordinates  $(u_l, v_l)$  and  $(u_r, v_r)$  and the unknown essential matrix  $E$ , but also  $z_l$  and  $z_r$ , which are unknown. However, since  $z_l$  and  $z_r$  are the depths of the same scene point measured in the camera frames, and the scene lies in front of the two cameras, we know that  $z_l \neq 0$  and  $z_r \neq 0$ . Therefore, we can eliminate  $z_l$  and  $z_r$  to get expression  $\boxed{2}$ , where the essential matrix is the only unknown quantity.

Since  $K_l$ ,  $K_r$ , and  $E$  are 3x3 matrices,  $K_l^{-1^T} E K_r^{-1}$  in slide 21 is also a 3x3 matrix, which we refer to as the fundamental matrix  $F$ . We now have a simple expression for our epipolar constraint where the image coordinates  $(u_l, v_l)$  and  $(u_r, v_r)$  are known and the fundamental matrix  $F$  is unknown. If we can find  $F$ , we can easily compute the essential matrix  $E$  since  $K_l$  and  $K_r$  are known. Once we have  $E$ , we can find the translation  $T_x$  and rotation  $R$  using singular value decomposition.

### Fundamental Matrix $F$

Epipolar constraint:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Rewriting in terms of image coordinates:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

Fundamental Matrix  $F$

$$E = K_l^T F K_r$$

$$E = T_x R$$

[Fagueras 1992, Luong 1992]

22

For calibration, what remains to be done is to compute the fundamental matrix  $F$ . We now present a method for estimating  $F$  from a few pairs of corresponding points in the two images.

### Estimating Fundamental Matrix

Shree K. Nayar

Columbia University

Topic: Uncalibrated Stereo, Module: Reconstruction II  
First Principles of Computer Vision

23

We first find a small number of corresponding features in the two images given to us. This can be done by applying the SIFT detector to the two images and finding the strongest matches between the images. Let's denote the  $i^{th}$  pair of corresponding points as  $(u_l^{(i)}, v_l^{(i)})$  and  $(u_r^{(i)}, v_r^{(i)})$ .

### Stereo Calibration Procedure

Find a set of corresponding features in left and right images (e.g. using SIFT or hand-picked)

Left image



$$\bullet (u_l^{(1)}, v_l^{(1)})$$

$\vdots$

$$\bullet (u_l^{(m)}, v_l^{(m)})$$

Right image



$$\bullet (u_r^{(1)}, v_r^{(1)})$$

$\vdots$

$$\bullet (u_r^{(m)}, v_r^{(m)})$$

24

### Stereo Calibration Procedure

Step A: For each correspondence  $i$ , write out epipolar constraint.

$$\begin{array}{ccc} \begin{bmatrix} u_l^{(i)} & v_l^{(i)} & 1 \end{bmatrix} & \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} & \begin{bmatrix} u_r^{(i)} \\ v_r^{(i)} \\ 1 \end{bmatrix} = 0 \\ \text{Known} & \text{Unknown} & \text{Known} \end{array}$$

Expand the matrix to get linear equation:

$$(f_{11}u_r^{(i)} + f_{12}v_r^{(i)} + f_{13})u_l^{(i)} + (f_{21}u_r^{(i)} + f_{22}v_r^{(i)} + f_{23})v_l^{(i)} + f_{31}u_r^{(i)} + f_{32}v_r^{(i)} + f_{33} = 0$$

25

### Stereo Calibration Procedure

Step B: Rearrange terms to form a linear system.

$$\begin{array}{c} \begin{bmatrix} u_l^{(1)}u_r^{(1)} & u_l^{(1)}v_r^{(1)} & u_l^{(1)} & v_l^{(1)}u_r^{(1)} & v_l^{(1)}v_r^{(1)} & v_l^{(1)} & u_r^{(1)} & v_r^{(1)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_l^{(i)}u_r^{(i)} & u_l^{(i)}v_r^{(i)} & u_l^{(i)} & v_l^{(i)}u_r^{(i)} & v_l^{(i)}v_r^{(i)} & v_l^{(i)} & u_r^{(i)} & v_r^{(i)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_l^{(m)}u_r^{(m)} & u_l^{(m)}v_r^{(m)} & u_l^{(m)} & v_l^{(m)}u_r^{(m)} & v_l^{(m)}v_r^{(m)} & v_l^{(m)} & u_r^{(m)} & v_r^{(m)} & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ A \quad \mathbf{f} \\ \text{(Known)} \quad \text{(Unknown)} \end{array}$$

$$A \mathbf{f} = \mathbf{0}$$

26

We can plug each pair of corresponding image coordinates into our epipolar constraint. By expanding the matrix form of the epipolar constraint, we get the single linear equation shown at the bottom of slide 25. We can stack up the equations for all corresponding pairs and rewrite them in matrix form, as shown in slide 26. The matrix  $A$  is known since it is only comprised of the image coordinates in the left and right cameras. The vector  $\mathbf{f}$  has all the elements of the fundamental matrix. We now have the equation  $A \mathbf{f} = \mathbf{0}$ , which is a form we have seen before.

Notice that the image coordinates in the epipolar constraint are homogeneous coordinates. Therefore, multiplying the fundamental matrix by any scalar  $k$  does not affect the epipolar constraint. In other words,  $F$  and  $kF$  describe the same epipolar geometry. Therefore, the fundamental matrix  $F$  is only defined up to a scale factor. Another way to look at this is that if we double the size of the world and the stereo system, we end up getting the same left and right images. Therefore, we can arbitrarily fix the scale of  $F$ . To make it easier to solve for  $F$ , we set  $\|\mathbf{f}\|^2 = 1$ .

### The Tale of Missing Scale

Fundamental matrix acts on homogenous coordinates.

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 = \begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} kf_{11} & kf_{12} & kf_{13} \\ kf_{21} & kf_{22} & kf_{23} \\ kf_{31} & kf_{32} & kf_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix}$$

Fundamental Matrix  $F$  and  $kF$  describe the same epipolar geometry. That is,  $F$  is defined only up to a scale.

Set Fundamental Matrix to some arbitrary scale.

$$\|\mathbf{f}\|^2 = 1$$

27

Solving for  $F$  then turns out to be a constrained least-squares problem, where we want to find the  $\mathbf{f}$  that minimizes  $\|A\mathbf{f}\|^2$  (from our epipolar constraint) such that  $\|\mathbf{f}\|^2 = 1$  (from setting the scale). As seen in the lectures on image stitching and camera calibration, solving this problem is equivalent to solving an eigenvalue problem. In the end, we get a solution for the vector  $\mathbf{f}$  and rearrange its elements to get the fundamental matrix  $F$ .

### Solving for $F$

Step C: Find least squares solution for fundamental matrix  $F$ .

We want  $A\mathbf{f}$  as close to 0 as possible and  $\|\mathbf{f}\|^2 = 1$ :

$$\min_{\mathbf{f}} \|A\mathbf{f}\|^2 \quad \text{such that } \|\mathbf{f}\|^2 = 1$$

Constrained linear least squares problem

Like solving Projection Matrix during Camera Calibration.

Or, Homography Matrix for Image Stitching.

Rearrange solution  $\mathbf{f}$  to form the fundamental matrix  $F$ .

28

Once we have the fundamental matrix, we can compute the essential matrix, since we know the intrinsic camera matrices. Then, using singular value decomposition, we can decompose the essential matrix into the translation matrix and the rotation matrix. We have now fully calibrated our stereo system.

### Extracting Rotation and Translation

Step D: Compute essential matrix  $E$  from known left and right intrinsic camera matrices and fundamental matrix  $F$ .

$$E = K_l^T F K_r$$

Step E: Extract  $R$  and  $\mathbf{t}$  from  $E$ .

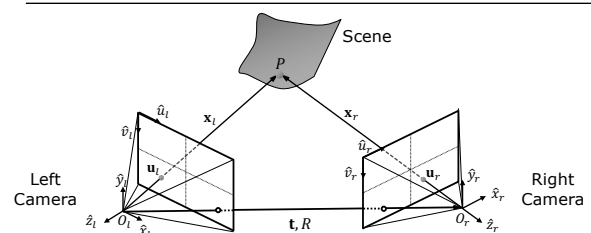
$$E = T_{\times} R$$

(Using Singular Value Decomposition)

29

The next step is to find dense correspondences between the left and right images. Ideally, for every point in the left image, we want to find the corresponding point in the right image. As discussed in the lecture on camera calibration, stereo works only for image regions that are well-textured. Therefore, we are not guaranteed correspondences for all image pixels.

### Uncalibrated Stereo

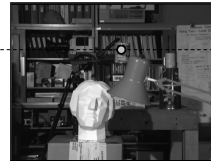


- ✓ 1. Assume Camera Matrix  $K$  is known for each camera
- ✓ 2. Find a few Reliable Corresponding Points
- ✓ 3. Find Relative Camera Position  $\mathbf{t}$  and Orientation  $R$
4. Find Dense Correspondence
5. Compute Depth using Triangulation

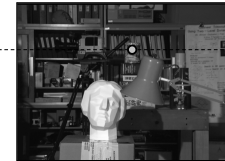
30

Let us review how we found correspondences in the case of simple stereo, where the right camera is displaced with respect to the left camera along the horizontal direction by a distance  $b$  (the baseline). In this special case we showed that for a point in the left image, the corresponding point in the right image must lie on the same horizontal scan line as the point in the left image, which reduces stereo matching to a 1D search.

### Simple Stereo: Finding Correspondences



Left Camera Image



Right Camera Image

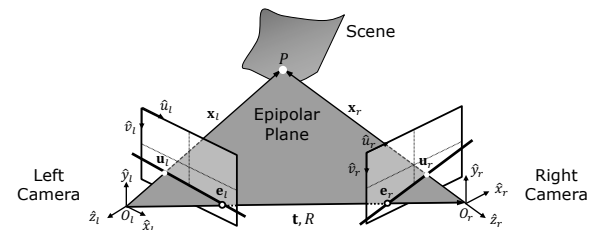
Corresponding scene points lie on the same horizontal scan-line.  
Finding correspondence is a 1D search.

REVIEW

32

It turns out that in the case of uncalibrated stereo as well the stereo matching problem remains a 1D search. Let us now look at how to find this 1D search space in the right image for any point in the left image. This brings us back to epipolar geometry. Shown here is the epipolar plane which we know is unique for any given scene point  $P$ . The epipolar plane intersects with the two image planes to produce two lines called the epipolar lines. Thus, every scene point has two corresponding epipolar lines, one in each of the two images.

### Epipolar Geometry: Epipolar Line



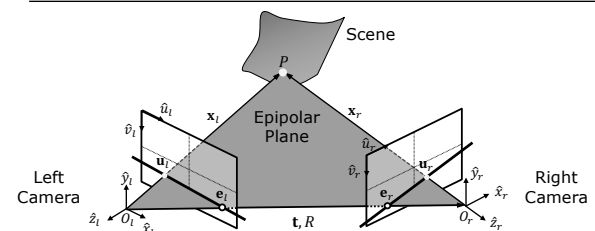
Epipolar Line: Intersection of image plane and epipolar plane.

Every scene point has two corresponding epipolar lines, one each on the two image planes.

33

Consider the coordinates of a scene point  $P$  in the left image. The corresponding point in the right image must lie on the epipolar line in the right image corresponding to  $P$ . As shown here, the left image point corresponds to a single outgoing ray, and the projection of this ray onto the right image is the epipolar line we are looking for. So, to find the corresponding point in the right image we only need to search along this epipolar line.

### Epipolar Geometry: Epipolar Line



Given a point in one image, the corresponding point in the other image must lie on the epipolar line.

Finding correspondence reduces to a 1D search.

34



Now, let us discuss how to find epipolar lines. We have calibrated our stereo system and hence know its fundamental matrix. We are given a single point  $(u_l, v_l)$  in the left image and want to find the epipolar line in the right image. We can use our epipolar constraint again, but in this case,  $(u_l, v_l)$  and  $F$  are known and we have an expression for  $(u_r, v_r)$ . As shown at the bottom, the result is the equation of a straight line in  $u_r$  and  $v_r$ . Similarly, if we start with a point in the right image, we can find its epipolar line in the left image.

### Finding Epipolar Lines

Given: Fundamental matrix  $F$  and point on left image  $(u_l, v_l)$

Find: Equation of Epipolar line in the right image

Epipolar Constraint Equation:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

Expanding the matrix equation gives:

$$(f_{11}u_l + f_{21}v_l + f_{31})u_r + (f_{12}u_l + f_{22}v_l + f_{32})v_r + (f_{13}u_l + f_{23}v_l + f_{33}) = 0$$

Equation for right epipolar line:  $a_l u_r + b_l v_r + c_l = 0$

Similarly we can calculate epipolar line in left image for a point in right image.

35

### Finding Epipolar Lines: Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

and the left image point

$$\tilde{u}_l = \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix}$$

The equation for the epipolar line in the right image is

$$\begin{bmatrix} u_r & v_r & 1 \end{bmatrix} \begin{bmatrix} -.003 & -.003 & 2.97 \\ -.028 & -.008 & 56.38 \\ 13.19 & -29.2 & -9999 \end{bmatrix} \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix} = 0$$

36

### Finding Epipolar Lines: Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

and the left image point

$$\tilde{u}_l = \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix}$$

The equation for the epipolar line in the right image is

$$.03u_r + .99v_r - 265 = 0$$

Epipolar Line

37

As an example, let us assume we are given the fundamental matrix for the two images shown here, as well as one point in the left image. We can simply plug these into the epipolar equation in slide 35 to get the straight-line equation in  $u_r$  and  $v_r$  shown at the bottom.

### Finding Correspondence



Left Image



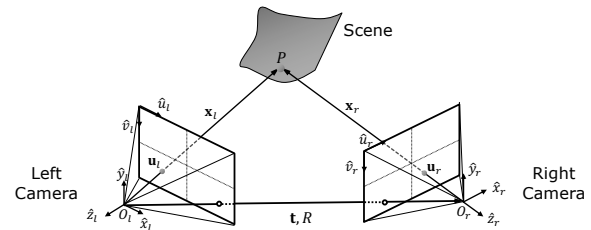
Right Image

Epipolar Line

Corresponding scene points lie on the epipolar lines.  
Finding correspondence is a 1D search.

38

### Uncalibrated Stereo



- ✓ 1. Assume Camera Matrix  $K$  is known for each camera
- ✓ 2. Find a few Reliable Corresponding Points
- ✓ 3. Find Relative Camera Position  $\mathbf{t}$  and Orientation  $R$
- ✓ 4. Find Dense Correspondence
5. Compute Depth using Triangulation

39

To find the matching point in the right image for a point in the left image, we use a small window around the point in the left image and apply template matching along the epipolar line to find its best match. This process is applied to all pixels in the left image.

At this point, we have dense correspondences between the left image and the right image. Now, we want to use each correspondence to estimate the 3D coordinates of the corresponding scene point. This can be done in either the left camera's or the right camera's coordinate frame. This process is referred to as computing depth.

### Computing Depth

Shree K. Nayar  
Columbia University

Topic: Uncalibrated Stereo, Module: Reconstruction II  
First Principles of Computer Vision

40

Shown here are the equations for the projections of a scene point onto the left and right images.

### Computing Depth

Given the intrinsic parameters, the projections of scene point on the two image sensors are:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \quad \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad 41$$

### Computing Depth

Left Camera Imaging Equation

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \quad [1]$$

Right Camera Imaging Equation

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

We also know the relative position and orientation between the two cameras.

$$\begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad [2]$$

42

### Computing Depth

Left Camera Imaging Equation:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad [3]$$

$$\tilde{\mathbf{u}}_l = P_l \tilde{\mathbf{x}}_r$$

Right Camera Imaging Equation:

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad [4]$$

$$\tilde{\mathbf{u}}_r = M_{int_r} \tilde{\mathbf{x}}_r$$

43

Let us refer to these projections as the left and right camera imaging equations [1]. The matrices in these imaging equations are the intrinsic matrices, which are known to us. Using  $\mathbf{u}_l$  and  $\mathbf{u}_r$ , we want to find either  $\mathbf{x}_l$  or  $\mathbf{x}_r$ . Since we have calibrated our stereo system, we know the translation and rotation between the left and right cameras, which gives us the relationship [2] between  $\mathbf{x}_r$  and  $\mathbf{x}_l$ . We can substitute equation [2] into equation [1] to get equation [3]. The product of the extrinsic and intrinsic matrices in [3] can be denoted as the projection matrix  $P_l$ . We keep the right camera imaging equation [4] the same as in equation [1]. The right camera's intrinsic matrix is denoted as  $M_{int_r}$ .

Here, we show the two imaging equations again. Note that  $\mathbf{u}_l$ ,  $\mathbf{u}_r$ ,  $P_l$ , and  $M_{int_r}$  are all known, while  $\mathbf{x}_r$  and  $\mathbf{x}_l$  are unknown. We can rearrange these two equations to get the system of equations in [1]. This system has 4 equations and 3 unknowns— $x_r$ ,  $y_r$ , and  $z_r$ .

### Computing Depth

The imaging equations:

$$\tilde{\mathbf{u}}_r = M_r \tilde{\mathbf{x}}_r \quad \tilde{\mathbf{u}}_l = P_l \tilde{\mathbf{x}}_r$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

Known                      Unknown                      Known                      Unknown

Rearranging the terms:

$$\begin{bmatrix} u_r m_{31} - m_{11} & u_r m_{32} - m_{12} & u_r m_{33} - m_{13} \\ v_r m_{31} - m_{21} & v_r m_{32} - m_{22} & v_r m_{33} - m_{23} \\ u_l p_{31} - p_{11} & u_l p_{32} - p_{12} & u_l p_{33} - p_{13} \\ v_l p_{31} - p_{21} & v_l p_{32} - p_{22} & v_l p_{33} - p_{23} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} m_{14} - u_r m_{34} \\ m_{24} - v_r m_{34} \\ p_{14} - u_l p_{34} \\ p_{24} - v_l p_{34} \end{bmatrix} \quad [1]$$

44

We now have a 4x3 known matrix  $A$ , our unknown scene point coordinates  $\mathbf{x}_r$ , and a known 4x1 vector  $\mathbf{b}$ . This is an overdetermined system of linear equations of the type we have seen before, and we can solve for  $\mathbf{x}_r$  using the pseudo-inverse method. We repeat this process for every pair of corresponding points in the left and right image, which gives us a complete 3D depth map of the scene.

### Computing Depth: Least Squares Solution

$$\begin{bmatrix} u_r m_{31} - m_{11} & u_r m_{32} - m_{12} & u_r m_{33} - m_{13} \\ v_r m_{31} - m_{21} & v_r m_{32} - m_{22} & v_r m_{33} - m_{23} \\ u_l p_{31} - p_{11} & u_l p_{32} - p_{12} & u_l p_{33} - p_{13} \\ v_l p_{31} - p_{21} & v_l p_{32} - p_{22} & v_l p_{33} - p_{23} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} m_{14} - u_r m_{34} \\ m_{24} - v_r m_{34} \\ p_{14} - u_l p_{34} \\ p_{24} - v_l p_{34} \end{bmatrix}$$

$A_{4 \times 3}$                        $\mathbf{x}_r$                        $\mathbf{b}_{4 \times 1}$   
(Known)                      (Unknown)                      (Known)

Find least squares solution using pseudo-inverse:

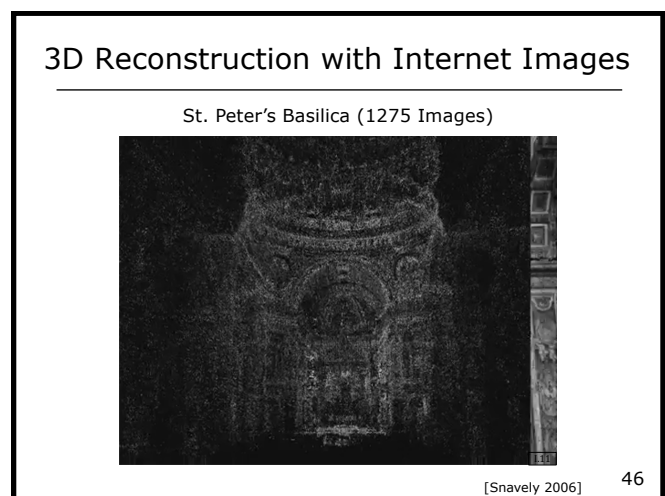
$$\mathbf{A} \mathbf{x}_r = \mathbf{b}$$

$$\mathbf{A}^T \mathbf{A} \mathbf{x}_r = \mathbf{A}^T \mathbf{b}$$

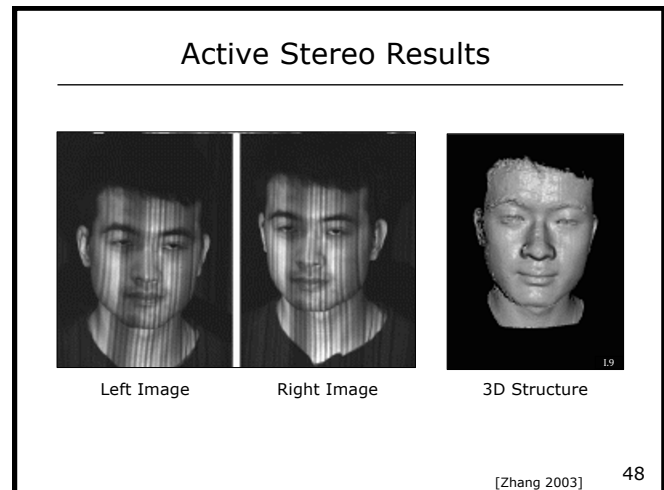
$$\mathbf{x}_r = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

45

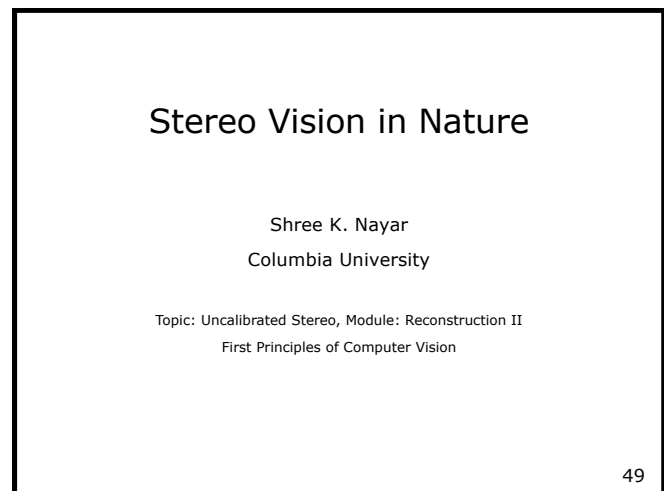
Uncalibrated stereo has a long history and has been used to create some impressive systems. One such system is called Photo Tourism. The system grabs a bunch of arbitrary images of a tourist location, for instance the Eiffel tower, from the internet and applies the calibration technique between all pairs of these images to compute a dense, depth map of the scene. On the right is an impressive reconstruction of St. Peter's Basilica, which was computed using 1275 images. The quality of the reconstruction is best seen in the video of this lecture.



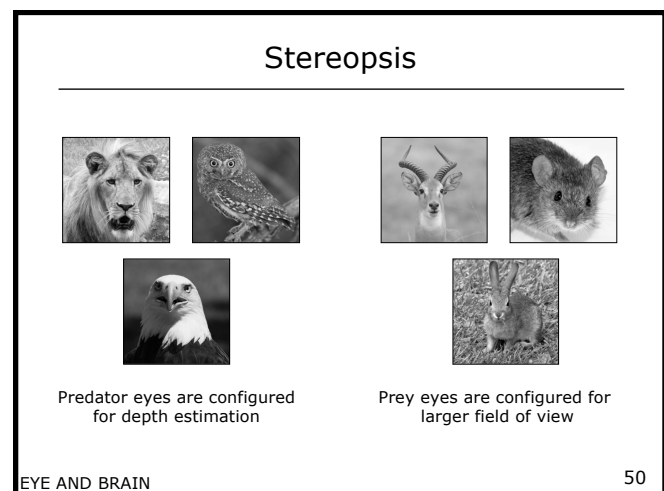
We know that stereo depends on the existence of texture in order to find matches between the two images. One way to make stereo more robust is to use active illumination. Shown here is a stereo system that projects an illumination pattern onto the scene. The pattern is not static but rather changes over time to create a texture that varies over space and time, which makes stereo matching very robust. As can be seen from the computed depth map on the right, the reconstruction of the face is accurate and includes fine details.



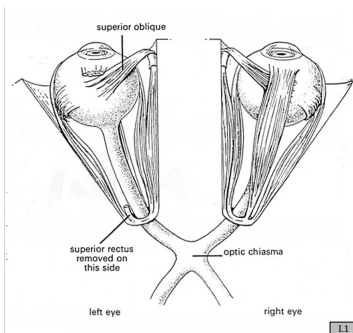
Let us look at some stereo vision systems found in nature. In the natural sciences, stereo is referred to as stereopsis. In Greek, *stereo* means solid and *opsis* means appearance.



Stereopsis manifests in many different ways in animals. For instance, in the case of predators, the two eyes overlap substantially in terms of field of view. This overlap allows predators to perceive depth, which is important in estimating the distance of a prey. In contrast, in the case of prey, the fields of view of the two eyes have less overlap, enabling them to capture a wider field of view. This allows prey to be more aware of whether there is a predator in their surroundings.



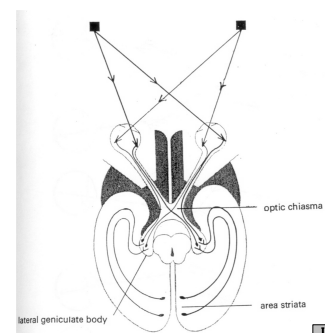
### The Human Stereo System



EYE AND BRAIN

51

### Stereopsis in Humans



EYE AND BRAIN

52

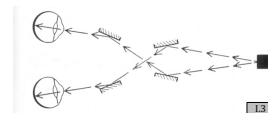
Let us now take a closer look at the human visual system. The two eyes are separated by an interocular distance; in adults, the interocular distance is, on average, around 64 millimeters. When we look at an object, the two eyes rotate such that their optical axes intersect at the object. This process is called vergence, and is made possible by a sophisticated control system that uses six ocular muscles.

On the right we see how an object is projected onto the retinæ (sensors) of the two eyes. The visual cortex, which performs depth perception, is distributed over both sides of our brain. For this reason, the right halves of the two images are directed to the right side of the brain and the left halves to the left side of the brain. This routing of the images is done by a relay system called lateral geniculate nucleus.

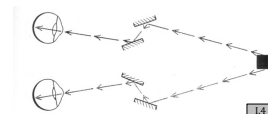
Many contraptions have been developed to understand how human stereo vision works. One such device is called the pseudoscope. It consists of a set of mirrors that reflect light from the object, such that the rays of light that should have traveled from the object into the right eye end up being redirected into the left eye, and vice-versa. In effect, the eyes are swapped causing a reversal of depth—convex objects appear concave, and vice versa. In the case of the telestereoscope, mirrors change the paths of light rays so that the effective interocular distance (between the two eyes) is altered. This has the effect of making objects appear smaller or larger than they really are.

### Human Stereo Experiments

A *pseudoscope* gives reversed depth



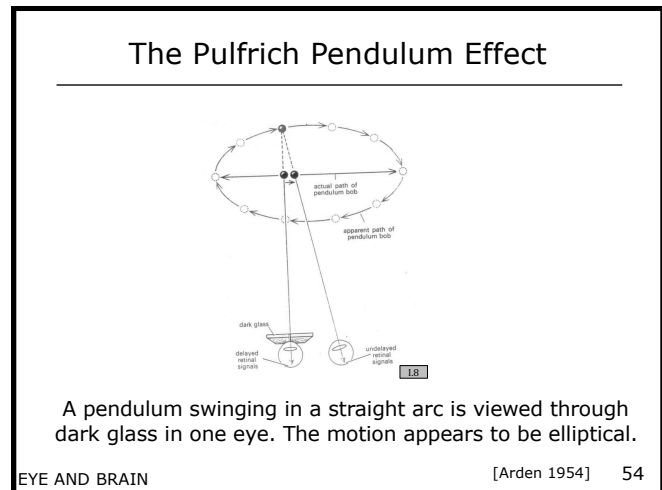
A *telestereoscope* increases separation of the eyes



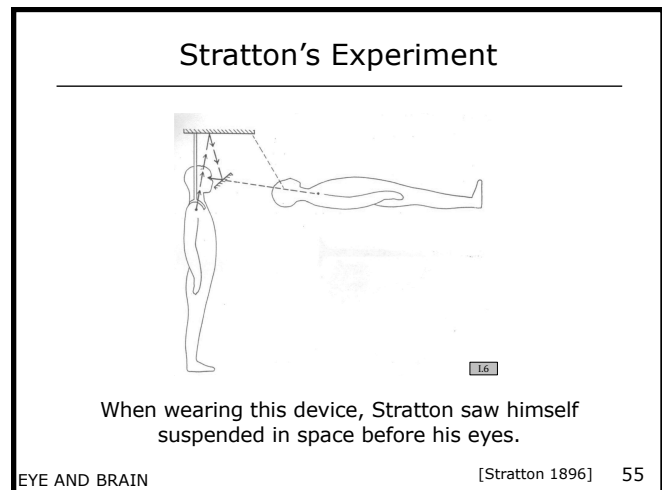
EYE AND BRAIN

53

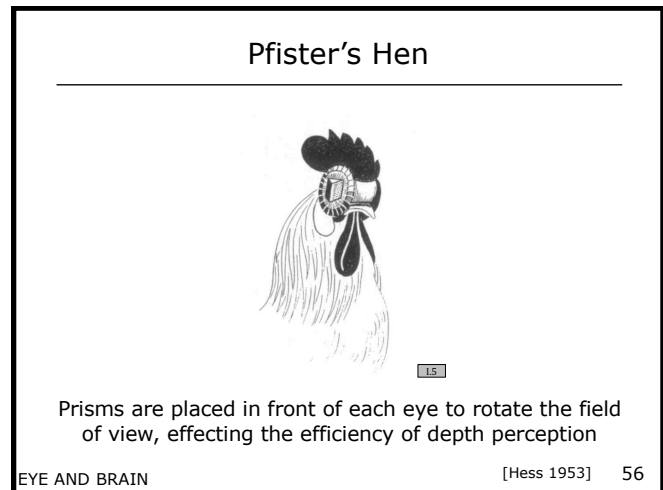
Here is an interesting experiment that produces a phenomenon called the Pulfrich pendulum effect. In front of the left eye is placed a very dark glass (with low transmittance). In front of the observer is a ball that moves back and forth, along a line. Because the left eye is covered by the dark glass, it produces an image of the ball with a slight delay. As a result, by the time the image of the ball is produced in the left eye, the ball has moved to a new location and is perceived in that location by the right eye. Because of the delay in the left eye, the ball appears to be as moving along an ellipse even though it is really moving along a line.



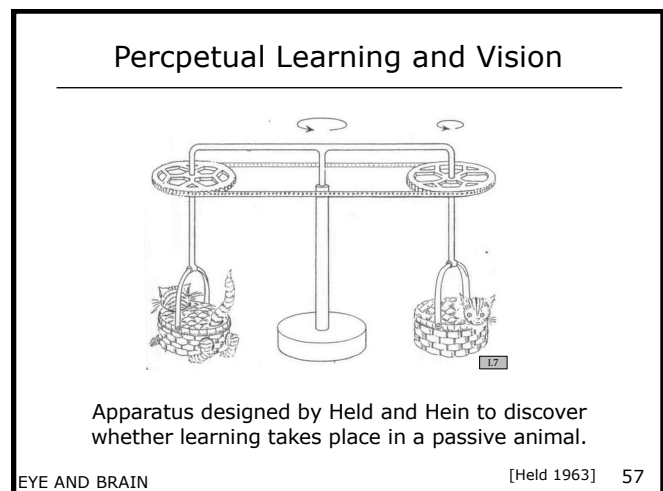
G.M. Stratton performed a fascinating set of experiments in 1896. He was trying to understand what happens to our visual perception if we shift and rotated images before they entered the eyes. The contraption shown here, when worn by the observer, makes them view the world downward from a point above them. Stratton wore this device for many days, to see if he could get used to this new perspective of the world. He only took them off when he went sleep, during which time he wore blinders. In his notes, he documented that he had an out of body experience and tried to teach himself to handle various objects, tie his shoelaces, and walk around. Although difficult, over time, he was able to get used to seeing the world from this unusual perspective. Most interestingly, he reported that when he stopped wearing the device, he quickly got used to the “normal” perspective. His experiment demonstrated that the human visual system has some ability to adapt to new perspectives of the world.



This ability to adapt is not true for all animals. In this experiment by Pfister, prisms were placed in front of the eyes of hens to rotate the fields of view of the eyes by 7, 10, or even 15 degrees. When trying to peck on a grain, for instance, the hen ended up pecking in the wrong place. While humans and monkeys can adapt to shifted or even inverted images, less intelligent animals such as the hen cannot, even after many months.



Finally, shown here is a fascinating experiment done by Held and Hein, which sought to determine whether we need to interact physically with the world in order to learn to see. Two kittens were raised in darkness initially, and then they were placed in the two baskets shown here. One of the baskets has holes in it allowing the kitten in it to touch the ground. The kitten in the other basket has no contact with the ground. The kitten that could make contact with the ground was able to use its feet spin the contraption. This kitten became an active observer while the other one remained a passive observer. Both kittens received the same visual stimulus—they were exposed to the same environment as the contraption spins.



Over time, the kitten that was able to make contact with the ground and hence spin the contraption ended up developing vision and was able to navigate and walk around when it was removed from the basket. However, the passive observer remained essentially blind. This result suggests that in order to develop visual perception, an organism needs to be an active learner, i.e., physically interact with the world.



## References: Textbooks

- Robot Vision (Chapter 13)  
Horn, B. K. P., MIT Press
- Computer Vision: A Modern Approach (Chapter 10)  
Forsyth, D and Ponce, J., Prentice Hall
- Multiple View Geometry (Chapters 8-10)  
Hartley, R. and Zisserman, A., Cambridge University Press
- Computer Vision: Algorithms and Applications (Chapter 7)  
Szeliski, R., Springer
- An introduction to 3D Computer Vision (Chapter 3)  
Cyganek, B., Siebert, J. P., Wiley Pub

59

## References: Papers

- [Arden 1954] G. B. Arden and R. A. Weale. "Variations of the Latent Period of Vision." Proc. Royal Society of London, 1954.
- [Fagueras 1992] O. Fagueras. "What can be seen in three dimensions with an uncalibrated stereo rig?." European Conference on Computer Vision, 1992.
- [Furukawa 2010] Y. Furukawa, B. Curless, S. M. Seitz and R. Szeliski. "Towards Internet-scale Multi-view Stereo" CVPR 2010.
- [Held 1963] R. Held and A. Hein. "Movement-produced stimulation in the development of visually guided behavior." Journal of comparative and physiological psychology, 1963.
- [Longuet-Higgins 1981] H.C. Longuet-Higgins. "A computer algorithm for reconstructing a scene from two projections." Nature, 1981.
- [Luong 1992] Q. T. Luong. Matrice fondamentale et auto-calibration en vision par ordinateur. PhD Thesis, University of Paris, Orsay, 1992.
- [Snavely 2006] N. Snavely, S. M. Seitz and R. Szeliski. "Photo tourism: Exploring photo collection in 3D," ACM SIGGRAPH, 2006.
- [Stratton 1896] G. M. Stratton. "Some preliminary experiments on vision without inversion of the retinal image." Psychological Review, 1896.
- [Zhang 2003] L. Zhang, B. Curless and S. M. Seitz. "Spacetime Stereo: Shape Recovery for Dynamic Scenes," CVPR 2003.

60

## Image Credits

- I.1 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.2 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.3 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.4 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.5 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.6 "Some Preliminary Experiments on Vision without Inversion of the Retinal Image." G.M. Stratton. Psychological Review, 1896. Public Domain.
- I.7 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.8 Eye and Brain. R.L. Gregory, Princeton University Press, 4th Edition, 1990. Used with permission.
- I.9 Li Zhang. Used with permission.

61

## Image Credits

- I.11 Noah Snavely. Used with permission.
- I.12 Yasutaka Furukawa. Used with permission.

62

**Acknowledgements:** Thanks to Pranav Sukumar, Kevin Chen and Nikhil Nanda for their help with transcription, editing and proofreading.

## References

- [Szeliski 2022] Computer Vision: Algorithms and Applications, Szeliski, R., Springer, 2022.
- [Forsyth and Ponce 2003] Computer Vision: A Modern Approach, Forsyth, D. and Ponce, J., Prentice Hall, 2003.
- [Horn 1986] Robot Vision, Horn, B. K. P., MIT Press, 1986.
- [Hartley and Zisserman 2000] Multiple View Geometry, Hartley, R. and Zisserman, A., Cambridge University Press, 2000.
- [Cyganek and Siebert 2009] An introduction to 3D Computer Vision, Cyganek, B. and Siebert, J. P., Wiley, 2009.
- [Tsai 1986] An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, Tsai, R. Y., CVPR, 1986.
- [Arden and Weale 1954] Variations of the Latent Period of Vision, Arden, G. B. and Weale, R. A., Proc. Royal Society of London, 1954.
- [Fagueras 1992] What can be seen in three dimensions with an uncalibrated stereo rig?, Fagueras, O., European Conference on Computer Vision, 1992.
- [Furukawa 2010] Towards Internet-scale Multi-view Stereo, Furukawa, Y., Curless, B., Seitz, S. M. and Szeliski, R., CVPR 2010.
- [Held and Hein 1963] Movement-produced stimulation in the development of visually guided behavior, Held, R. and Hein, A., Journal of comparative and physiological psychology, 1963.
- [Longuet-Higgins 1981] A computer algorithm for reconstructing a scene from two projections, Longuet-Higgins, H. C., Nature, 1981.
- [Luong 1992] Matrice fondamentale et auto-calibration en vision par ordinateur, Luong, Q. T., PhD Thesis, University of Paris, Orsay, 1992.
- [Snavely 2006] Photo tourism: Exploring photo collection in 3D,” Snavely, N., Seitz, S. M. and Szeliski, R., ACM SIGGRAPH, 2006.

[Stratton 1896] Some preliminary experiments on vision without inversion of the retinal image, Stratton, G. M., Psychological Review, 1896.

[Zhang 2003] Spacetime Stereo: Shape Recovery for Dynamic Scenes, Zhang, L., Curless B. and Seitz, S. M., CVPR 2003.