



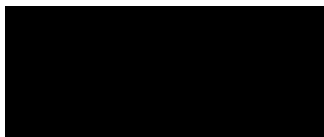
THE UNIVERSITY OF
WESTERN AUSTRALIA

CITS 4402 Computer Vision

Ajmal Mian

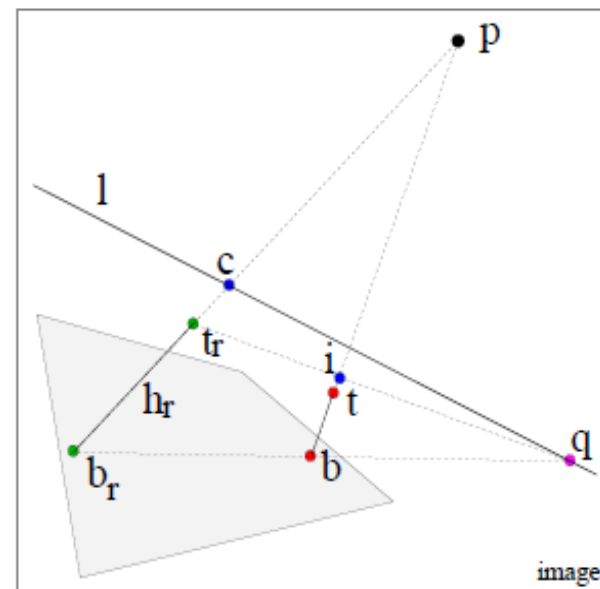
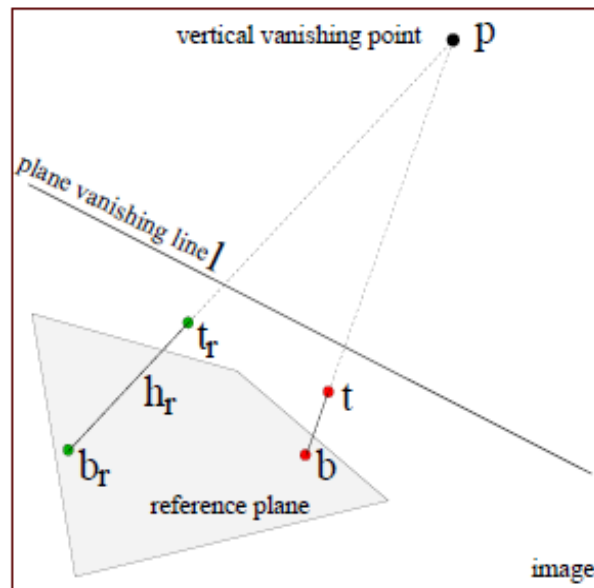
Lecture 10 – Stereo Reconstruction

Revision: Height Measurement with Cross-Ratio



$$t = \frac{i(R_W - 1)}{R_W}$$

$$h = \frac{h_r(R_I - 1)}{R_I}$$



As i approaches t , cross ratio becomes extremely large.
But then, i is already the solution i.e. $t = i$ or $h = h_r$



Objectives of this lecture

- ↘ Understanding stereo vision
- ↘ Understanding epipolar geometry
- ↘ Calculating the Essential and Fundamental matrices
- ↘ How to perform stereo image rectification
- ↘ 3D reconstruction using stereo cameras
- ↘ 3D reconstruction using structured light

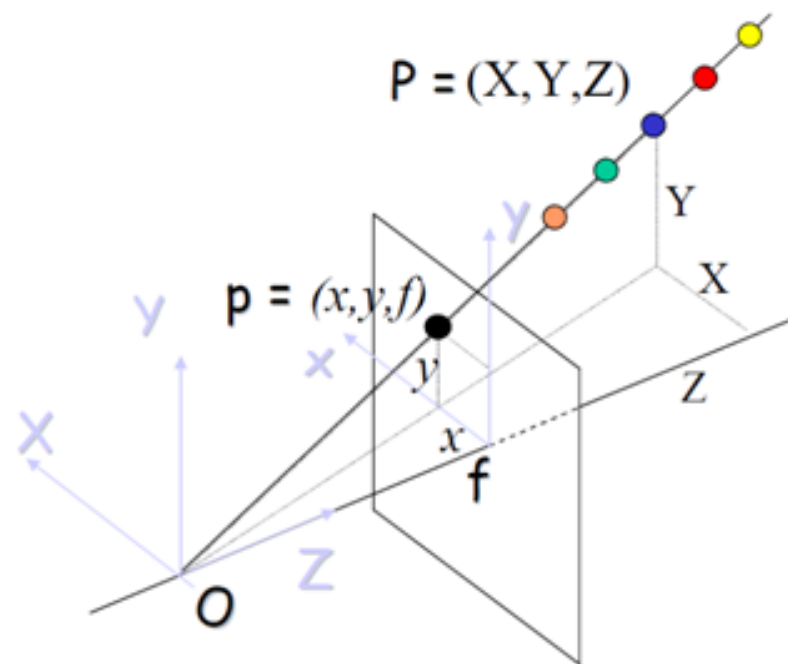
Depth ambiguity in an image

↘ All points on the ray OP has the same image coordinates p .

$$x = \frac{fX}{Z} = \frac{fkX}{kZ}$$

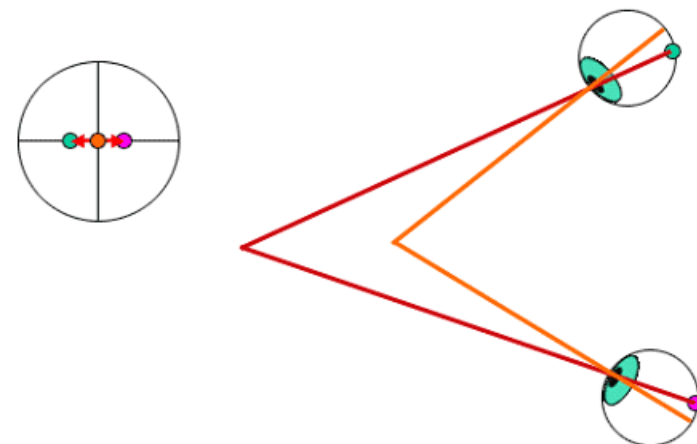
$$y = \frac{fY}{Z} = \frac{fkY}{kZ}$$

For any $k \neq 0$



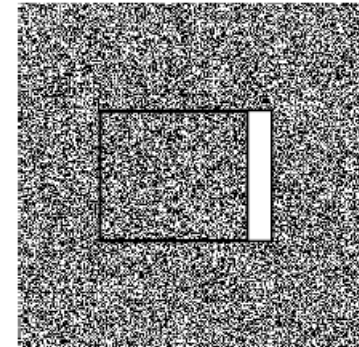
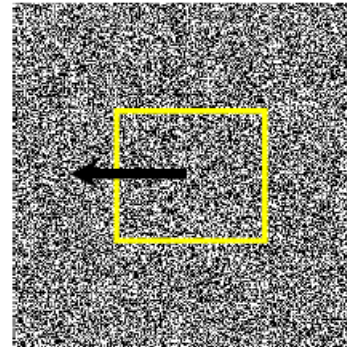
Human's have two eyes for stereo vision

- ↘ The disparity between the left and right retina images gives us a perception of depth
- ↘ The disparity in images is proportional to the object's distance
- ↘ In 3D movies, two videos are captured in a similar way
- ↘ And shown to us using glasses such that the left eye sees only the left image and the right eye sees only the right image



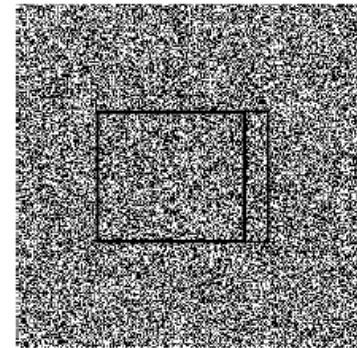
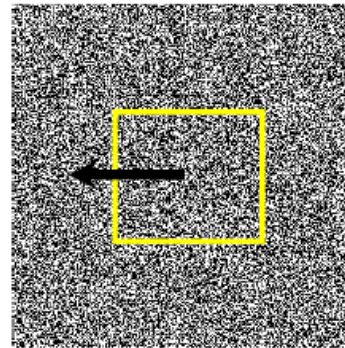
Random Dot Stereogram

- ↘ Generate an image with random dots
- ↘ Duplicate the image
- ↘ Select a rectangular region inside one of the images and shift it horizontally



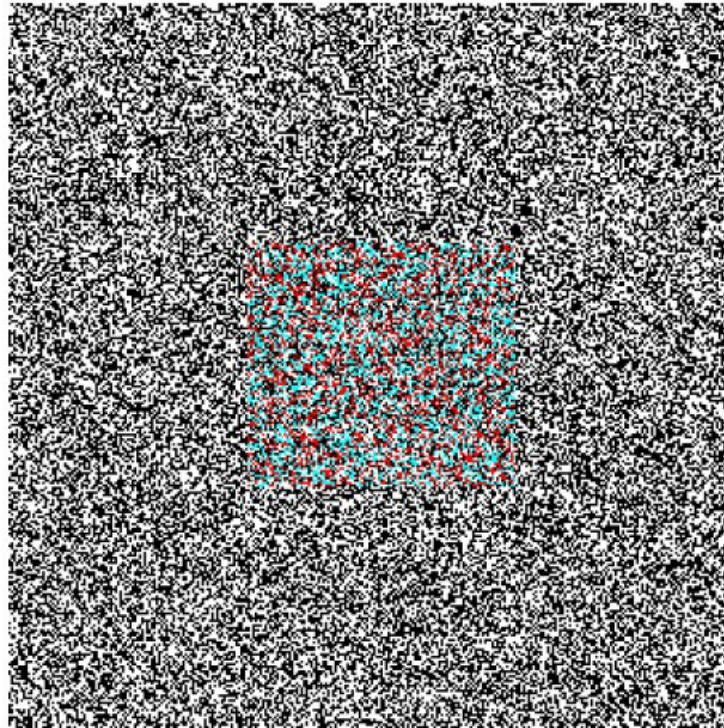
Random Dot Stereogram

- Generate an image with random dots
- Duplicate the image
- Select a rectangular region inside one of the images and shift it horizontally
- Fill the hole with more random dots
- View with a stereoscope OR focus on a point behind the image
- You will see the rectangular region to appear in front or behind the rest of the surface giving a 3D perception





You can use colour coding and view with colour 3D glasses





Random Dot Autostereogram

- Invented by Christopher Tyler 1979.
- Uses a single image to create 3D effect

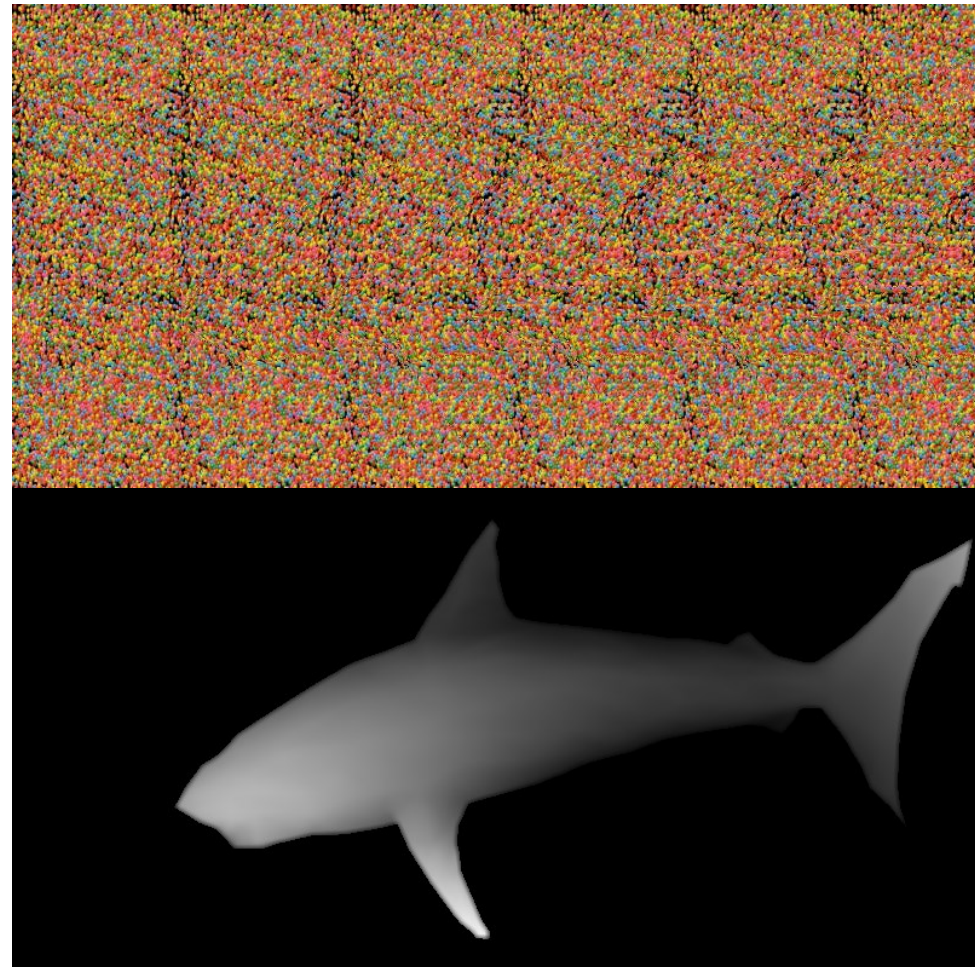


Image courtesy of Fred Hsu, March 2005

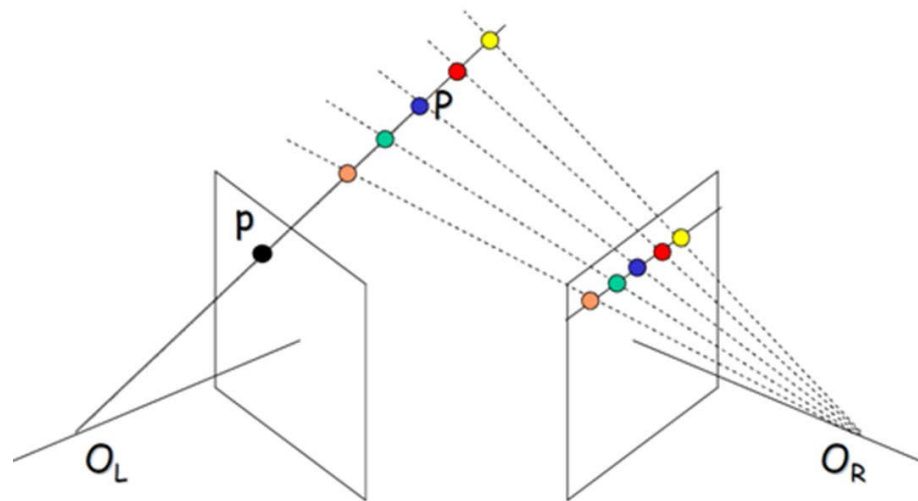


Back to camera based stereo

- ↘ The human mind is excellent at resolving stereo image disparity to perceive depth and does not even need to interpret the image to do so
- ↘ Can we make a computer do the same?

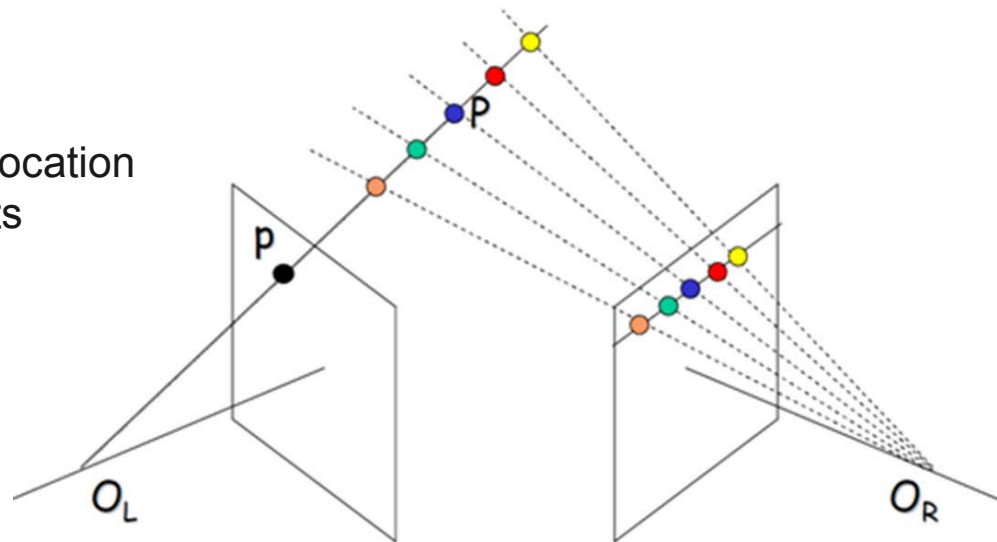
Add a second camera

- ↘ Stereo cameras can resolve depth ambiguity
- ↘ Notice how the points project on the second image plane
- ↘ A single world point P and two optical centers form a triangle
- ↘ We can find the location of P in world coordinates using triangulation



How can we resolve the triangulation?

- ↘ We can find O_L and O_R in world coordinates through camera calibration
- ↘ Let's say, we want to find the world coordinates of the pixel p
- ↘ In other words, we want to find the location of P in world coordinates since p is its projection on the left image plane
- ↘ To complete the triangle, we need to find the projection of P on the right image plane
- ↘ This is called finding the corresponding point of p



OR the correspondence problem



Two tasks

- ↘ For each point in the left image, find a corresponding point in the right image
[CORRESPONDENCE]
- ↘ Calculate the 3-D $[X, Y, Z]$ coordinates of each

A simple stereo system

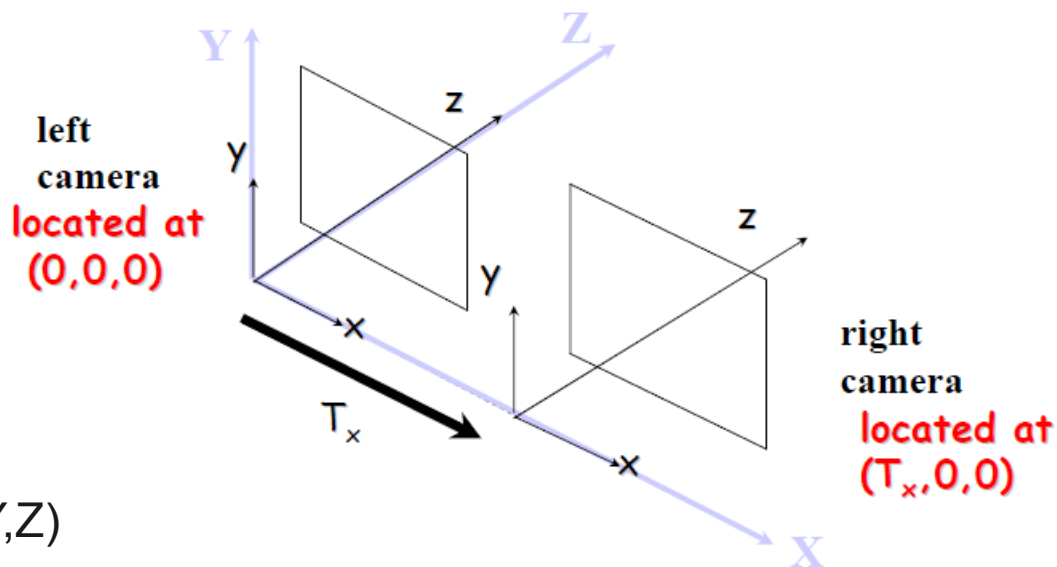


Image coordinates of point (X,Y,Z)
in Left Camera

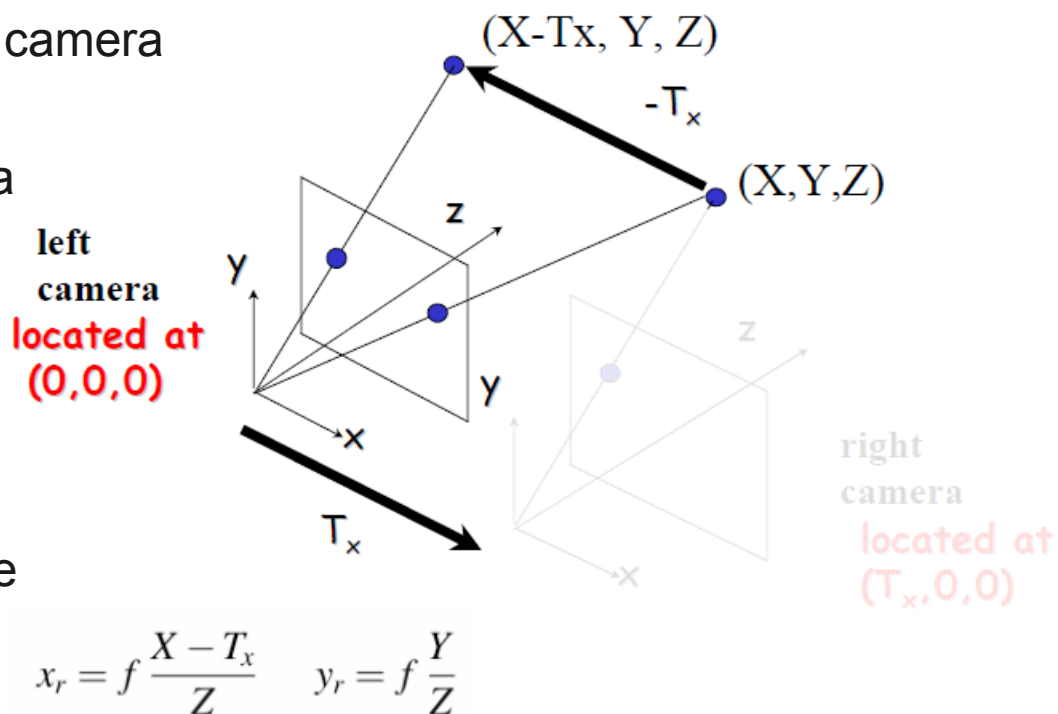
$$x_l = \frac{fX}{Z}, \quad y_l = \frac{fY}{Z}$$

A simple stereo system

↘ Move the point and the right camera

↘ Moving the point and camera together does not affect where it is projected

↘ The image coordinates of the point in the right camera are



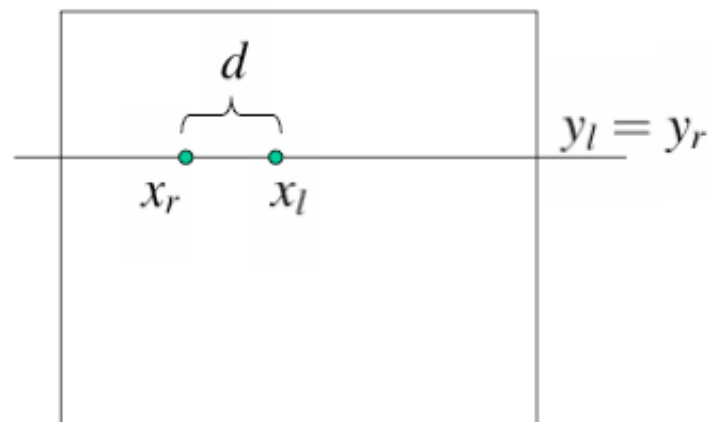
Stereo disparity

Left camera

$$x_l = f \frac{X}{Z} \quad y_l = f \frac{Y}{Z}$$

Right camera

$$x_r = f \frac{X - T_x}{Z} \quad y_r = f \frac{Y}{Z}$$



Stereo Disparity

$$d = x_l - x_r = f \frac{X}{Z} - \left(f \frac{X}{Z} - f \frac{T_x}{Z} \right)$$

$$d = \frac{f T_x}{Z}$$

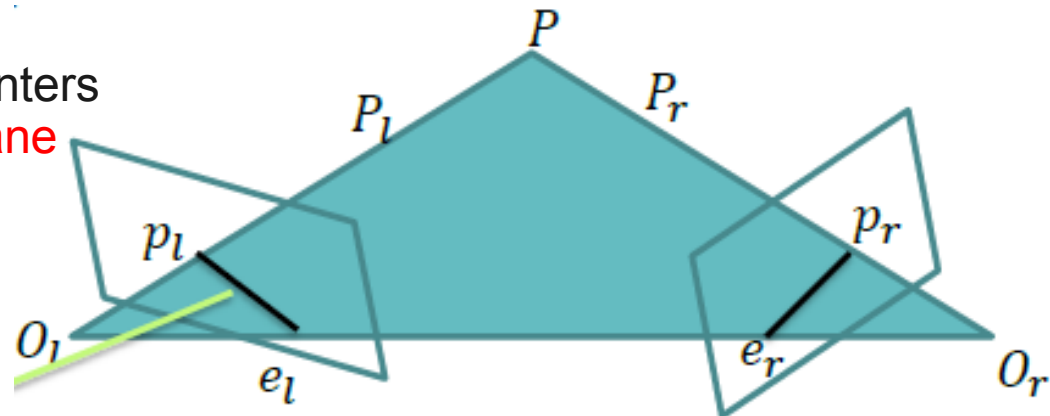
depth $Z = \frac{\text{baseline } f T_x}{\text{disparity } d}$

Important equation!

Note that the depth and stereo disparity are inversely proportional

Epipolar Geometry : Definitions

- ↘ A 3D point P and two optical centers O_l and O_r define an **epipolar plane**
- ↘ The intersection of the epipolar plane with the image planes defines **epipolar lines**
- ↘ All epipolar lines in an image plane intersect at the **epipoles** (e_l or e_r)
- ↘ **Epipolar line** is the image in one camera of the ray through the optical center and image point in the other camera
- ↘ **Epipole** is the image of one camera in the other.
- ↘ Can you find the epipole in the previous example?





The Essential matrix and the Fundamental matrix

- ↘ Both are 3x3 matrices that encode the epipolar geometry of the two views
- ↘ Multiplying a point in one image by the essential/fundamental matrix gives us the epipolar line in the second image to search along
- ↘ Does not give the exact location of the point!
- ↘ We need to search for the exact location using template matching for example.
- ↘ Search for the matching point is constrained to the epipolar line. This is known as the **epipolar constraint**.

The Essential Matrix

$$P_r = R(P_l - T)$$

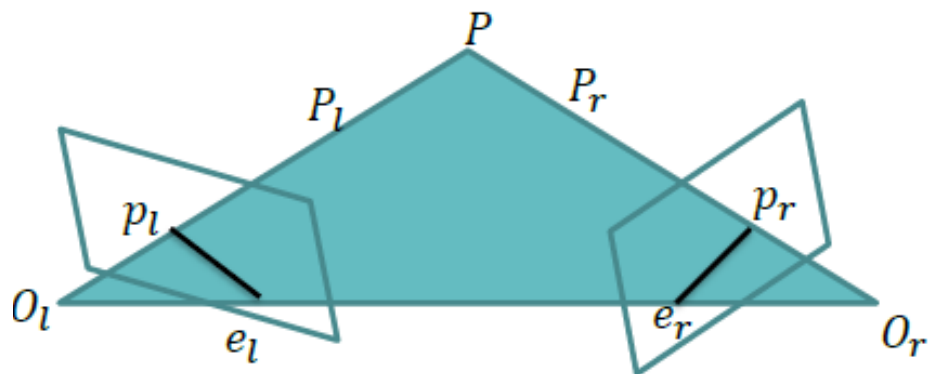
$$\Rightarrow P_r^T R S P_l = 0$$

$$\Rightarrow P_r^T E P_l = 0$$

R is the rotation matrix and
 T the translation vector

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

$E = RS$ is the *Essential Matrix*



The essential matrix has rank 2 and depends only on the extrinsic camera parameters



Longuet-Higgins equations

$$P_r^T E P_l = 0 \quad \text{Relates two rays in world coordinates}$$

$$p_l = \frac{f_l}{Z_l} P_l \quad p_r = \frac{f_r}{Z_r} P_r$$

$$\left(\frac{Z_r}{f_r} p_r \right)^T E \left(\frac{Z_l}{f_l} p_l \right) = 0$$

$$p_r^T E p_l = 0$$

Relates two 2D image points



Longuet-Higgins equation

- ↘ An image point can also be thought of as a viewing ray
 - (u, v) 2D image point
 - (u, v, f) 3D point on image plane
 - $k(u, v, f)$ viewing ray into the scene
 - $k(X, Y, Z)$ ray through the point P in the scene

- ↘ $k = f/Z$, $u = fX/Z$ and $v = fY/Z$



Epipolar lines

- ↘ Consider the line $l = [a \ b \ c]^T$ in an image i.e. $au + bv + c = 0$
- ↘ Now consider a point $p = [u \ v \ 1]^T$ in homogeneous coordinates
- ↘ Recall that if p lies on l , then $p^T l = l^T p = 0$
- ↘ Remember $p_r^T E p_l = 0$
- ↘ The epipolar line $l_r = E p_l$ corresponds to the point p_r
- ↘ Similarly, the epipolar line $l_l = E^T p_r$ corresponds to the point p_l
- ↘ A point in the right image corresponds to a line in the left image. A point in the left image corresponds to a line in the right image.



Epipoles

- ↘ Epipoles lie on epipolar lines
- ↘ $e_r^T E p_l = 0$ and $p_r^T E e_l = 0$
- ↘ In fact, all epipolar lines contain the epipole since they intersect at the epipole
- ↘ Thus $e_r^T E = 0$, and $E e_l = 0$
- ↘ Notice that E has a right and left null vector
- ↘ These two null vectors are the epipoles



Fundamental matrix

- ↘ The essential matrix uses camera coordinates
- ↘ We want to perform computations in pixel or image coordinates
- ↘ Only possible if we know the intrinsic camera parameters
- ↘ The fundamental matrix is similar to the essential matrix because it gives the relationships between points and epipolar lines in image coordinates



Fundamental matrix

- ↘ The pixel coordinates and camera coordinates are related by an affine transformation matrix

$$\bar{p}_l = M_l p_l, \Rightarrow p_l = M_l^{-1} \bar{p}_l$$

$$\bar{p}_r = M_r p_r, \Rightarrow p_r = M_r^{-1} \bar{p}_r$$

$$p_r^T E p_l = 0$$

Substituting p_r and p_l

$$(M_r^{-1} \bar{p}_r)^T E (M_l^{-1} \bar{p}_l) = 0$$

$$\bar{p}_r^T (M_r^{-T} E M_l^{-1}) \bar{p}_l = 0$$

$$\bar{p}_r^T F \bar{p}_l = 0$$



Fundamental matrix

- ↘ $F = M_r^{-T} R S M_l^{-1}$
- ↘ Has rank 2
- ↘ Depends on intrinsic and extrinsic camera parameters
- ↘ Analogous to the essential matrix
- ↘ Thus all previous derivations with E hold in this case too
 - $e_r^T F = 0$, and $F e_l = 0$
- ↘ Additionally, pixel relationships hold i.e. $[u_r \ v_r \ 1] F [u_l \ v_l \ 1]^T = 0$



Estimating the fundamental matrix

$$\begin{bmatrix} u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = 0$$

$$\begin{aligned} &u_i u'_i f_{11} + u_i v'_i f_{21} + u_i f_{31} + \\ &v_i u'_i f_{12} + v_i v'_i f_{22} + v_i f_{32} + \\ &u'_i f_{13} + v'_i f_{23} + f_{33} = 0 \end{aligned}$$

↘ OR

$$u_i u'_i f_{11} + u_i v'_i f_{21} + u_i f_{31} + v_i u'_i f_{12} + v_i v'_i f_{22} + v_i f_{32} + u'_i f_{13} + v'_i f_{23} + f_{33} = 0$$

↘ One set of corresponding points gives one equation



Estimating the fundamental matrix

$$\begin{bmatrix} u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} u_i u'_i & u_i v'_i & u_i & v_i u'_i & v_i v'_i & v_i & u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

Minimum 8 points are required
but you can use more points

This is a system of linear equations $Ax = 0$ A has rank 8

Solution: Find the eigenvector of $A^T A$ with the minimum eigenvalue



Eight-Point Algorithm

- ↘ Construct the $m \times 9$ matrix A where $m \geq 8$ (eight or more points)
- ↘ Find Singular Value Decomposition (SVD) of A i.e. $USV^T = A^T A$
- ↘ The last column of U contains the entries of the F matrix
- ↘ F must have rank 2
 - (so it can have a left and right null space i.e. the epipoles)
- ↘ Find SVD of F i.e. $U_f S_f V_f^T = F$
- ↘ Set the smallest singular value in S_f to zero to create S_f'
- ↘ Recompute $F = U_f S_f' V_f^T$



Numerical Stability

- ↘ Image correspondences can have a wide range of values
- ↘ It is better to normalize the corresponding points so that they have zero mean and unit standard deviation

$$\hat{x}_i = T x_i, \text{ and } \hat{x}'_i = T' x'_i$$

$$T = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & -\mu_x \\ 0 & \frac{1}{\sigma^2} & -\mu_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$F = (T')^{-1} F_n T$$

F_n is the fundamental matrix calculated from the normalized set of corresponding points



Calculating the 3D coordinates of a point

If we have a pair of corresponding points in stereo images and we know the camera parameters of both cameras, we can find its $[X, Y, Z]$ location

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$u = \frac{q_{11}X + q_{12}Y + q_{13}Z + q_{14}}{q_{31} + q_{32} + q_{33} + q_{34}}$$

$$v = \frac{q_{21}X + q_{22}Y + q_{23}Z + q_{24}}{q_{31} + q_{32} + q_{33} + q_{34}}$$

- We get another two equations from the second camera.
- Both sets of equations share the same three variables $[X Y Z]$
- Four equations and three variables – can be solved with linear least squares

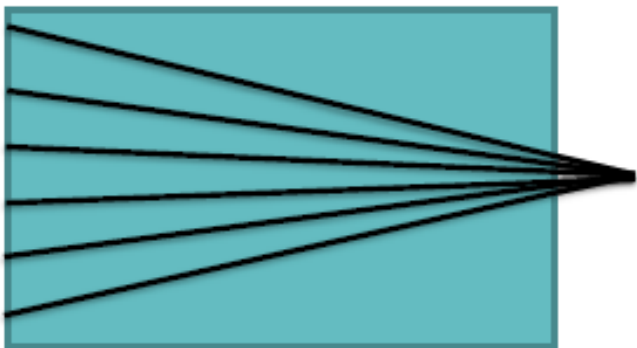
$$u' = \frac{q'_{11}X + q'_{12}Y + q'_{13}Z + q'_{14}}{q'_{31} + q'_{32} + q'_{33} + q'_{34}}$$

$$v' = \frac{q'_{21}X + q'_{22}Y + q'_{23}Z + q'_{24}}{q'_{31} + q'_{32} + q'_{33} + q'_{34}}$$

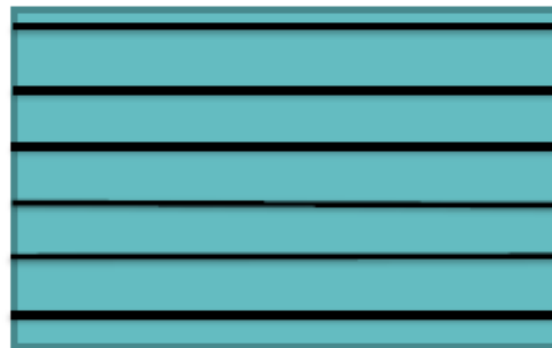


BREAK

Image rectification



Epipolar lines are usually like this

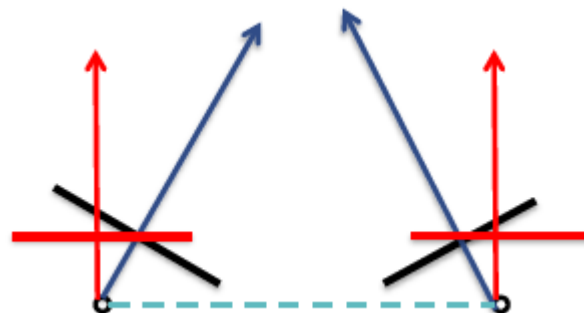


We want to make them like this

- To simplify for correspondence points.
- Search is easier along a horizontal line.

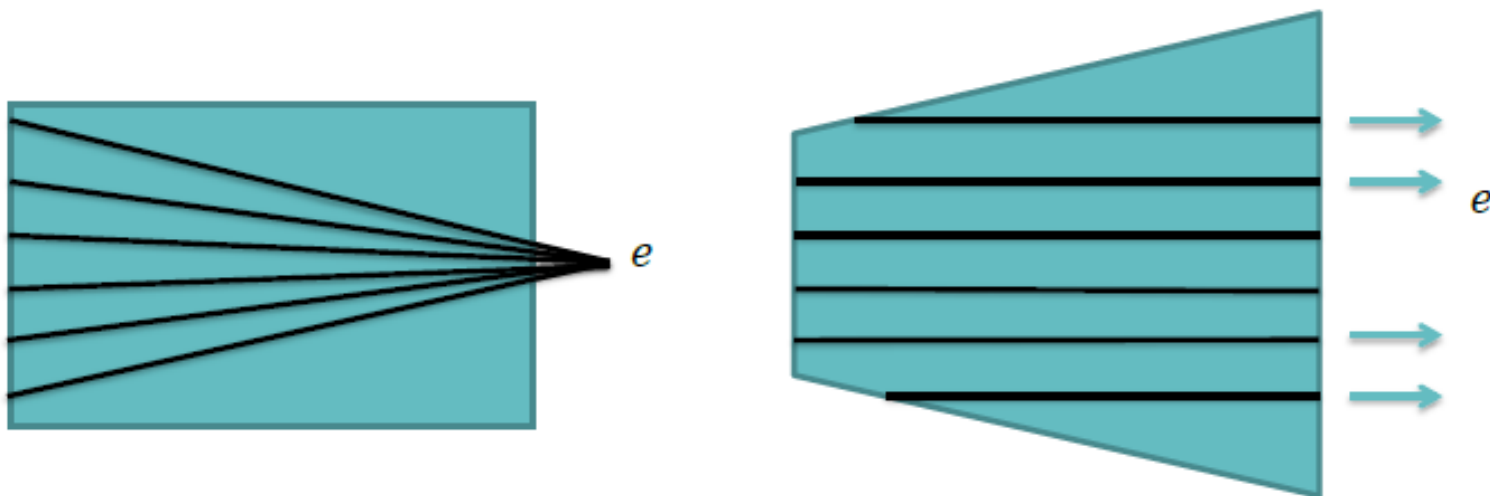
Stereo image rectification

- ↘ Reproject the image planes onto a common plane parallel to the line between the optical centers
- ↘ Requires a rotation around the optical center i.e. the focal point
- ↘ A rotation around the optical center is simply a 2D homography in the image



Basic idea

↘ We need to apply homography so that the epipolar lines become horizontal



↘ The epipole will be at infinity $e = [1 \ 0 \ 0]^T$

↘ Rectified image does not remain rectangular



Stereo rectification – Properties of homography matrix

$$\mathbf{H}\mathbf{e} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

- Assuming each row of \mathbf{H} to be a line

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \mathbf{h}_3^T \end{bmatrix}$$

$$\mathbf{H}\mathbf{e} = [\mathbf{h}_1^T \mathbf{e} \quad \mathbf{h}_2^T \mathbf{e} \quad \mathbf{h}_3^T \mathbf{e}]^T = [1 \quad 0 \quad 0]^T$$

- Lines \mathbf{h}_2 and \mathbf{h}_3 pass through the epipole \mathbf{e}
- The first one is irrelevant



Stereo rectification – Properties of homography matrix

- ↘ Let H and H' be the homographies to be applied to rectify the two images
- ↘ Lines h'_2 and h'_3 must pass through epipole e'
- ↘ Lines h_2 and h'_2 , and h_3 and h'_3 are corresponding epipolar lines
- ↘ Lines h_3 and h'_3 define the rectifying plane

Remember that the pair of homography matrices are not unique
i.e. many pairs are possible.

Need to find the pair that minimizes image distortion.



Stereo image rectification

↘ Assuming extrinsic parameter R and T are known

$$R_l = \begin{bmatrix} e_1^T \\ e_2^T \\ e_3^T \end{bmatrix}$$

$$e_1 = \frac{T}{\|T\|}$$

$$e_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} \begin{bmatrix} -T_y \\ T_x \\ 0 \end{bmatrix}$$

$$e_3 = e_1 \times e_2$$



Algorithm – stereo image rectification

- ↘ Computer R_l

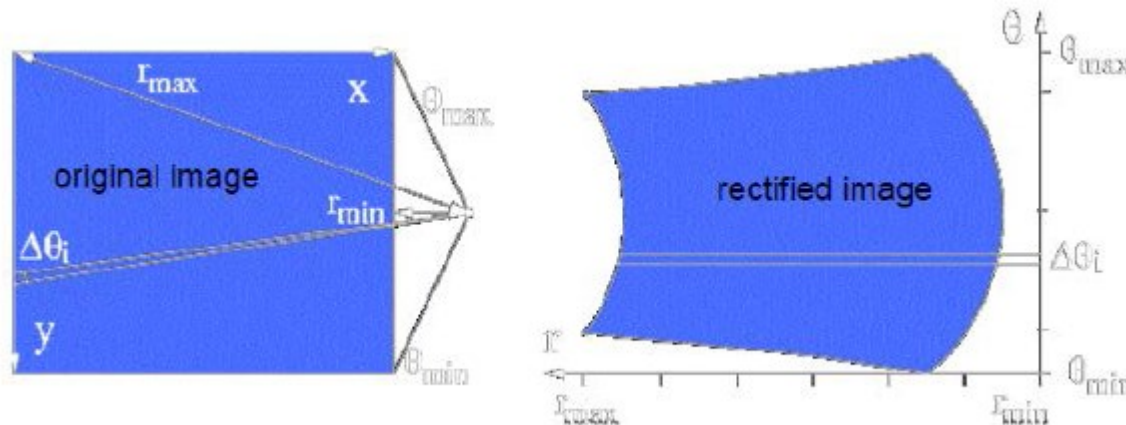
- ↘ $R_r = RR_l$

- ↘ For each left point $p = [x \ y \ f]^T$
 - Compute $R_l p = [x' \ y' \ z']^T$
 - Compute $p' = f/z'[x' \ y' \ z']^T$

- ↘ Repeat the above for the right camera point using R_r

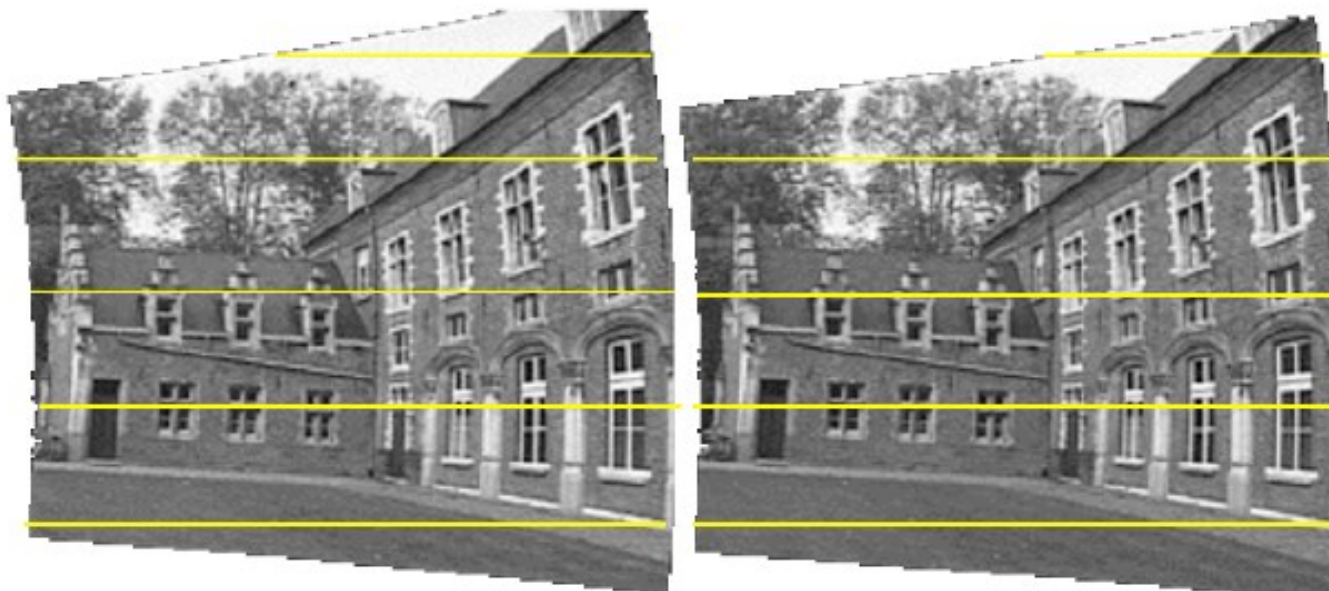
A better approach

- ↘ The above conventional approach does not work when the epipoles are inside (or very close to) the image
- ↘ For example a camera translating forward
- ↘ Solution: Perform polar rectification around the epipoles





Polar rectification example





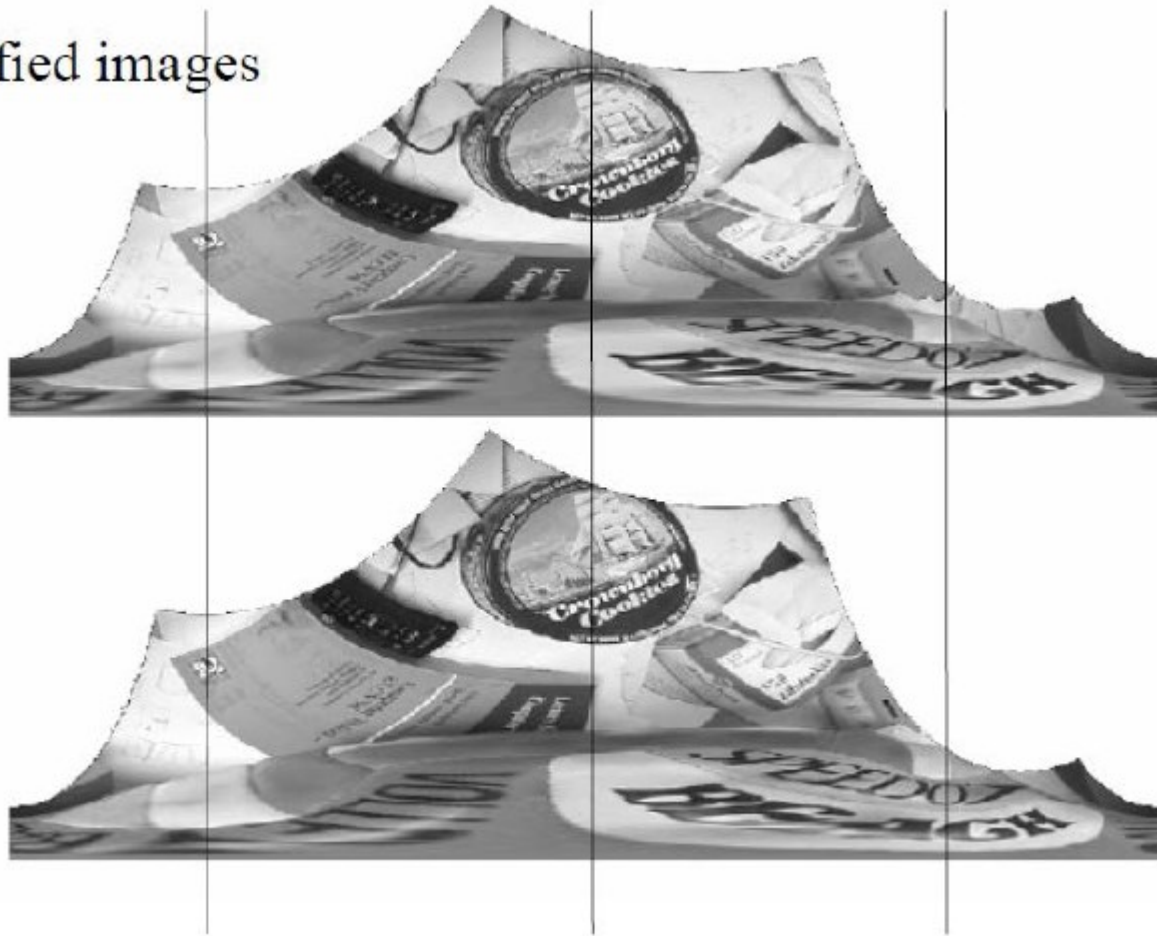
Another example



Input images



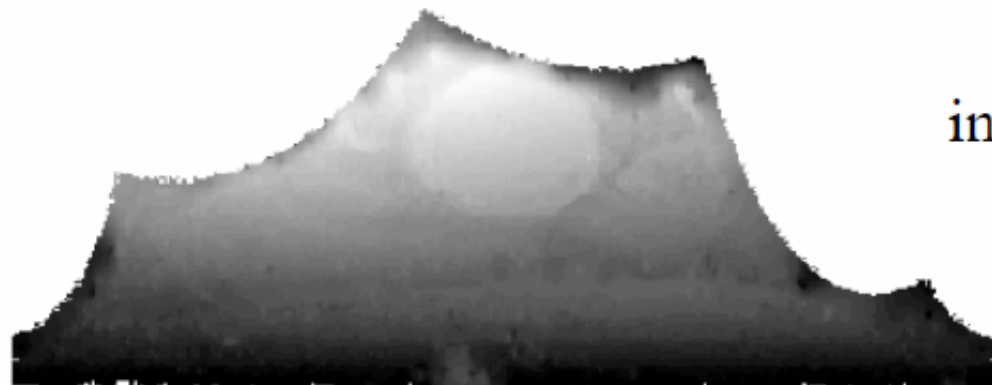
Rectified images



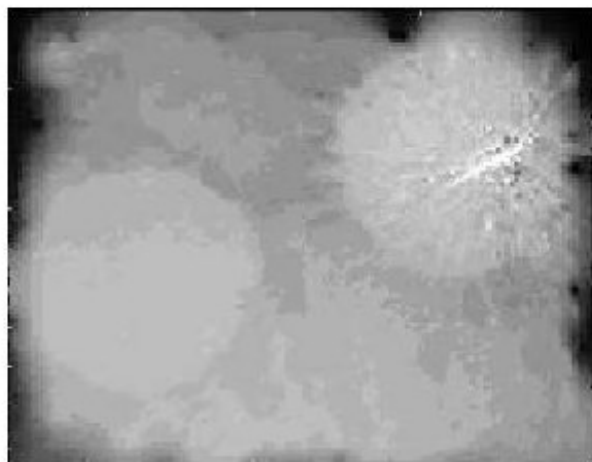


Depth Maps

sparse



interpolated



Depth map in
pixel coords



Views of texture mapped
depth surface



Correspondence problem

- ↘ What we have discussed so far reduces the correspondence search to the epipolar line
- ↘ We made the epipolar lines horizontal for simplicity
- ↘ To find corresponding points in two (rectified) image pairs is still a challenging problem due to
 - Lack of texture
 - Repetitive texture
 - Occlusions around the object boundaries
 - The selection of optimal correspondence search window is difficult



Putting constraints on correspondence search

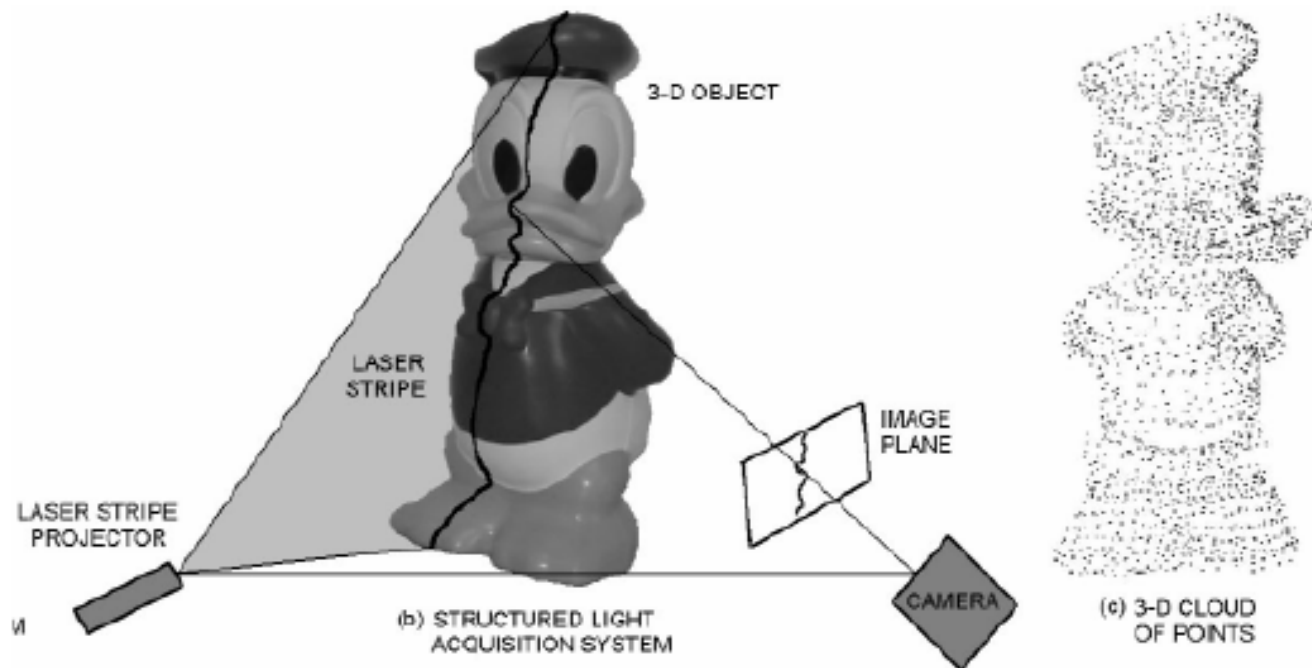
- ↘ **Epipolar:** Corresponding points must lie on corresponding epipolar lines
- ↘ **Uniqueness:** A point in one image can have only one corresponding point in the other image
- ↘ **Ordering:** Points in the second image must appear in the same order as in the first image
- ↘ **Smoothness:** Disparity values are expected to change slowly except at object boundaries



Perfecting correspondences

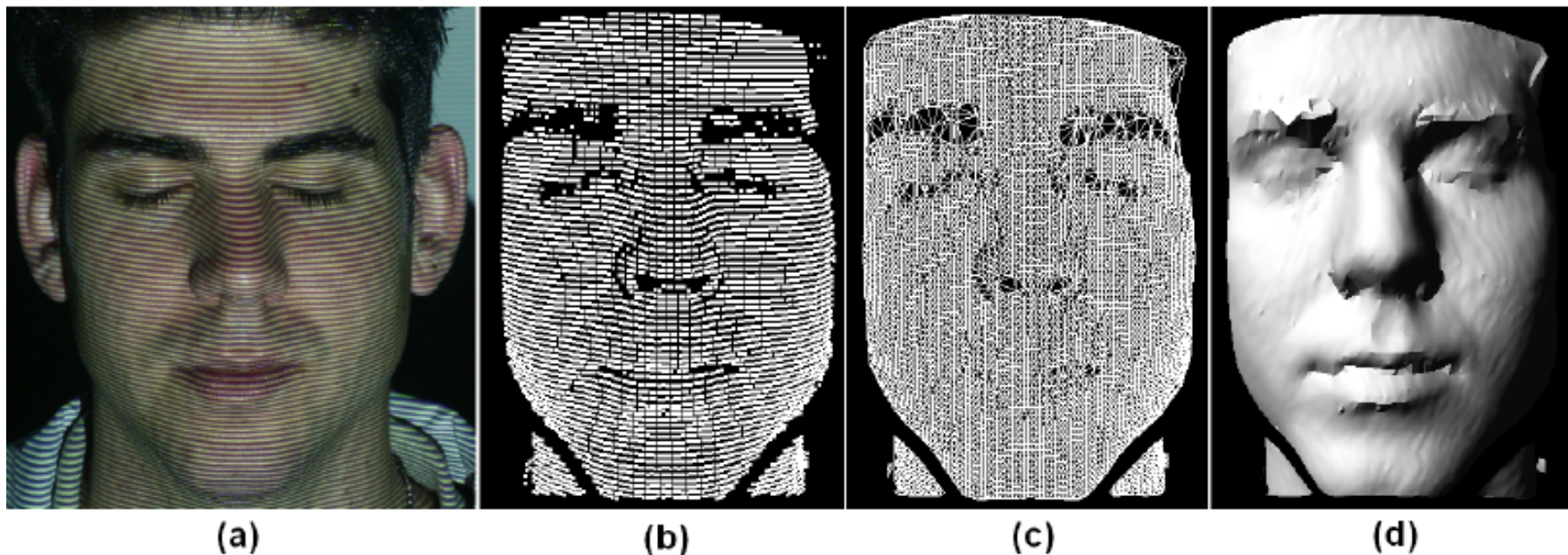
- ↘ Active stereo can make the correspondence search more accurate
- ↘ The idea is to project coded light onto the scene
- ↘ The code determines which points correspond in the two images
- ↘ In a basic setting, we need one camera and one projector
- ↘ The projector can possibly illuminate the entire scene at once OR scan a line through it

Laser stripe scanner



Structured light system (time coding)

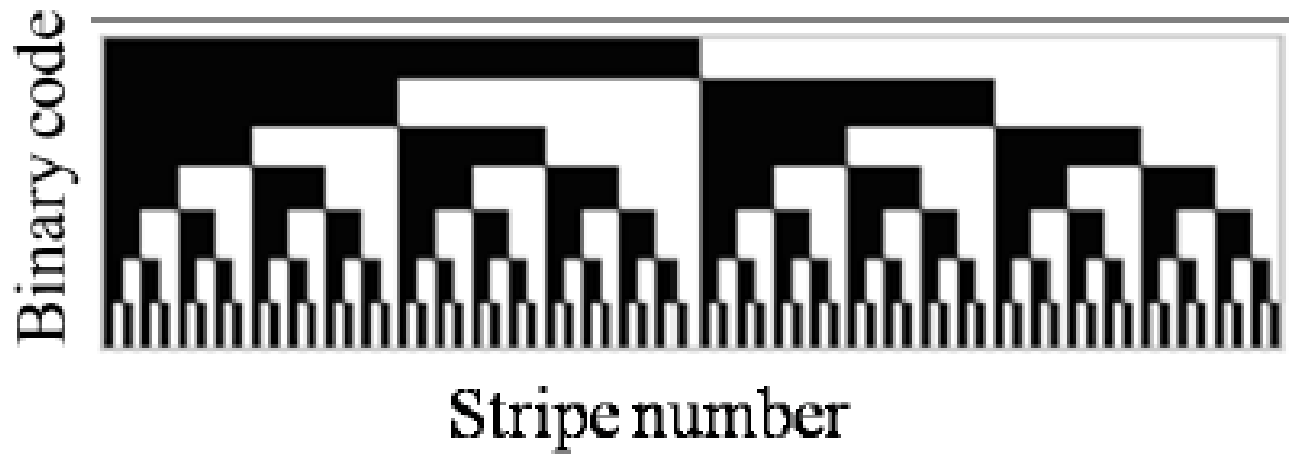
- ↘ Project light stripes onto the scene
- ↘ Calibrate the projector
- ↘ Code the stripe in the time domain using a binary pattern
- ↘ Need $\log_2 N$ images where N is the number of possible codes



Mian et al, "3D face recognition by matching shape descriptors", IVCNZ, 2004.



Binary code





Projector calibration

- Need to find the equation of every sheet (plane) of light. The equation of light sheet i is given by

$$a_i X + b_i Y + c_i Z + d_i = 0$$

- Given an image point (u, v) , it must satisfy the camera projection equation:

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} su \\ sv \\ s \end{bmatrix}$$

for some unknown X, Y, Z .

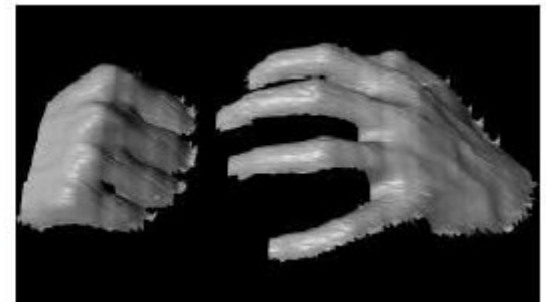
- If (u, v) has been identified to be illuminated by the i^{th} sheet of light then the unknowns the above two equations. Thus,

$$\begin{bmatrix} q_{11} - uq_{31} & q_{12} - uq_{32} & q_{13} - uq_{33} \\ q_{21} - vq_{31} & q_{22} - vq_{32} & q_{23} - vq_{33} \\ a_i & b_i & c_i \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} uq_{34} - q_{14} \\ vq_{34} - q_{24} \\ -d_i \end{bmatrix}$$

which is again a linear problem.

Structured light – colour coding

- Similar to the previous technique except that the stripes are colour coded
- Colour codes are such that they are very different in the local neighborhood
- Single shot (image) acquisition of 3D shape is possible





Structured light – spatial pattern

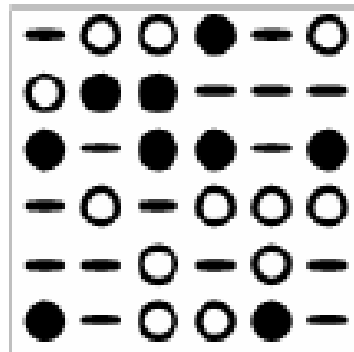
- Rather than colour pattern, use spatial pattern
- Single shot acquisition of 3D
- Can use monochrome camera
- Can use infra-red so it not visible to the human eye





Example: Spatial pattern

↘ The spatial pattern is unique within a neighborhood





Summary

- ↘ We learned about stereo vision and epipolar geometry
- ↘ We learned the essential and fundamental matrices and their use in stereo reconstruction
- ↘ We learned two techniques to perform image rectification for stereo
- ↘ We learned how to calculate the 3D coordinates of a point using stereo
- ↘ We learned about active stereo and how to calibrate a projector

Acknowledgement: Material for this lecture was taken from Rober Collins Computer Vision course, and previous lectures of Du Huynh and Peter Kovesi.