Lecture 11 and 12: Stereo Vision

Dr. V Masilamani
masila@iiitdm.ac.in

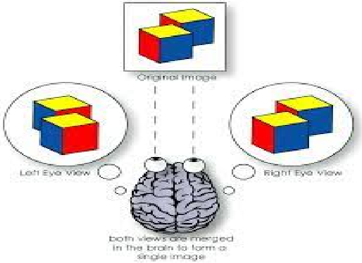
Department of Computer Science and Engineering
IIITDM Kancheepuram
Chennai-127

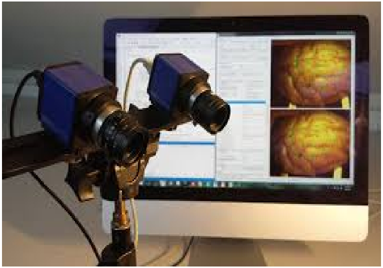
Overview

1. Goal of stereo vision
2. Visual cues for 3D
3. Estimating depth with stereo

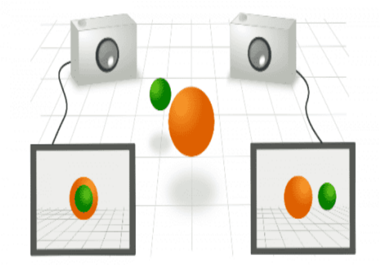
Need of Stereo Vision

To mimic human vision system

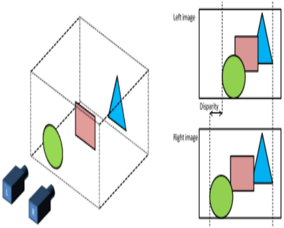




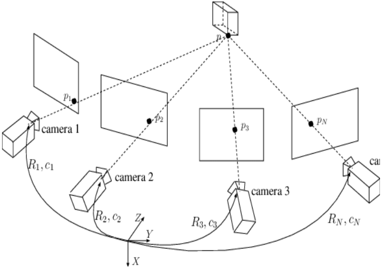
What extra information captured by two cameras, compared to one



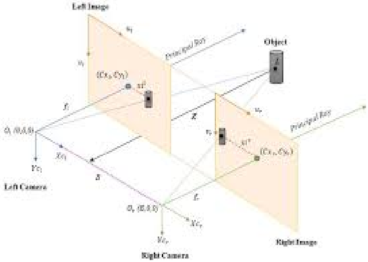
Need of Stereo Vision

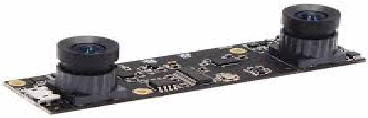


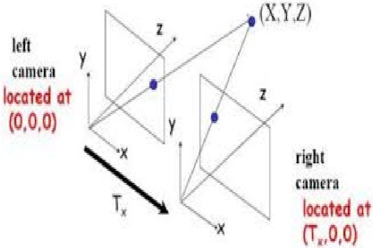
1. Goal of stereo Vision: The recovery of the 3D structure(3D points, shape etc) of a scene using two or more images of the 3D scene, each acquired from a different viewpoint in space.

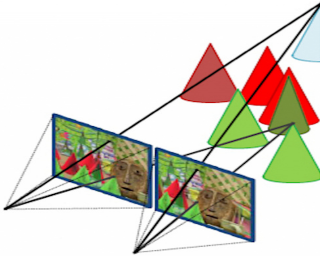


2. The term binocular vision is used when two cameras are employed.


Left Image
Object
Right Image
Left Camera
Right Camera




(X,Y,Z)
left
camera
located at
(0,0,0)
right
camera
located at
(T_x,0,0)
T_x



Visual cues for 3D

When we look at image, what properties indicate the differences in depth or provide hints about object's shape?

Shading

Textures

Focus

Motion

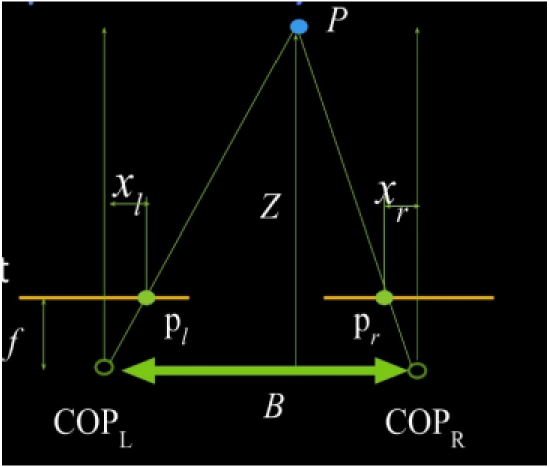
Shape from X

- X = shading, texture, focus, motion, ...
- We'll focus on the motion cue

Estimating depth with stereo

- Stereo : Shape from motion between the two views
- We need to consider
  - Reconstruction
  - Correspondence
- Assuming the following constraints, let us find depth
  - CCS of Camara 1 is aligned with WCS
  - X and Y axes of CCS1 and CCS2 are the same
  - Origin of CCS 2 is displaced from the origin of CCS 1 by $(B, 0, 0)$

Geometry for a stereo system

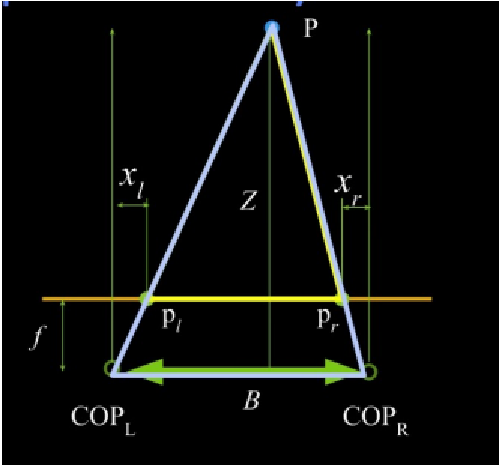


- B is baseline width,distance between centres of the cameras
- **Defn of depth:** The disance between a point p and the baseline is called as depth of p

Observe that:

- Two Image planes are coplanar
- Image planes at the front of image camera
- Origins are at the center of the planes
- Displacement(from origin) is positive in left image and negative in right image
- Z is the depth to be obtained from the two formed images.
- f is the focal length.

Geometry for a stereo system (cont.)



Geometry for a stereo system (cont.)

- Similar triangles (P_l,P,P_r) and (COP_L,P,COP_R)

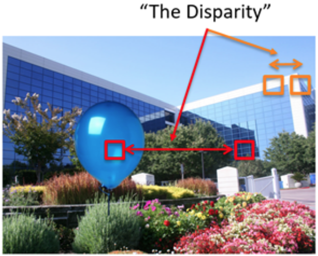
$$\frac{B \; - \; x_l + x_r}{Z \; - \; f} = \frac{B}{Z}$$

$$Z = \frac{f \; B}{x_l - x_r}$$

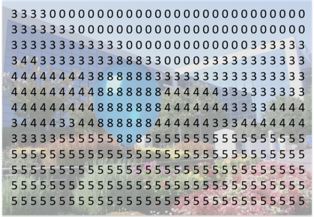
- x_l-x_r is called as disparity
- Depth is inversely proportional to Disparity

Disparity map

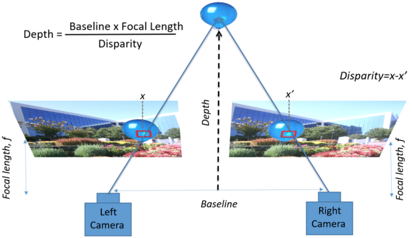
Left camera



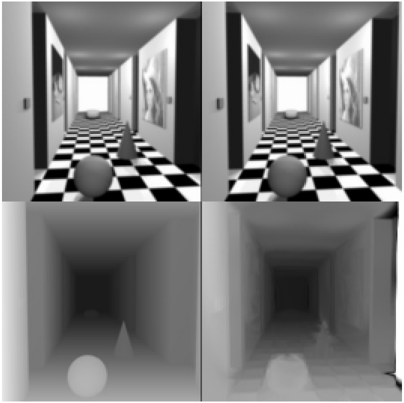
Right camera



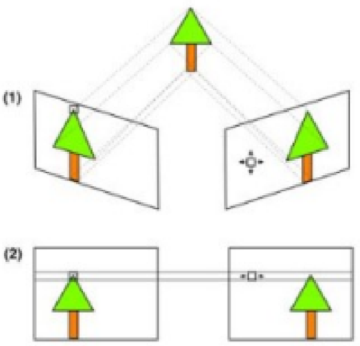
Disparity Map referenced to Right camera



Disparity map

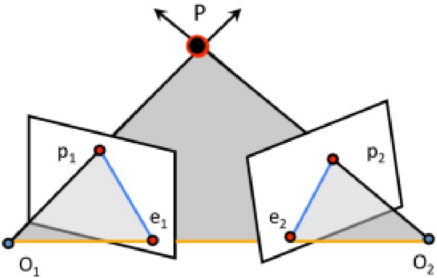


Stereo Correspondence



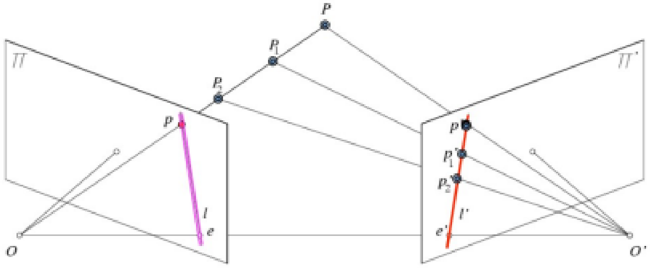
Search problem: Given an element in the left image, we search for the element in the right image. This involves two decisions:

- Which image element to match
- Which similarity measure to adopt



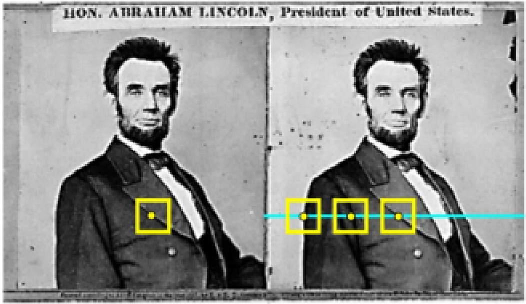
- The standard epipolar geometry setup involves two cameras observing the same 3D point P, whose projection in each of the image planes is located at p1 and p2.
- e1 and e2 are called epipoles.
- Epipolar line is the intersection of an epipolar plane with the image plane

Epipolar Constraint



- Potential matches for p have to lie on the corresponding epipolar line r
- Potential matches for p' have to lie on the corresponding epipolar line l

Why is the epipolar constraint useful?



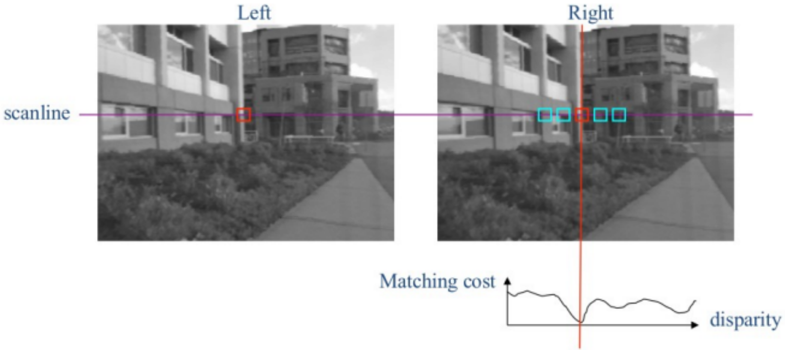
The epipolar constraint reduces the correspondence problem to a 1D search along the epipolar line.

Correspondence problem

Beyond the constraint by epipolar geometry , there are other constraints to help identify corresponding points.

- **Similarity** - Image patch from the left should match with the right
- **Uniqueness** - There is no more than one match for the pixel in right image
- **Ordering** - If pixels go a,b,c in left, they go a,b,c in right
- **Disparity Gradient is limited** - Depth doesn't change too quickly

Correspondence search with similarity constraint



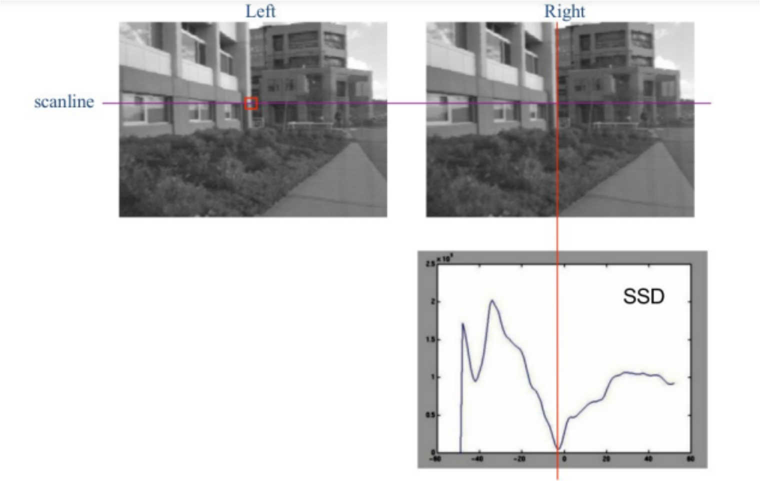
- Sum of Squares difference(SSD)
- Normalized Correlation

$$r_{ij} = \frac{\sum_m \sum_n [f(m+i, n+j) - \bar{f}][g(m+i, n+i) - \bar{g}]}{\sqrt{\sum_m \sum_n (f(m+i, n+j) - \bar{f})^2 \sum_m \sum_n (g(m+i, n+i) - \bar{g})^2)}}$$

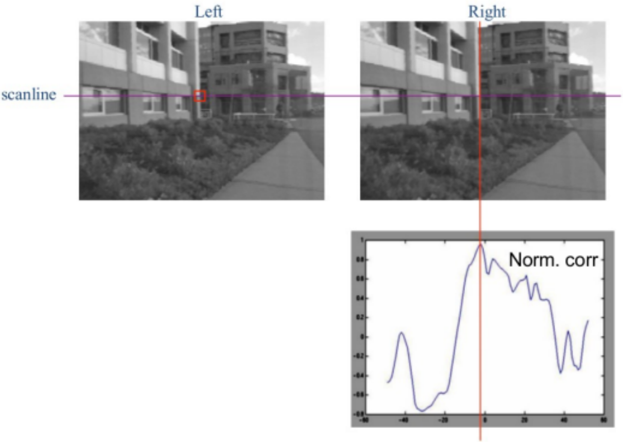
- Mutual Information

$$I(X;Y) = \int_y \int_x P(x,y) log\left(\frac{P(x,y)}{P(x)P(y)}\right) dxdy$$

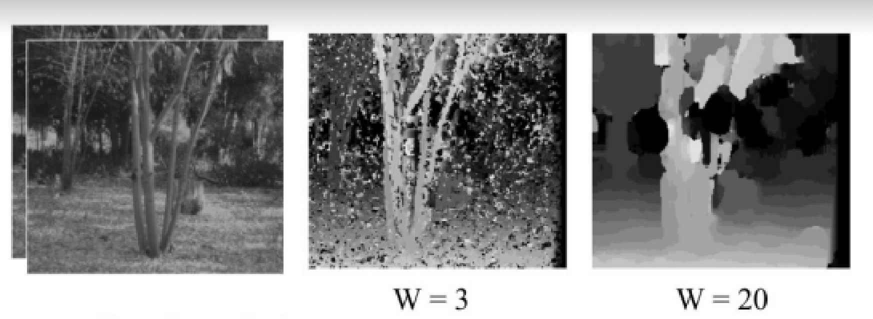
SSD



Normalized Correlation



Effect of Window Size

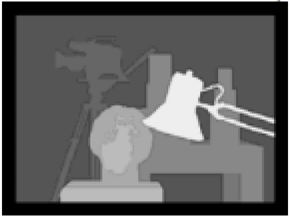


- Smaller window
  - More detail
  - Less noise
- Larger window
  - Smoother disparity maps
  - Less detail

Disparity Map

(a)

(b)



(c)

Challenges in Stereo Vision

- Low-contrast; Texture less image regions
- Occlusions
- Large base lines
- Camera calibration errors etc.,