

---

# Stereo Vision – A simple system

Dr. Gerhard Roth  
Winter 2012

# Stereo

---

- Stereo
  - Ability to infer information on the 3-D structure and distance of a scene from two or more images taken from different viewpoints
  - Humans use only two eyes/images (try thumb trick)
- Two important problems in stereo
  - Correspondence and reconstruction
- Correspondence
  - What parts of left and right images are parts of same object?
- Reconstruction
  - Given correspondences in left and right images, and possibly information on stereo geometry, compute the 3D location and structure of the observed objects

# What is stereo vision?

---

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image



Left image



Right image

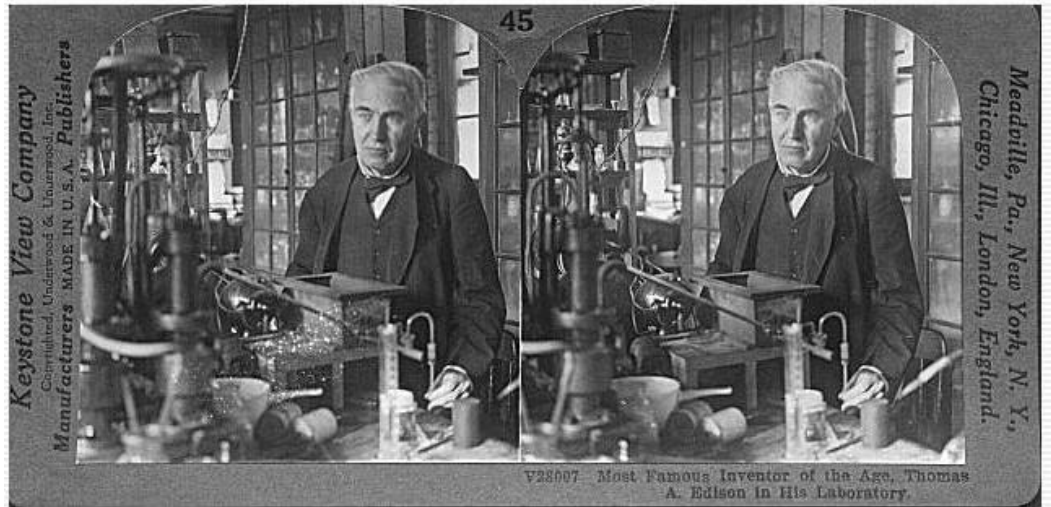
Dense depth map



# What is stereo vision?

---

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image
  - Humans can do it



Stereograms: Invented by Sir Charles Wheatstone, 1838

# What is stereo vision?

---

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image



Autostereograms: [www.magiceye.com](http://www.magiceye.com)

# What is stereo vision?

---

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image



Autostereograms: [www.magiceye.com](http://www.magiceye.com)

# Anaglyphs – One way for stereo

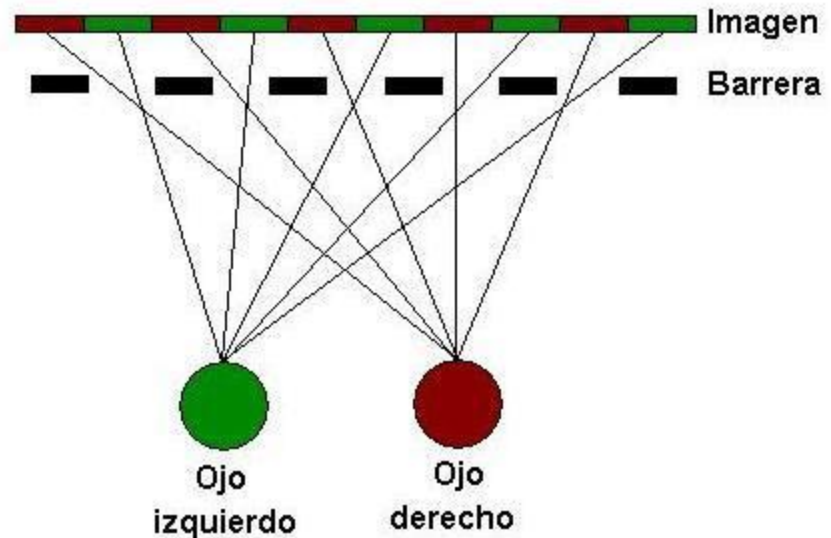
---



Many slides adapted from Steve Seitz

# Nintendo 3DS – 3d vision for gaming

---



- Autostereogram with parallax barrier – each eye gets different view
- But you must be at proper location (right in the middle) to see 3d
- First mass produced hand held 3d autostereogram gaming system

# Application of stereo: Robotic exploration

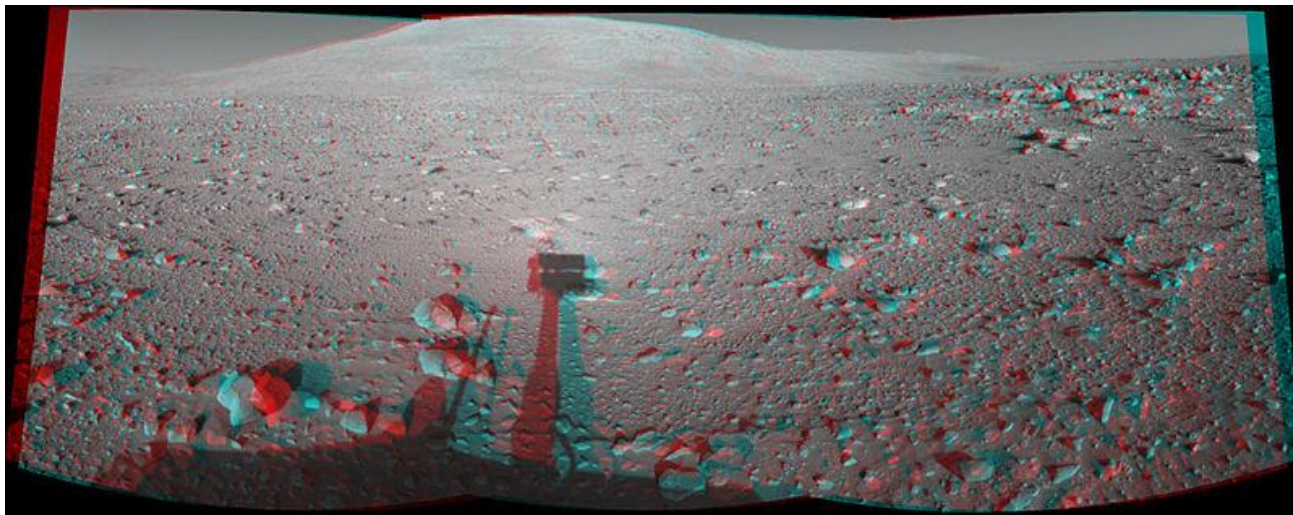
---



[Nomad robot](#) searches for meteorites  
in Antarctica



Real-time stereo on Mars



# Application: View Interpolation

---



Right Image

# Application: View Interpolation

---



Left Image

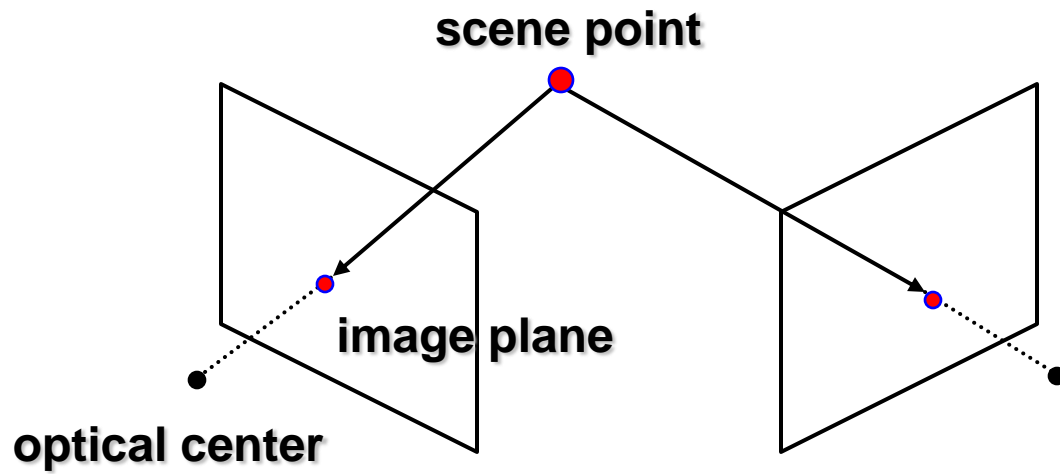
# Application: View Interpolation

---



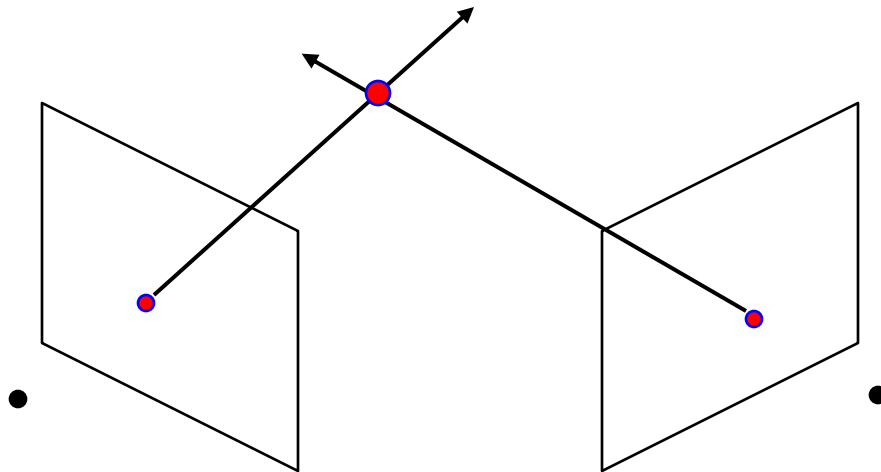
# Stereo

---



# Stereo

---

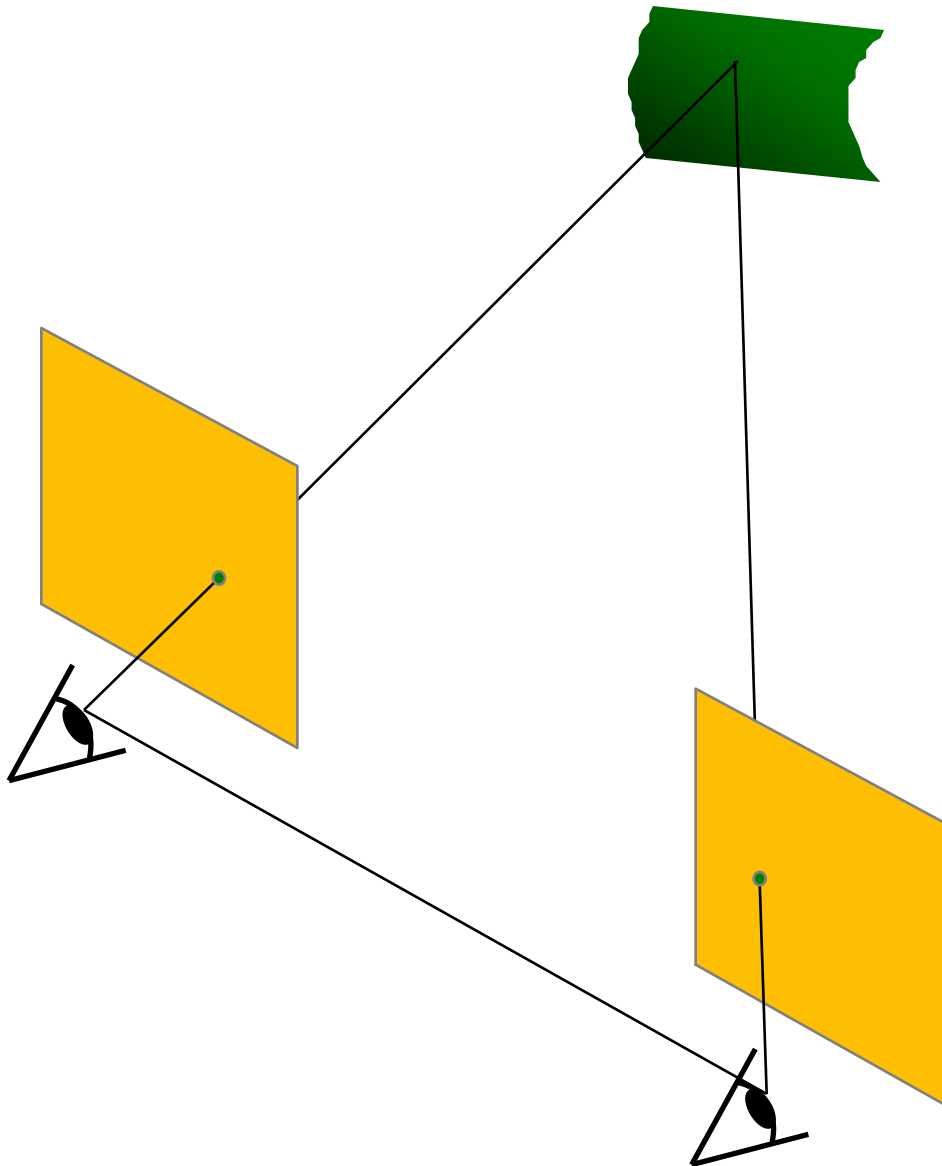


## Basic Principle: Triangulation

- Gives reconstruction as intersection of two rays
- Requires
  - Camera calibration
  - Point correspondence

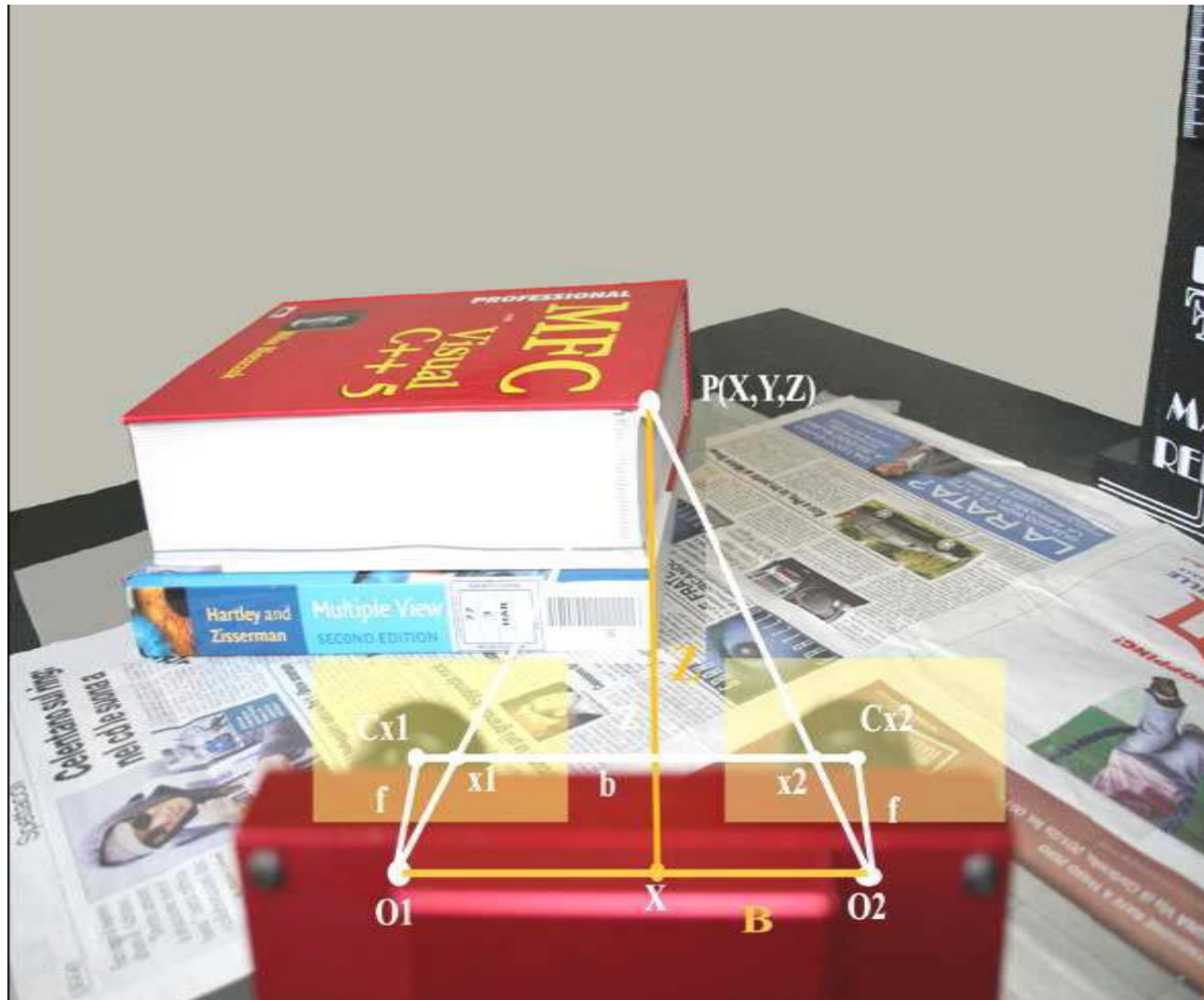
# Simplest Case: Parallel images

---



- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same

# Simple Stereo Camera



# Simple Stereo System

---

- Left and right image planes are coplanar
  - Represented by  $I_L$  and  $I_R$
- So this means that all matching features are on the same horizontal line
  - So the search for matches can proceed on the same line



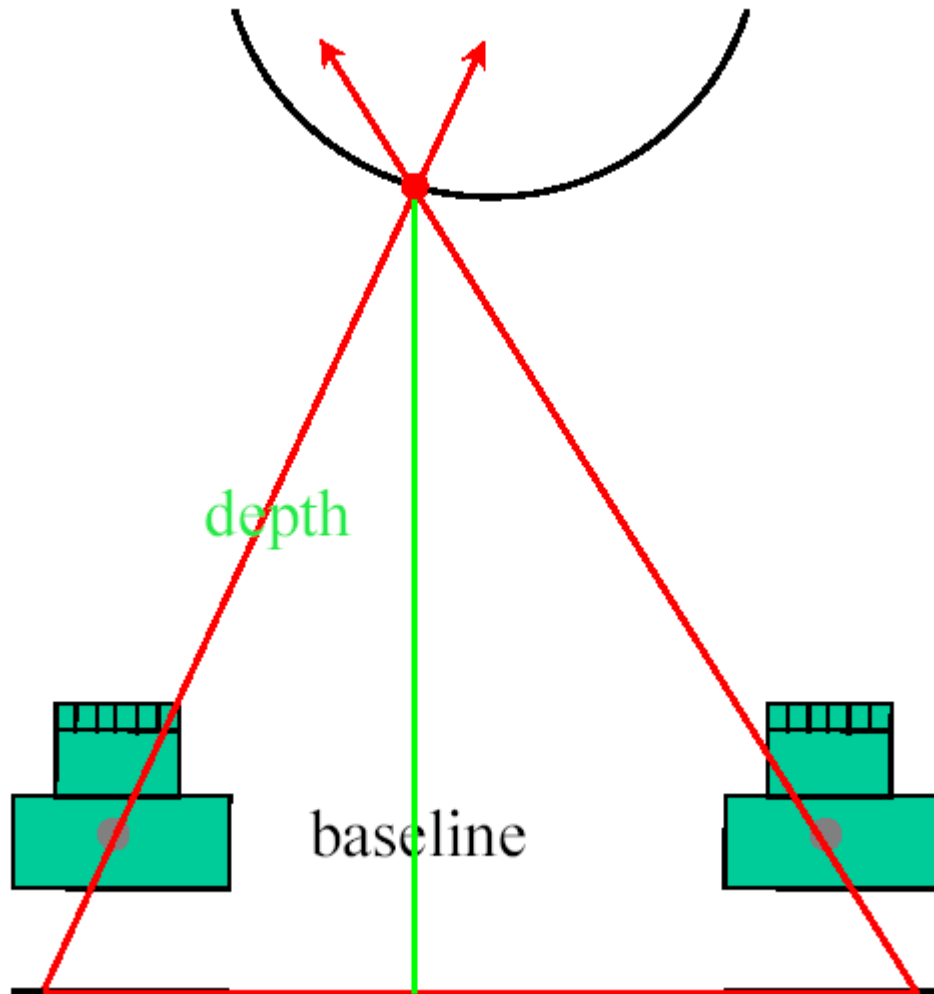
# Simple Stereo System (2D)

---

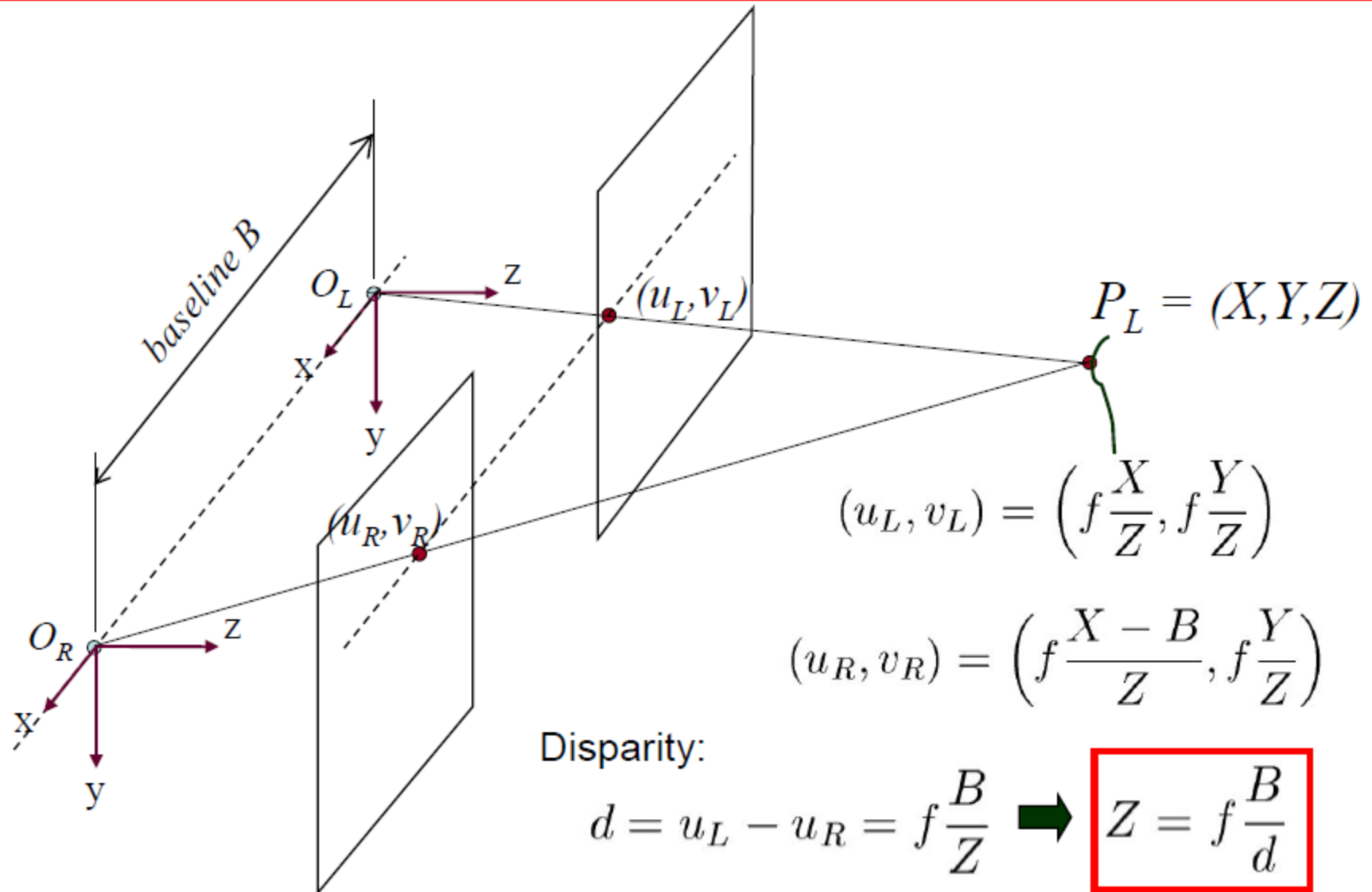
- Distance between centers of projection is called the baseline  $T$
- Centers of projection of cameras  $C_L$  and  $C_R$
- Point  $P$  in 3D space projects to  $P_L$  and  $P_R$
- $X_L$  and  $X_R$  are co-ordinates of  $P_L$  and  $P_R$  with respect to principal points  $C_L$  and  $C_R$ 
  - These are camera co-ordinates in millimeters
- $Z$  is the difference between point  $P$  and the baseline
  - $Z$  is called the depth

# Simple Stereo System

---

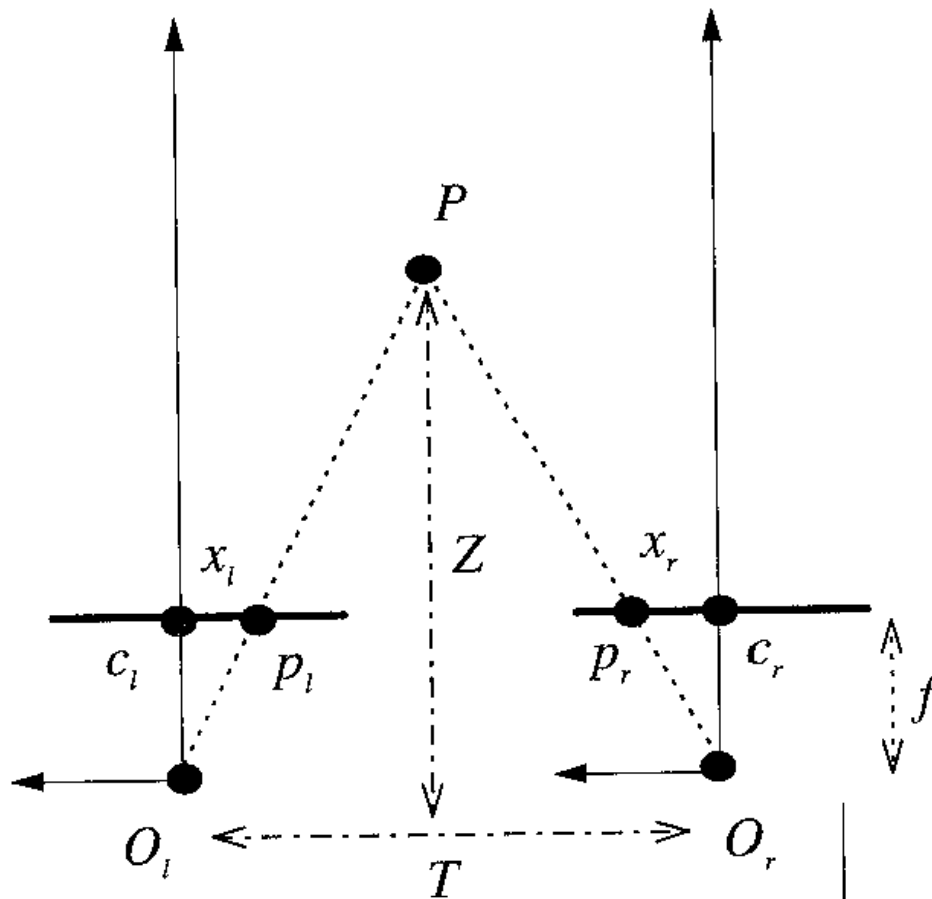


# Simple Stereo System



# Basic Stereo Derivations

---



Derive expression for  $Z$  as a function of  $x_l$ ,  $x_r$ ,  $f$  and  $B$

Here  $x_r$  is projection on rightmost camera looking out from cameras

# Stereo Derivations (camera coords)

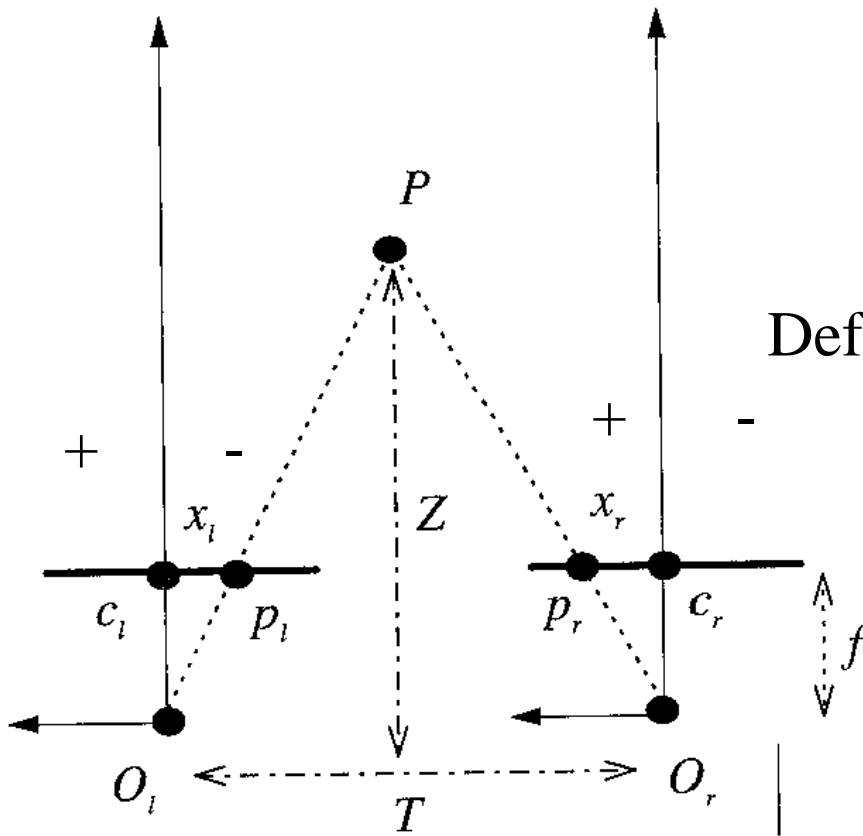
Similar triangles  
( $P_L, P, P_R$ ) and ( $O_L, P, O_R$ )

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

Define the disparity:  $d = x_r - x_l$

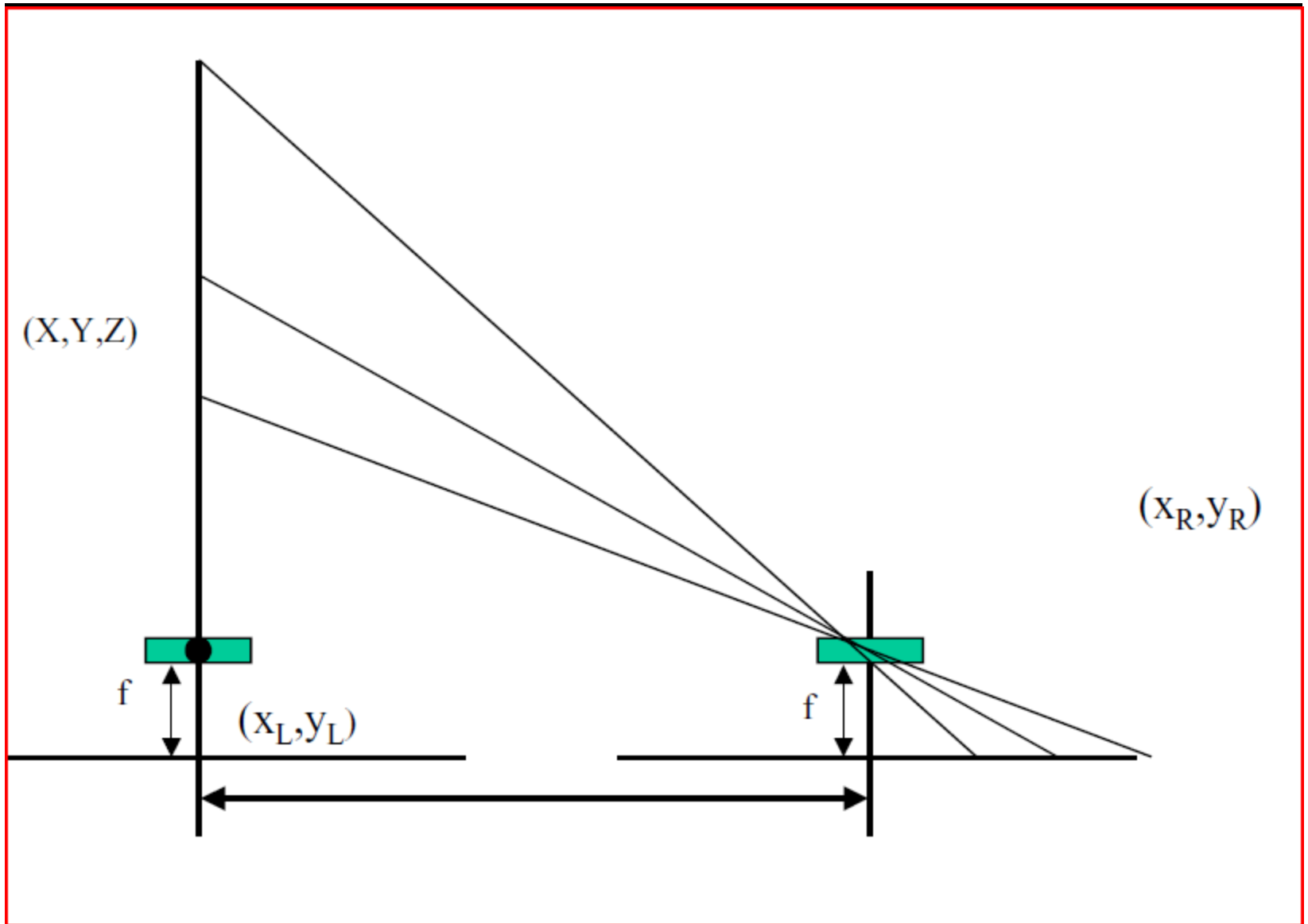
$$Z = f \frac{T}{d} \quad \text{Since } f, T \text{ are constant}$$

$$Z \propto \frac{1}{d}$$



Note that  $x_l$  is always to left of  $x_r$  so disparity  $d \geq 0$  in all cases

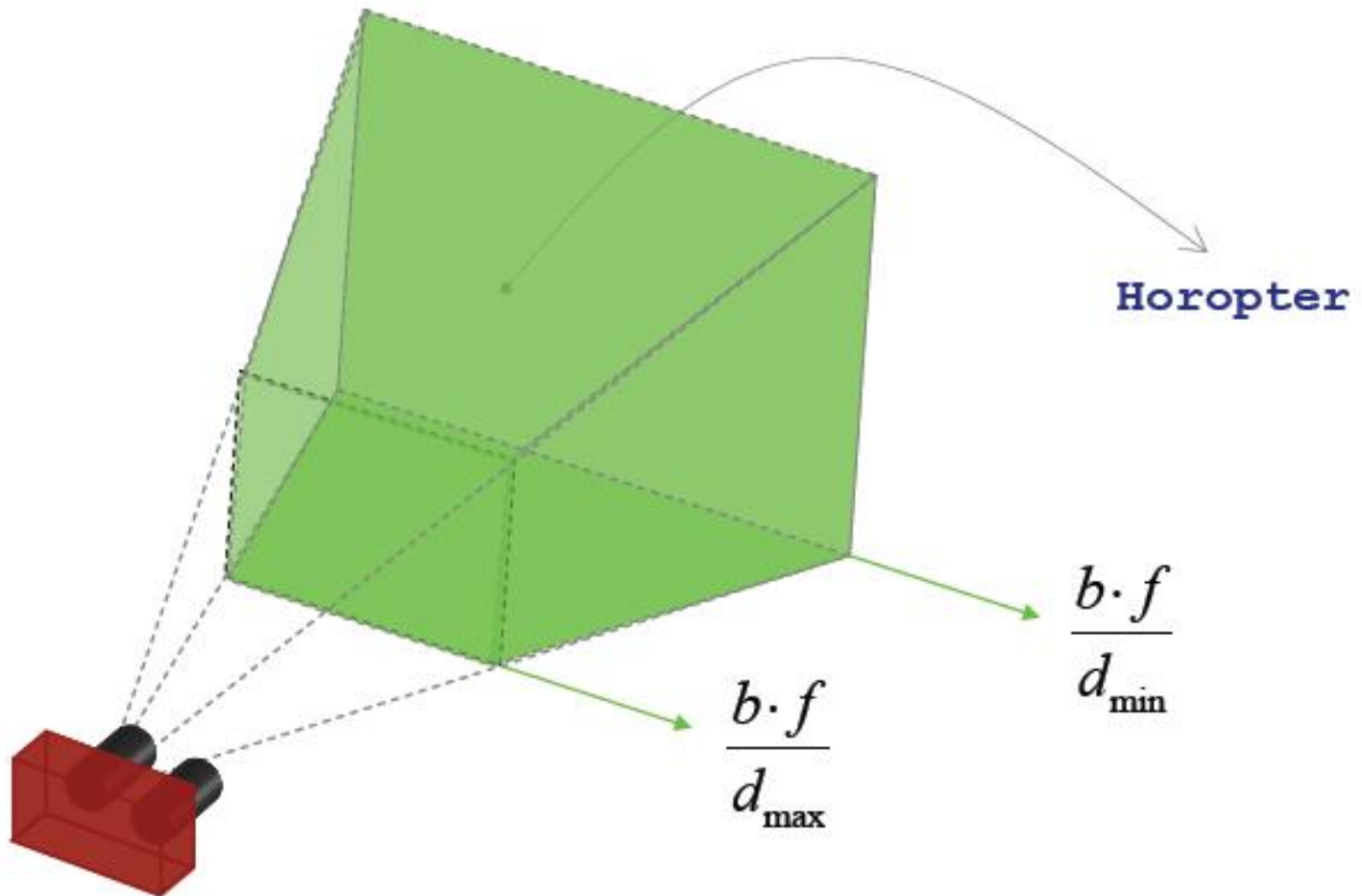
# Range Versus Disparity – $z$ fixed for $d$



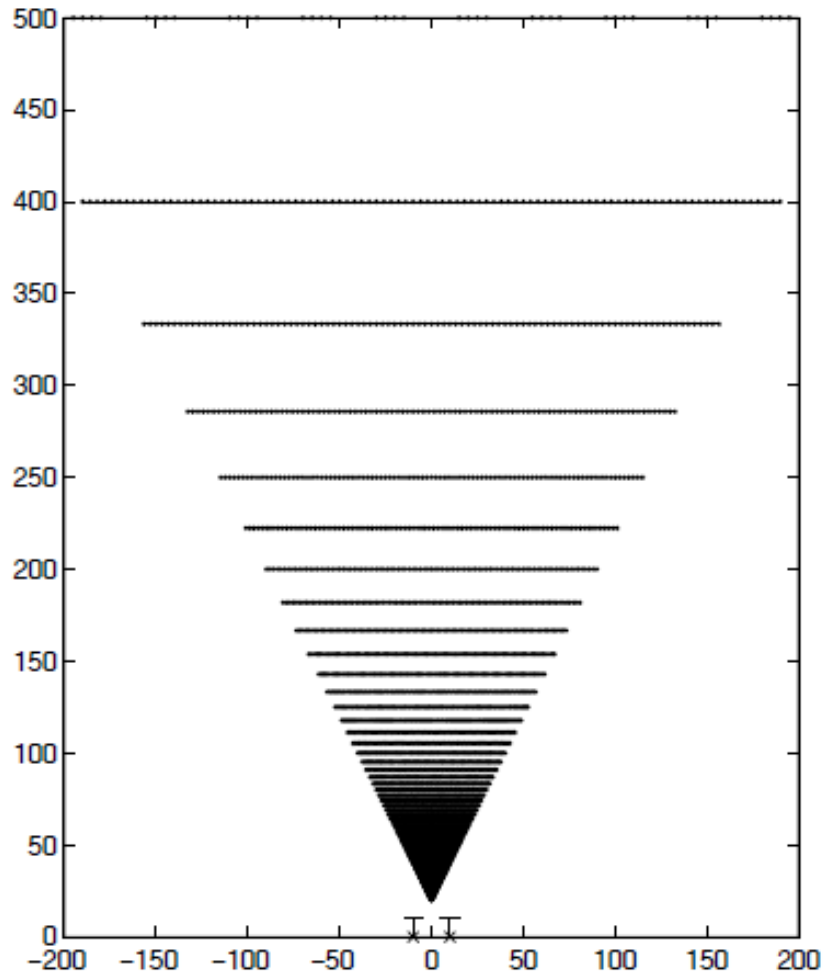
# Range field (in Z) of Stereo

---

b, f and d\_min, and d\_max define the range in Z



# Implicit Iso-Disparity for Simple Stereo



Lines where the disparity value has the same value (see stereo-res.jpg)

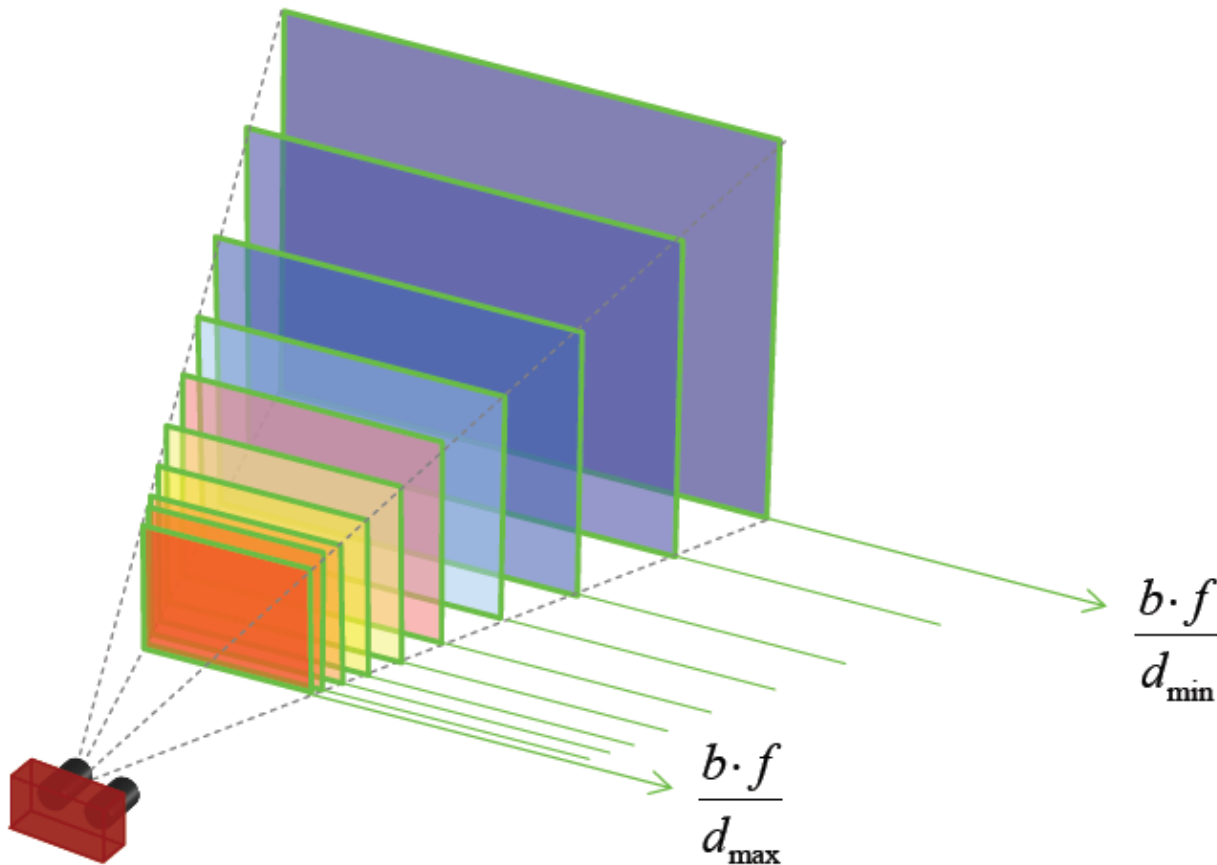
Notice the rapid Increase in spacing

Implication is that the resolution of stereo depth depends on the disparity

# Discrete Depth according to Disparity

---

Depth discretized into parallel planes, one for each possible disparity value, more accuracy requires sub-pixel precision – which is costly



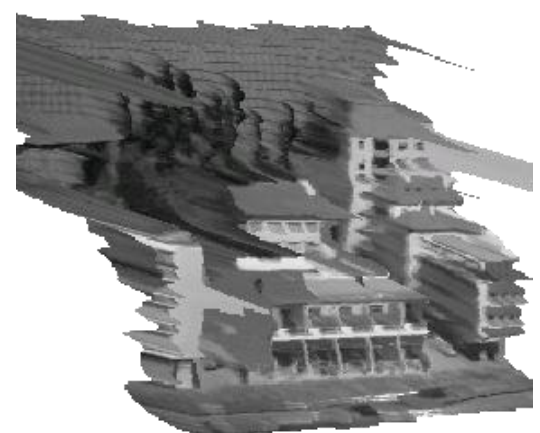
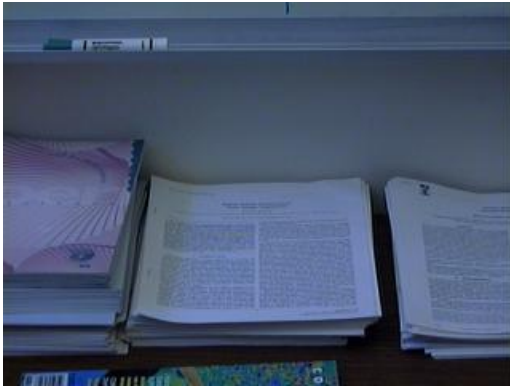
# Disparity Map

---

- Disparity  $d$  was in mm (camera co-ordinates)
- If  $x_l$  and  $x_r$  in pixels then  $d = x_l - x_r > 0$  but only for simple stereo (origin top left)
  - Usually disparity is stated as number of pixels  $d = ||x_l - x_r||$
  - For points at infinity the disparity  $d$  is zero
  - Maximum possible disparity  $d$  depends on how close we can get to the camera (there is a maximum value)
- Depth still inversely proportional to disparity
  - If we compute the disparity for the entire images then we have a disparity map (computed relative to the left image)
- Often show disparity in image form
  - Bright points have highest disparity (closest)
  - Dark points have lowest disparity (farthest)

# Disparity Map

---



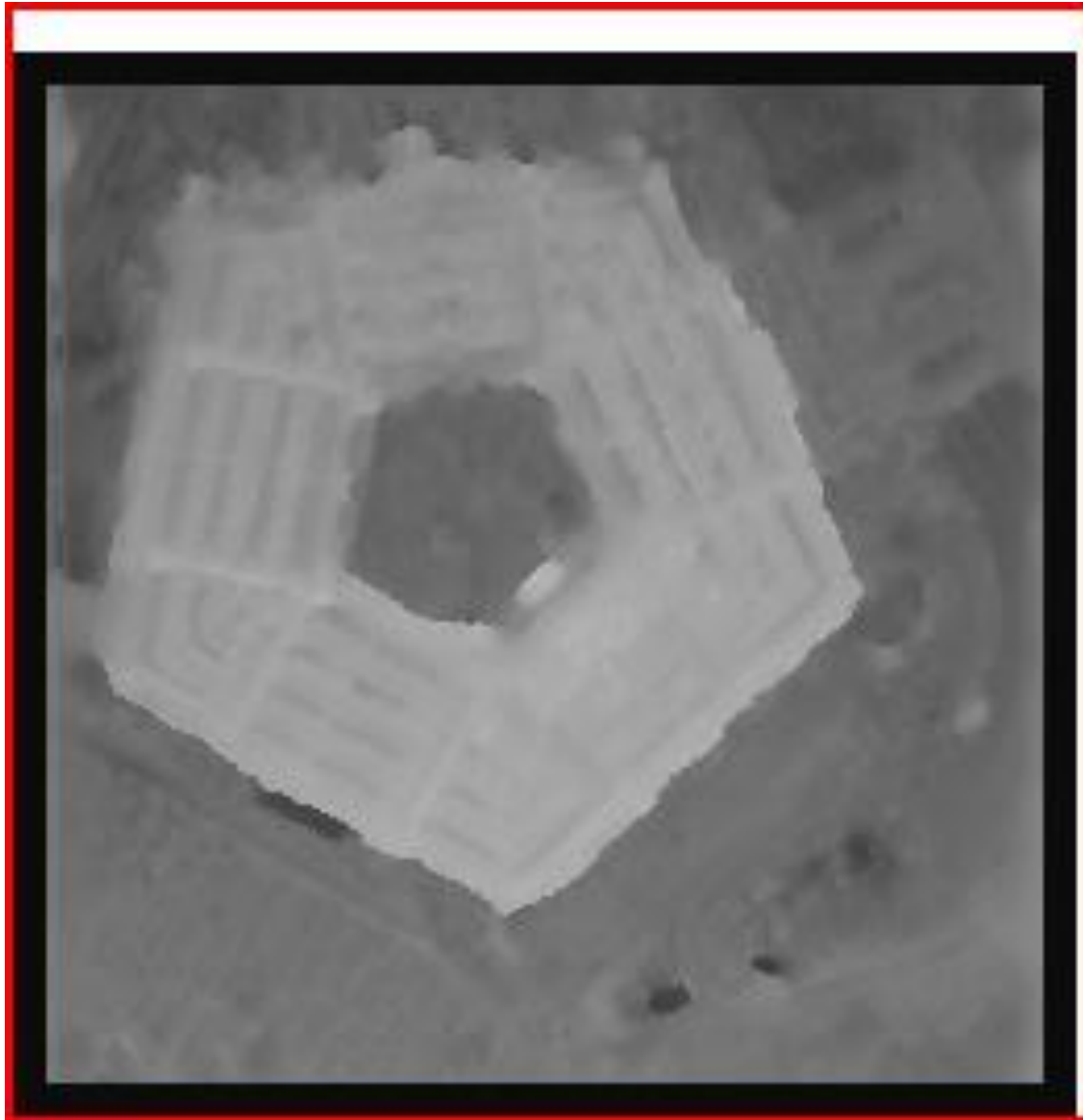
# Example Pentagon Stereo Pair

---



# Example Pentagon Disparity Map

---



# Pentagon example

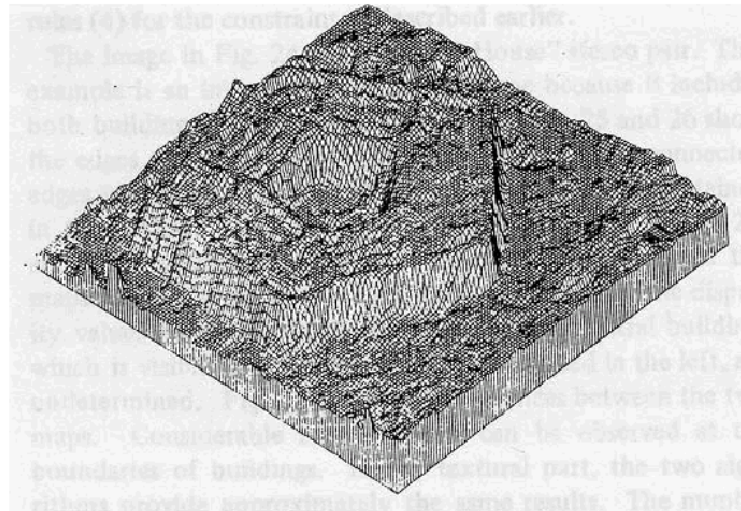
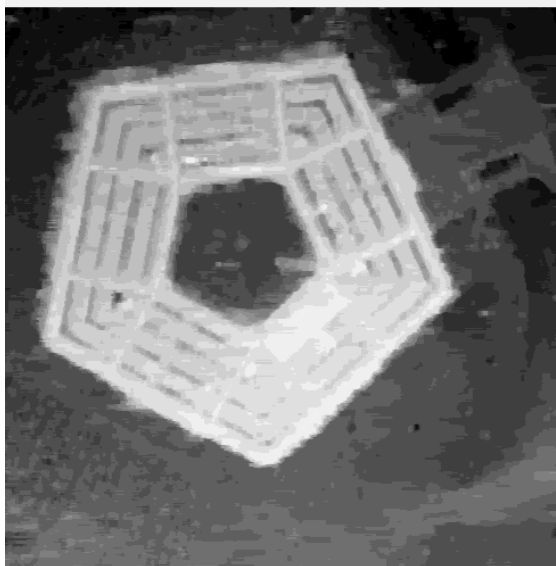
left image



right image



disparity map



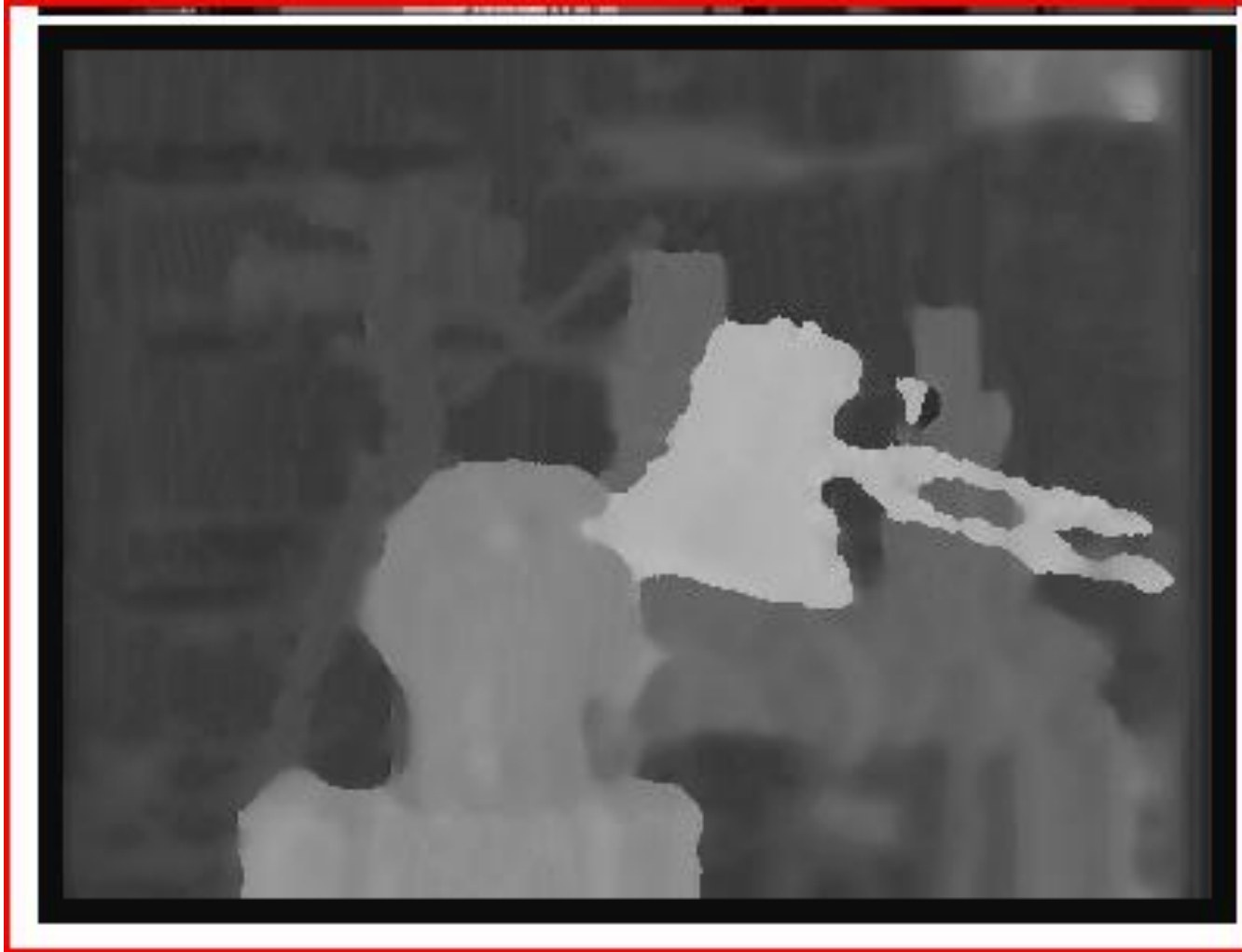
# Example Tskuba Stereo Pair

---



# Example Tskuba Disparity Map

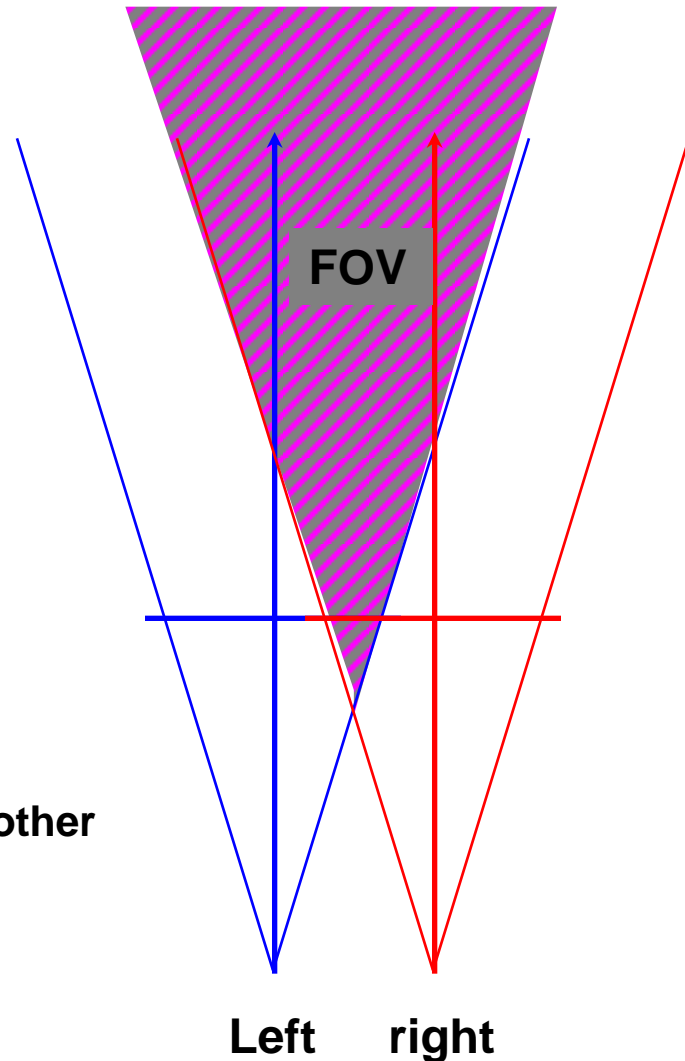
---



# Characteristics of Simple Stereo

---

- FOV is field of view of cameras
  - Overlap of the two cameras
- Baseline is a system parameter
  - It is also a tradeoff
- If B is the Baseline
  - Depth Error  $\propto 1/B$
- PROS of Longer baseline
  - **better depth estimation**
- CONS
  - **smaller common FOV**
  - **Correspondence harder due to increased chance of occlusion**
  - **Occlusion means that a feature is visible in one image but not in another because something occludes it**
  - **Occlusion more likely as baseline increases**



# Baseline Tradeoff

---

- **Short Baseline**
  - Better matches are likely (similarity constraint)
  - Fewer occlusions are likely (in general)
  - Less precision (depth not computed as accurately)
- **Larger Baseline**
  - Poorer matches are likely (similarity constraint)
  - More occlusions are likely (in general)
  - More precision (depth is computed more accurately)
- Use smallest baseline you need to get the depth precision that you want!

# Real-Time Stereo Systems

---

- There are a number of systems that can compute disparity maps
- In practice systems only work if there is texture in the regions that must be matched
- Often such systems return sparse depth
  - A few thousand images in regions where there is texture
  - Do some interpolation when there is no texture
- Point Grey research makes such a camera
  - A successful Canadian company
- Produces a variety of stereo cameras

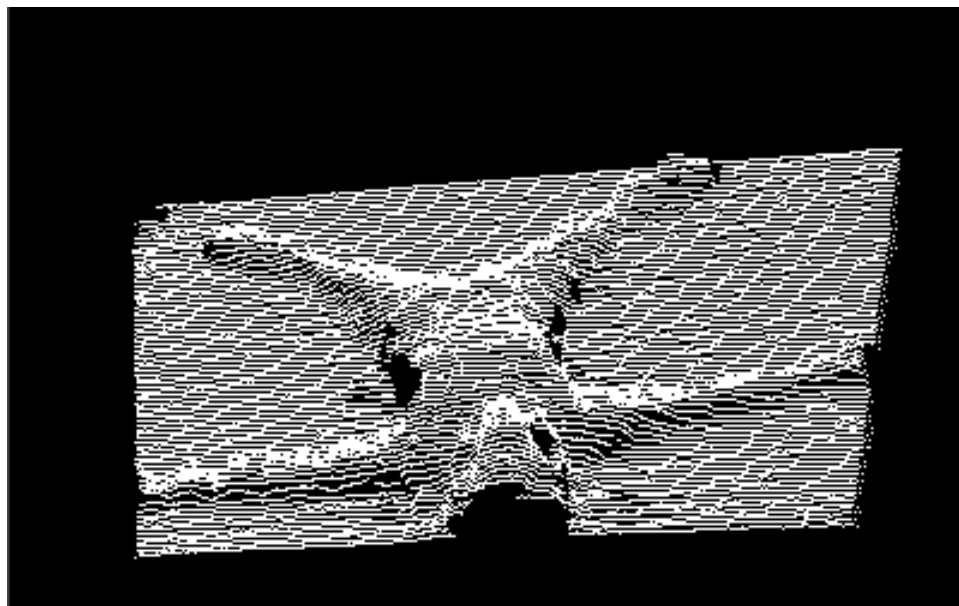
# BumbleBee

---



# Example image from BumbleBee

---



# Passive stereo systems

---

- If system only receives light rays and does not emit any radiation it is a passive systems
- Your visual system is passive
- Minus
  - More likely to get bad matches and false depth
  - Can not work in regions without any texture
  - Results depend on the ambient lighting
- Plus
  - Requires only two calibrated cameras and is simple
  - If cameras are in simple configuration (mechanically aligned) then there are well know algorithms to do matching and compute a disparity map in real-time

# Active stereo systems

---

- Emits radiation then it is an active system
- With two cameras an active stereo system
  - Simplest example is just use a laser pointer with stereo!
  - Can have active systems with just one camera!
  - An example is Kinect which has self-identifying patterns
- Minus
  - More complex hardware, active sensor can be dangerous
  - Need a way to spread active radiation over scene quickly
    - Sweep a laser beam, or use a grating to spread the laser out
  - Usually more complex calibration than a passive system
- Plus
  - Can work well in many different lighting conditions
  - When it returns depth we can be very confident that this depth is correct (fewer false matches)