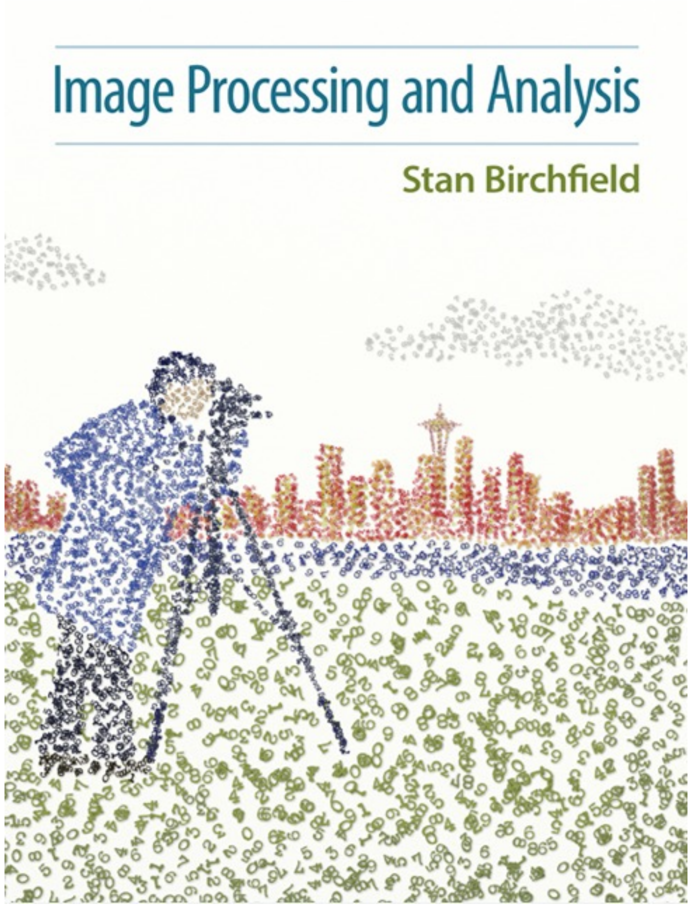
University of Stavanger

Prof. Kjersti Engan

ELE510 Image processing and computer vision

Stereopsis and correspondence problem, (chap 13.1, 13.2 Birchfield) 2023

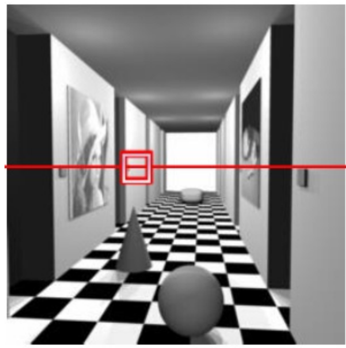


Parts in course presentations is material from Cengage learning. It can be used for teaching, and it can be share it with students on access controlled web-sites (Canvas) for use in THIS course. **But it should not be copied or distributed further in any way** (by you or anybody).

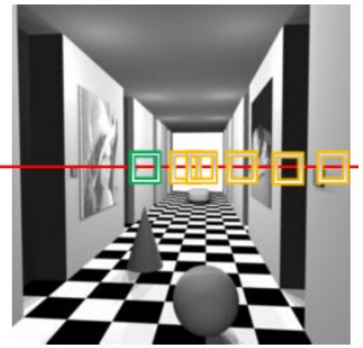
Human stereopsis and correspondence problem

Three points from the topic:

1. How does the human visual system give us perception of depth?
2. From a computational standpoint, a stereo system must solve two main problems. Which?
3. What are the stereo correspondance constraints?



Right Image of Stereo vison

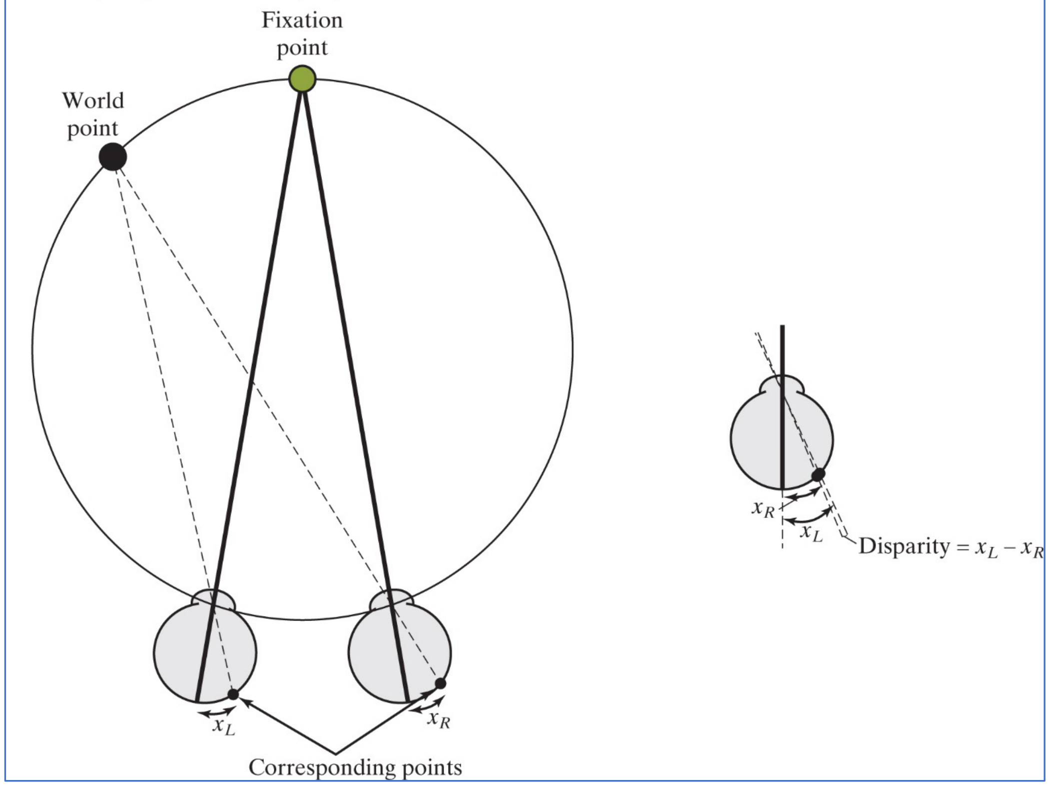


Left Image of Stereo vision

(13.1) Human Stereopsis

- Depth is perceived by the **retinal disparity -** the horizontal difference in the retinal locations of two projections of the same scene point.
- Beyond a few meters, the retinal disparity is too small to be detectable.
- Human use also other cues such as:
  - relative size,
  - persepctive,
  - object overlap,
  - contrast, light
  - motion parallax

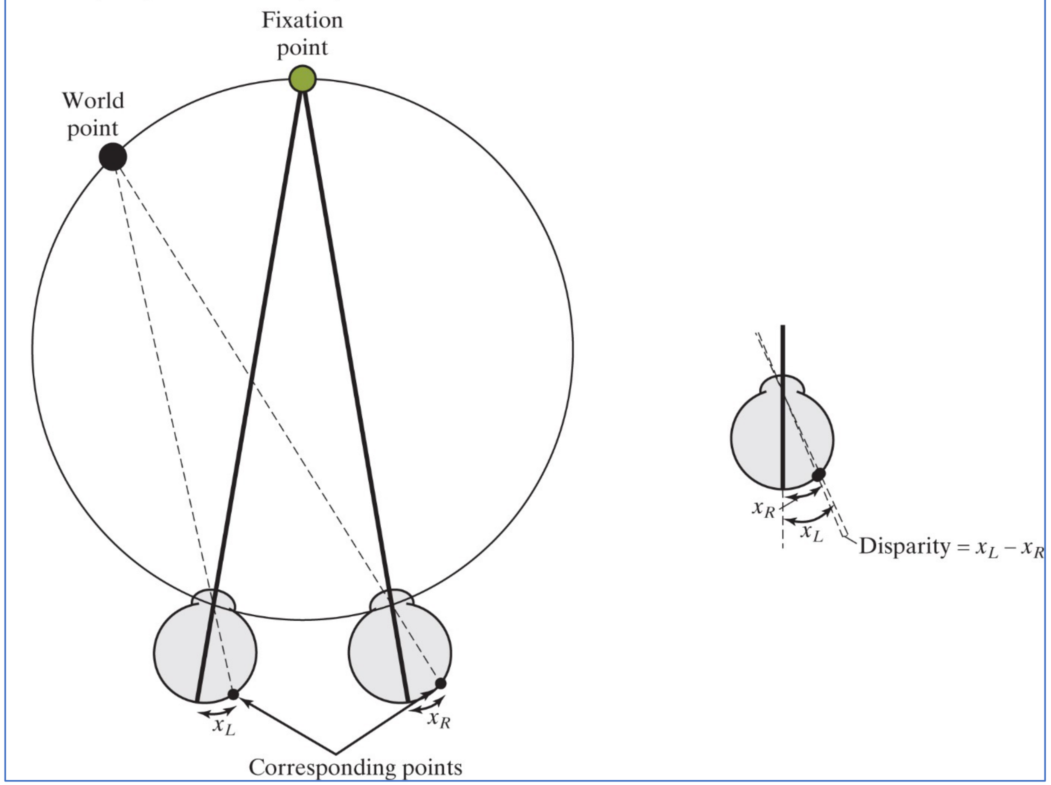
Figure 2.5 Retinal disparity is defined as the distance between corresponding points on the two retinas, after the retinas have been overlaid on top of one another and rotated so that their optical axes are coincident. Based on B. A.Wandell. *Foundations of Vision*. Sunderland, Mass., Sinauer Associates, Inc., 1995.



Human stereopsis

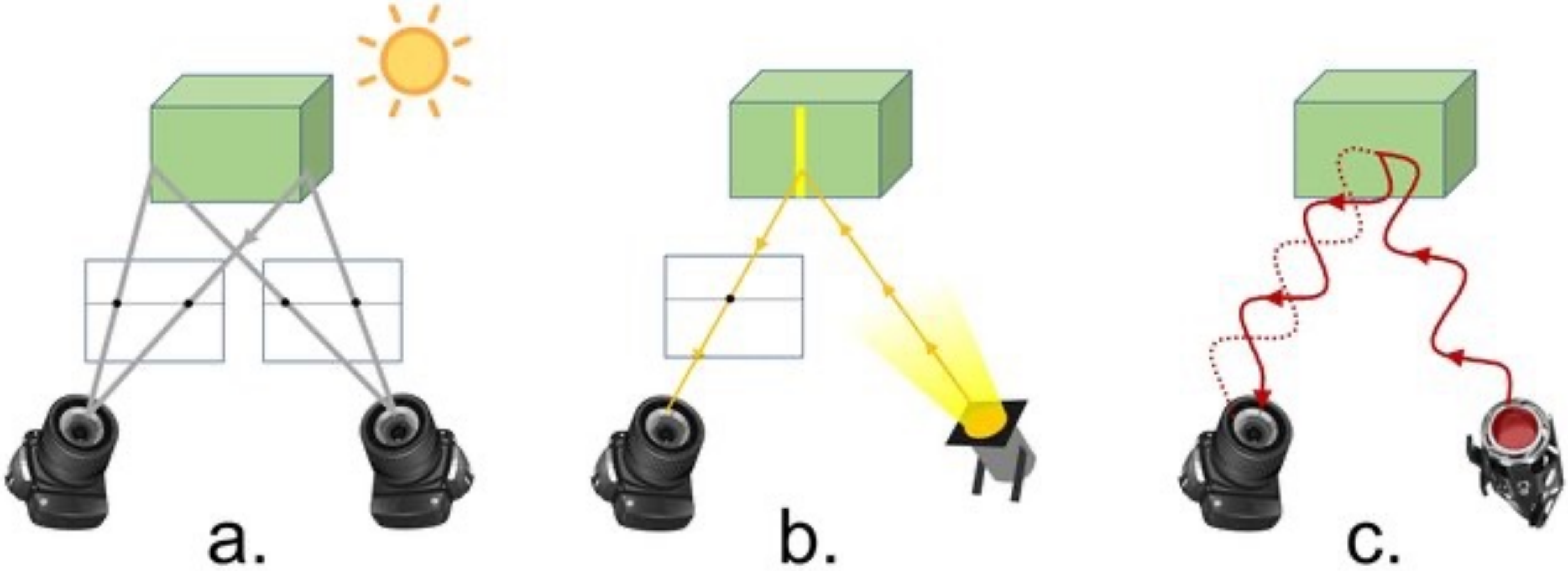
- Stereo vision, or **stereopsis**, refers to the process of recovering 3D information about the world from multiple images of a scene taken at the same time by different imaging devices.

Figure 2.5 Retinal disparity is defined as the distance between corresponding points on the two retinas, after the retinas have been overlaid on top of one another and rotated so that their optical axes are coincident. Based on B. A.Wandell. *Foundations of Vision*. Sunderland, Mass., Sinauer Associates, Inc., 1995.



Passive and Active depth

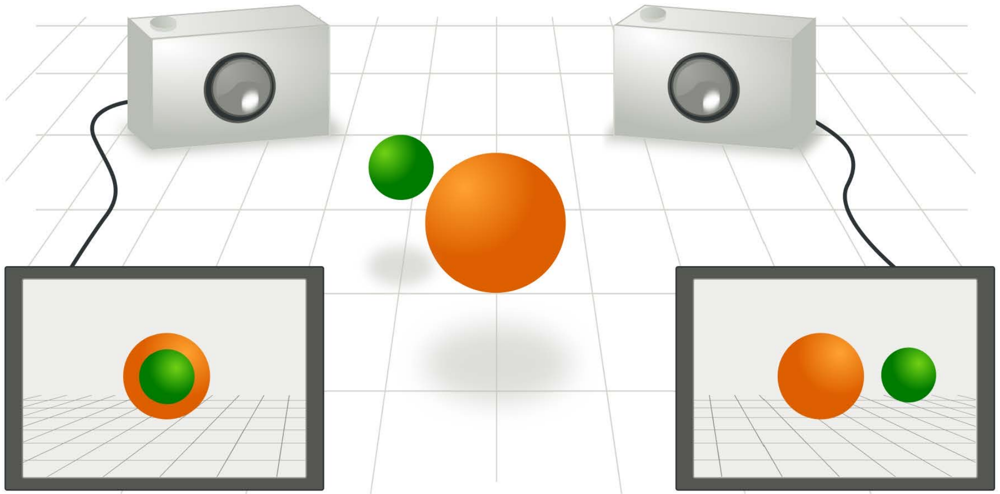
- Passive:
  - Mimic what humans do, like two cameras to estimte depth by disparity
  - Photometric stereo: two images taken by same camera, same location, but different lighting conditions
  - Depth from defocus; multiple images taken with different focal lengts
- Active
  - Light is projected onto the scene, and the light is sensed in some way
  - Laser range finder (single point)
  - Laser scanner
  - Time of flight (TOF) camera (light from LED or diode. Entire scene captured simultaneously
  - Structured light (example Kinect)



Depth Sensing technologies. (a) Passive stereo. (b) Structured light. (c) Time of Flight.

A.Haider “What can we learn from depth camera sensor noise”, 2022, https://www.mdpi.com/1424-8220/22/14/5448

Two cameras – Two views



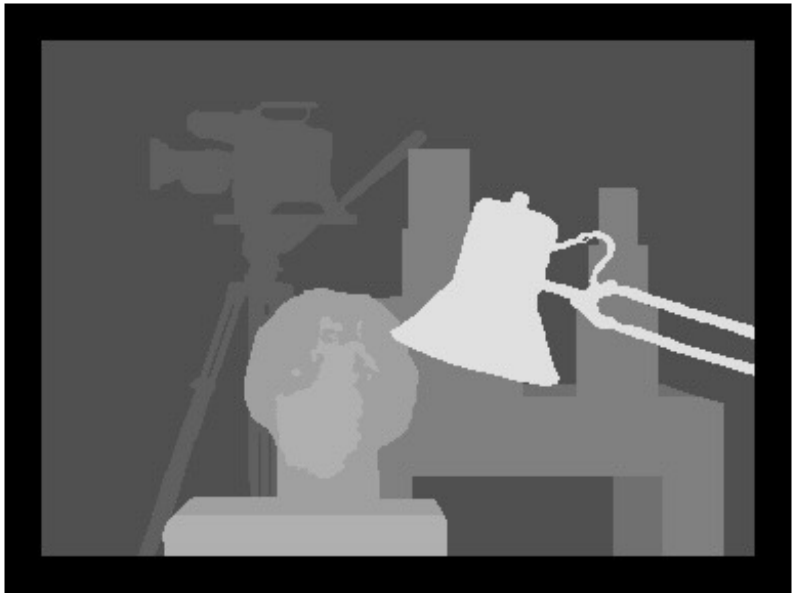
The images for the two views are different with respect to the relationship between objects.

In the left image, the green ball hides (occludes) the central part of the orange ball.

http://upload.wikimedia.org/wikipedia/commons/thumb/7/7c/Aufnahme_mit_zwei_Kameras.svg/2000px-Aufnahme_mit_zwei_Kameras.svg.png[27.09.2013 10:16:16]

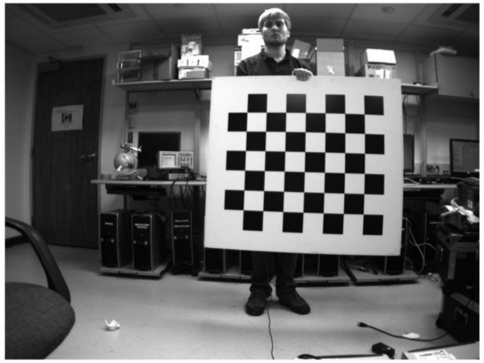
Some databases exist with stereo images and "true depth maps". Her an example from the Middelbury database.

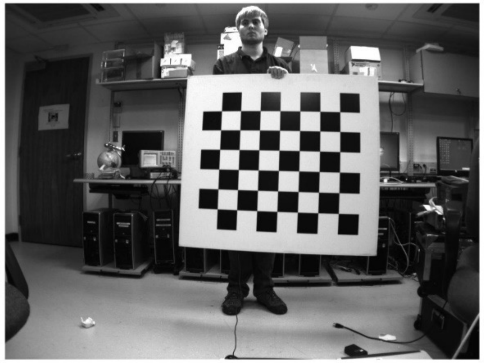
https://vision.middlebury.edu/stereo/data/scenes2001/



Stereopsis, Left and Right image.

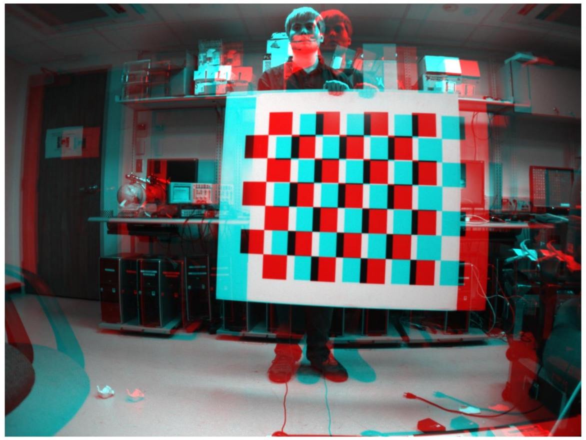
Left Image





Right Image

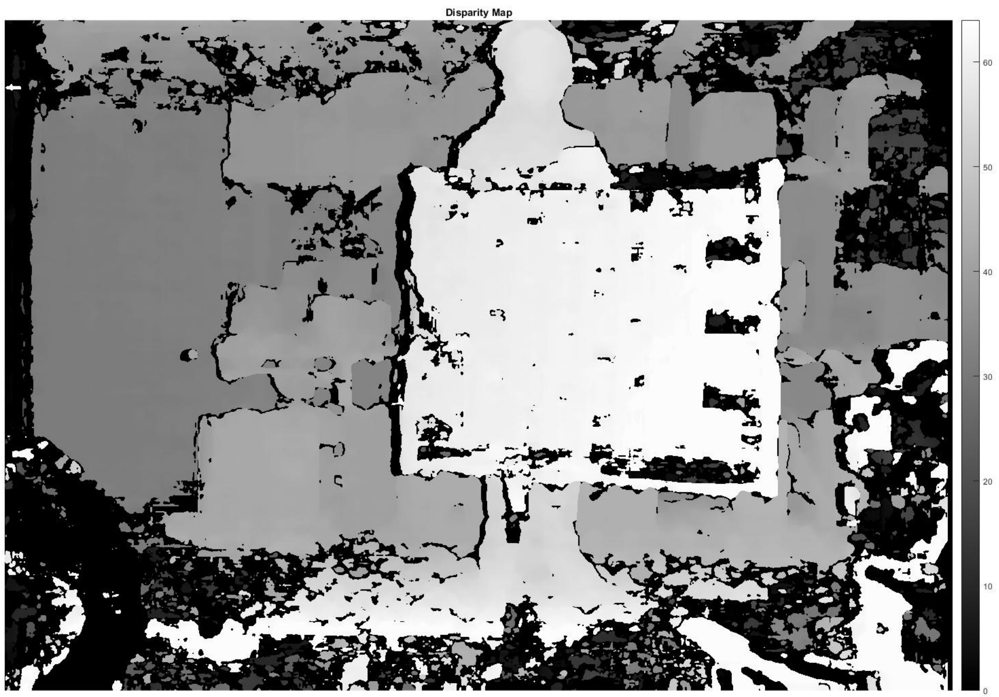
Example from Matlab, Computer Vision System Toolbox



Stereo anaglyph

Anaglyph 3D images contain two differently filtered colored images, one for each eye. When viewed through "color-coded anaglyph glasses", each of the two images reaches the eye it's intended for, revealing an integrated

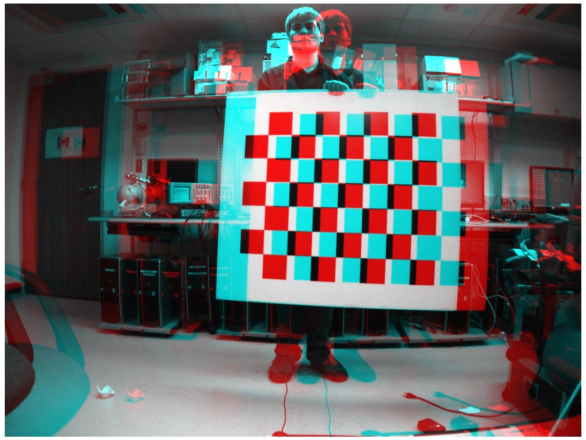
Disparity map - Depth map



Disparity map

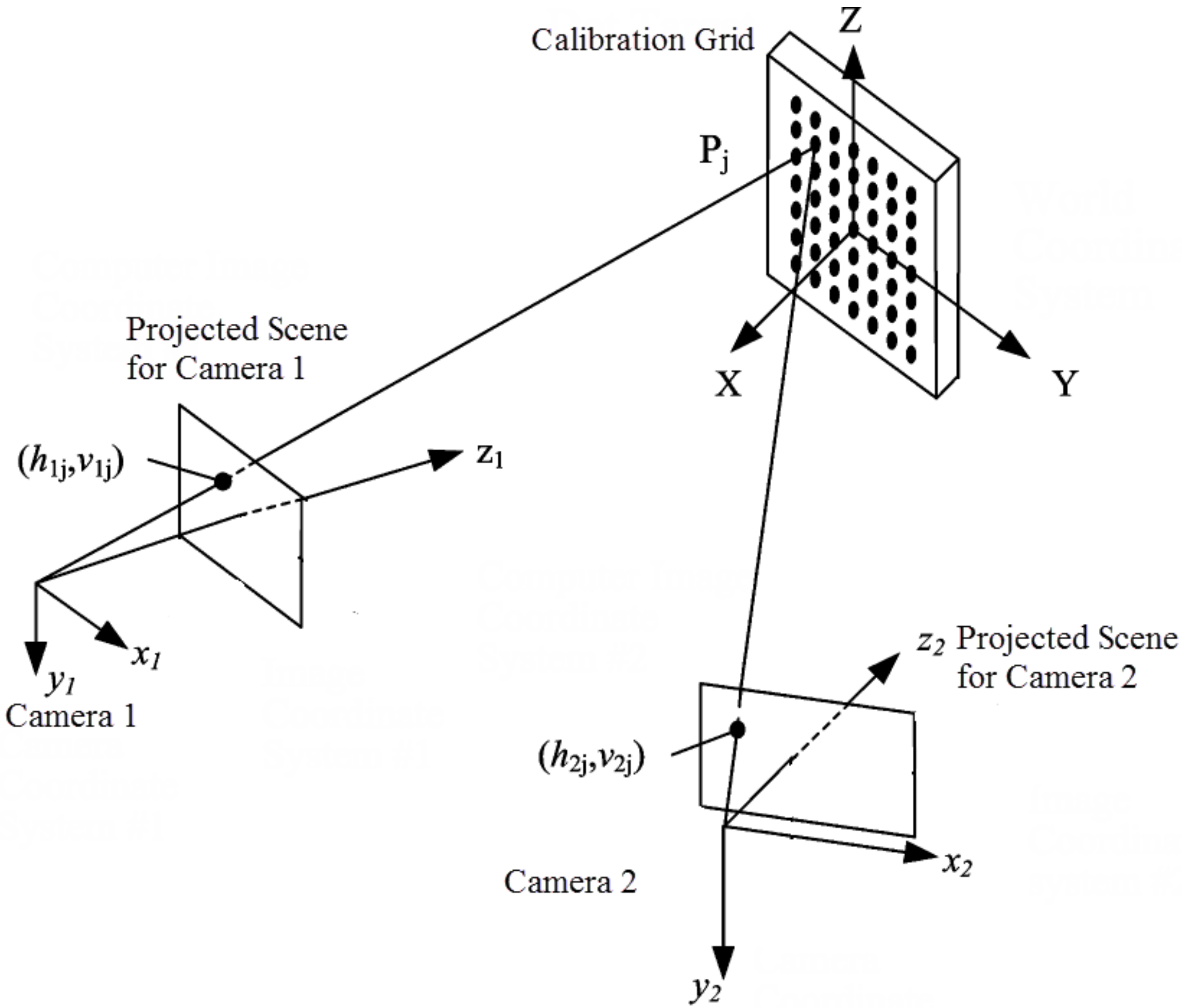
Disparity: The difference between corresponding points in the Left and Right image.

Depth map: The distances to the object points computed from the disparity and the geometry of the system.



binocular vision is based on *triangulation*.

The 3D location of any visible object point must lie on the straight line that passes through the centre of projection and the image of the object point. The determination of the intersection of two such lines generated from two independent images is called triangulation.



Rectified cameras

When cameras are rectified the image planes of the two cameras are coplanar. The camera positions are related by a translation parallel to the scanlines

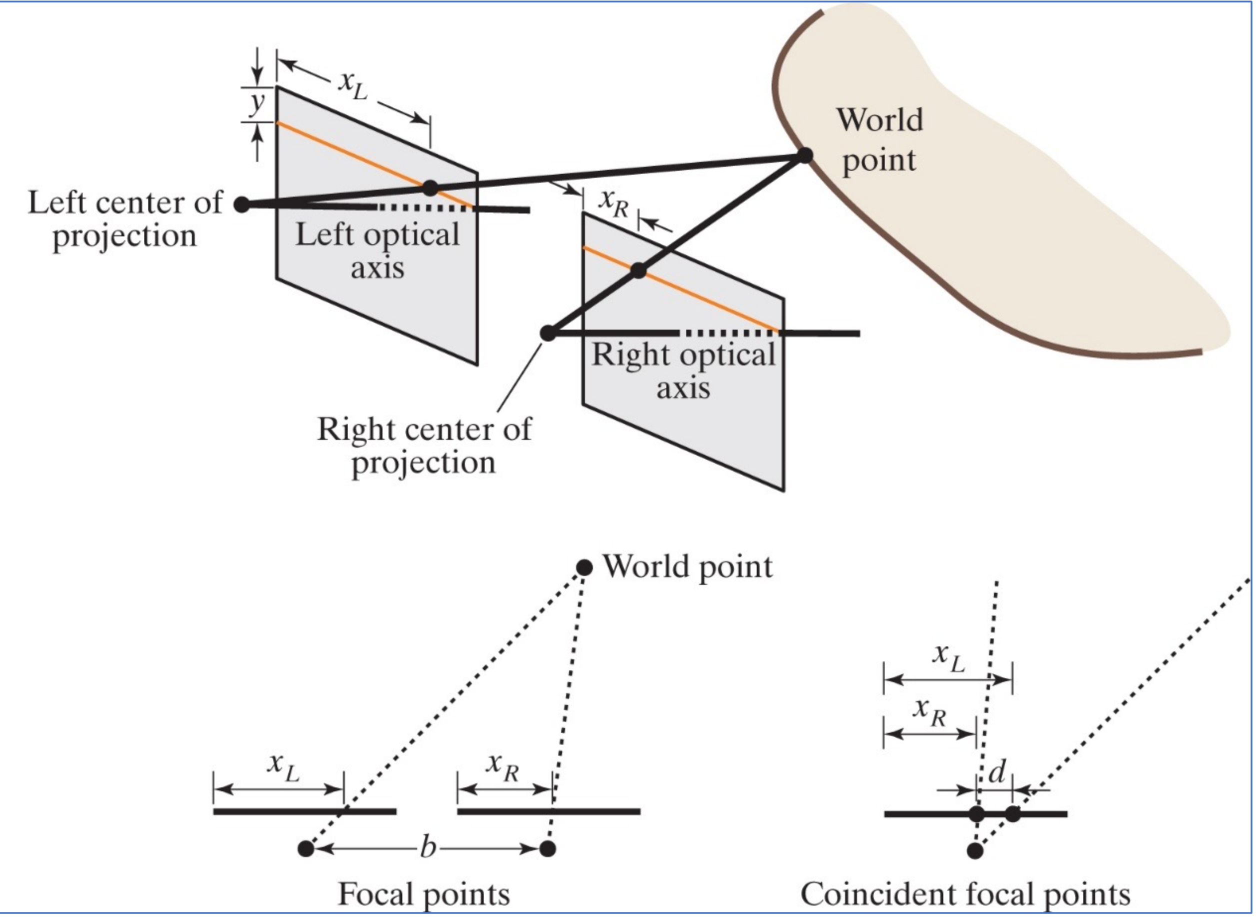
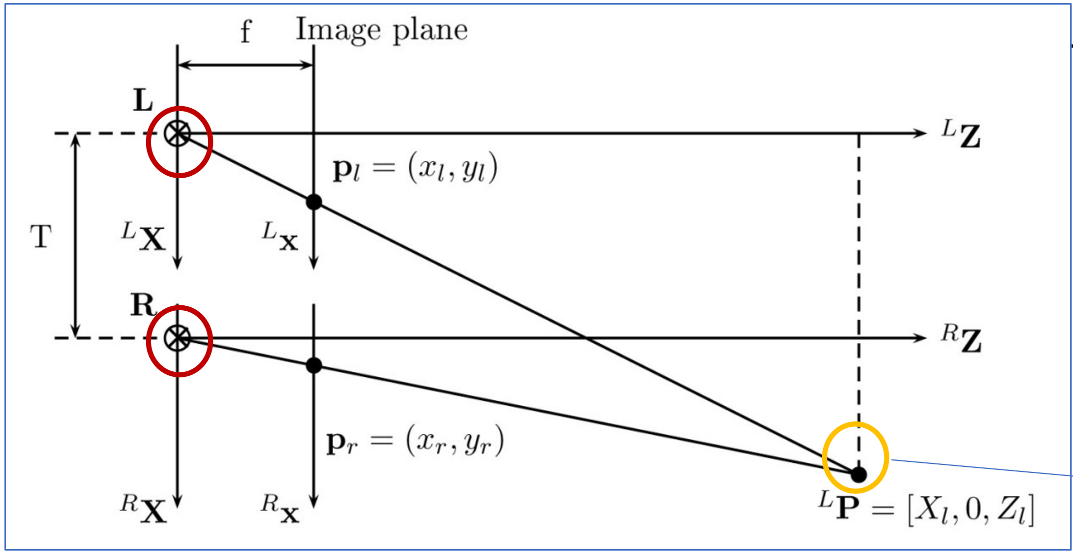


Figure 13.4 Rectified stereo geometry. TOP: A world point is imaged at point (x_L, y) in the left image and (x_R, y) in the right image, with respect to coordinate systems aligned with each image and placed in the top-left corner, as usual. BOTTOM LEFT: The same scene viewed in 2D. (The y axis, going into the page, is not shown.) BOTTOM RIGHT: Overlapping the two imaging rays onto a single (virtual) sensor, the distance $x_L - x_R$ between the two coordinates is the disparity d.

A simple binocular stereo system



Physical point

This is a 2D case that can be considered as a cross section of a 3D case with *parallel optical axes, i.e. rectified*

$$Z_l = Z_r = Z, \qquad X_r = X_l - T$$

$$\begin{cases} \dfrac{Z}{f} = \dfrac{X_l}{x_l} \\ \dfrac{Z}{f} = \dfrac{X_l - T}{x_r} \end{cases} \Rightarrow \begin{cases} X_l = \dfrac{Z}{f} x_l \\ X_l = \dfrac{Z}{f} x_r + T \end{cases}$$

$$\frac{Z}{f}(x_l - x_r) = T \quad \Rightarrow \quad Z = \frac{fT}{x_l - x_r} = \frac{fT}{d}$$

The **disparity**: *d (here rectified cameras)*

In general the disparity is a 2D vector: $\mathbf{d} = \left[d_x \; d_y \right]^{\mathrm{T}}$

Disparity – rectified cameras – book notation

- The disparity is inversely proportional to depth. For rectified cameras:

$$d = x_L - x_R = f\frac{x_w + b}{z_w} - f\frac{x_w}{z_w} = \frac{fb}{z_w}$$

where *b* is the distance between the two focal points, called the **baseline.**

Stereopsis, Imaging from Two Views

Stereo vision refers to the ability to infer information on the 3-D structure and distance of a scene, from two images taken from different viewpoints.

From a computational standpoint, a stereo system must solve two problems:

1) **The correspondence problem**. Finding corresponding points in two images.

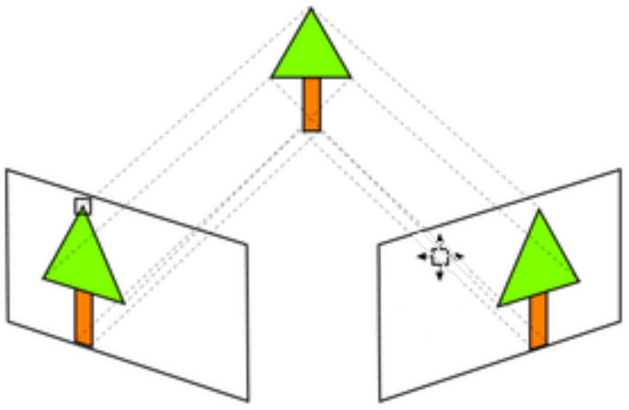
2) **The reconstruction problem**. As a result from the first step we get a disparity map. This is used to reconstruct the scene by finding world points and the structure of imaged objects.

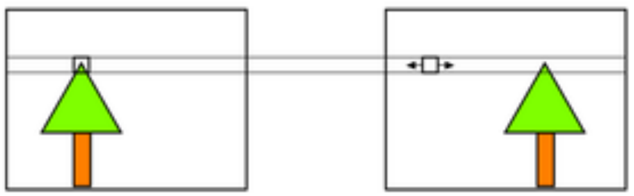
(13.2) Matching stereo images – the correspondence problem

- **We want to infer depth** by matching the pixels in two images.

- **Correspondence problem:** to determine for each point in one image its corresponding point in the other image.

- Two pixels are said to **correspond** if both pixels are projections along lines of sight of the same physical scene element.





Corresponding points - example

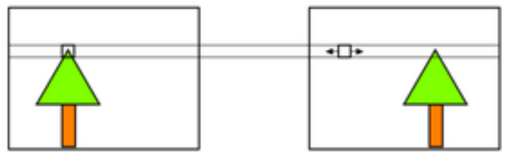
Left image

Right image

Disparity for the marked point: (210 - 206, 344 - 344) = (4,0)

From the Middlebury Database, http://vision.middlebury.edu/flow/data/

Correspondence problem



- Given (xL,yL) where can the corresponding (xR,yR) be?
  - It is constrained to be along a line called epipolar line (the epipolar constrain)
  - For rectified cameras or rectified images, the scanlines are the epipolar lines.

- Need to find the corresponding points from the **matching space** (possible matches)

- If d=0 -> zw (depth) goes to infinity, like stars in the sky
- If d is large, zw (depth) becomes small. This means object close to camera.
- usually we have d<dmax

- Frontoparallell: an object parallell to both image planes, and thus of constant depth and disparity

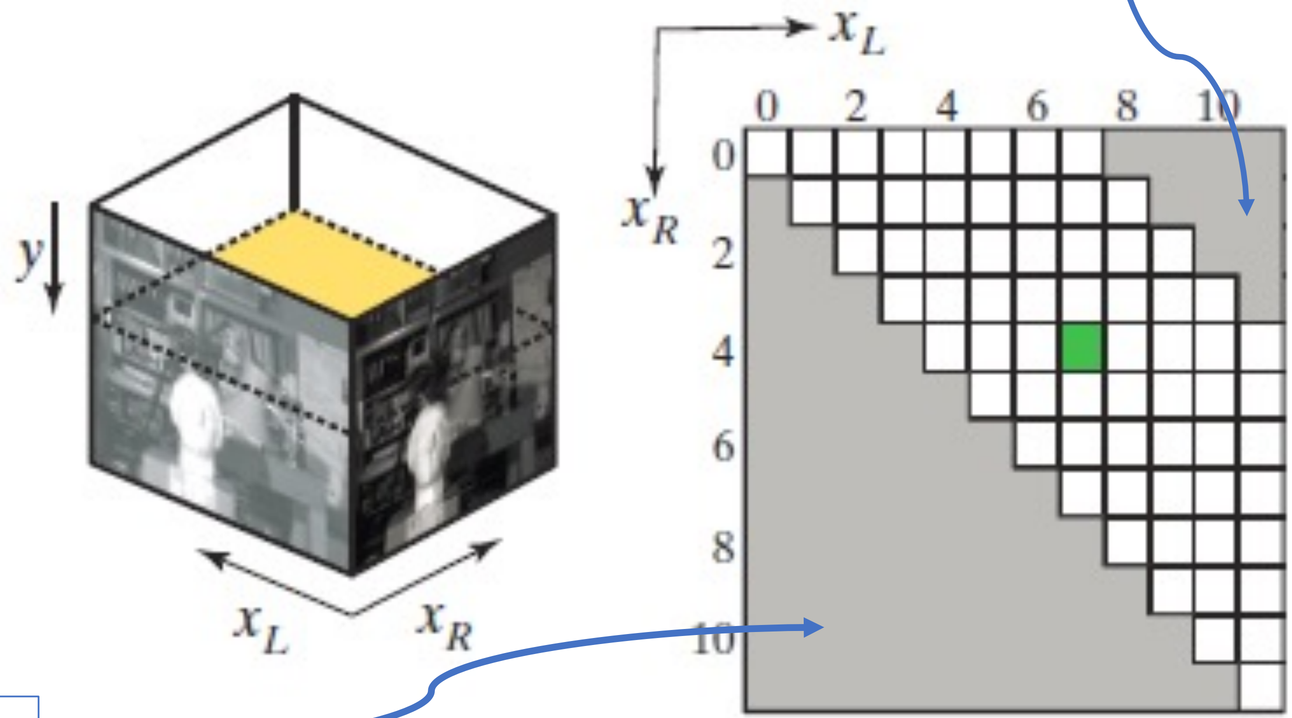
Correspondence - matching space

Maximum allowed disparity constricts this corner (here 7)

Example, matching space for rectified images.

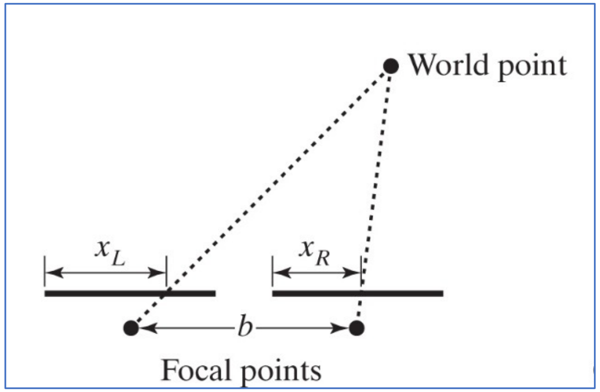
Green represent a match between pixel xl=7 and xR=4, so that d=3. The possible disparity is bounded by the shaded region.

A frontoparallell object will have all matches at a diagonal (constant disparity)



XR > XL – impossible, thus forbidden region.

Stereo constraints

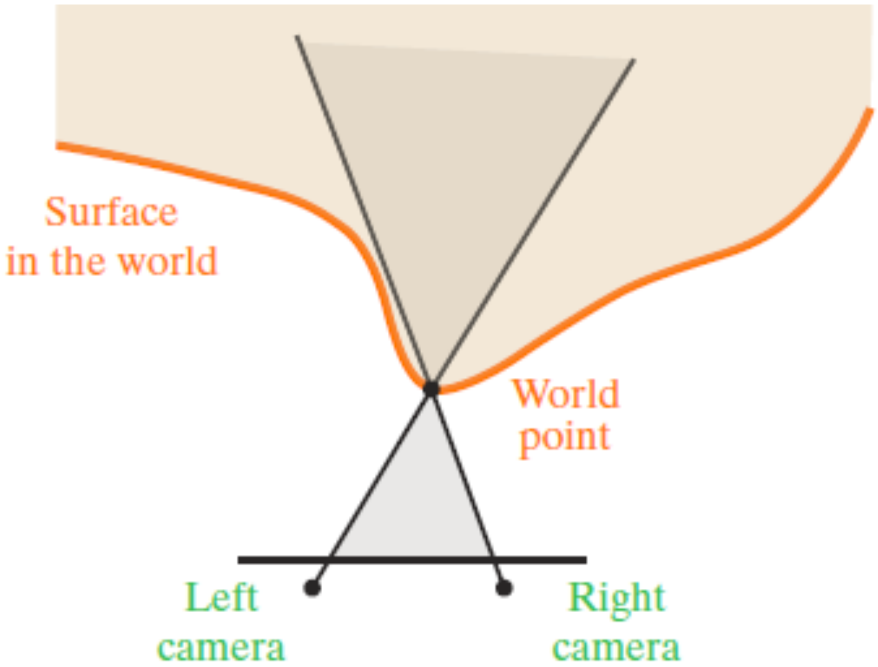


- The **cheirality constraint** requires $x_L \geq x_R$ for matching pixels since only objects in front of the camera can be visible.

- The **maximum disparity constraint** forbids matches whose disparity exceeds a certain amount, which enforces a minimum distance from the camera to the surface being viewed.

- The **uniqueness constraint** says that if $x_L \leftrightarrow x_R$ is a match, then there is no other match $x_L \leftrightarrow x$ where $x \neq x_R$, and there is no other match $x \leftrightarrow x_R$ where $x \neq x_L$.

- **Forbidden zone:** when a point on a continuous surface is viewed by both cameras, it is not physically possible for another point on the same surface to also be visible in both cameras if it lies within the region defined by two lines passing through the centers of projection and the point.
- The forbidden zone is taken care of by the **ordering constraint**.



Stereo constraints

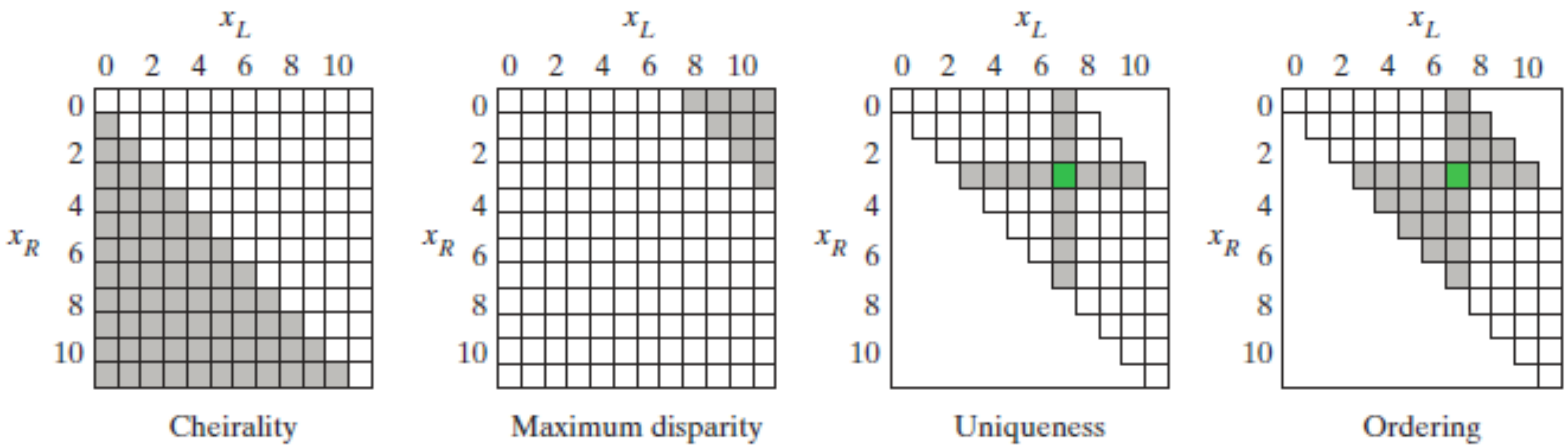


Figure 13.6 Stereo constraints. The gray cells indicate matches that are simply not allowed (left two grids) or that are illegal when the green cell indicates a match (right two grids). Cheirality precludes matches with $x_L < x_R$, which would refer to points behind the camera. Maximum disparity precludes matches whose disparity exceeds a threshold. Uniqueness prevents a pixel in either scanline from matching more than one pixel. Ordering ensures that the pixel coordinates of the matches are monotonically increasing as the pixels along either scanline are traversed. Note that the gray cells in the right grid are the forbidden zone.

Correspondance - Block Matching

- **Block matching** is an *area-based* approach that relies upon a statistical correlation between local intensity regions.
- For each pixel (*x,y*) in the left image, the right image is searched for the best match among all possible disparities $0 \leq d \leq d_{max}$ (in the matching space)
- A window of possible matches is searched, and a similarity measure is used.

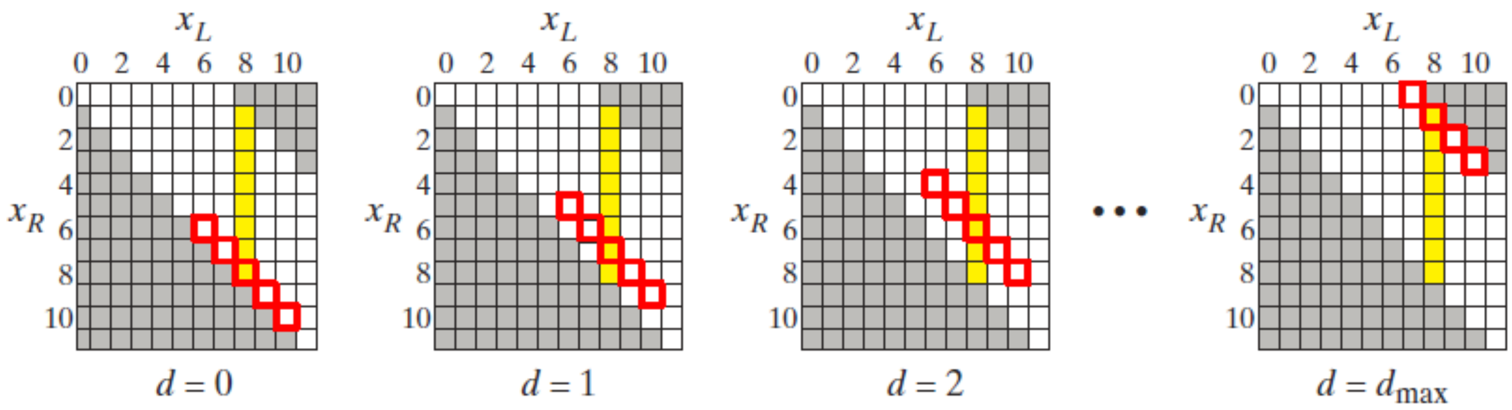


Figure 13.10 Block matching algorithm. For every pixel in the left image, a search is performed to find the disparity yielding the lowest cost. The red cells indicate the dissimilarities that are aggregated in Lines 5–6 of BLOCKMATCH1, while the yellow cells indicate the matches considered during the search. Shown is the pixel $x_L = 8$ with a window size of $w = 5$.

$$d_L(x, y) = \arg \min_{0 \le d \le d_{\max}} dissim(I_L(x, y), I_R(x - d, y))$$

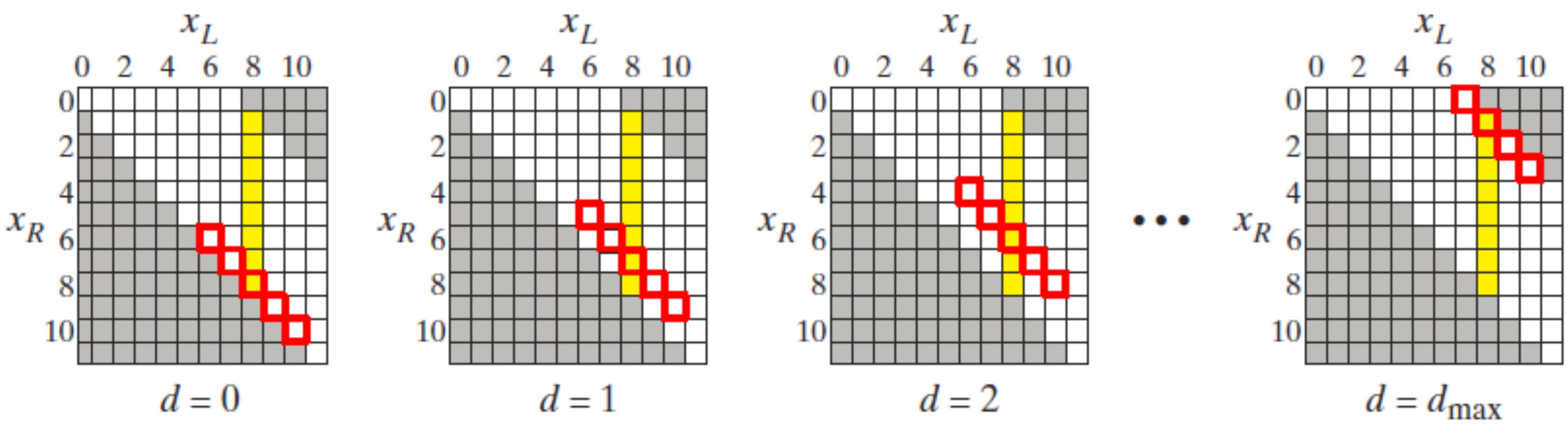


Figure 13.10 Block matching algorithm. For every pixel in the left image, a search is performed to find the disparity yielding the lowest cost. The red cells indicate the dissimilarities that are aggregated in Lines 5–6 of BLOCKMATCH1, while the yellow cells indicate the matches considered during the search. Shown is the pixel $x_L = 8$ with a window size of $w = 5$.

dL: left disparity map, i.e. with respect to the left image. Can also find dR, and do left-right disparity check.

Agree? -> OK!

Disagree? -> unrealiable

Dissimilarity Measures

- **Sum of absolute differences (SAD):**

$$dissim(I_L(\mathbf{x}_L), I_R(\mathbf{x}_R)) \quad = \quad |I_L(\mathbf{x}_L) - I_R(\mathbf{x}_R)| \qquad (\text{SAD})$$

- **Sum of squared differences (SSD)**:

$$dissim(I_L(\mathbf{x}_L), I_R(\mathbf{x}_R)) \quad = \quad (I_L(\mathbf{x}_L) - I_R(\mathbf{x}_R))^2 \quad (\text{SSD})$$

- **Crosscorrelation, the product of their intensities:**

$$dissim(I_L(\mathbf{x}_L), I_R(\mathbf{x}_R)) \quad = \quad -I_L(\mathbf{x}_L) I_R(\mathbf{x}_R) \qquad (\text{cross correlation})$$

- Example 1:

Rectified cameras, maximum disparity is 20.

Left image pixel position: (52,3)

Which of the following right image coordinates could be a possible match?

a) (26,3)
b) (48,13)
c) (64,3)
d) (48,3)
e) (59,6)
f) (52,10)
g) (36,3)

- Example 2:

Rectified cameras, maximum disparity is 20.

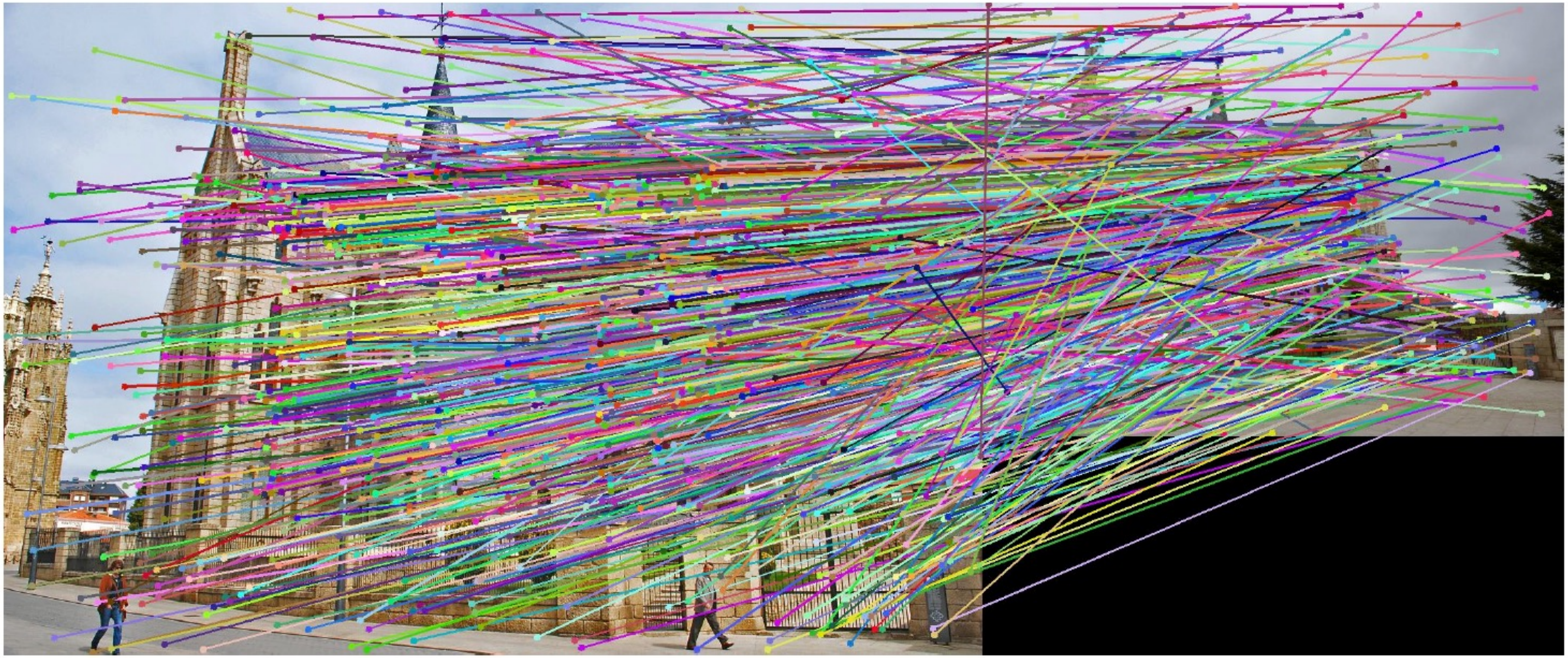
Left image pixel position: (14,12) is matched with right image (10,12)

What is the forbidden zone?

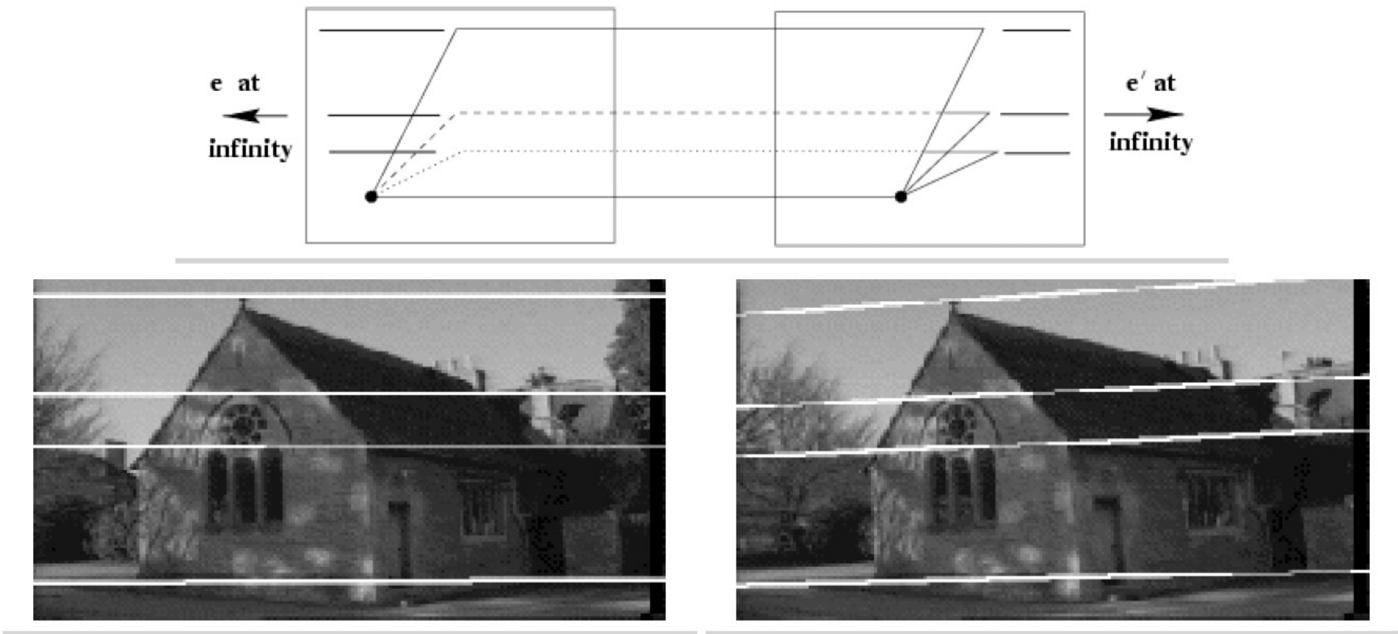
Which of the following right image coordinates could be a possible match for XL: (12,12)?

a) (30,12)
b) (10,12)
c) (11,12)
d) (8,12)

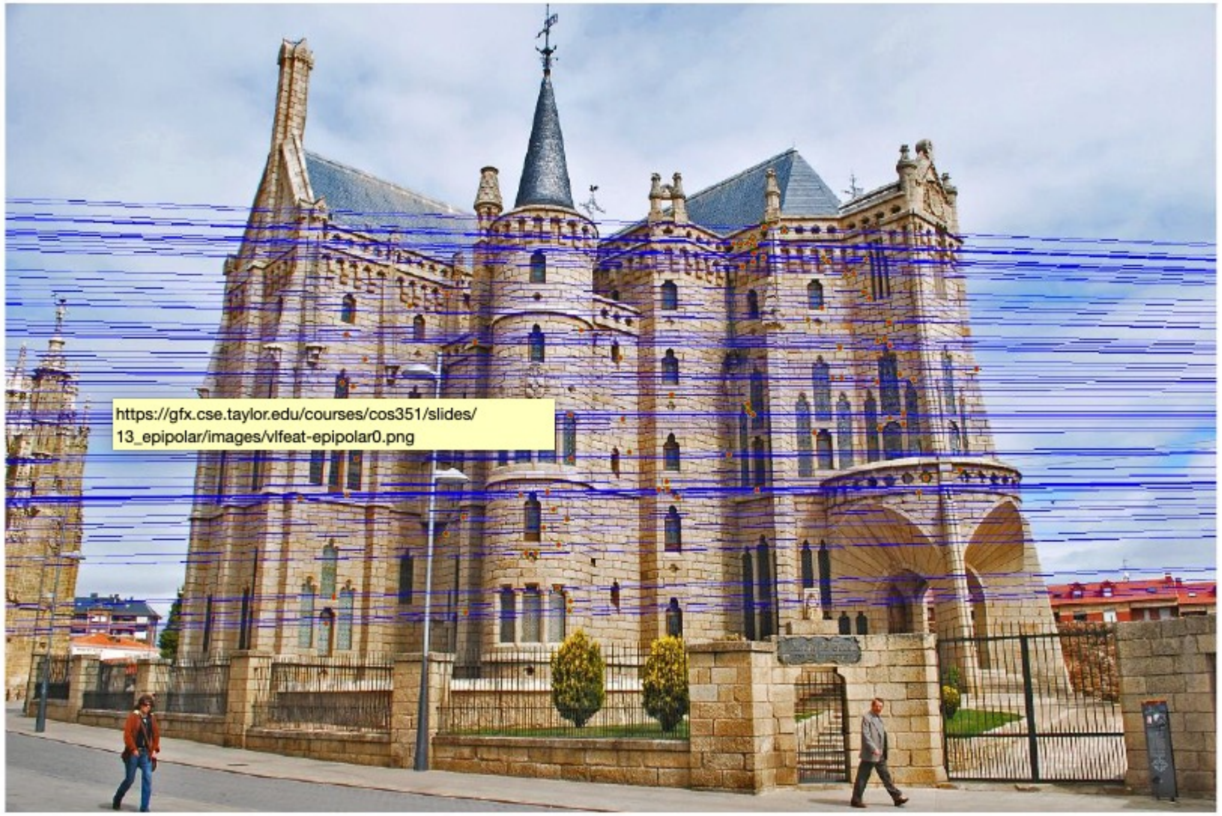
Where are possible corresponding points?

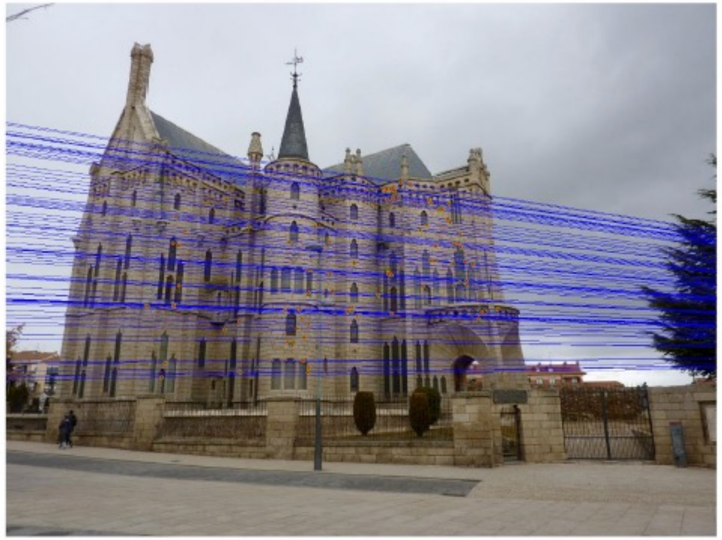


- Remember rectified cameras: corresponding points are on the scan lines (same y coordinate).
- Wouldn't it be nice to know where matches can live (matching space) also for unrectified cameras?
- We can constrain our 2D search to 1D to find corresponding points
- This 1D line of possible corresponding points is called the epipolar line

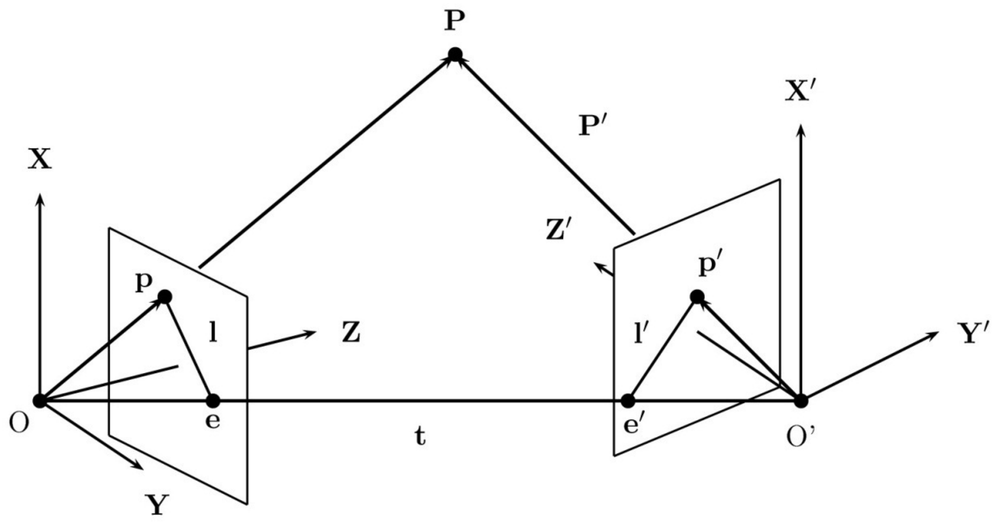


1D line of possible corresponding points - epipolar line





Epipolar Geometry



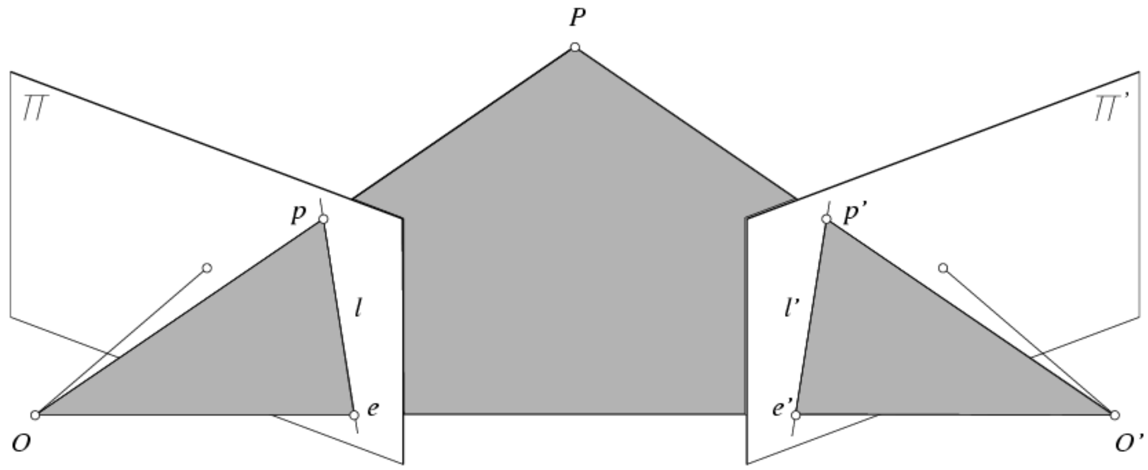
The *left camera coordinates are used as reference, and the world coordinate system in stereo vision*.

Right camera coordinates are denoted by primed symbols ‘

The *Epipolar Plane*: OO’P
Epipoles: e, e’
Epipolar lines: l, l’

Baseline: OO’ = t (here)
Camera centers: O, O’
Optical axises: Z, Z’
Image points: p, p’

Epipolar Geometry



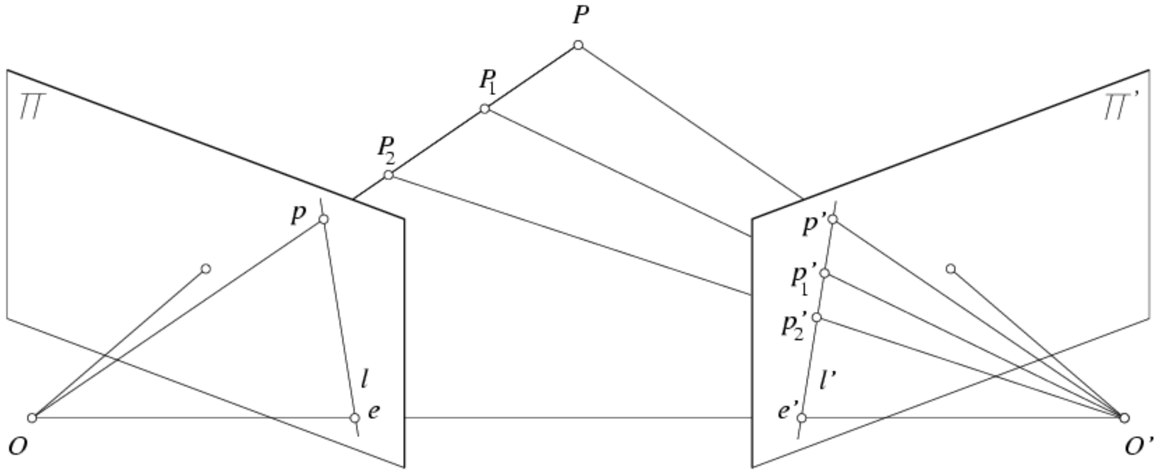
The *Epipolar Plane*: OO'P
Epipoles: e, e'
Epipolar lines: l, l'
Image points: p, p'

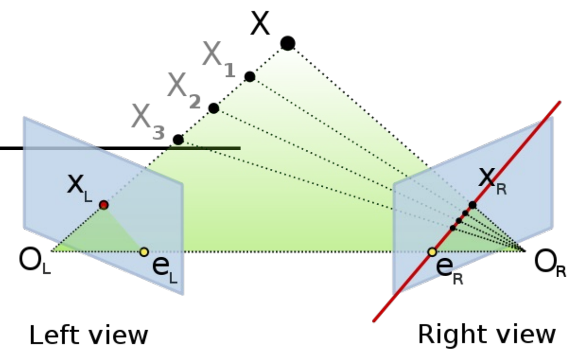
Note! There is a separate *epipolar plane* for each point in the scene.

The optical centers of the cameras lenses are distinct, thus each center projects onto a distinct point into the other camera's image plane.
These two image points, here denoted by **e** and **e´**, are called *epipoles* or *epipolar points*.

Both epipoles **e** and **e´** in their respective image planes and both optical centers **O** and **O´** lie on a single 3D line.

Epipolar Constraint

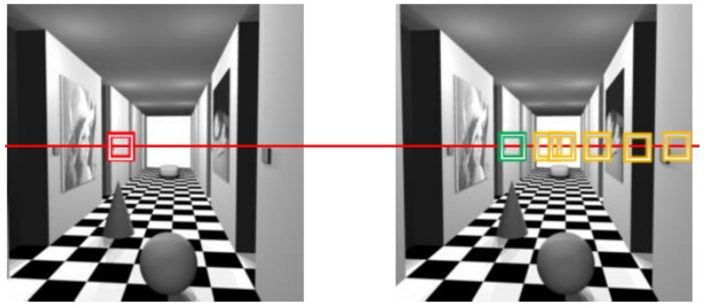




Source: https://commons.wikimedia.org/wiki/File:Epipolar_geometry.svg#/media/File:Epipolar_geometry.svg, Licensed under: CC BY-SA 3.0

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

Human stereopsis and correspondence problem



Right Image of Stereo vison

Left Image of Stereo vision

Three points from the topic:

1. How does the human visual system give us perception of depth?
   - ✓retinal disparity, relative size, motion parallax ….
2. From a computational standpoint, a stereo system must solve two main problems. Which?
   - ✓The correspondence problem and the reconstruction problem
3. What are the stereo correspondance constraints?
   - ✓Cheirality, maximum disparity, uniqueness, ordering, epipolar lines