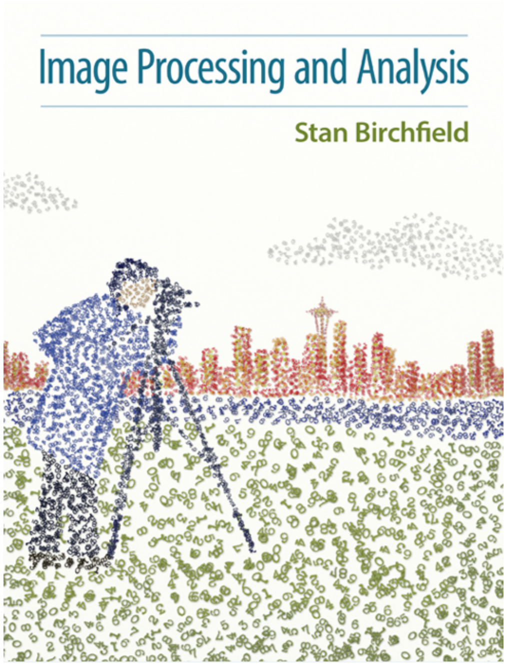
University of Stavanger

Prof. Kjersti Engan

ELE510 Image processing and computer vision

Stereopsis and correspondence problem, (chap 13.1, 13.2 Birchfield) 2020



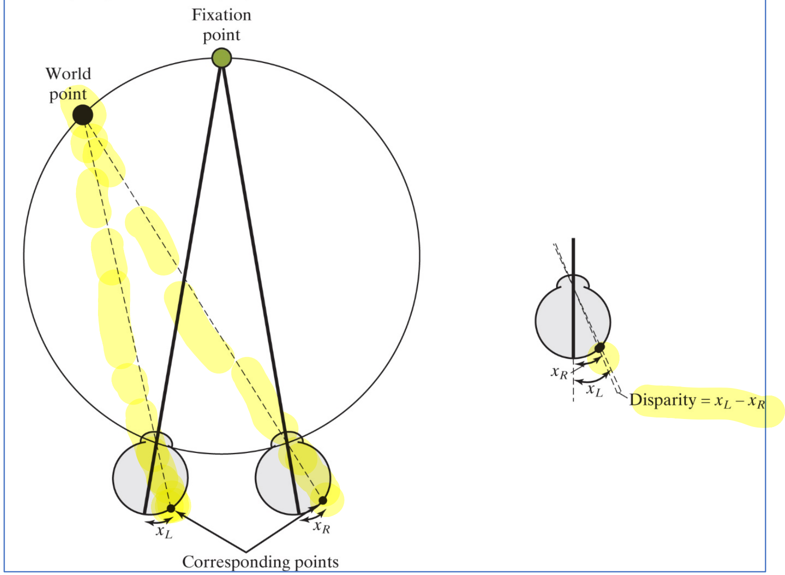
Parts in course presentations is material from Cengage learning. It can be used for teaching, and it can be share it with students on access controlled web-sites (Canvas) for use in THIS course. **But it should not be copied or distributed further in any way** (by you or anybody).

(13.1) Human Stereopsis

- Depth is perceived by the **retinal disparity** in two dissimilar images of the same object, which is the horizontal difference in the retinal locations of two projections of the same scene point.

- 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.



Human stereopsis

- 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 etc. (in addition to the retinal disparity)

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.

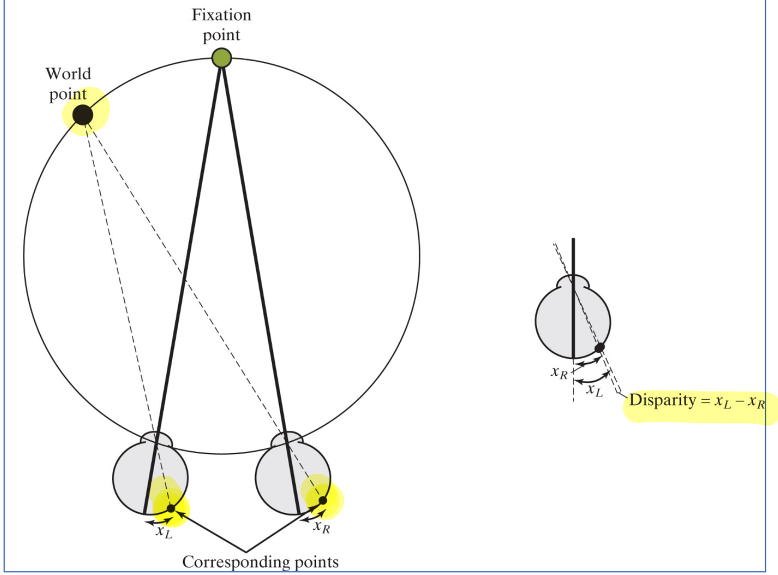
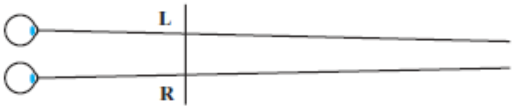
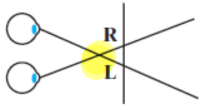


Figure 13.1 A pair of stereo images. To fuse, relax your eyes so that you are seeing through the paper, and try to bring the two images together so that they come into alignment. Be patient, as it may take some time initially. If you experience difficulty fusing the images, it may help to place a vertical divider (such as a piece of cardboard) between the images so that each eye can see only one image. (Note that fusion is impossible if the distance between the two left edges of the images exceeds the interpupillary distance, which may occur if this page was enlarged by photocopying or viewed on screen at a large zoom setting.)

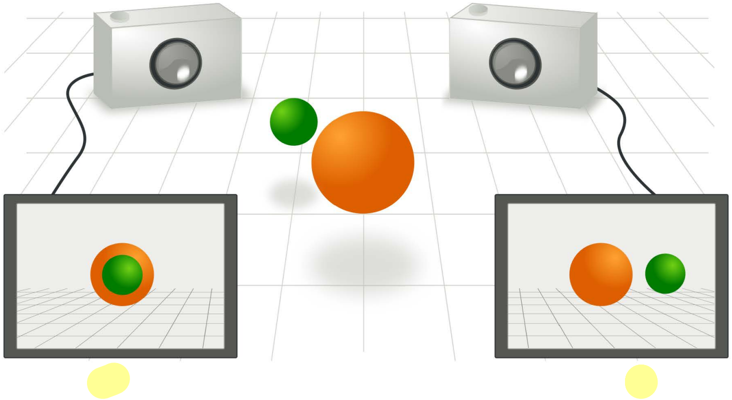
Left image Right image

Right image Left image





Two cameras – Two views

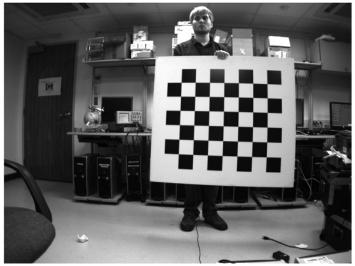


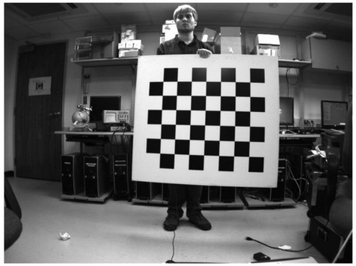
The images for the two views are different with respect to the relationship between objects. In the left image, the green ball hides (occludes) the central part of the orange ball.

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]

Stereopsis, Left and Right image.

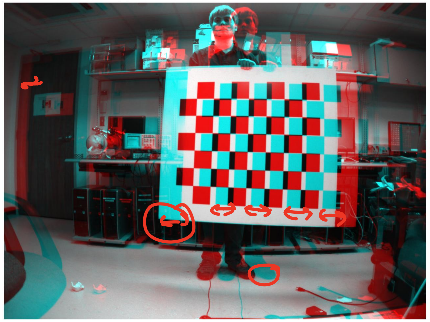
Left
Image





Right
Image

Example from
Matlab,
Computer
Vision System
Toolbox



Stereo
Anaglyph

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

- **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.
- The **disparity** between two points is defined as the difference in their image coordinates.

Corresponding points

Left image

Right image

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

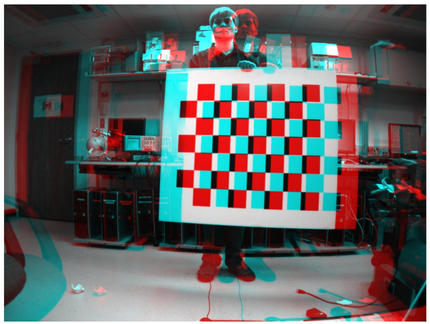
From the Middlebury Database, http://vision.middlebury.edu/flow/data/

Disparity map - Depth map

Disparity map

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

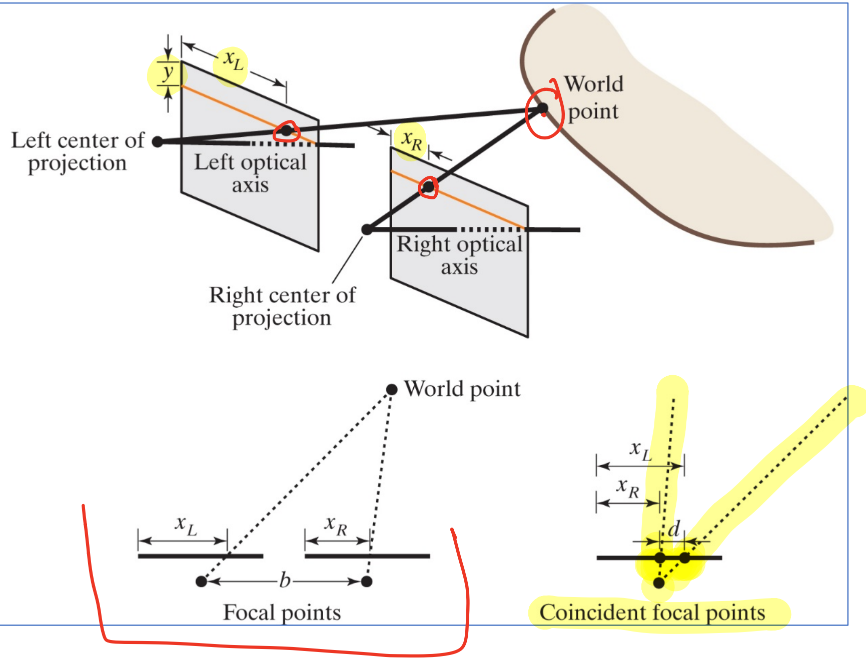
Depth map: The distances to the object points computed from the disparity and the geometry of the system.



Rectified cameras

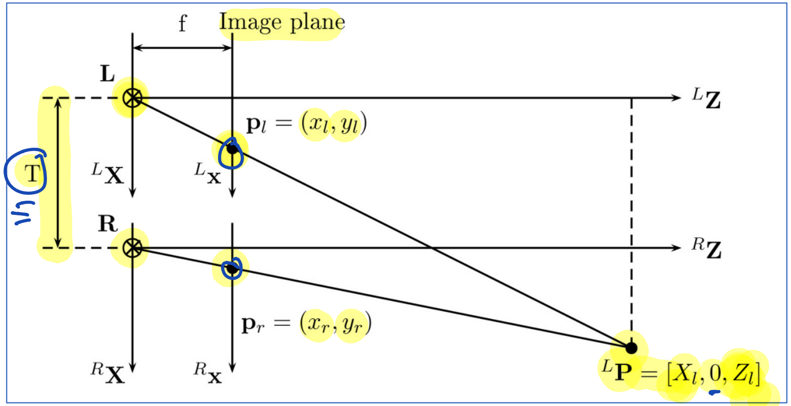
When cameras are rectified the image planes of the two cameras are coplanar. The camera positions are related by a translation parallel to the scanlines

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.



$x_L > x_R$

A simple binocular stereo system



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

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

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

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

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

In general the disparity is a 2D vector: $\mathbf{d} = \begin{bmatrix} d_x & d_y \end{bmatrix}^{\mathrm{T}}$

Disparity – rectified cameras – book notation

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

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

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