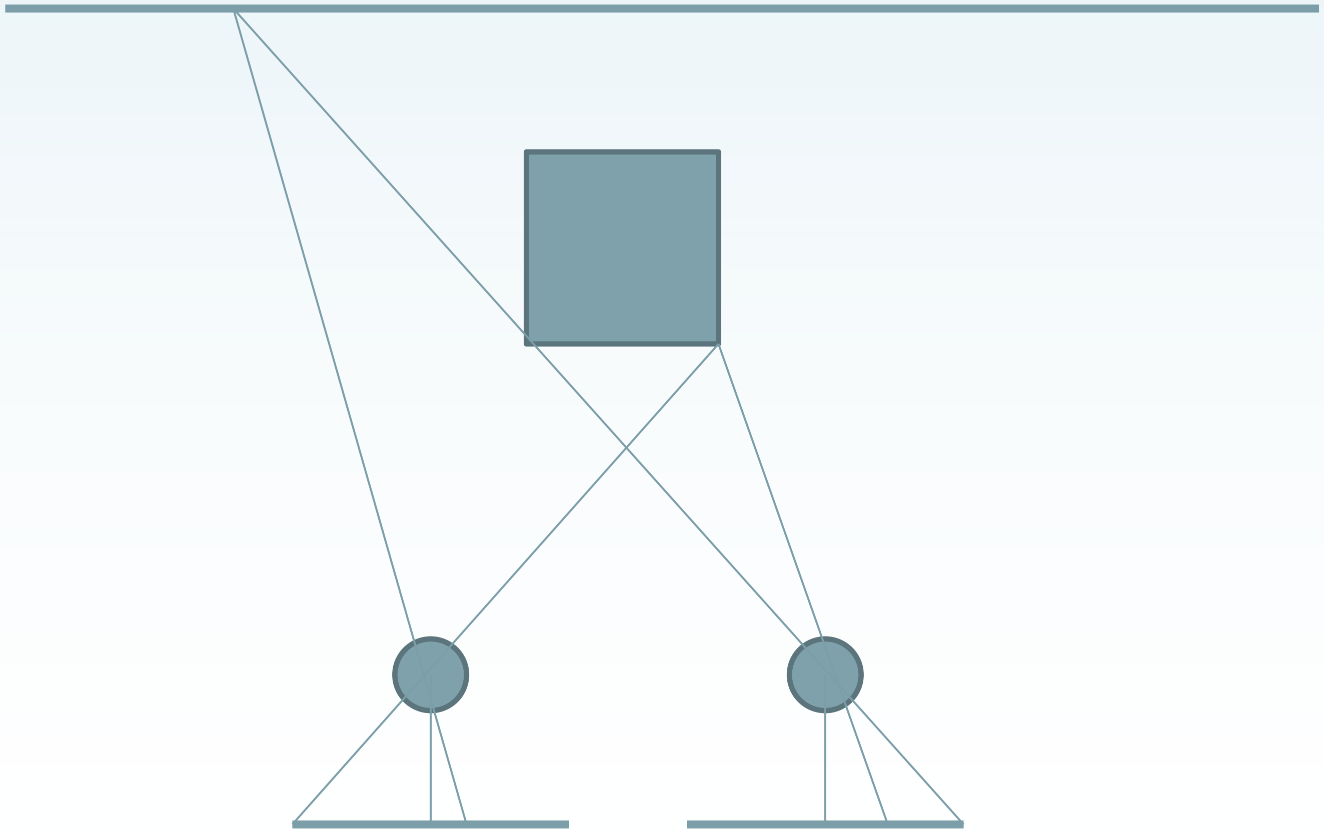
Algorithms Courework #2

Ingemar J. Cox

Stereo vision

- Stereo vision is the primary means by which humans infer the depth of objects in a scene.



A stereo pair

Right image

Left image

More details

- These images are epipolar aligned
- So a pixel in the ***i***th row of the left image can only match a pixel in the ***i***th row of the right image
- This reduces the 2D matching problem to N 1D matching problems, one for each of the N rows

Stereo matching by dynamic programing

Please read the paper “Stereo Without Disparity Gradient Smoothing: a Bayesian Sensor Fusion Solution”

http://www.bmva.org/bmvc/1992/bmvc-92-035.pdf

The basic idea

- We need to match each pixel in the left row to no more than one pixel in the right row
- "… no more than one …" Either
  - Pixel L[i] matches pixel R[j], or
  - Pixel L[i] is occluded, i.e. not visible to the right eye, or
  - Pixel R[j] is occluded, I,e, not visible to the left eye.

The basic idea

We need:

1. A cost for matching L[i] to R[j], and
2. A cost for occlusion

L[i] and R[j] are the pixel intensities. The closer their values, the more likely they are corresponding points, i.e. they match. So use squared error.

Note, since images are greyscale, L[i] and R[j] are scalar values. The math in the paper then simplifies

Occlusion cost is given in the paper. Use 3.8

Matching cost

Equation (3) reduces to:

$$\hat{z} = \frac{z_{1,i1} + z_{2,i2}}{2}$$

and the matching cost in Equation (6) becomes

$$(\hat{z} - z_1)(\hat{z} - z_2)/\sigma^2$$

The paper assumes $\sigma^2 = 16$

Determining the optimum match