3D Vision 3

Week 9

Two-view Geometry: Triangulation

Two-view Geometry: Stereo

Announcements

- **Survey 2 – Week 8|9** is live on Wattle, please provide a response to improve the course for future years
- **Class Representatives:** another mechanism to provide feedback
  - Arjun Raj (u7526852@anu.edu.au) [COMP4528]
  - Ruxuan Li (u7763307@anu.edu.au) [ENGN4528]
  - Jiabao Han (u7765516@anu.edu.au) [COMP6528]
  - Xingchen Zhang (u7670173@anu.edu.au) [COMP6528]
- **Exam:** 9:00am–12:15pm Saturday 1 June 2024
  - 7-11 Barry Drive, First Floor Left Side

Weekly Study Plan: Overview

Wk	Starting	Lecture	Lab	Assessment
1	19 Feb	Introduction	✗	
2	26 Feb	Low-level Vision 1	1	
3	4 Mar	Low-level Vision 2	1	
		Mid-level Vision 1		
4	11 Mar	Mid-level Vision 2	1	CLab1 report due Friday
		High-level Vision 1		
5	18 Mar	High-level Vision 2	2	
6	25 Mar	High-level Vision 3[1]	2	
	1 Apr	Teaching break	✗	
	8 Apr	Teaching break	✗	
7	15 Apr	3D Vision 1	2	CLab2 report due Friday
8	22 Apr	3D Vision 2	3	
9	29 Apr	3D Vision 3	3	
10	6 May	3D Vision 4	3	
		Mid-level Vision 3		
11	13 May	High-level Vision 4	✗	CLab3 report due Friday
12	20 May	Course Review	✗	

Weekly Study Plan: Part B

Wk	Starting	Lecture	By
7	15 Apr	3D vision: introduction, camera model, single-view geometry	Dylan
8	22 Apr	3D vision: camera calibration, two-view geometry (homography)	Dylan
9	29 Apr	3D vision: two-view geometry (epipolar geometry, triangulation, stereo)	Dylan
10	6 May	3D vision: multiple-view geometry	Weijian
		Mid-level vision: optical flow, shape-from-X	Dylan
11	13 May	High-level vision: self-supervised learning, detection, segmentation	Dylan
12	20 May	Course review	Dylan

Outline

1. Two-view Geometry: Triangulation
2. Two-view Geometry: Stereo
3. Single-view Geometry: Monocular Depth
4. Single-view Geometry: Optical flow

Essential & Fundamental Matrices: Summary

- Algebraic representations of epipolar geometry:
  - Projection matrices (given intrinsics + extrinsics)
  - **Essential matrix** (given intrinsics): $\mathbf{E} = [\boldsymbol{t}]_\times \mathbf{R}$
  - **Fundamental matrix**: $\mathbf{F} = \mathbf{K}'^{-\top}[\boldsymbol{t}]_\times \mathbf{R}\mathbf{K}^{-1} = [\boldsymbol{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}$ [HZ p.244]
    - Using identity $[\boldsymbol{v}]_\times \mathbf{M} = \mathbf{M}^{-\top}[\mathbf{M}^{-1}\boldsymbol{v}]_\times$ and $\boldsymbol{e}' = \mathbf{K}'\boldsymbol{t}$ (image of camera centre 1)
- Epipolar constraint for corresponding points $\{\boldsymbol{x}, \boldsymbol{x}'\}$:

$$\boldsymbol{x}'^\top \mathbf{F}\boldsymbol{x} = \mathbf{0}$$

  - $\mathbf{F}$ is rank 2 and is known only up to scale → 7 DoF
  - What is $\mathbf{F}\boldsymbol{x}$?
- **Estimation:** DLT (again!); assemble a matrix $\mathbf{A}$, compute the SVD, enforce rank 2 (with another SVD); or use a nonlinear solver

The Fundamental Matrix F: Recovering P & P'

- Since we can set the world coordinate system arbitrarily, set

$$P = [I \mid 0]$$
$$P' = \big[[e']_\times F \mid e'\big]$$

where epipole can be found by solving

$$F^\top e' = 0$$

via SVD/eigendecomposition [HZ p. 256]

Triangulation

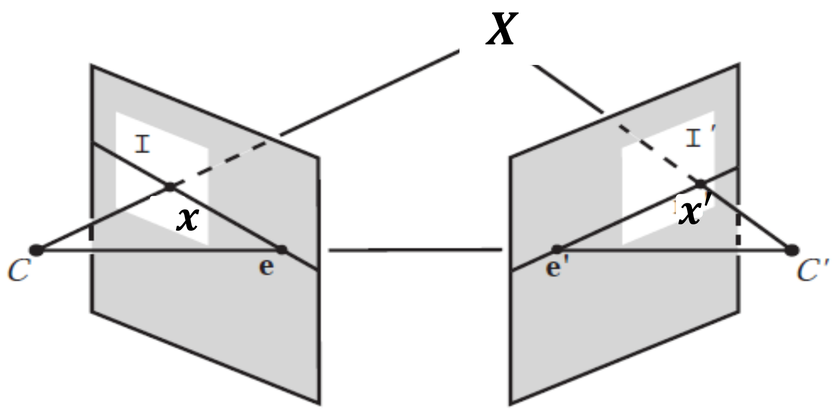
Two-view Geometry

Triangulation

- **Problem:** Given image points $\boldsymbol{x}$ and $\boldsymbol{x}'$ in *correspondence* across two or more images (taken from calibrated cameras with known intrinsics and extrinsics), compute the 3D location $\boldsymbol{X}$

Triangulation: DLT Linear Solution

- Generally, rays $\overrightarrow{Cx}$ and $\overrightarrow{C'x'}$ will not intersect exactly
- Can solve via the DLT algorithm (again!), finding a least squares solution to a system of equations
- Further reading:
  - HZ pp. 312–313



$$\boldsymbol{x} \times (\mathbf{P}\boldsymbol{X}) = 0 \text{ and } \boldsymbol{x}' \times (\mathbf{P}'\boldsymbol{X}) = 0$$

$$\therefore \mathbf{A}\boldsymbol{X} = \begin{bmatrix} u\mathbf{p}_3^\top - \mathbf{p}_1^\top \\ v\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ u'\mathbf{p}_3'^\top - \mathbf{p}_1'^\top \\ v'\mathbf{p}_3'^\top - \mathbf{p}_2'^\top \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0$$

Triangulation: DLT Linear Solution

Given $\mathbf{P}$, $\mathbf{P}'$, $\boldsymbol{x}$, $\boldsymbol{x}'$

1. Precondition points and projection matrices
2. Create matrix $\boldsymbol{A}$
3. SVD: $\mathbf{A} = \mathbf{U\Sigma V}^\top$
4. $\boldsymbol{X} = \mathbf{V}_{:,-1}$ (last column of $\mathbf{V}$)
5. Then refine with respect to a geometric error

$$\boldsymbol{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad \boldsymbol{x}' = \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1^\top \\ \mathbf{p}_2^\top \\ \mathbf{p}_3^\top \end{bmatrix} \quad \mathbf{P}' = \begin{bmatrix} \mathbf{p}_1'^\top \\ \mathbf{p}_2'^\top \\ \mathbf{p}_3'^\top \end{bmatrix}$$

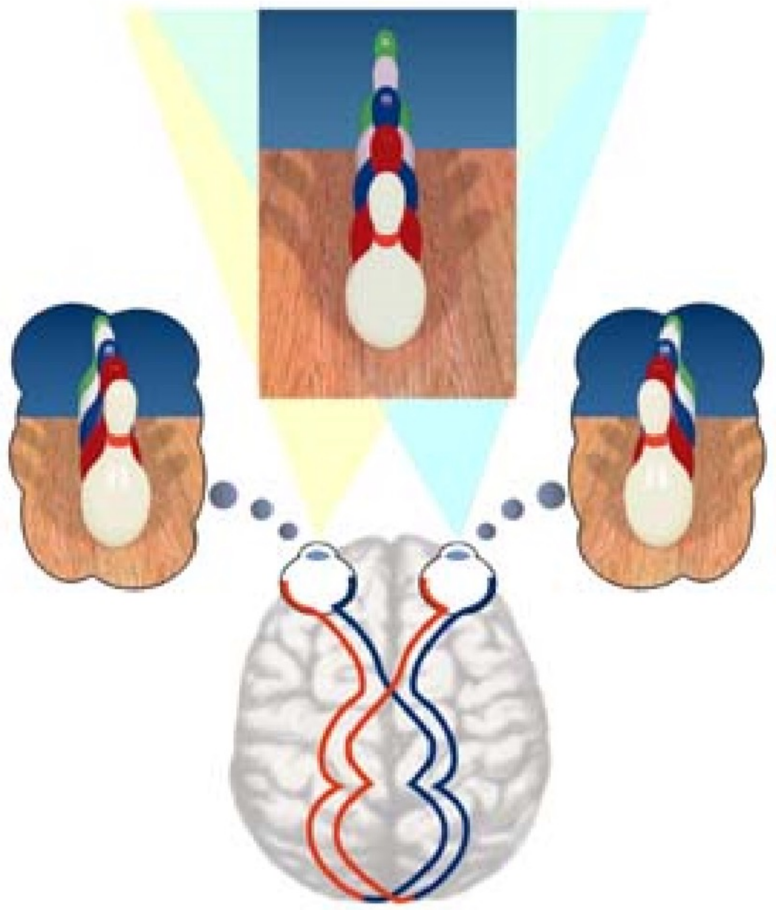
$$\mathbf{A} = \begin{bmatrix} u\mathbf{p}_3^\top - \mathbf{p}_1^\top \\ v\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ u'\mathbf{p}_3'^\top - \mathbf{p}_1'^\top \\ v'\mathbf{p}_3'^\top - \mathbf{p}_2'^\top \end{bmatrix}$$

Code: http://www.robots.ox.ac.uk/~vgg/hzbook/code/vgg_multiview/vgg_X_from_xP_lin.m

Stereo Vision

Two-view Geometry

Binocular Stereo Vision



3D Perception Using Stereo Vision

- Humans do it well from a *single* image and very well from *stereo* (binocular) images
- Mechanism is not well understood:
  - Biological components: some understanding
  - Algorithm: little understanding
- An intermediate goal of computer vision was to mimic the functionality of the biological system
  - *Not* to mimic its mechanics

Stereo Vision

- Stereo matching computes depth information from two images
- Depth information can be used to…
  - Differentiate objects from the background
  - Differentiate objects from one another
  - Expose camouflaged objects
  - Navigate in environment avoiding obstacles

Stereo Vision: Two Sub-Problems

Correspondence Problem

- The problem of measuring the disparity of each point in the two eye (camera) projections

Interpretation Problem

- The use of disparity information to recover the orientation and distance of surfaces in the scene

Stereo Vision: Algorithmic Steps

- Basic steps to be performed in any stereo imaging system:
  - Image Acquisition
  - Camera Modelling
  - Feature Extraction
  - Image Matching
  - Depth Interpolation

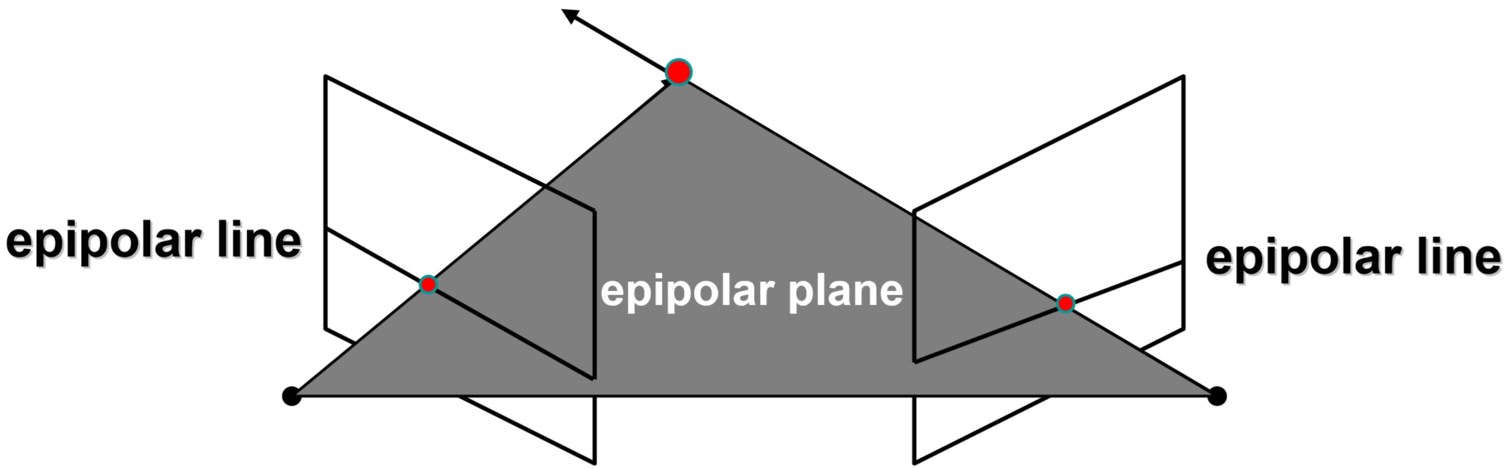
Stereo Image Acquisition

- Capturing two images with a very specific camera geometry

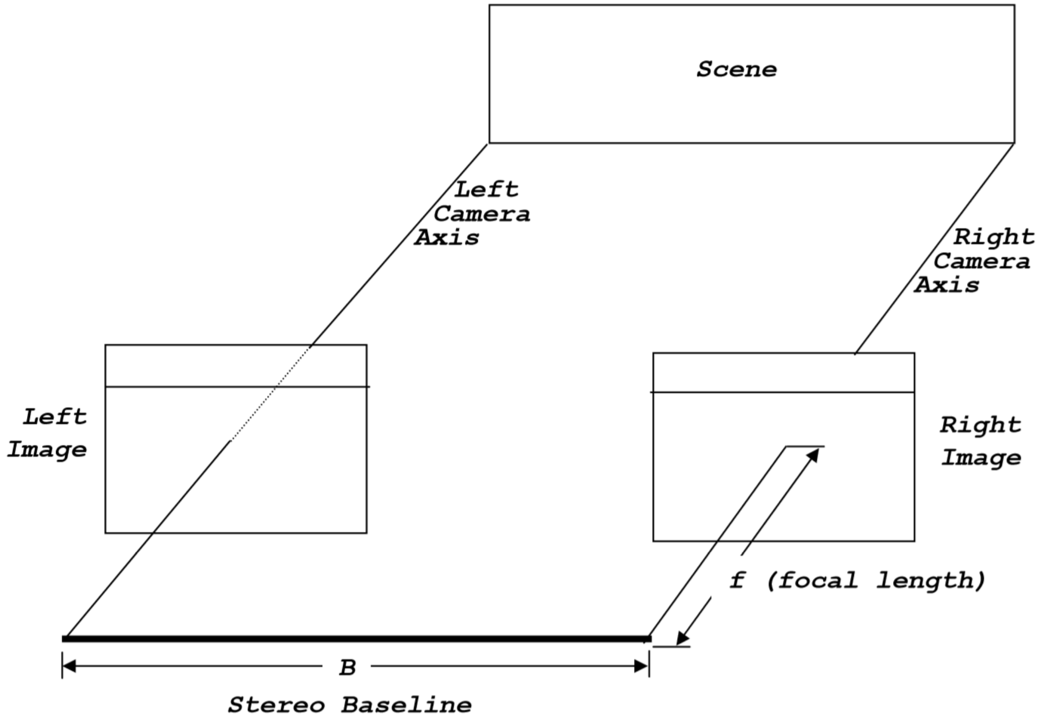
Camera Modelling

- For accurate depth results the camera parameters must be known
  - Intrinsic parameters: focal length, pixel skew/shape, distortion, etc.
  - Extrinsic parameters: relative rotation and translation

General Epipolar Constraint



Simple (Ideal) Stereo Geometry



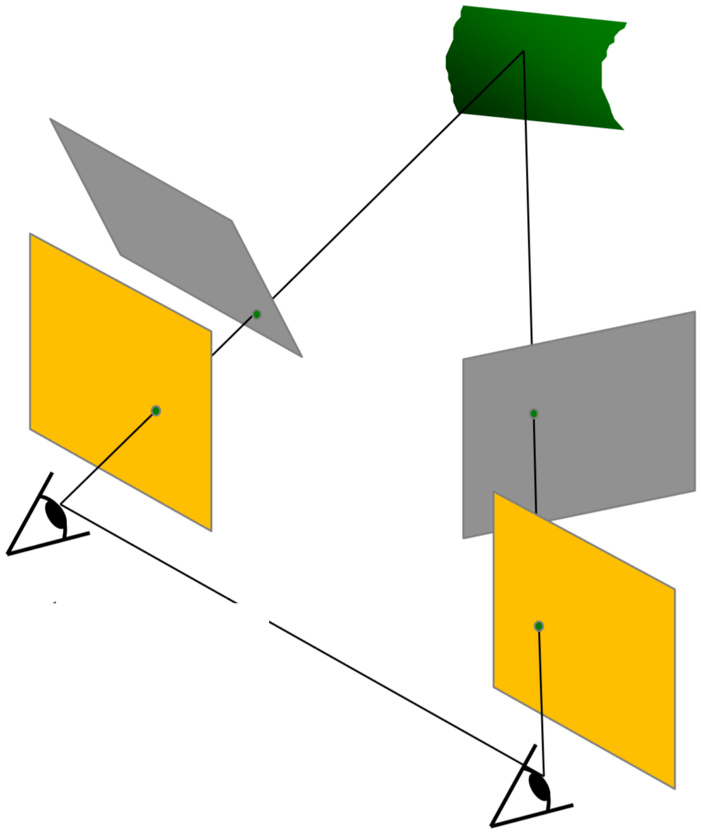
- Two slightly different images
- Extrinsics: identical rotations, translation has an offset in the X direction

Simplest Case: Rectified Stereo Images

- Assumptions:
  - Image planes of cameras are parallel
  - Principal axes are at same height
  - Identical focal lengths
- Then the epipolar lines fall along the horizontal scan lines of the images
- We will assume images have been rectified so that epipolar lines correspond to scan lines:
  - Simplifies algorithms
  - Improves efficiency

Simplest Case: Rectified Stereo Images

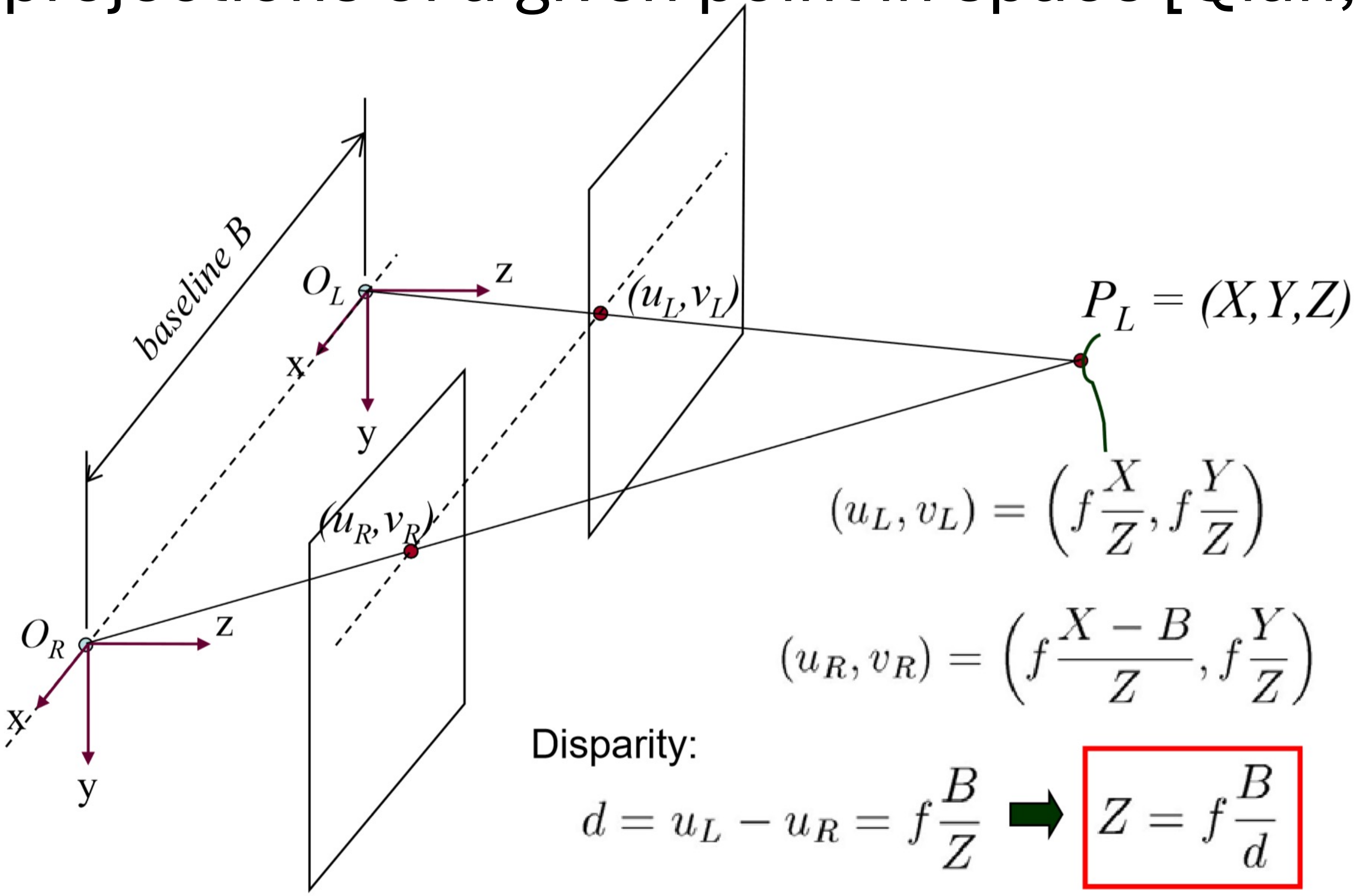
- We can always achieve this ideal geometry with stereo-rectification
- Image reprojection:
  - Reproject image planes onto a common plane parallel to the line between optical centres using suitable planar homographies
  - E.g., using homography estimation
  - Not covered in detail here



(Seitz)

Disparity-to-Depth Computation

- Disparity (human vision): the positional difference between the two retinal projections of a given point in space [Qian, Neuron '97]



(in O_L coordinate system)

$$(u_L, v_L) = \left(f\frac{X}{Z}, f\frac{Y}{Z}\right)$$

$$(u_R, v_R) = \left(f\frac{X-B}{Z}, f\frac{Y}{Z}\right)$$

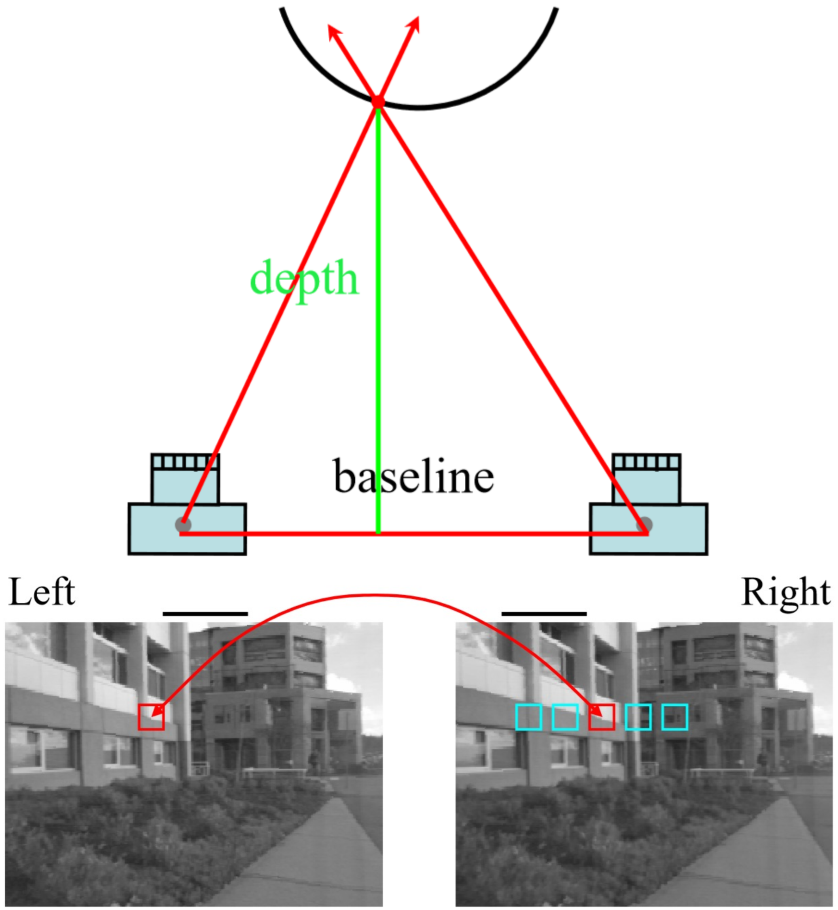
Disparity:

$$d = u_L - u_R = f\frac{B}{Z} \Rightarrow \boxed{Z = f\frac{B}{d}}$$

The Correspondence Problem

- What should we match?
  - Objects?
  - Edges?
  - Pixels?
  - Super-pixels (set of pixels)?
  - …
  - DINO features?

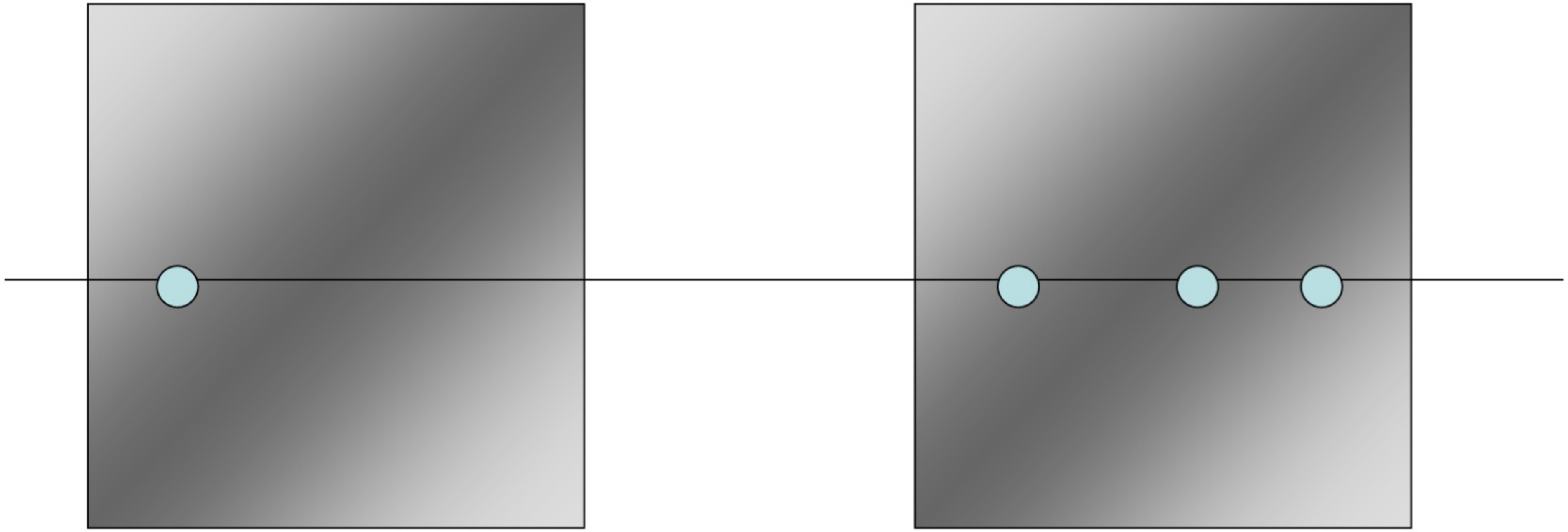
The Correspondence Problem



- Triangulate on two images of the same point to recover depth
  - Feature matching across views
  - Calibrated cameras
- Matching windows across horizontal scan lines

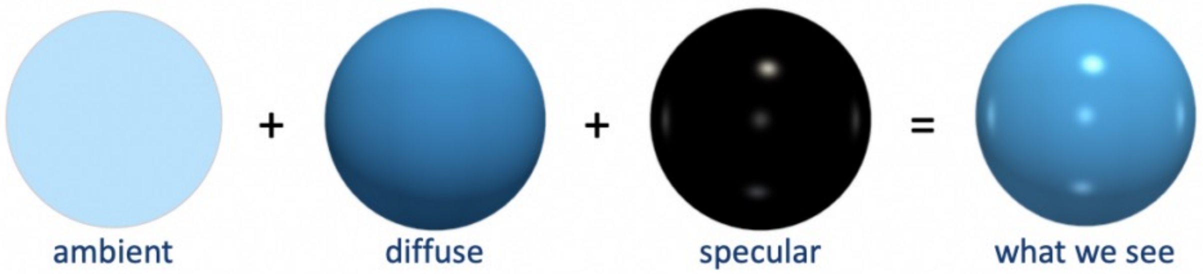
The Correspondence Problem: The Epipolar Constraint

- The epipolar constraint removes some ambiguity

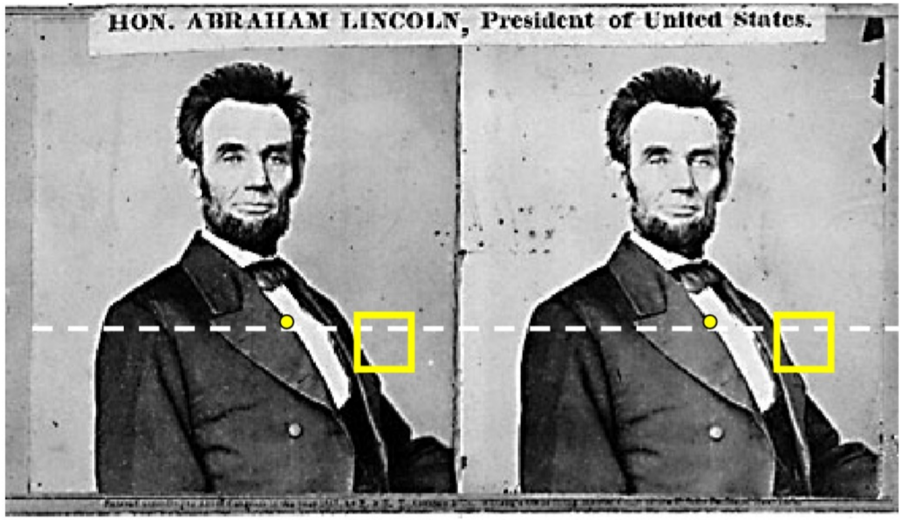


The Correspondence Problem: The Colour Constancy "Constraint"

- Same world point has same intensity (or colour) in both images
  - True for Lambertian surfaces
    - A Lambertian surface has a brightness that is independent of viewing angle
  - Violations:
    - Noise
    - Specularity
    - Non-Lambertian materials
    - Pixels that contain multiple surfaces
    - Occlusions



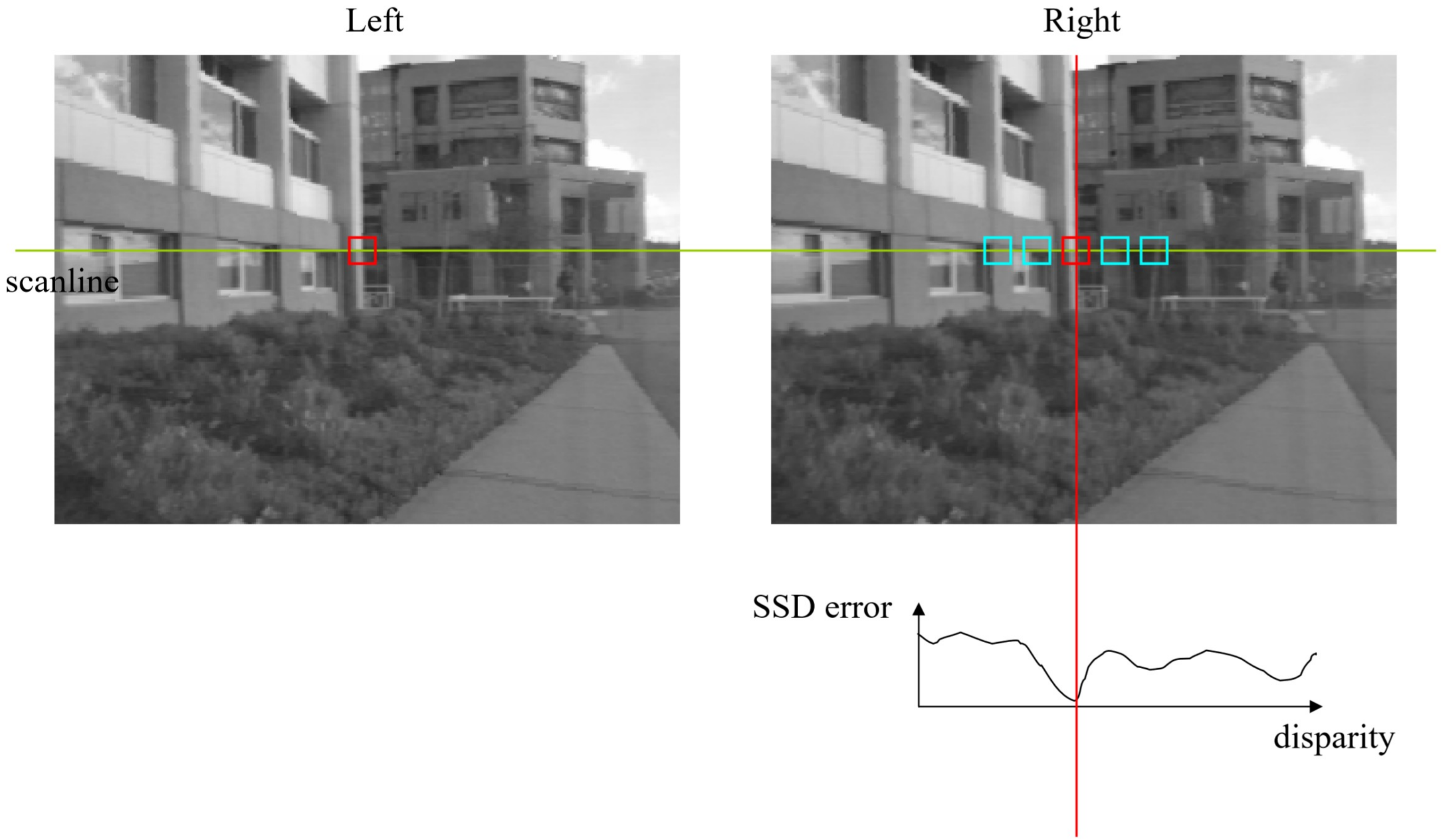
The Correspondence Problem: Pixel Matching



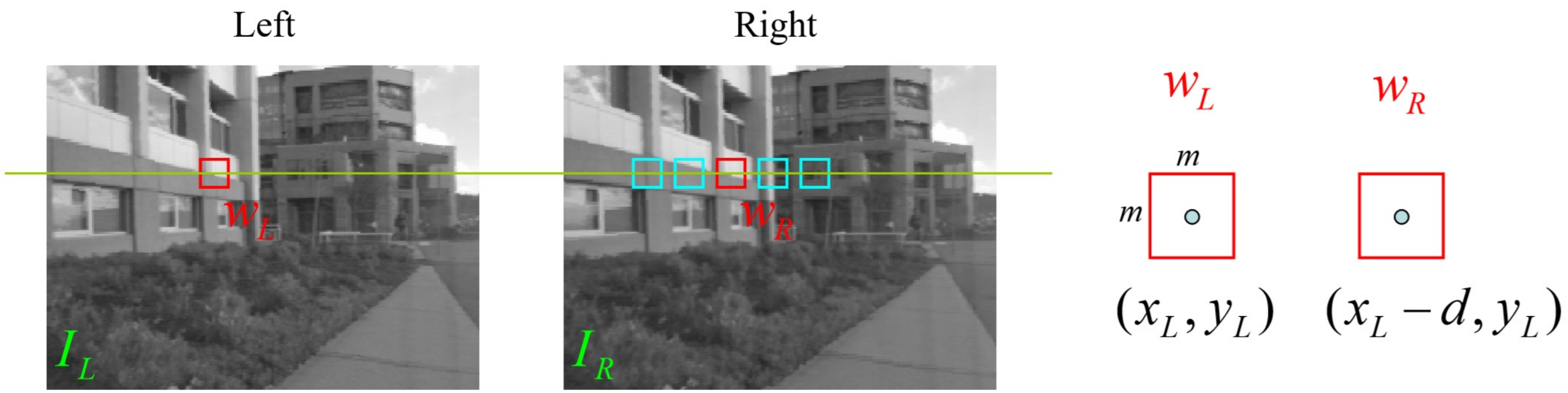
- For each epipolar line:
  - For each pixel in the left image:
    - Compare with every pixel on same epipolar line in right image
    - Pick pixel with minimum matching cost
- Still too much ambiguity: match windows instead

Image Credit: Seitz

The Correspondence Problem: Correspondence Using Correlation



Sum of Squared (Pixel) Differences



w_L and w_R are corresponding m by m windows of pixels.

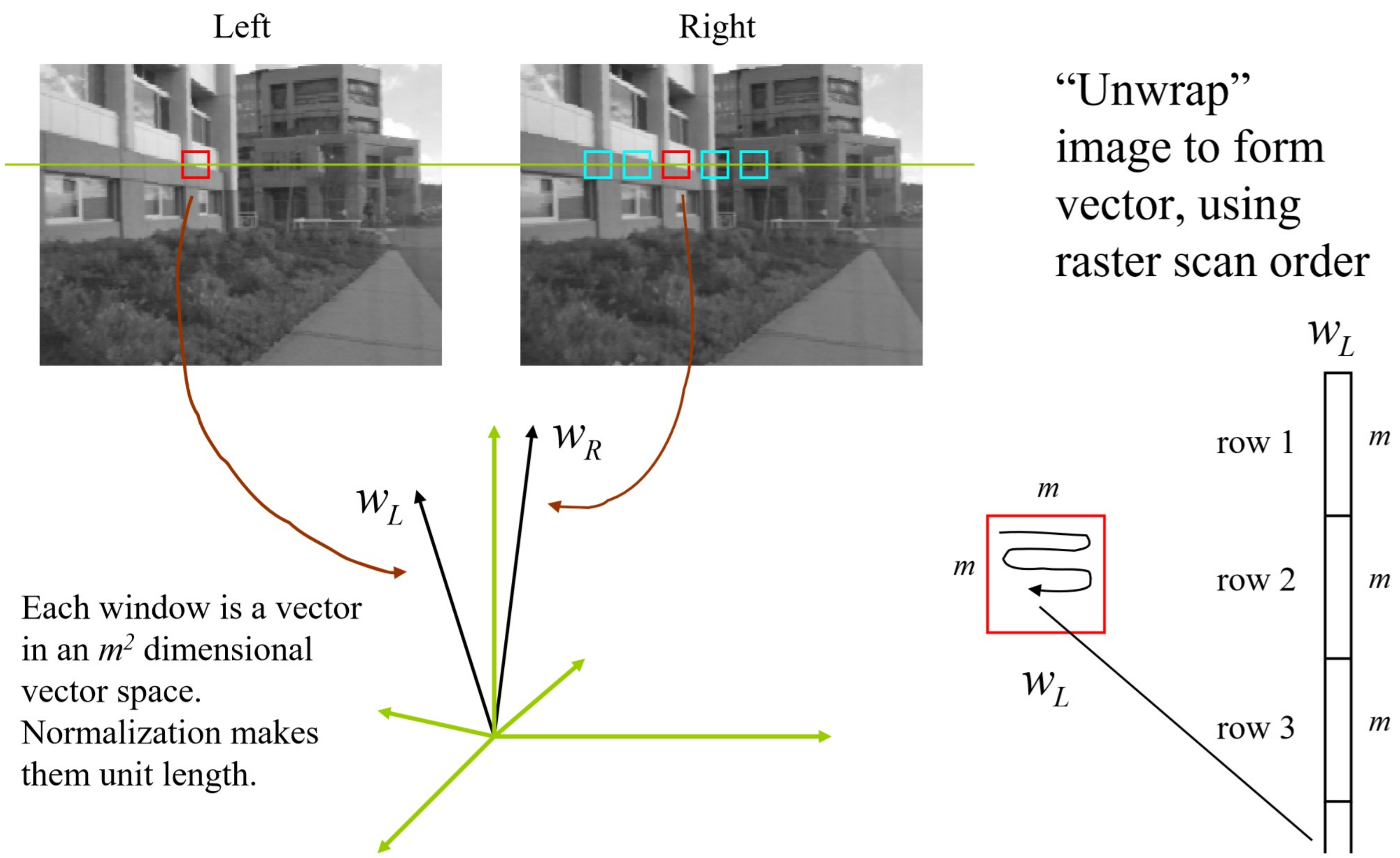
We define the window function :

$$W_m(x,y) = \{u,v \mid x - \tfrac{m}{2} \le u \le x + \tfrac{m}{2}, y - \tfrac{m}{2} \le v \le y + \tfrac{m}{2}\}$$

The SSD cost measures the intensity difference as a function of disparity :

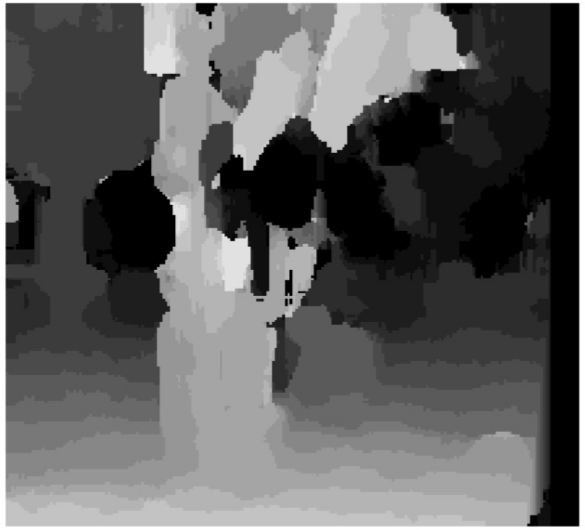
$$C_r(x,y,d) = \sum_{(u,v)\in W_m(x,y)} [I_L(u,v) - I_R(u-d,v)]^2$$

Image Patches as Vectors



Effect of Window Size

W = 3



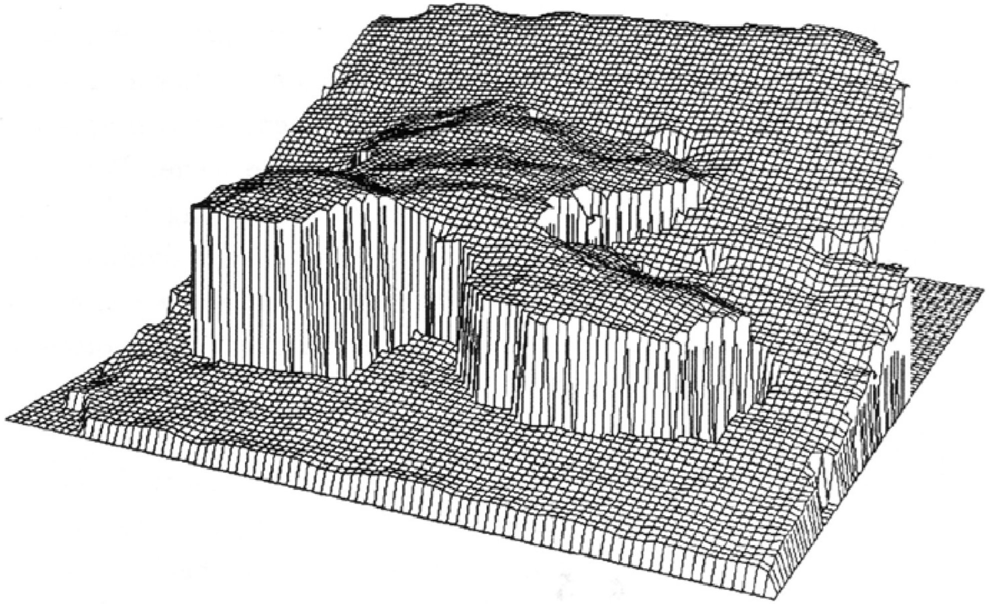
W = 20

- Improvement: use an adaptive window size (try multiple sizes and select best match)

Image Credit: Seitz

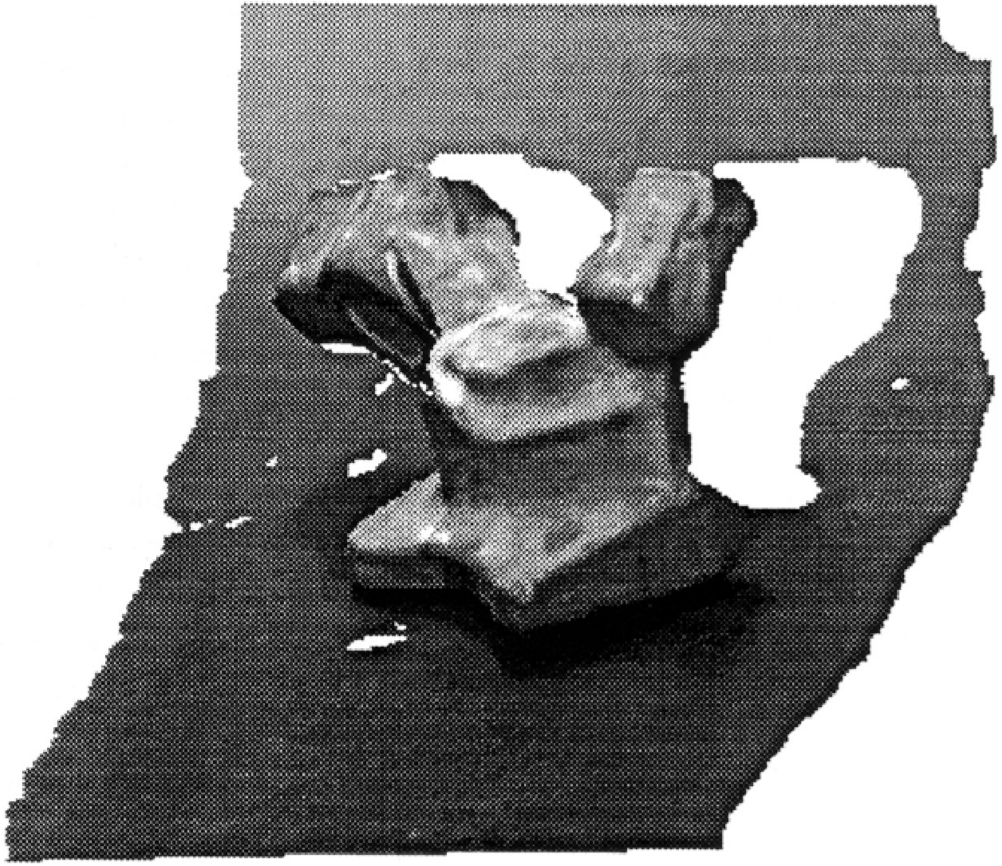
Example Stereo Pair

Left Camera



- Obtained depth map in 3D

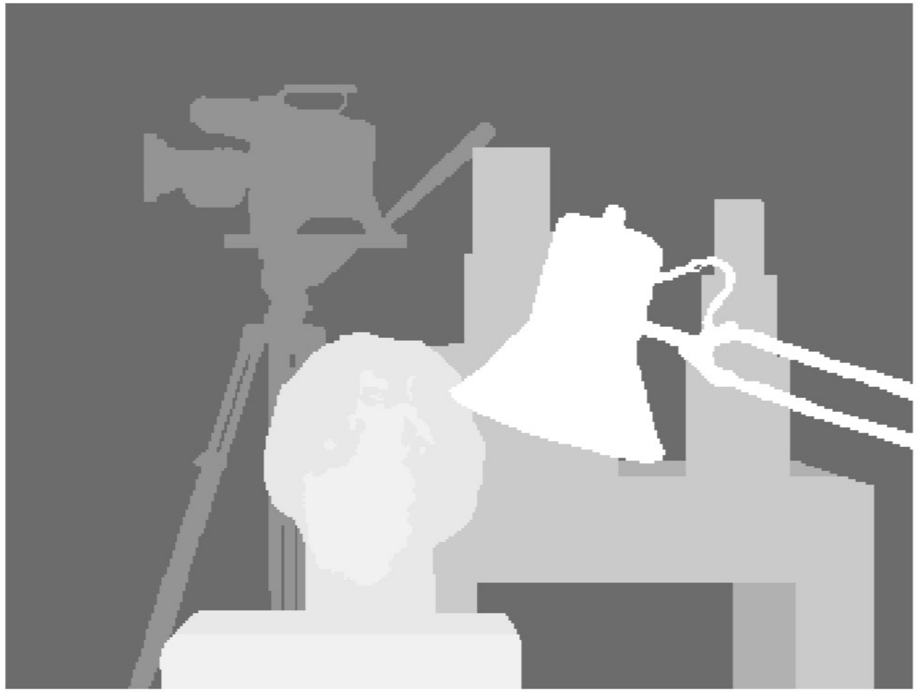
Novel View Synthesis



A Taxonomy of Stereo Algorithms

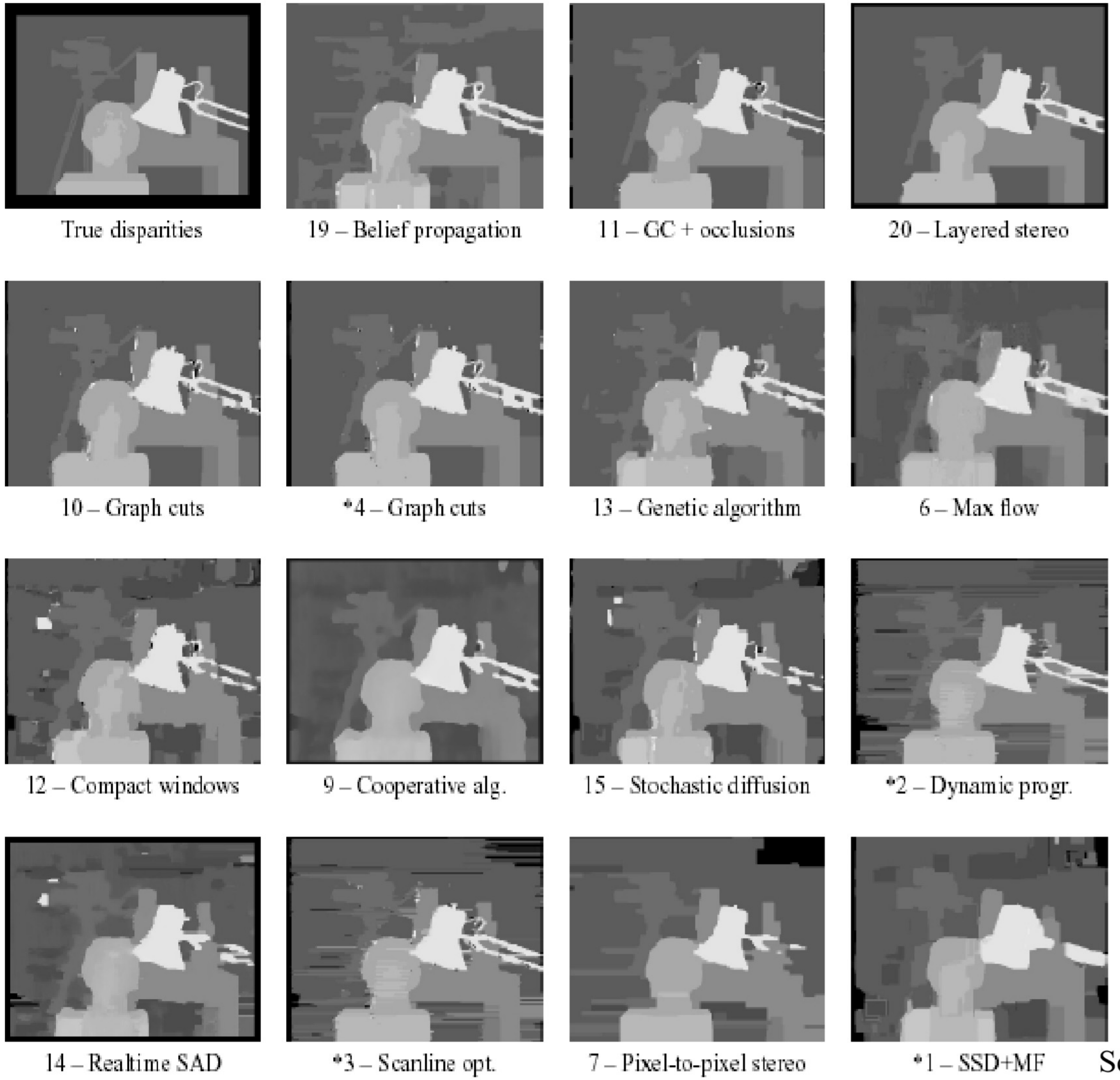
- D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, International Journal of Computer Vision, 47 (2002), pp. 7-42

Scene



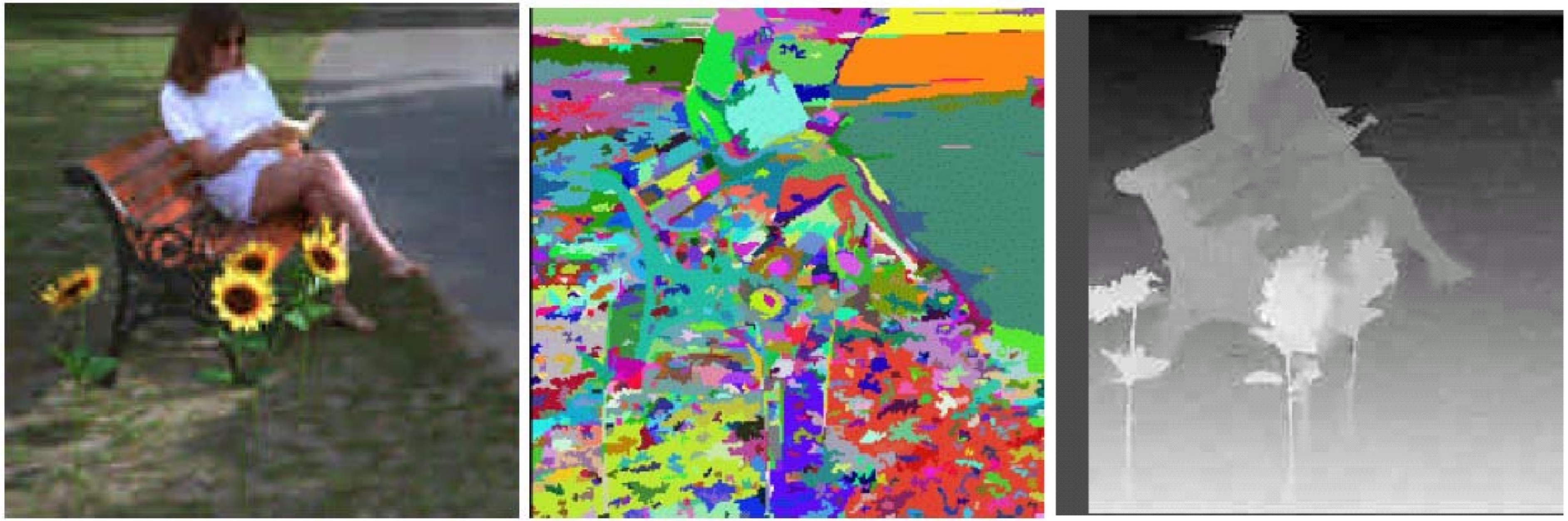
Ground truth

A Taxonomy of Stereo Algorithms



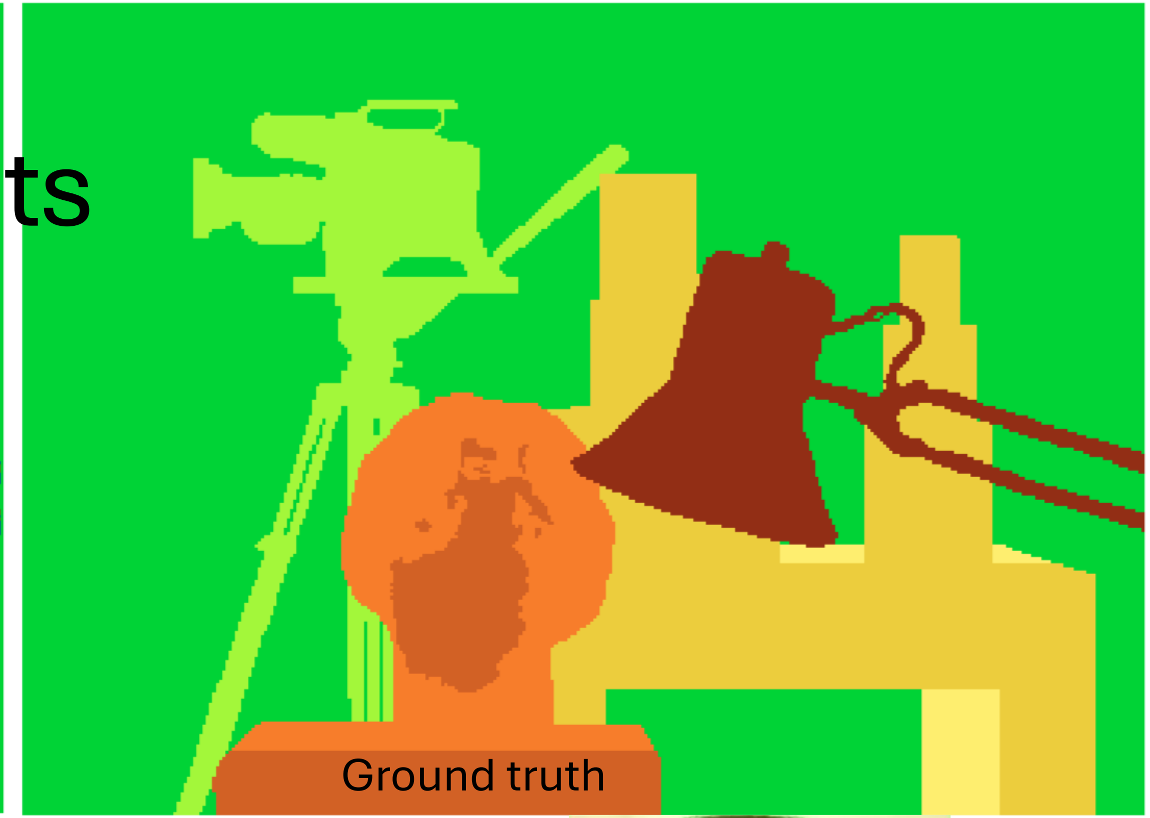
Scharstein and Szeliski

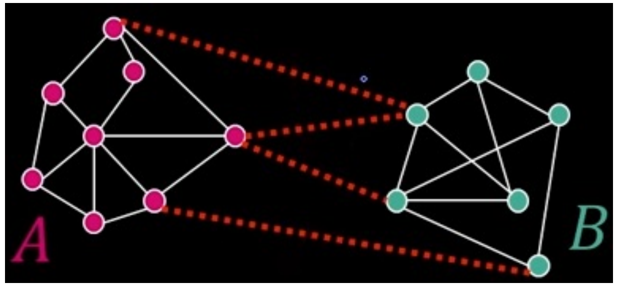
Stereo from Segmentation



Stereo from Graph-Cuts


Graph cuts


Ground truth

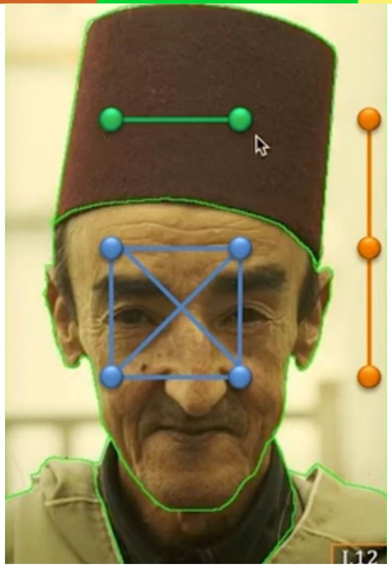


An image is represented by a graph

$$Cut(A,B) = \sum_{p\in A, q\in B} w_{pq}$$

Minimise the cost of a cut, also called min-cut

Each subgraph is an image segment



Summary of Stereo Vision

- Stereo matching + triangulation
  - The correspondence problem and the interpretation problem
- Constraints:
  - Geometry: epipolar constraint
  - Photometric: brightness constancy “constraint”

Next Week

- Multiple-view geometry: Structure-from-Motion
- Multiple-view geometry: Radiance fields (fancy triangulation)
- Multiple-view geometry: Learning-based pose-free reconstruction
- Mid-level vision: Optical flow
- Mid-level vision / single-view geometry: Shape-from-X