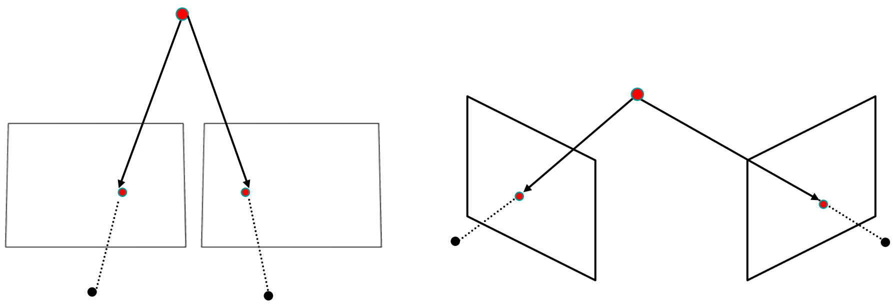
This lecture: Stereo

- Stereo Problem
- Cameras
- Geometry of two views
- Correspondence Search
- Extensions

Slide Credits: A. Zisserman, K. Grauman, B. Leibe, S. Seitz, S. Lazebnik, R. Szelisky, M. Pollefeys, K. Grauman, I. Kokkinos

General Case With Calibrated Cameras

- The two cameras need not have parallel optical axes.

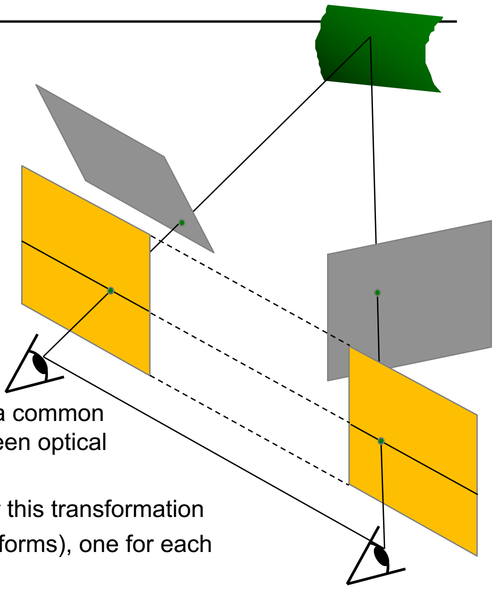


vs.

Slide credit: Kristen Grauman, Steve Seitz

Stereo Image Rectification

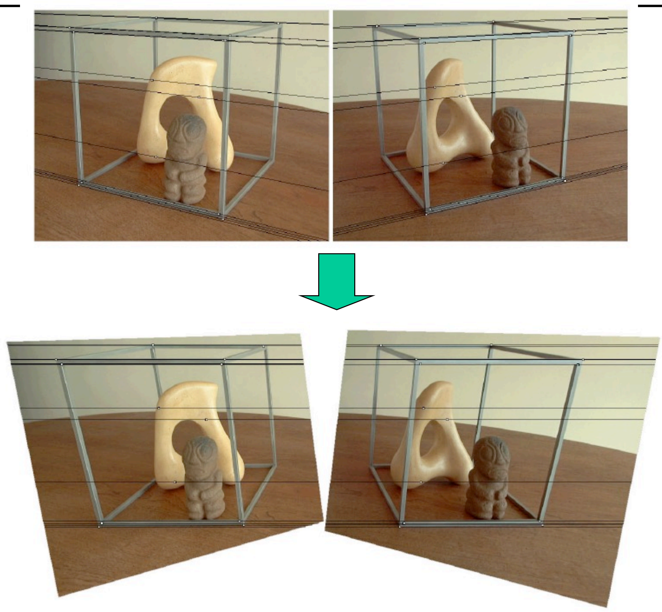
- In practice, it is convenient if image scanlines are the epipolar lines.



- Algorithm
  - Reproject image planes onto a common plane parallel to the line between optical centers
  - Pixel motion is horizontal after this transformation
  - Two homographies (3x3 transforms), one for each input image reprojection

3

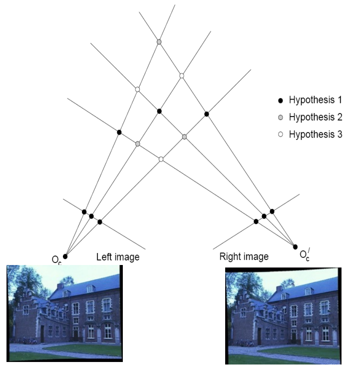
Stereo Image Rectification: Example



4

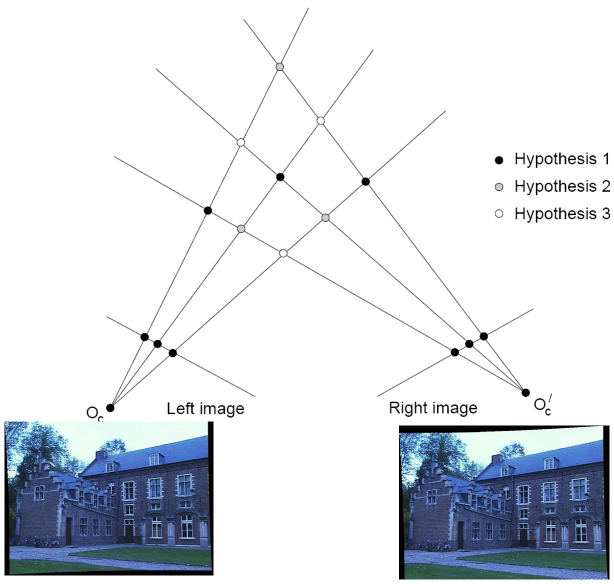
This lecture: Stereo

- Stereo Problem
- Cameras
- Geometry of two views
- Correspondence Search
  - Similarity constraint
  - Additional Constraints
- Extensions



5

Correspondence Problem



Multiple match hypotheses satisfy epipolar constraint, but which is correct?

6

Problem statement

Given: two images and their associated cameras compute corresponding image points.

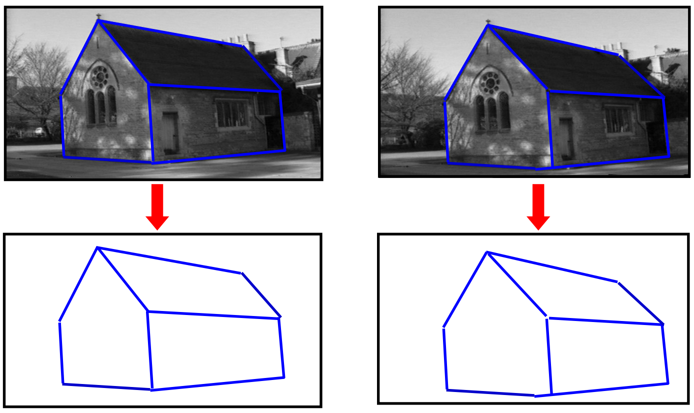
Algorithms may be classified into two types:

1. Dense: compute a correspondence at every pixel
2. Sparse: compute correspondences only for features

The methods may be top down or bottom up

7

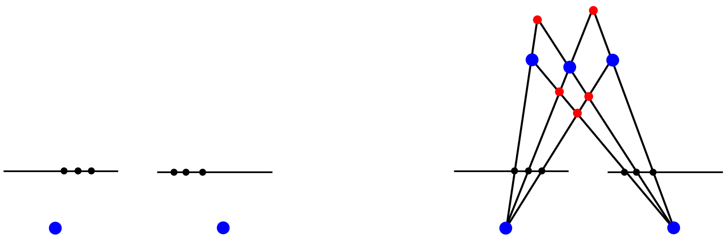
Top down matching



1. **Group model (house, windows, etc) independently in each image**
2. **Match points (vertices) between images**

8

Bottom up matching



9

Example image pair – parallel cameras

10

First image

11

Second image

12

Dense correspondence algorithm

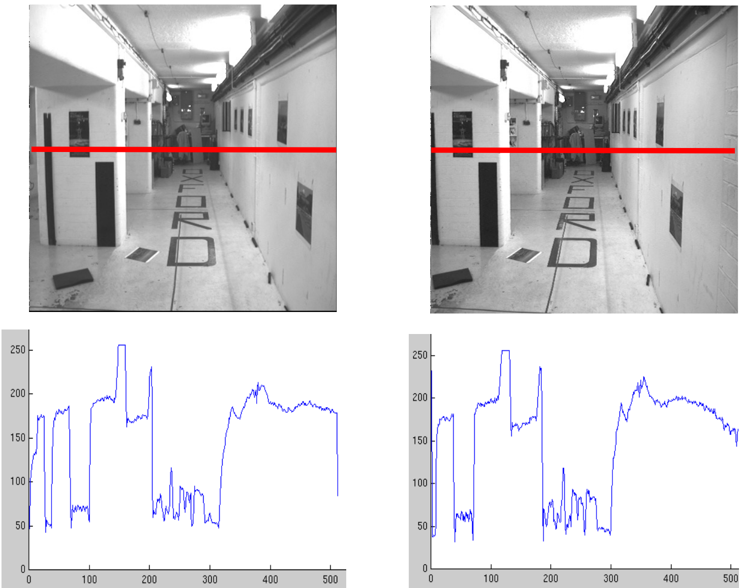
epipolar lines are corresponding rasters

Search problem (geometric constraint): **for each point in the left image, the corresponding point in the right image lies on the epipolar line (1D ambiguity)**

Disambiguating assumption (photometric constraint): **the intensity** neighbourhood **of corresponding points are similar across images**

Measure **similarity of neighbourhood intensity by cross-correlation**

13



Clear correspondence between intensities, but also noise and ambiguity

14

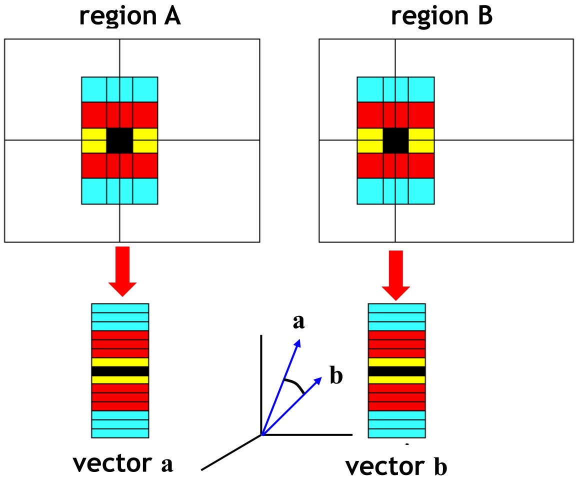
Normalized Cross Correlation

$$\text{NCC} = \frac{\sum_i \sum_j A(i,j)B(i,j)}{\sqrt{\sum_i \sum_j A(i,j)^2}\sqrt{\sum_i \sum_j B(i,j)^2}}$$

$$\mathtt{A} \rightarrow \mathbf{a}, \;\; \mathtt{B} \rightarrow \mathbf{b}$$

$$\text{NCC} = \frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

$$-1 \leq \text{NCC} \leq 1$$



15

Cross-correlation of neighbourhood regions

regions $\mathtt{A}, \mathtt{B}$, write as vectors $\mathbf{a}, \mathbf{b}$

translate so that mean is zero

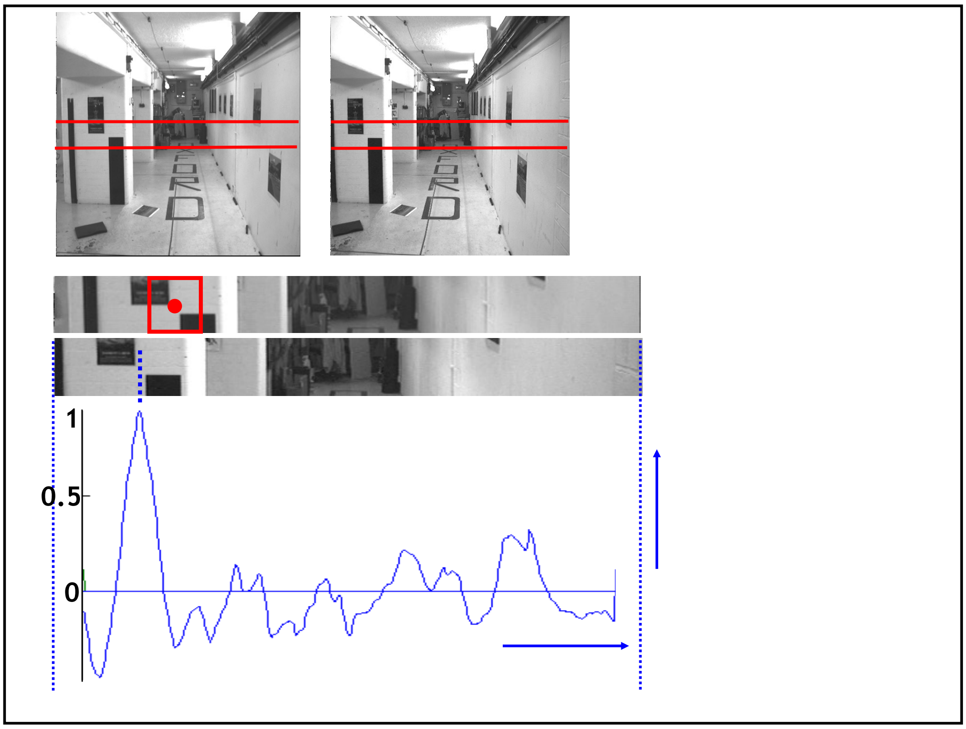
$$\mathtt{a} \rightarrow \mathtt{a} - \langle \mathbf{a} \rangle, \;\; \mathtt{b} \rightarrow \mathtt{b} - \langle \mathbf{b} \rangle$$

$$\text{cross correlation} = \frac{\mathbf{a}.\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

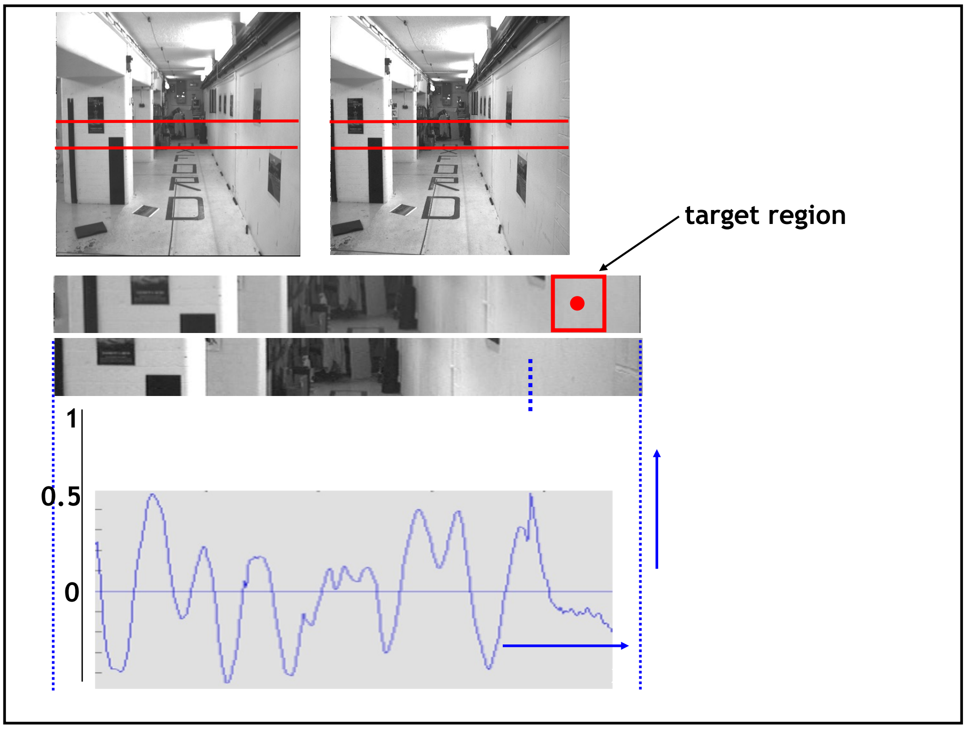
Invariant to $I \rightarrow \alpha I + \beta$

(exercise)

16



17



18

Effect of Window Size

W = 3　　W = 20

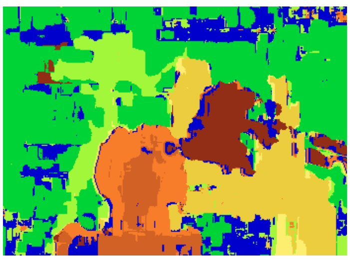
Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

19

Results with window search

Data

Window-based matching

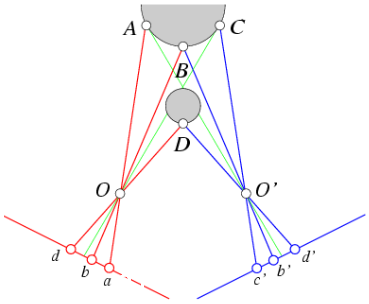


Ground truth

20

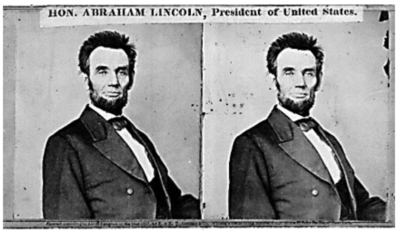
This lecture: Stereo

- Stereo Problem
- Cameras
- Geometry of two views
- Correspondence Search
  - Similarity constraint
  - Additional Constraints
- Extensions



21

Limitations of similarity constraint



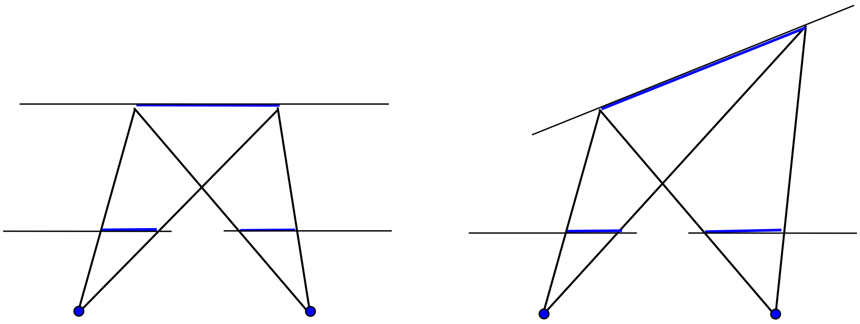
Textureless surfaces

Occlusions, repetition

Specularities

22

Foreshortening effects



fronto-parallel surface

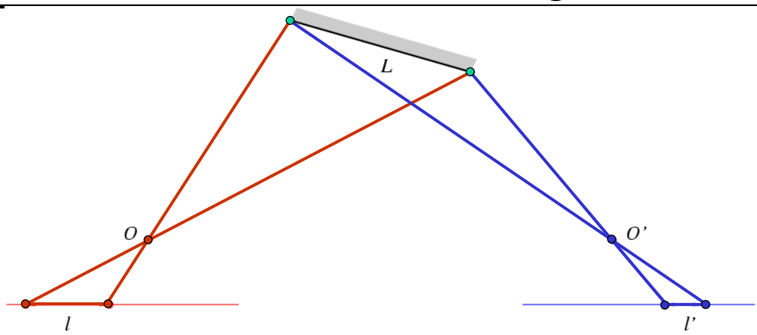
imaged length the same

slanting surface

imaged lengths differ

23

Correlation Methods: Foreshortening Problems



Solution: add a second pass using disparity estimates to warp the correlation windows, e.g. Devernay and Faugeras (1994).

Reprinted from "Computing Differential Properties of 3D Shapes from Stereopsis without 3D Models," by F. Devernay and O. Faugeras, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (1994). © 1994 IEEE.

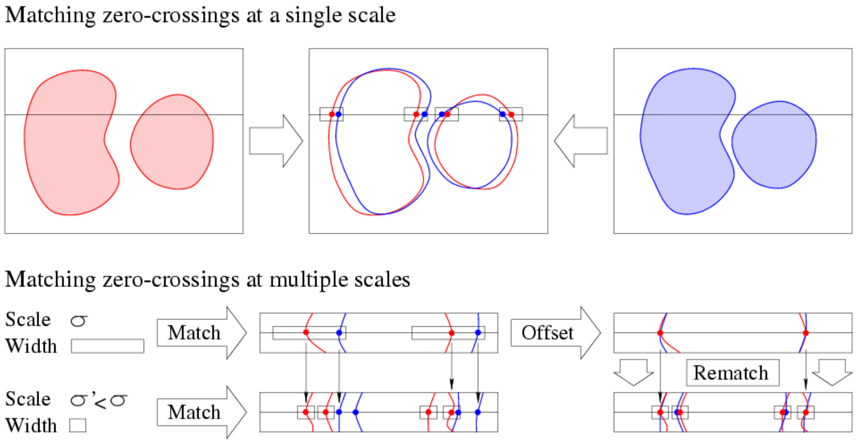
24

Feature based approaches

- Instead of trying to match every single pixel in the left image to its mate in the right image we could simply extract features such as edges in the left images and look for matches for those features only

25

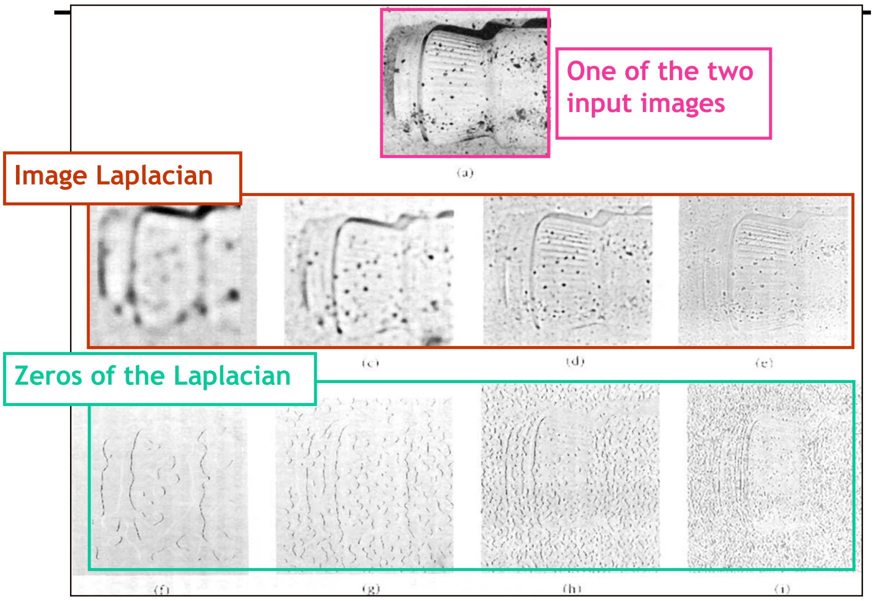
Multi-Scale Edge Matching (Marr, Poggio and Grimson, 1979-81)



·Edges are found by repeatedly smoothing the image and detecting the zero crossings of the second derivative (Laplacian).
· Matches at coarse scales are used to offset the search for matches at fine scales (equivalent to eye movements).

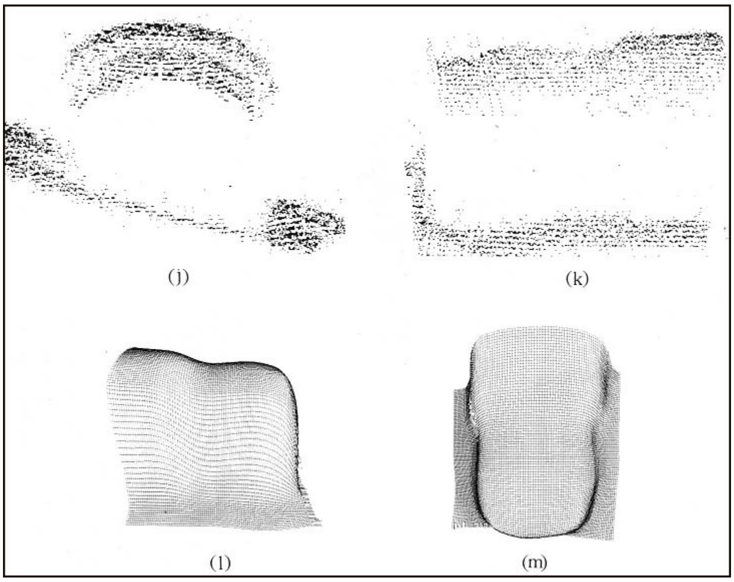
26

Multi-Scale Edge Matching (Marr, Poggio and Grimson, 1979-81)



27

Multi-Scale Edge Matching (Marr, Poggio and Grimson, 1979-81)



28

Correspondence

- Since stereo matching is an ill-posed problem global constraints are often employed
- These constraints effectively represent prior knowledge about the scenes

29

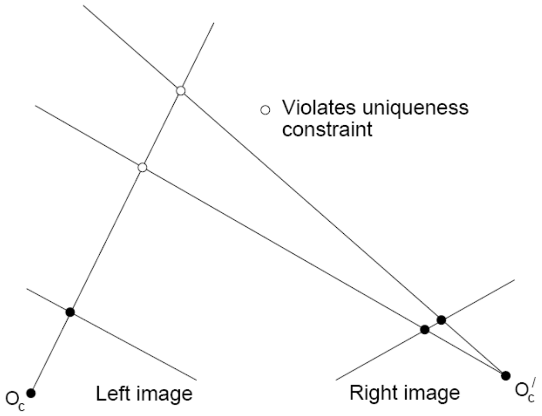
Correspondence Problem

- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Image regions for the matches are similar in appearance
- Additional constraints!

30

Uniqueness

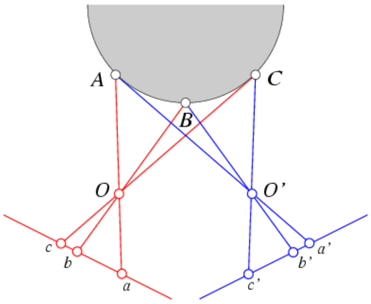
- For opaque objects, up to one match in right image for every point in left image



31

Non-Local Constraints

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views



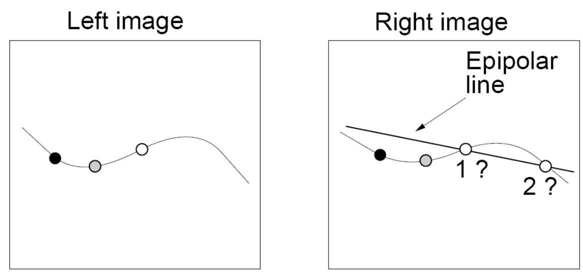
32

Ordering constraint may not always hold

32

Disparity Gradient

- Assume piecewise continuous surface, so want disparity estimates to be locally smooth

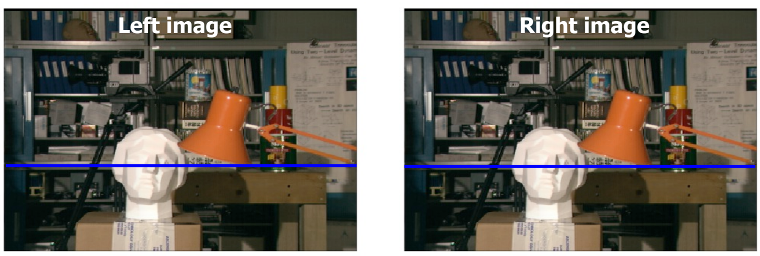


Given matches ● and ◉, point ○ in the left image must match point 1 in the right image. Point 2 would exceed the disparity gradient limit.

33

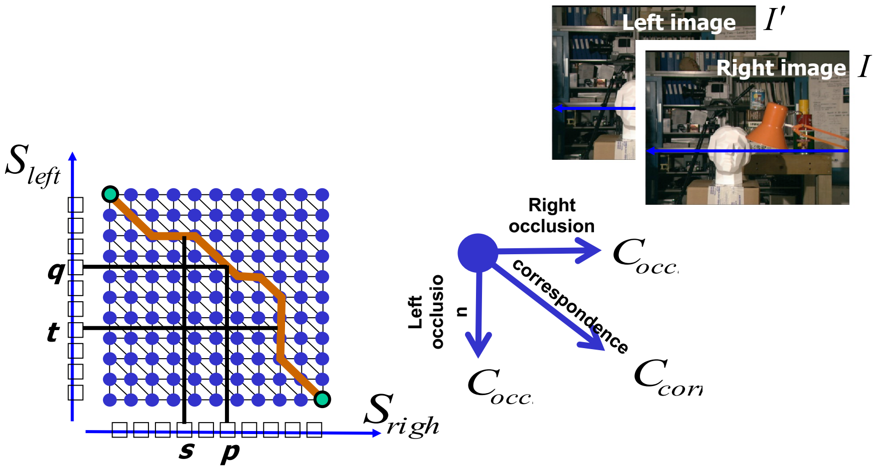
Scanline Stereo

- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently



34

“Shortest Path” for Scanline Stereo



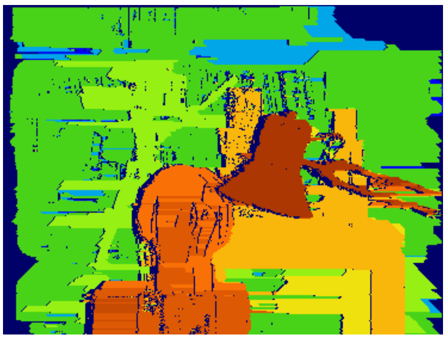
Can be implemented with dynamic programming

35 **[Ohta & Kanade ’85, Cox et al. ’96]**

35

Coherent Stereo on 2D Grid

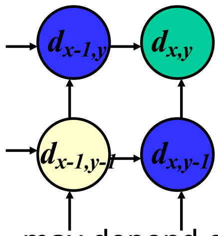
- Scanline stereo generates streaking artifacts



36

Dynamic Programming

- Can we apply this trick in 2D as well?

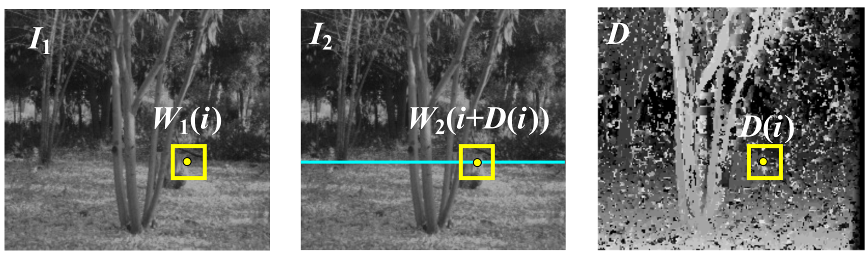


- No: $d_{x,y-1}$ and $d_{x-1,y}$ may depend on different values of $d_{x-1,y-1}$

⇒ Can't use dynamic programming to find spatially coherent disparities/ correspondences on a 2D grid!

⇒ But… we will later see methods for performing inference on 2D Markov Random Fields.

Stereo Matching as Energy Minimization



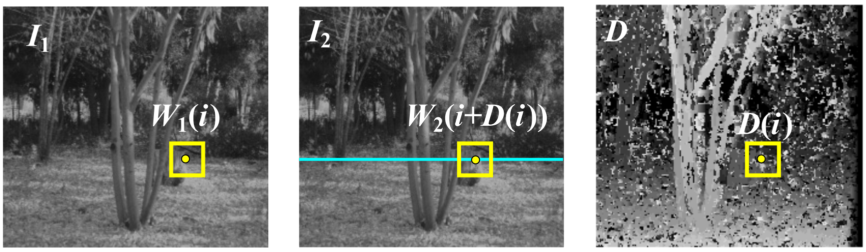
MAP estimate of disparity image *D*: $P(D \mid I_1, I_2) \propto P(I_1, I_2 \mid D) P(D)$

$$-\log P(D \mid I_1, I_2) \propto -\log P(I_1, I_2 \mid D) - \log P(D)$$

$$\boxed{E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)}$$

$$\boxed{E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2} \quad \boxed{E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)}$$

Stereo Matching as Energy Minimization



$$E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)$$

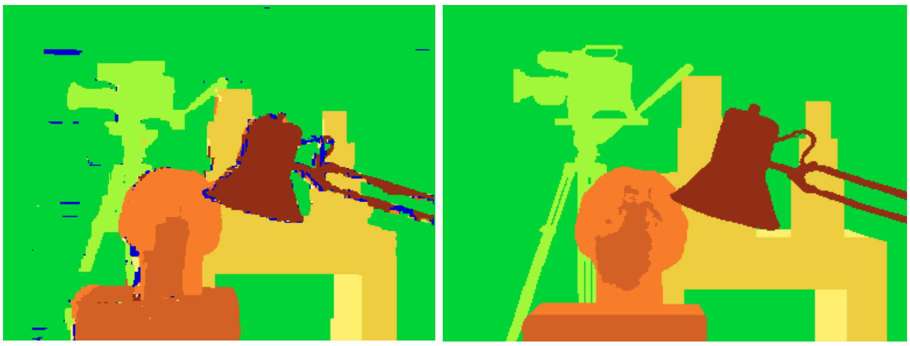
$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2$$

$$E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

- Energy functions of this form can be minimized using *Graph Cuts.*

39

Graph Cuts Results



Graph cuts　　　Ground truth

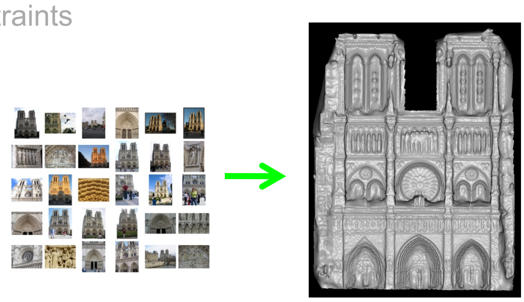
Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](), PAMI 2001.

40

40

This lecture: Stereo

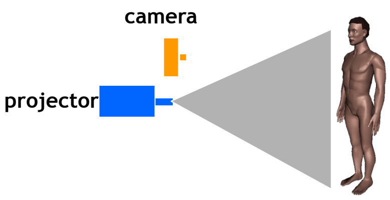
- Stereo Problem
- Cameras
- Geometry of two views
- Correspondence Search
  - Similarity constraint
  - Additional Constraints
- Extensions



41

Active stereo with structured light

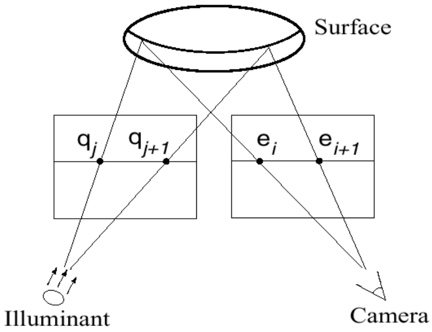
- Project "structured" light patterns onto the object
  - simplifies the correspondence problem
  - Allows us to use only one camera

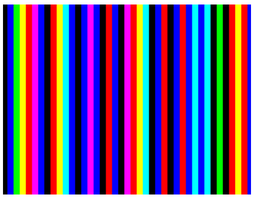


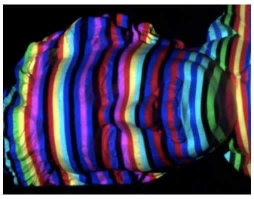
L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.] *3DPVT* 2002

42

Active stereo with structured light



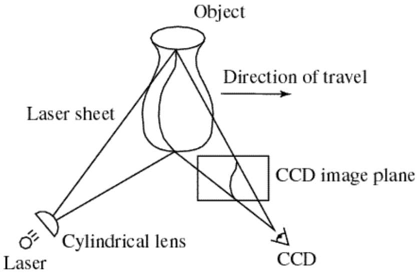


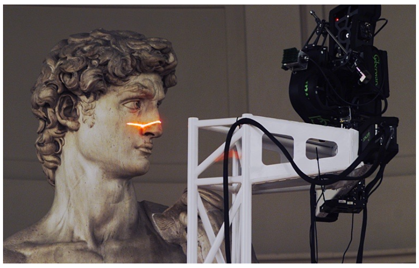


L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002

43

Laser scanning





Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning

Source: S. Seitz

44

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

45

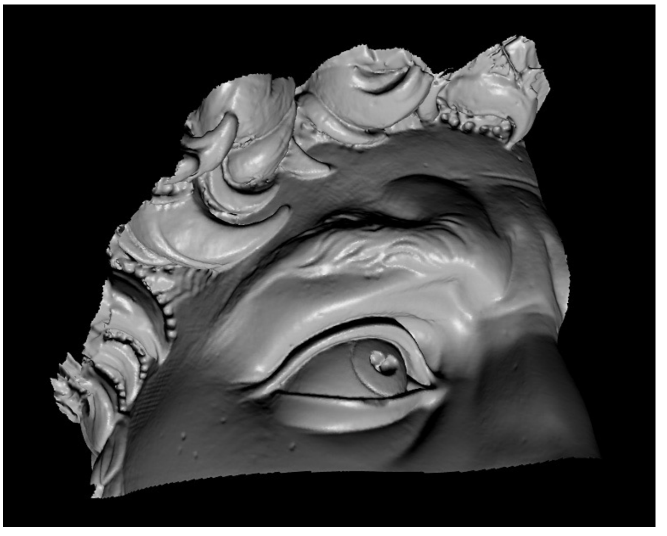
Laser sca

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

46

Laser scanned models

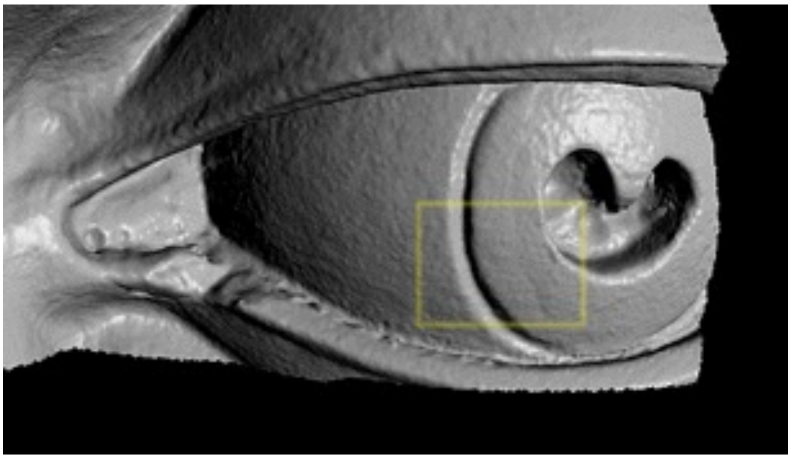


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

47

Laser scanned models

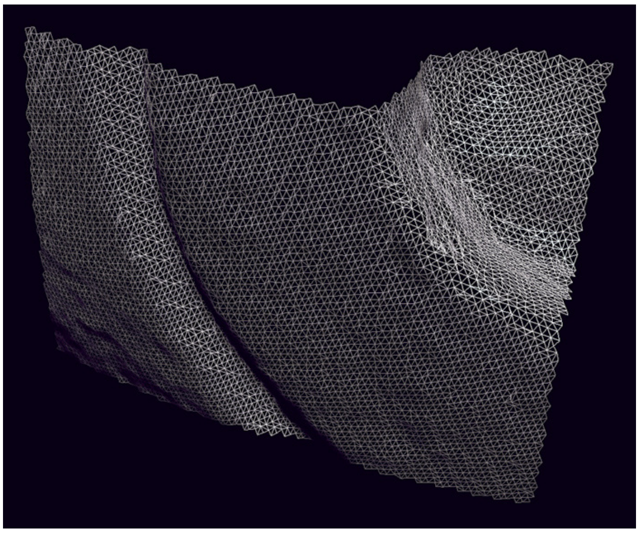


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

48

Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

49

Real-time stereo

Nomad robot searches for meteorites in Antartica
http://www.frc.ri.cmu.edu/projects/meteorobot/index.html

- **Used for robot navigation (and other tasks)**
  - **Several software-based real-time stereo techniques have been developed (most based on simple discrete search)**

50

50

Stereo reconstruction pipeline

- **Steps**
  - **Calibrate cameras**
  - **Rectify images**
  - **Compute disparity**
  - **Estimate depth**

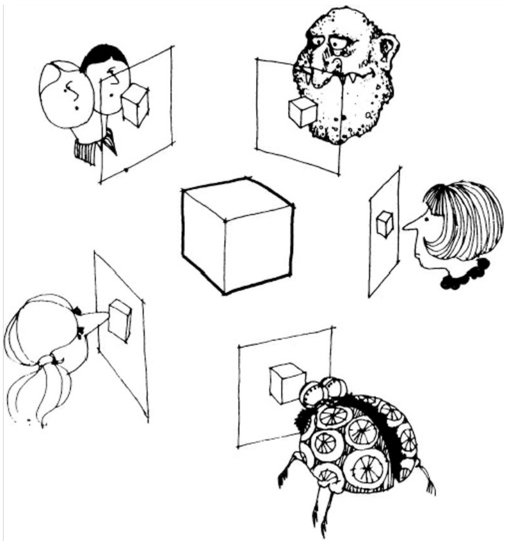
What will cause errors?

- Camera calibration errors
- Poor image resolution
- Occlusions
- Violations of brightness constancy (specular reflections)
- Large motions
- Low-contrast image regions

51

51

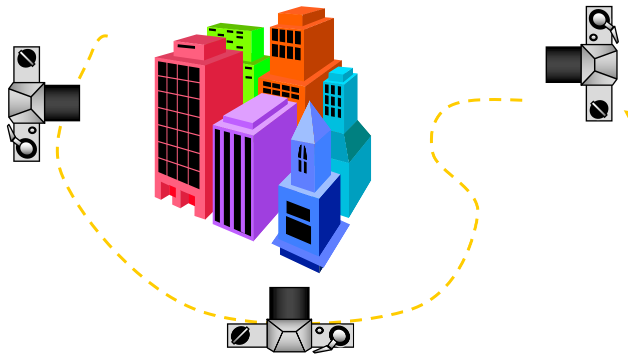
Multi-view stereo ?



52

52

Later in this class: structure from motion



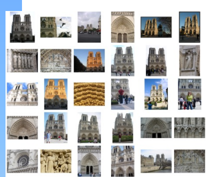
Unknown camera viewpoints

Recover both 3D shape and camera motion
Huge importance for robotics

53

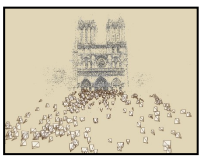
53

Structure from motion + the internet



Input photographs

Scene reconstruction



Relative camera positions and orientations

Point cloud

Sparse correspondence

Photo Explorer

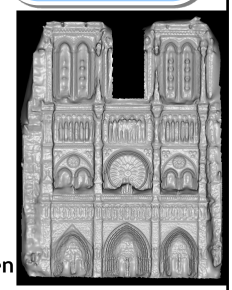


Photo Tourism: Exploring Photo Collections in 3D, Noah Snavely, Steven Seitz, Richard Szeliski, SIGGRAPH 2006
Multi-View Stereo for Community Photo Collections Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, Steven M. Seitz, ICCV 2007

54

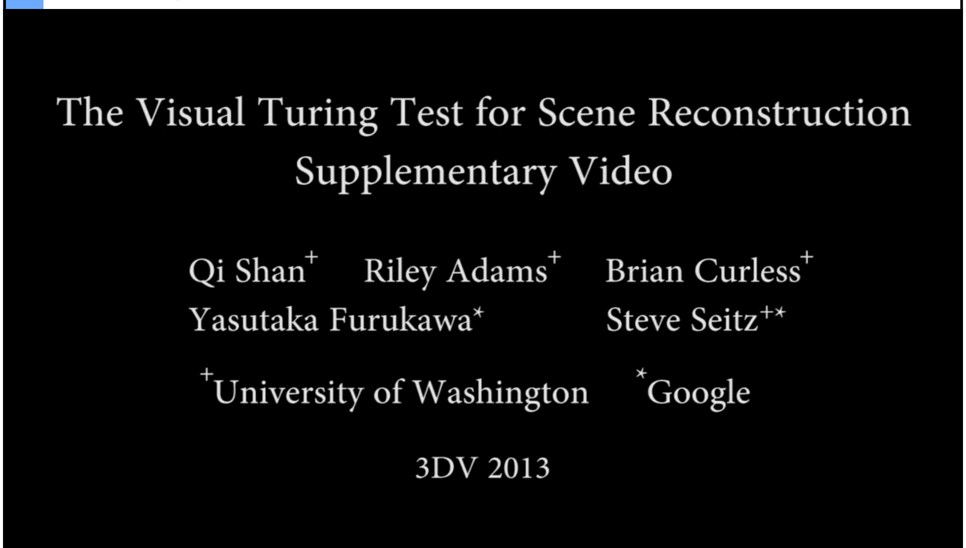
Something to play with:

- Finding Paths through the World's Photos
  - http://phototour.cs.washington.edu/findingpaths/
  - http://www.youtube.com/watch?v=gLLzV5qeKyk&fmt=18
- Navigating the world's photographs
  - Google Tech Talk:
    - http://video.google.com/videoplay?docid=-5778605234686979545

55

55

Fundamental Matrix + Dense correspondence



56

SIFT + Fundamental Matrix + RANSAC

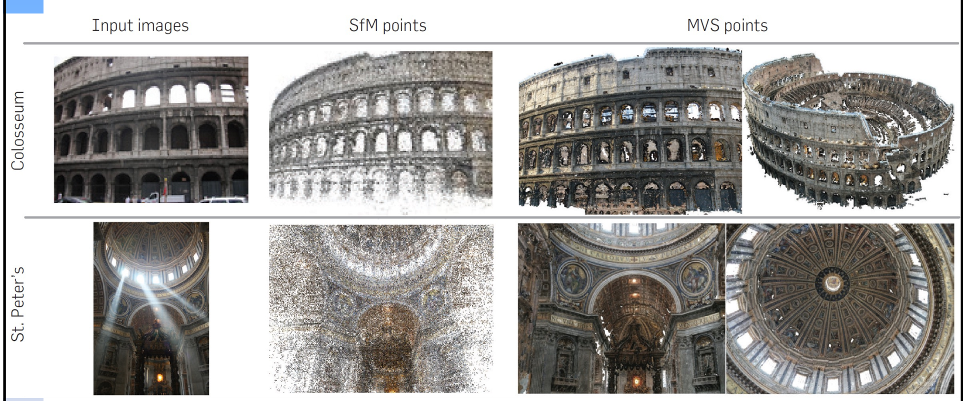
Despite their scale invariance and robustness to appearance changes, SIFT features are *local* and do not contain any global information about the image or about the location of other features in the image. Thus feature matching based on SIFT features is still prone to errors. However, since we assume that we are dealing with rigid scenes, there are strong geometric constraints on the locations of the matching features and these constraints can be used to clean up the matches. In particular, when a rigid scene is imaged by two pinhole cameras, there exists a 3×3 matrix F, the *Fundamental matrix*, such that corresponding points x_{ij} and x_{ik} (represented in homogeneous coordinates) in two images j and k satisfy[10]:

$$x_{ij}^{\top} F x_{ij} = 0. \tag{3}$$

A common way to impose this constraint is to use a greedy randomized algorithm to generate suitably chosen random estimates of F and choose the one that has the largest support among the matches, i.e., the one for which the most matches satisfy (3). This algorithm is called Random Sample Consensus (RANSAC)[6] and is used in many computer vision problems.

57

Sparse to Dense Correspodence



58

Unsupervised Monocular Depth Estimation with Left-Right Consistency

Clément Godard[1] Oisin Mac Aodha[2] Gabriel J. Brostow[1]

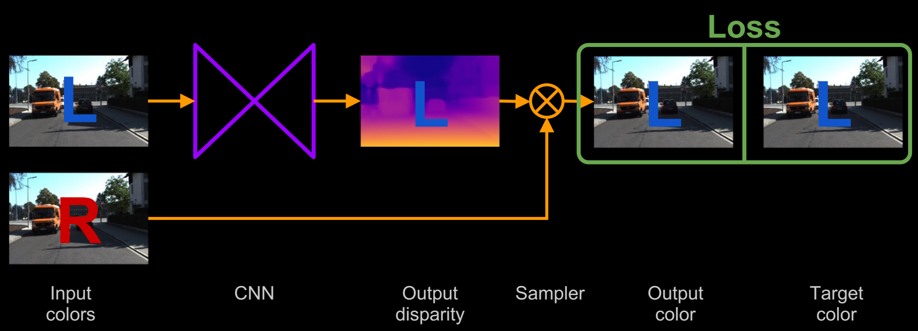
[1]University College London [2]Caltech

CVPR July 21-26 HONOLULU 2017

UCL

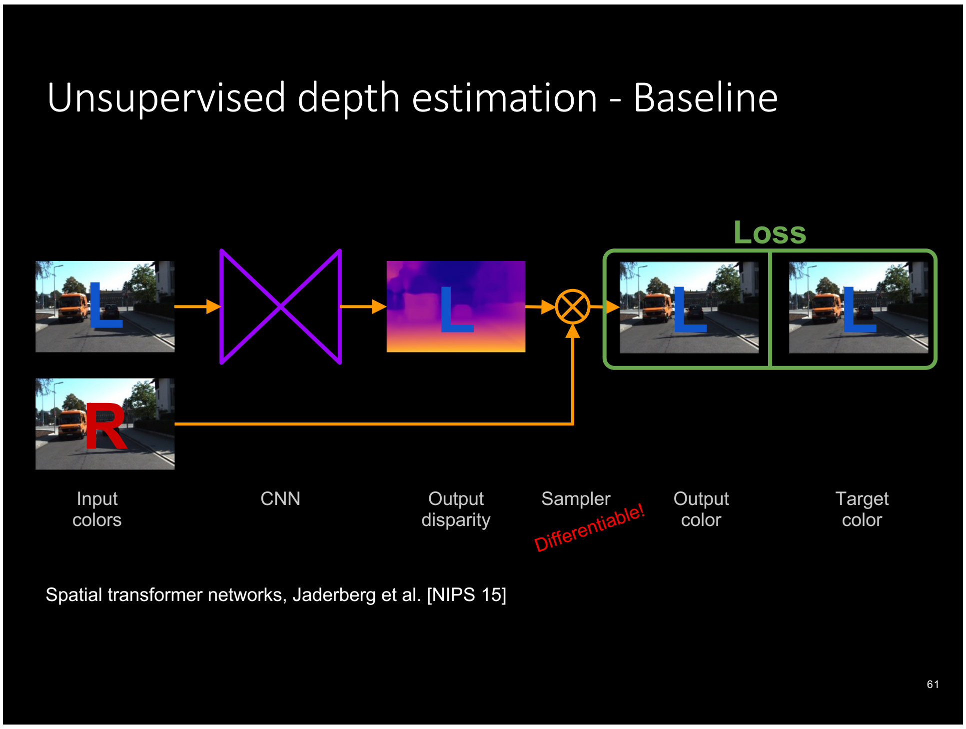
59

Unsupervised depth estimation - Concept

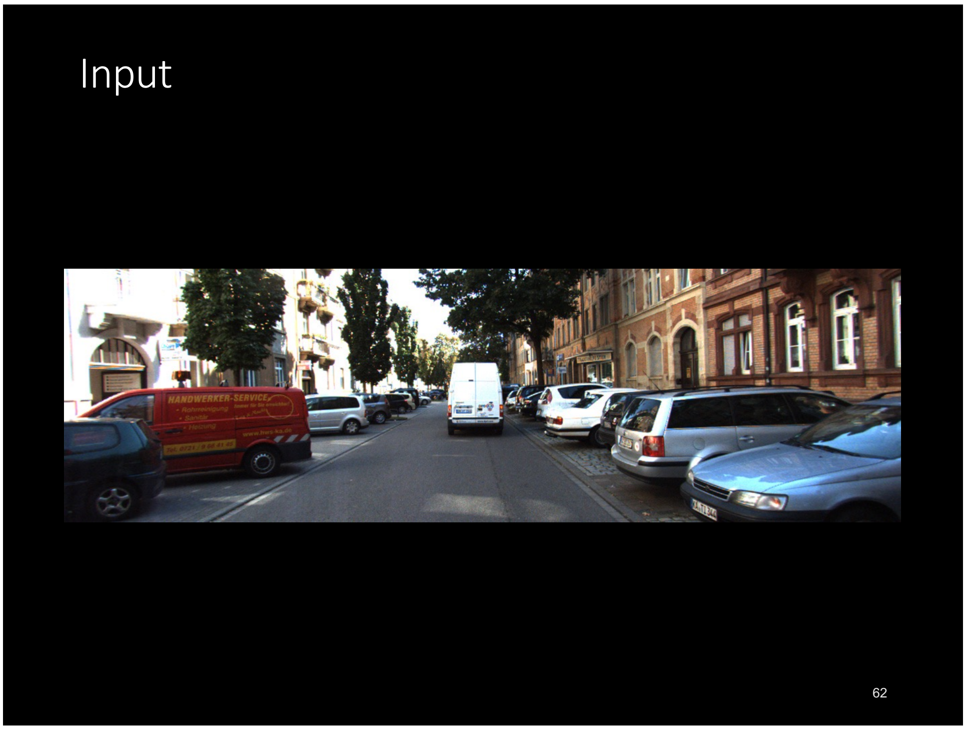


60

60


Unsupervised depth estimation - Baseline
Loss
Input
colors
CNN
Output
disparity
Sampler
Differentiable!
Output
color
Target
color
Spatial transformer networks, Jaderberg et al. [NIPS 15]
61

61


Input
62

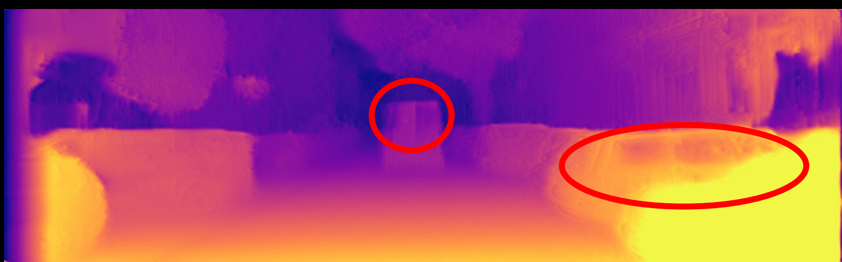
62

Baseline

63

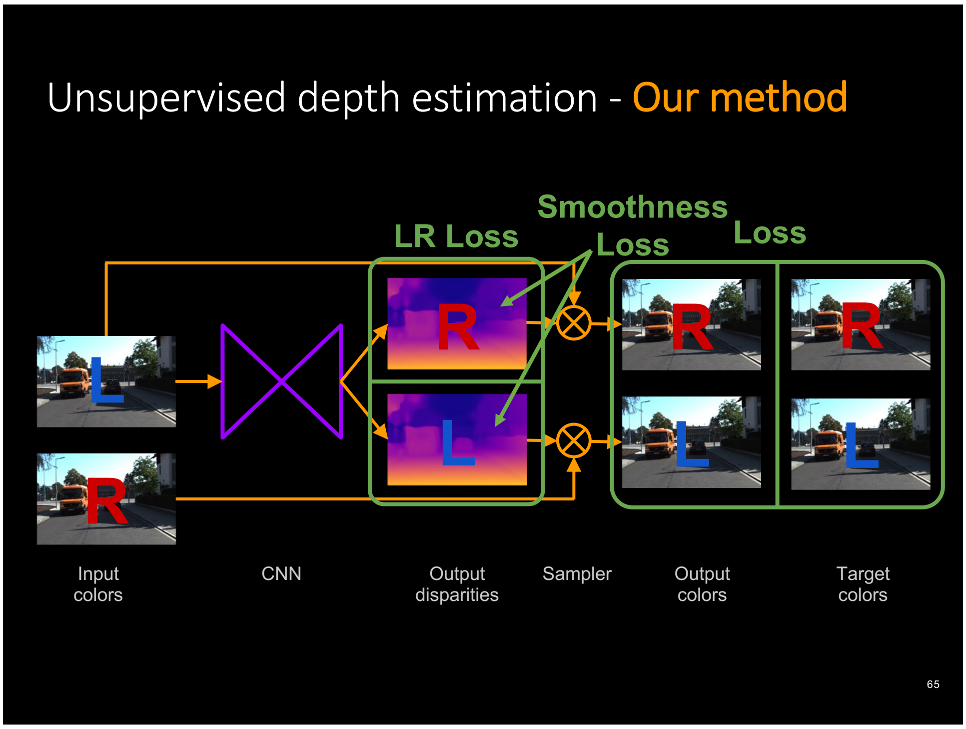
63

Our method



64

64


Unsupervised depth estimation - Our method
Smoothness
Loss
LR Loss
Loss
R
L
R
L
R
L
R
L
Input
colors
CNN
Output
disparities
Sampler
Output
colors
Target
colors
65

65