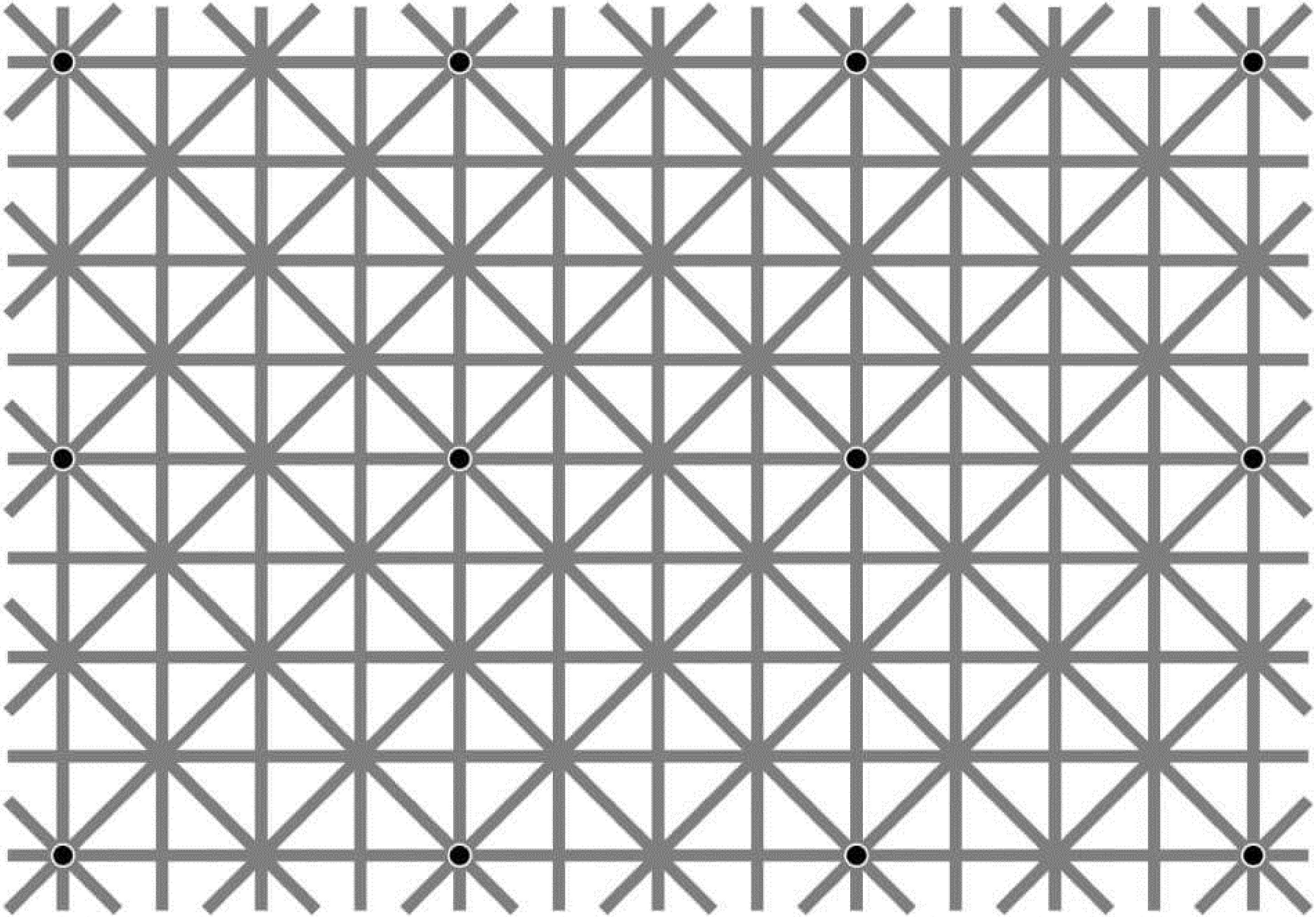


Ninio, J. and Stevens, K. A. (2000) Variations on the Hermann grid: an extinction illusion. Perception, 29, 1209-1217.

Perception, 2000, volume 29, pages 1209 – 1217

DOI:10.1068/p2985

Variations on the Hermann grid: an extinction illusion

Jacques Ninio

Laboratoire de Physique Statistique[1], École Normale Supérieure, 24 rue Lhomond, 75231 Paris cedex 05, France; e-mail: jacques.ninio@lps.ens.fr

Kent A Stevens

Department of Computer Science, Deschutes Hall, University of Oregon, Eugene, OR 97403, USA; e-mail: kent@cs.uoregon.edu

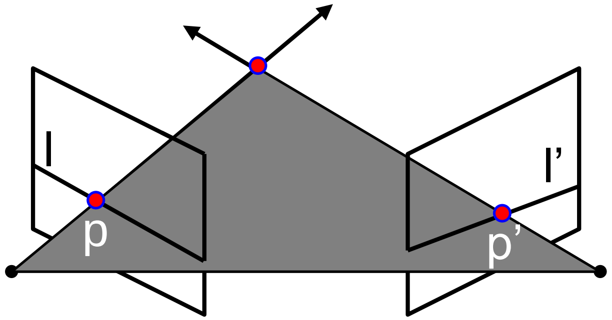
Received 21 September 1999, in revised form 21 June 2000

Abstract. When the white disks in a scintillating grid are reduced in size, and outlined in black, they tend to disappear. One sees only a few of them at a time, in clusters which move erratically on the page. Where they are not seen, the grey alleys seem to be continuous, generating grey crossings that are not actually present. Some black sparkling can be seen at those crossings where no disk is seen. The illusion also works in reverse contrast.

The Hermann grid (Brewster 1844; Hermann 1870) is a robust illusion. It is classically presented as a two-dimensional array of black squares, separated by rectilinear alleys. It is thought to be caused by processes of local brightness computation in arrays of

Fundamental matrix

Let *p* be a point in left image, *p'* in right image



Epipolar relation

- *p* maps to epipolar line *l'*
- *p'* maps to epipolar line *l*

Epipolar mapping described by a 3x3 matrix *F*

$$p'^T F p = 0$$

Fundamental matrix

This matrix F is called

- the “Essential Matrix”
  - when image intrinsic parameters are known
- the “Fundamental Matrix”
  - more generally (uncalibrated case)

Can solve for F from point correspondences

- Each (p, p’) pair gives one linear equation in entries of F

$$p'^T F p = 0$$

- F has 9 entries, but really only 7 or 8 degrees of freedom.

The scale of algorithm name quality

better

RANSAC

SIFT

Deep Learning

Optical Flow

Hough Transform

Neural Networks

Essential and Fundamental Matrix

Dynamic Programming

worse

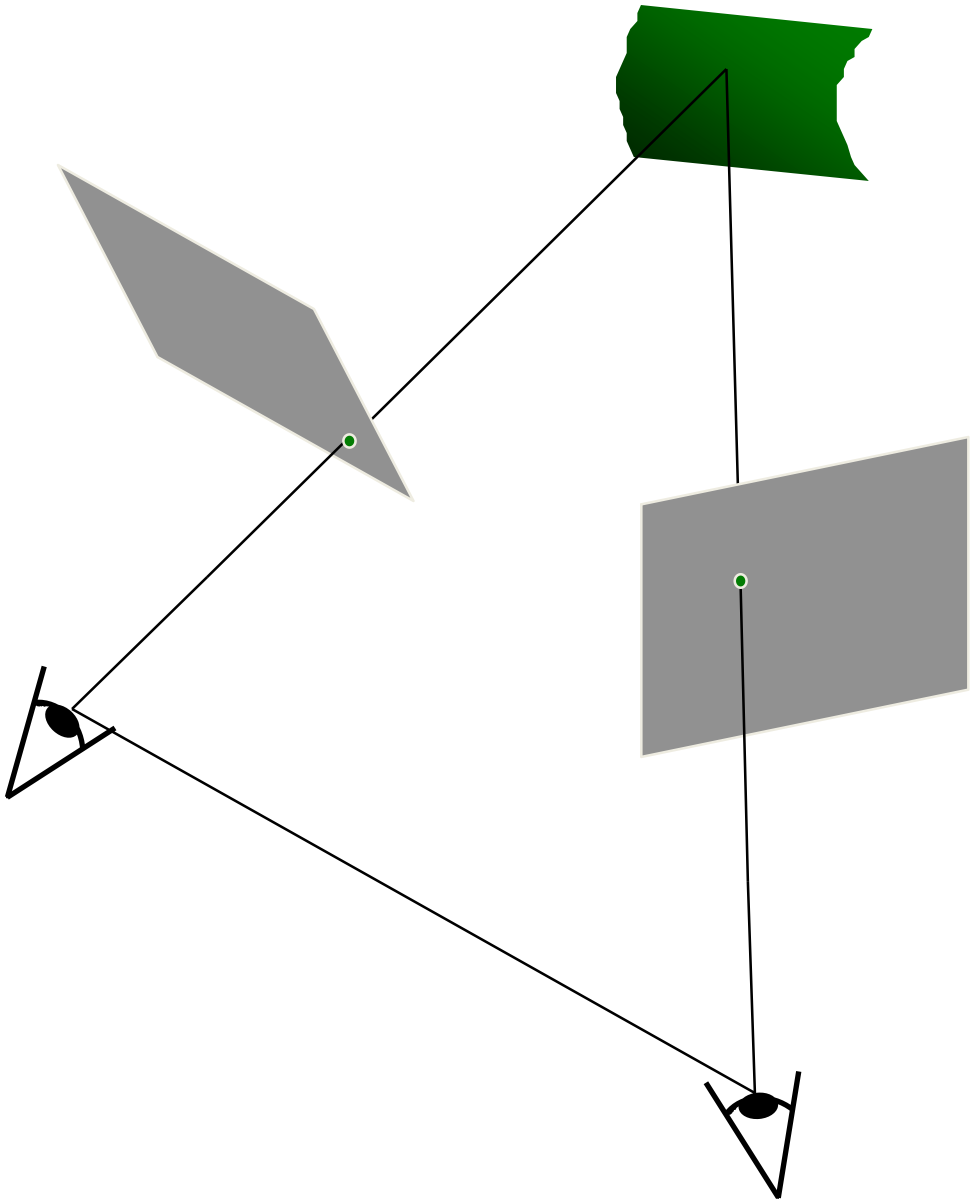
Today's lecture

- Stereo Matching (Sparse correspondence to Dense Correspondence)

- Next lecture: Optical Flow (Dense motion estimation)

Stereo Matching

Stereo image rectification



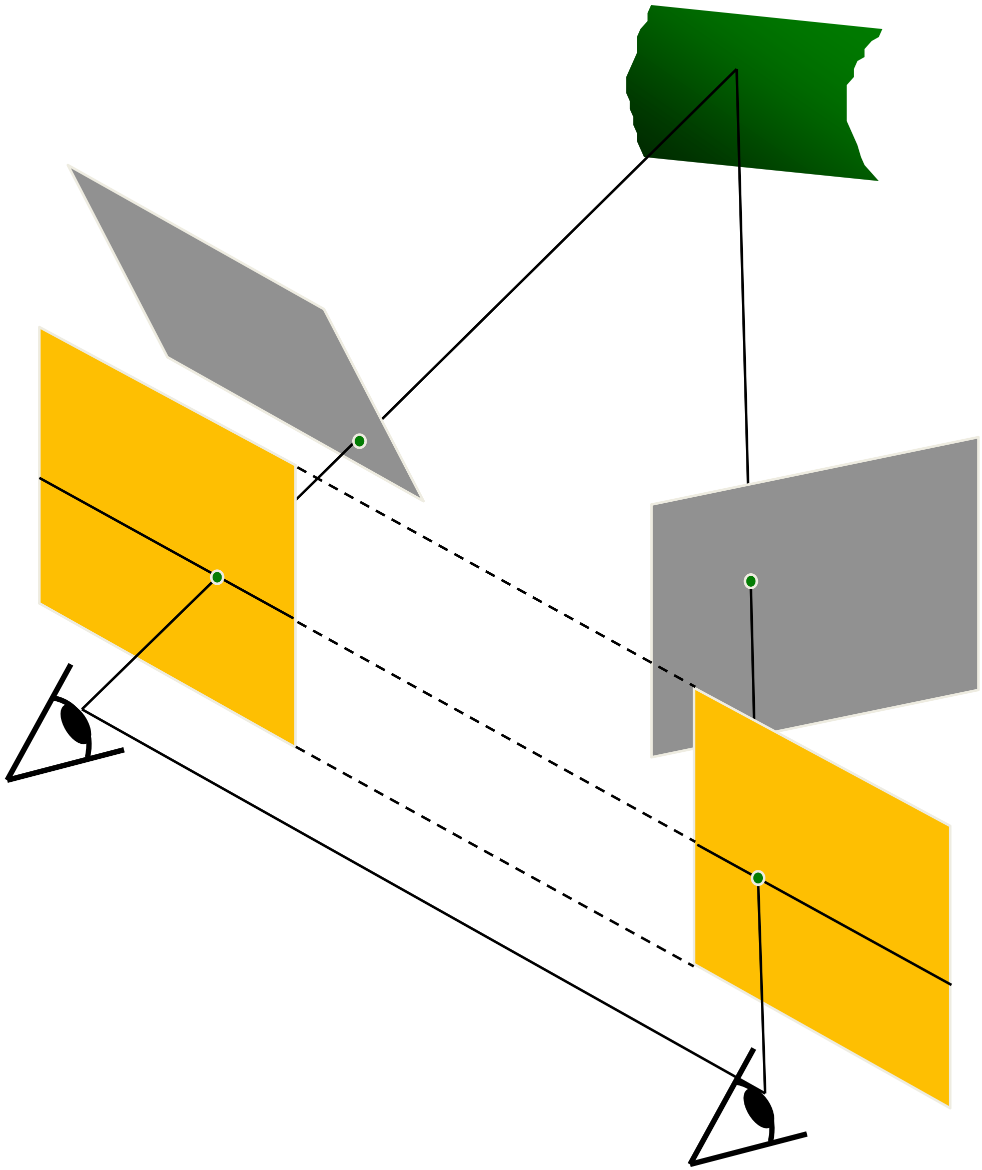
Stereo image rectification

- Reproject image planes onto a common plane parallel to the line between camera centers

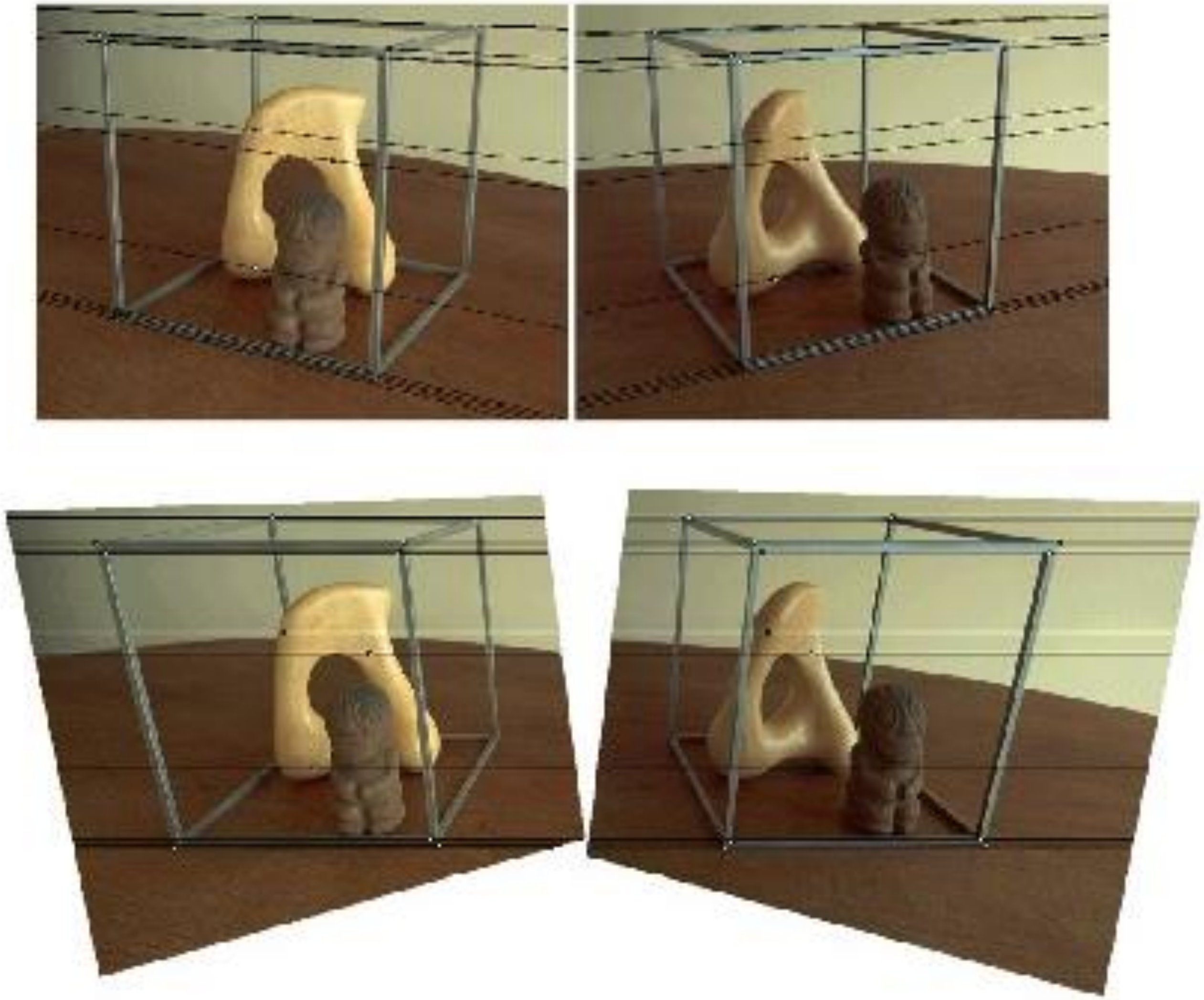
- Pixel motion is horizontal after this transformation

- Two homographies (3x3 transform), one for each input image reprojection

➢ C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision](). IEEE Conf. Computer Vision and Pattern Recognition, 1999.



Rectification example

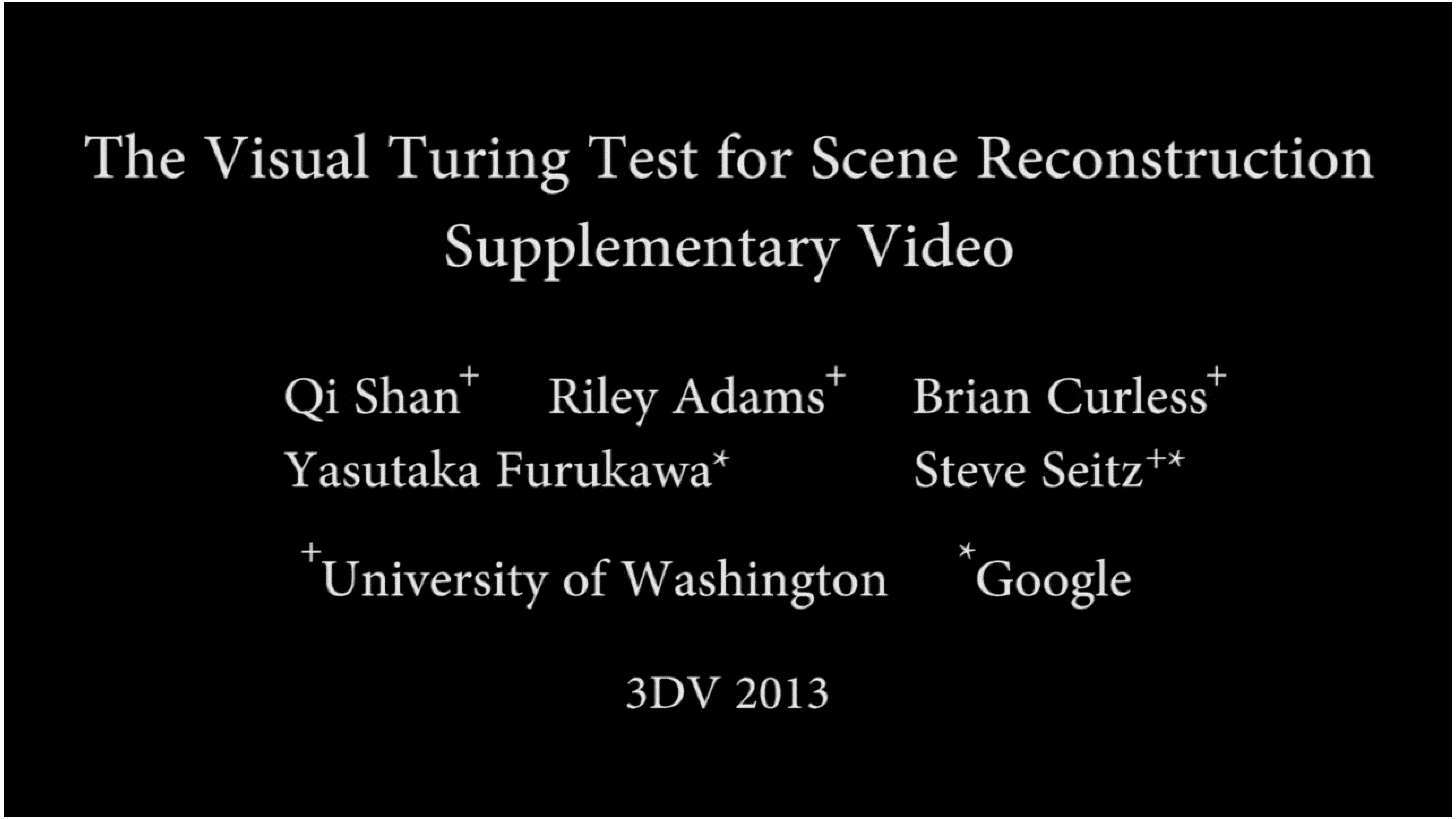


The correspondence problem

- Epipolar geometry constrains our search, but we still have a difficult correspondence problem.

Fundamental Matrix + Sparse correspondence

Fundamental Matrix + Dense correspondence



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]:

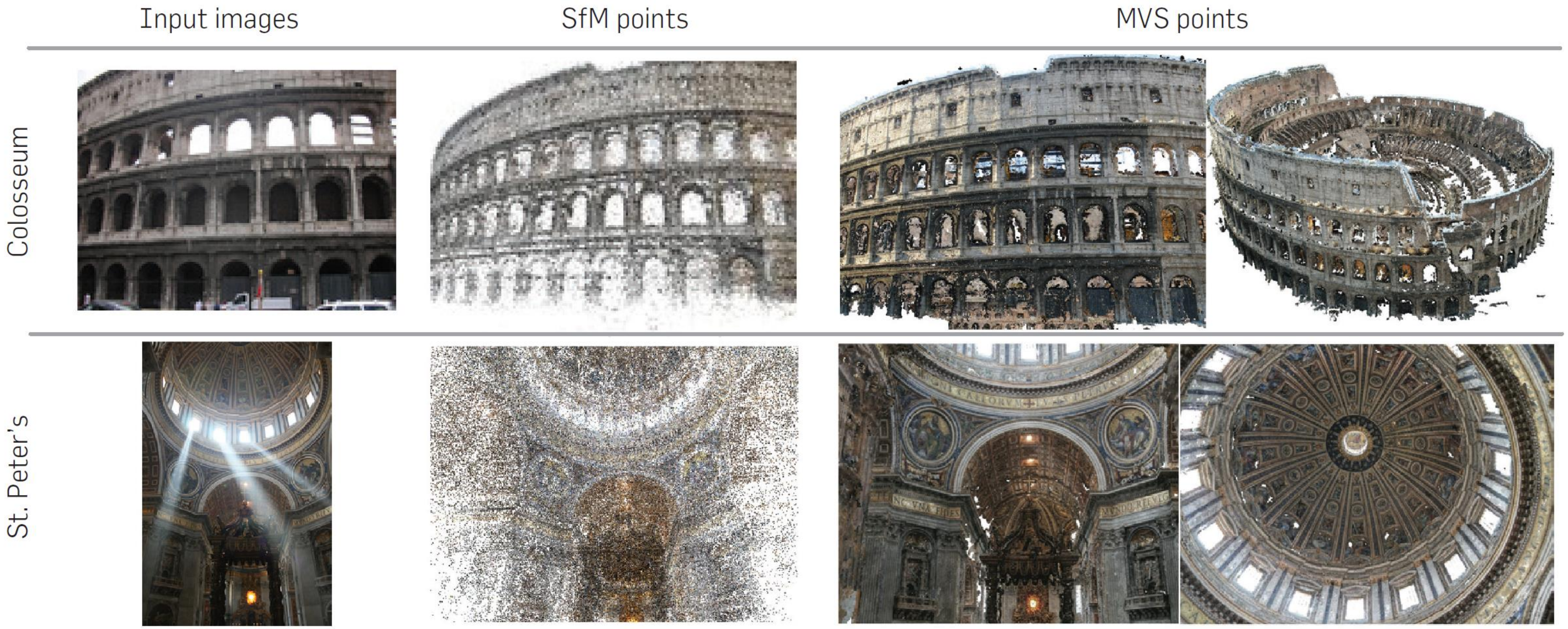
$$x_{ij}^{\top} F x_{ij} = 0. \tag{3}$$

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.

Building Rome in a Day

By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112

Sparse to Dense Correspodence

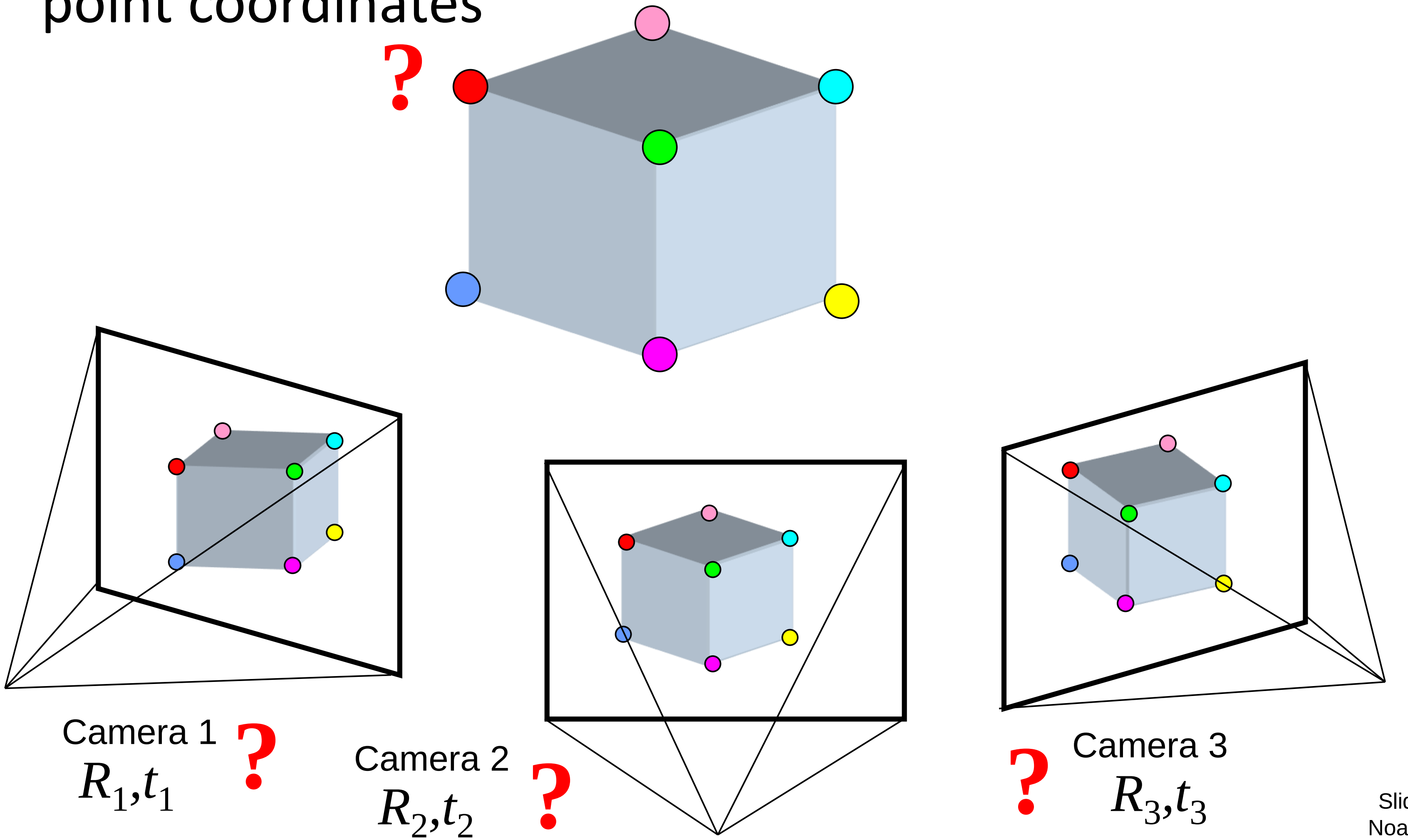


Building Rome in a Day

By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112

Structure from motion (or SLAM)

- Given a set of corresponding points in two or more images, compute the camera parameters and the 3D point coordinates



Slide credit:
Noah Snavely

Structure from motion ambiguity

- If we scale the entire scene by some factor *k* and, at the same time, scale the camera matrices by the factor of 1/*k*, the projections of the scene points in the image remain exactly the same:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\frac{1}{k} \mathbf{P} \right) (k\mathbf{X})$$

It is impossible to recover the absolute scale of the scene!

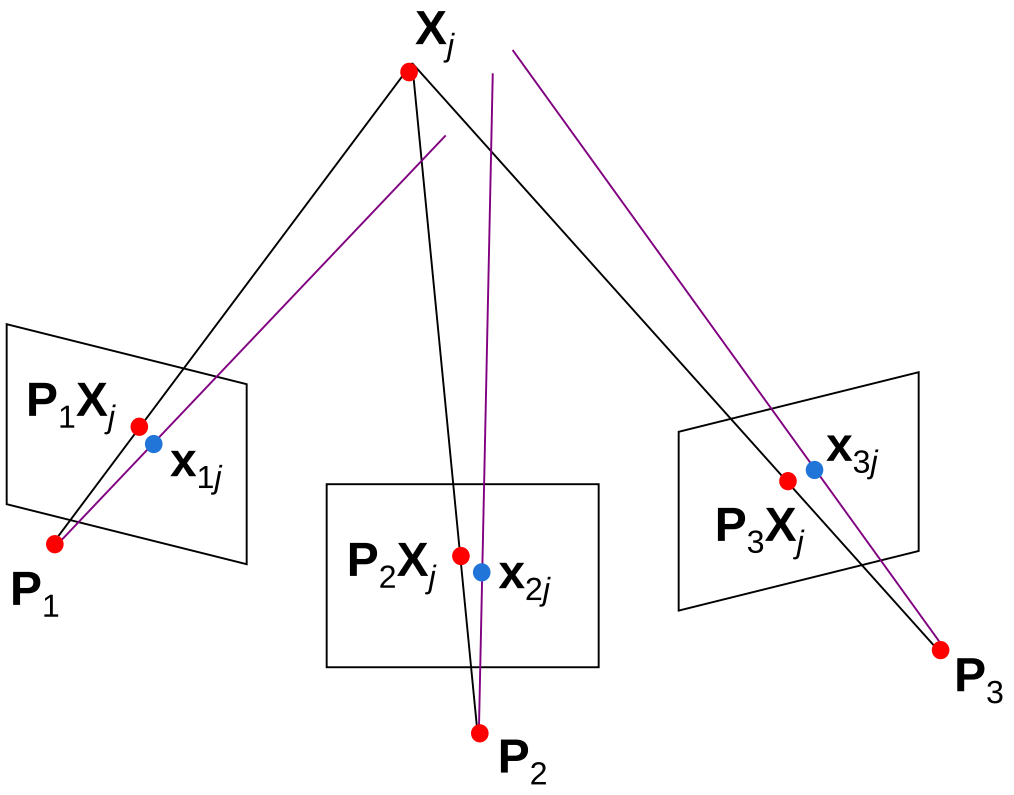
How do we know the scale of image content?


Photography
Your Eyes Deserve It
TOTAL
RECLAIMS
TOTAL
RECLAIMS
TOTAL
RECLAIMS
TOTAL

Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing reprojection error

$$E(\mathbf{P},\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} D\left(\mathbf{x}_{ij}, \mathbf{P}_i\mathbf{X}_j\right)^2$$



Correspondence problem

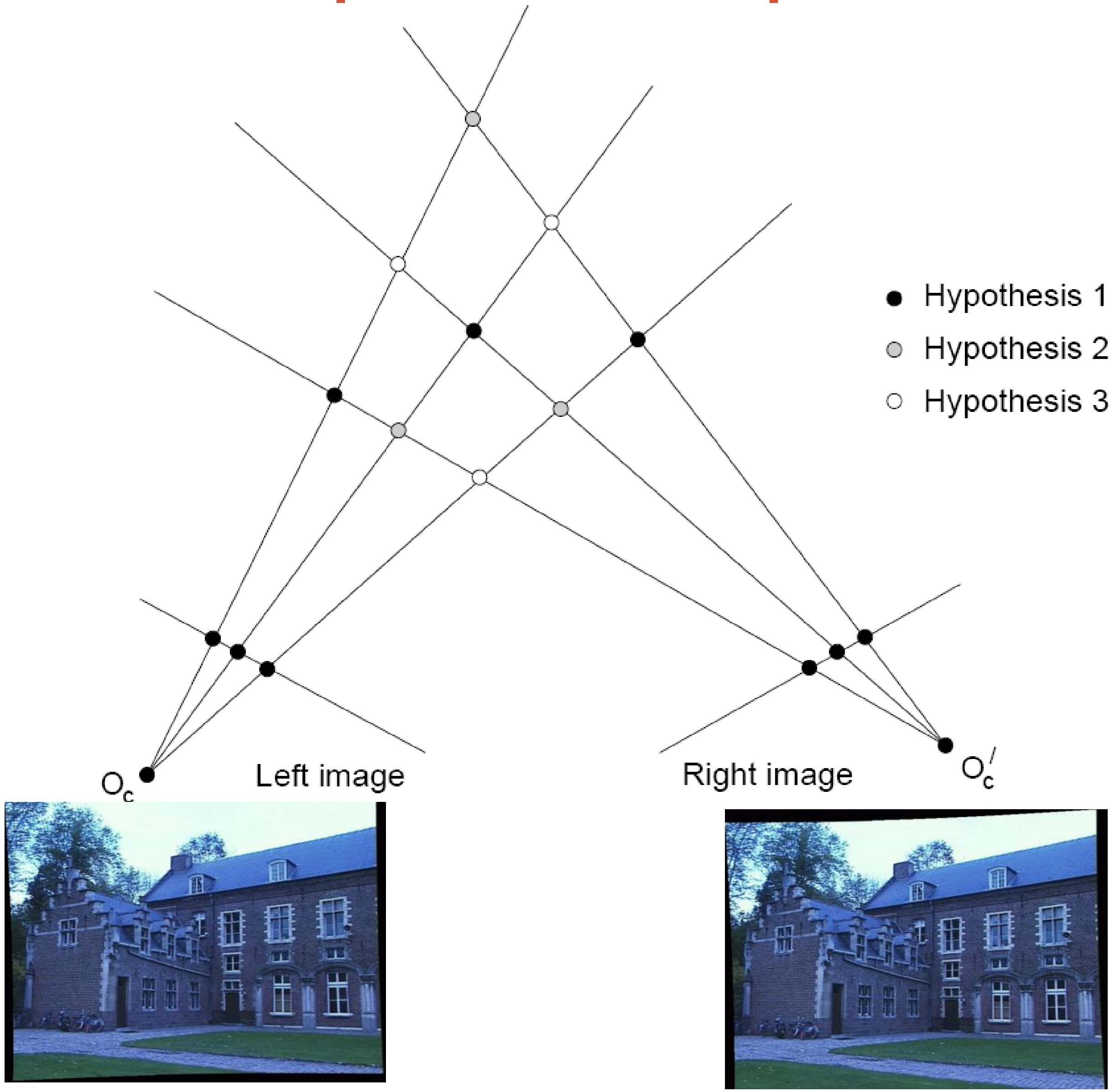


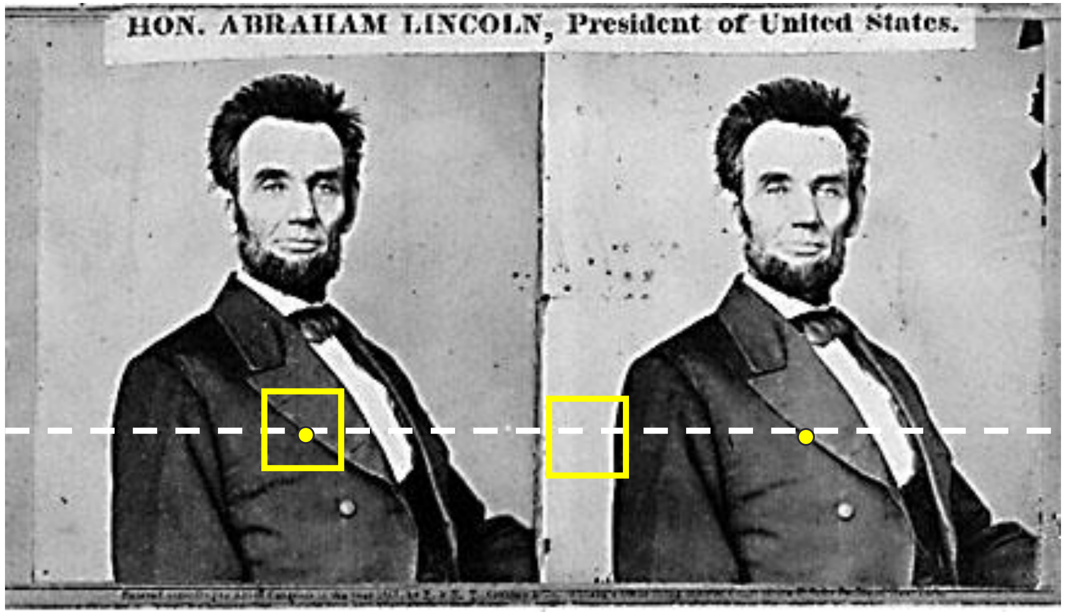
Figure from Gee & Cipolla 1999

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

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are “soft” constraints to help identify corresponding points
  - Similarity
  - Uniqueness
  - Ordering
  - Disparity gradient
- 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

Dense correspondence search



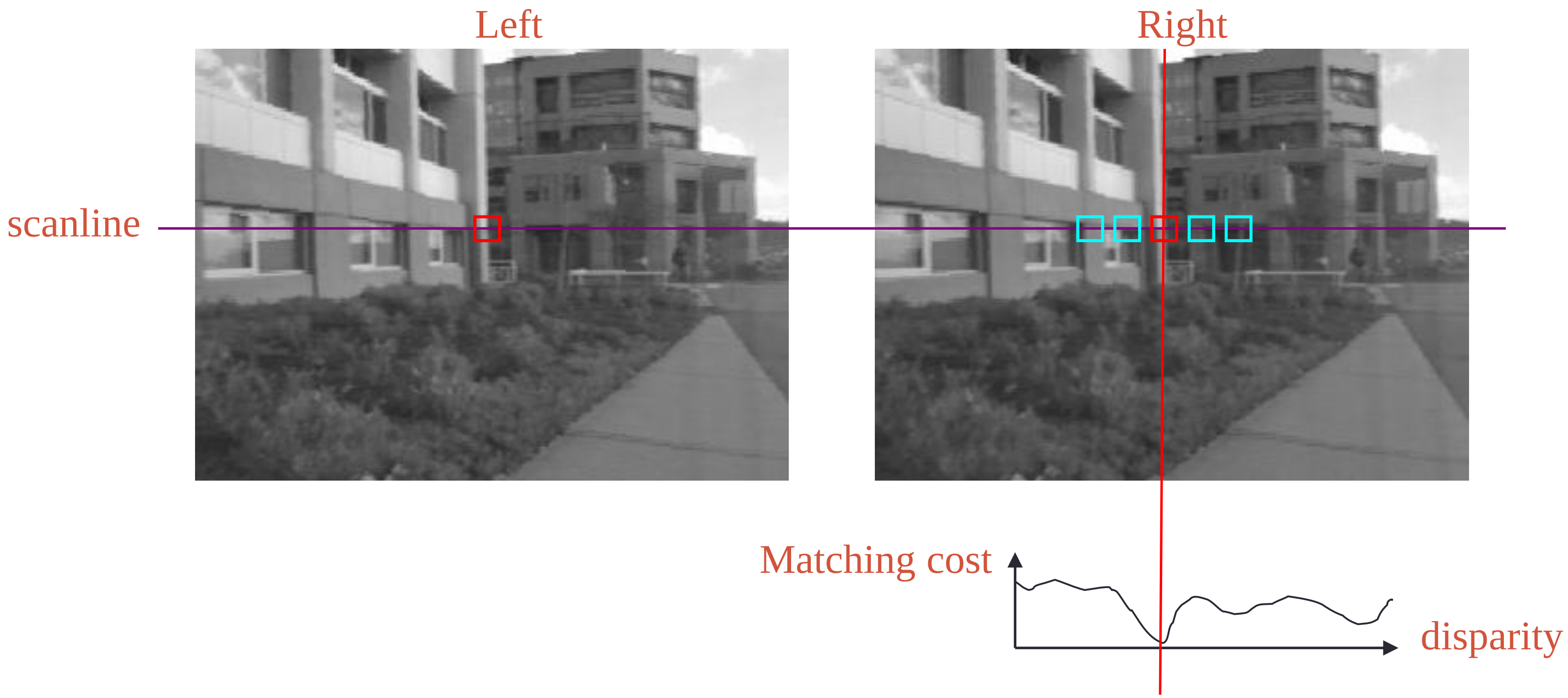
For each epipolar line

For each pixel / window in the left image

- compare with every pixel / window on same epipolar line in right image
- pick position with minimum match cost (e.g., SSD, normalized correlation)

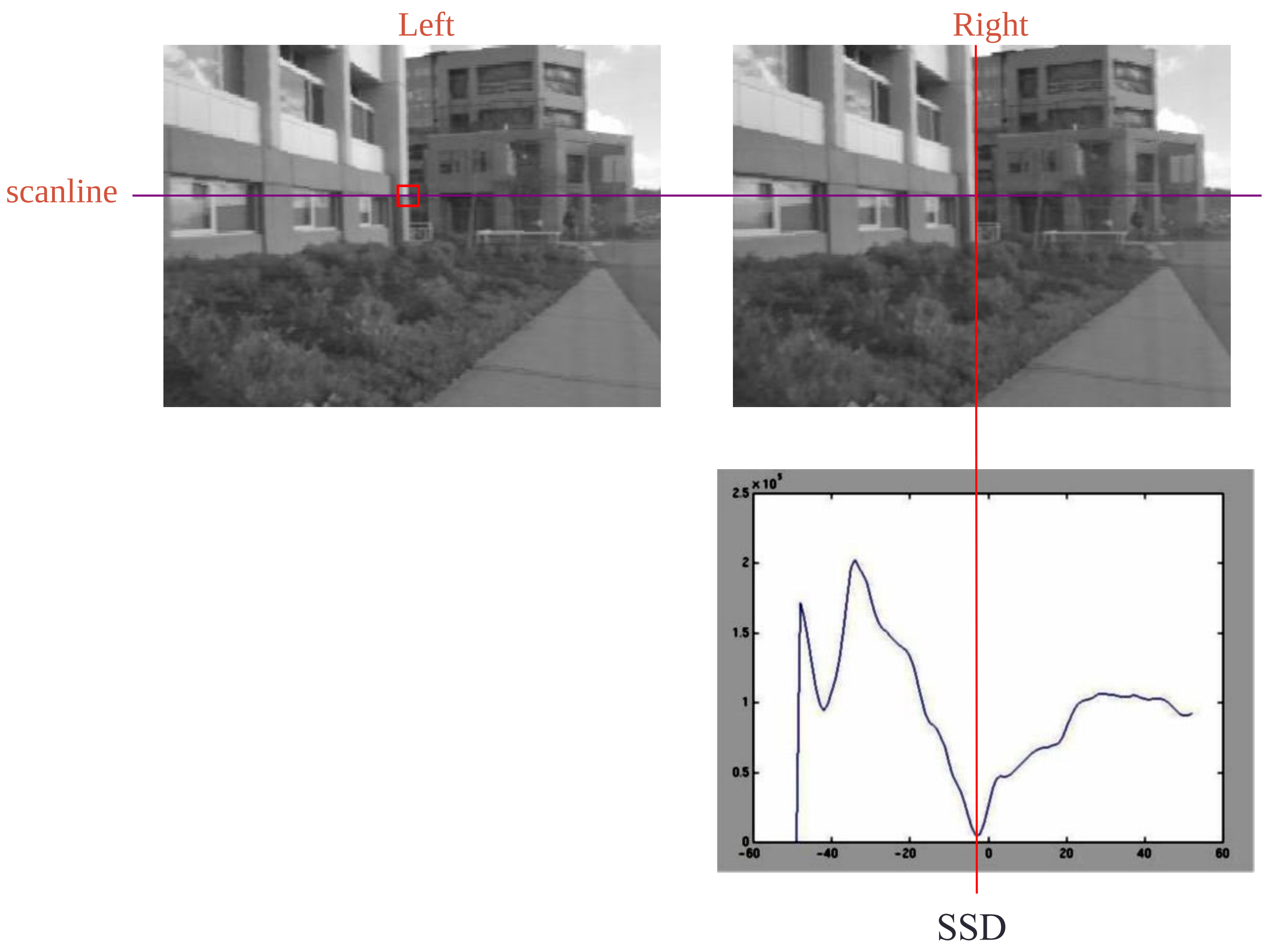
Adapted from Li Zhang

Correspondence search with similarity constraint

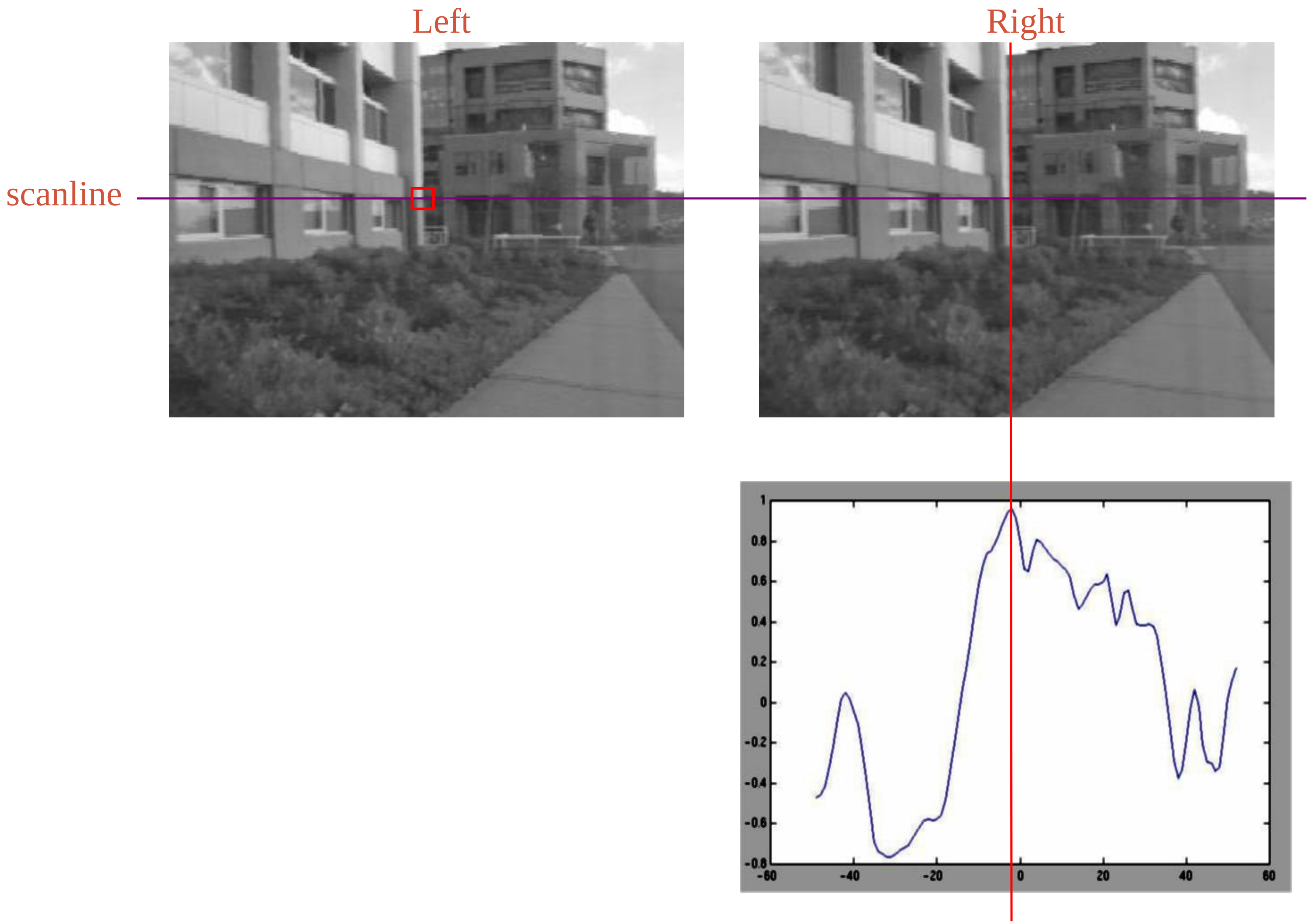


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search with similarity constraint



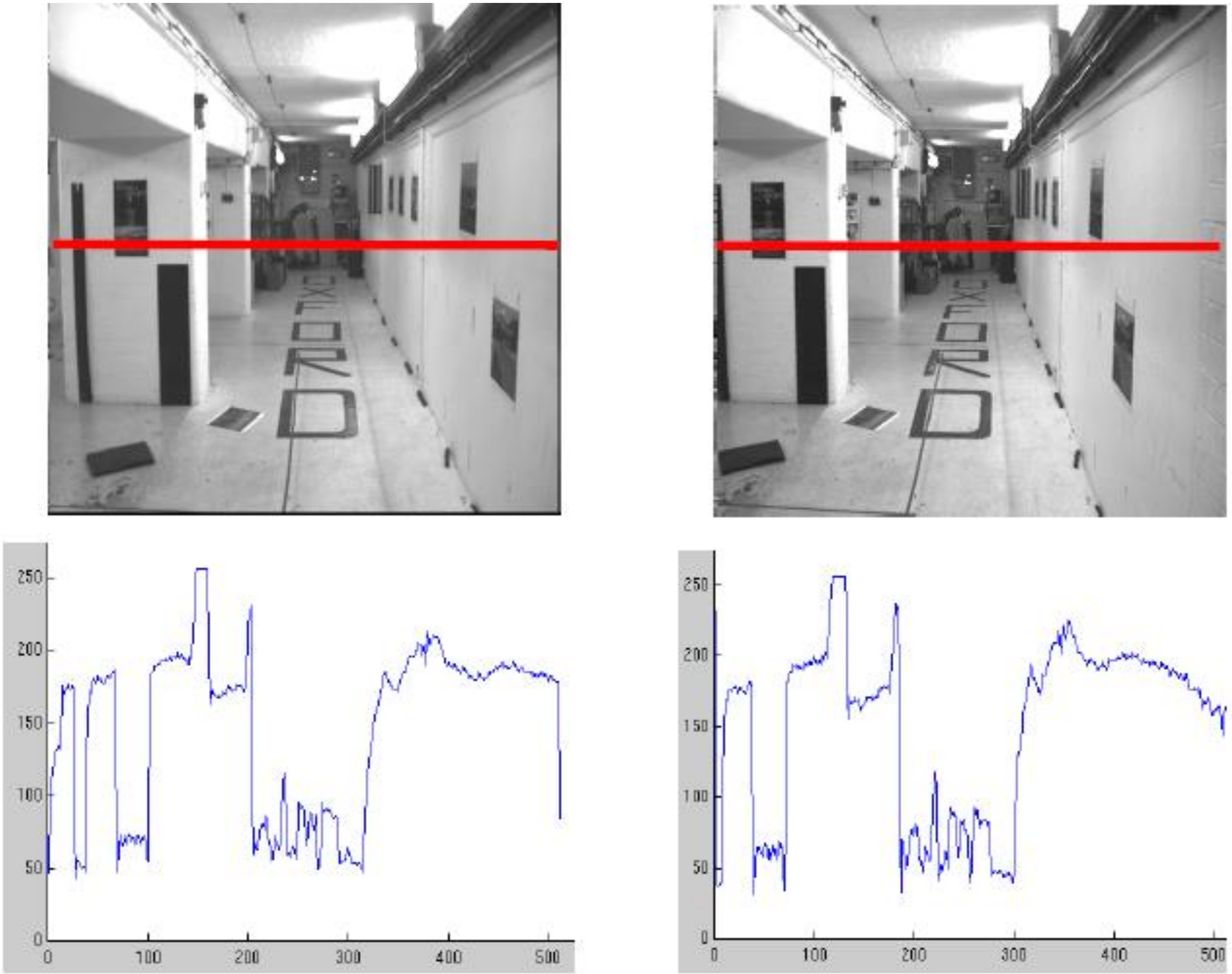
Correspondence search with similarity constraint



Norm. corr

Correspondence problem

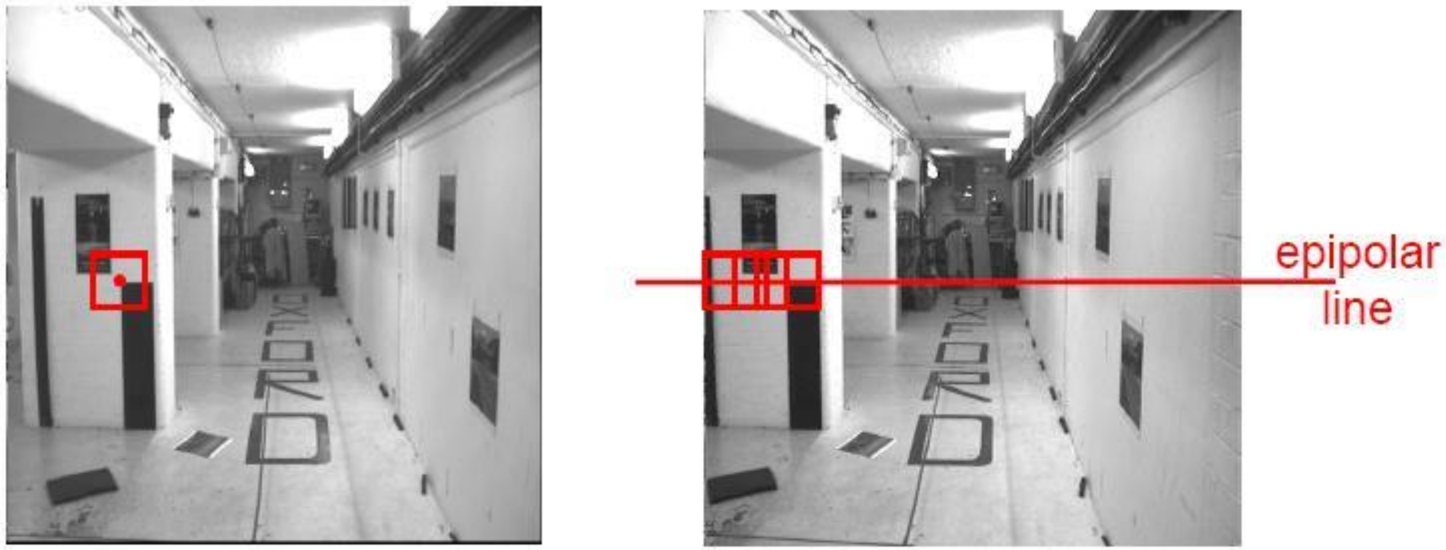
Intensity profiles



- Clear correspondence between intensities, but also noise and ambiguity

Source: Andrew Zisserman

Correspondence problem



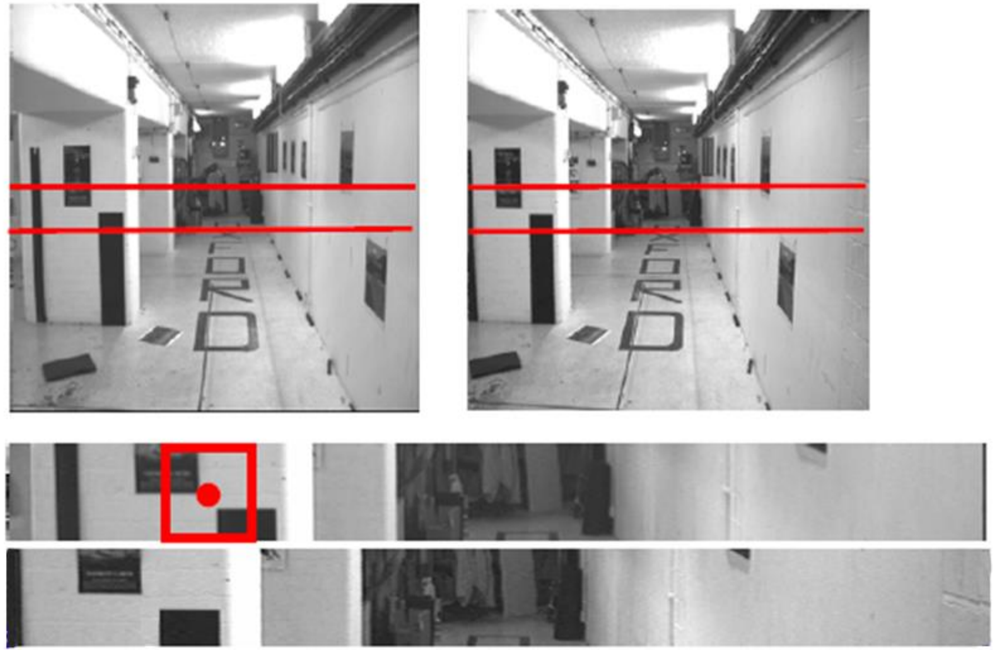
Neighborhoods of corresponding points are similar in intensity patterns.

Source: Andrew Zisserman

Correlation-based window matching

left image band (x)

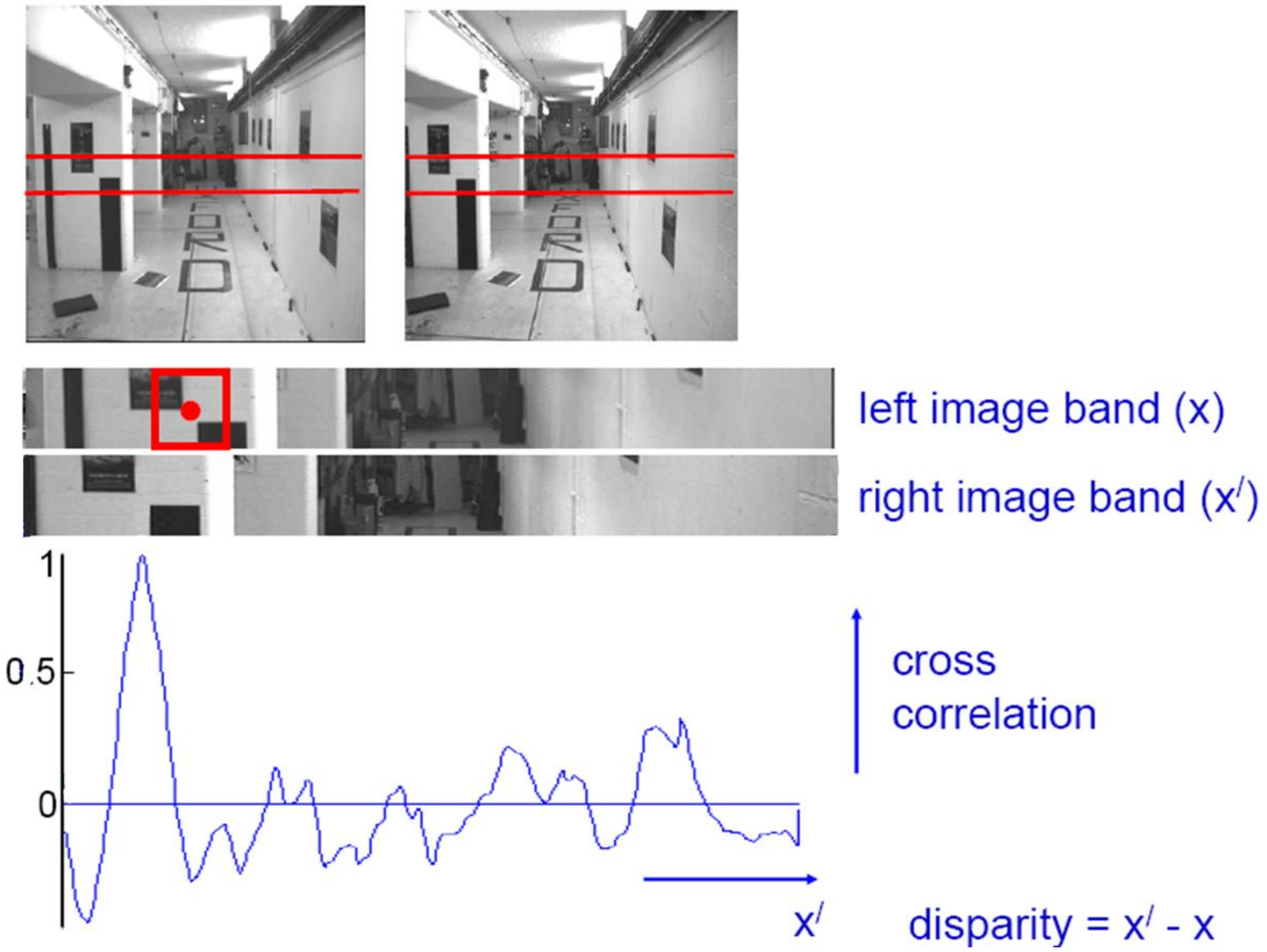
Correlation-based window matching



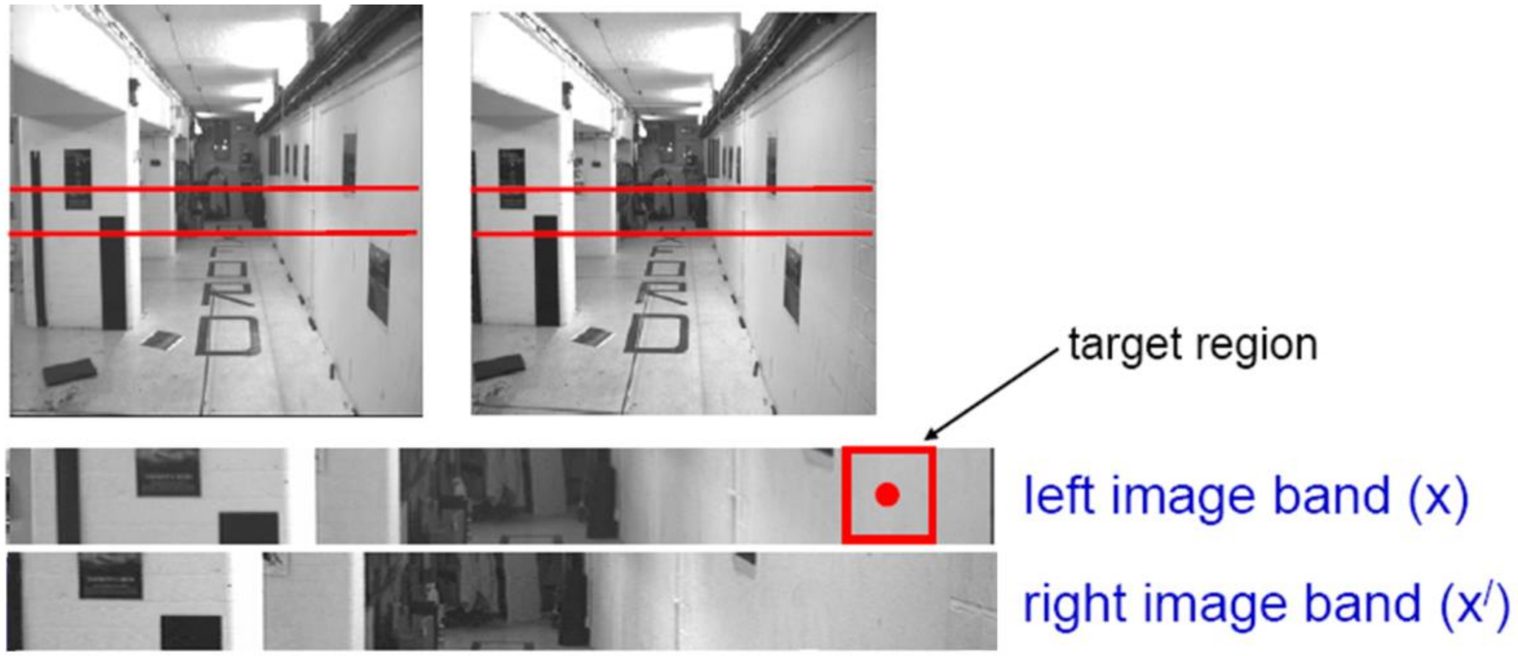
left image band (x)

right image band (x')

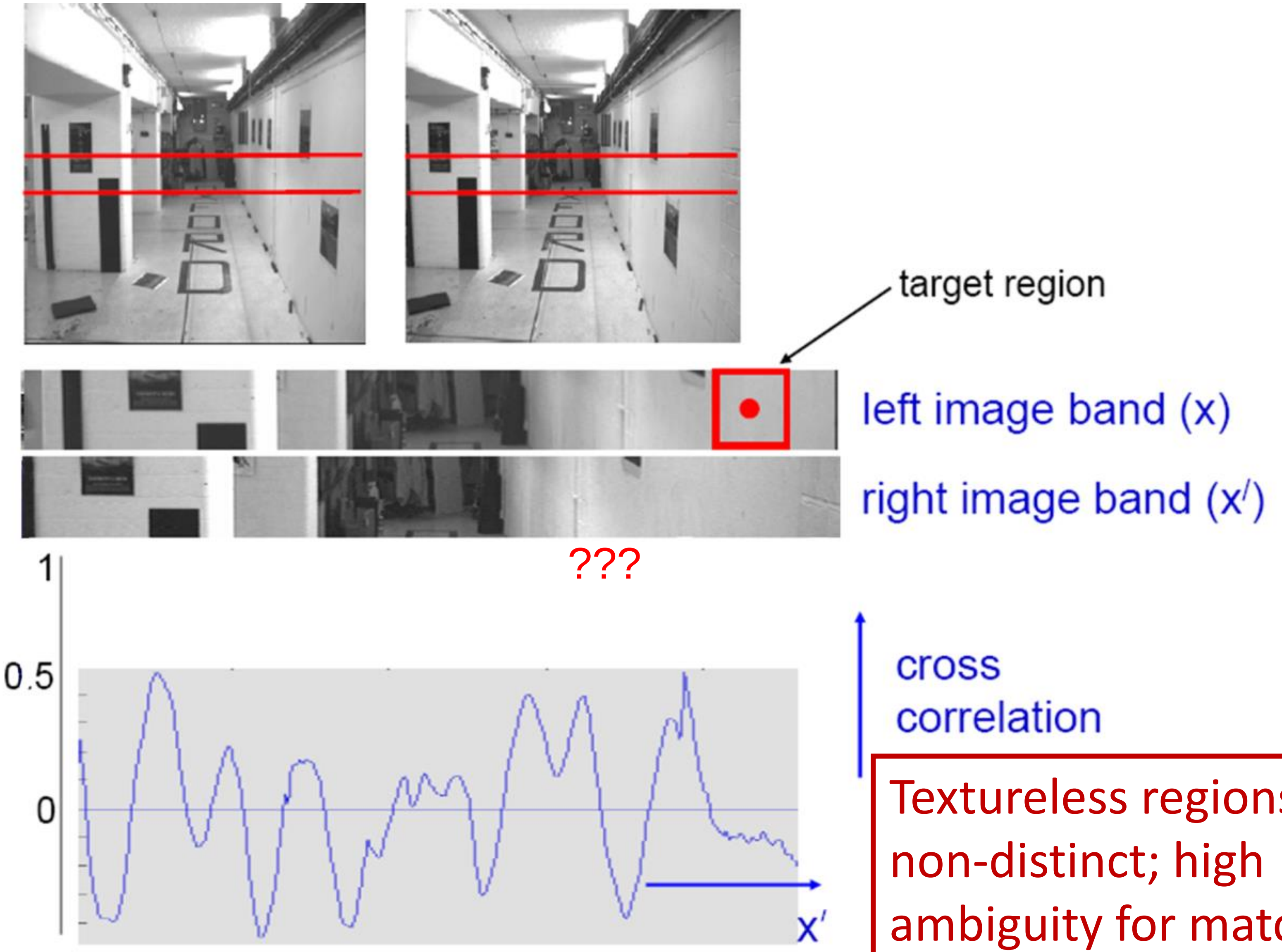
Correlation-based window matching



Correlation-based window matching

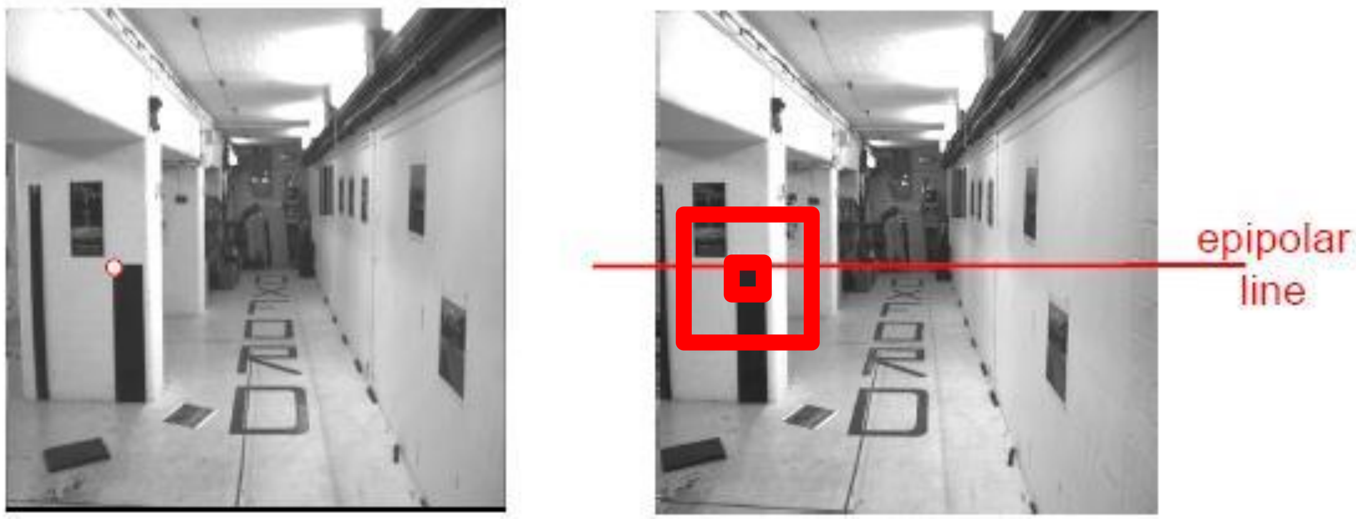


Correlation-based window matching



Textureless regions are non-distinct; high ambiguity for matches.

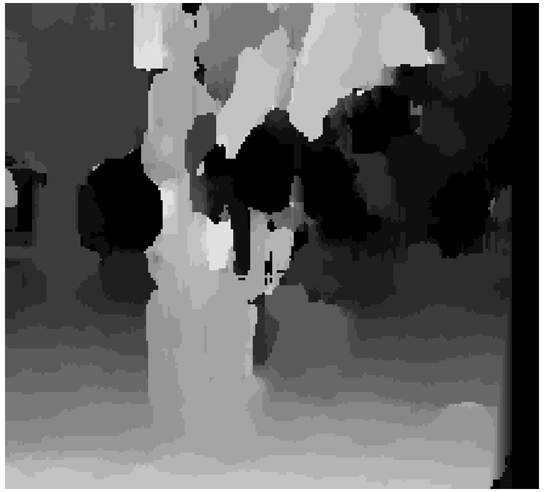
Effect of window size



Source: Andrew Zisserman

Effect of window size

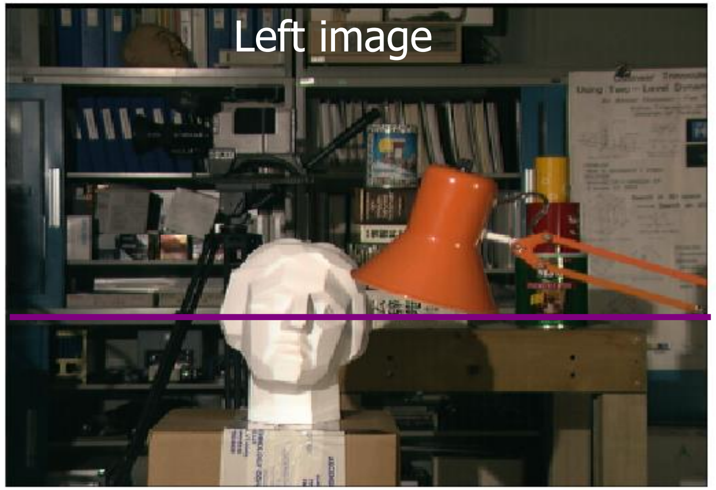
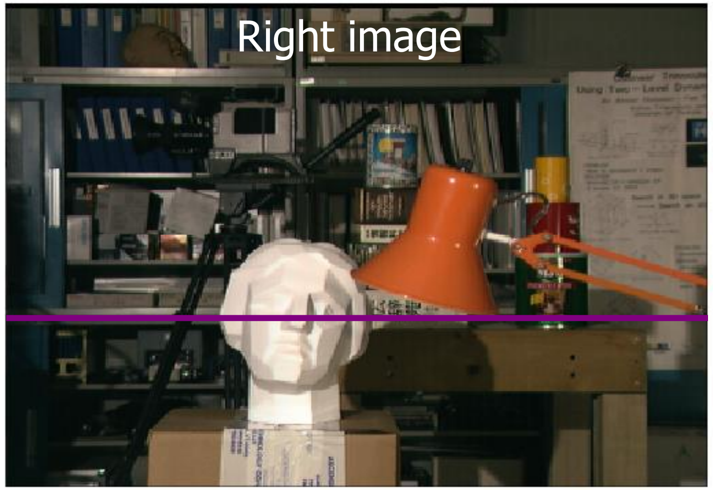
W = 3



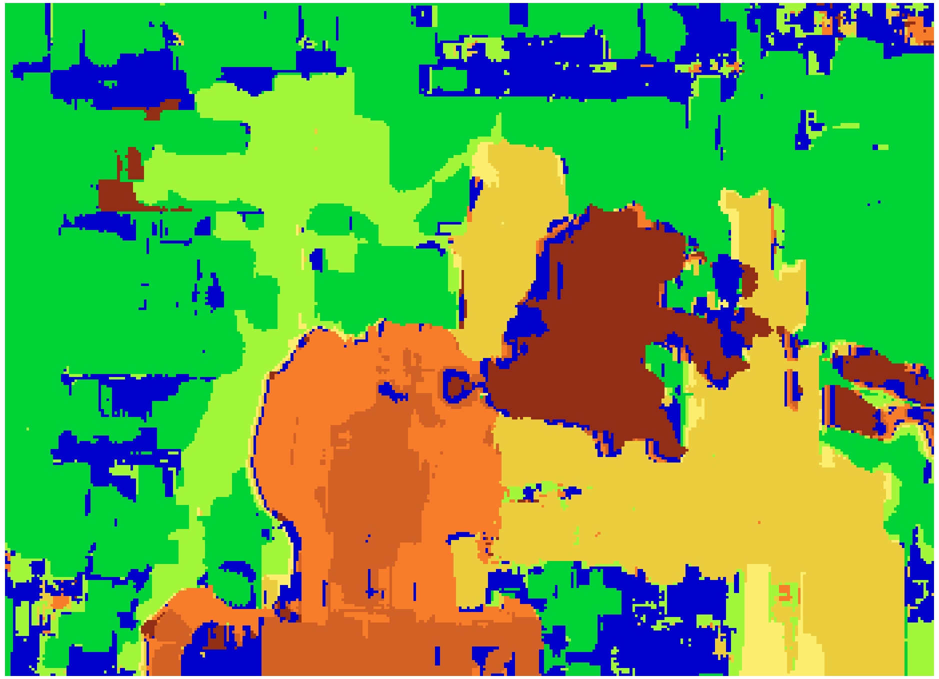
W = 20

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

Figures from Li Zhang


Left image

Right image

Results with window search



Window-based matching
(best window size)

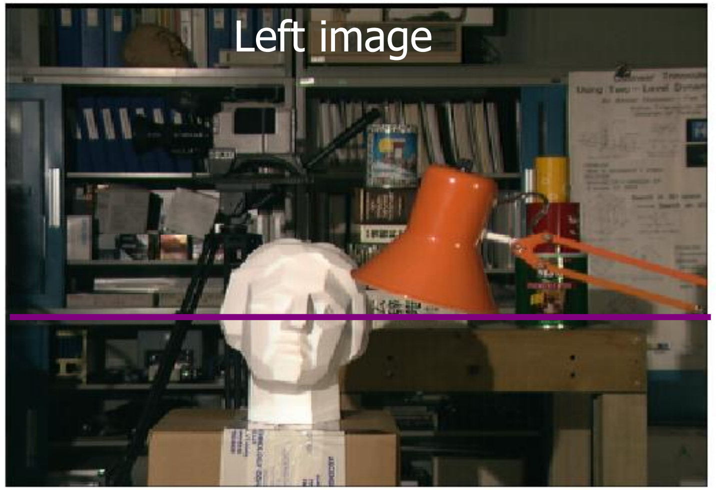
Ground truth

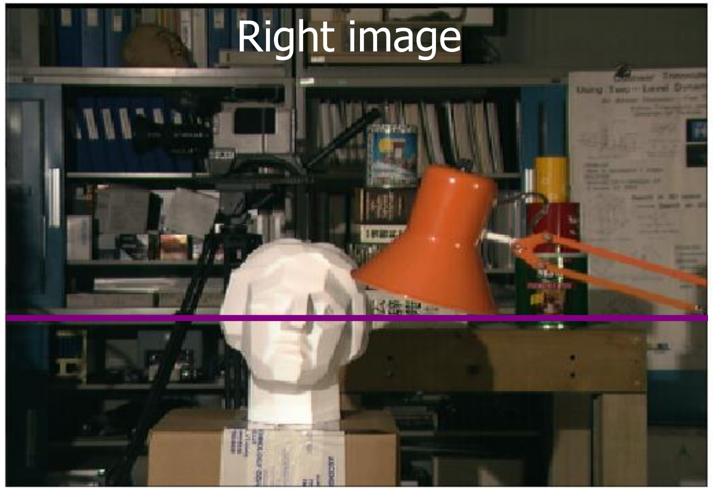
Better solutions

- Beyond individual correspondences to estimate disparities:
- Optimize correspondence assignments jointly
  - Scanline at a time (e.g. dynamic programming)
  - Full 2D grid (e.g. graph cuts)
  - Approximate 2D solution (e.g. semi-global matching)

Scanline stereo

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



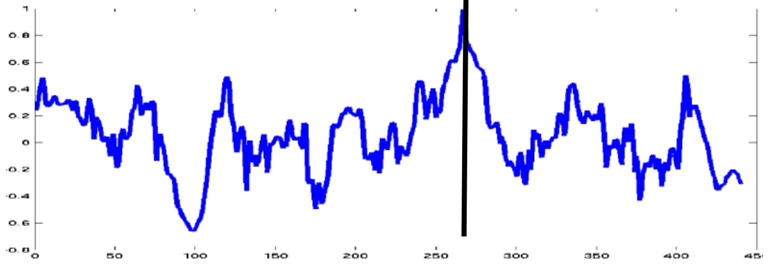


Robert Collins
CSE486, Penn State

Matching using Epipolar Lines

Left Image

Right Image



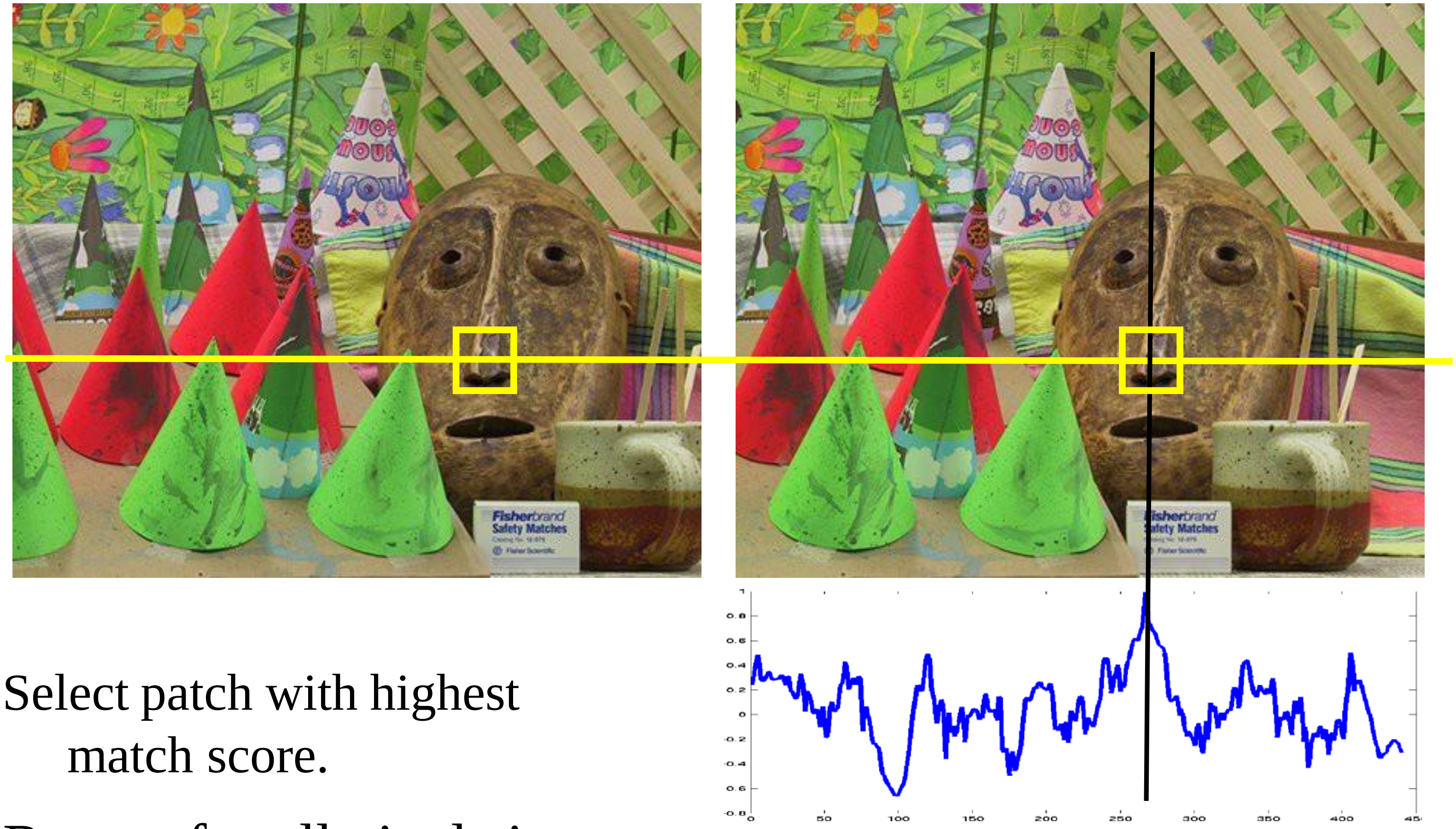
Match Score Values

For a patch in left image

Compare with patches along same row in right image

Matching using Epipolar Lines

Left Image Right Image



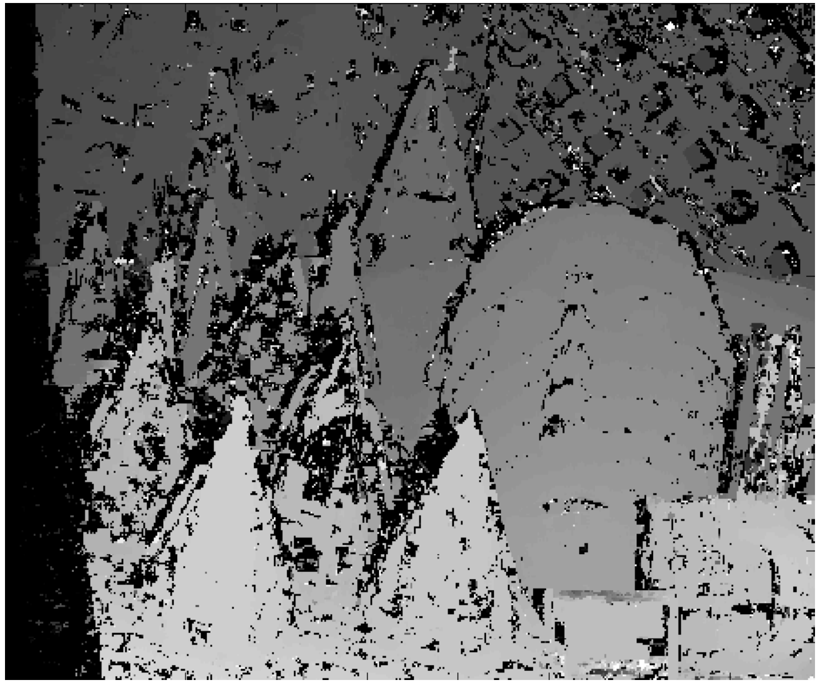
Select patch with highest match score.

Repeat for all pixels in left image.

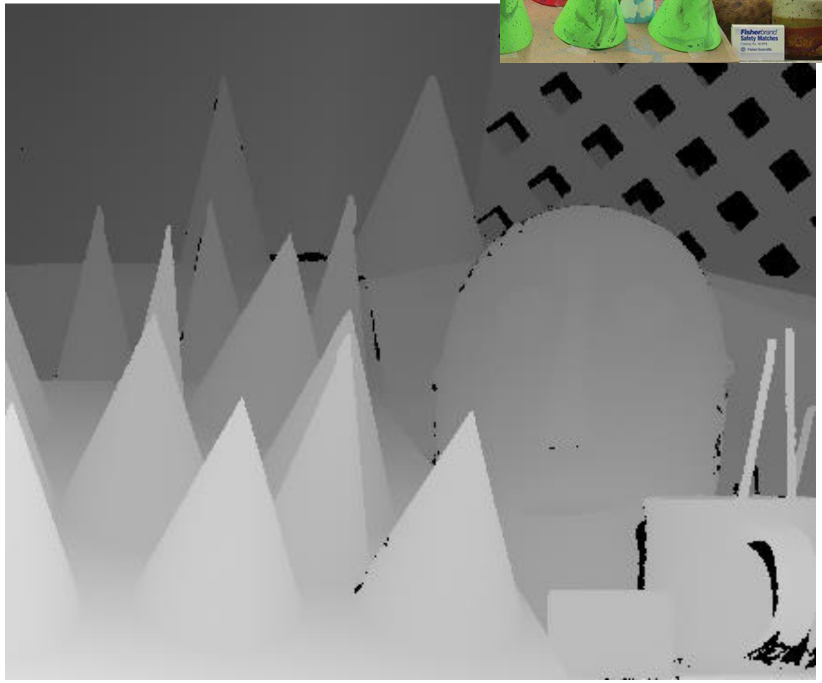
Match Score Values

Robert Collins
CSE486, Penn State

Example: 5x5 windows NCC match score



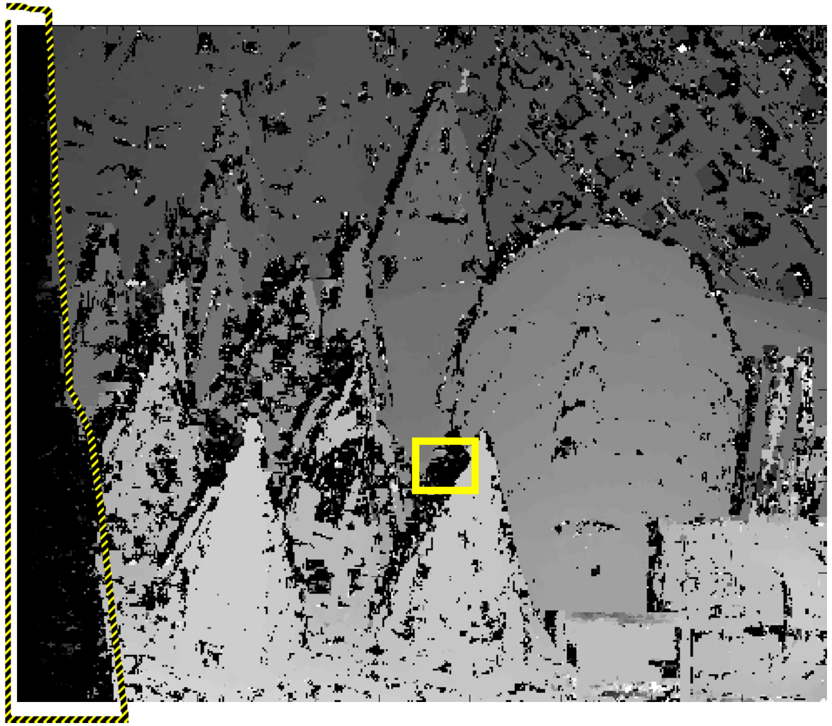
Computed disparities

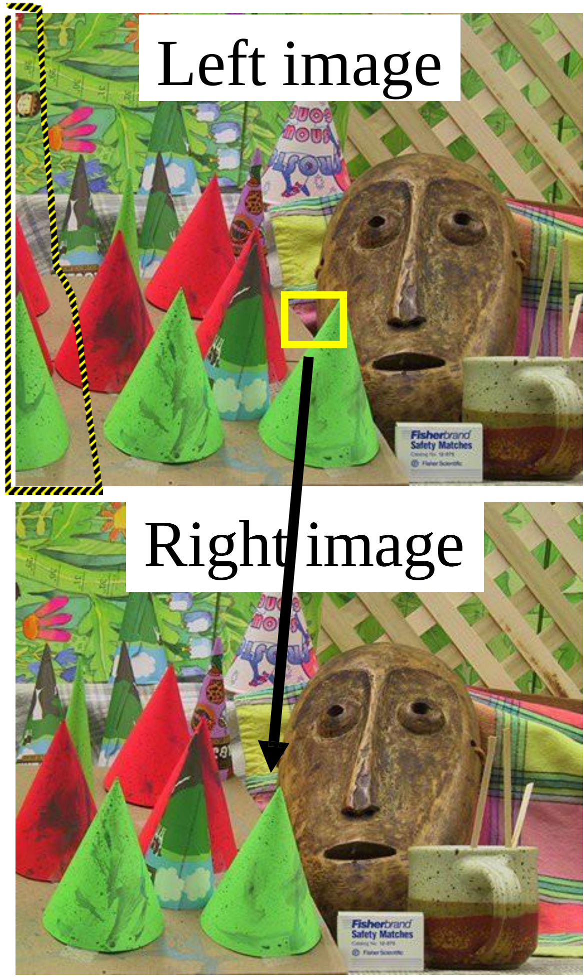


Ground truth

Black pixels: bad disparity values, or no matching patch in right image

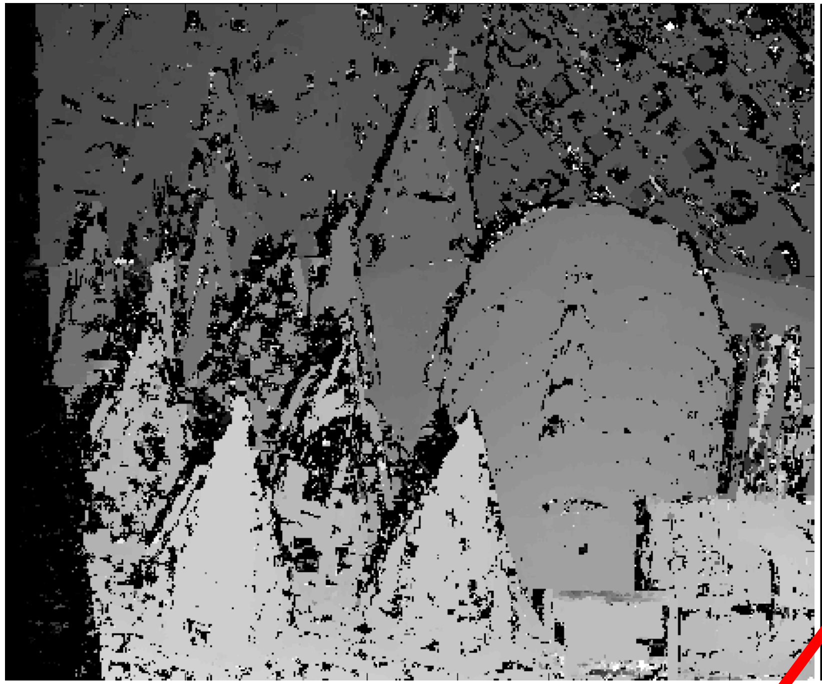
Occlusions: No matches



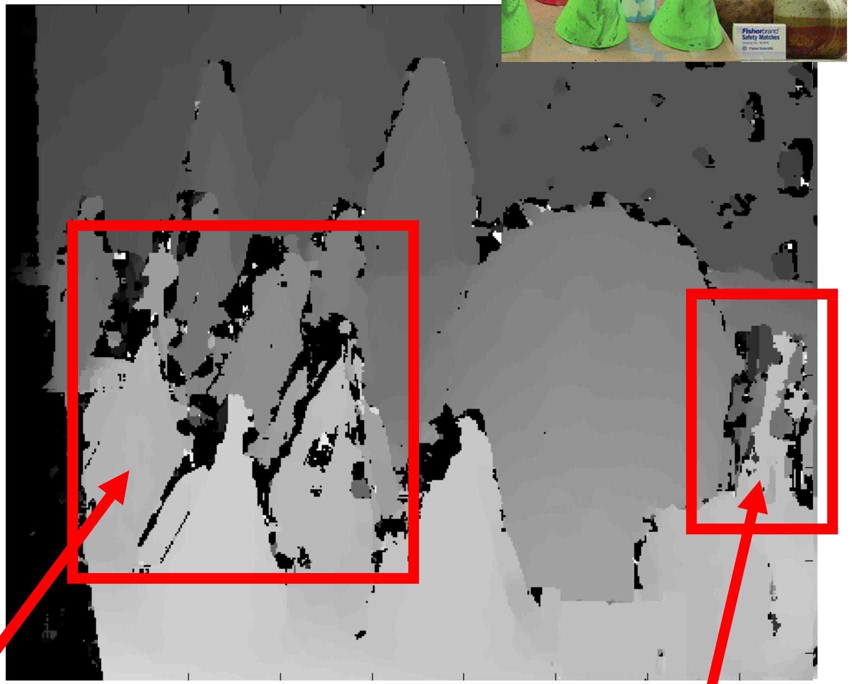


Robert Collins
CSE486, Penn State

Effects of Patch Size



5x5 patches



11x11patches

Smoother in some areas

Loss of finer details

Adding Intra-Scanline Consistency

So far, each left image patch has been matched independently along the right epipolar line.

This can lead to errors.

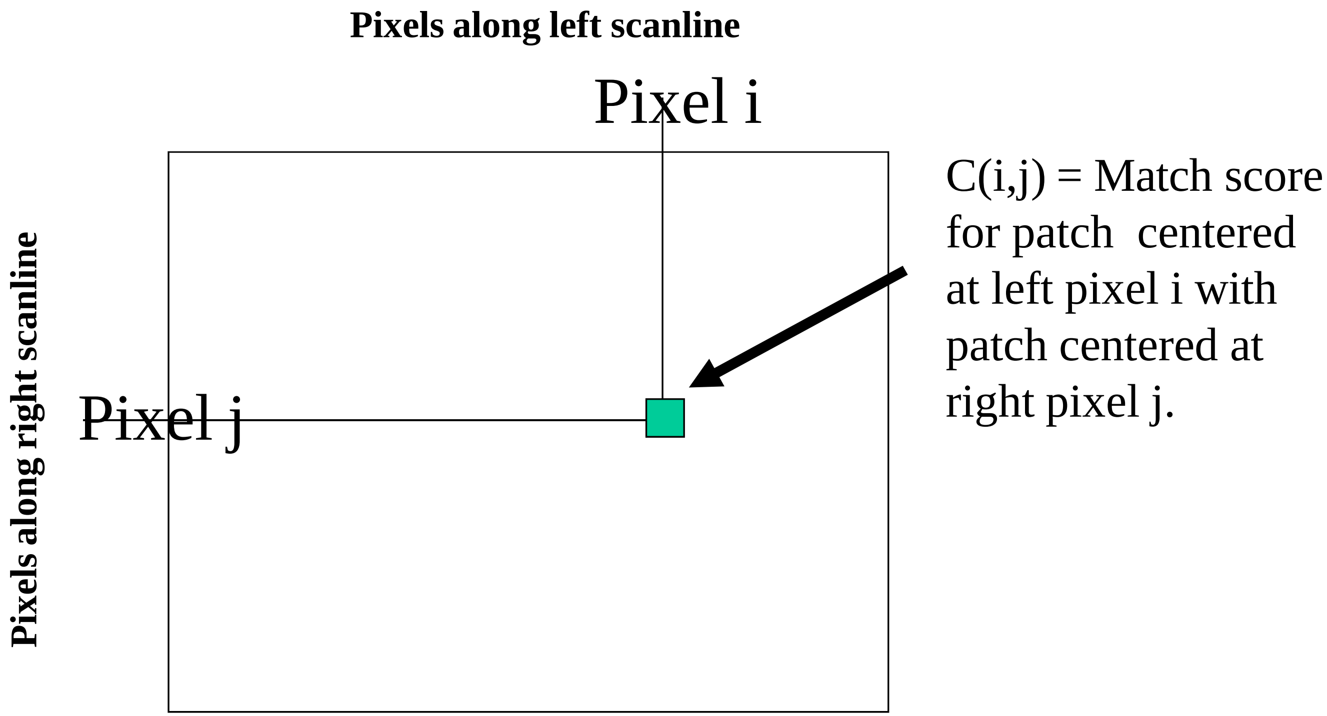
We would like to enforce some consistency among matches in the same row (scanline).

Robert Collins
CSE486, Penn State

Disparity Space Image

First we introduce the concept of DSI.

The DSI for one row represents pairwise match scores between patches along that row in the left and right image.

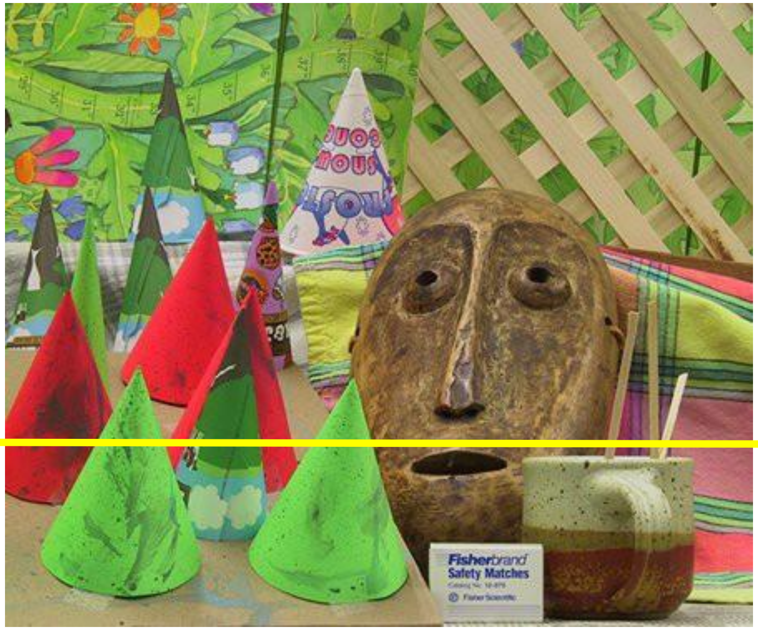


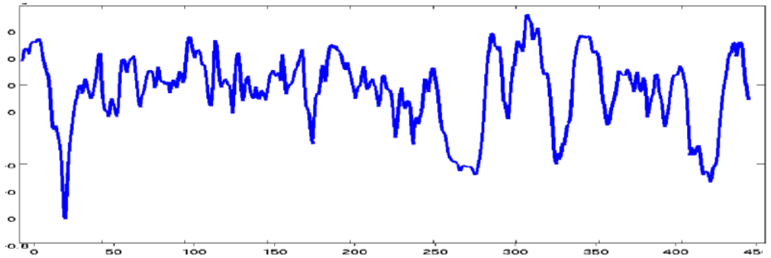
Robert Collins
CSE486, Penn State

Disparity Space Image (DSI)

Left Image

Right Image





Dissimilarity Values

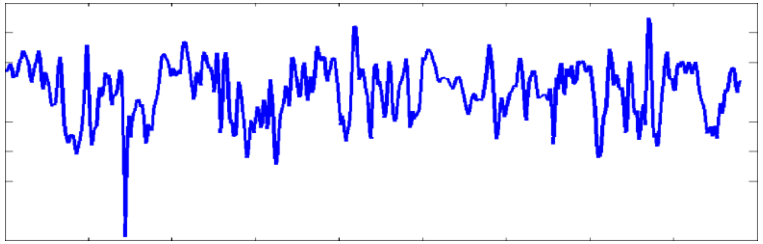
(1-NCC) or SSD

Robert Collins
CSE486, Penn State

Disparity Space Image (DSI)

Left Image

Right Image



Dissimilarity Values

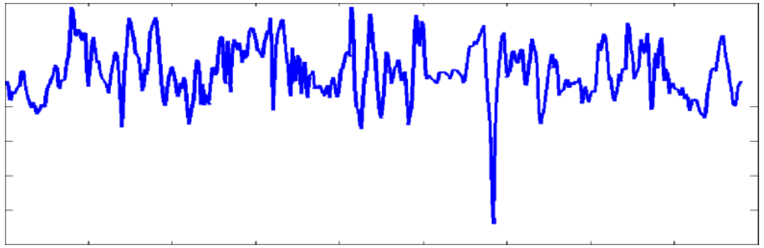
(1-NCC) or SSD

Robert Collins
CSE486, Penn State

Disparity Space Image (DSI)

Left Image

Right Image

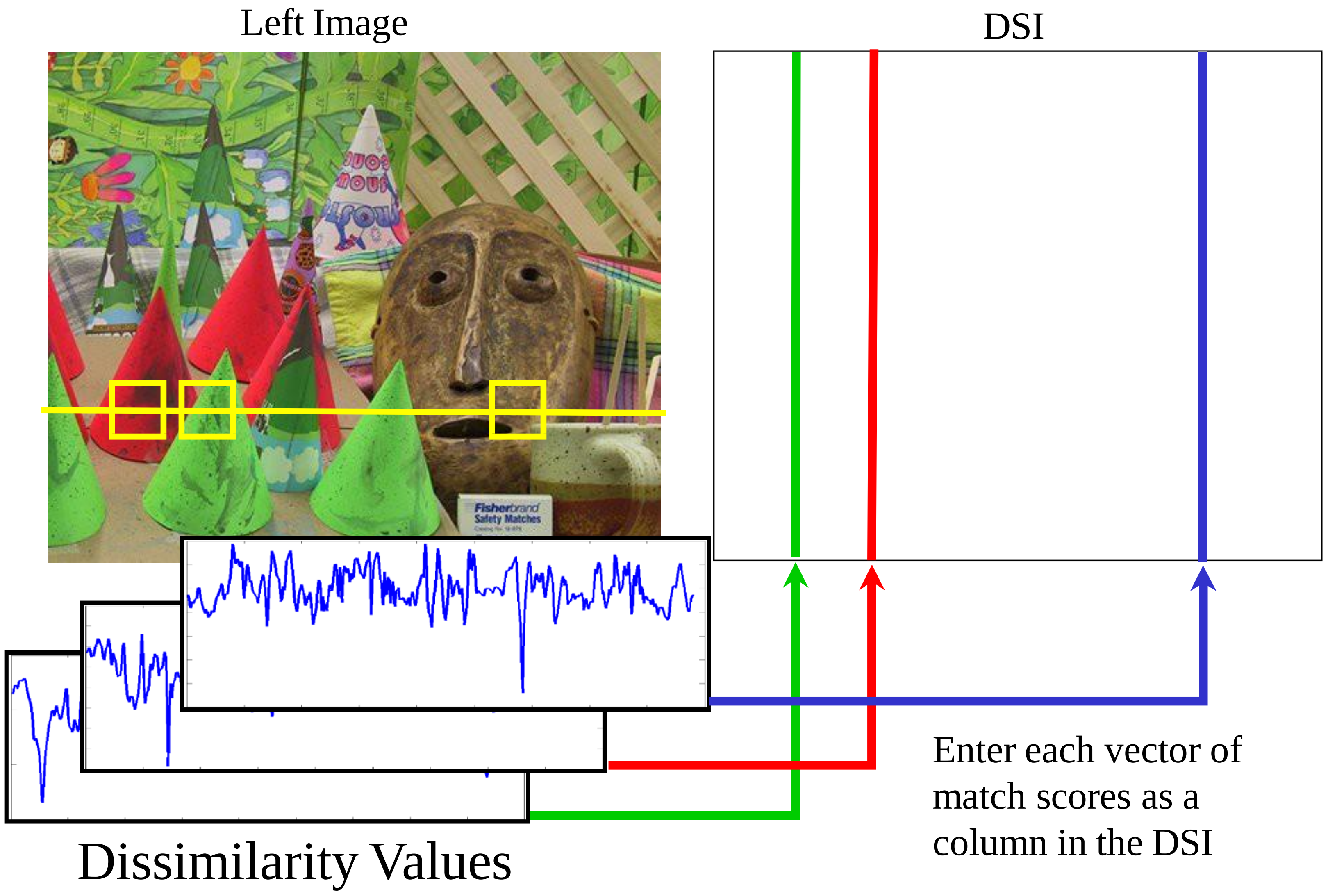


Dissimilarity Values

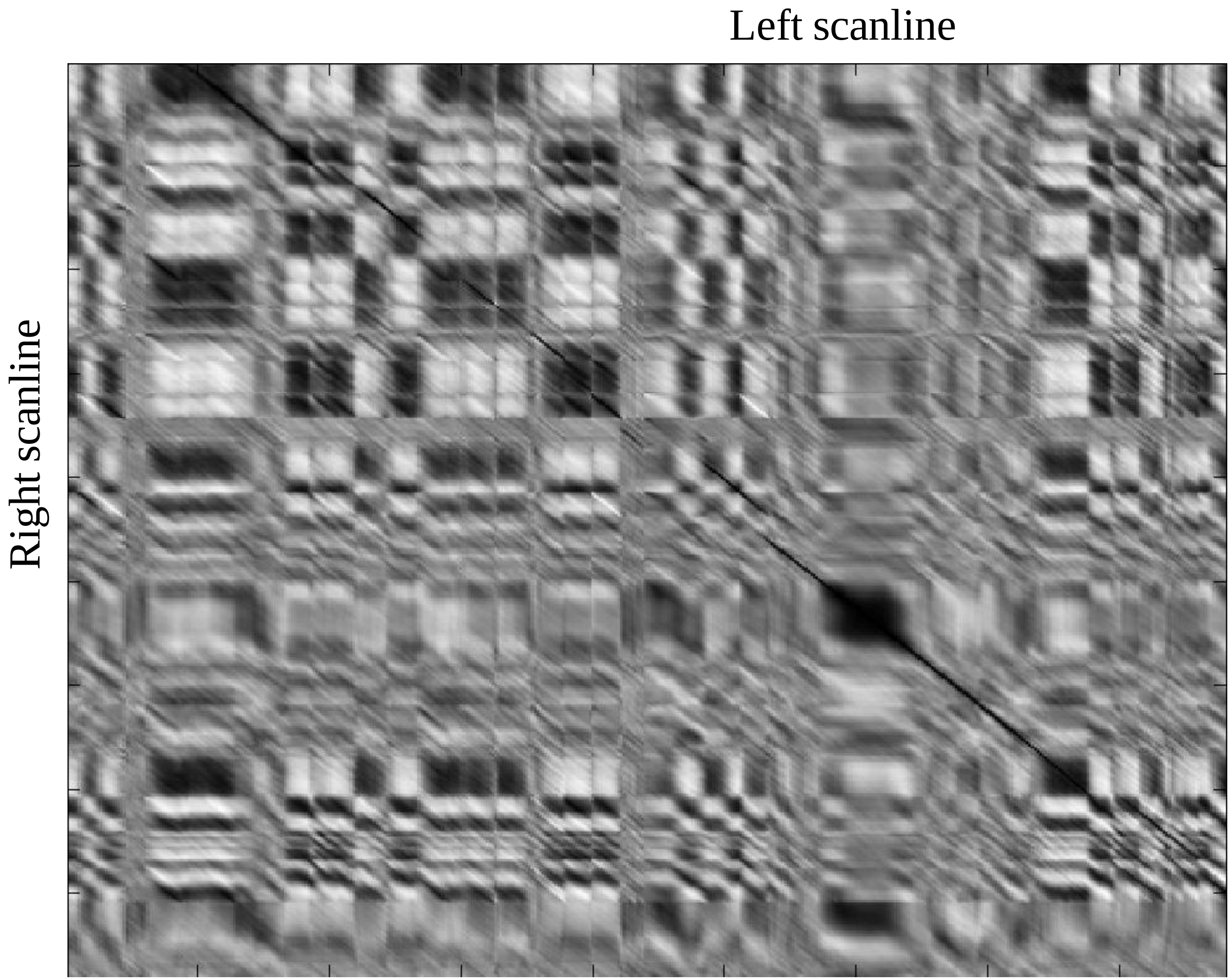
(1-NCC) or SSD

Robert Collins
CSE486, Penn State

Disparity Space Image (DSI)

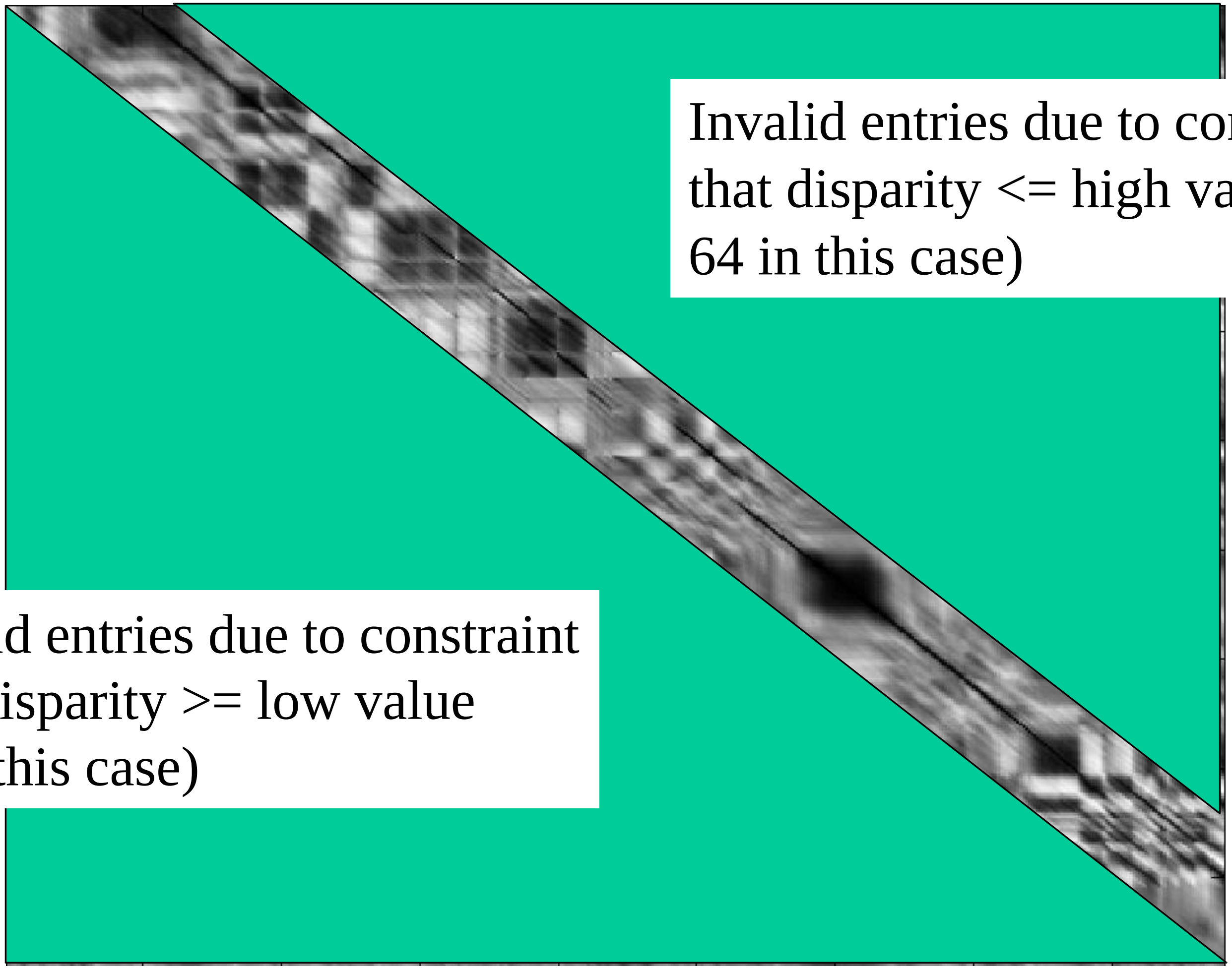


Disparity Space Image



Robert Collins
CSE486, Penn State

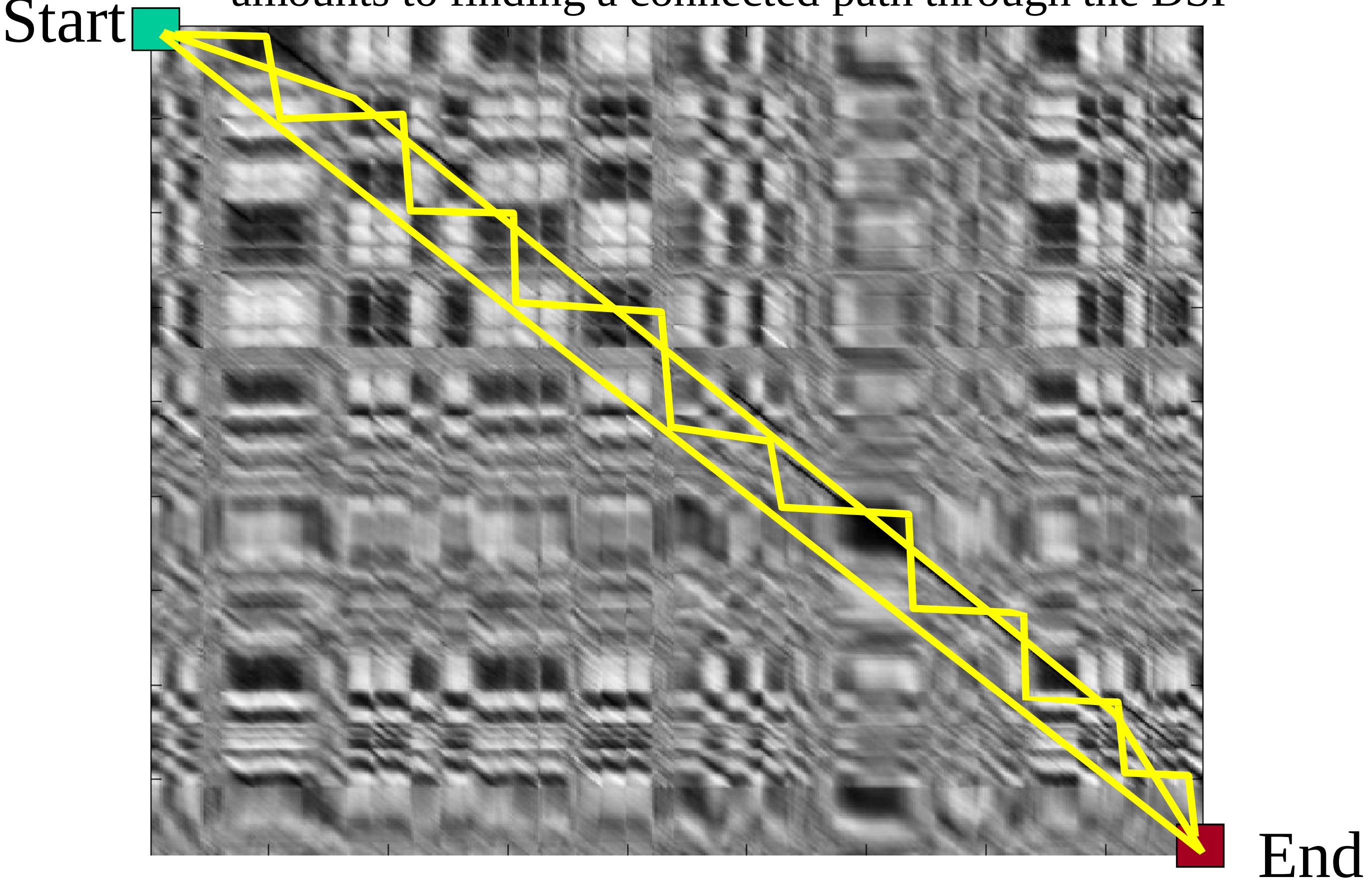
Disparity Space Image



Robert Collins
CSE486, Penn State

DSI and Scanline Consistency

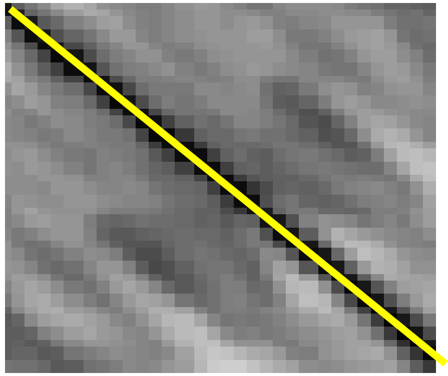
Assigning disparities to all pixels in left scanline now amounts to finding a connected path through the DSI

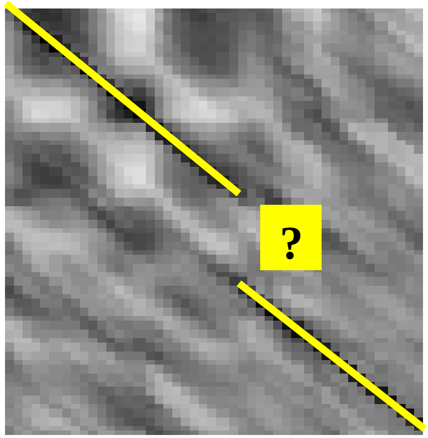


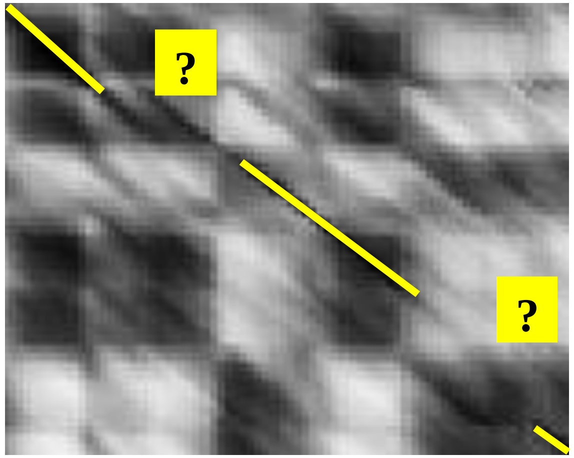
Lowest Cost Path

We would like to choose the “best” path.

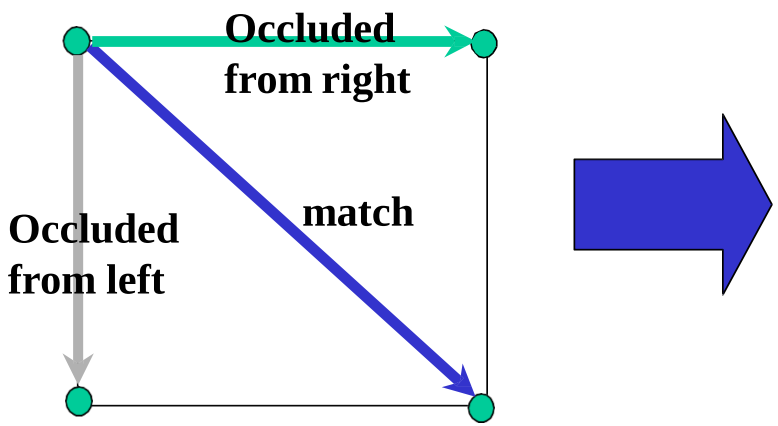
Want one with lowest “cost” (Lowest sum of dissimilarity scores along the path)

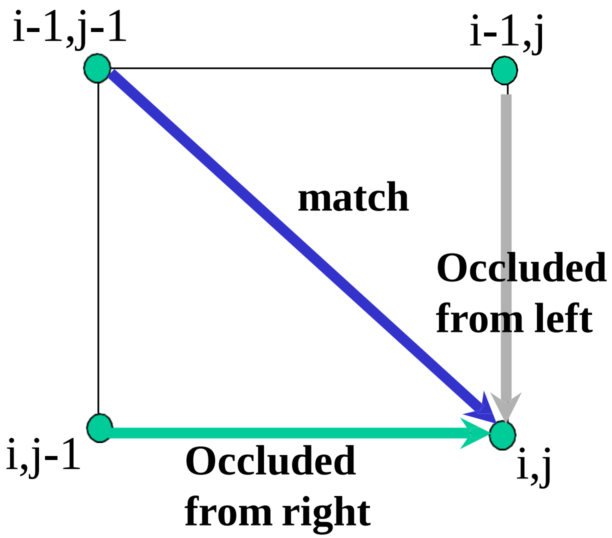






Cox et.al. Stereo Matching



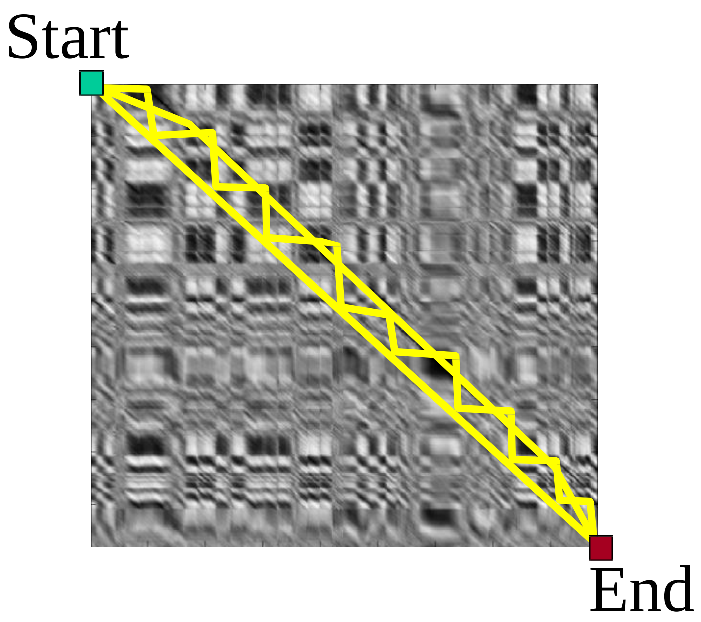


Three cases:

- Matching patches. Cost = dissimilarity score
- Occluded from right. Cost is some constant value.
- Occluded from left. Cost is some constant value.

```
C(i,j)= min([ C(i-1,j-1) + dissimilarity(i,j)
              C(i-1,j) + occlusionConstant,
              C(i,j-1) + occlusionConstant]);
```

Cox et.al. Stereo Matching



Recap: want to find lowest cost path from upper left to lower right of DSI image.

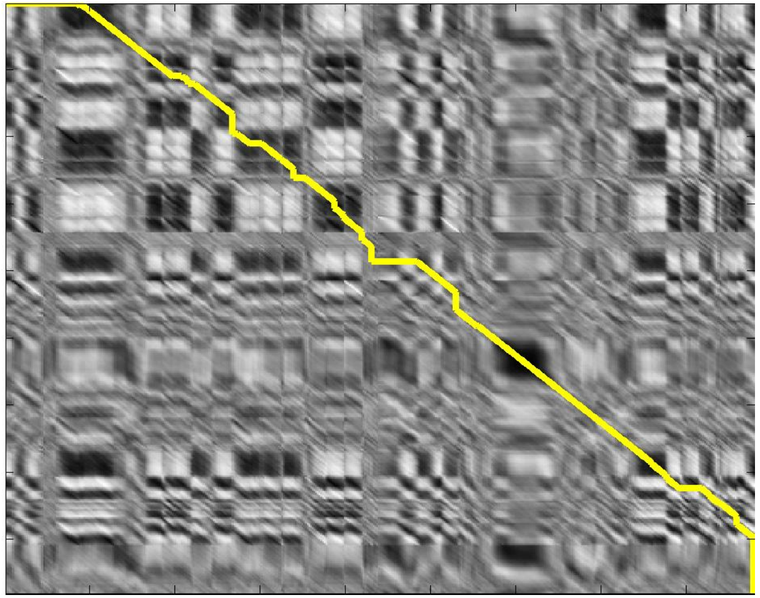
At each point on the path we have three choices: step left, step down, step diagonally.

Each choice has a well-defined cost associated with it.

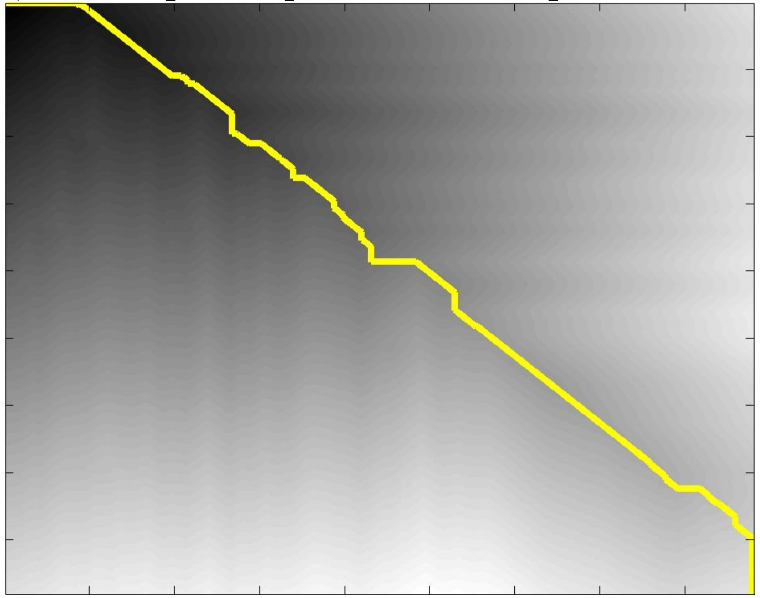
This problem just screams out for Dynamic Programming!
(which, indeed, is how Cox et.al. solve the problem)

Real Scanline Example

DSI



DP cost matrix
(cost of optimal path from each point to END)

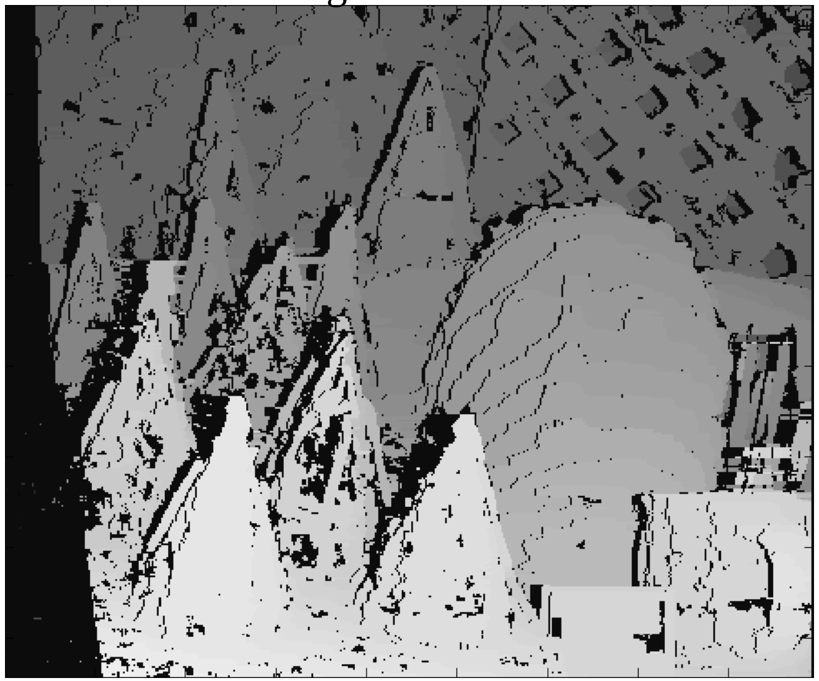


Every pixel in left column now is marked with either a disparity value, or an occlusion label.

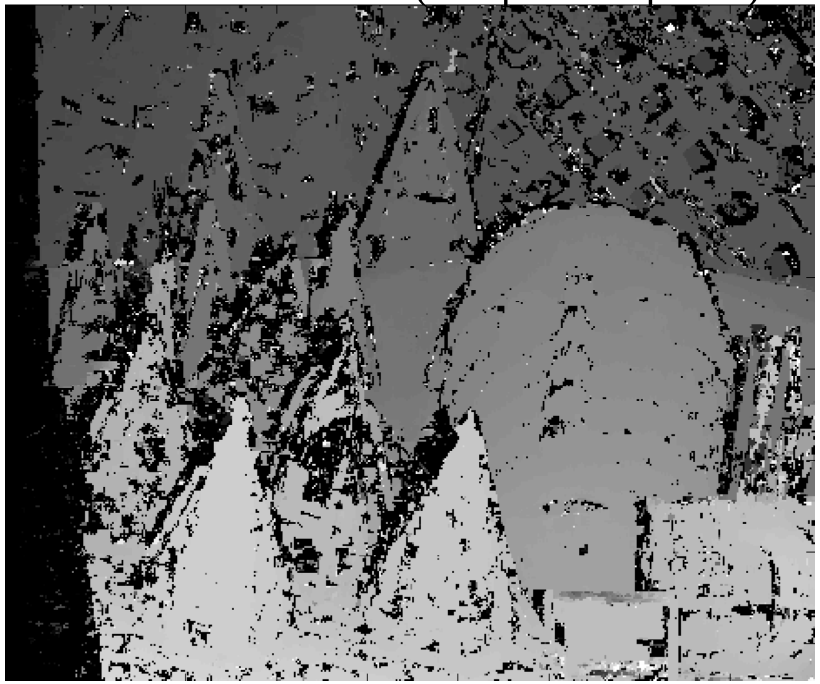
Proceed for every scanline in left image.

Example

Result of DP alg



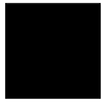
Result without DP (independent pixels)



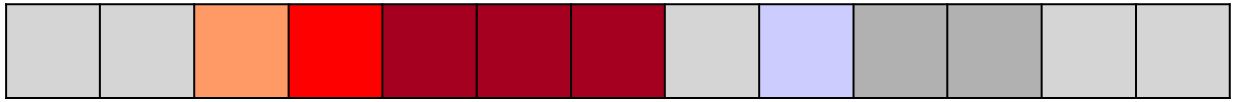
Result of DP alg. Black pixels = occluded.

Occlusion Filling

Simple trick for filling in gaps caused by occlusion.

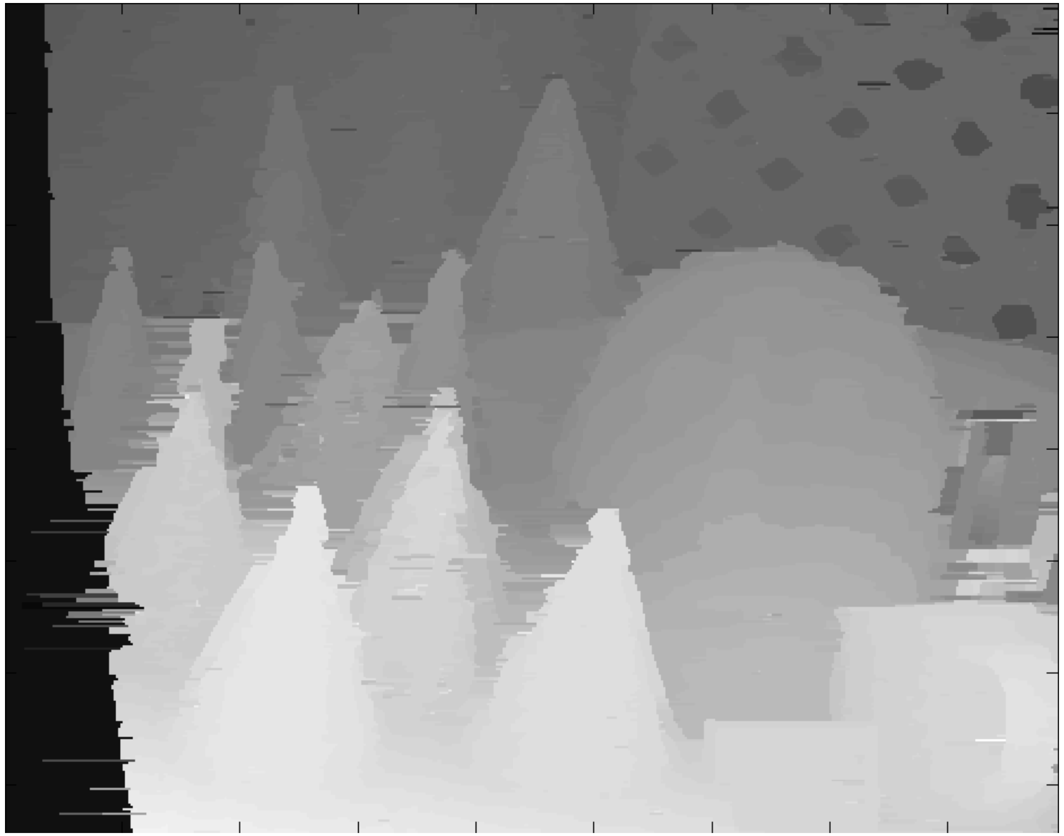
 = left occluded

Fill in left occluded pixels with value from the nearest valid pixel preceding it in the scanline.



Similarly, for right occluded, look for valid pixel to the right.

Example

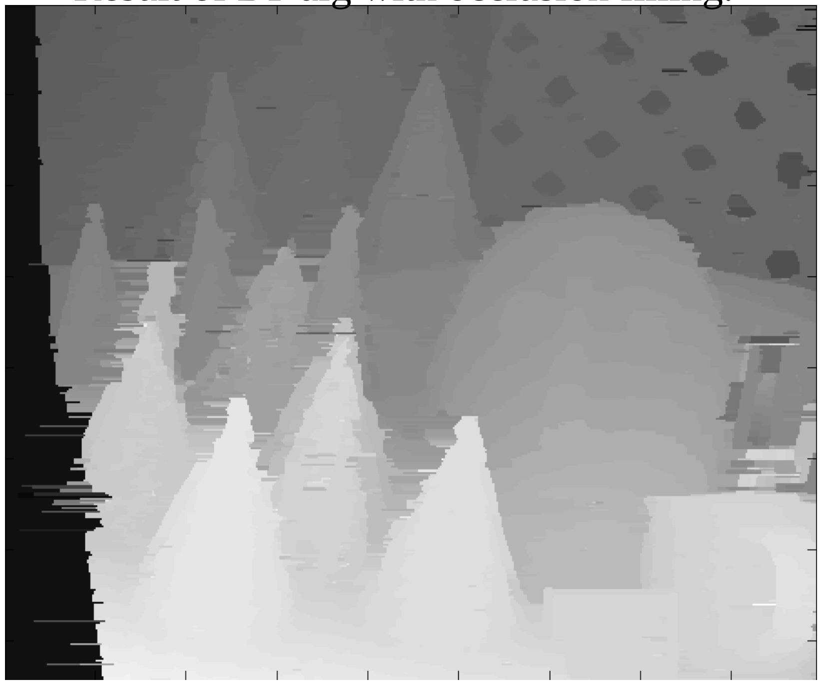


Result of DP alg with occlusion filling.

Robert Collins
CSE486, Penn State

Example

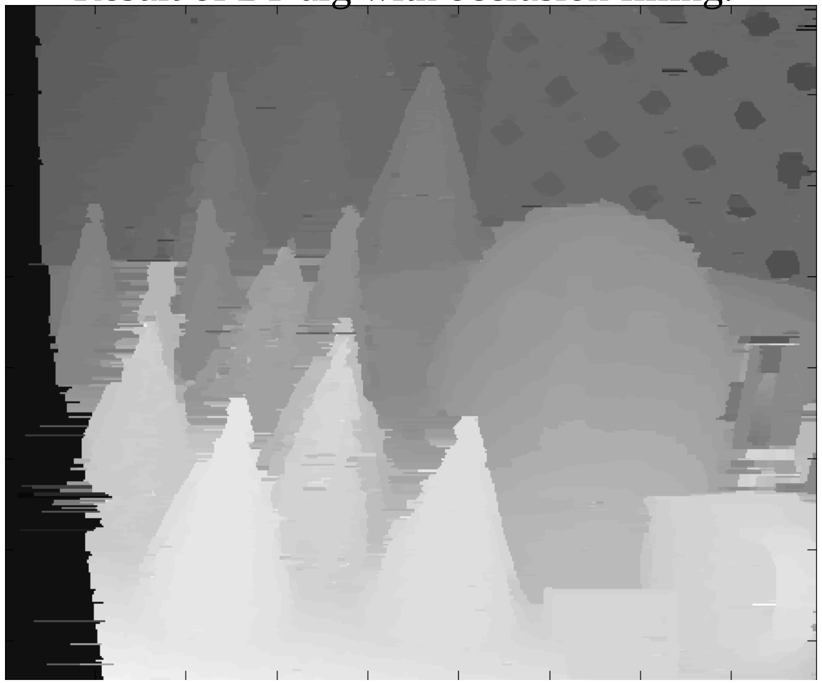
Result of DP alg with occlusion filling.



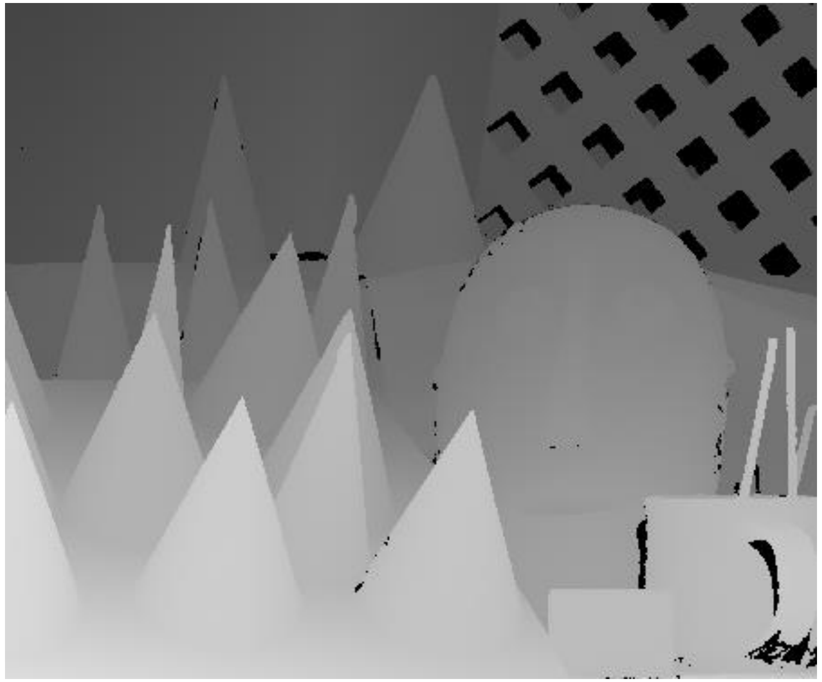
Result without DP (independent pixels)

Example

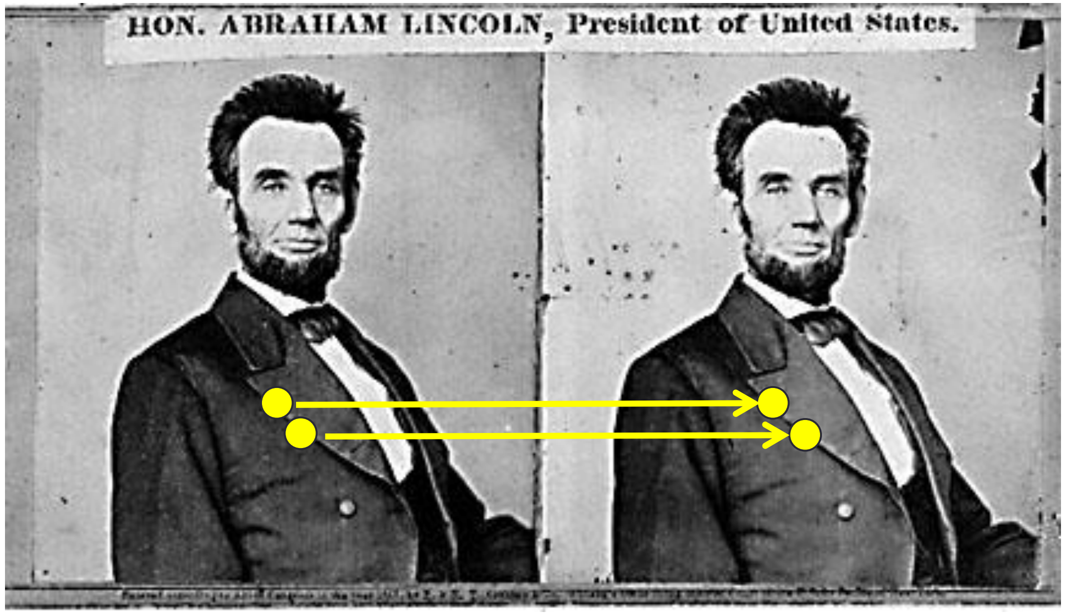
Result of DP alg with occlusion filling.



Ground truth

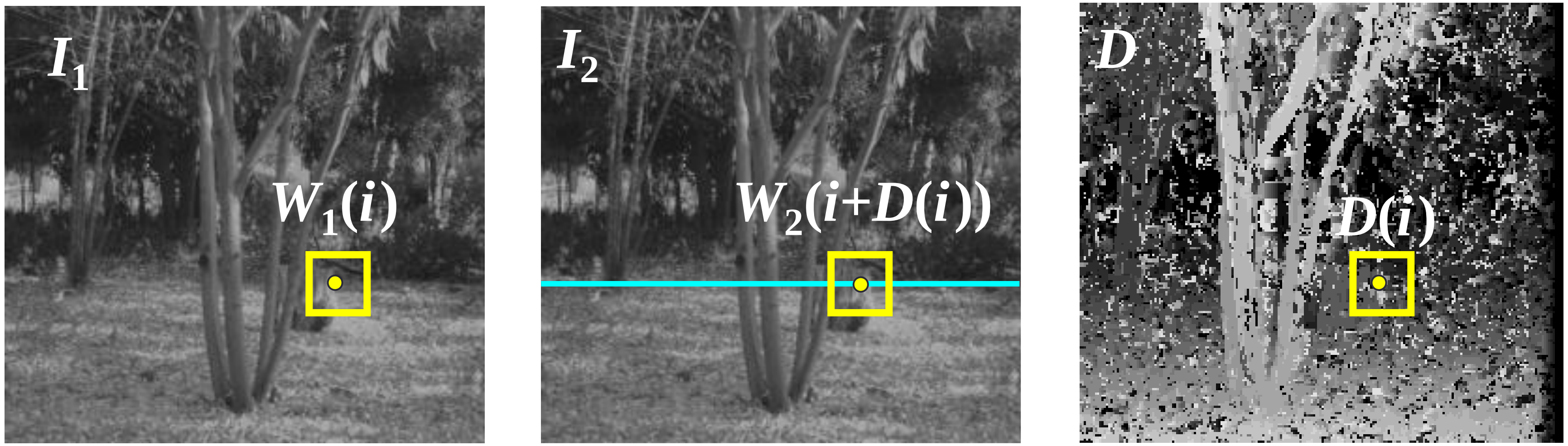


Stereo with 2D smoothness constraint



- What defines a good stereo correspondence?
  1. Match quality
     - Want each pixel to find a good match in the other image
  2. Smoothness
     - If two pixels are adjacent, they should (usually) move about the same amount

Optimizing for match quality *and* smoothness (in any direction)



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

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

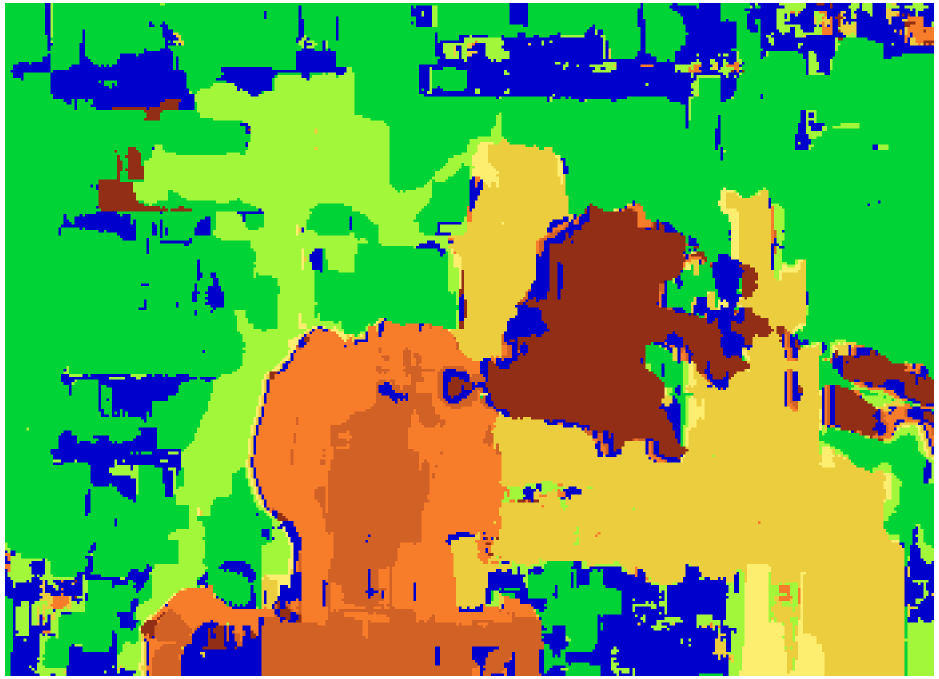
$$\boxed{E_{\text{smooth}} = \sum_{\text{neighbors}\, i,j} \rho\left(D(i) - D(j)\right)}$$

- Energy functions of this form can be minimized using *graph cuts*

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

Source: Steve Seitz

Results with window search



Window-based matching
(best window size)

Ground truth

Better results…

Graph cut method

Ground truth

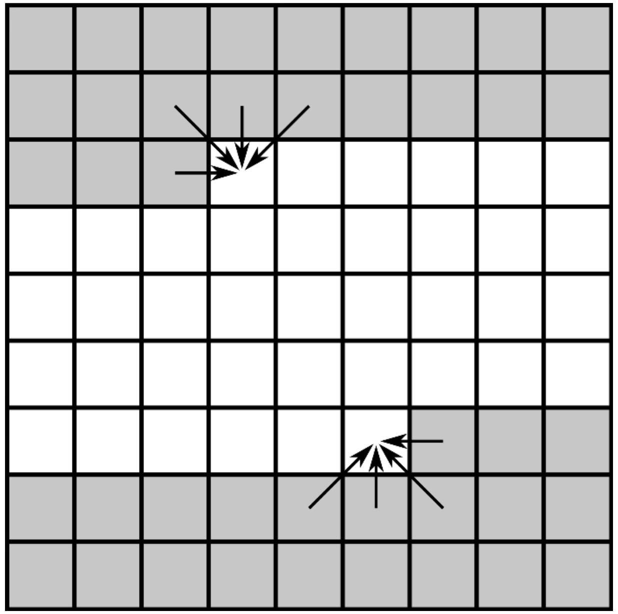
Boykov et al., [Fast Approximate Energy Minimization via Graph Cuts](), International Conference on Computer Vision, September 1999.

For the latest and greatest: https://vision.middlebury.edu/stereo/eval3/

Semi-global matching

$$E(D) = \sum_{\mathbf{p}} \Big(C(\mathbf{p}, D_{\mathbf{p}}) + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_1 \, \mathrm{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_2 \, \mathrm{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] \Big)$$

- Approximate the full smoothness optimization by considering 8 or 16 directions in two or three passes.
- Optimization looks like scanline, dynamic programming stereo, but with a 2d notion of smoothness



Stereo Processing by Semi-Global Matching and Mutual Information. Hirschmuller, PAMI 2007. **3500+ citations**

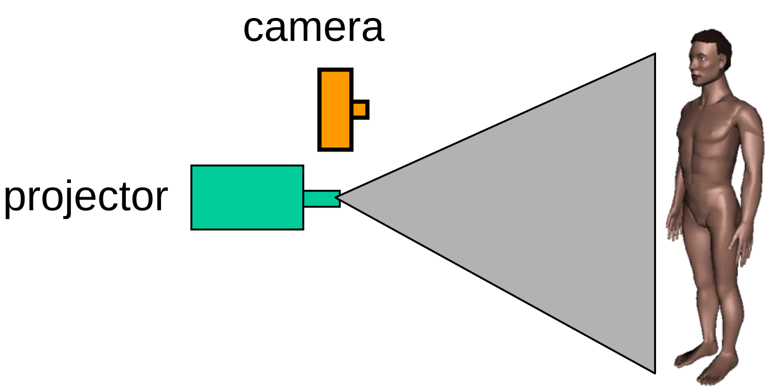
Semi-global matching

Stereo Depth Estimation Challenges

- Low-contrast ; textureless image regions
- Occlusions
- Violations of brightness constancy (e.g., specular reflections)
- Really large baselines (foreshortening and appearance change)
- Camera calibration errors

Active stereo with structured light

- Project “structured” light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera

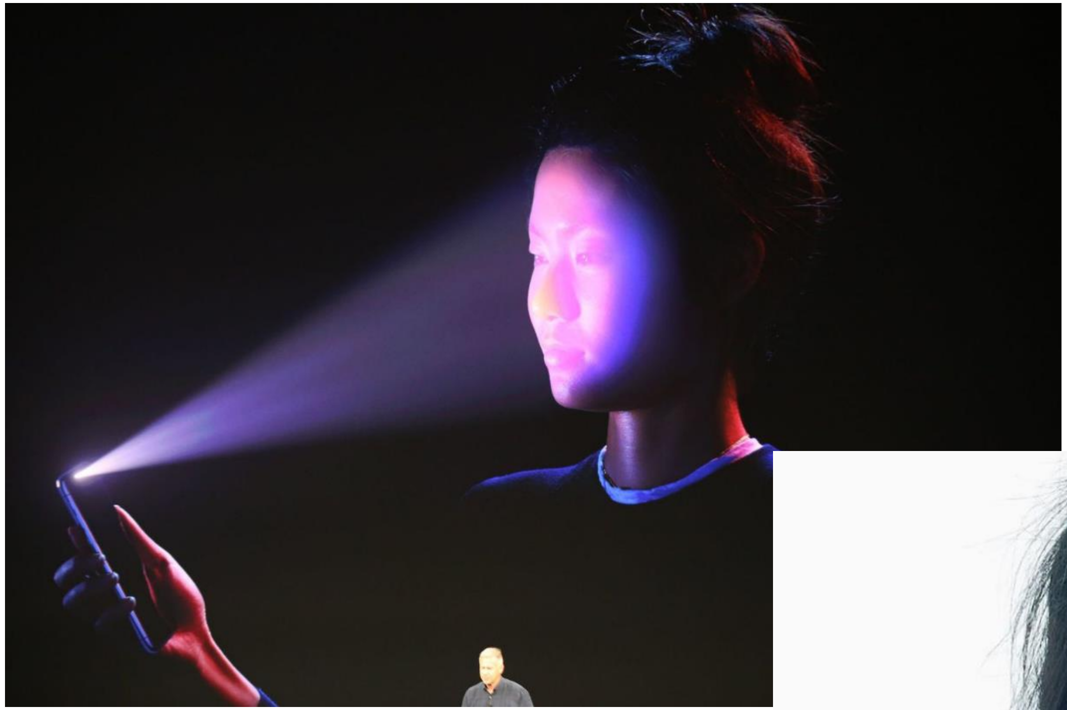


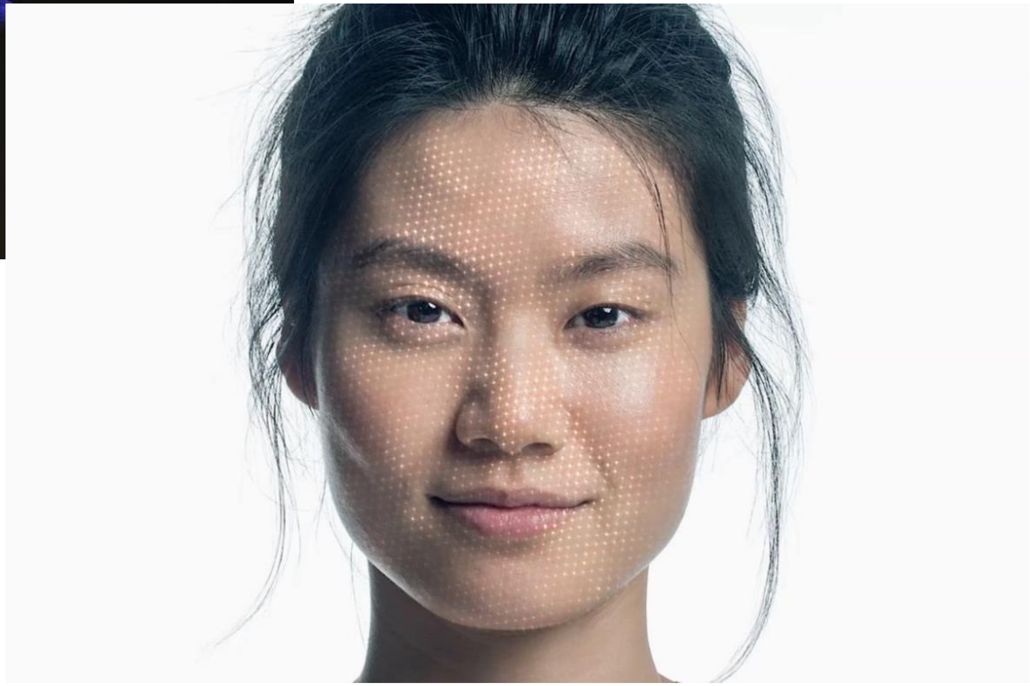
L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.]() *3DPVT* 2002

Kinect: Structured infrared light

http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

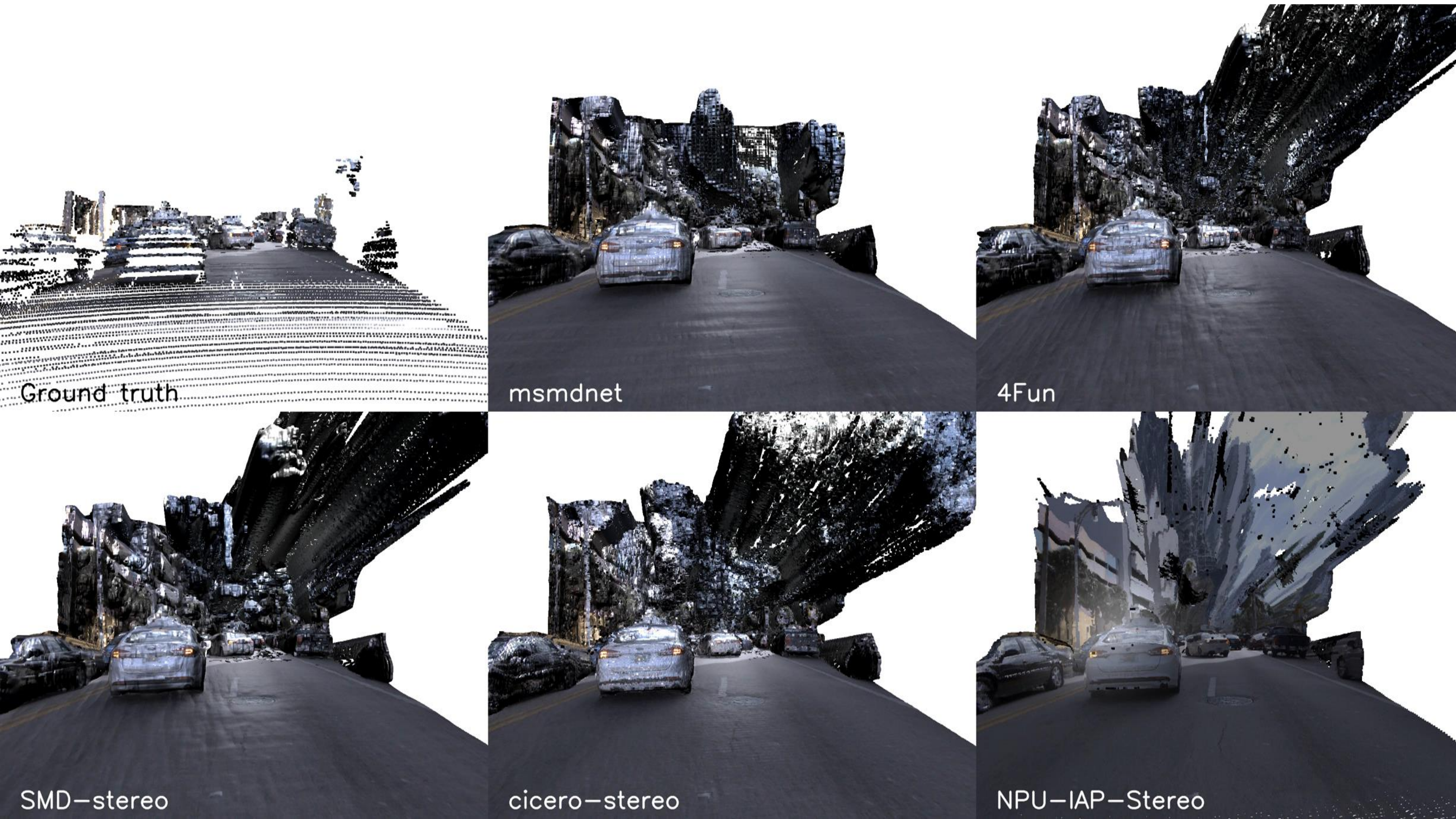
iPhone X





iPhone 12 switched to lidar
(time of flight)

Self-driving efforts use both lidar and stereo


Ground truth
msmdnet
4Fun
SMD-stereo
cicero-stereo
NPU-IAP-Stereo