CEng 583 - Computational Vision

2011-2012 Spring

Week – 4

18th of March, 2011

Tentative Schedule:

Week & Date		Topic
1		**Introduction to Vision.** What is vision? What are its goals and problems? What are the main processing stages?
2		**Low-level Vision**. Cameras. Projective geometry. Calibration.
3		**Early Vision**. Edges. Corners. Texture. Segmentation. Optic Flow.
4		**3D Vision**. Monocular and binocular cues. 3D reconstruction.
5		**Applications**. Video surveillance. Human behaviour understanding. Object recognition. Image/video retrieval. Image annotation.
6		**Paper presentations with theme**: Monocular depth estimation.
7		**Paper presentations with theme**: Image annotation.
8		**Paper presentations with theme**: Object/shape modelling. Object recognition.
9		**Paper presentations with theme**: Feature Descriptors.
10		**Paper presentations with theme**: Context. Saliency. Attention.
11		**Project Presentations**
12		**Project presentations**
13		**Project presentations**
14		**Project presentations**

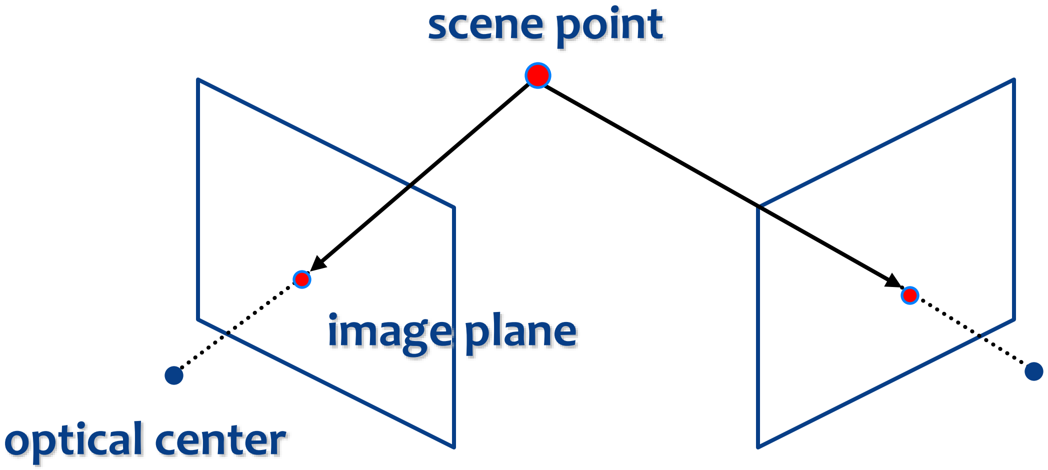
Today

* 3D Vision
  * Binocular (Multi-view) cues:
    * Stereopsis
    * Motion
  * Monocular cues
    * Shading
    * Texture
    * Familiar size
    * etc.

"God must have loved depth cues, for He made so many of them." -- (Yonas & Ganrud, 1985)

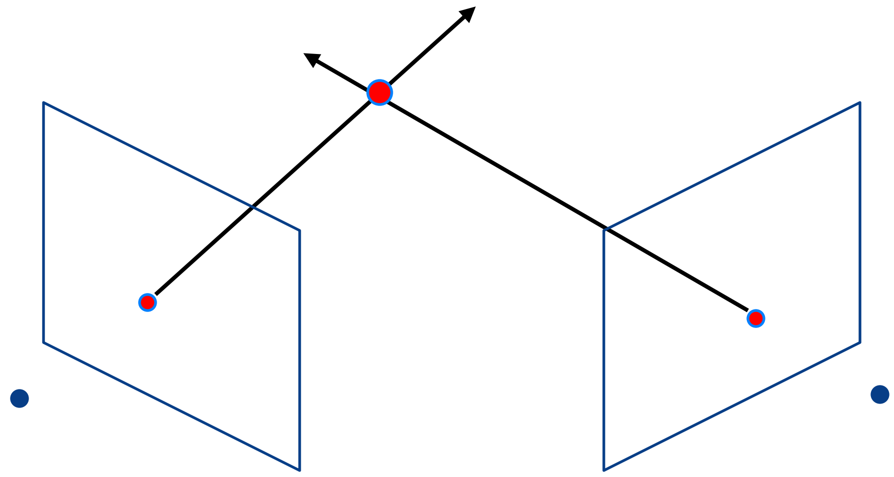
Binocular Cues: Stereopsis

Depth with stereo: basic idea



Source: Steve Seitz

Depth with stereo: basic idea

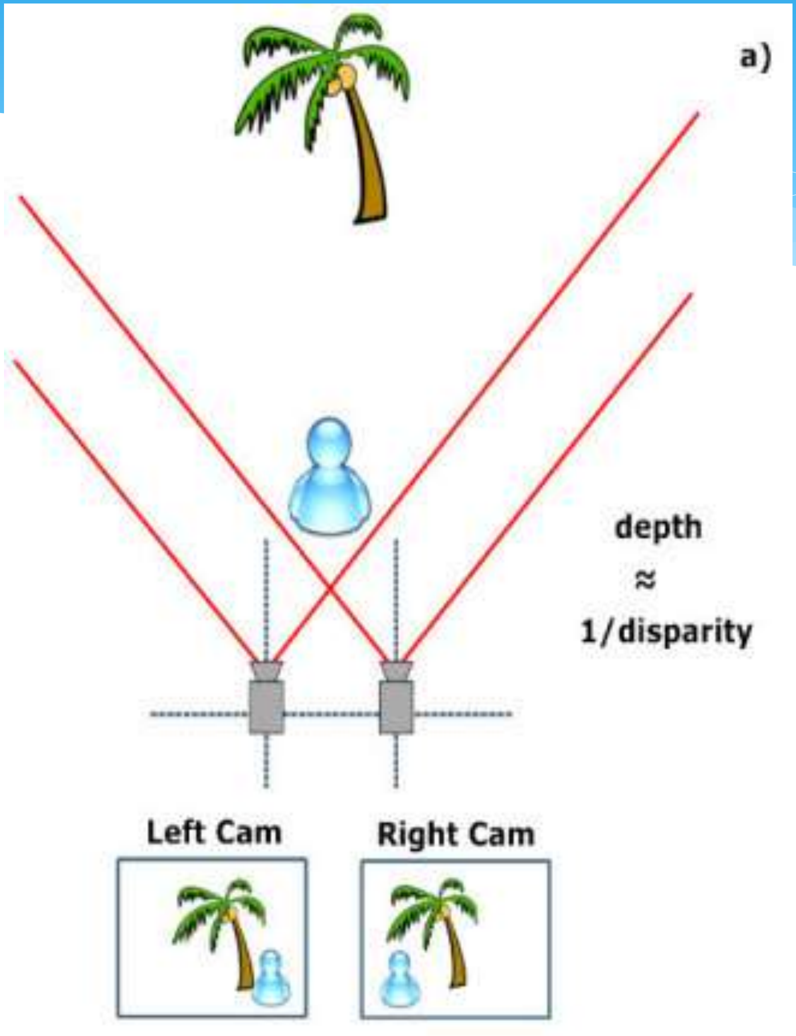


Basic Principle: Triangulation

- Gives reconstruction as intersection of two rays
- Requires
  - camera pose (calibration)
  - ***point correspondence***

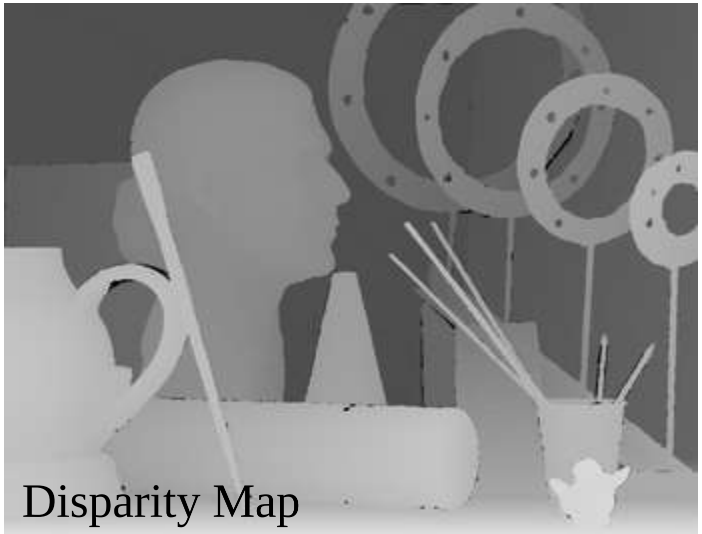
Source: Steve Seitz

The Problem



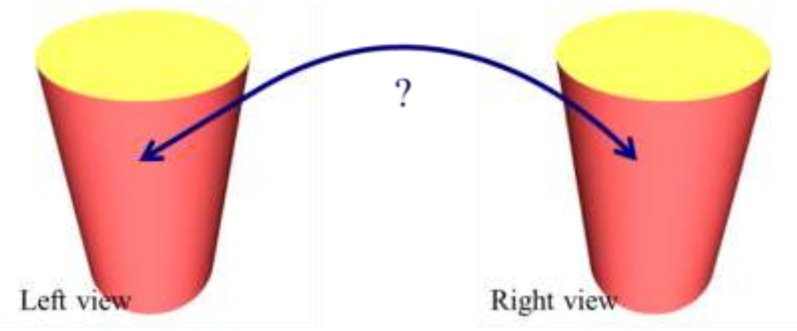
Picture: http://www.imec.be/ScientificReport/SR2007/html/1384302.html

The Problem

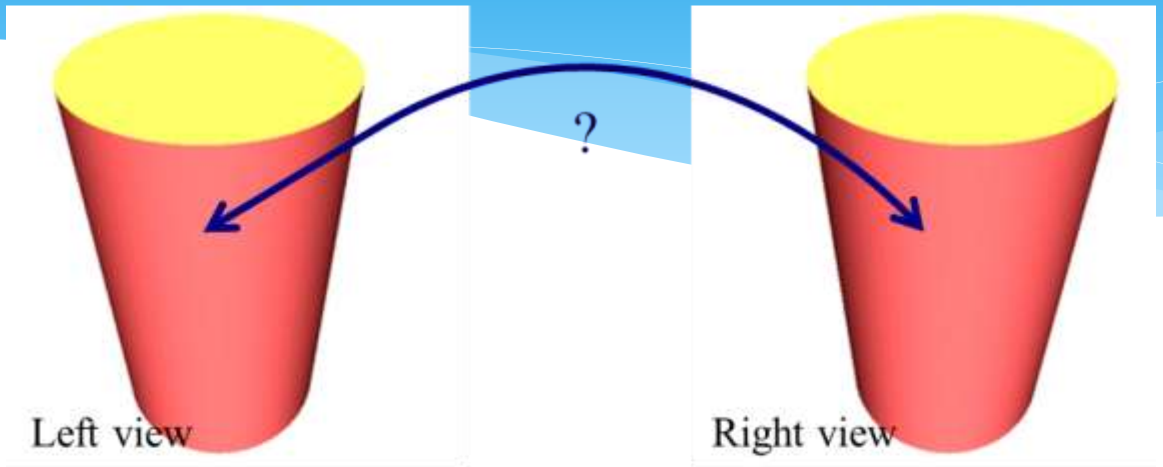


The Problem

* Calibration
  * If you are interested in 3D reconstruction or utilizing the epipolar line
* Matching
  * Computing Similarities
  * Finding the “best” match for each pixel/feature
  * Gives us the disparities
* 3D Reconstruction



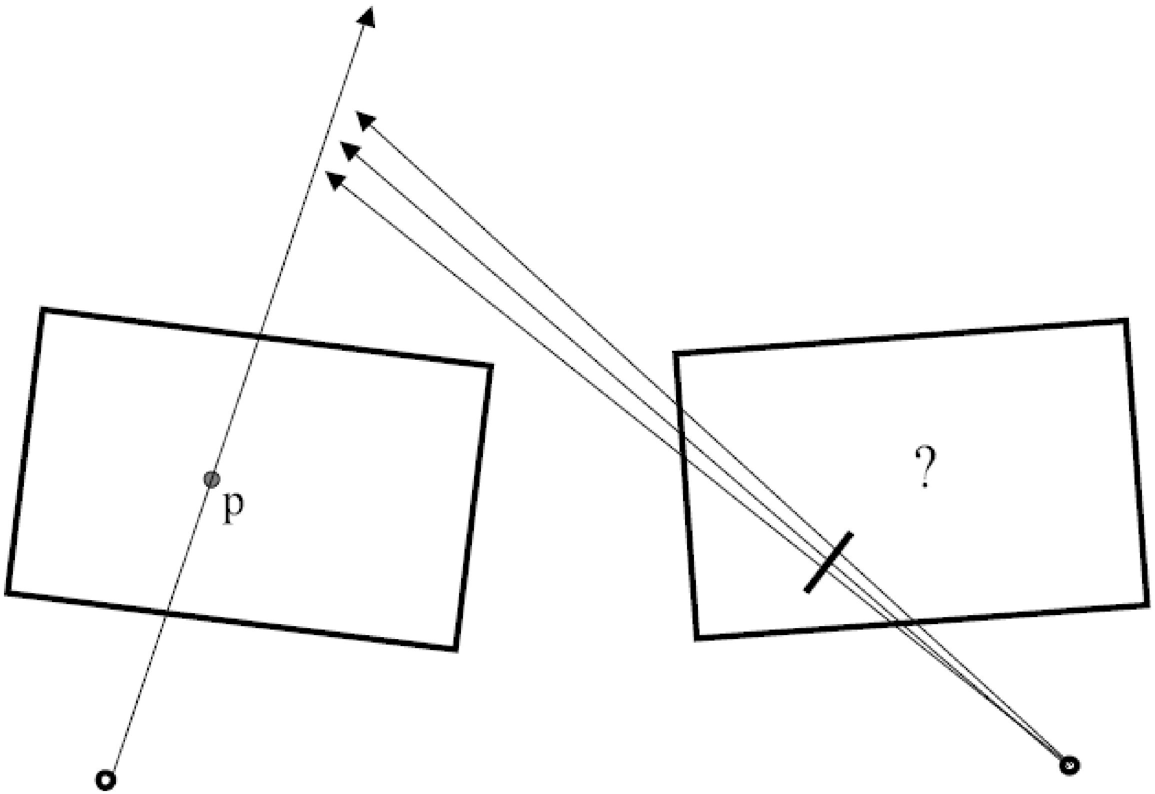
Correspondence Problem



- How can we match pixels?
  - Local versus Global Matching
- Especially homogeneous ones?
- What if we cannot find a match?
  - → Interpolation, Filling-in

(Barrow&Tenenbaum, 1981)

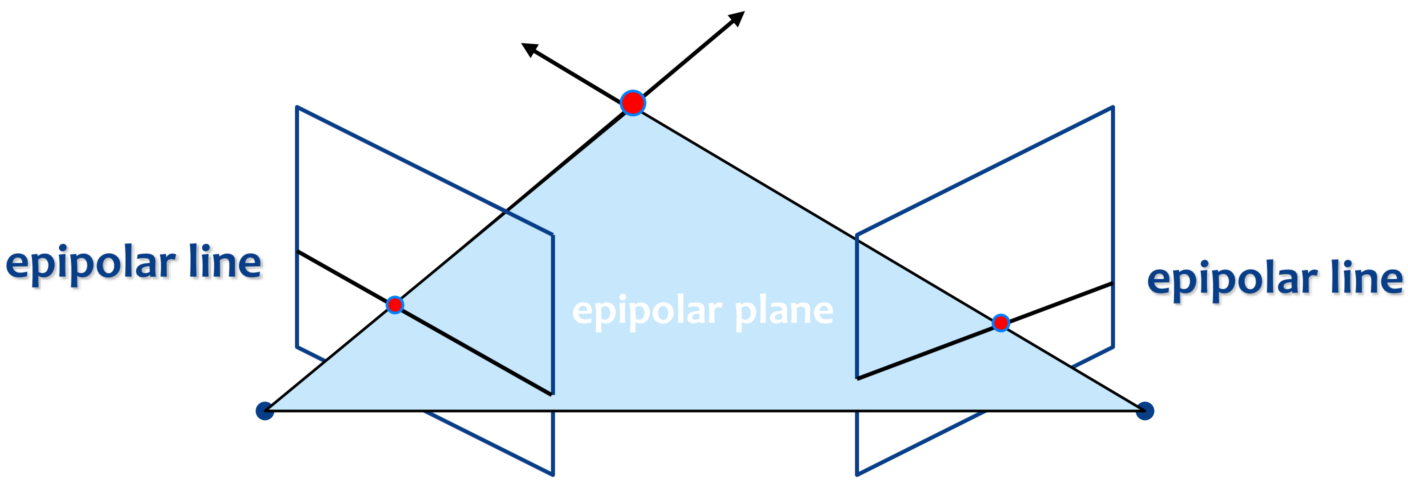
Stereo correspondence constraints



Trevor Darrell

Stereo correspondence constraints

- Geometry of two views allows us to constrain where the corresponding pixel for some image point in the first view must occur in the second view.

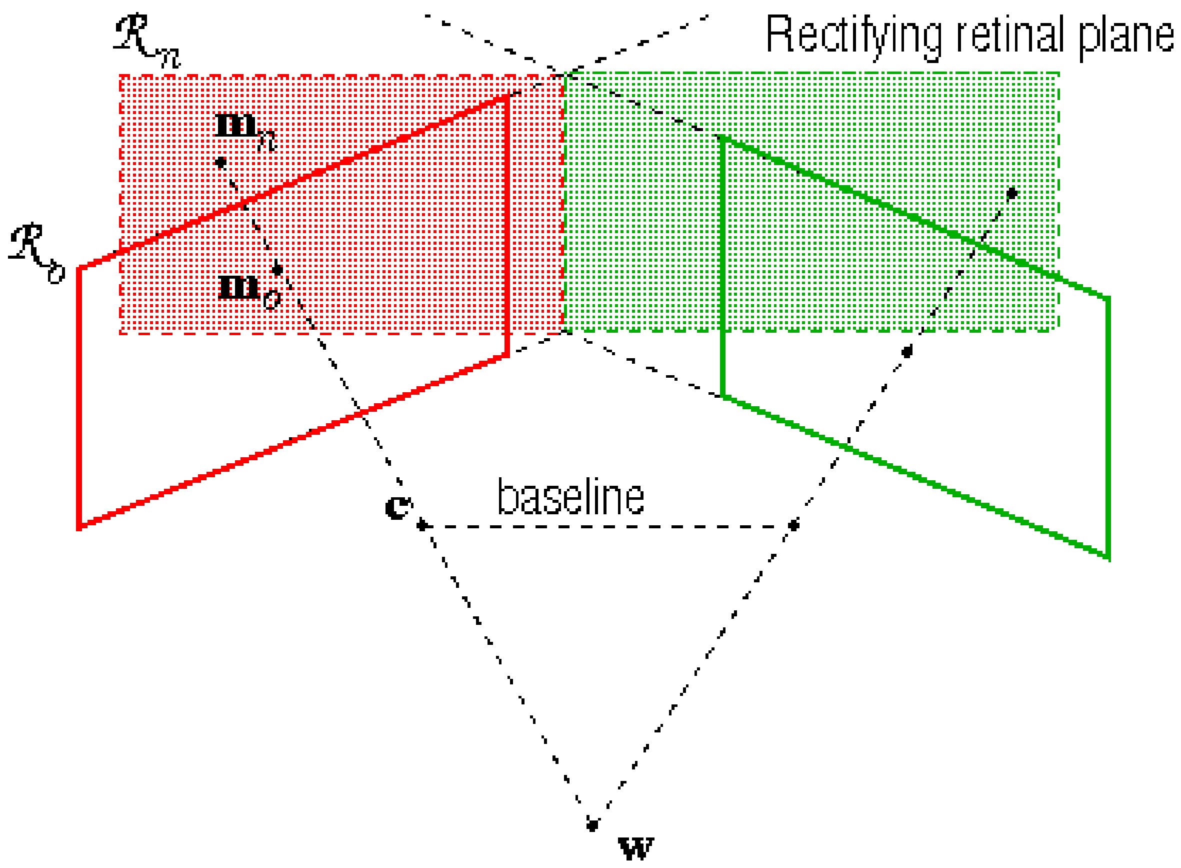


Epipolar constraint: Why is this useful?

- Reduces correspondence problem to 1D search along *conjugate epipolar lines*

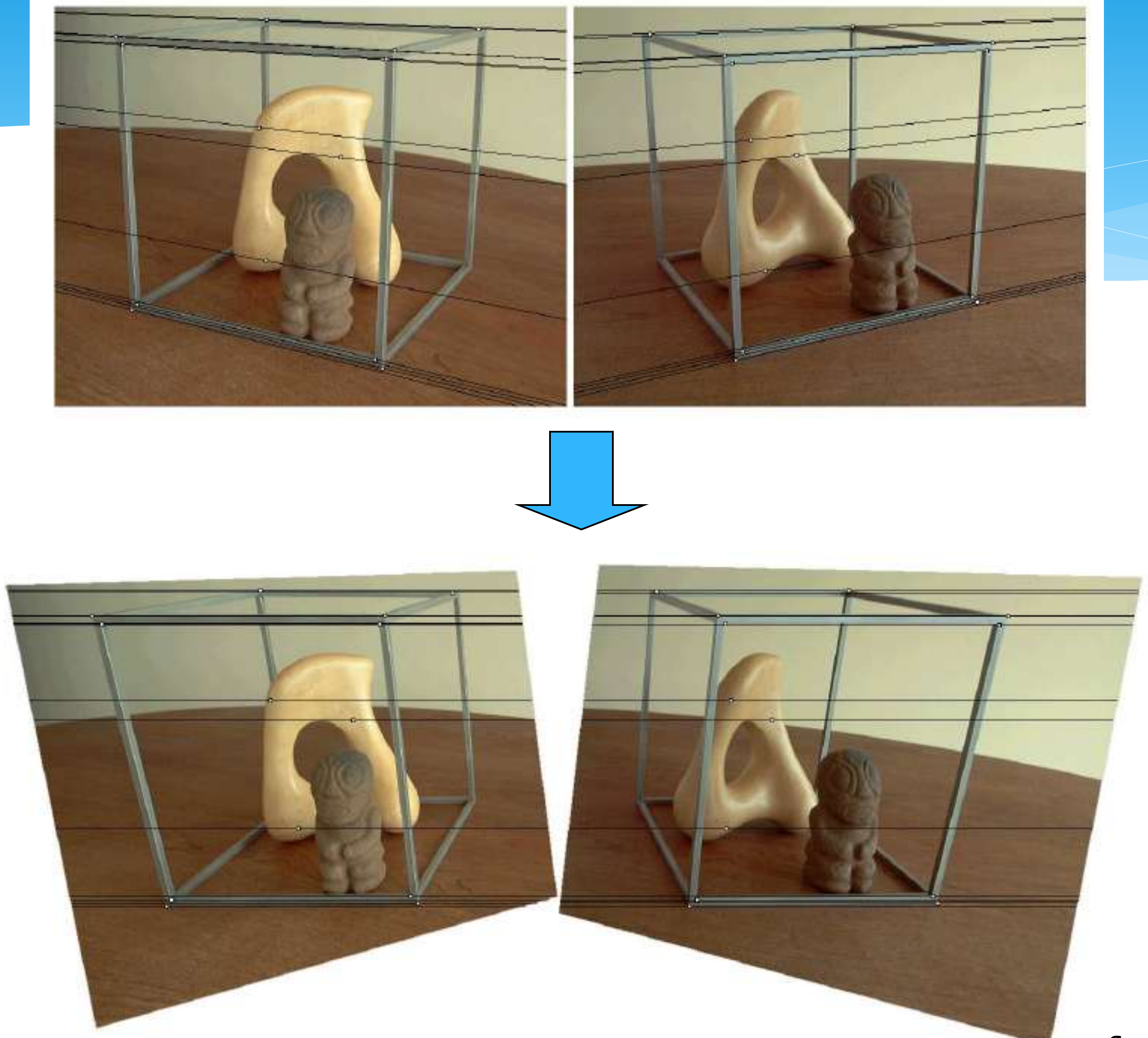
Adapted from Steve Seitz

Stereo image rectification



http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FUSIELLO/tutorial.html

Stereo image rectification: example



Source: Alyosha Efros

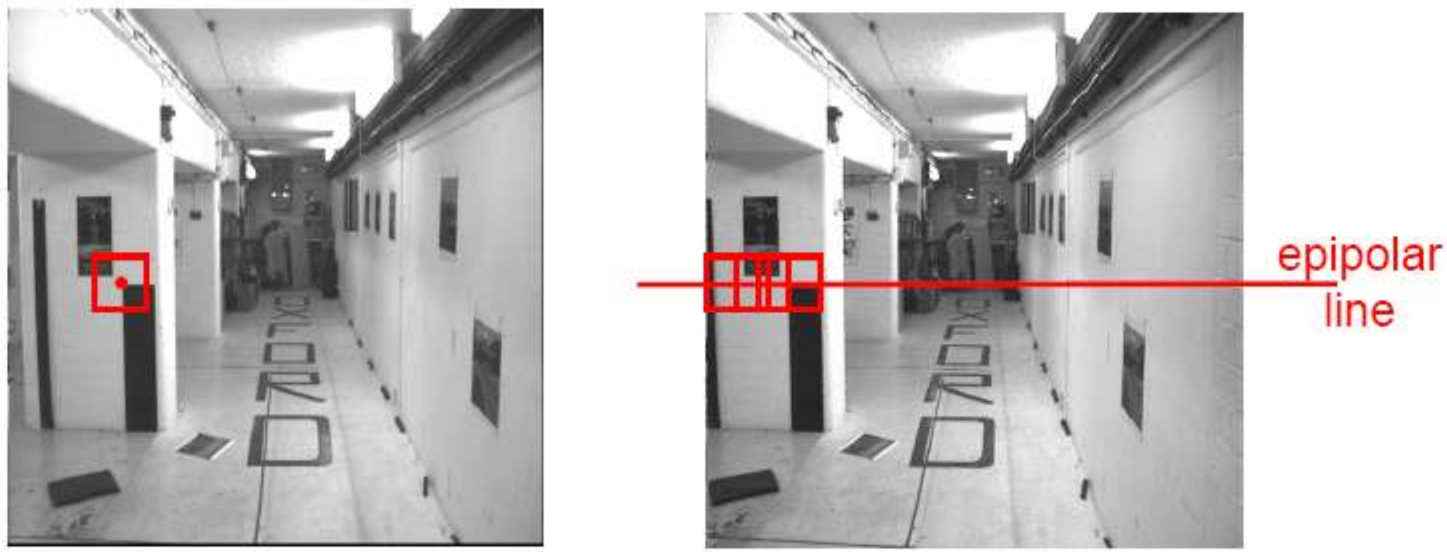
Correspondence problem

* 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

Grauman

Correspondence problem



Neighborhood of corresponding points are similar in intensity patterns.

Source: Andrew Zisserman

Computing Similarity

TABLE 2
Common Block-Matching Methods (See Fig. 4 for Visual Description of Terms)

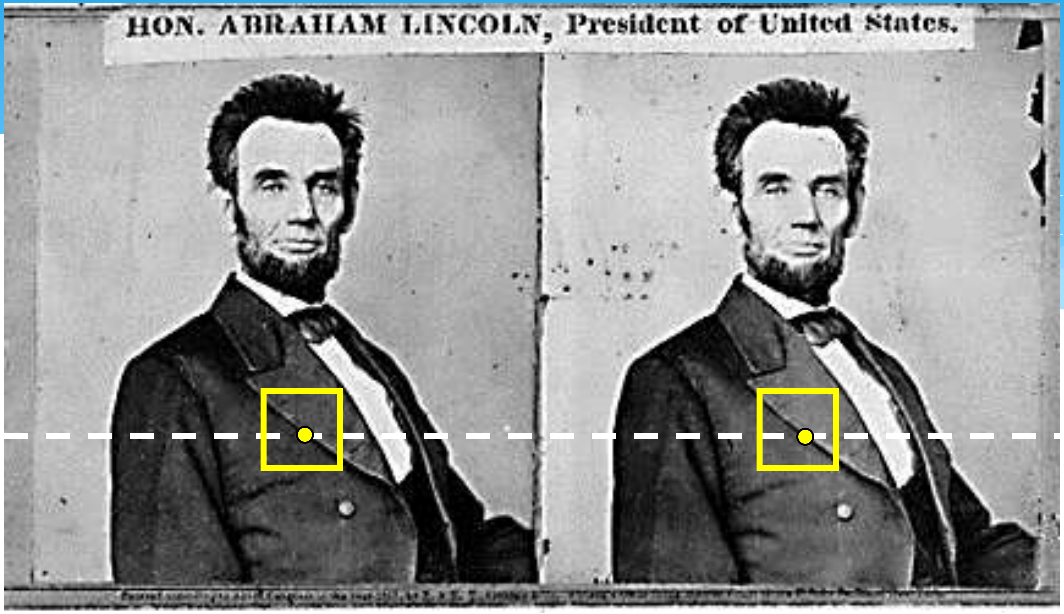
MATCH METRIC	DEFINITION
Normalized Cross-Correlation (NCC)	$\dfrac{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)\cdot\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2\cdot\left(I_2(u+d,v)-\bar{I}_2\right)^2}}$
Sum of Squared Differences (SSD)	$\sum_{u,v}\left(I_1(u,v)-I_2(u+d,v)\right)^2$
Normalized SSD	$\sum_{u,v}\left(\dfrac{\left(I_1(u,v)-\bar{I}_1\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2}}-\dfrac{\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_2(u+d,v)-\bar{I}_2\right)^2}}\right)^2$
Sum of Absolute Differences (SAD)	$\sum_{u,v}\left\|I_1(u,v)-I_2(u+d,v)\right\|$
Rank	$\sum_{u,v}\left(I_1'(u,v)-I_2'(u+d,v)\right)$ $I_k'(u,v)=\sum_{m,n}I_k(m,n)<I_k(u,v)$
Census	$\sum_{u,v}HAMMING\left(I_1'(u,v),I_2'(u+d,v)\right)$ $I_k'(u,v)=BITSTRING_{m,n}\left(I_k(m,n)<I_k(u,v)\right)$

Correlation-based window matching

left image band (x)

Source: A

Dense correspondence search



For each epipolar line

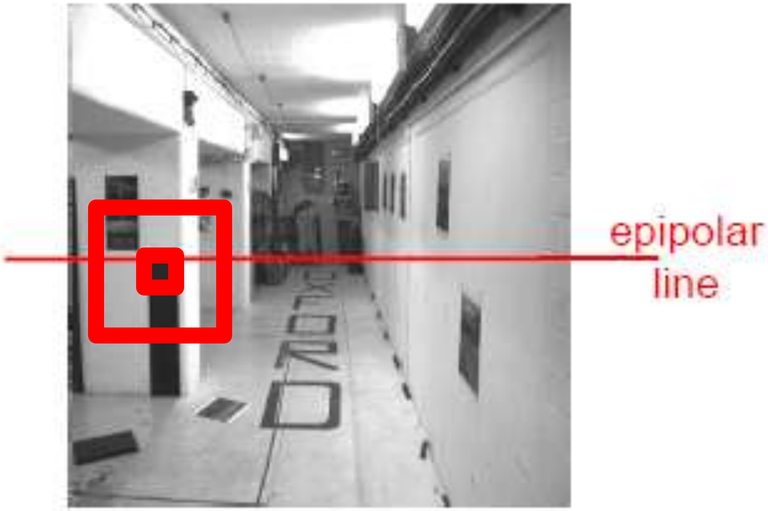
For each pixel / window in the left image

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

Adapted from Li Zhang

Grauman

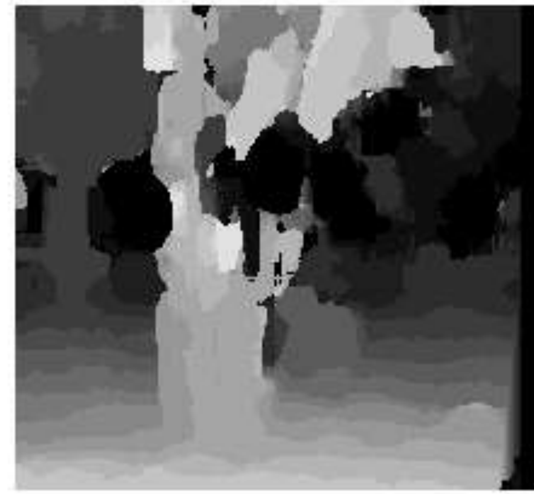
Effect of window size



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

Grauman

Uniqueness

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

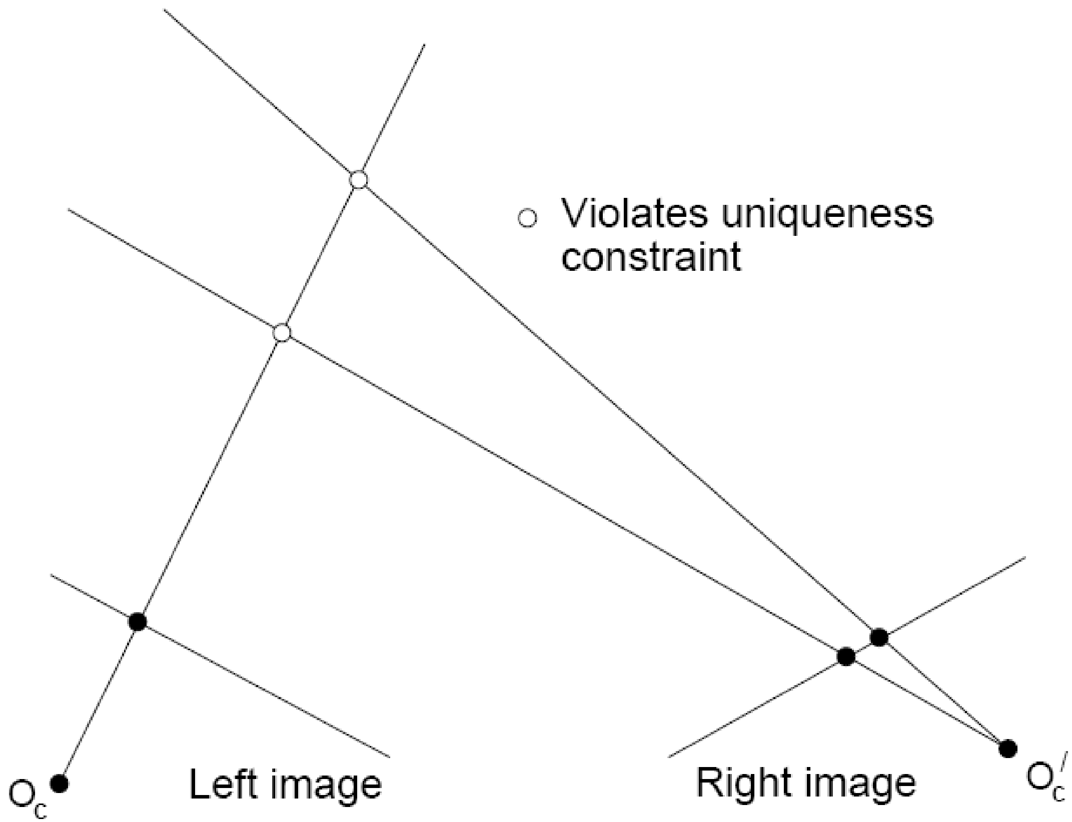


Figure from Gee & Cipolla 1999

Grauman

Ordering constraint

* Points on **same surface** (opaque object) will be in same order in both views

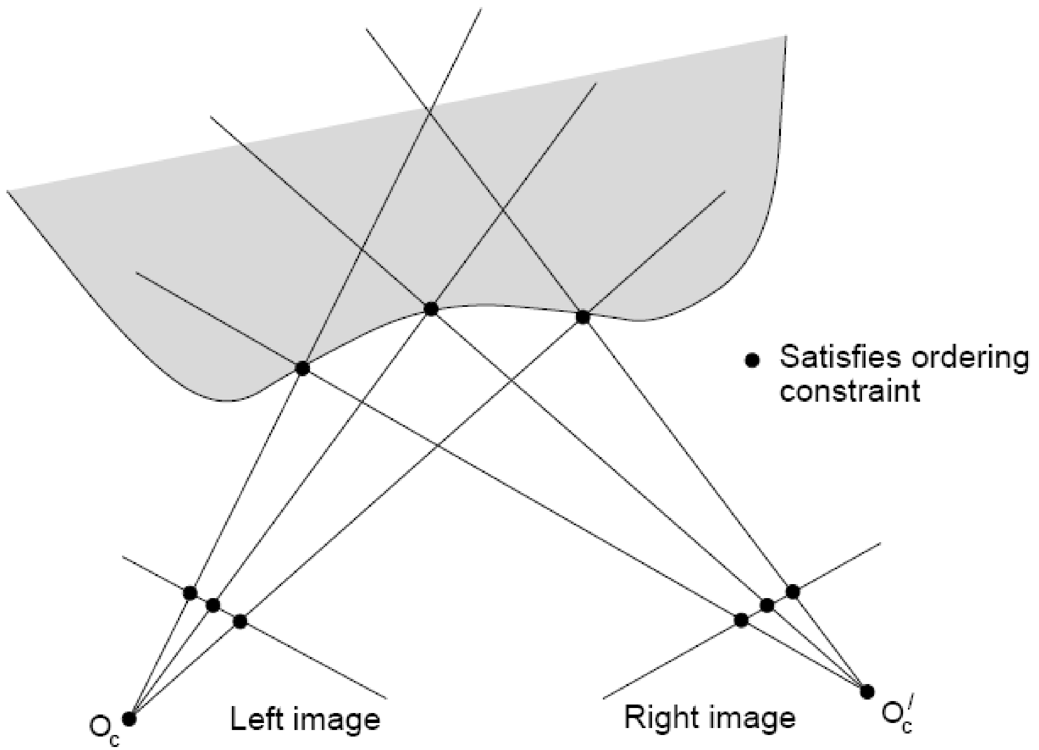
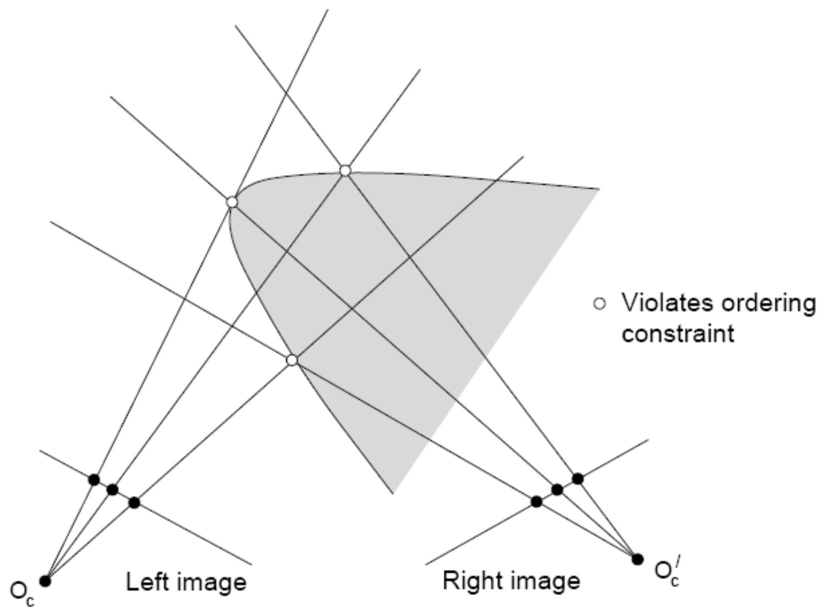


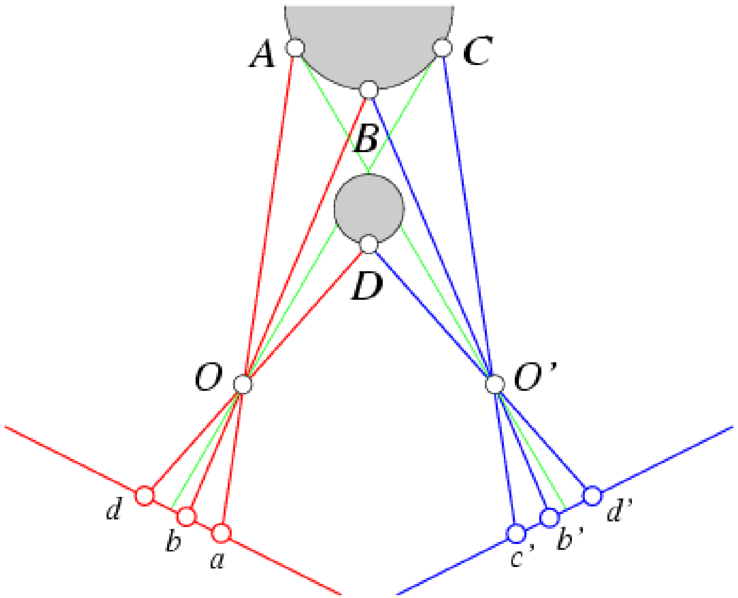
Figure from Gee & Cipolla 1999

Grauman

Ordering constraint

- Won't always hold, e.g. consider transparent object, or an occluding surface

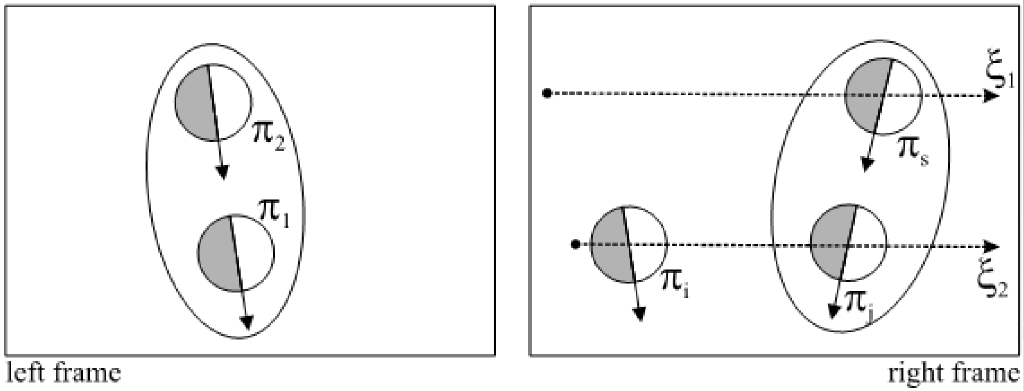




Figures from Forsyth & Ponce

Grauman

Grouping Constraint



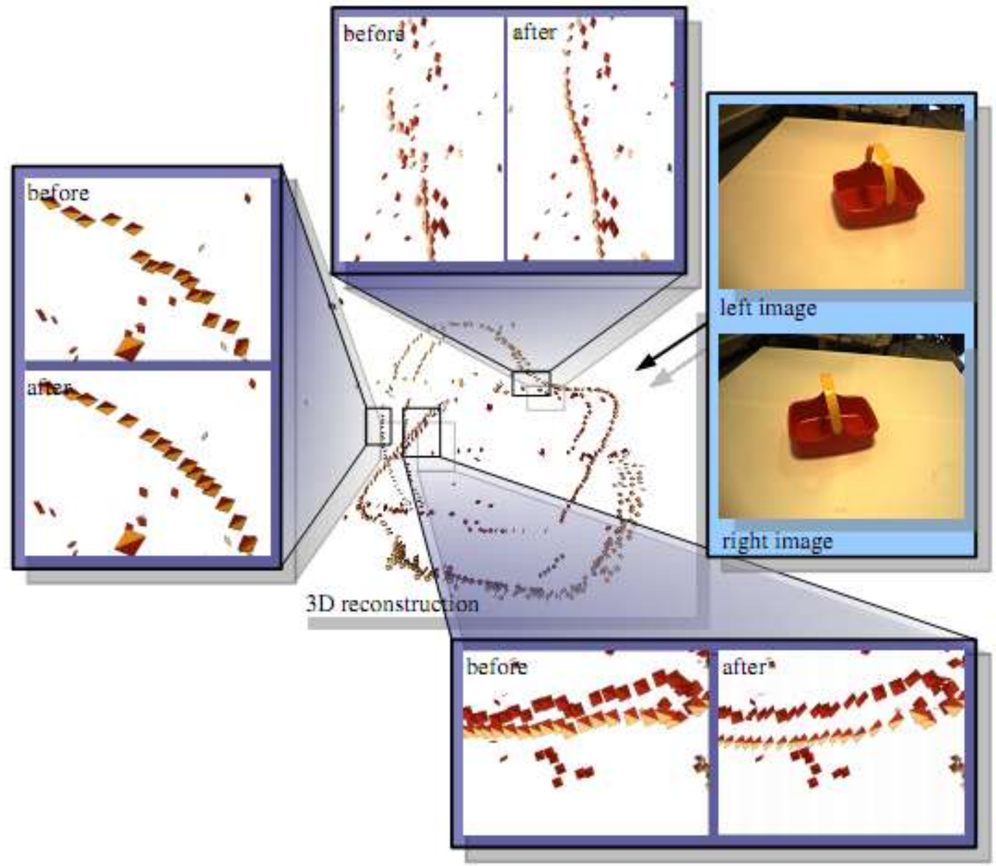
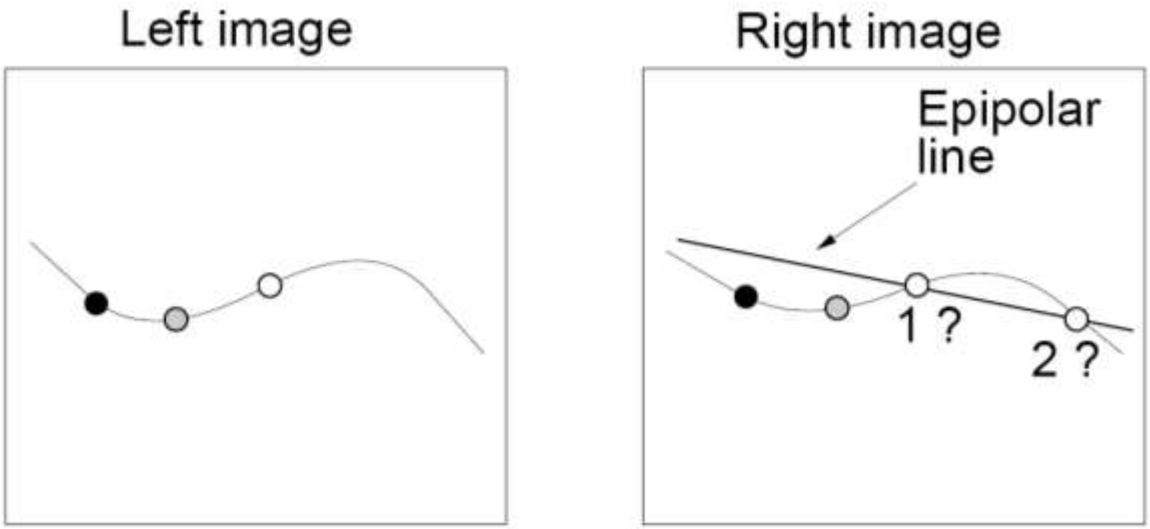


Figure 5.6: Illustration of the effects of the 3D–primitives' correction using interpolation.

Pugeault et al., 2006; 2008.

Disparity gradient

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



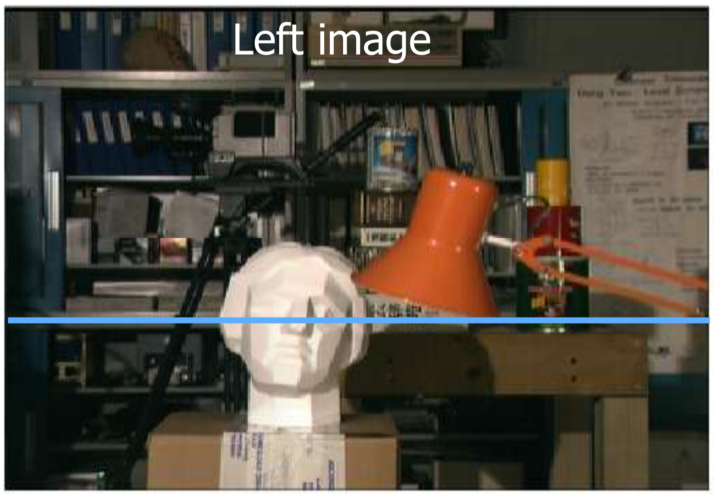
Given matches ● and ◉, point ○ in the left image must match point 1 in the right image. Point 2 would exceed the disparity gradient limit.

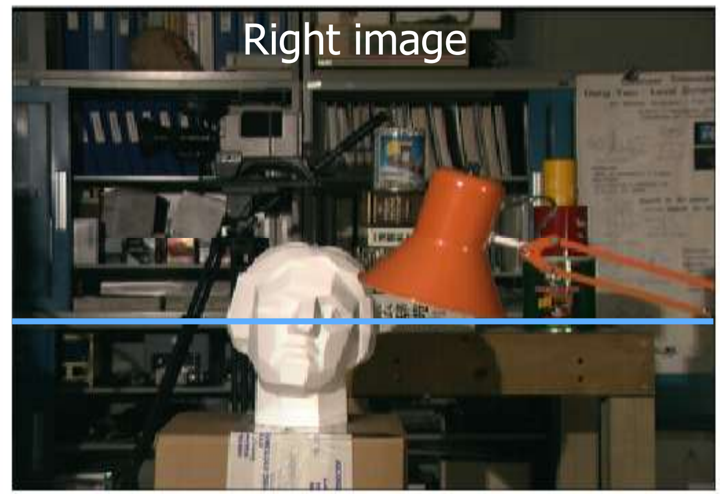
Figure from Gee & Cipolla 1999

Grauman

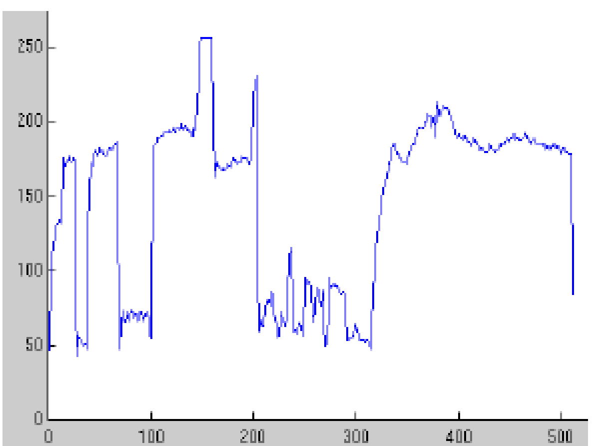
Scanline stereo

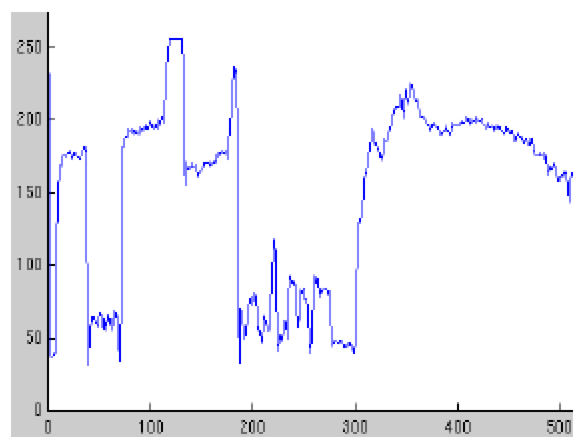
- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently





intensity

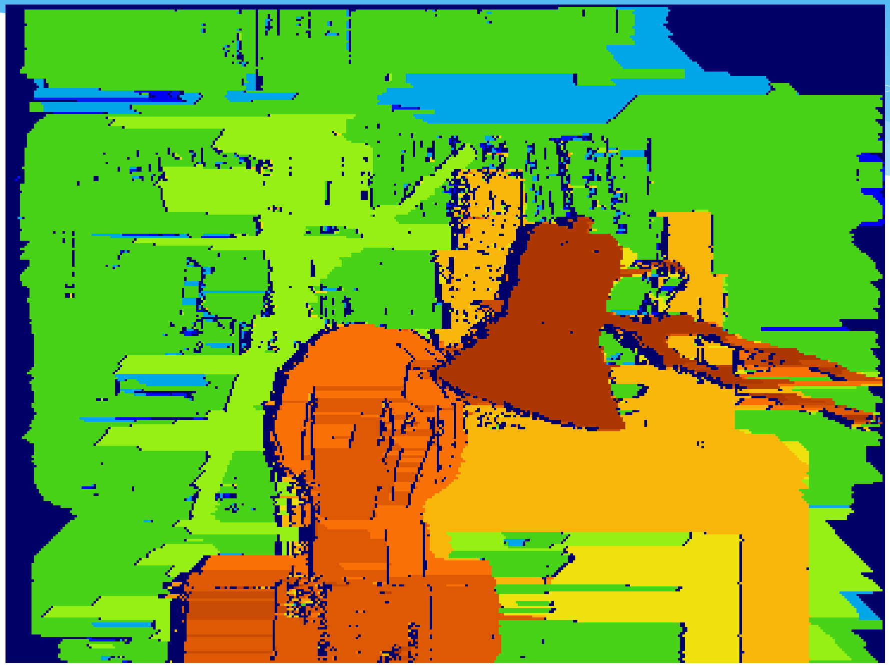




Grauman

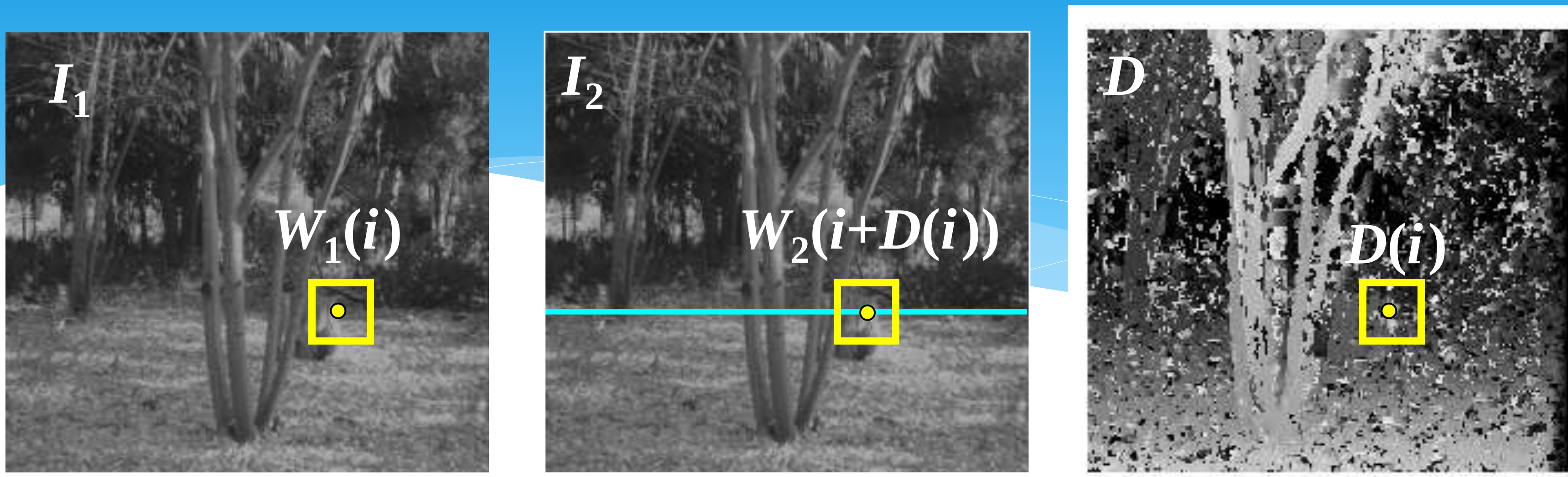
Coherent stereo on 2D grid

- Scanline stereo generates streaking artifacts



- Can’t use dynamic programming to find spatially coherent disparities/ correspondences on a 2D grid

As energy minimization…



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

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

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

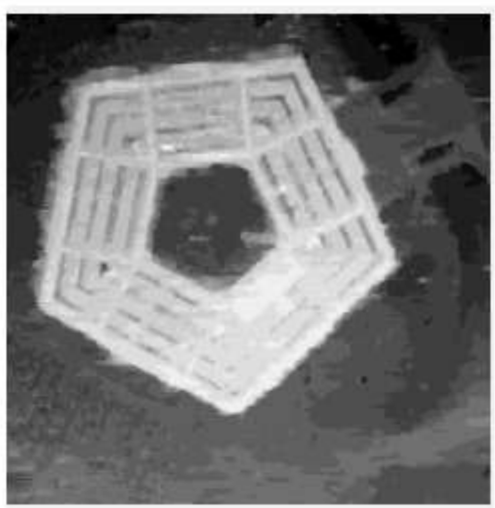
Grauman

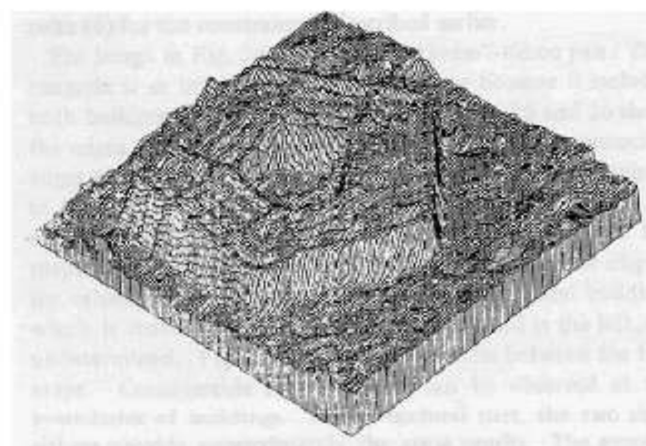
Examples…

left image

right image

range map





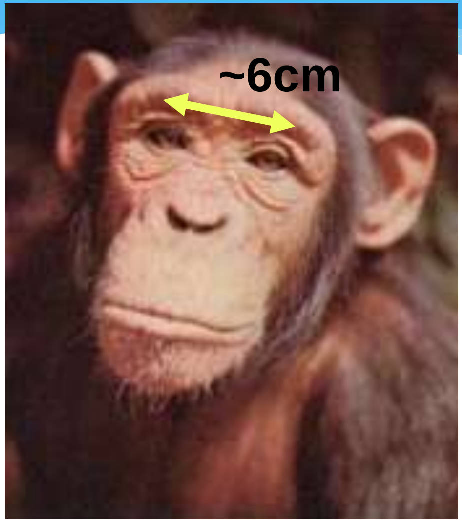
left image

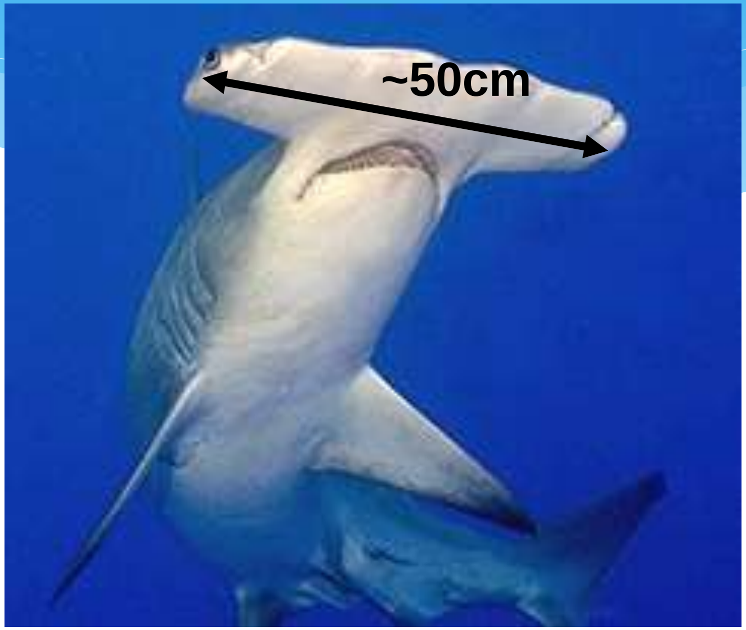
right image

depth map
intensity = depth

Grauman

Stereo vision

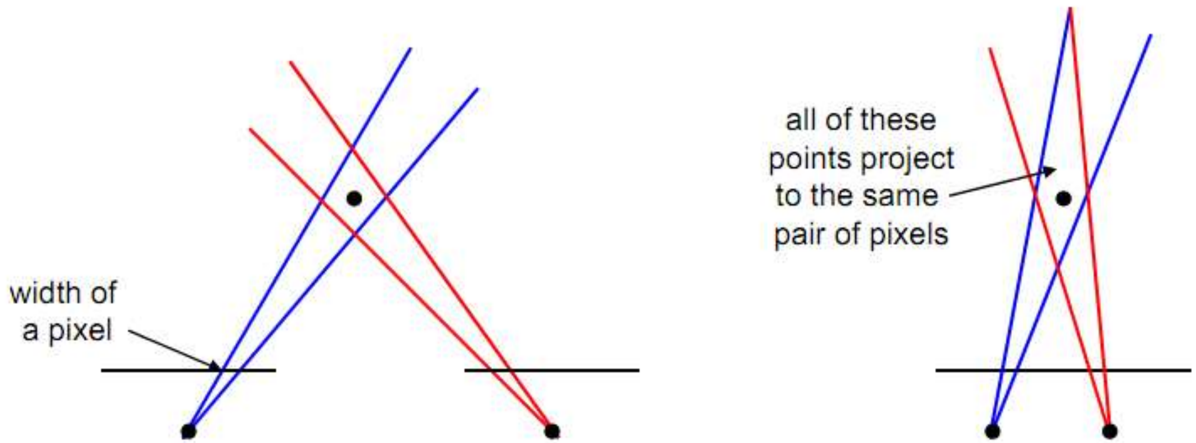




After 30 feet (10 meters) disparity is quite small and depth from stereo is unreliable…

Slide: A. Torralba

Choosing the stereo baseline



Large Baseline **Small Baseline**

What's the optimal baseline?

- Too small: large depth error
- Too large: difficult search problem

[Seitz]

Multibaseline Stereo

- Basic Approach
  - Choose a reference view
  - Use your favorite stereo algorithm BUT
    - replace two-view SSD with SSD over all baselines
- Limitations
  - Must choose a reference view
  - Visibility: select which frames to match [Kang, Szeliski, Chai, CVPR'01]

Active stereo with structured light

Li Zhang's one-shot stereo

- Project "structured" light patterns onto the object
  - simplifies the correspondence problem

Szeliski

http://vision.middlebury.edu/stereo/

vision.middlebury.edu

stereo • mview • MRF • flow • color

Stereo

Evaluation • Datasets • Code • Submit

Daniel Scharstein • Richard Szeliski

Welcome to the Middlebury Stereo Vision Page, formerly located at www.middlebury.edu/stereo. This website accompanies our taxonomy and comparison of two-frame stereo correspondence algorithms [1]. It contains:

- An on-line evaluation of current algorithms
- Many stereo datasets with ground-truth disparities
- Our stereo correspondence software
- An on-line submission script that allows you to evaluate your stereo algorithm in our framework

How to cite the materials on this website:
We grant permission to use and publish all images and numerical results on this website. If you report performance results, we request that you cite our paper [1]. Instructions on how to cite our datasets are listed on the datasets page. If you want to cite this website, please use the URL "**vision.middlebury.edu/stereo/**".

References:

[1] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002. Microsoft Research Technical Report MSR-TR-2001-81, November 2001.

Support for this work was provided in part by NSF CAREER grant 9984485 and NSF grant IIS-0413169. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Problems with Stereo

* Calibration
* Matching is difficult.
  * Deciding on what to match:
    * Pixels vs. features.
  * How to match:
    * Local vs. global.
* Accuracy of depth is limited by the baseline.

Further Reading

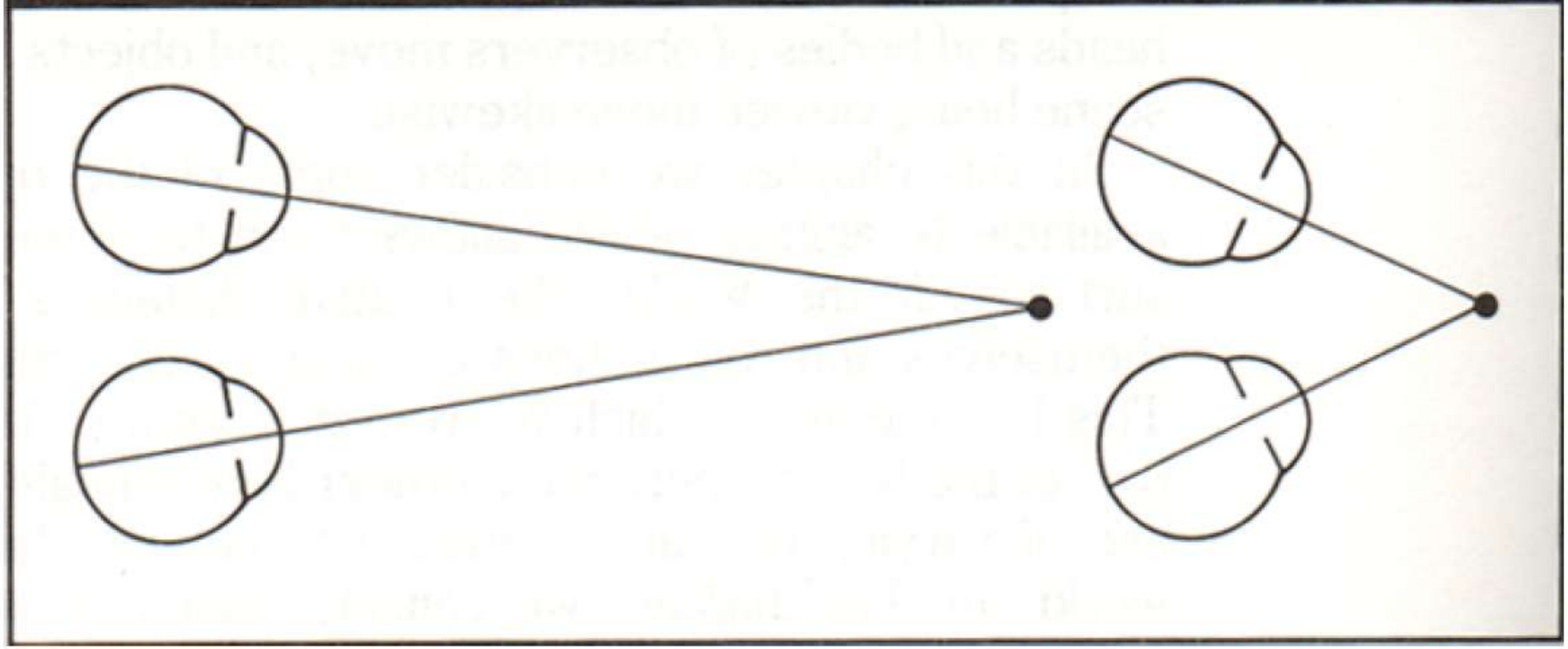
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 8, AUGUST 2003 993

Advances in Computational Stereo

Myron Z. Brown, *Member, IEEE*, Darius Burschka, *Member, IEEE*, and
Gregory D. Hager, *Senior Member, IEEE*

Human Stereo Vision: Fixation, convergence

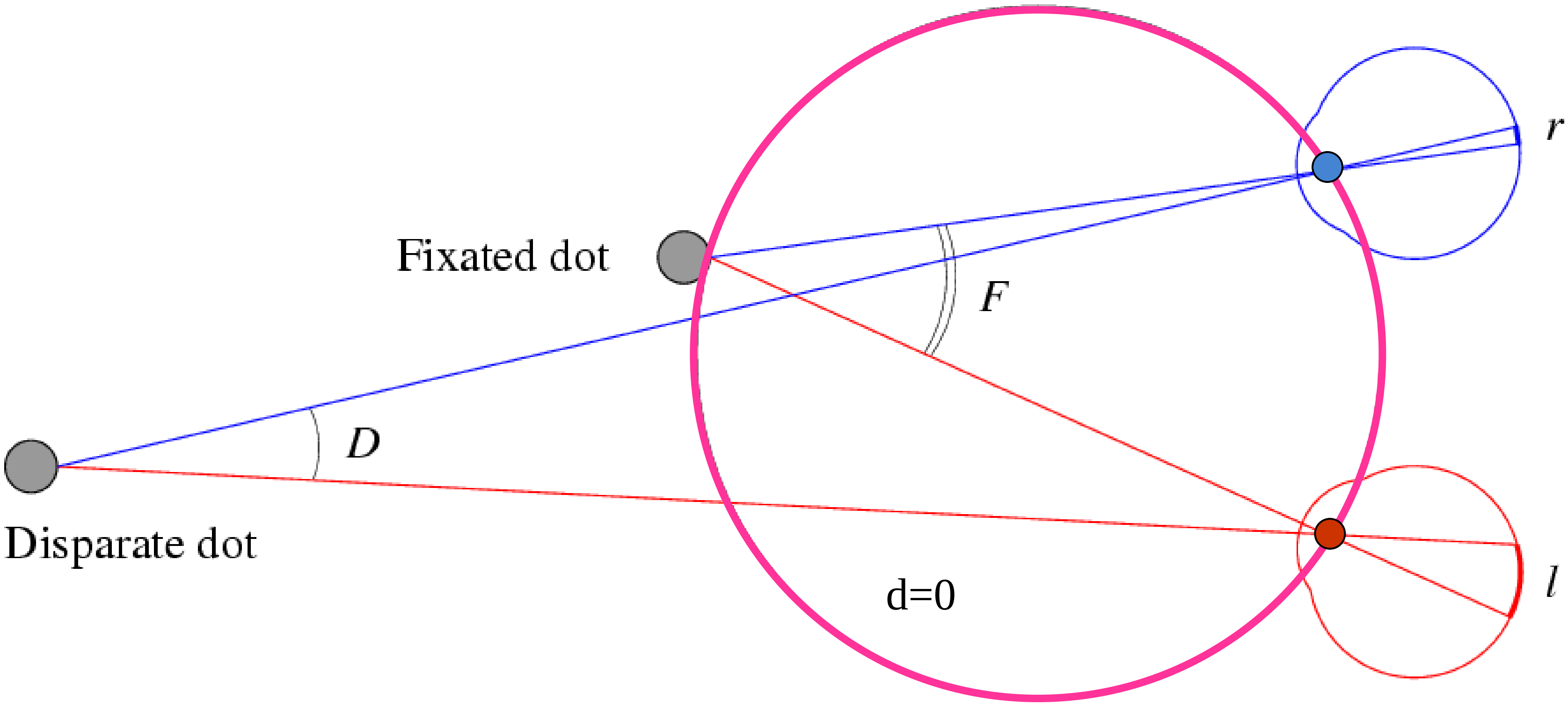
FIGURE 7.1



From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Grauman

Human stereopsis: disparity



Disparity: *d = r-l = D-F.*

Adapted from M. Pollefeys

Do you have stereo vision?

THE FRAMING GAME

In order to see 3D your brain has to use the visual information from both eyes. If the two eye views are too different and cannot be matched up, the brain will be forced to make a choice. It will reject all or part of the information from one eye. The brain can suppress or turn off visual information it cannot use. The Framing Game can tell you whether both your eyes are **TURNED ON** at the same time. The illustration to the left demonstrates what should happen.

- Center your nose over the brown eye below.
- Focus your eyes on the single brown eye.
- Put your free thumb in front of your nose.
- Continue to focus on the eye. If both eyes are on, you will see two thumbs framing one eye.
- Now, switch your focus to your thumb. You should see two eyes framing one thumb.

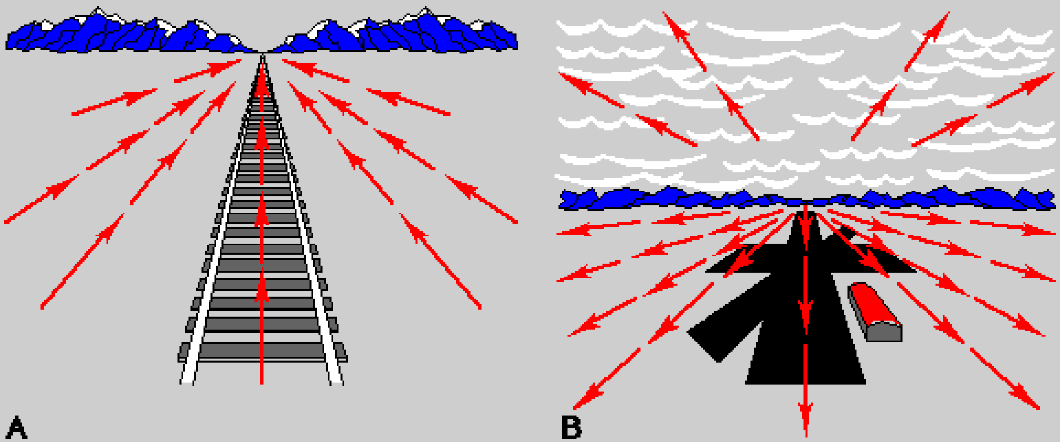
SUCCESSFUL?

Both your eyes are **ON** and you are an excellent candidate for 3D viewing fun. Continue with this guide and enjoy!

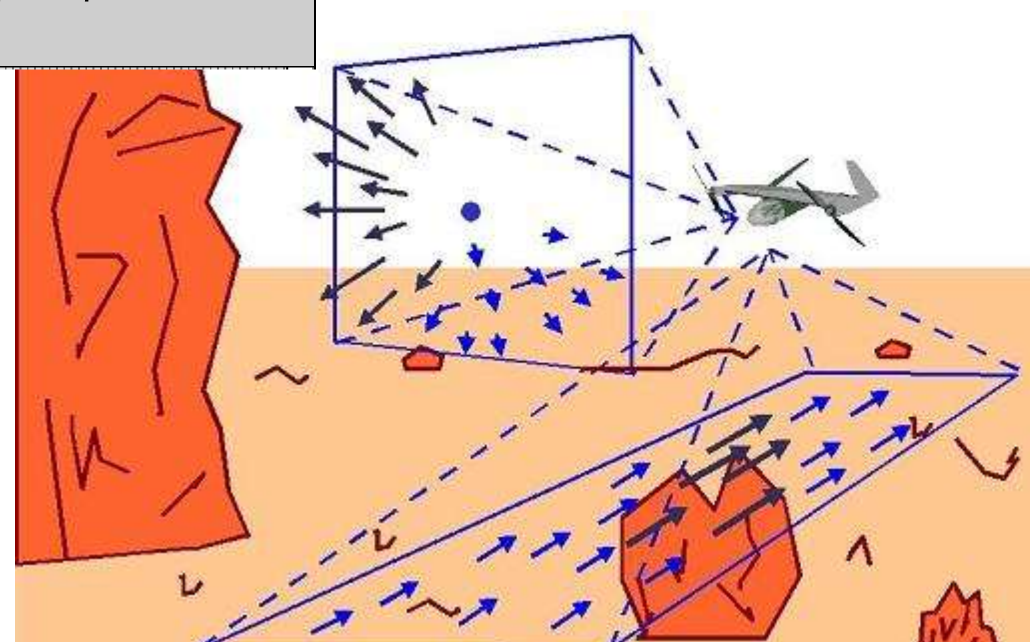
http://www.vision3d.com/frame.html

Binocular Cues: Motion

Depth from optical flow



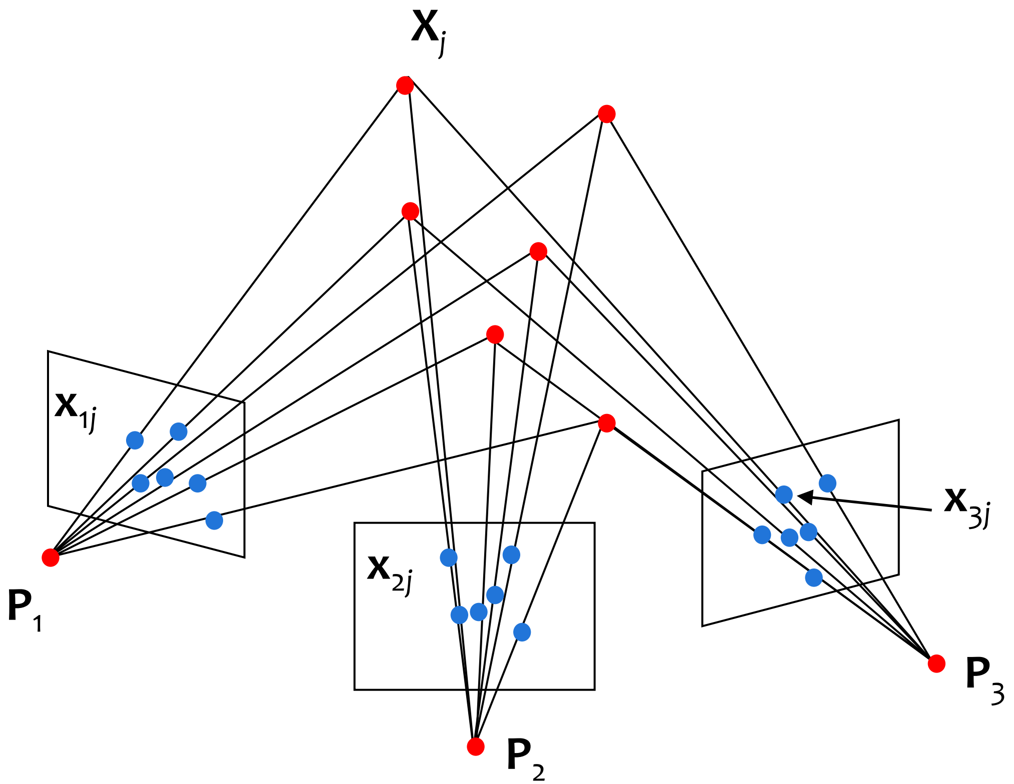
http://cns.bu.edu/vislab/projects/buk/



http://www.pages.drexel.edu/~weg22/opticFlow.html

Structure from motion

- Given: *m* images of *n* fixed 3D points
  - $\mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$
- Problem: estimate *m* projection matrices $\mathbf{P}_i$ and *n* 3D points $\mathbf{X}_j$ from the *mn* correspondences $\mathbf{x}_{ij}$

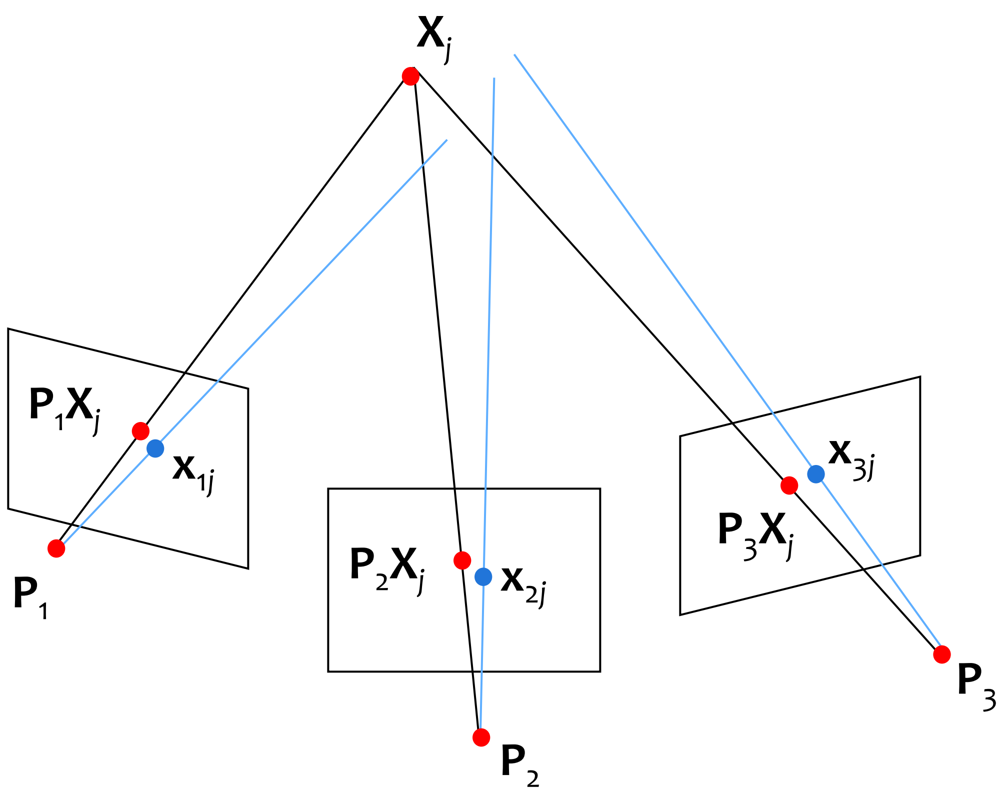


Lazebnik

Bundle adjustment

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

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



Lazebnik

Building Rome in a Day

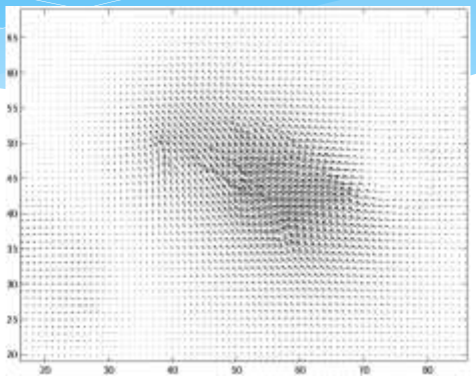
Sameer Agarwal[1,*] Noah Snavely[2] Ian Simon[1] Steven M. Seitz[1] Richard Szeliski[3]

[1]University of Washington [2]Cornell University [3]Microsoft Research

http://grail.cs.washington.edu/rome/

Problems with motion

* Structure from optic flow:
  * Estimation of optic flow is not easy: Flow field is usually over-smooth, noisy and incomplete.
  * Gives a rough estimate only.



* Structure from Motion:
  * Requires too many views/frames
  * Matching is now more difficult due to many views
  * Illumination becomes a bigger problem

Monocular Cues

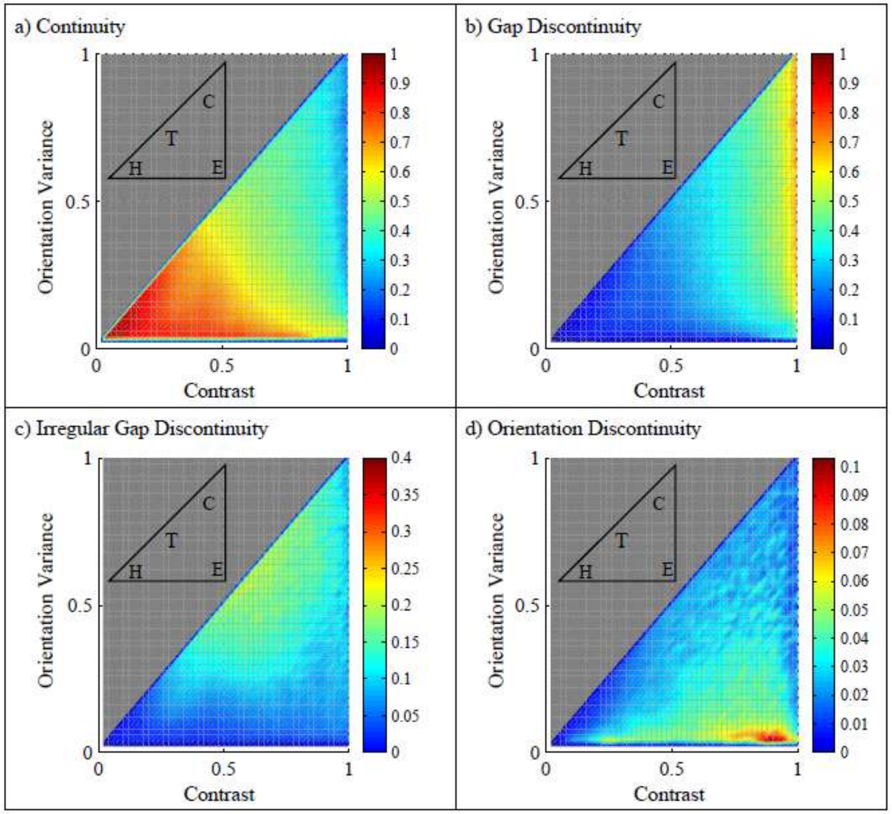
An important fraction of people don’t use stereo vision.

Monocular cues

Figure 7.3: Line drawing of a scene. Picture courtesy of [van Diepen and Graef, 1994].

‘No news is good news’ [W.E.L. Grimson]

- No contrast in 2D means continuity in 3D
- Utilized a lot in surface interpolation & dense stereo methods.
- Quantified & extended in (Kalkan et al., 2006)



Examples for monocular cues

Monocular cues to depth

- **Relative depth cues:**
  - provide relative information about depth between elements in the scene

- **Absolute depth cues:**
  - (assuming known camera parameters) these cues provide information about the absolute depth between the observer and elements of the scene

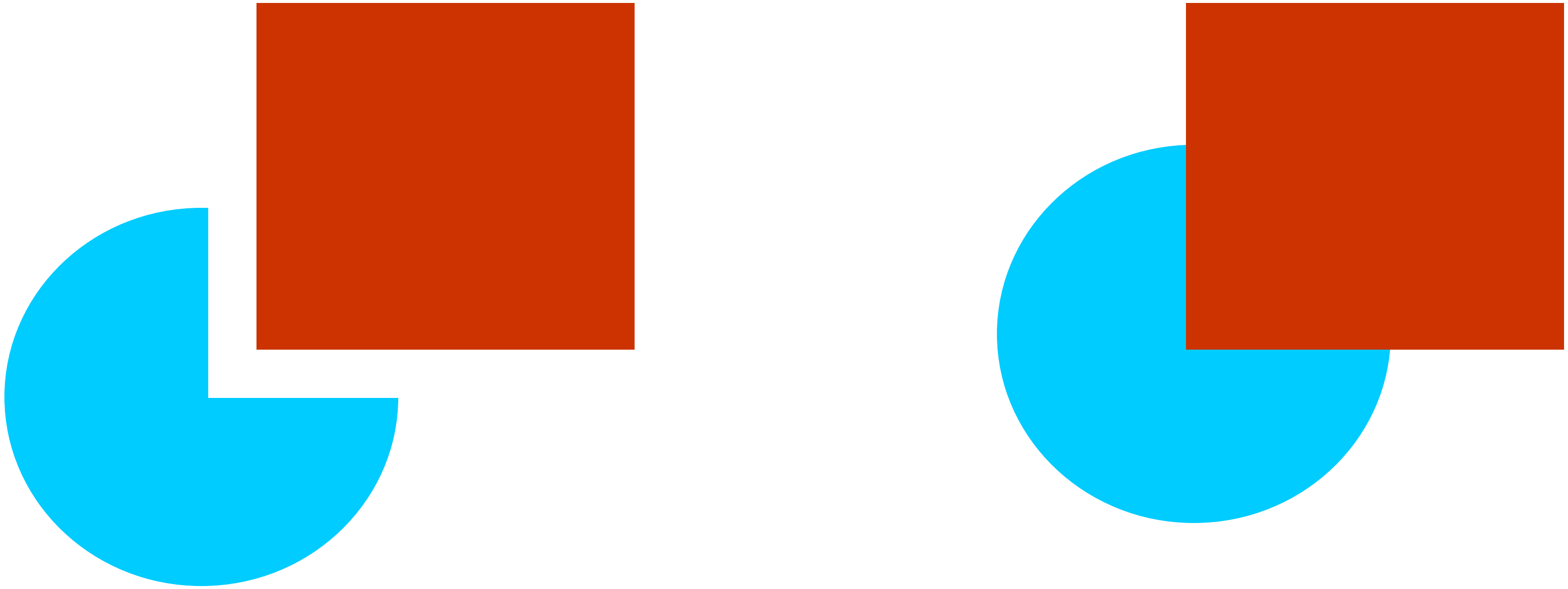
Slide: A. Torralba

Relative depth cues

Simple and powerful cue, but hard to make it work in practice…

Slide: A. Torralba

Interposition / occlusion



Slide: A. Torralba

Texture Gradient

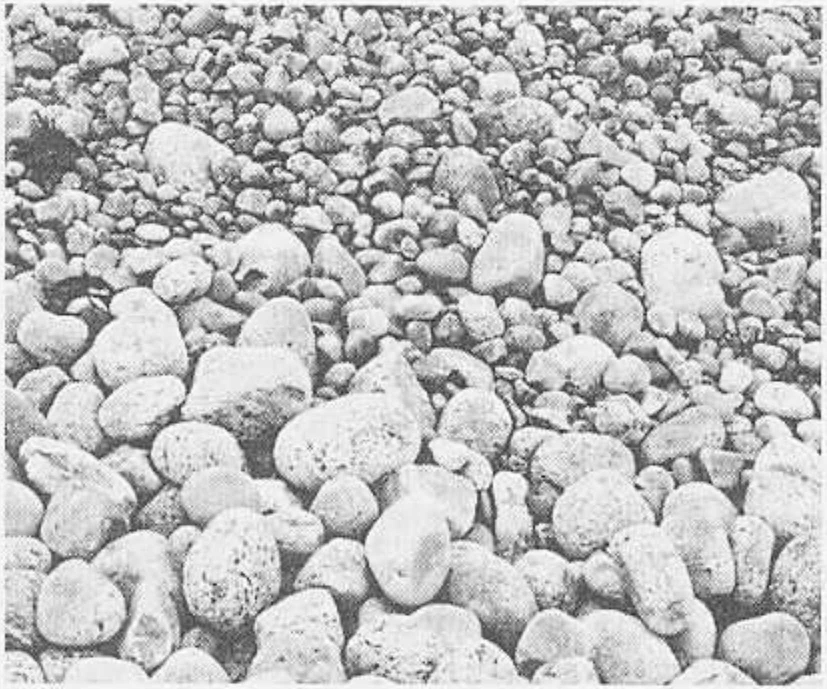


FIGURE 8.27
Texture gradients provide information about depth. (Frank Siteman/Stock, Boston.)
© Frank Sitman/Stock Boston

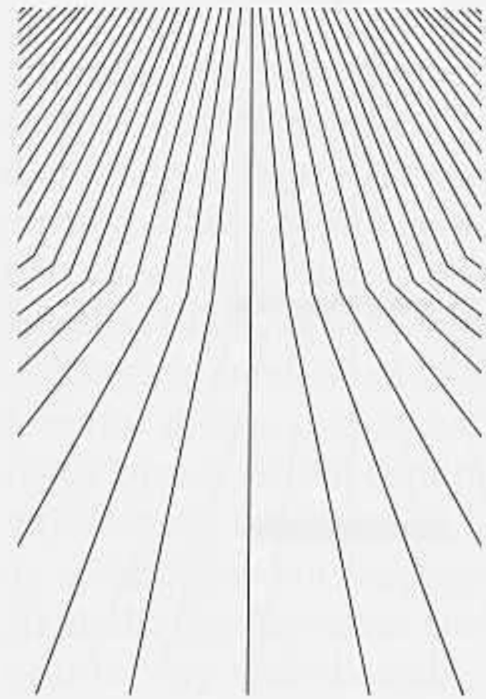


FIGURE 8.28
Texture discontinuity signals the pre
corner.

A Witkin**.** Recovering Surface Shape and Orientation from Texture (1981)

Slide: A. Torralba

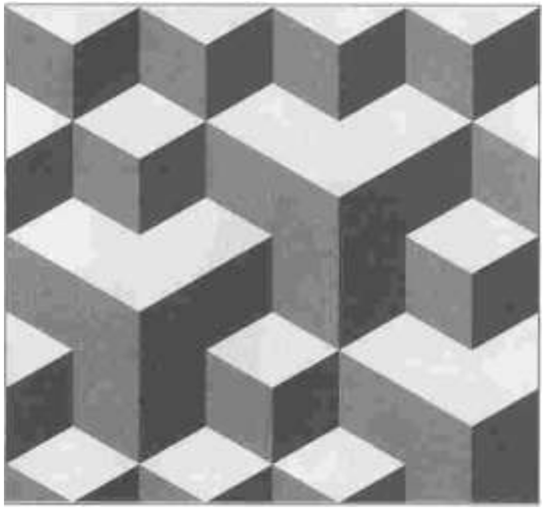
Illumination

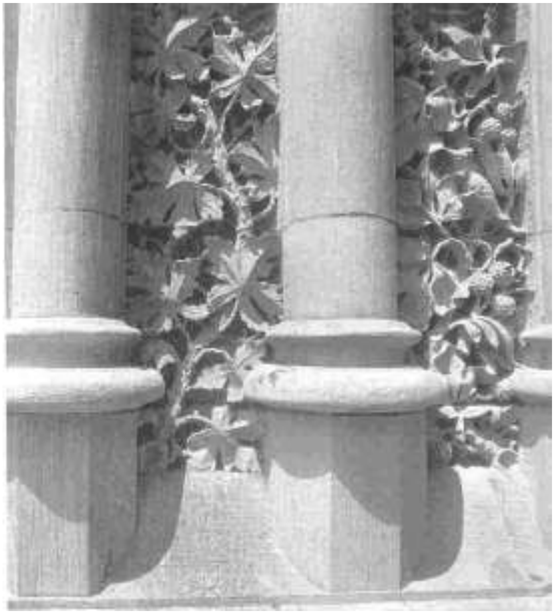
* Shading
* Shadows
* Inter-reflections

Slide: A. Torralba

Shading

* Based on 3 dimensional modeling of objects in light, shade and shadows.

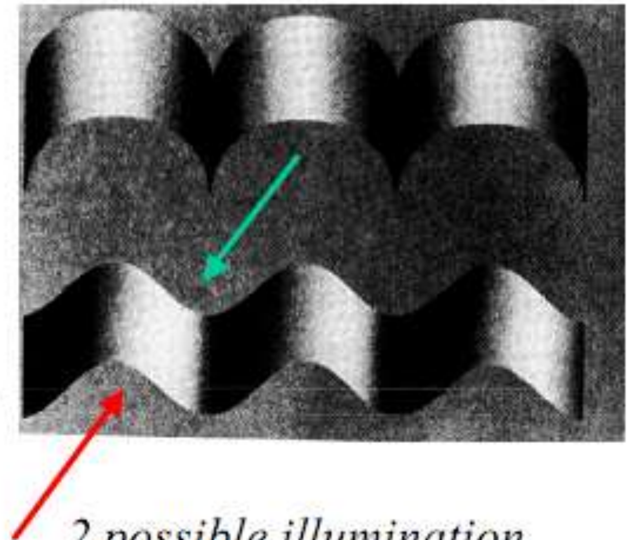




Source: A. Torralba

Does Shading Play a Central Role?

- Contour plays a more important role
  - Variations in intensity are same on both shapes
  - Upper region is perceived as composed of three cylindrical pieces illuminated from above
  - Lower region is perceived as sinusoidal, illuminated from one side
    - Note the ambiguities of the surface perceptions, depending on assumed illumination direction

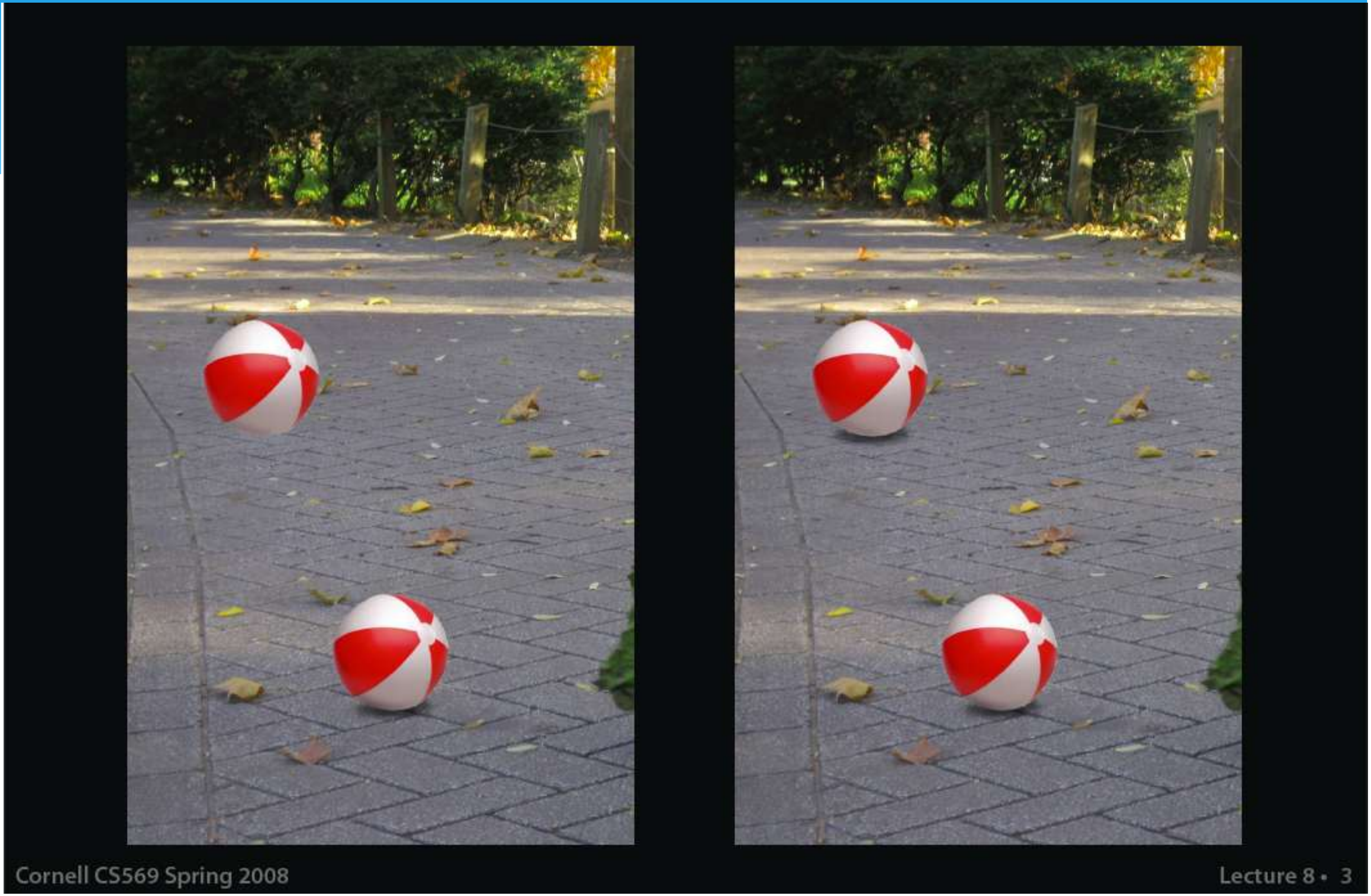


2 possible illumination hypotheses

5

[Larry Davis](), [Ramani Duraiswami](), [Daniel DeMenthon](), and [Cornelia Fermüller]()

Shadows



Slide by Steve Marschner

http://www.cs.cornell.edu/courses/cs569/2008sp/schedule.stm

Atmospheric perspective

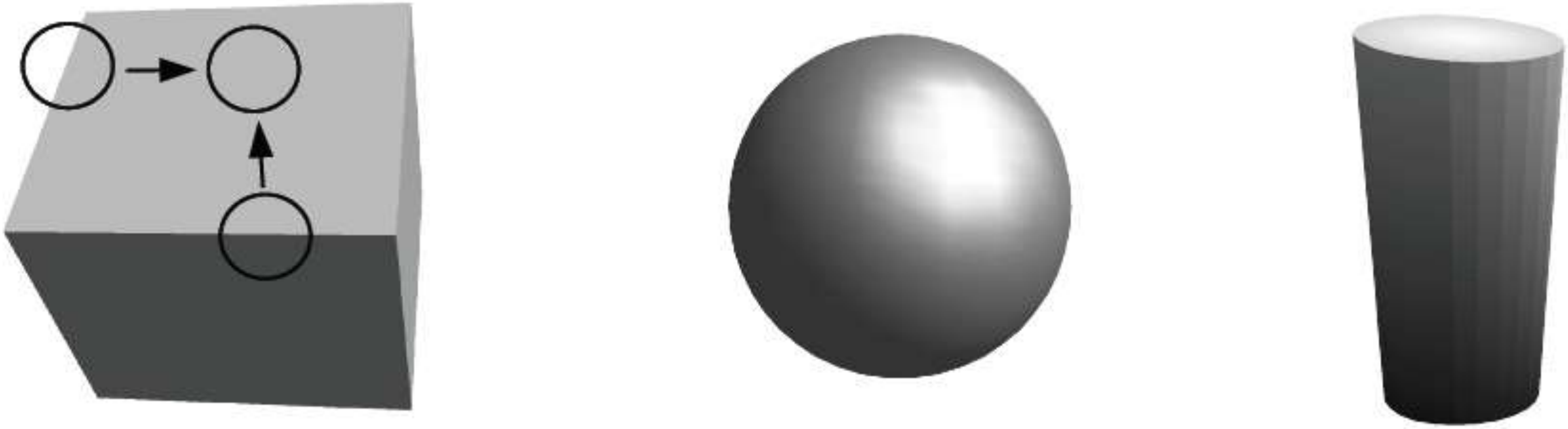
Far objects:

* Bluish
* Lower contrast

Predicting Depth from Existing Depth

* Combination of different depth cues.

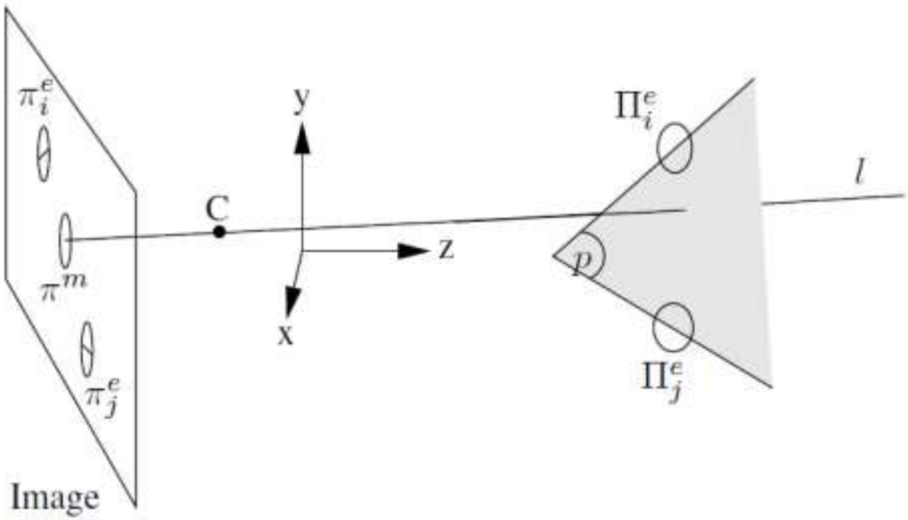
Depth Prediction from Edges

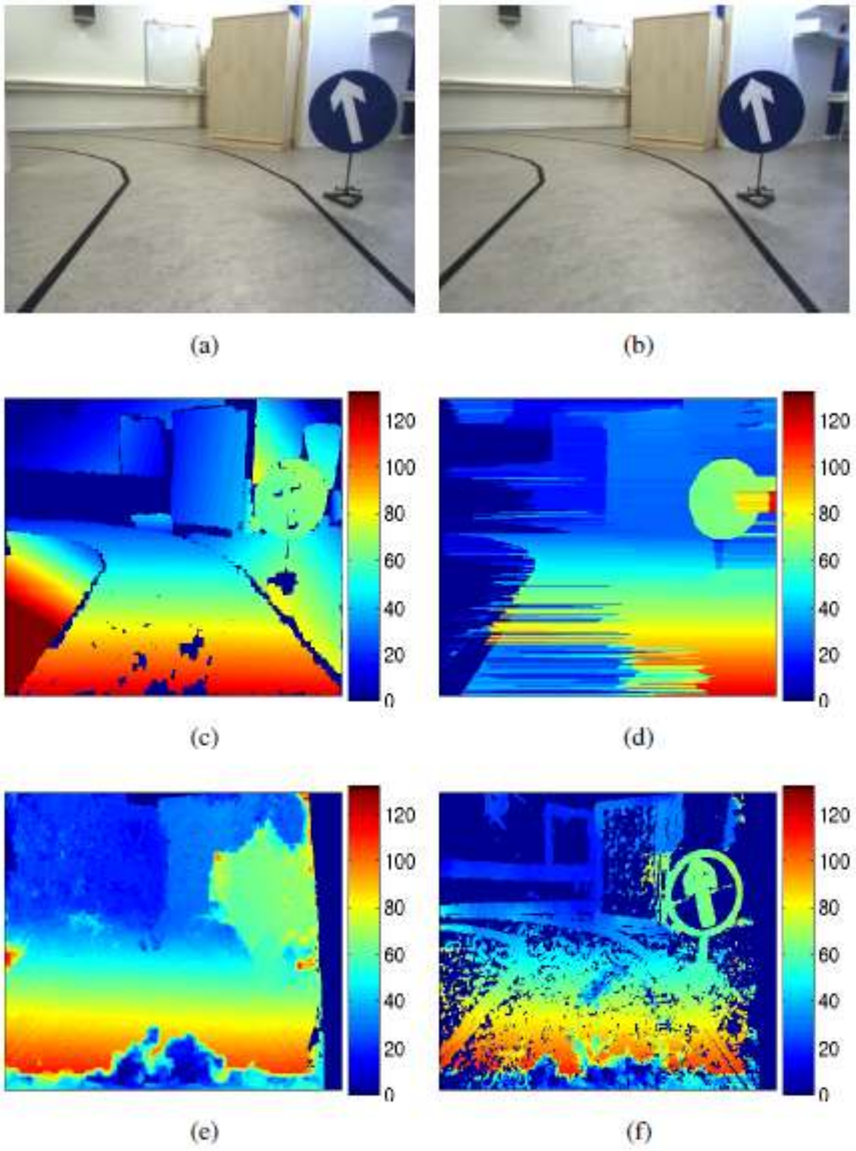


(a) (b) (c)

Kalkan et al., 2008.

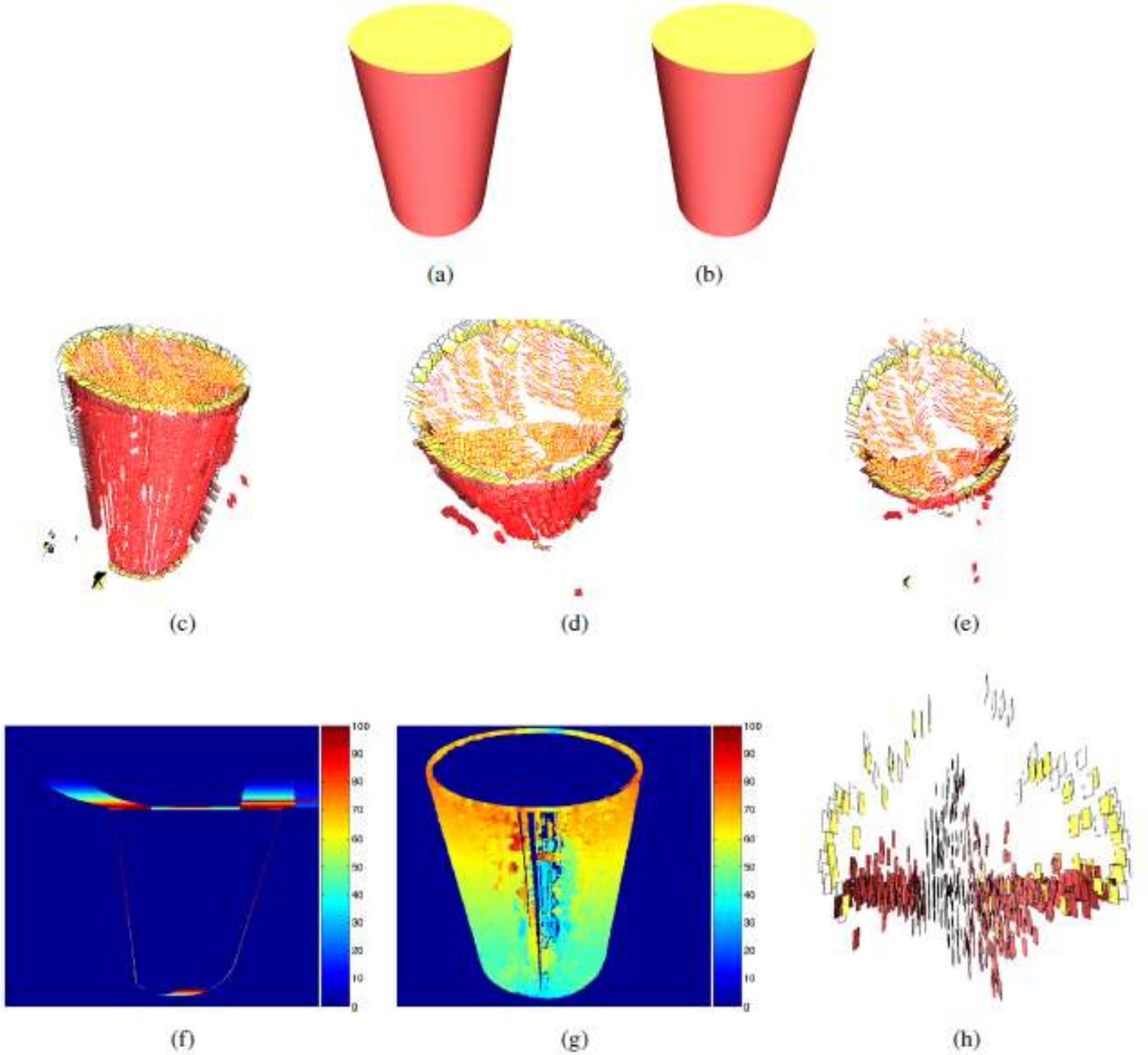
Depth Prediction from Edges





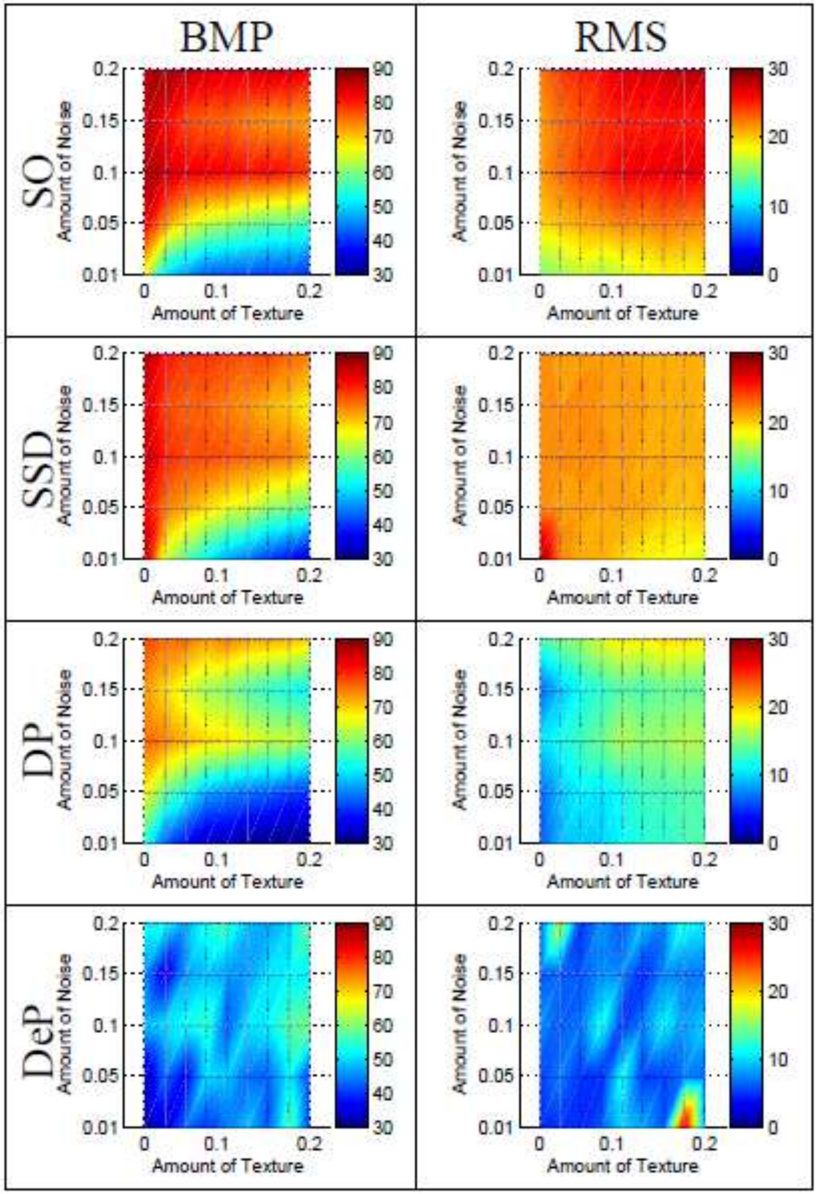
Kalkan et al., 2008.

Depth Prediction from Edges



Kalkan et al., 2008.

Depth Prediction from Edges



Kalkan et al., 2008.

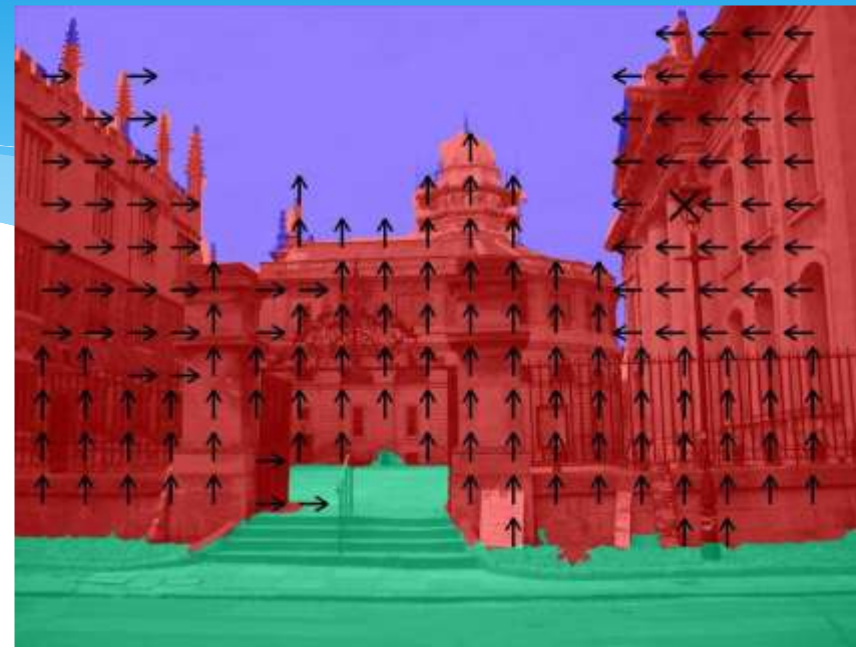
Learning Monocular Cues from Labeled Data

Learn to Estimate Surface Orientations

- **Learn structure of the world from labeled examples**

Slides by Efros

Label Geometric Classes



- **Goal:** learn labeling of image into 7 Geometric Classes:
- **Support (ground)**
- **Vertical**
  - Planar: facing **Left** (←), **Center** (↑), **Right** (→)
  - Non-planar: **Solid** (**X**), **Porous** or wiry (**O**)
- **Sky**

Slides by Efros

What cues to use?

Vanishing points, lines

Color, texture, image location

Texture gradient

Slides by Efros

The General Case (outdoors)

* Typical outdoor photograph off the Web
  * Got 300 images using Google Image Search keyboards: “outdoor”, “scenery”, “urban”, etc.
* Certainly not random samples from world
  * 100% horizontal horizon
  * 97% pixels belong to 3 classes -- ground, sky, vertical (gravity)
  * Camera axis usually parallel to ground plane
* Still very general dataset!

Slides by Efros

Let's use many weak cues

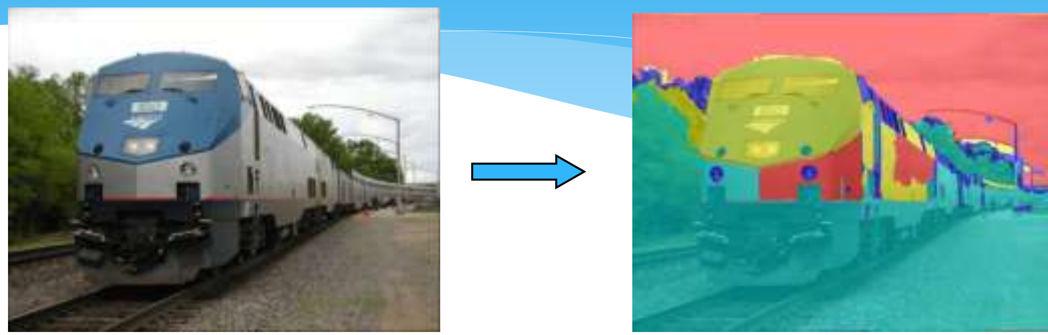
* Material
* Image Location
* Perspective

SURFACE CUES
Location and Shape
L1. Location: normalized x and y, mean
L2. Location: norm. x and y, 10^{th} and 90^{th} pctl
L3. Location: norm. y wrt estimated horizon, 10^{th}, 90^{th} pctl
L4. Location: whether segment is above, below, or straddles estimated horizon
L5. Shape: number of superpixels in segment
L6. Shape: normalized area in image
Color
C1. RGB values: mean
C2. HSV values: C1 in HSV space
C3. Hue: histogram (5 bins)
C4. Saturation: histogram (3 bins)
Texture
T1. LM filters: mean abs response (15 filters)
T2. LM filters: hist. of maximum responses (15 bins)
Perspective
P1. Long Lines: (num line pixels)/sqrt(area)
P2. Long Lines: % of nearly parallel pairs of lines
P3. Line Intersections: hist. over 8 orientations, entropy
P4. Line Intersections: % right of center
P5. Line Intersections: % above center
P6. Line Intersections: % far from center at 8 orientations
P7. Line Intersections: % very far from center at 8 orientations
P8. Vanishing Points: (num line pixels with vertical VP membership)/sqrt(area)
P9. Vanishing Points: (num line pixels with horizontal VP membership)/sqrt(area)
P10. Vanishing Points: percent of total line pixels with vertical VP membership
P11. Vanishing Points: x-pos of horizontal VP - segment center (0 if none)
P12. Vanishing Points: y-pos of highest/lowest vertical VP wrt segment center
P13. Vanishing Points: segment bounds wrt horizontal VP
P14. Gradient: x, y center of gradient mag. wrt. image center

Slides by Efros

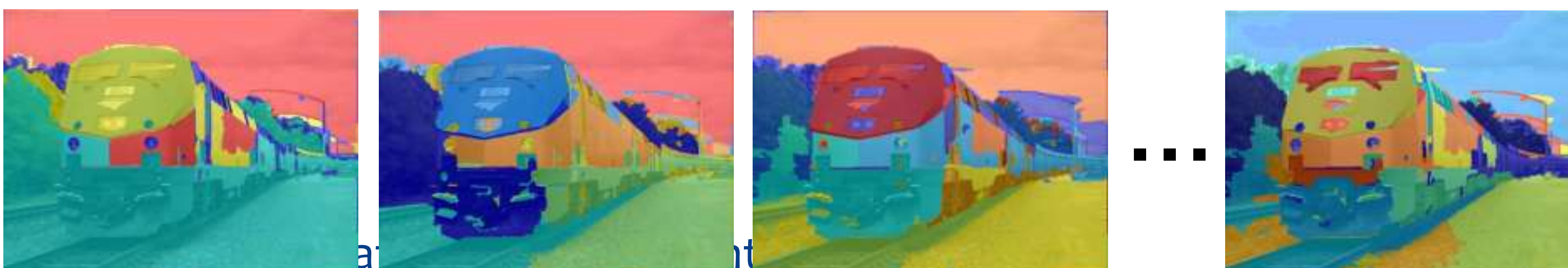
Image Segmentation

* Naïve Idea #1: segment the image



  * Chicken & Egg problem

* Naïve Idea #2: multiple segmentations



Slides by Efros

Estimating surfaces from segments

- We want to know:
  - Is this a good (coherent) segment?

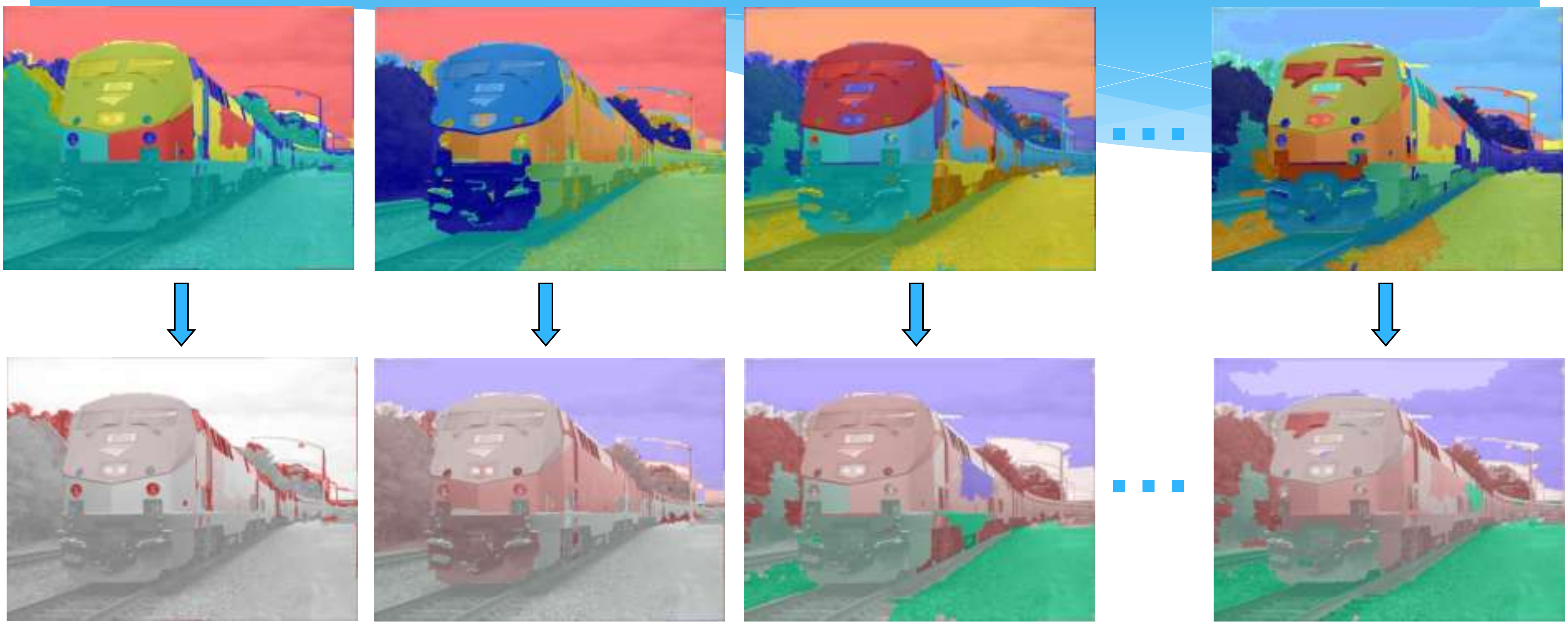
    P(*good segment* | *data*)
  - If so, what is the surface label?

    P(*label* | *good segment*, *data*)

- *Learn* these likelihoods from training images

Slides by Efros

Labeling Segments



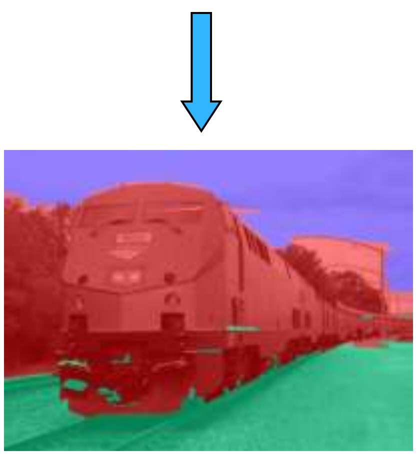
For each segment:

- Get P(*good segment* | *data*) P(*label* | *good segment*, *data*)

Slides by Efros

Image Labeling

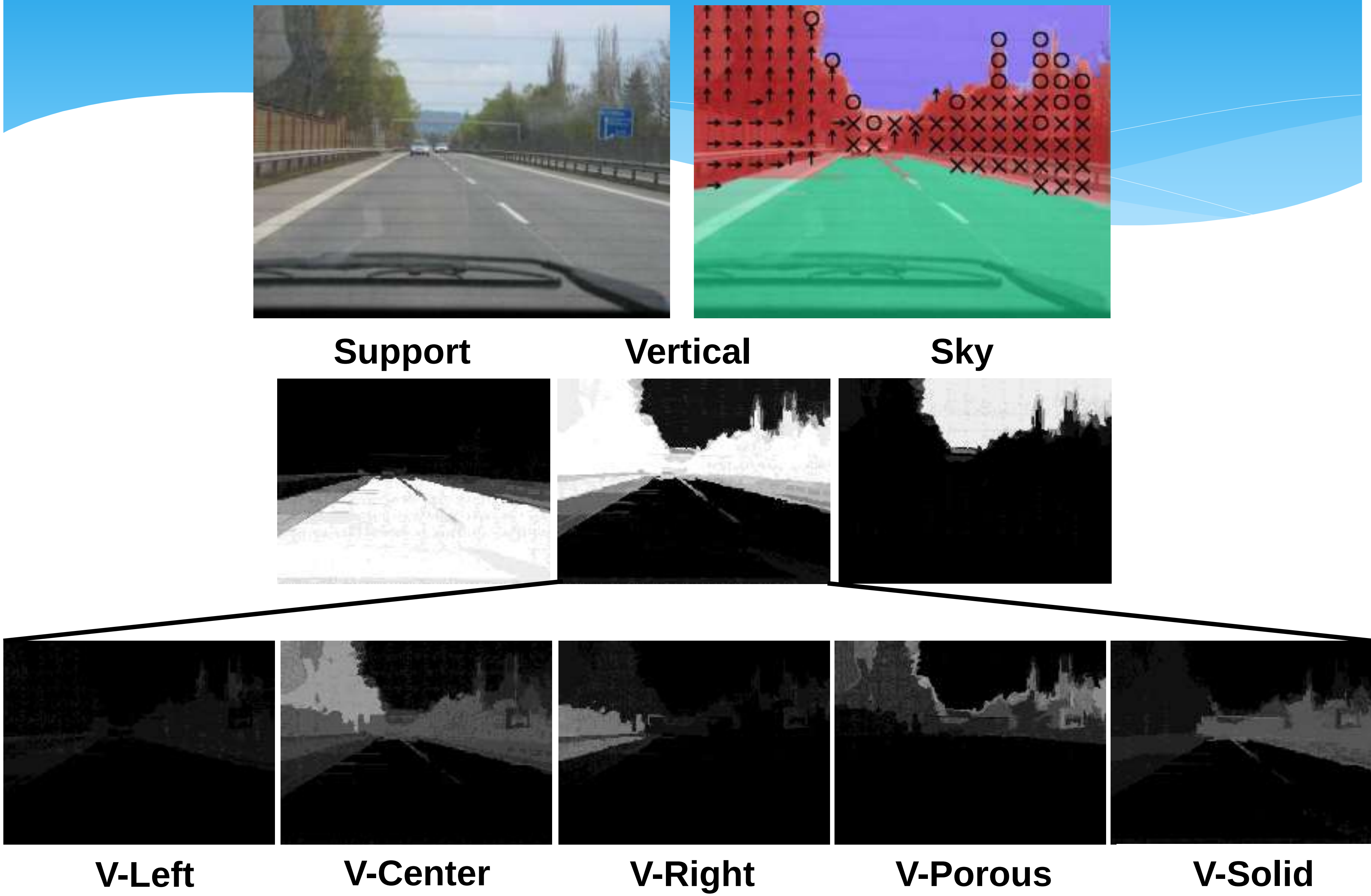
Labeled Segmentations



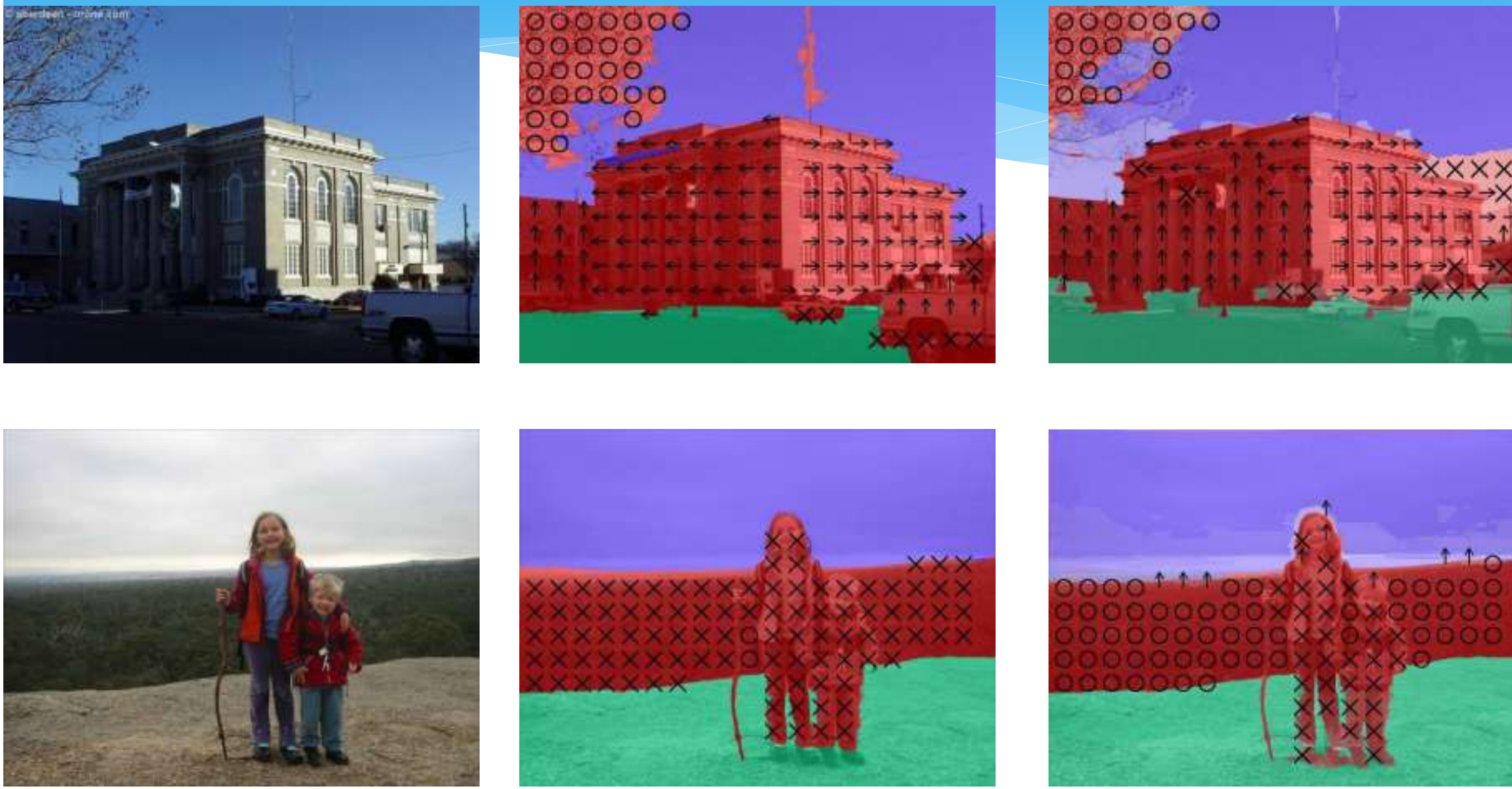
Labeled Pixels

Slides by Efros

No Hard Decisions



Labeling Results



Input image **Ground Truth** **Our Result**

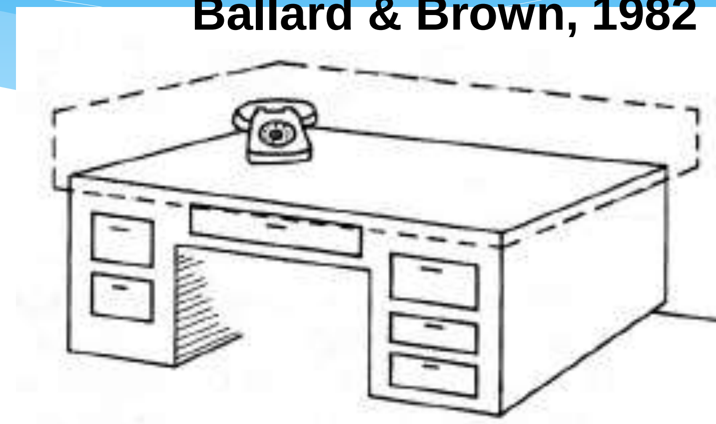
Slides by Efros

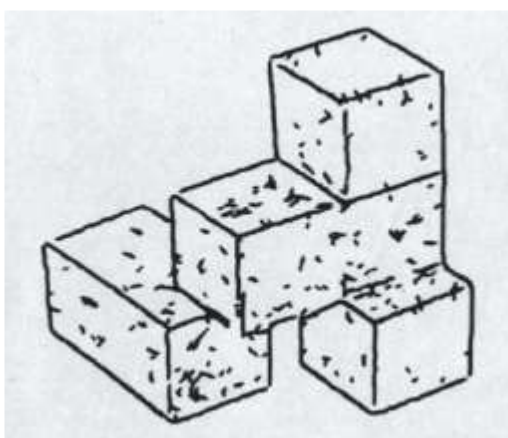
Reasoning about spatial relationships between objects

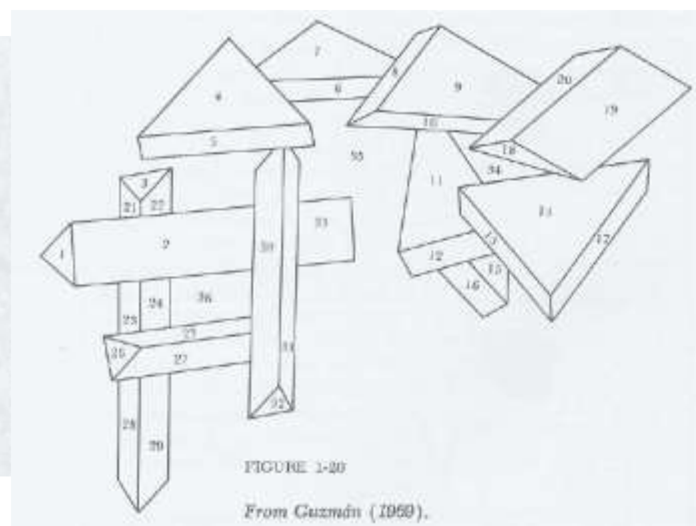
1. LEFT OF
2. RIGHT OF
3. BESIDE (alongside, next to)
4. ABOVE (over, higher than, on top of)
5. BELOW (under, underneath, lower than)
6. BEHIND (in back of)
7. IN FRONT OF
8. NEAR (close to, next to?)
9. FAR
10. TOUCHING
11. BETWEEN
12. INSIDE (within)
13. OUTSIDE

Freeman, 1974

Ballard & Brown, 1982



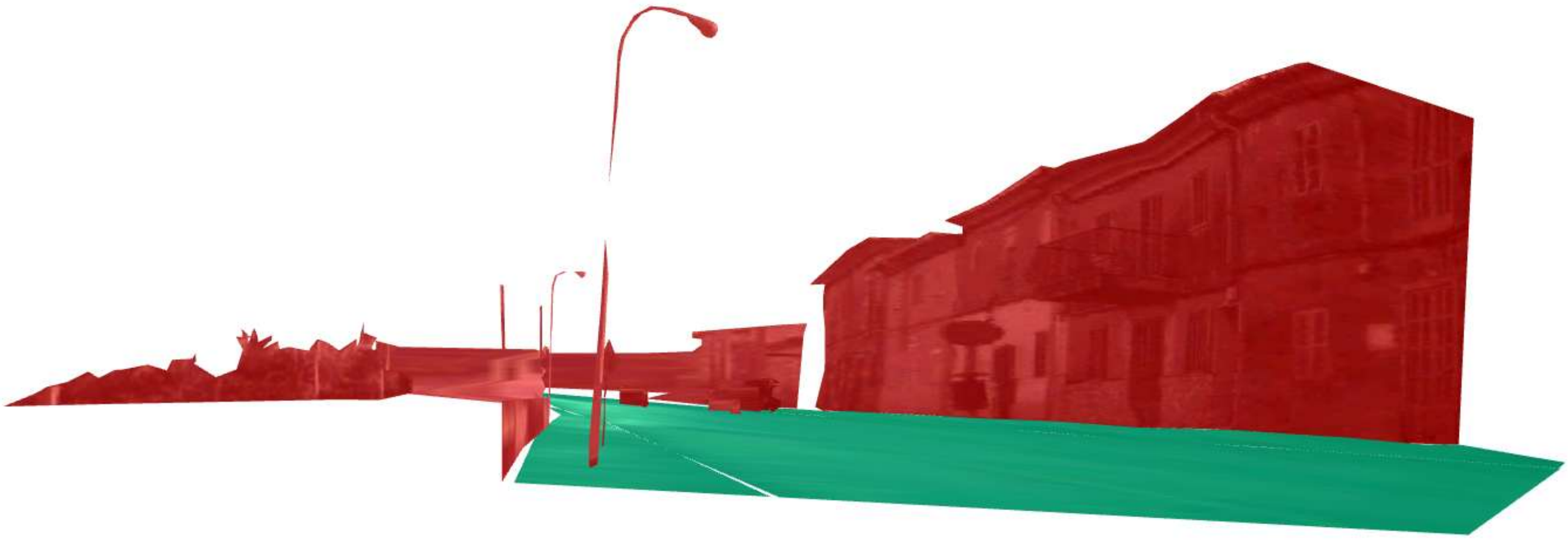




Guzman, 1969

A. Torralba

Scene layout assumptions



A. Torralba

Recovering scene geometry

* Polygon types
  * Ground
  * Standing
  * Attached
* Edge types
  * Contact
  * Attached
  * Occluded
* Camera parameters

A. Torralba

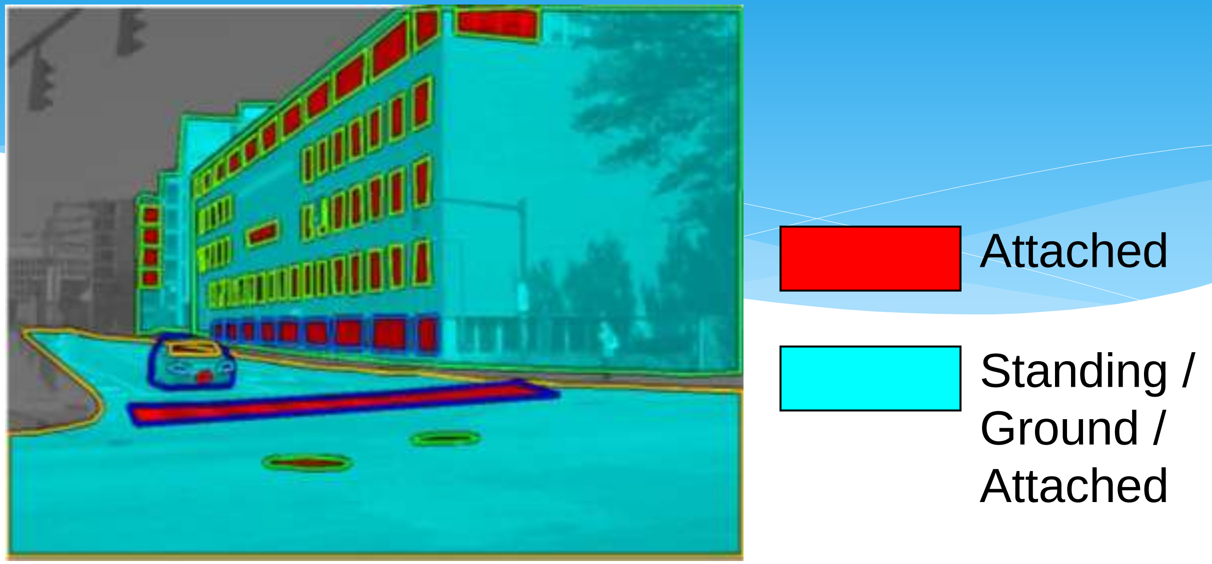
Recovering scene geometry

* Polygon types
  * Ground
  * Standing
  * Attached
* Edge types
  * Contact
  * Attached
  * Occluded
* Camera parameters

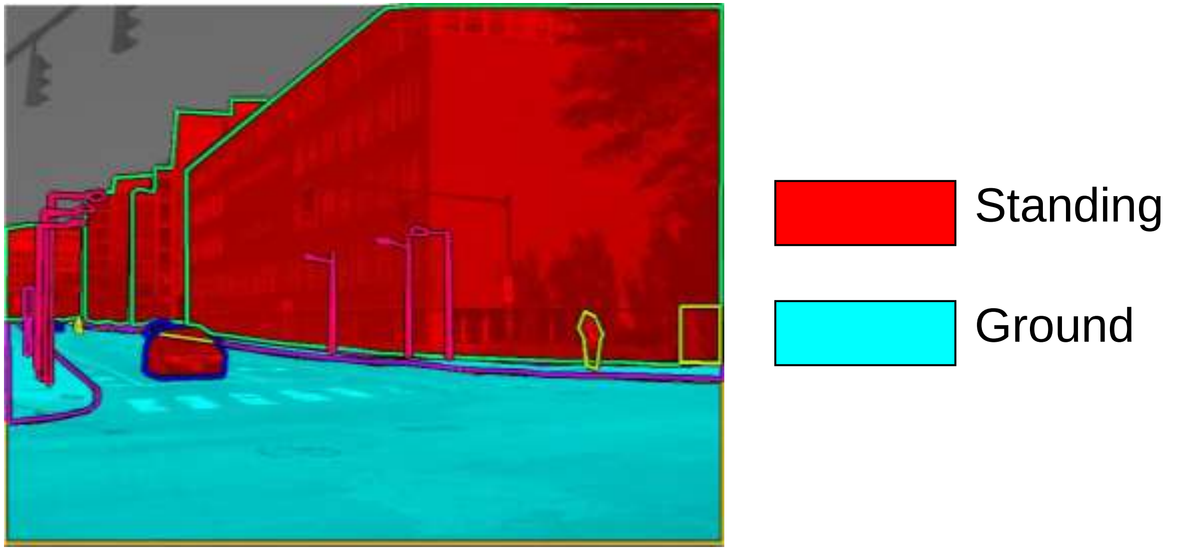
A. Torralba

Relationships between polygons

Part-of



Supported-by



A. Torralba

Recovering scene geometry

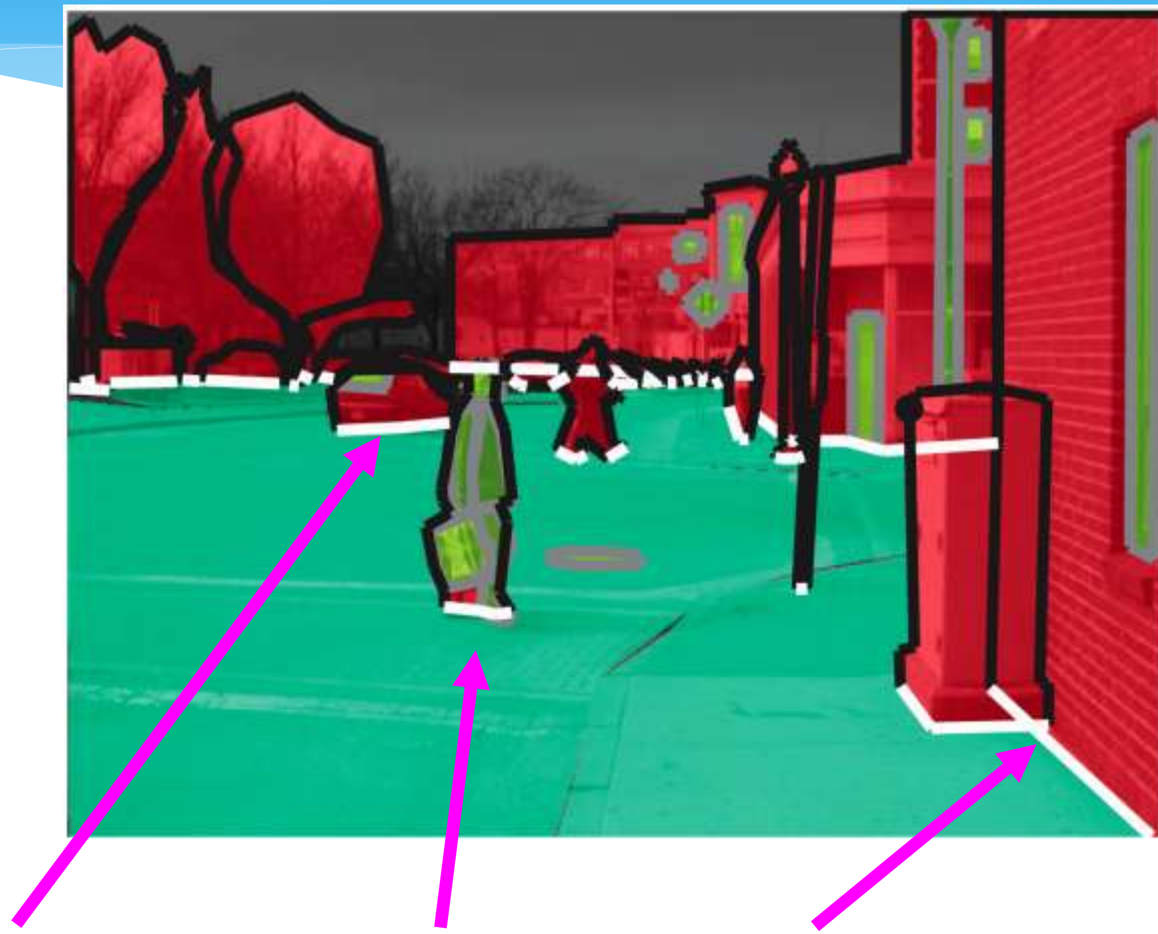
* Polygon types
  * Ground
  * Standing
  * Attached
* Edge types
  * Contact
  * Attached
  * Occluded
* Camera parameters

A. Torralba

Edge types

Ground and attached objects have attached edges

Standing objects can have contact or occluding edges



Cues for contact edges:

Orientation

Proximity to ground

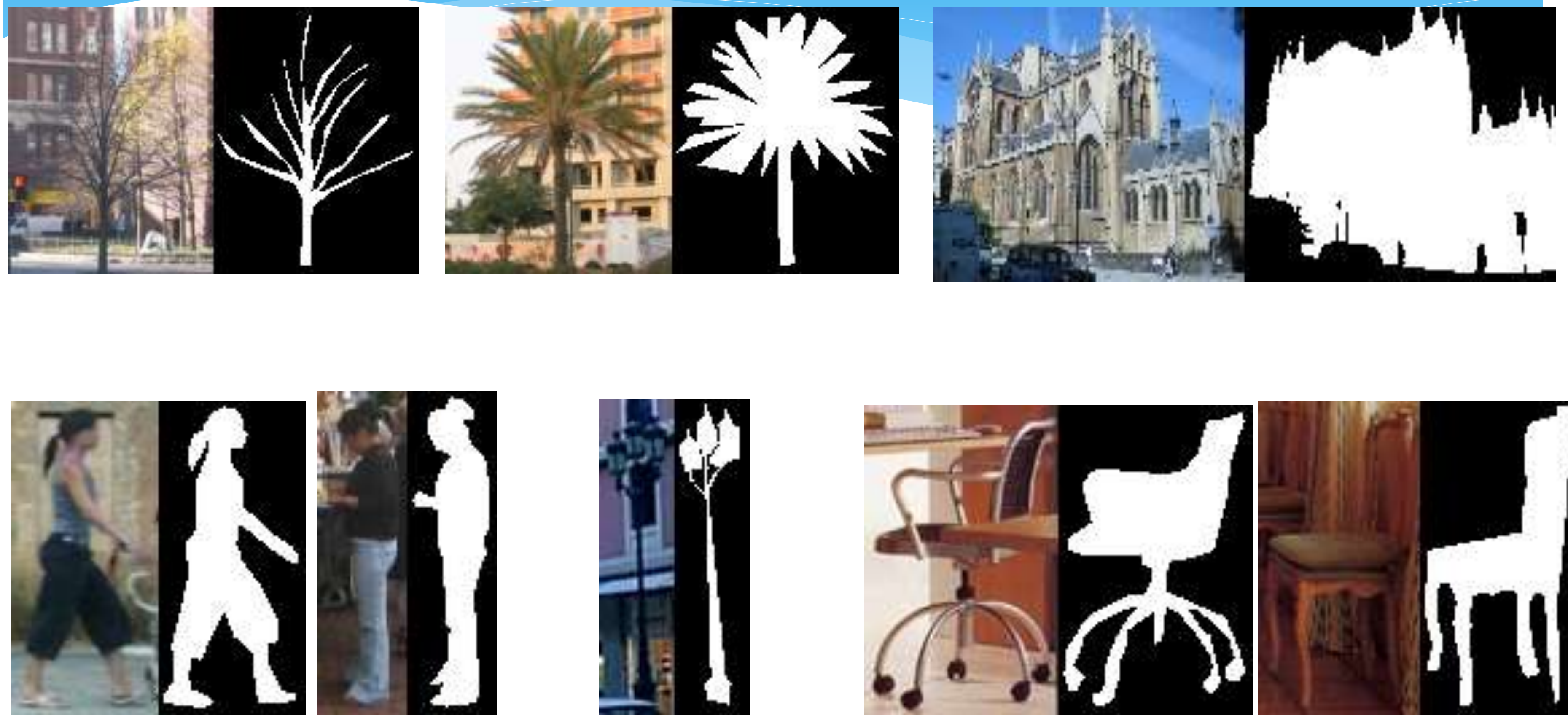
Length

A. Torralba


LabelMe
Please contact us if you find any bugs or have any suggestions.
Show me another image
Label as many objects and regions as you can in this image
Edit/delete object
window
Done
Delete
Sign in (why?)
With your help, there are 91348 labelled objects in the database (more stats)
Instructions (Get more help)
Use your mouse to click around the boundary of some objects in this image. You will then be asked to enter the name of the object (examples: car, window).
Good
Bad
Labeling tools
Erase segment
Zoom
Fit Image
Polygons in this image (XML)
door
door
road
stair
window
window
sidewalk
building region
house
window
window
window

A. Torralba

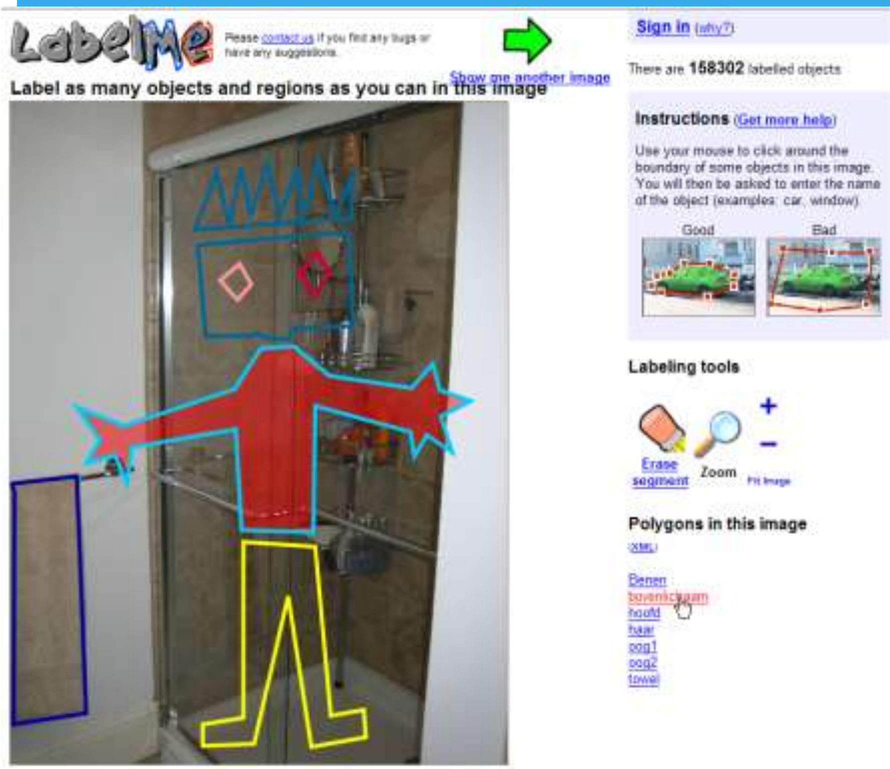
Polygon quality



A. Torralba

Online Hooligans

Do not try this at home



A. Torralba

Absolute (monocular) depth cues

Are there any monocular cues that can give us absolute depth from a single image?

Source: A. Torralba

Familiar size

**Which "object" is closer to the camera?
How close?**

Source: A. Torralba

Familiar size

* Apparent reduction in size of objects at a greater distance from the observer

* Size perspective is thought to be conditional, requiring knowledge of the objects.

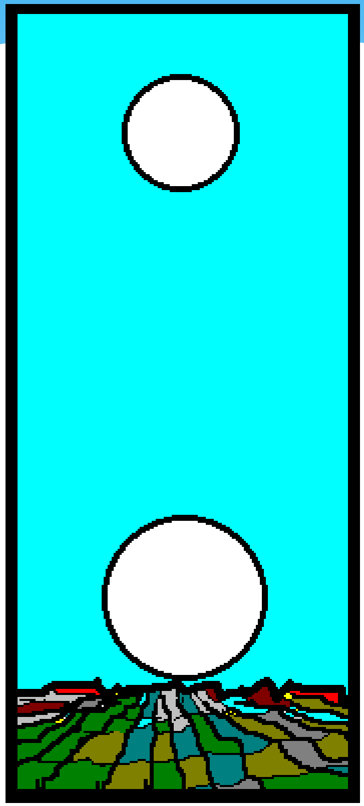
Source: A. Torralba

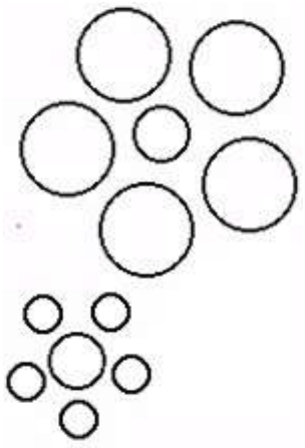
Distance from the horizon line

- Based on the tendency of objects to appear nearer the horizon line with greater distance to the horizon.
- Objects approach the horizon line with greater distance from the viewer.

Source: A. Torralba

Moon illusion





Ebbinghaus illusion

http://en.wikipedia.org/wiki/Moon_illusion

Adapted from: A. Torralba

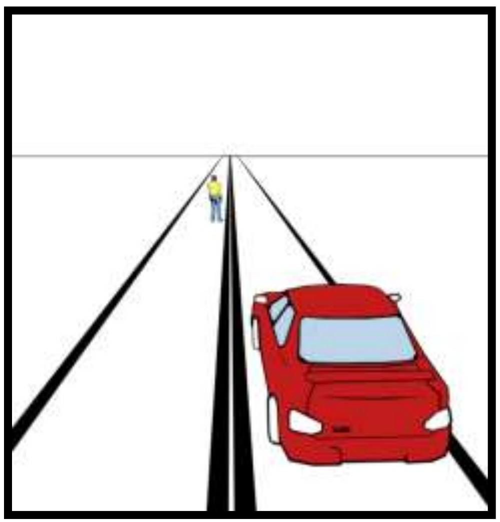
Relative height

* The object closer to the horizon is perceived as farther away, and the object further from the horizon is perceived as closer

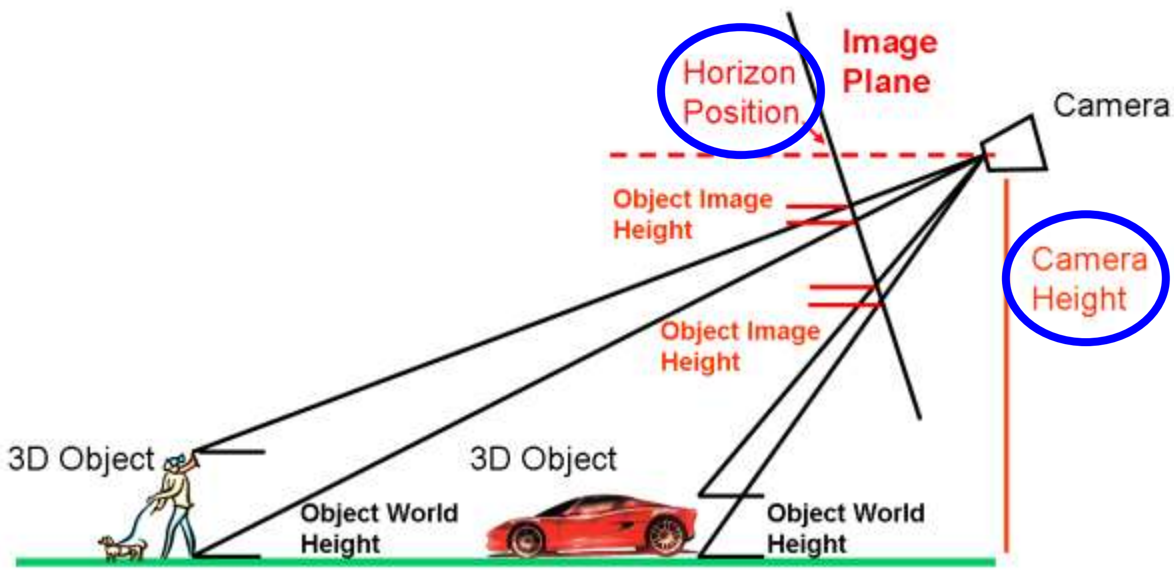
* If you know camera parameters: height of the camera, then we know real depth

Source: A. Torralba

Object Size in the Image



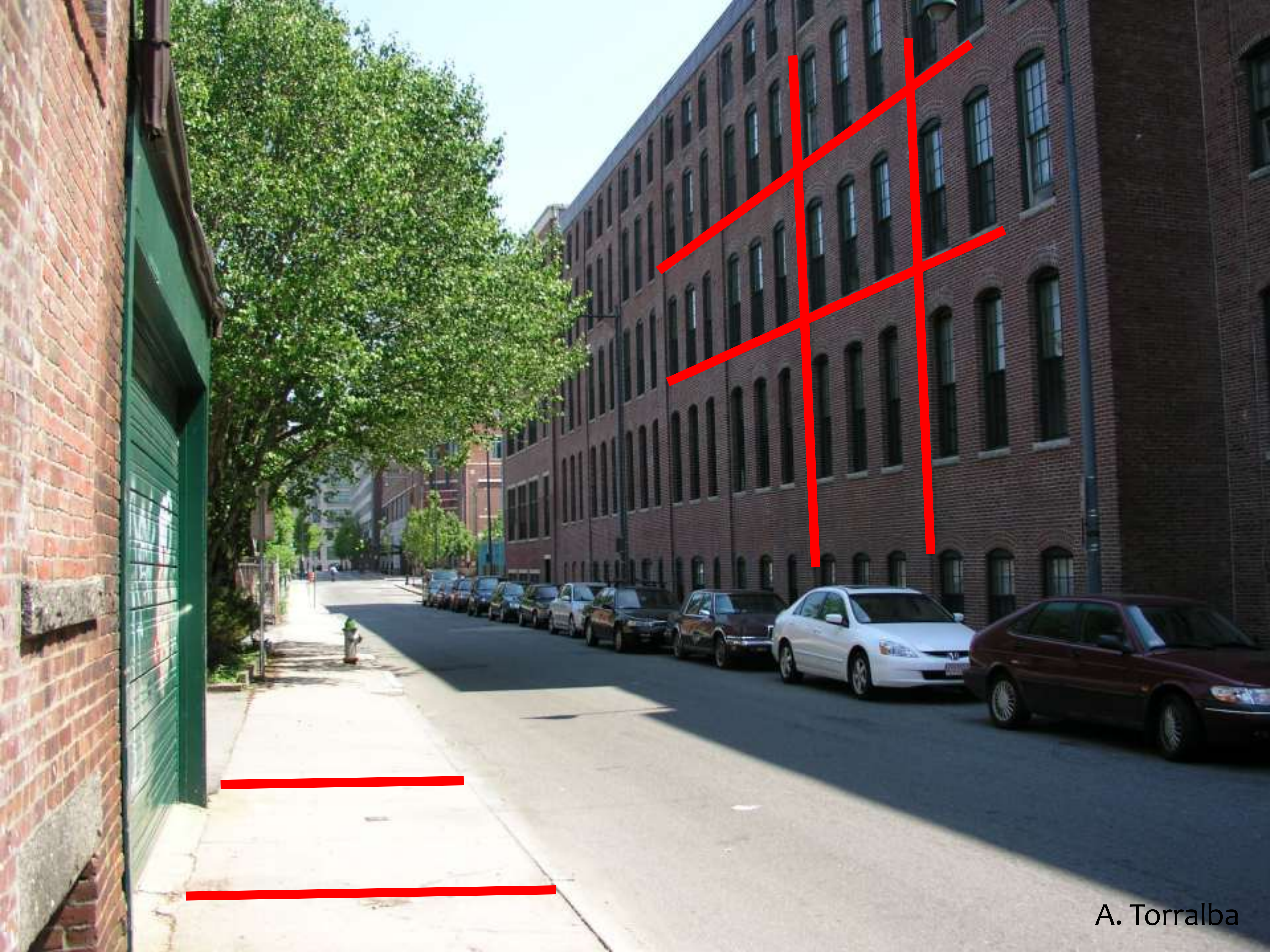
Image



World

Slide by Derek Hoiem

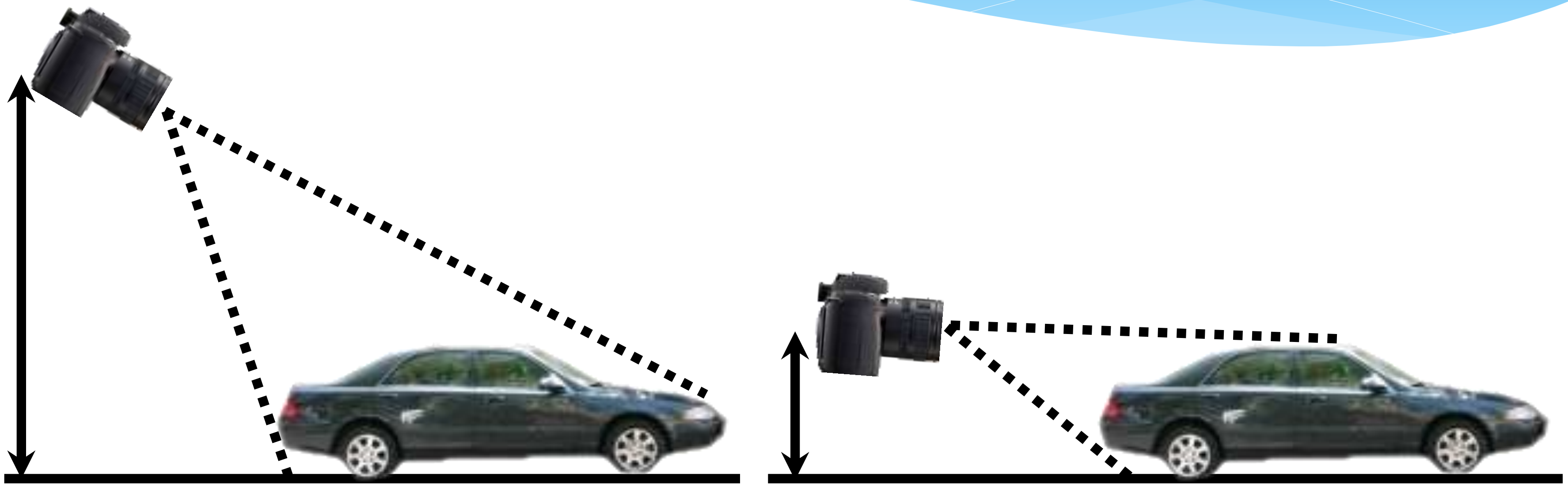
A. Torralba



A. Torralba


A. Torralba

Camera parameters



- Assume
  - flat ground plane
  - camera roll is negligible (consider pitch only)
- Camera parameters: height and orientation

Slide from J-F Lalonde

Camera parameters

v

$$\frac{t-b}{X} = \frac{v-b}{C}$$

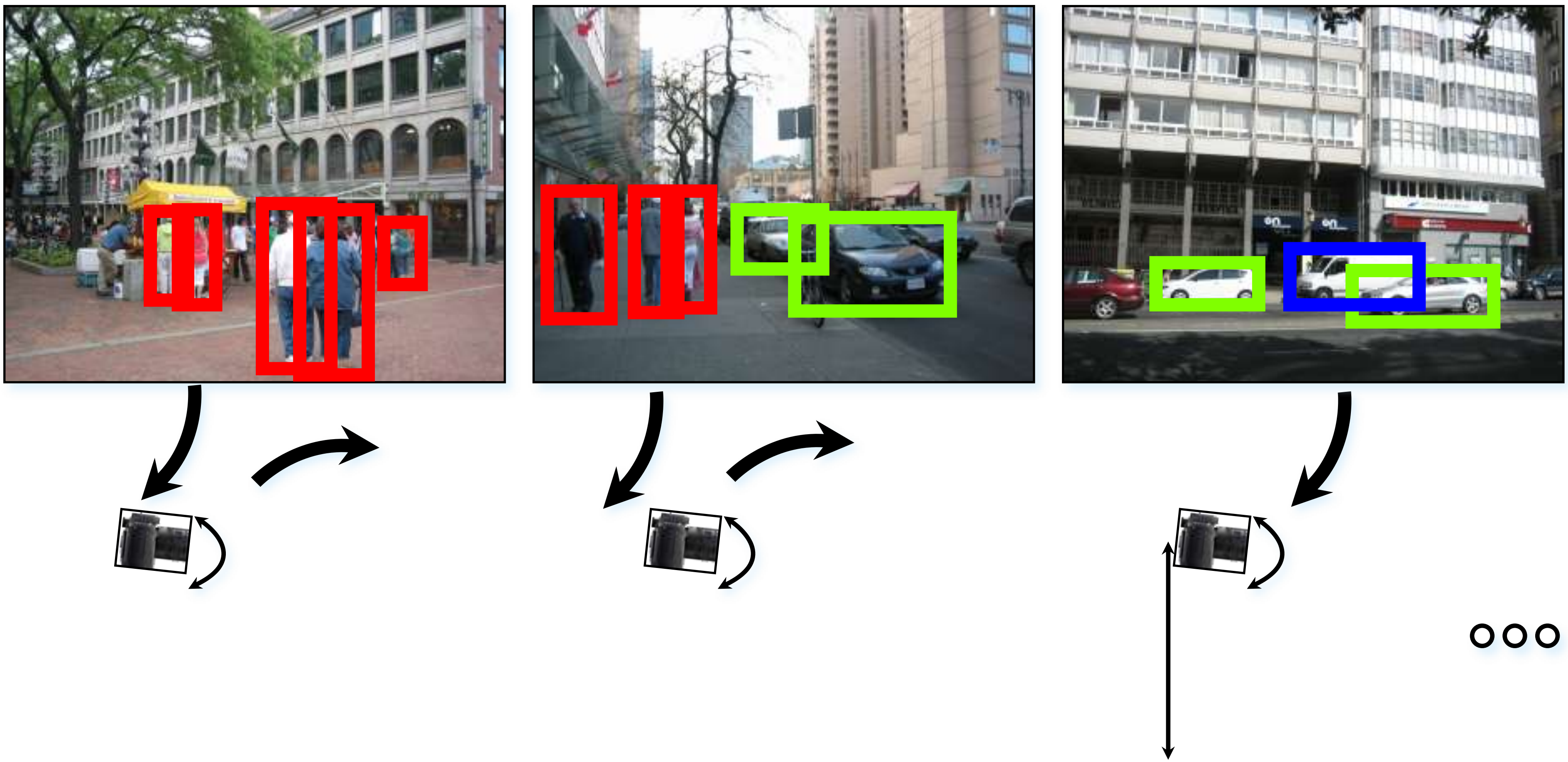
X – World object height (in meters)
C – World camera height (in meters)

A. Torralba

Camera parameters

Human height distribution
1.7 +/- 0.085 m
(National Center for Health Statistics)

Car height distribution
1.5 +/- 0.19 m
(automatically learned)



Slide from J-F Lalonde

Object heights

Database image

Pixel heights

Real heights

Slide from J-F Lalonde

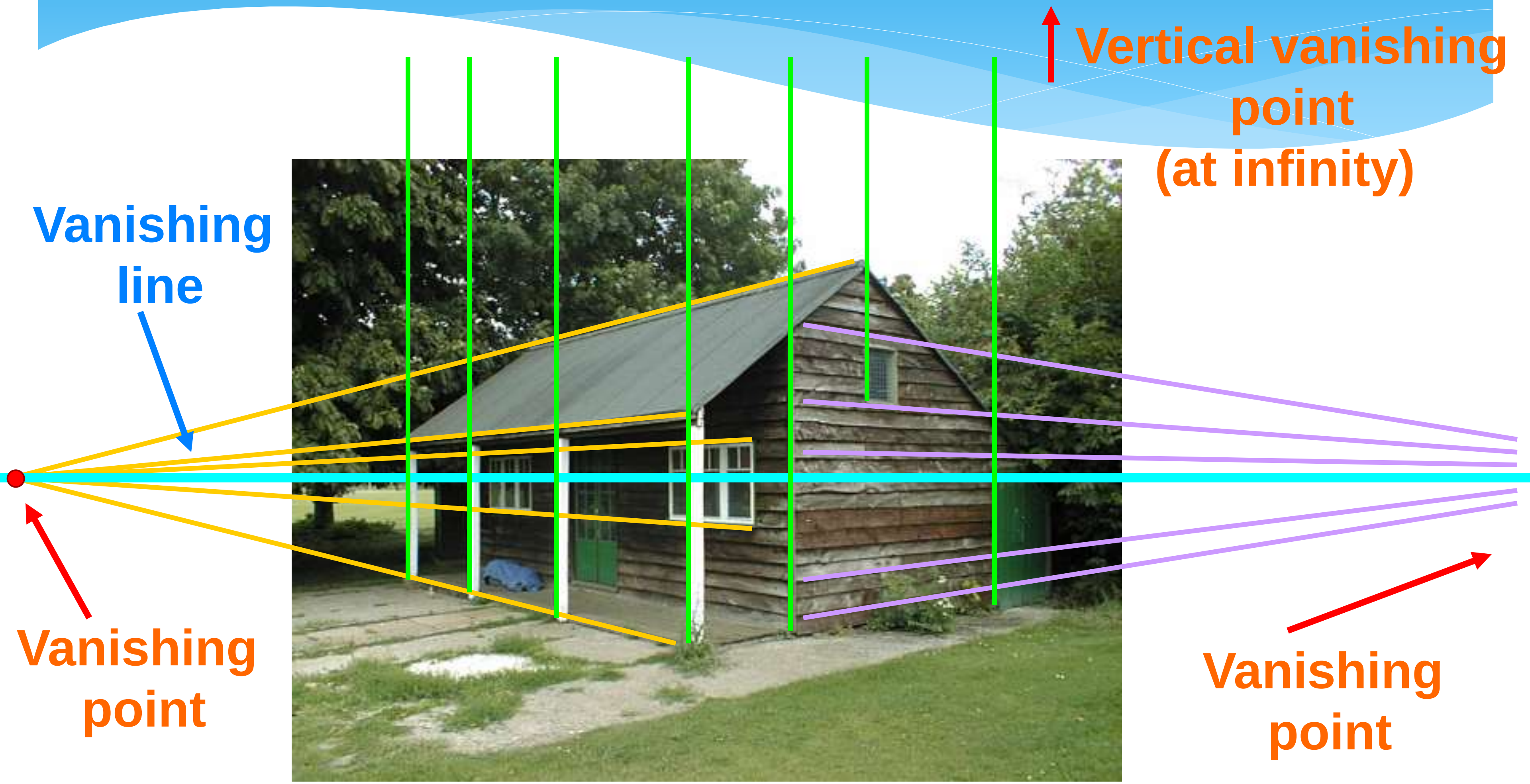
Depth from Vanishing Lines

Three-dimensional reconstruction from single and multiple images.

Antonio Criminisi

Microsoft Research, Cambridge, UK

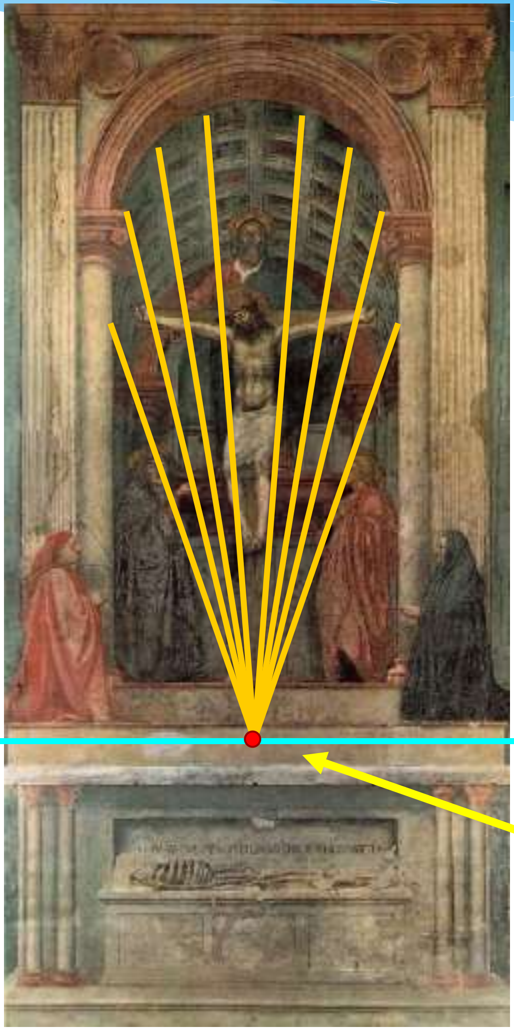
Visual cues



Source: A. Criminisi

Visual cues

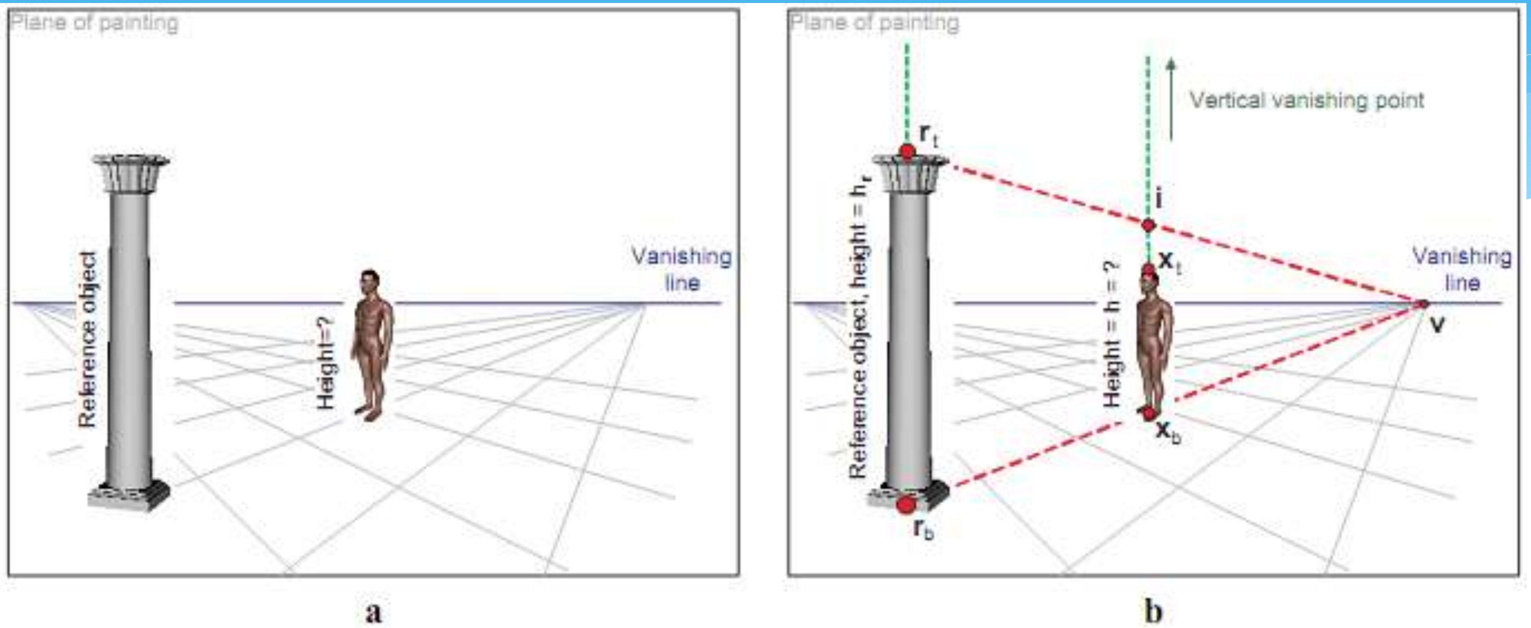
Masaccio's *Trinity*



Source: A. Criminisi

vanishing line (horizon)

vanishing point

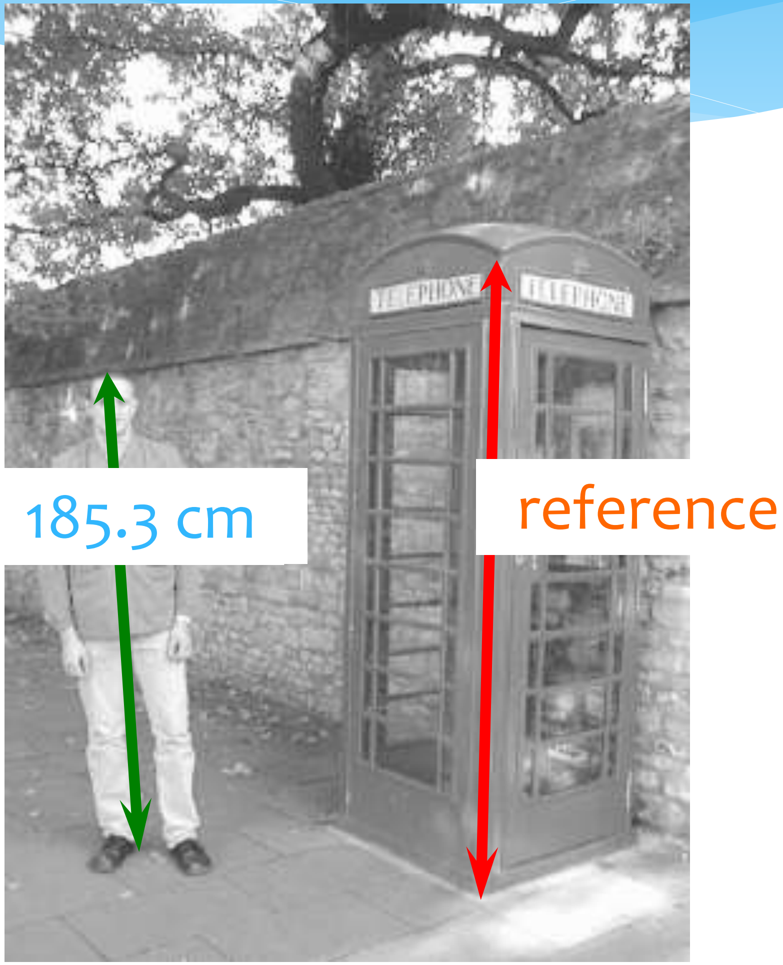


$$\frac{h}{h_r} = \frac{d(\mathbf{x}_t, \mathbf{x}_b)}{d(\mathbf{i}, \mathbf{x}_b)}$$

Source: A. Criminisi

Measuring heights in real photos

Problem: How tall is this person?

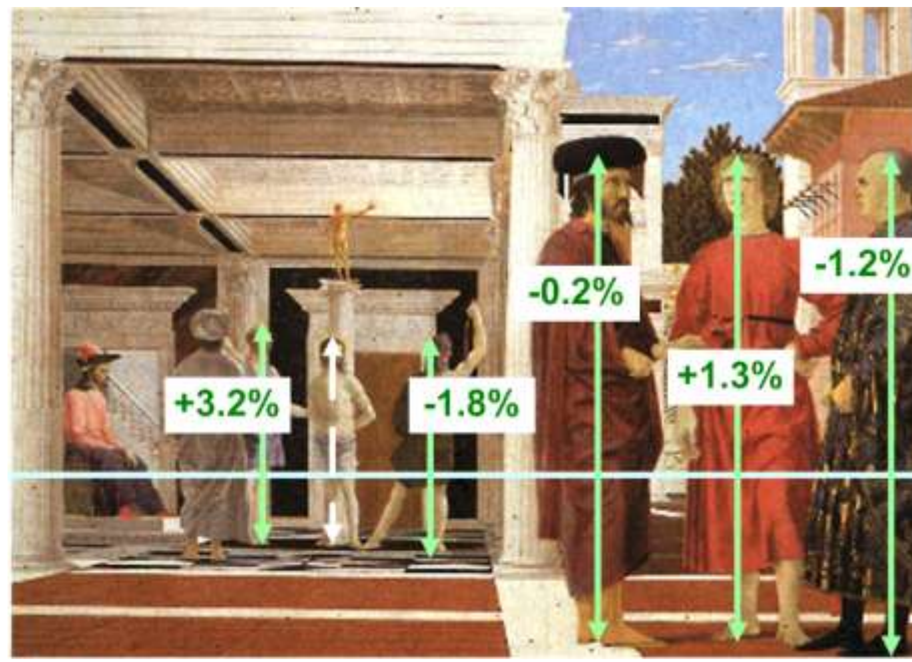


Source: A. Criminisi

Assessing geometric accuracy

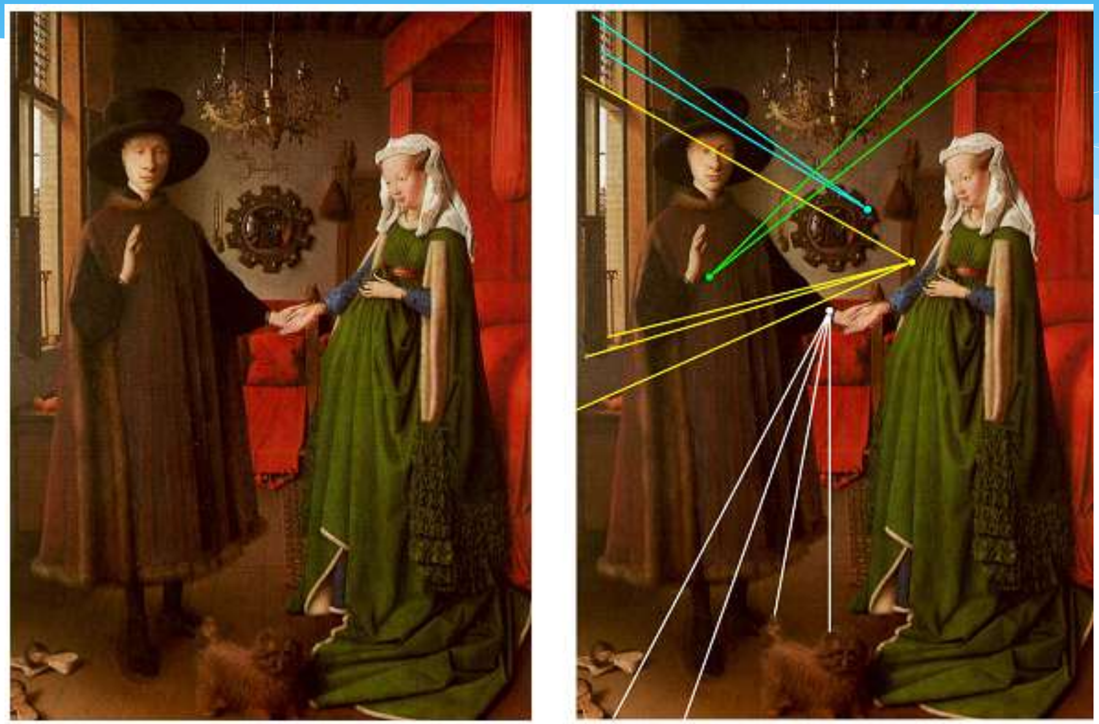
Problem:
Are the heights of the two groups of people consistent with each other?

Piero della Francesca, *Flagellazione di Cristo*, c.1460, Urbino



Measuring relative heights

Source: A. Criminisi



a b

Problems with Monocular Depth Cues

* They provide relative information.
* The ones that provide absolute information require a “reference”.
* What features/visual-information to investigate?
  * Usually hand-designed.
  * How can we also learn the features that lead to monocular cues?
* One cue is not sufficient.
  * Different cues should be combined.

What did I skip?

- Shape from silhouette.
- The details of most of the monocular cues (i.e., shading, shadow, occlusion, etc.).
- Reconstruction from disparity, especially for features like edges and corners.

Reading

* I will supply material for:
  * Stereo
  * Depth from motion
  * Monocular cues