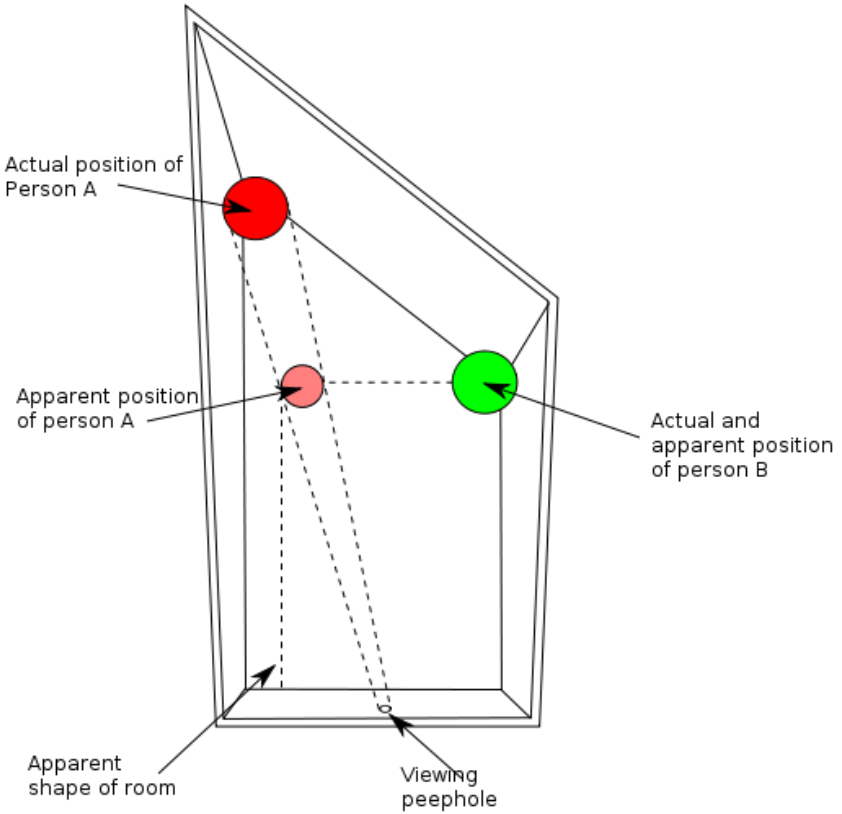
Single & Stereo View Geometry

<Vision System>

Department of Robot Engineering

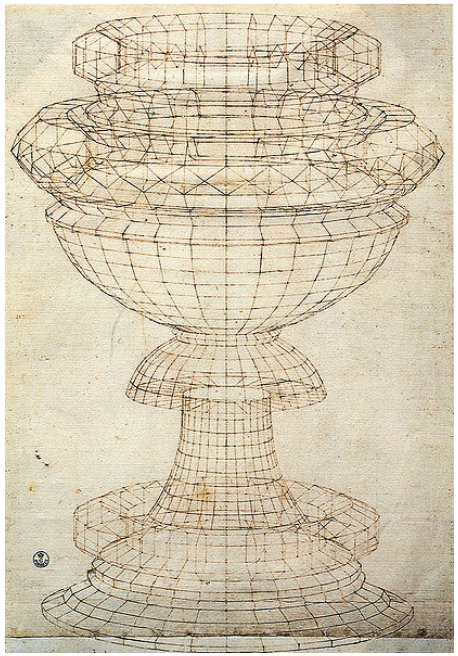
Prof. Younggun Cho

Ames Room


Actual position of Person A
Apparent position of person A
Actual and apparent position of person B
Apparent shape of room
Viewing peephole

Projective geometry—what's it good for?

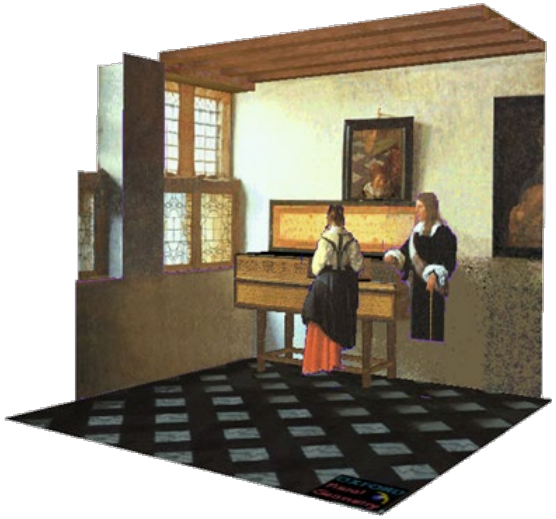
- Uses of projective geometry
  - Drawing
  - Measurements
  - Mathematics for projection
  - Undistorting images
  - Camera pose estimation
  - **Object recognition**



Paolo Uccello

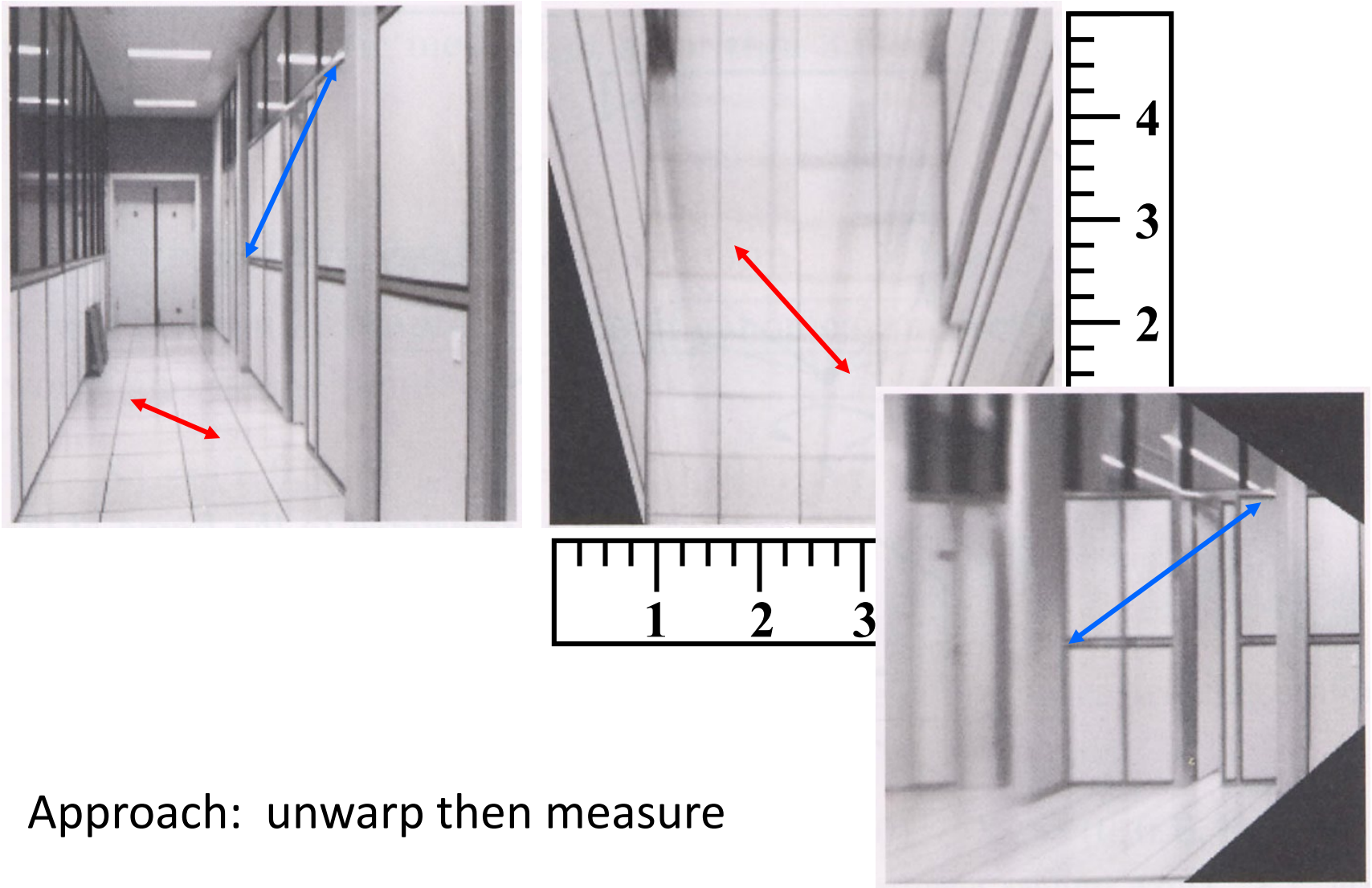
Applications of projective geometry

Vermeer's *Music Lesson*



Reconstructions by Criminisi et al.

Measurements on planes

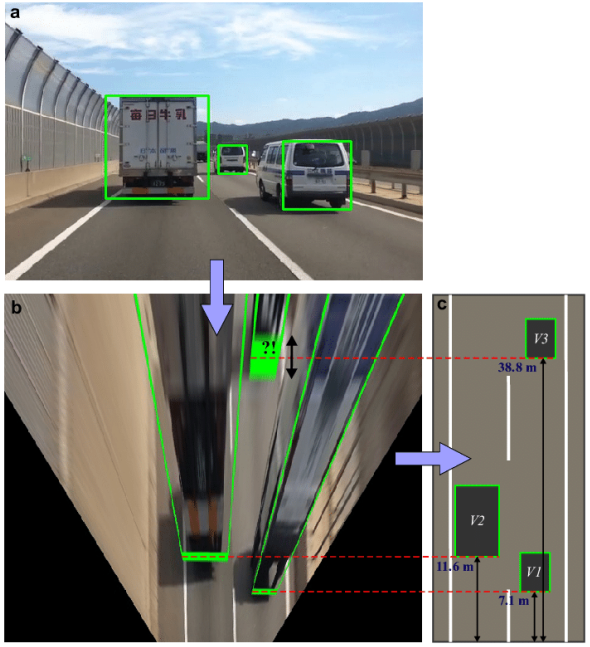


Approach: unwarp then measure

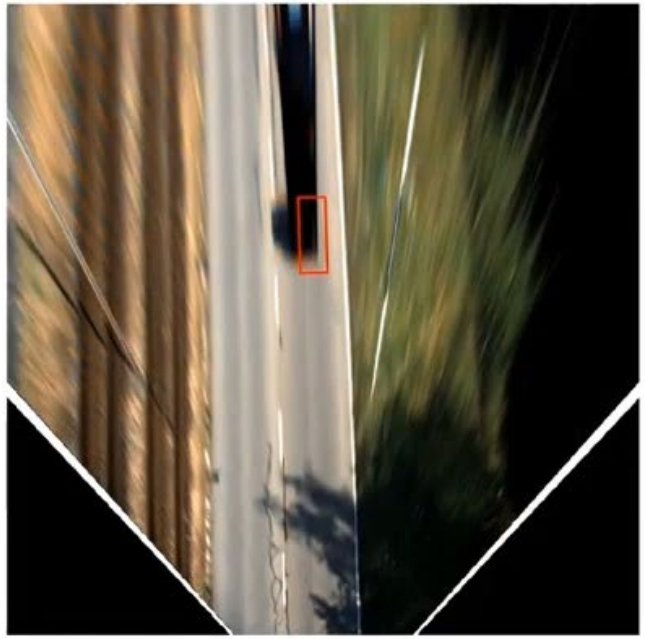
Application to Autonomous Car

https://www.youtube.com/watch?v=2zvS87d1png

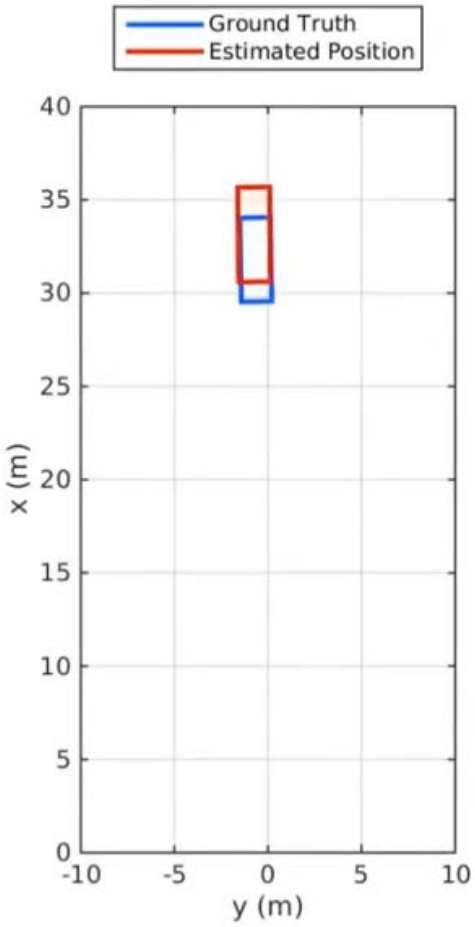
- IPM (inverse perspectiv mapping)



Projection of detection results onto FV image



Detection results on BEV image

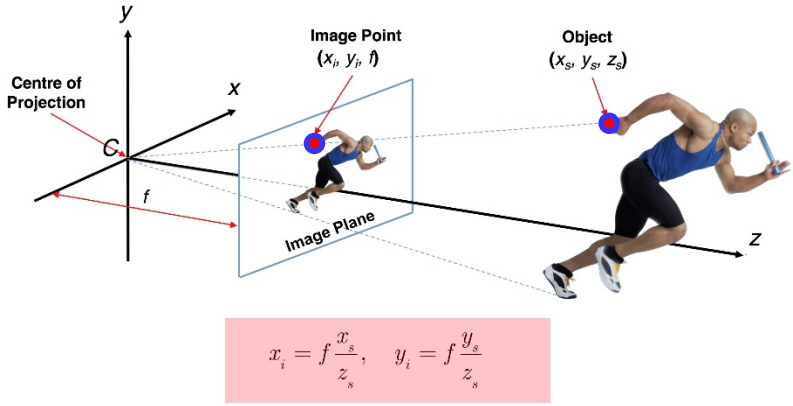


Estimated distance in meter unit

KITTI raw

(Basic) Homogeneous coordinates and lines

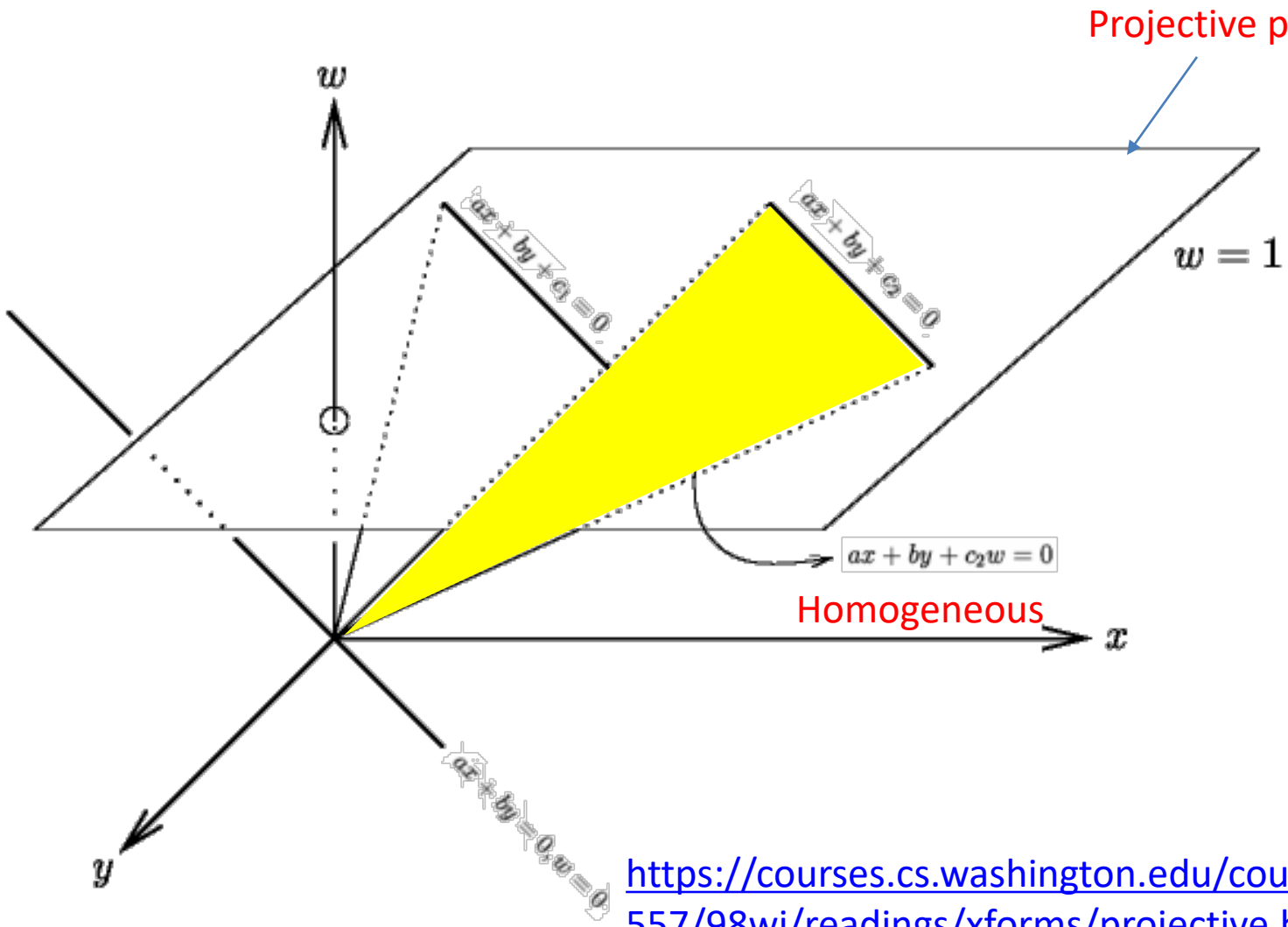
- Point: (X,Y,W) same as (SX,SY,SW)



- Line in projective plane
  - $ax + by + c = 0$ (w=1)
  - $aX + bY + cW = 0$ (general)
  - $u^T p = p^T u = 0$ (linear form)
    - $u = [a,b,c]^T$ and $p = [X,Y,W]^T$
    - (x,y) Euclidean = (X/W,Y/W)
    - W = 0? Ideal points, points at infinity (X,Y,0)

https://mc.ai/part-i-projective-geometry-in-2d/

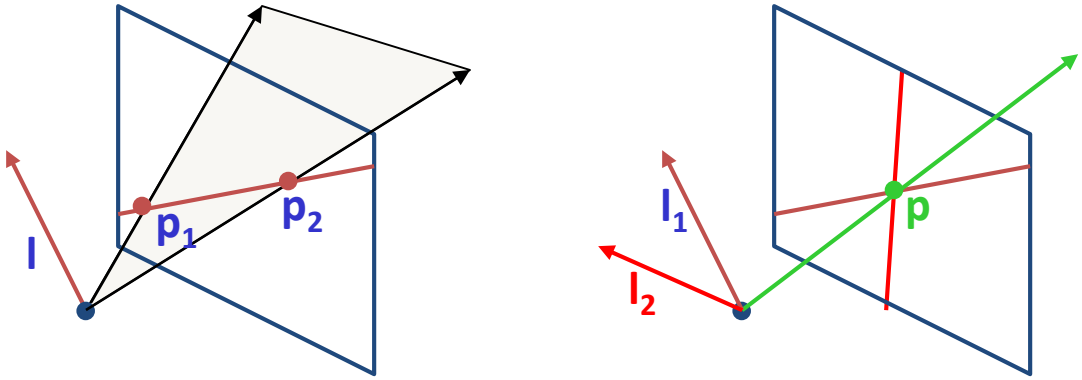
Projective vs. Homogeneous



https://courses.cs.washington.edu/courses/cse557/98wi/readings/xforms/projective.html

Point and line duality

- A line **l** is a homogeneous 3-vector
- It is ⊥ to every point (ray) **p** on the line: $\mathbf{l} \cdot \mathbf{p} = 0$



What is the line **l** spanned by rays $\mathbf{p}_1$ and $\mathbf{p}_2$?

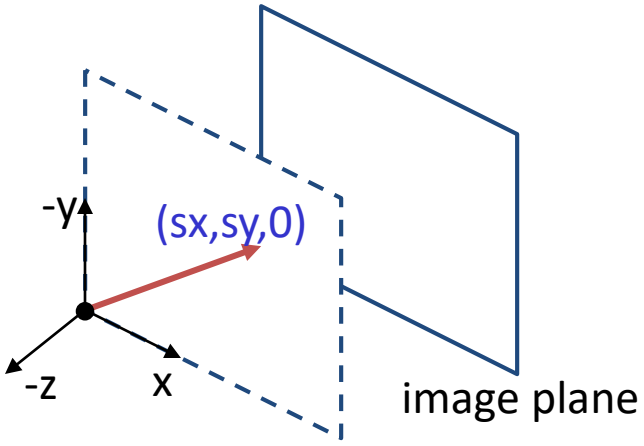
- **l** is ⊥ to $\mathbf{p}_1$ and $\mathbf{p}_2$ $\Rightarrow$ $\mathbf{l} = \mathbf{p}_1 \times \mathbf{p}_2$
- **l** can be interpreted as a *plane normal*

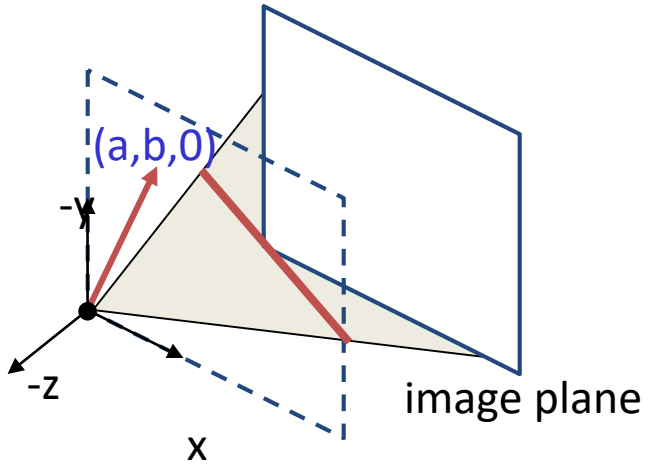
What is the intersection of two lines $\mathbf{l}_1$ and $\mathbf{l}_2$?

- **p** is ⊥ to $\mathbf{l}_1$ and $\mathbf{l}_2$ $\Rightarrow$ $\mathbf{p} = \mathbf{l}_1 \times \mathbf{l}_2$

Points and lines are *dual* in projective space

Ideal points and lines





- Ideal point ("point at infinity")
  - $p \cong (x, y, 0)$ – parallel to image plane
  - It has infinite image coordinates

Ideal line

- $l \cong (a, b, 0)$ – parallel to image plane
- Corresponds to a line in the image (finite coordinates)
  - goes through image origin (*principal point*)

3D projective geometry

- These concepts generalize naturally to 3D
  - Homogeneous coordinates
    - Projective 3D points have four coords: **P** = (X,Y,Z,W)
  - Duality
    - A plane **N** is also represented by a 4-vector
    - Points and planes are dual in 3D: **N P**=0
    - Three points define a plane, three planes define a point

3D to 2D: perspective projection

Projection:

$$\mathbf{p} = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{\Pi P}$$

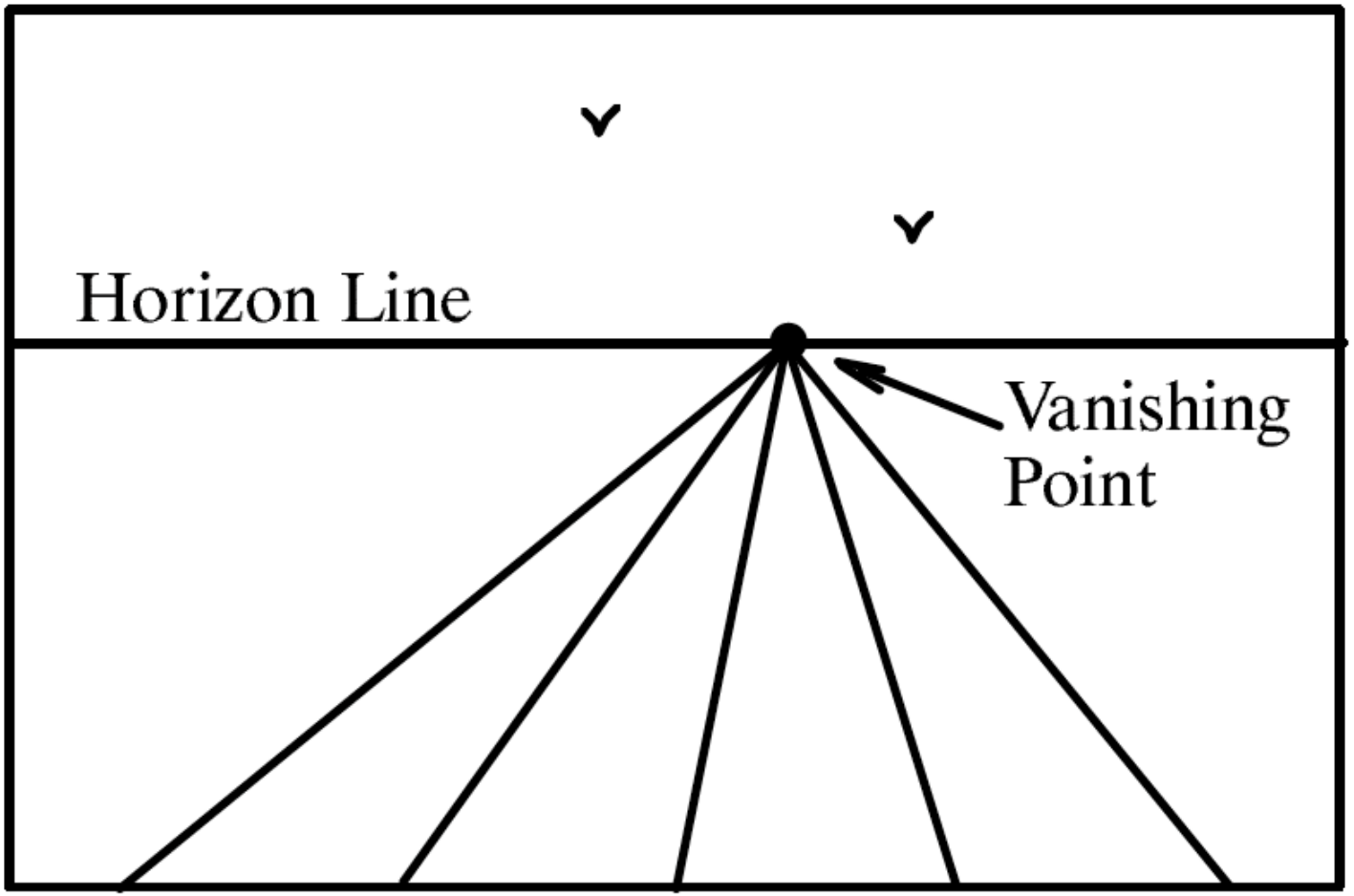
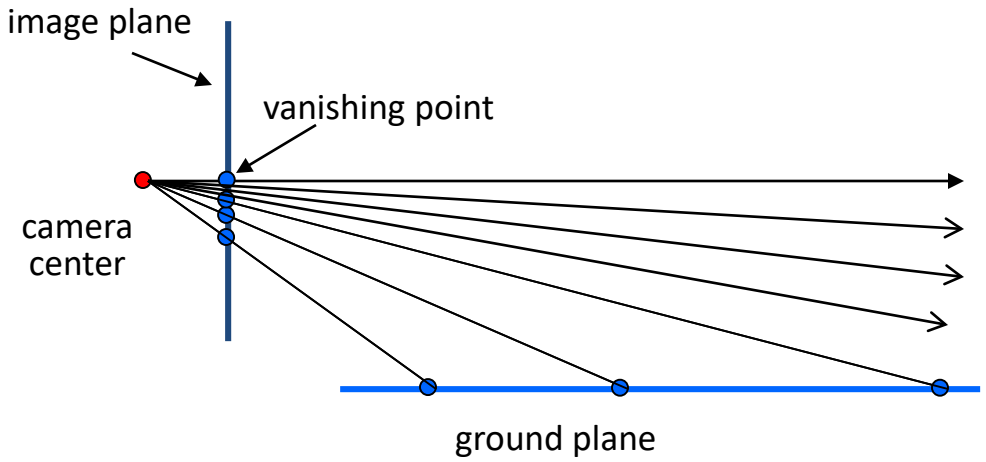
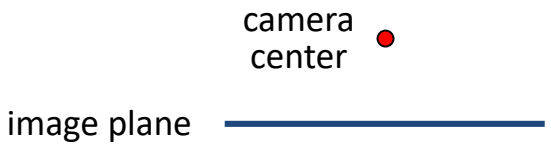


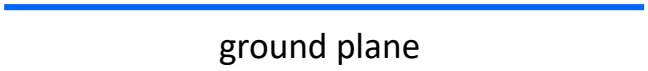
Figure 23.4
A perspective view of a set of parallel lines in the plane. All of the lines converge to a single vanishing point.

Vanishing points (1D)

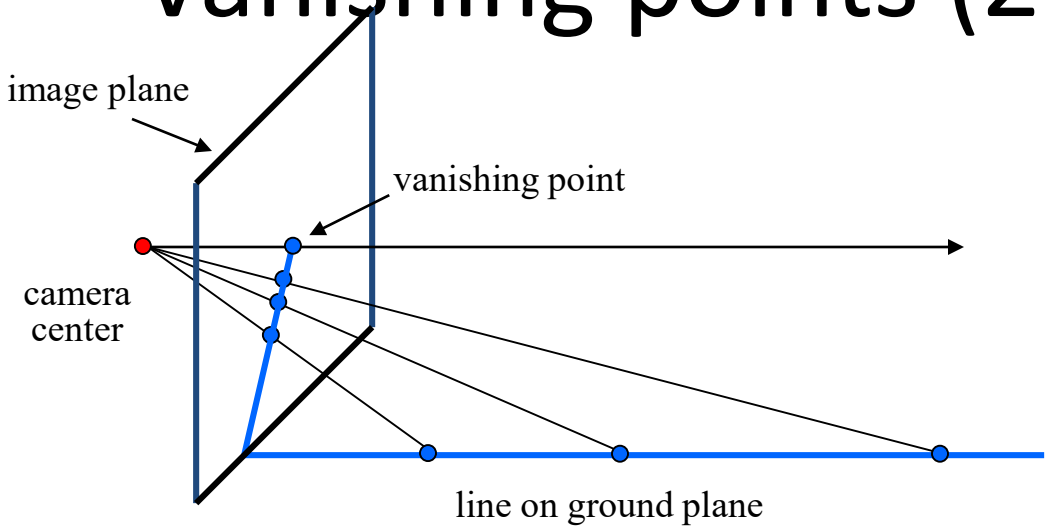


- Vanishing point
  - projection of a point at infinity
  - can often (but not always) project to a finite point in the image

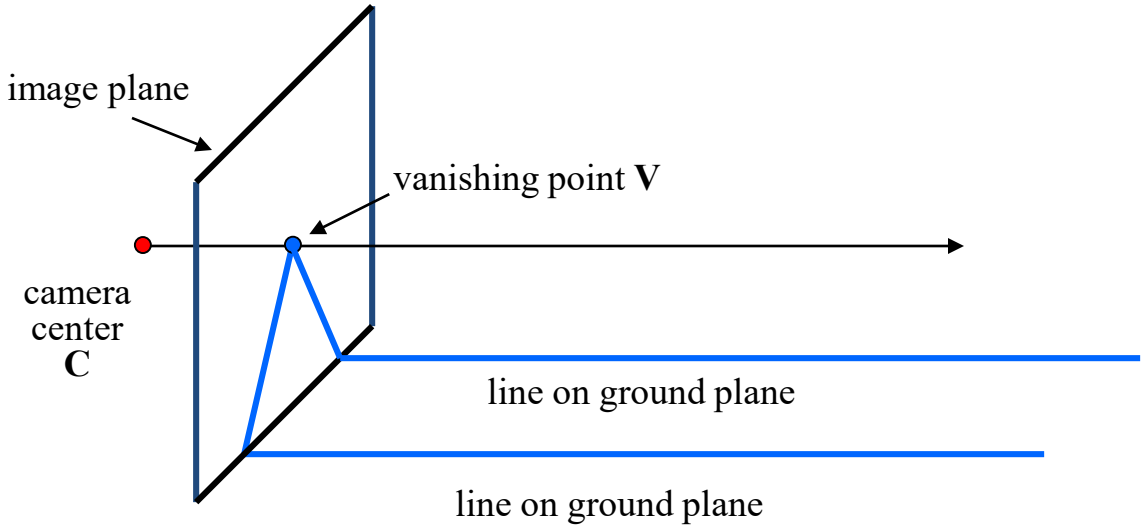




Vanishing points (2D)



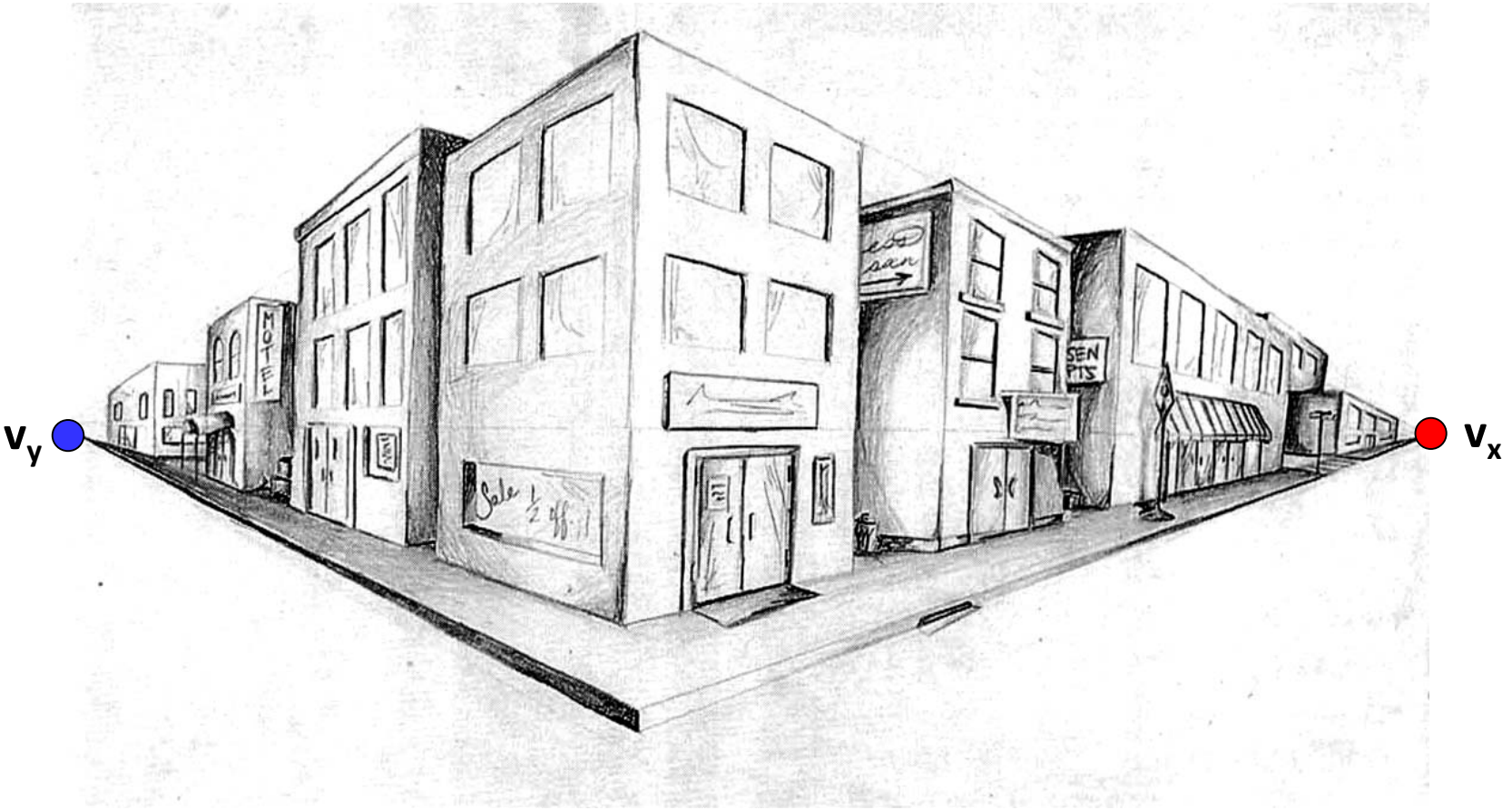
Vanishing points



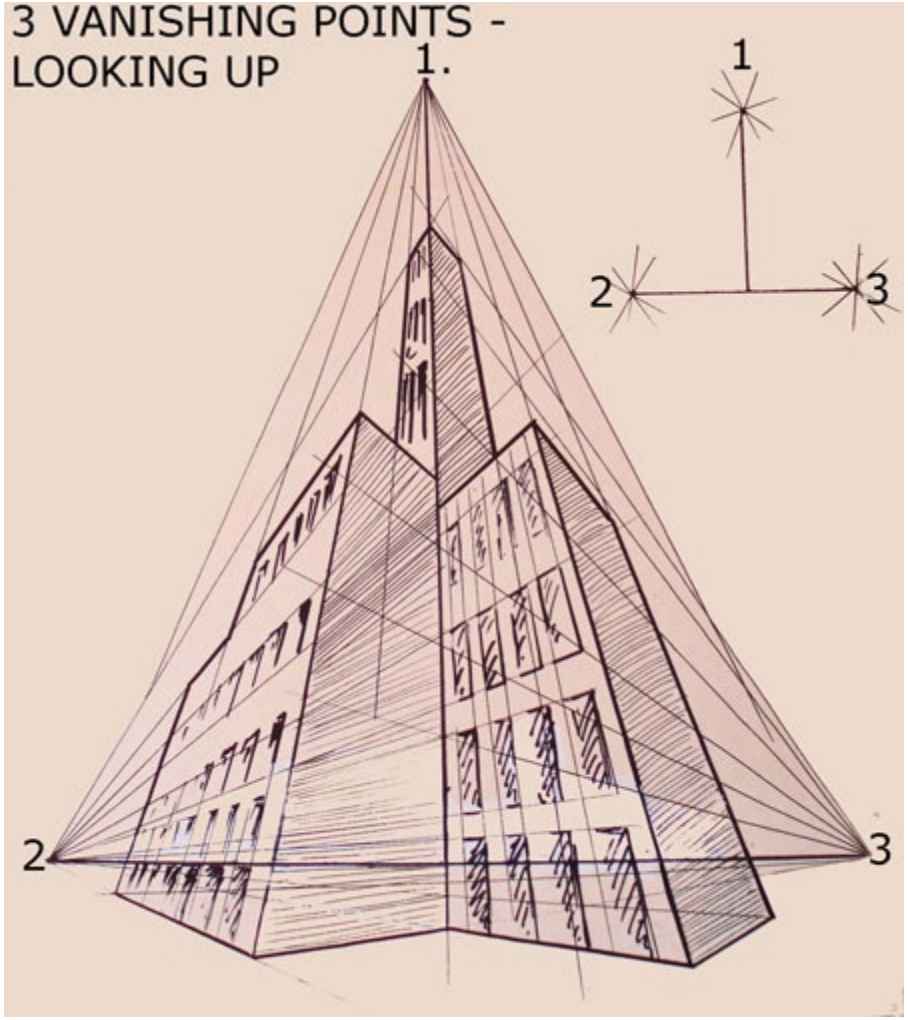
- Properties
  - Any two parallel lines (in 3D) have the same vanishing point **v**
  - The ray from **C** through **v** is parallel to the lines
  - An image may have more than one vanishing point
    - in fact, every image point is a potential vanishing point

One-point perspective

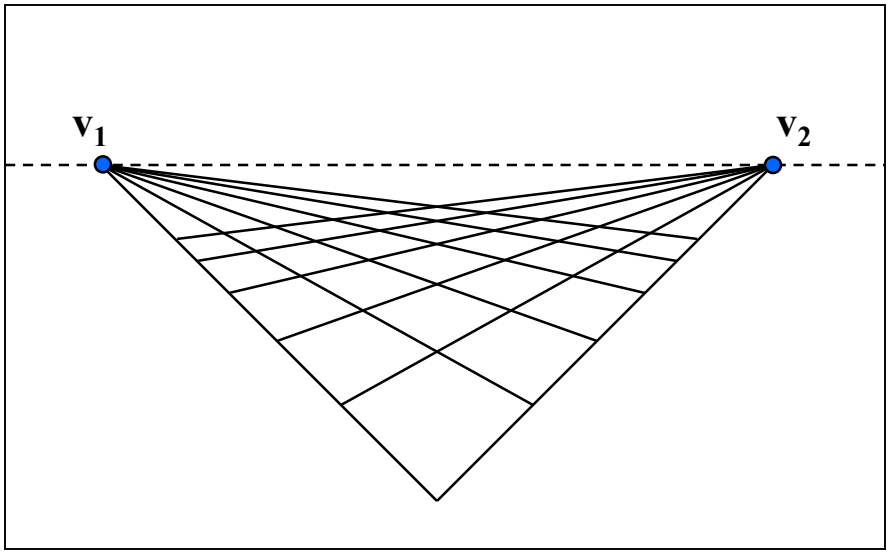
Two-point perspective



Three-point perspective

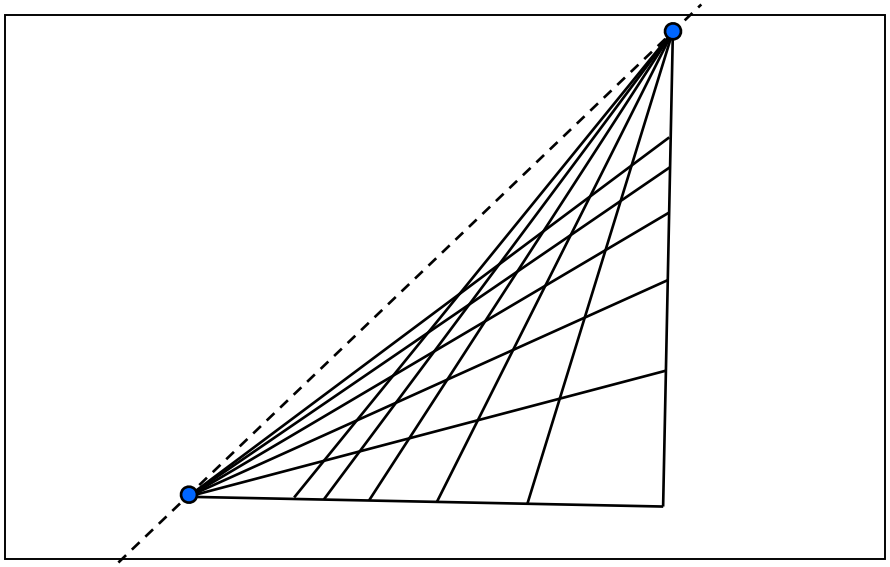


Vanishing lines



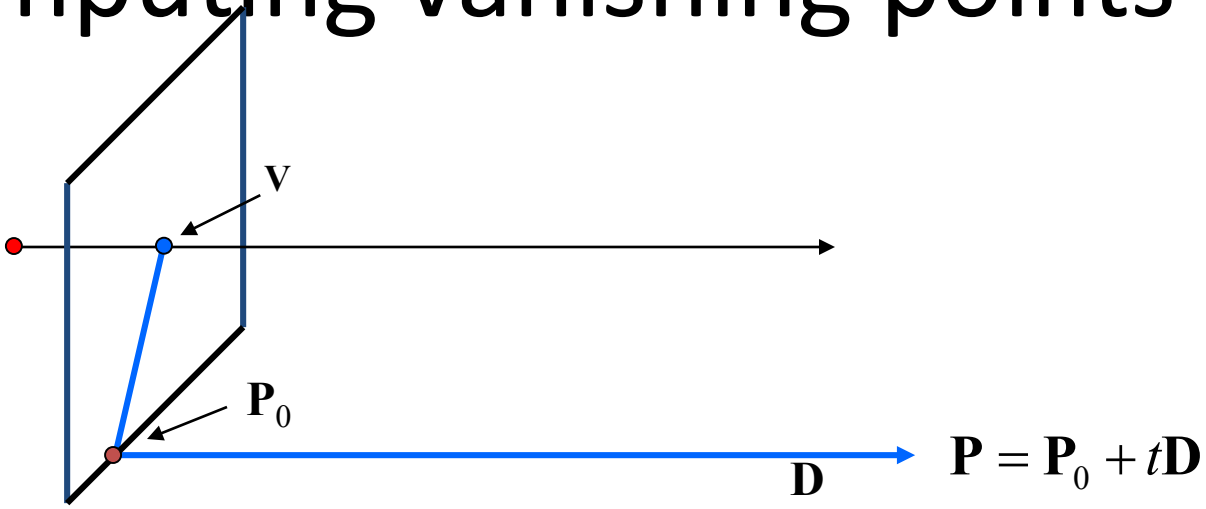
- Multiple Vanishing Points
  - Any set of parallel lines on the plane define a vanishing point
  - The union of all of these vanishing points is the *horizon line*
    - also called *vanishing line*
  - Note that different planes (can) define different vanishing lines

Vanishing lines



- Multiple Vanishing Points
  - Any set of parallel lines on the plane define a vanishing point
  - The union of all of these vanishing points is the *horizon line*
    - also called *vanishing line*
  - Note that different planes (can) define different vanishing lines

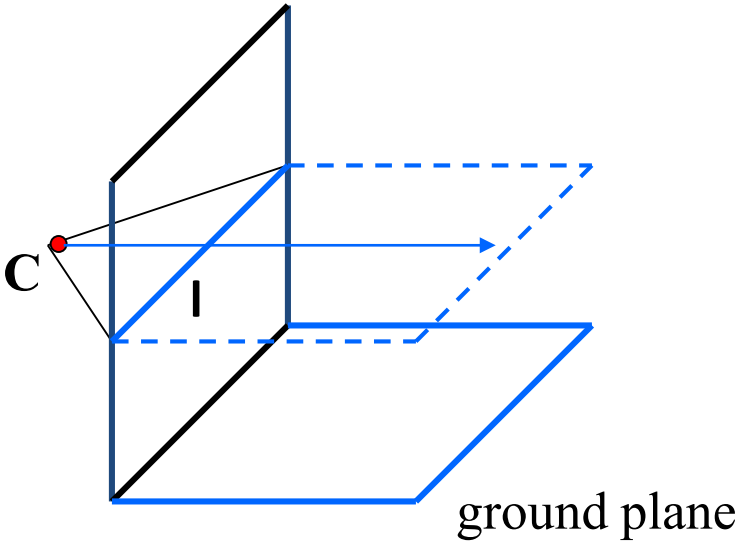
Computing vanishing points

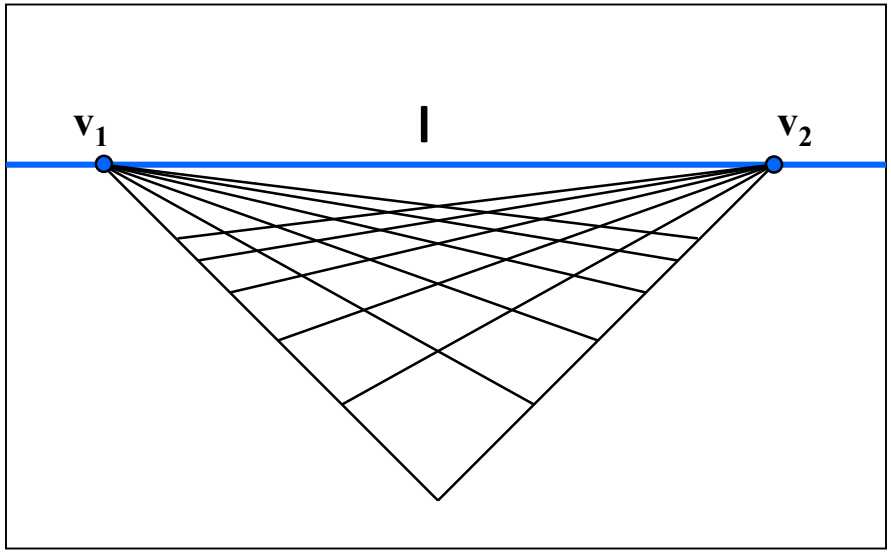


$$\mathbf{P}_t = \begin{bmatrix} P_X + tD_X \\ P_Y + tD_Y \\ P_Z + tD_Z \\ 1 \end{bmatrix} \cong \begin{bmatrix} P_X / t + D_X \\ P_Y / t + D_Y \\ P_Z / t + D_Z \\ 1/t \end{bmatrix}$$

- Properties $\mathbf{v} = \Pi \mathbf{P}_\infty$
  - $\mathbf{P}_\infty$ is a point at *infinity*, **v** is its projection
  - Depends only on line *direction*
  - Parallel lines $\mathbf{P}_0 + t\mathbf{D}$, $\mathbf{P}_1 + t\mathbf{D}$ intersect at $\mathbf{P}_\infty$

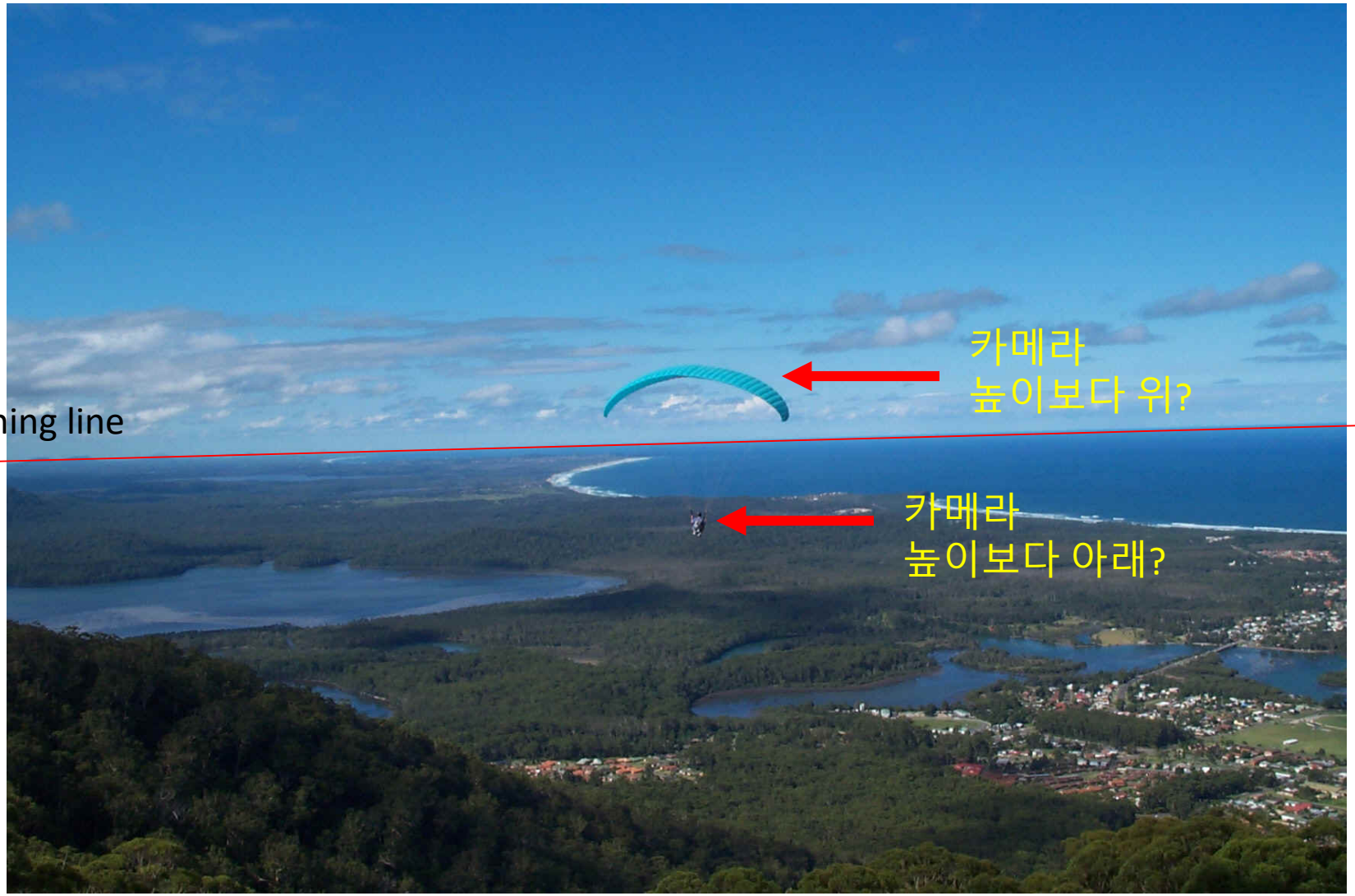
Computing vanishing lines



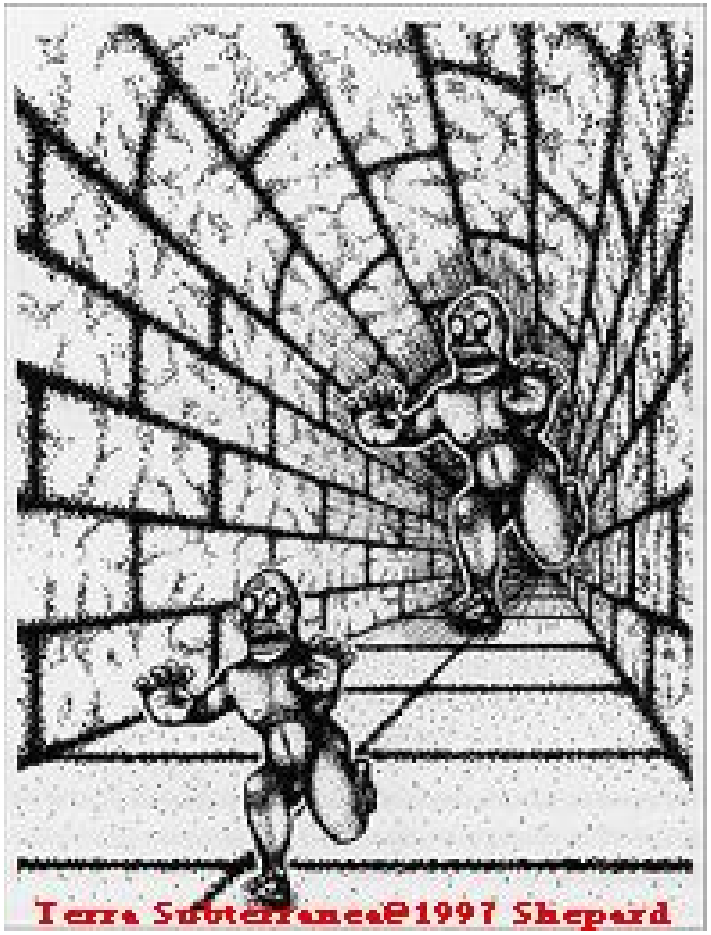


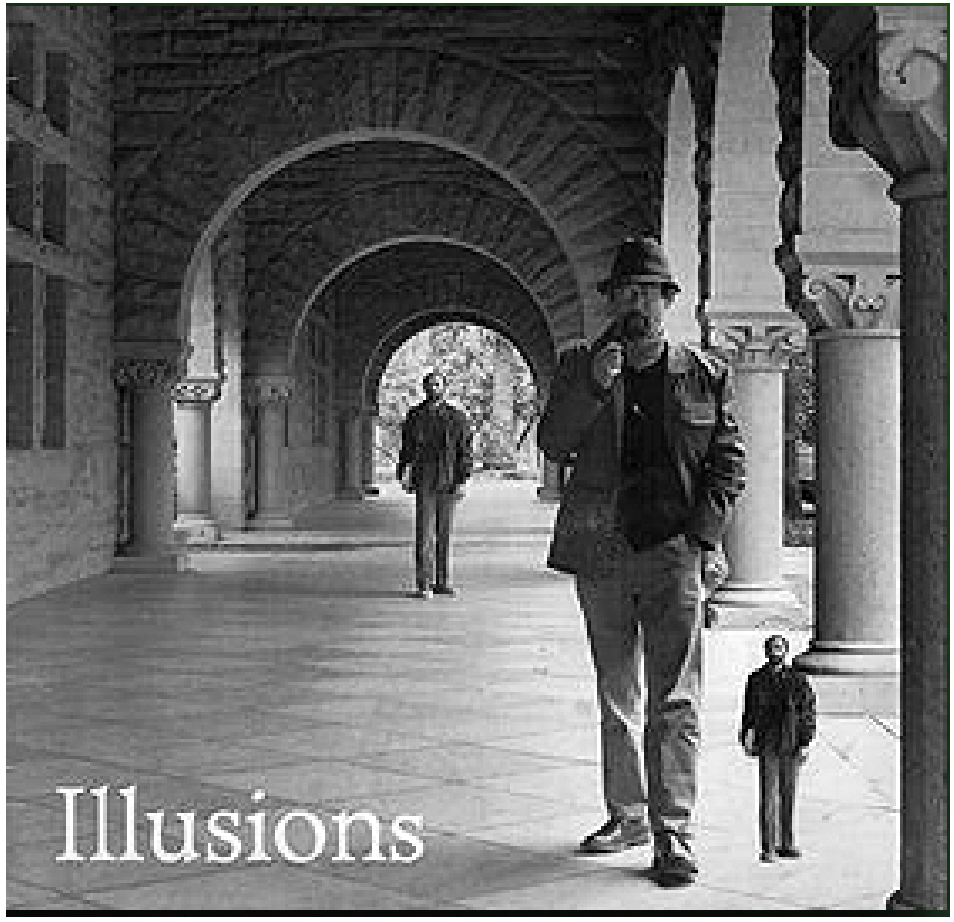
- Properties
  - **l** is intersection of horizontal plane through **C** with image plane
  - Compute **l** from two sets of parallel lines on ground plane
  - All points at same height as **C** project to **l**
    - points higher than C project above l
  - Provides way of comparing height of objects in the scene

Example



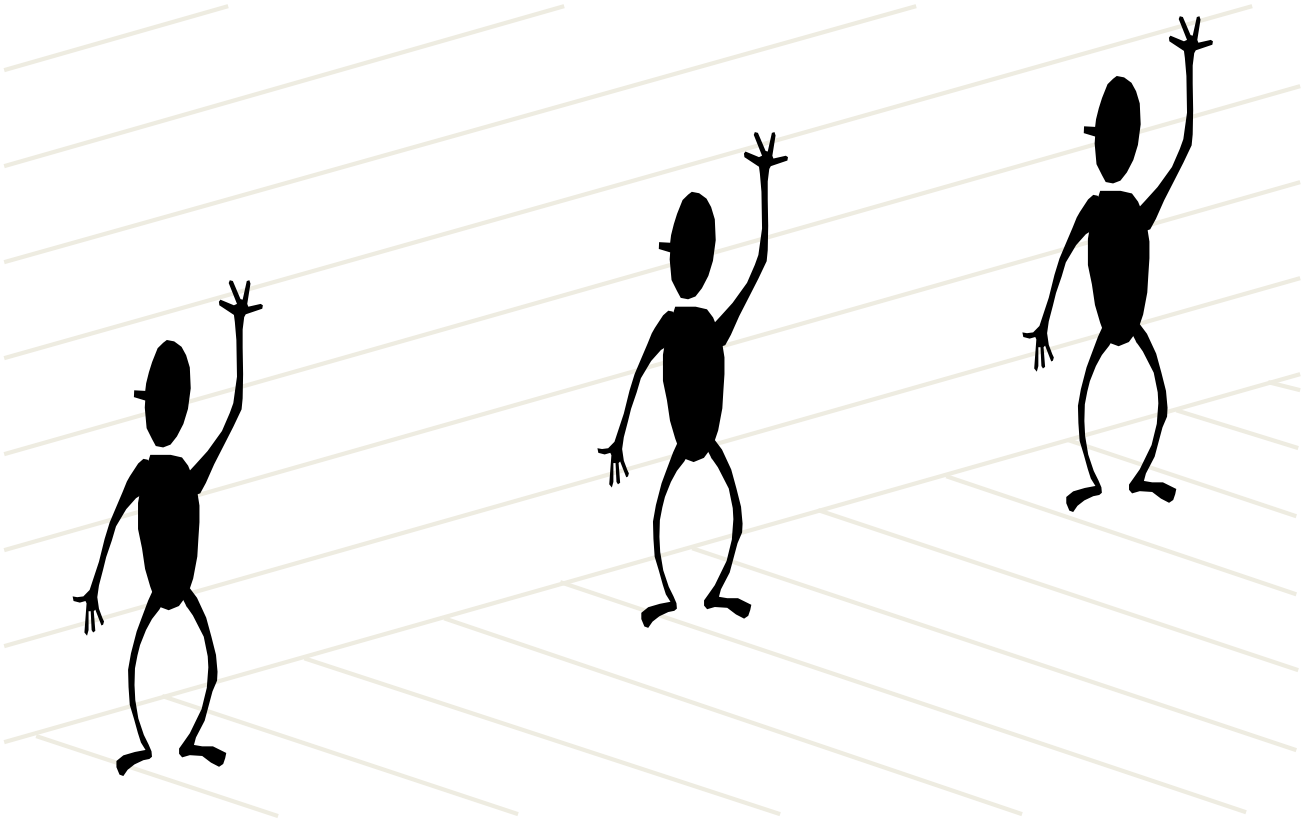
Fun with vanishing points



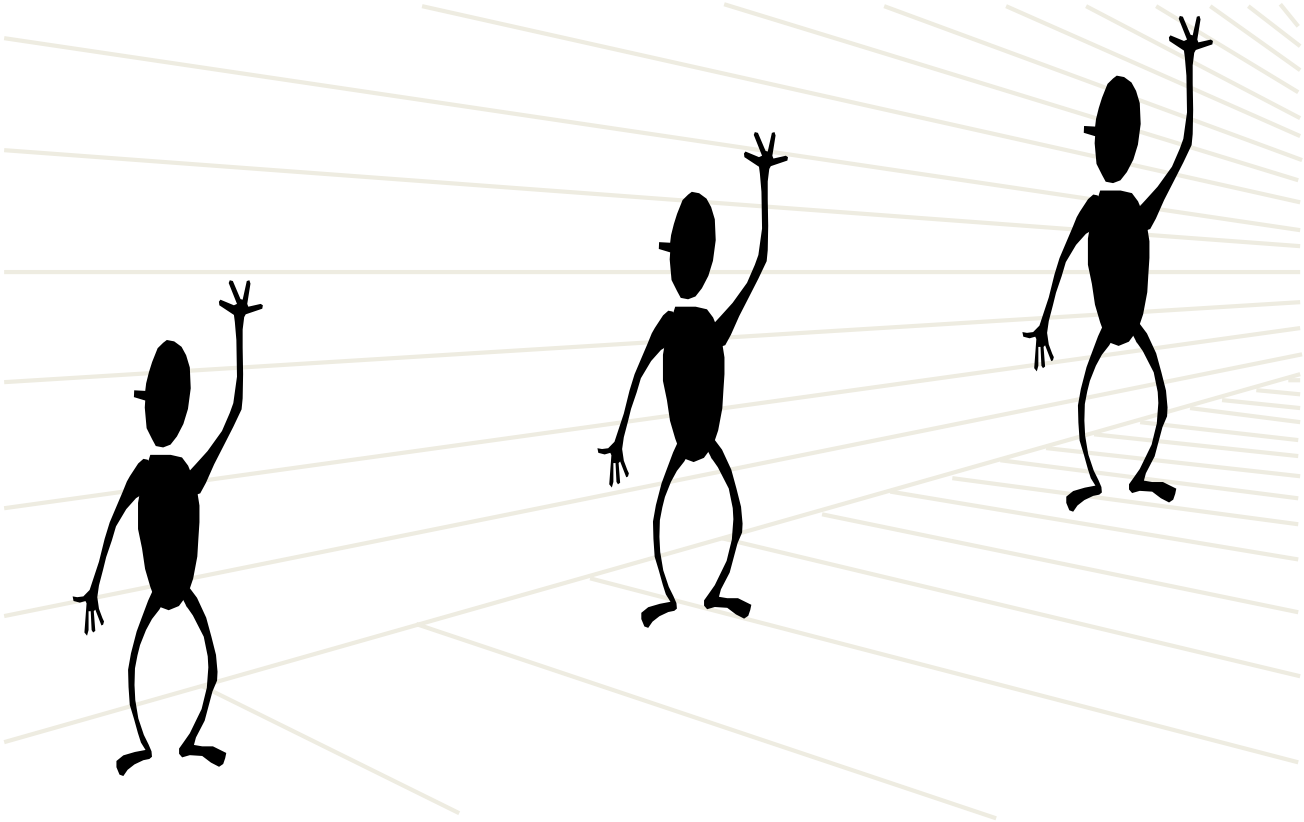


Lots of fun with vanishing points

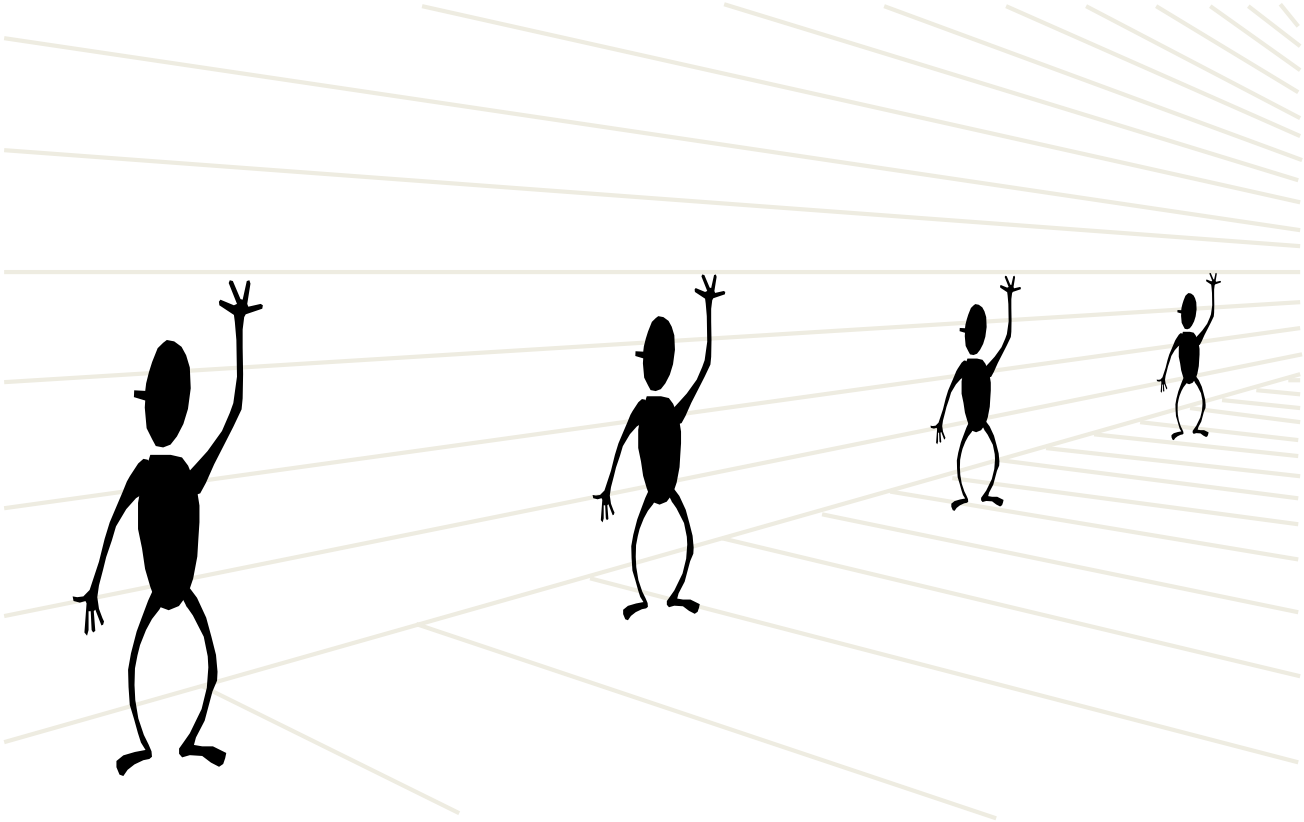
Perspective cues



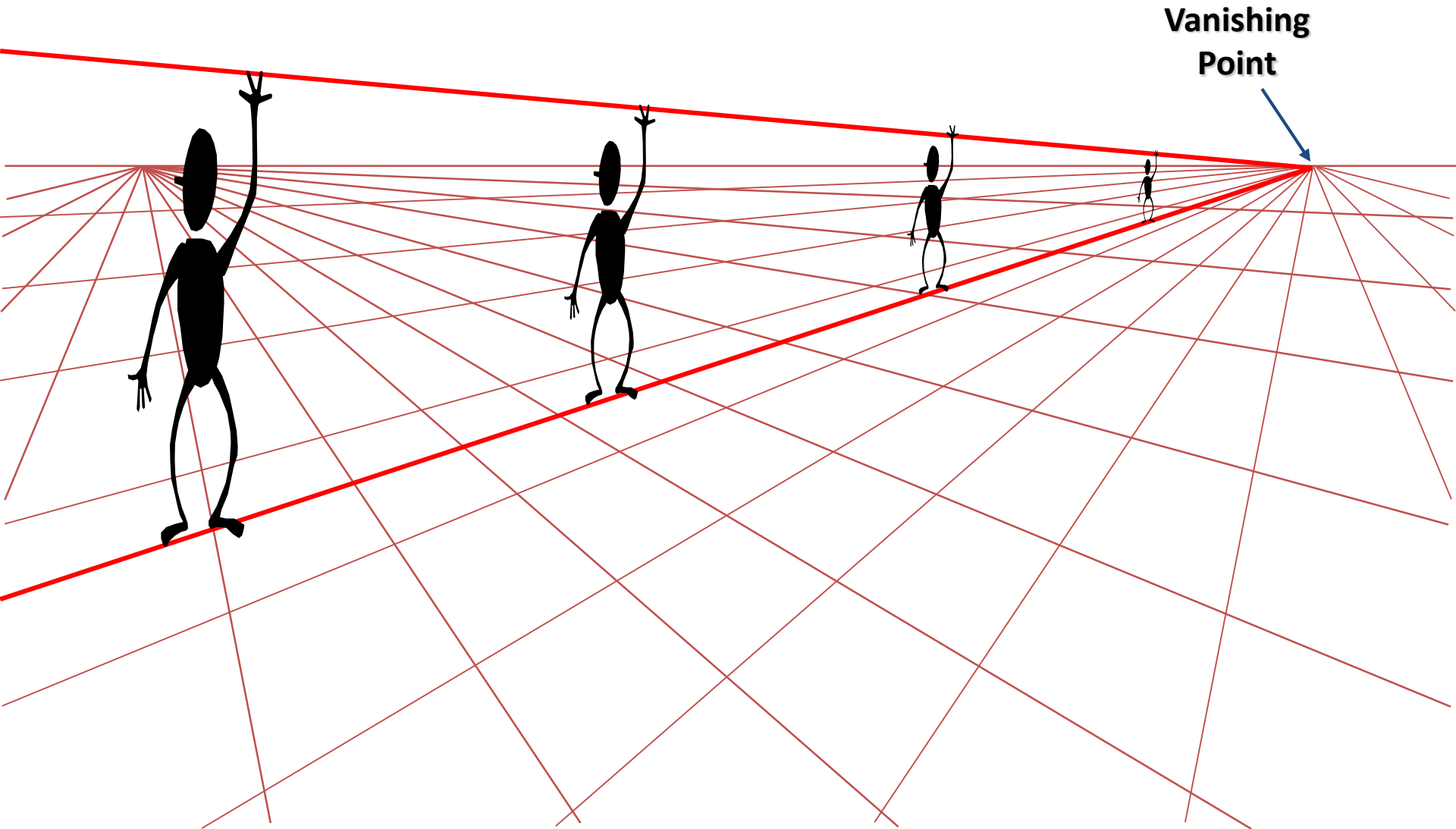
Perspective cues



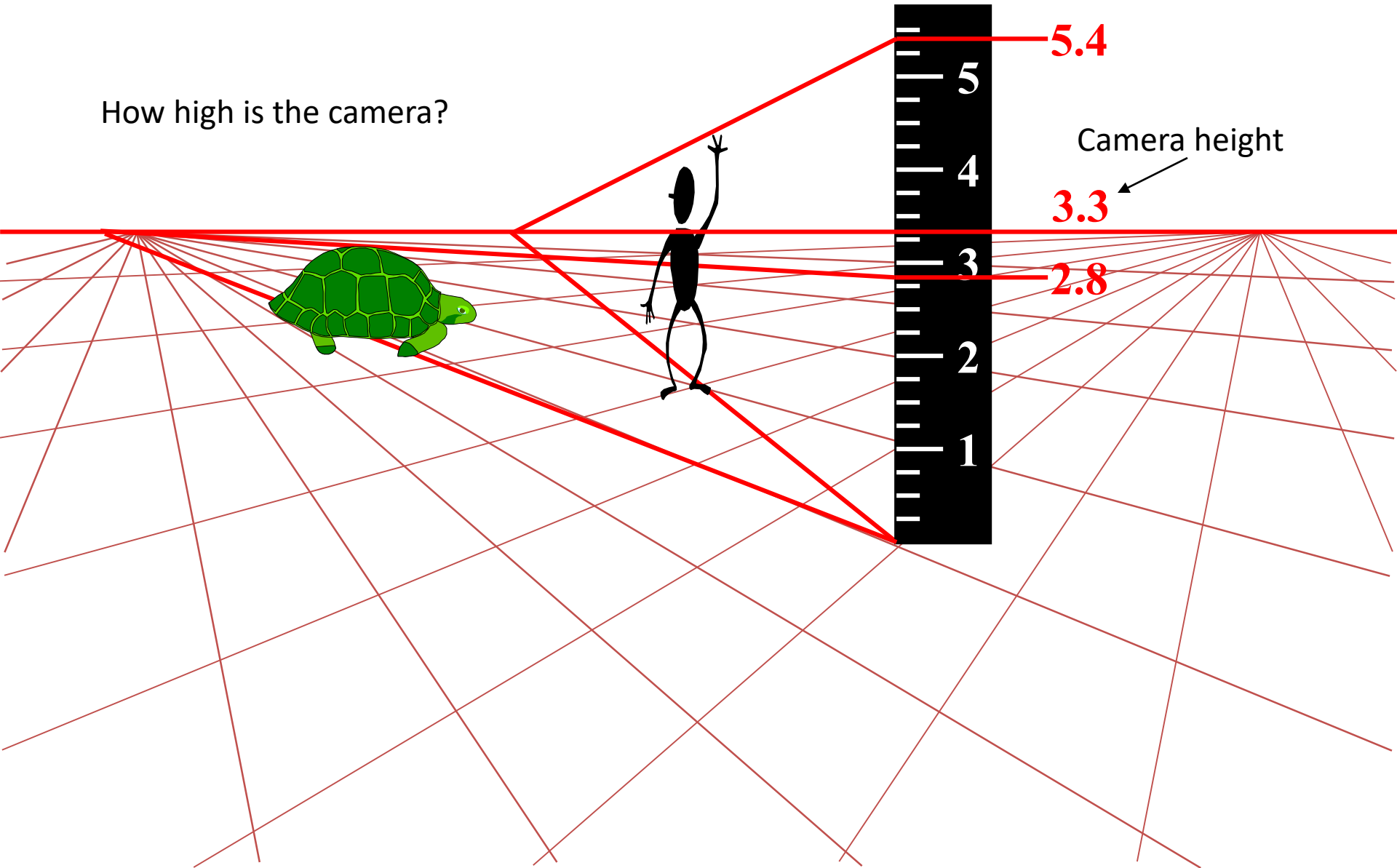
Perspective cues



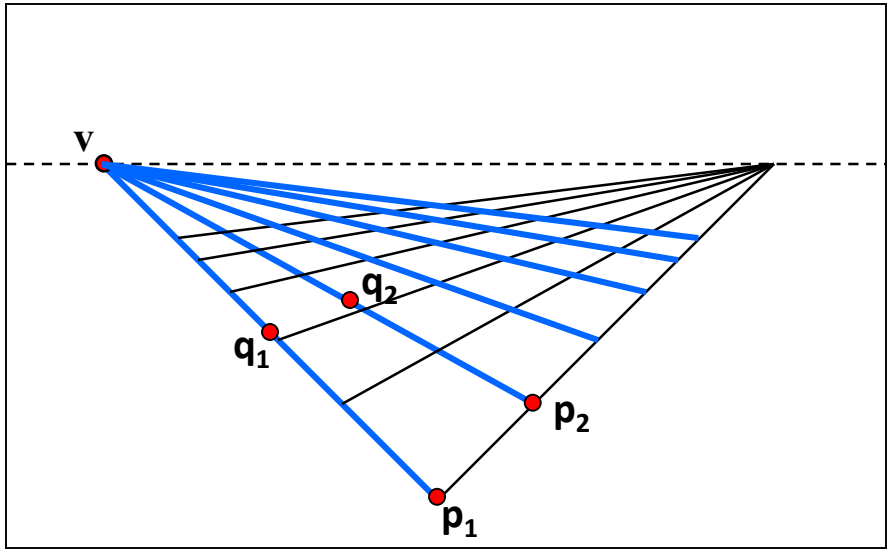
Comparing heights



Measuring height



Computing vanishing points (from lines)



- Intersect p_1q_1 with p_2q_2

$$v = (p_1 \times q_1) \times (p_2 \times q_2)$$

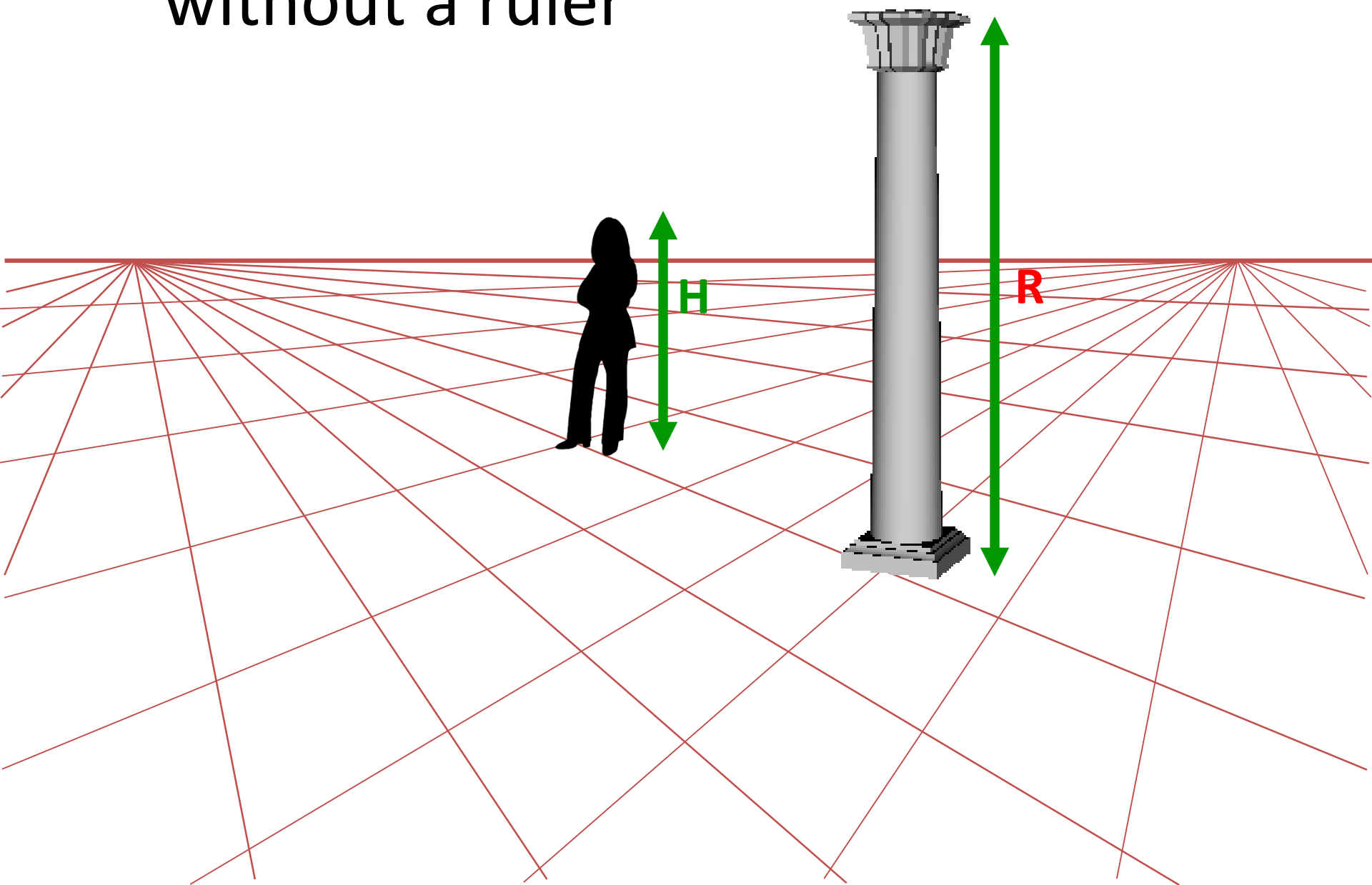
Plane normal Plane normal

See p.12

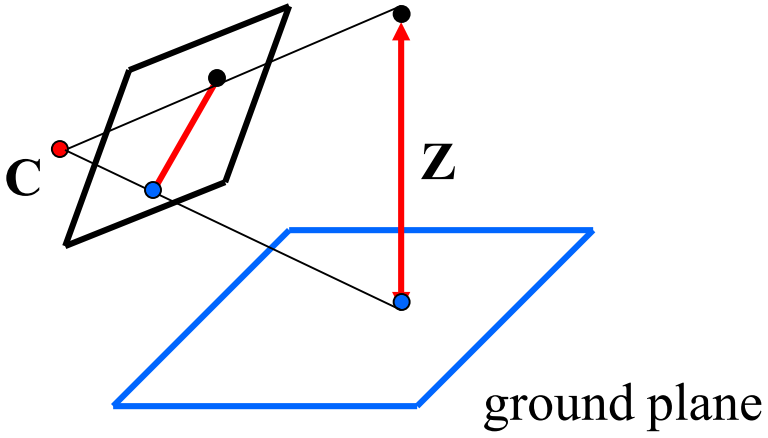
Least squares version

- Better to use more than two lines and compute the “closest” point of intersection
- See notes by Bob Collins for one good way of doing this:
  - http://www-2.cs.cmu.edu/~ph/869/www/notes/vanishing.txt

Measuring height without a ruler



Measuring height without a ruler



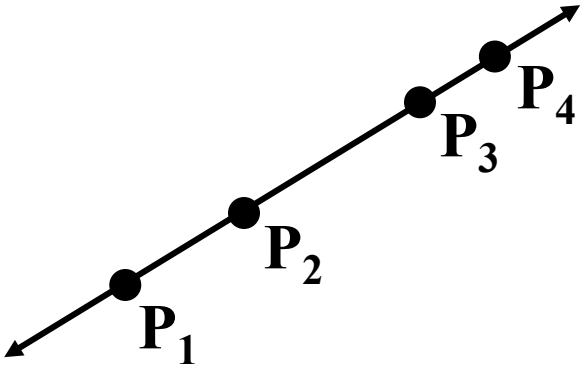
Compute Z from image measurements

- Need more than vanishing points to do this

The cross ratio

- A Projective Invariant
  - Something that does not change under projective transformations (including perspective projection)

The *cross-ratio* of 4 collinear points



$$\frac{\|\mathbf{P}_3 - \mathbf{P}_1\| \, \|\mathbf{P}_4 - \mathbf{P}_2\|}{\|\mathbf{P}_3 - \mathbf{P}_2\| \, \|\mathbf{P}_4 - \mathbf{P}_1\|}$$

$$\mathbf{P}_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

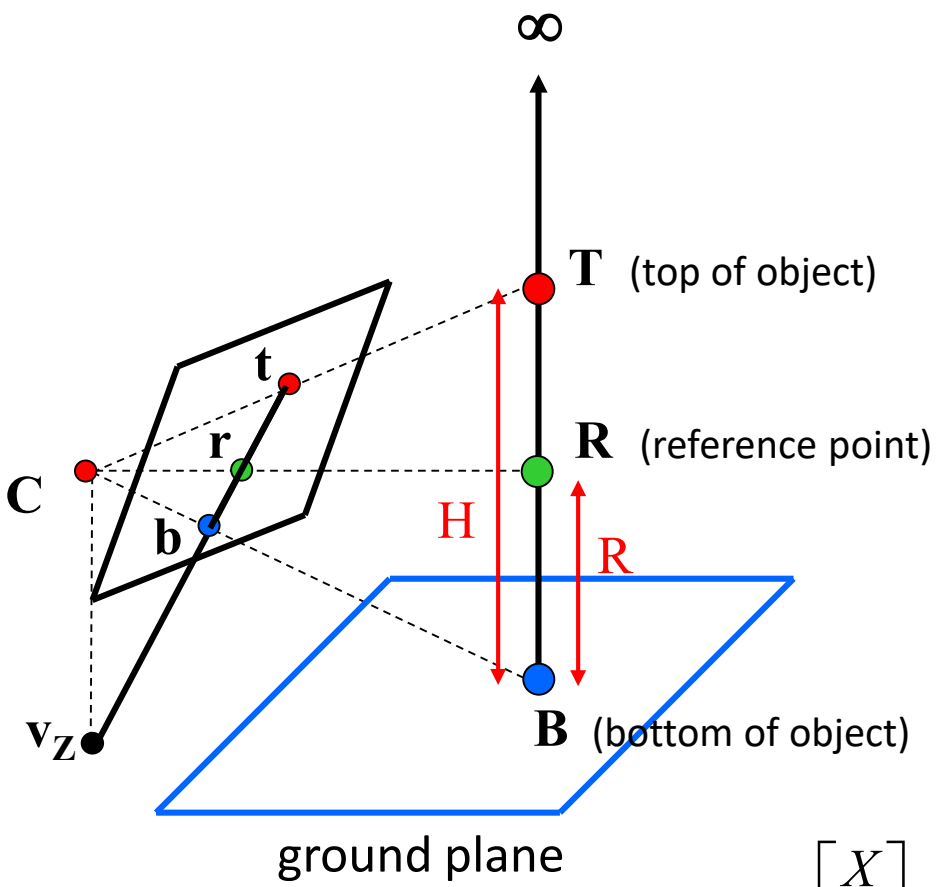
Can permute the point ordering

$$\frac{\|\mathbf{P}_1 - \mathbf{P}_3\| \, \|\mathbf{P}_4 - \mathbf{P}_2\|}{\|\mathbf{P}_1 - \mathbf{P}_2\| \, \|\mathbf{P}_4 - \mathbf{P}_3\|}$$

- 4! = 24 different orders (but only 6 distinct values)

This is the fundamental invariant of projective geometry

Measuring height



$$\frac{\|\mathbf{T}-\mathbf{B}\|\,\|\infty-\mathbf{R}\|}{\|\mathbf{R}-\mathbf{B}\|\,\|\infty-\mathbf{T}\|}=\frac{H}{R}$$

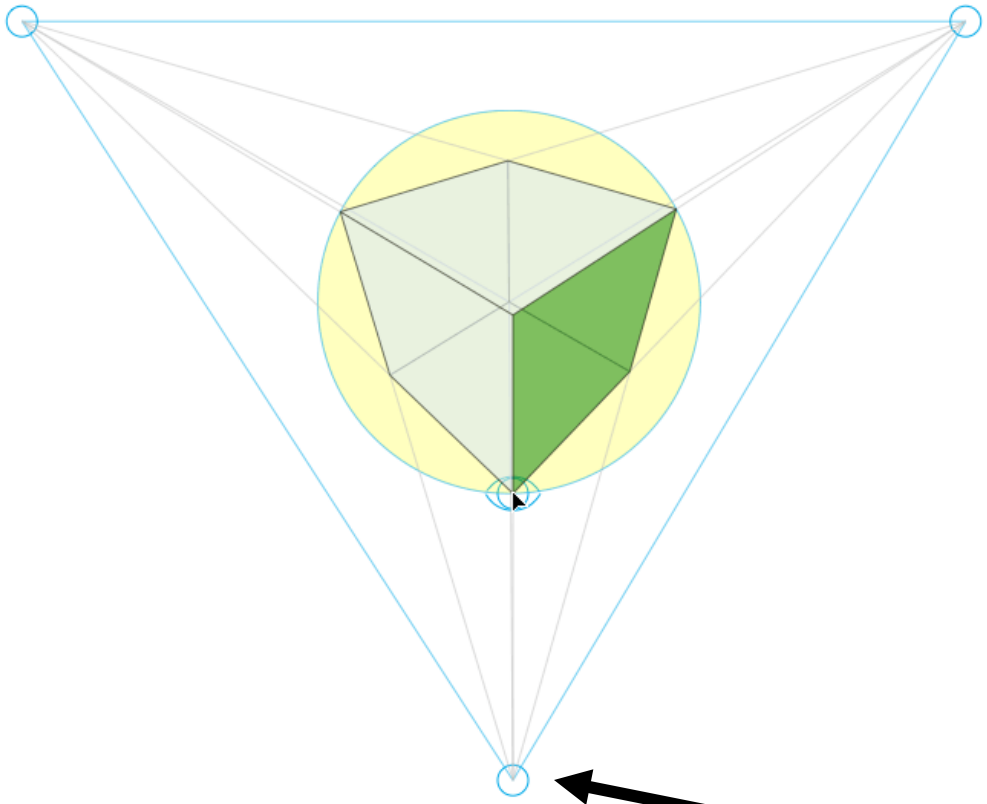
scene cross ratio

$$\frac{\|\mathbf{t}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{r}\|}{\|\mathbf{r}-\mathbf{b}\|\,\|\mathbf{v}_Z-\mathbf{t}\|}=\frac{H}{R}$$

image cross ratio

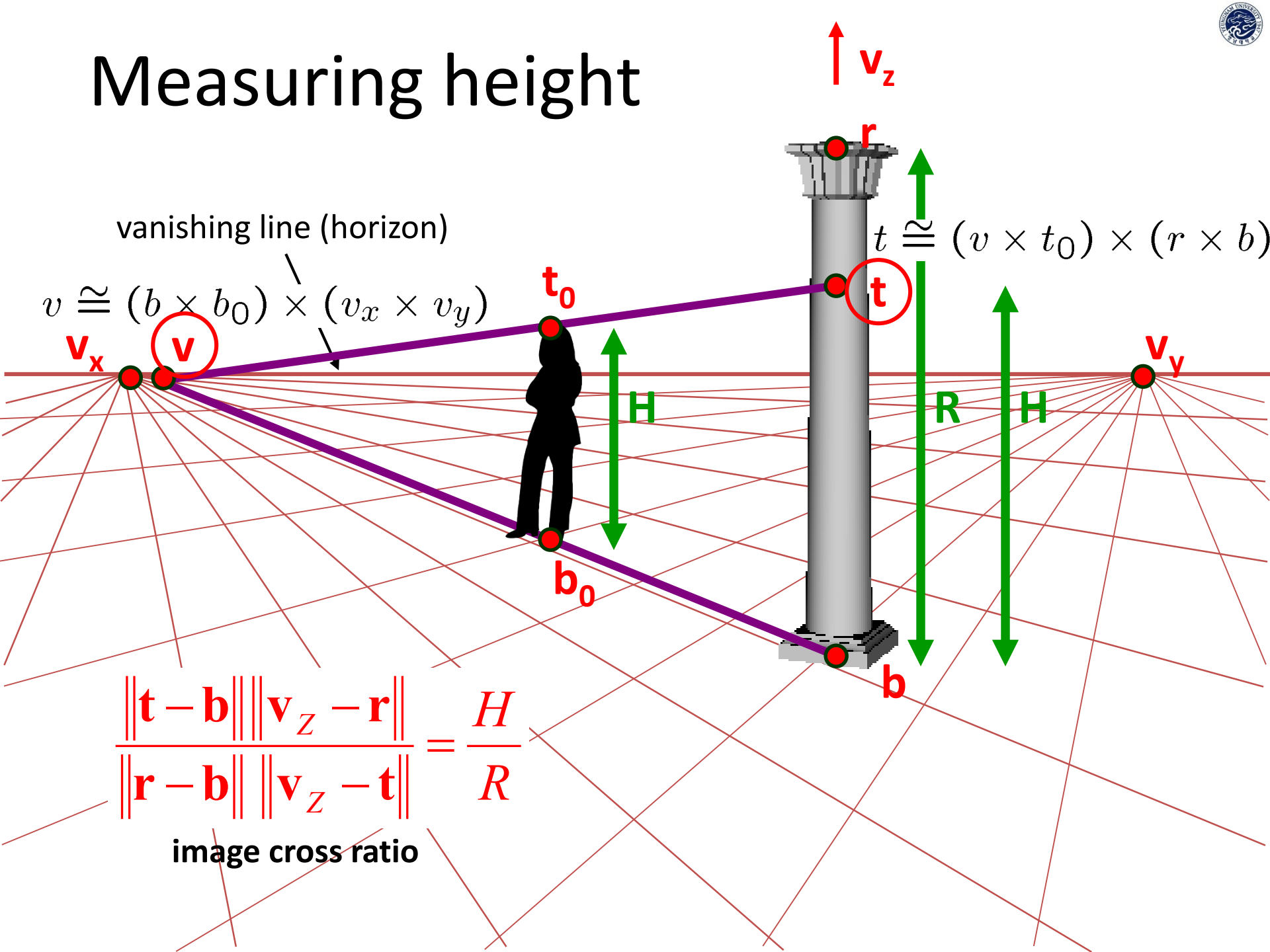
scene points represented as $\mathbf{P}=\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}$ image points as $\mathbf{p}=\begin{bmatrix}x\\y\\1\end{bmatrix}$

Finding the vertical (z) vanishing point



Find intersection of projections of vertical lines

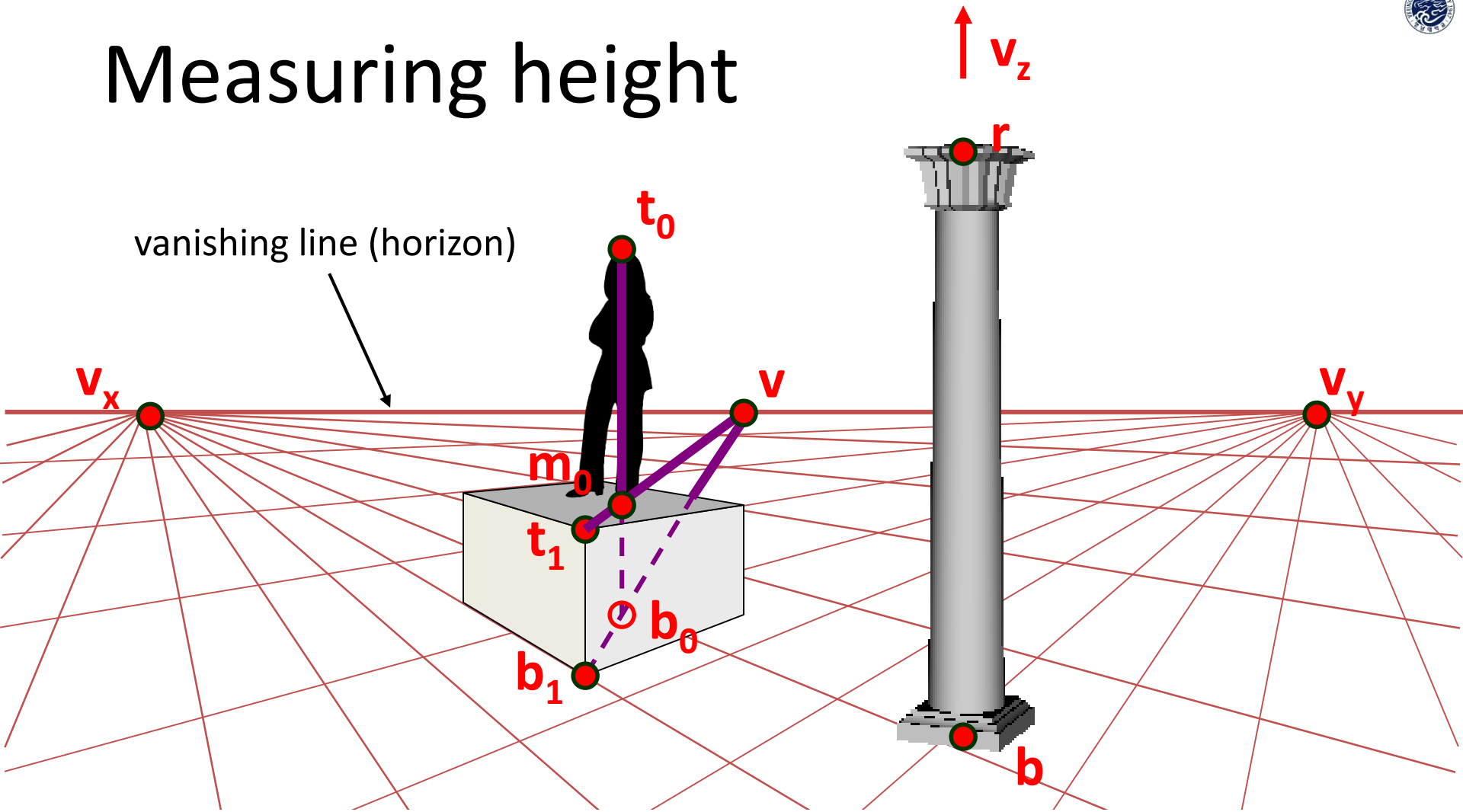
Measuring height



$$\frac{\|\mathbf{t}-\mathbf{b}\|\|\mathbf{v}_Z-\mathbf{r}\|}{\|\mathbf{r}-\mathbf{b}\|\|\mathbf{v}_Z-\mathbf{t}\|}=\frac{H}{R}$$

image cross ratio

Measuring height



What if the point on the ground plane $\mathbf{b_0}$ is not known?

- Here the person is standing on the box, height of box is known
- Use one side of the box to help find $\mathbf{b_0}$ as shown above

3D modeling from a photograph

St. Jerome in his Study, H. Steenwick

3D modeling from a photograph

3D modeling from a photograph

Flagellation, Piero della Francesca

Related problem: camera calibration

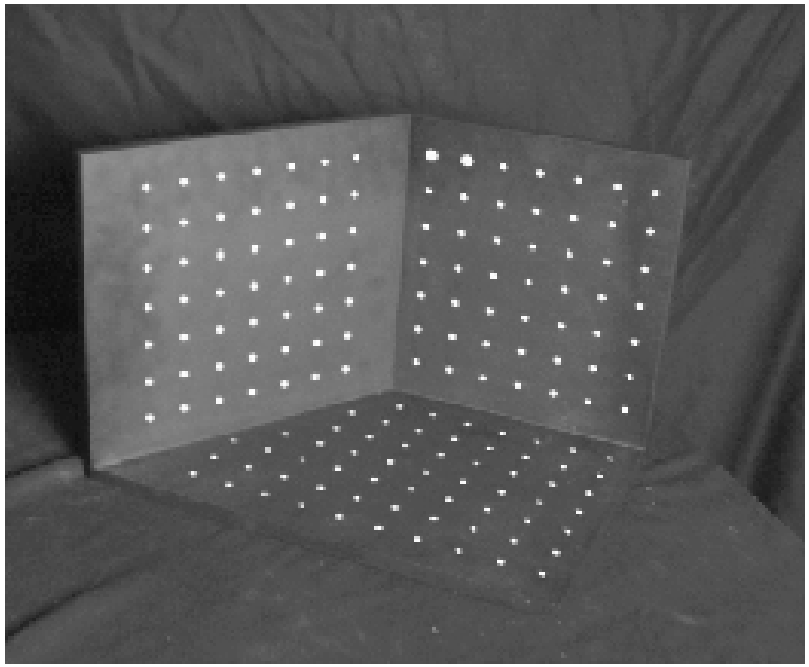
- Goal: estimate the camera parameters
  - Version 1: solve for 3x4 projection matrix

$$\mathbf{x} = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{\Pi X}$$

  - Version 2: solve for camera parameters separately
    - intrinsics (focal length, principal point, pixel size)
    - extrinsics (rotation angles, translation)
    - radial distortion

Calibration using a reference object

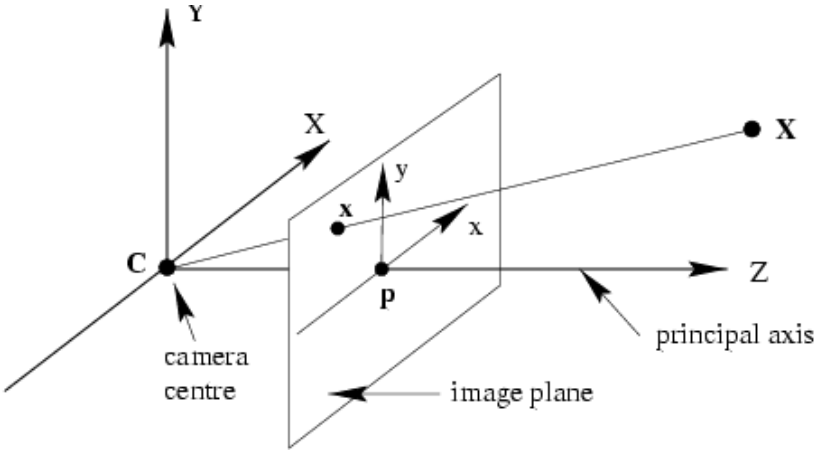
- Place a known object in the scene
  - identify correspondence between image and scene
  - compute mapping from scene to image

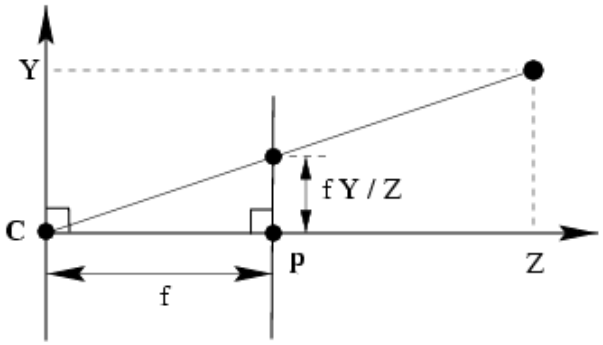


Issues

- must know geometry very accurately
- must know 3D->2D correspondence

Recall: Pinhole camera model



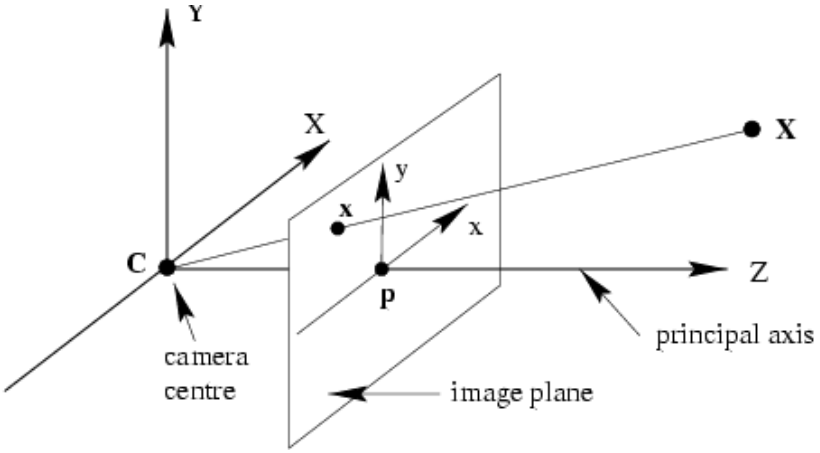


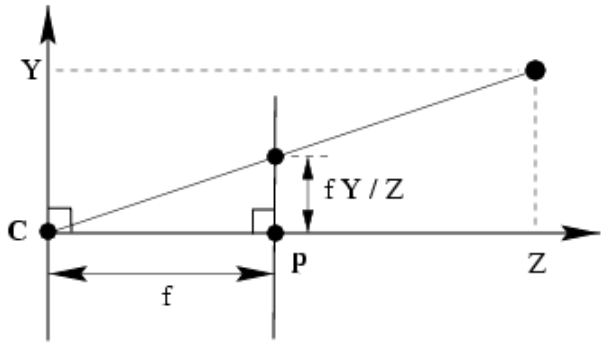
$$(X,Y,Z) \mapsto (f\,X/Z, f\,Y/Z)$$

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f\,X \\ f\,Y \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$\mathbf{x} = \mathrm{PX}$$

Pinhole camera model



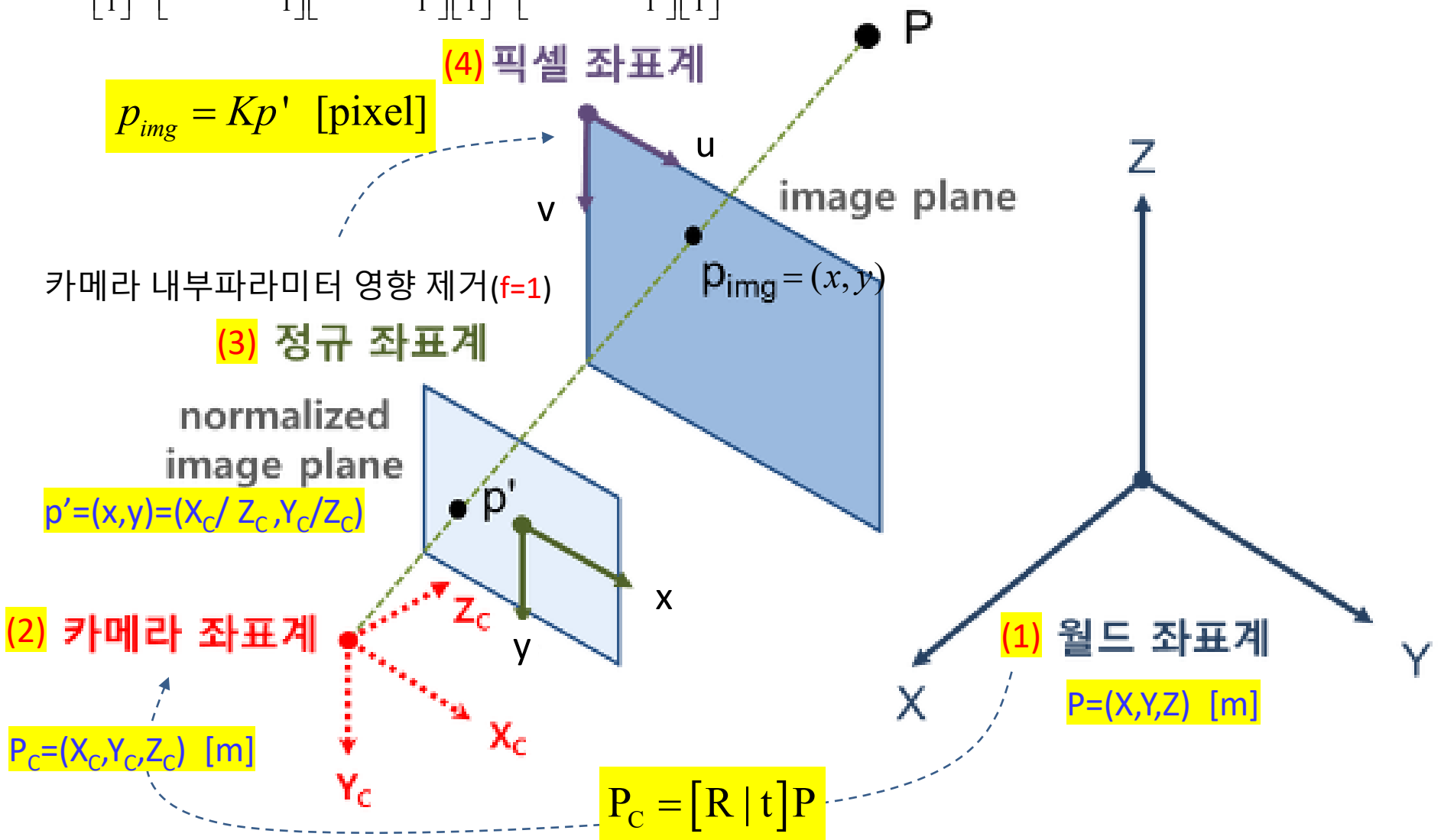


$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & \\ & f & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

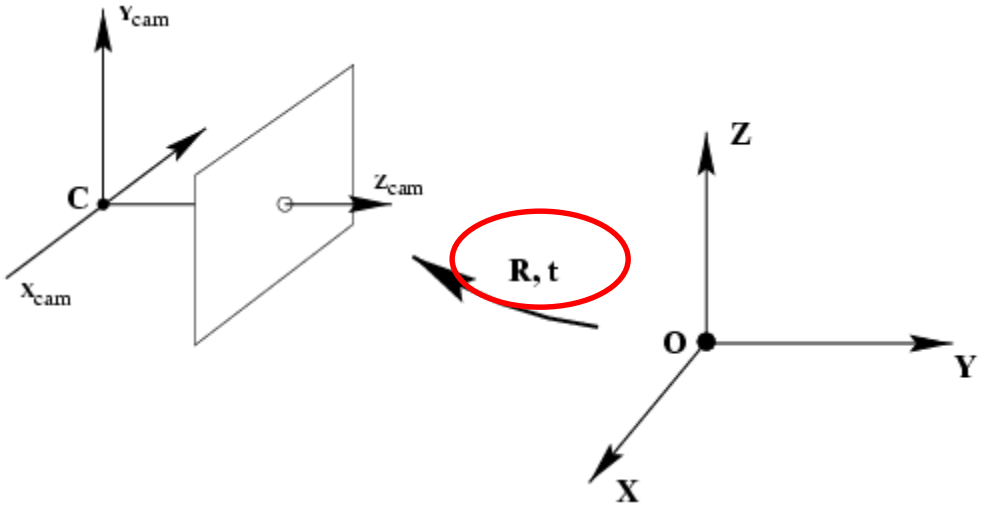
$$\mathrm{x} = \mathrm{PX} \qquad \mathrm{P} = \mathrm{diag}(f, f, 1)\left[\mathrm{I} \mid 0\right]$$

World, Camera, Normalized, Pixel Coordinates (중요)

$$p_{img} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & & \beta_x \\ & \alpha_y & \beta_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$



(1) →(2): World coord. → Camera coord. by Camera rotation and translation (extrinsic para.)



- In general, the camera coordinate frame will be related to the world coordinate frame by a rotation and a translation

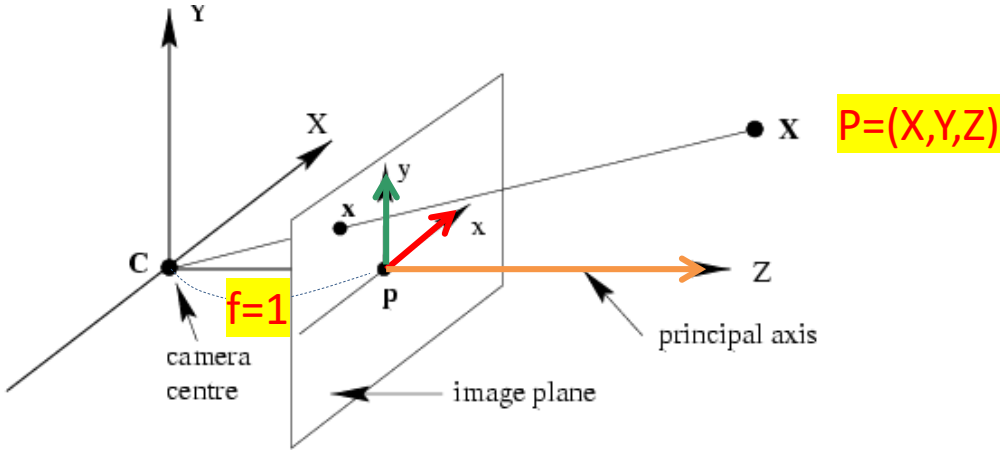
$$\tilde{X}_{cam} = R\left(\tilde{X} - \tilde{C}\right)$$

coords. of point in camera frame

coords. of a point in world frame (nonhomogeneous)

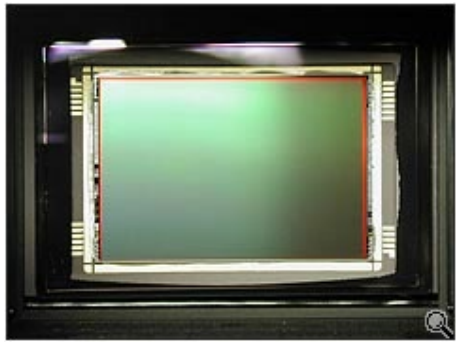
coords. of camera center in world frame

(2)→(3): Camera coord. → Normalized coord. (normalized by Z, f=1, 가상)



- **Principal axis:** line from the camera center perpendicular to the image plane
- **Normalized coordinate system:** camera center is at the origin and the principal axis is the z-axis (x,y)=(X/Z,Y/Z)
- **Principal point (p):** point where principal axis intersects the image plane (origin of normalized coordinate system)

(3)→(4): Normalized coord. → Pixel coord. (intrinsic camera parameter)



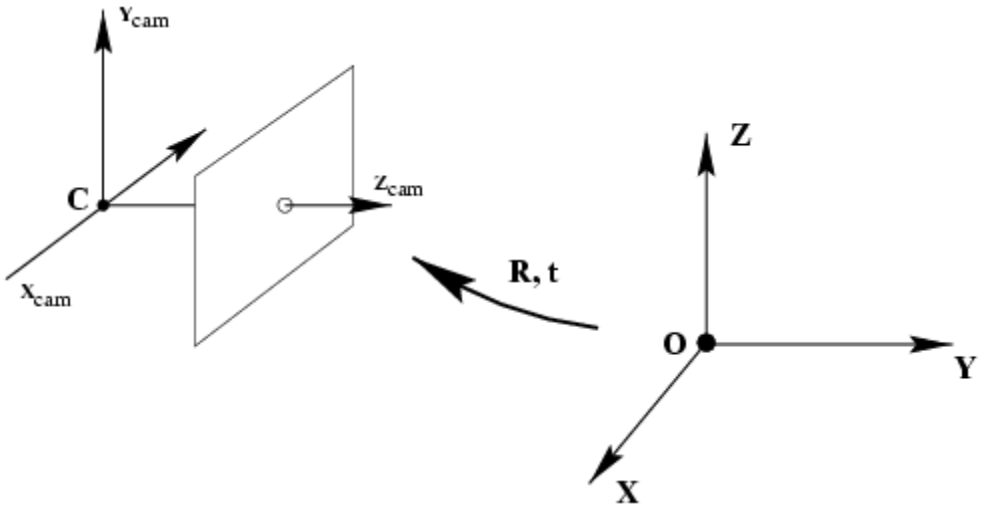
Pixel size: $\frac{1}{m_x} \times \frac{1}{m_y}$

- m_x pixels per meter in horizontal direction, m_y pixels per meter in vertical direction

$$K = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & & \beta_x \\ & \alpha_y & \beta_y \\ & & 1 \end{bmatrix}$$

pixels/m m pixels

Total: World → Pixel Coord.



In non-homogeneous coordinates:

$$\tilde{X}_{cam} = R\left(\tilde{X} - \tilde{C}\right)$$

$$X_{cam} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \tilde{X} \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} X$$

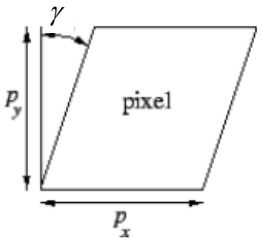
$$x = K[I \mid 0]X_{cam} = K[R \mid -R\tilde{C}]X \qquad P = K[R \mid t], \qquad t = -R\tilde{C}$$

Note: C is the null space of the camera projection matrix (PC=0)

Camera parameters

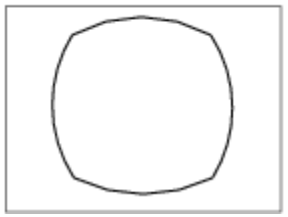
- Intrinsic parameters
  - Principal point coordinates (p_x, p_y)
  - Focal length (f)
  - Pixel magnification factors (m_x, m_y)
  - *Skew : non-rectangular pixels* (γ), usually 0
  - *Radial distortion*

$$K = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & \text{skew} & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & \gamma & \beta_x \\ & \alpha_y & \beta_y \\ & & 1 \end{bmatrix}$$

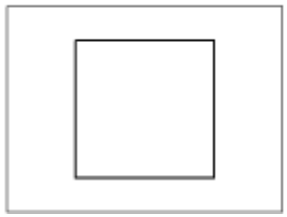


radial distortion

linear image

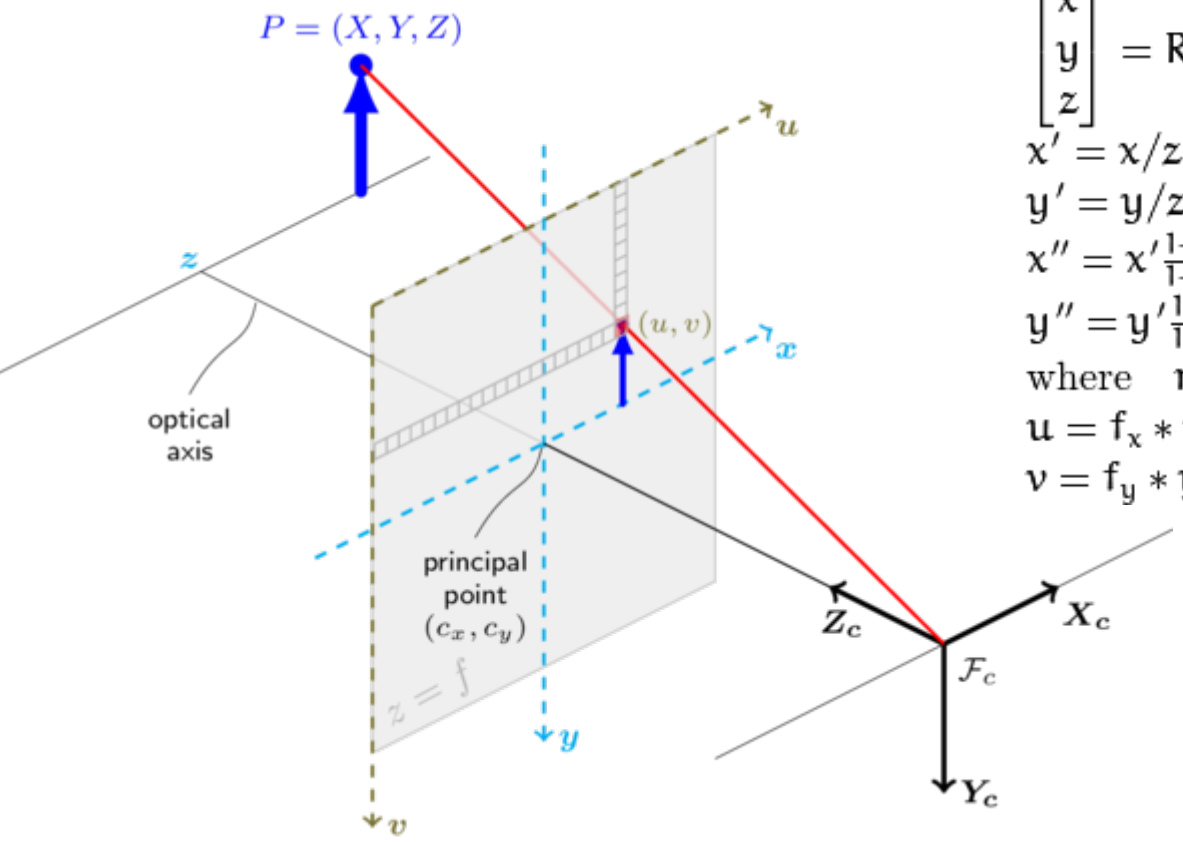


correction



Distortion Correction Summary

(x,y): normalized corrdinates



$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = x/z$$

$$y' = y/z$$

$$x'' = x' \frac{1+k_1 r^2+k_2 r^4+k_3 r^6}{1+k_4 r^2+k_5 r^4+k_6 r^6} + 2p_1 x' y' + p_2(r^2 + 2x'^2)$$

$$y'' = y' \frac{1+k_1 r^2+k_2 r^4+k_3 r^6}{1+k_4 r^2+k_5 r^4+k_6 r^6} + p_1(r^2 + 2y'^2) + 2p_2 x' y'$$

where $r^2 = x'^2 + y'^2$

$$u = f_x * x'' + c_x$$

$$v = f_y * y'' + c_y$$

k_1, k_2, k_3, k_4, k_5, and k_6 are radial distortion coefficients. p_1 and p_2 are tangential distortion coefficients. Higher-order coefficients are not considered in OpenCV.

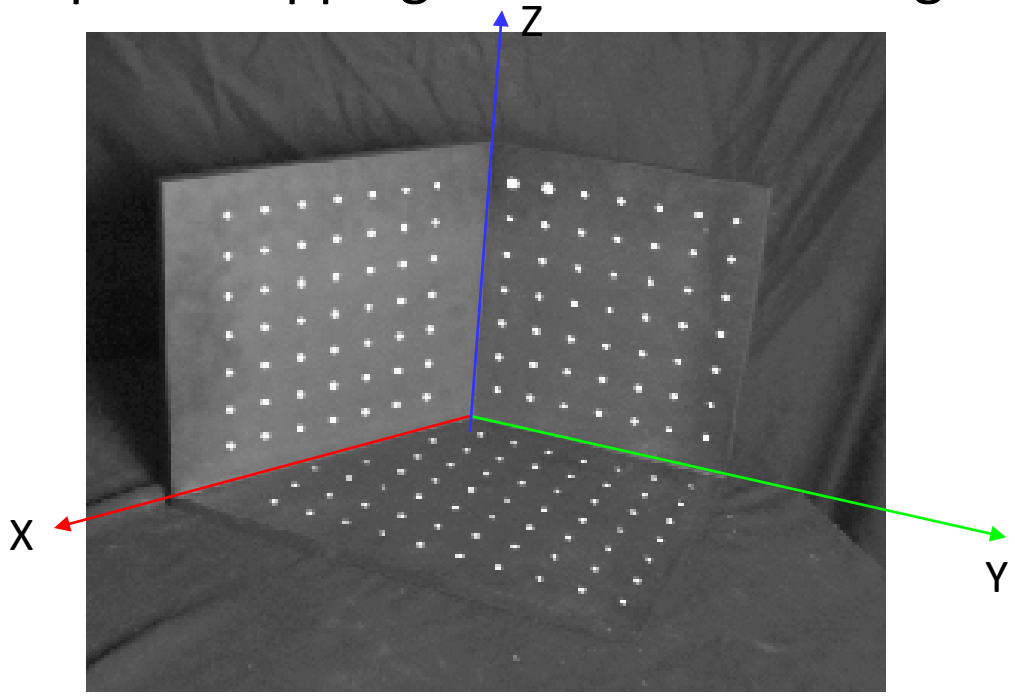
https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

Camera parameters

- Intrinsic parameters
  - Principal point coordinates
  - Focal length
  - Pixel magnification factors
  - *Skew (non-rectangular pixels)*
  - *Radial distortion*
- Extrinsic parameters
  - Rotation (R) and translation (t) relative to world coordinate system

Estimating the projection matrix

- Place a known object in the scene
  - identify correspondence between image and scene
  - compute mapping from scene to image



$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \cong \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

Direct linear calibration

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \cong \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

$$u_i = \frac{m_{00}X_i + m_{01}Y_i + m_{02}Z_i + m_{03}}{m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}}$$

$$v_i = \frac{m_{10}X_i + m_{11}Y_i + m_{12}Z_i + m_{13}}{m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}}$$

$$u_i(m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}) = m_{00}X_i + m_{01}Y_i + m_{02}Z_i + m_{03}$$
$$v_i(m_{20}X_i + m_{21}Y_i + m_{22}Z_i + m_{23}) = m_{10}X_i + m_{11}Y_i + m_{12}Z_i + m_{13}$$

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_iX_i & -u_iY_i & -u_iZ_i & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_iX_i & -v_iY_i & -v_iZ_i & -v_i \end{bmatrix} \begin{bmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{03} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{13} \\ m_{20} \\ m_{21} \\ m_{22} \\ m_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Direct linear calibration $\mathrm{Ap} = 0$

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ & & & & & & \vdots & & & & & \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{03} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{13} \\ m_{20} \\ m_{21} \\ m_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Can solve for m_{ij} by linear least squares

- use eigenvector trick that we used for homographies

- P has 11 degrees of freedom (12 parameters, but scale is arbitrary)
- One 2D/3D correspondence gives us two linearly independent equations
- Homogeneous least squares
- 6 correspondences needed for a minimal solution

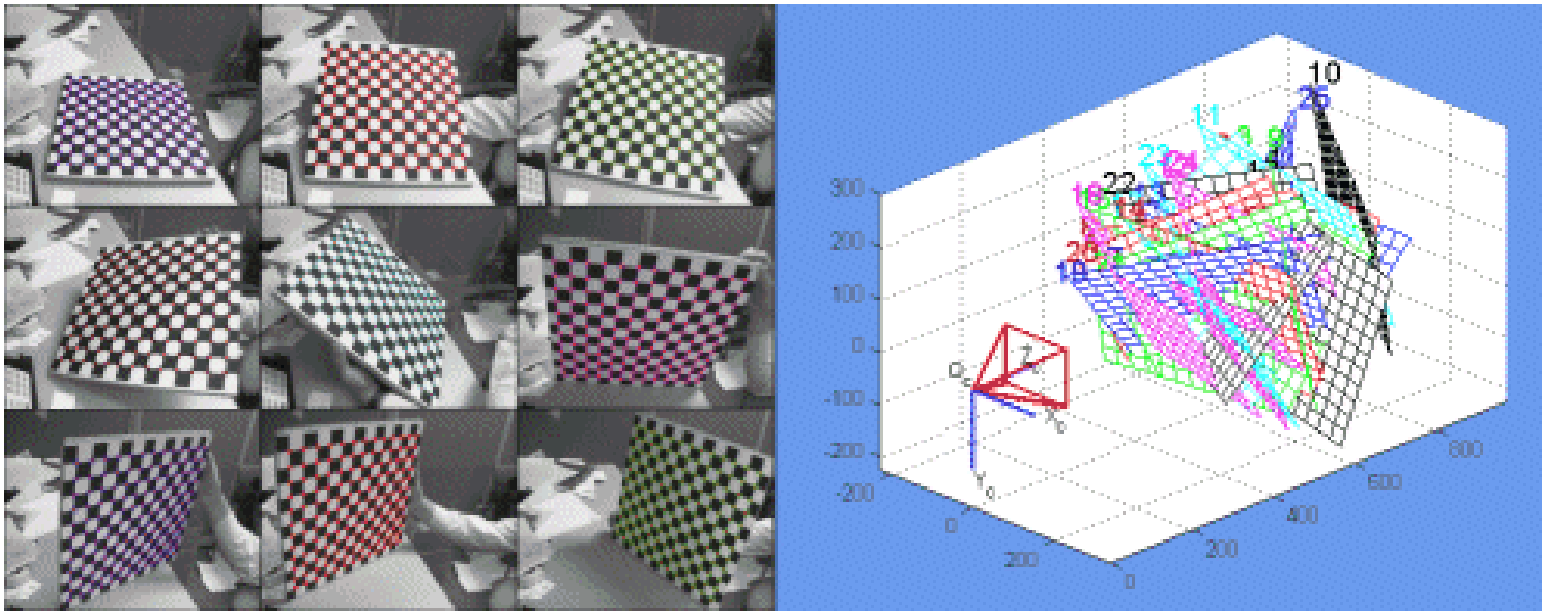
Direct linear calibration

- Advantage:
  - Very simple to formulate and solve
- Disadvantages:
  - Doesn't tell you the camera parameters
  - Doesn't model radial distortion
  - Hard to impose constraints (e.g., known *f*)
  - Doesn't minimize the right error function

For these reasons, *nonlinear methods* are preferred

- Define error function E between projected 3D points and image positions
  - E is nonlinear function of intrinsics, extrinsics, radial distortion
- Minimize E using nonlinear optimization techniques

Alternative: multi-plane calibration

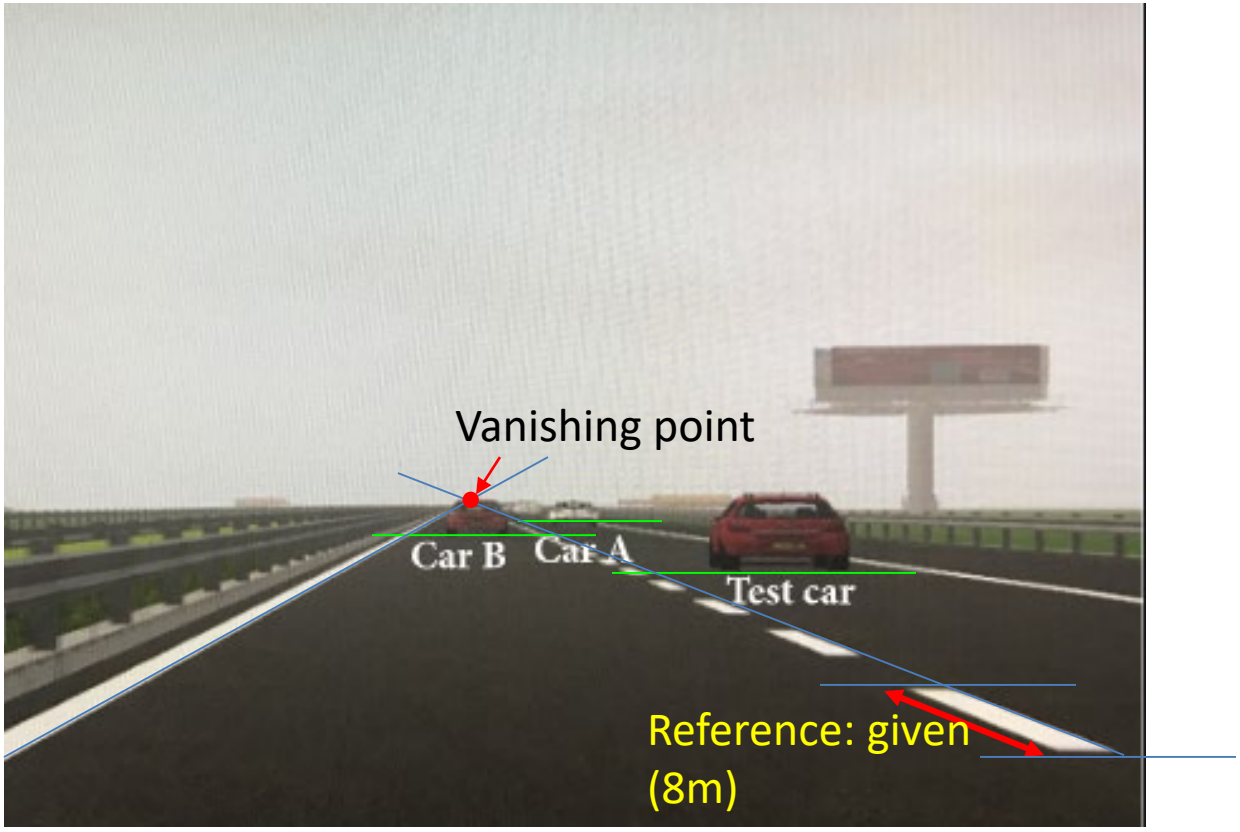


Advantage

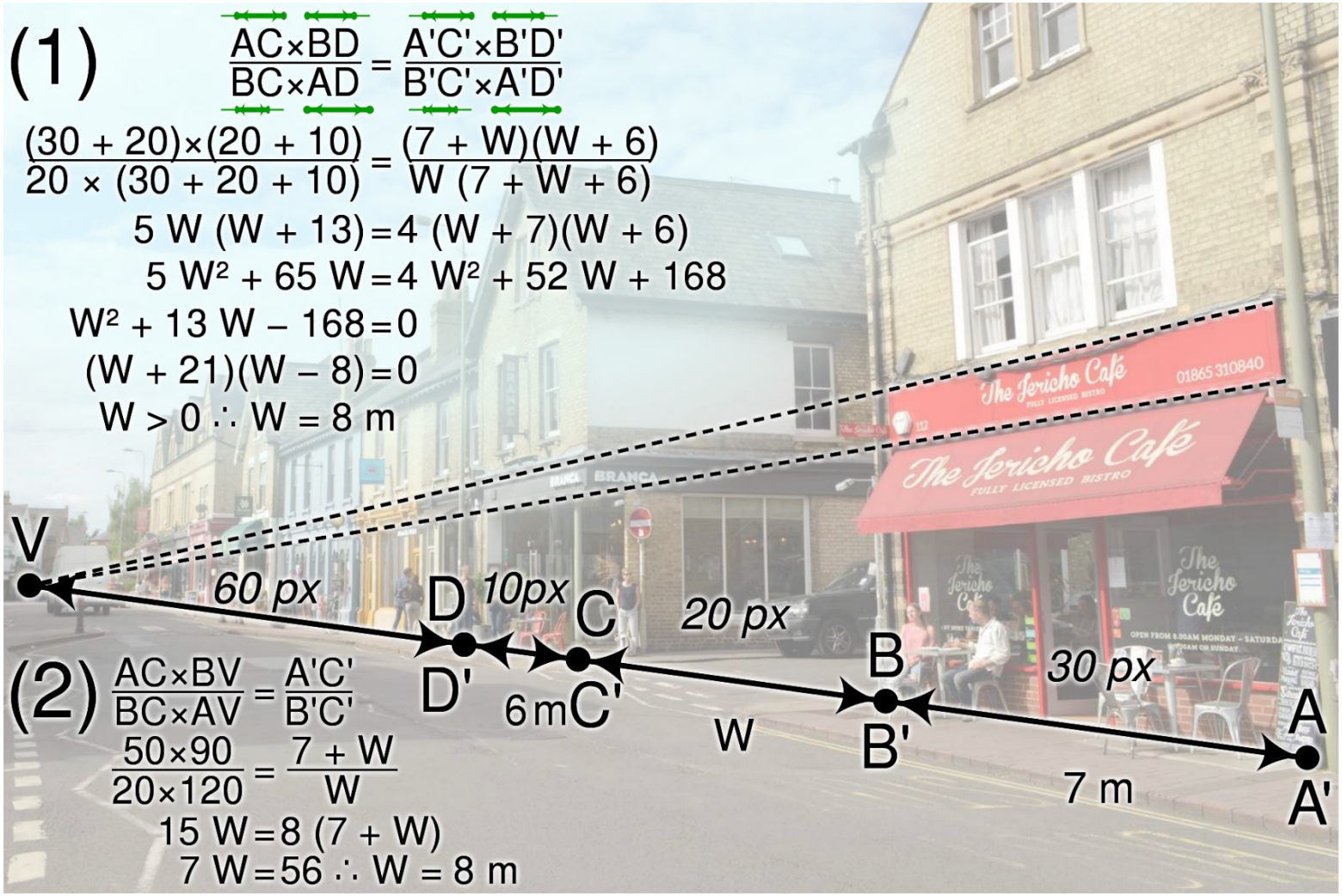
- Only requires a plane
- Don't have to know positions/orientations
- Good code available online! (including in OpenCV)
  - Matlab version by Jean-Yves Bouget: (recommend)
    http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
  - Zhengyou Zhang's web site: http://research.microsoft.com/~zhang/Calib/

Measuring Car Distance on Road

- Method: using vanishing point and cross ratio
  - Find distances of Test car, Car A, Car B



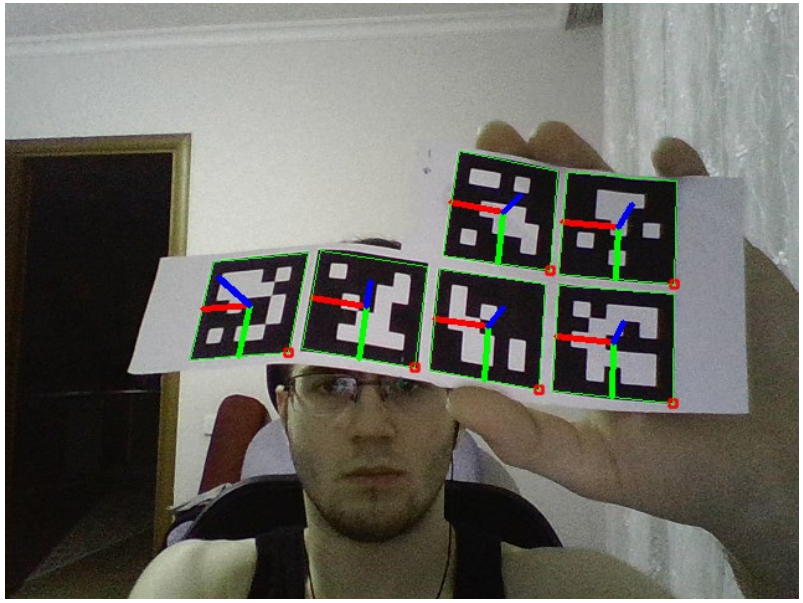
Hint


(1) $\frac{AC \times BD}{BC \times AD} = \frac{A'C' \times B'D'}{B'C' \times A'D'}$
$\frac{(30 + 20) \times (20 + 10)}{20 \times (30 + 20 + 10)} = \frac{(7 + W)(W + 6)}{W\,(7 + W + 6)}$
$5\,W\,(W + 13) = 4\,(W + 7)(W + 6)$
$5\,W^2 + 65\,W = 4\,W^2 + 52\,W + 168$
$W^2 + 13\,W - 168 = 0$
$(W + 21)(W - 8) = 0$
$W > 0 \therefore W = 8$ m
V
60 px
D
10px
C
20 px
B
30 px
A
D'
6m
C'
W
B'
7 m
A'
(2) $\frac{AC \times BV}{BC \times AV} = \frac{A'C'}{B'C'}$
$\frac{50 \times 90}{20 \times 120} = \frac{7 + W}{W}$
$15\,W = 8\,(7 + W)$
$7\,W = 56 \therefore W = 8$ m

What can we do with single camera?

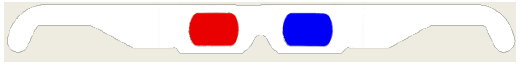
- For single camera (single-view geometry), we can use known scale objects for distance estimation
  - Example) Markers

https://docs.opencv.org/master/d5/dae/tutorial_aruco_detection.html



STEREO CAMERA

Stereo Geomtry



http://www.rainbowsymphony.com/freestuff.html

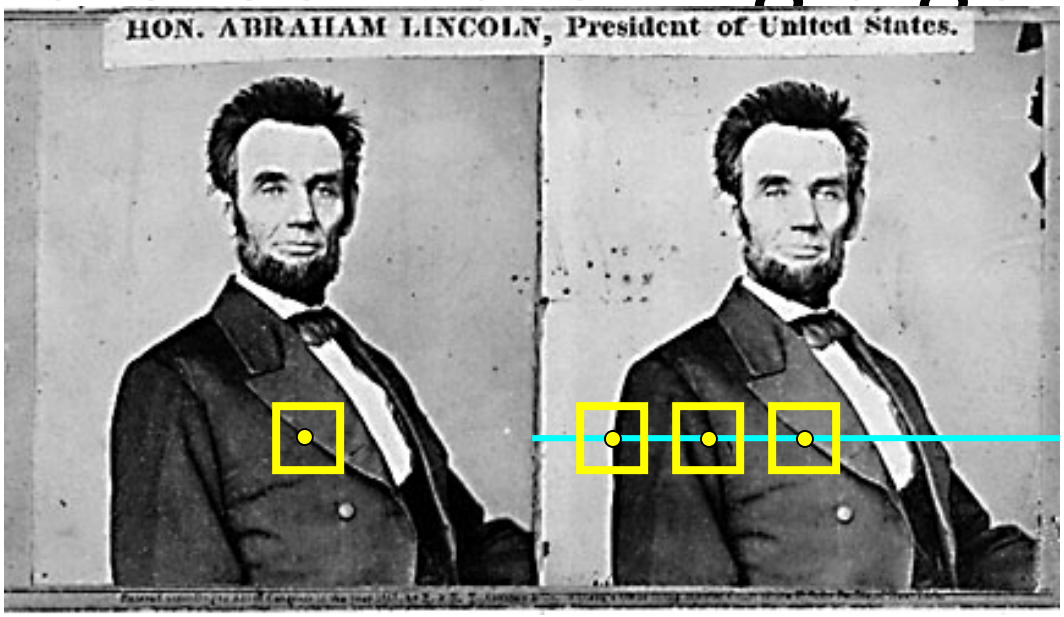
(Wikipedia for images)

Many slides adapted from Steve Seitz

Real World with Stereo Vision

https://www.youtube.com/watch?v=P-QEnnMHYt0

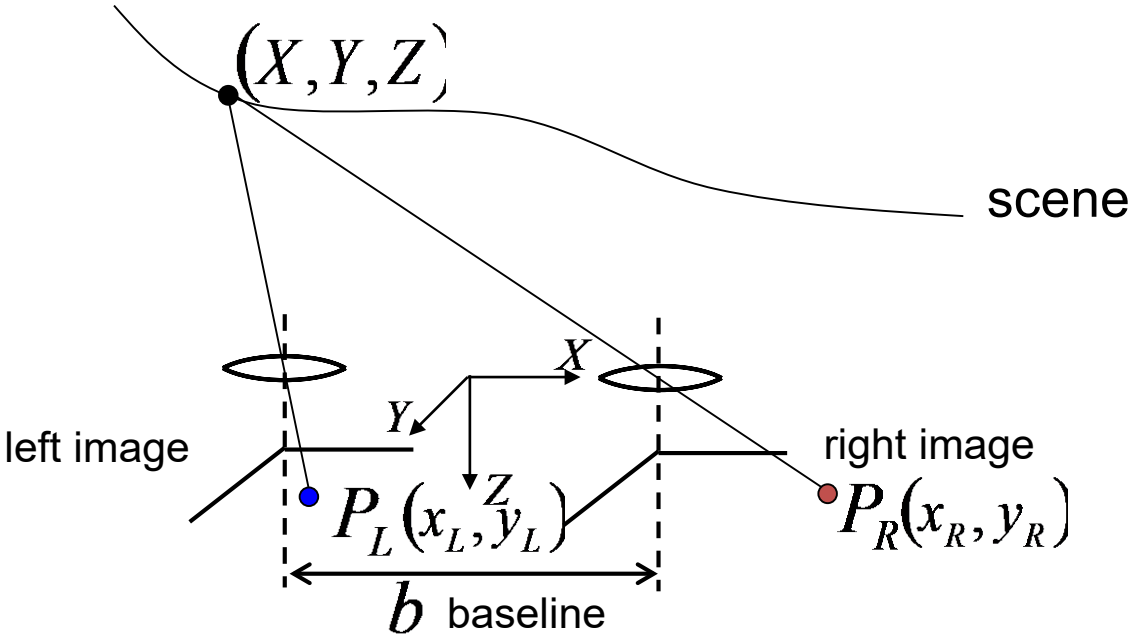
Basic stereo matching algorithm



- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information

- Simplest case: epipolar lines are scanlines
  - When does this happen?

Disparity and Depth

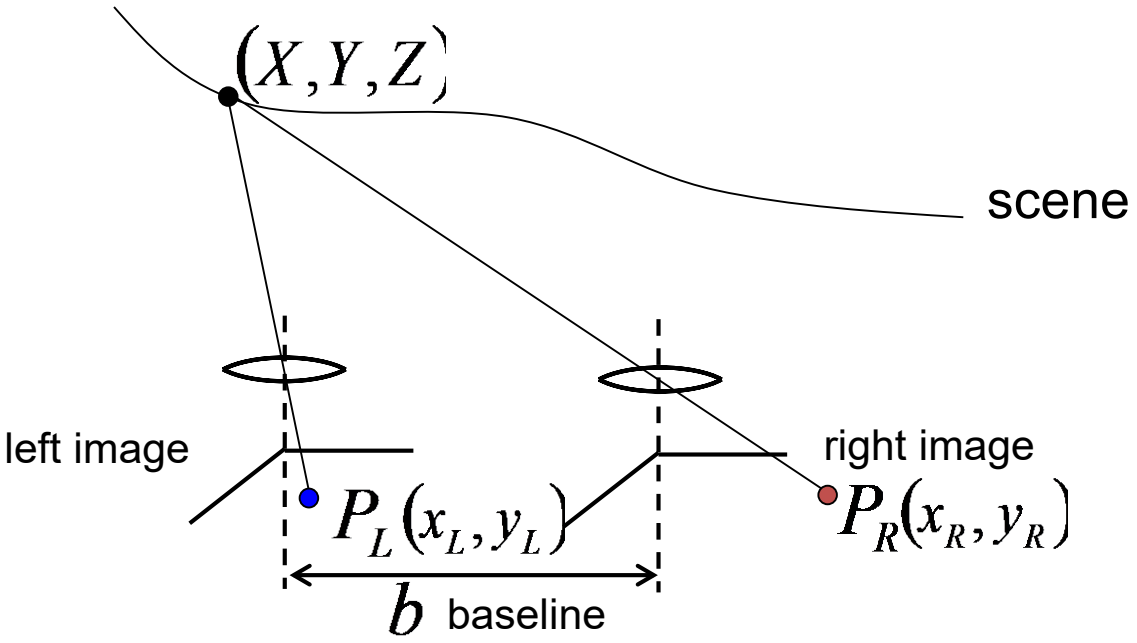


Assume that we know P_L corresponds to P_R

From perspective projection (define the coordinate system as shown above)

$$\frac{x_L}{f} = \frac{X + b/2}{Z} \qquad \frac{x_R}{f} = \frac{X - b/2}{Z} \qquad \frac{y_L}{f} = \frac{y_R}{f} = \frac{Y}{Z}$$

Disparity and Depth



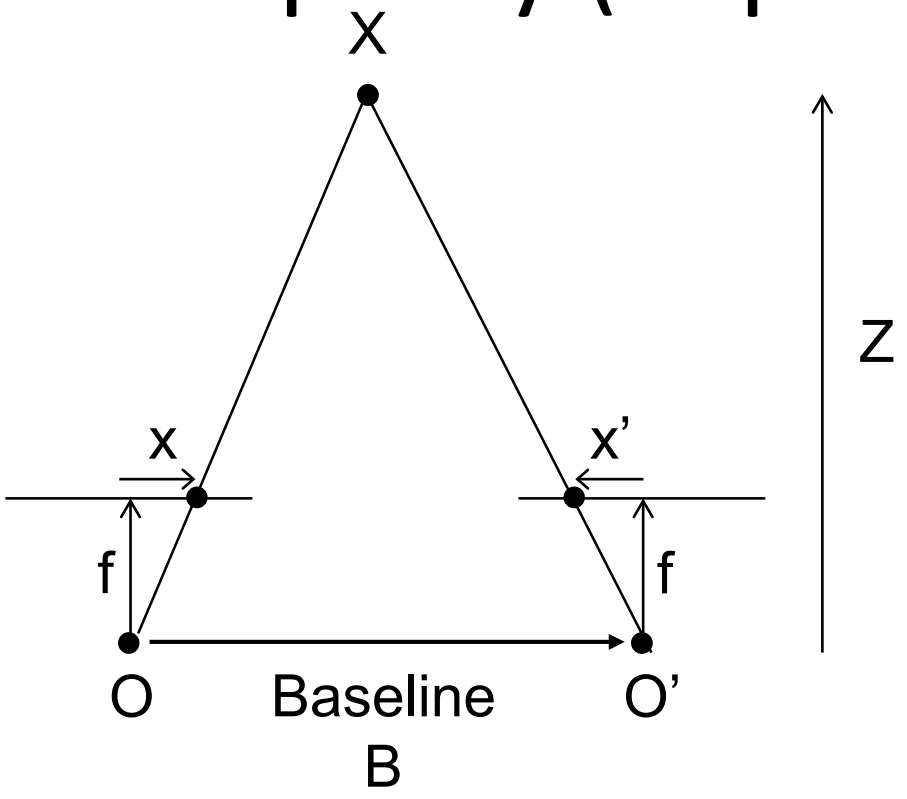
$$\frac{x_L}{f} = \frac{X + b/2}{Z} \qquad \frac{x_R}{f} = \frac{X - b/2}{Z} \qquad \frac{y_L}{f} = \frac{y_R}{f} = \frac{Y}{Z}$$

$$\Rightarrow \quad X = \frac{b(x_L + x_R)}{2(x_L - x_R)} \qquad Y = \frac{b(y_L + y_R)}{2(x_L - x_R)} \qquad \boxed{Z = \frac{bf}{(x_L - x_R)}}$$

$d = x_L - x_R$ is the **disparity** between corresponding left and right image points

- **inverse proportional to depth**
- **disparity increases with baseline b**

Depth from disparity (depth only)

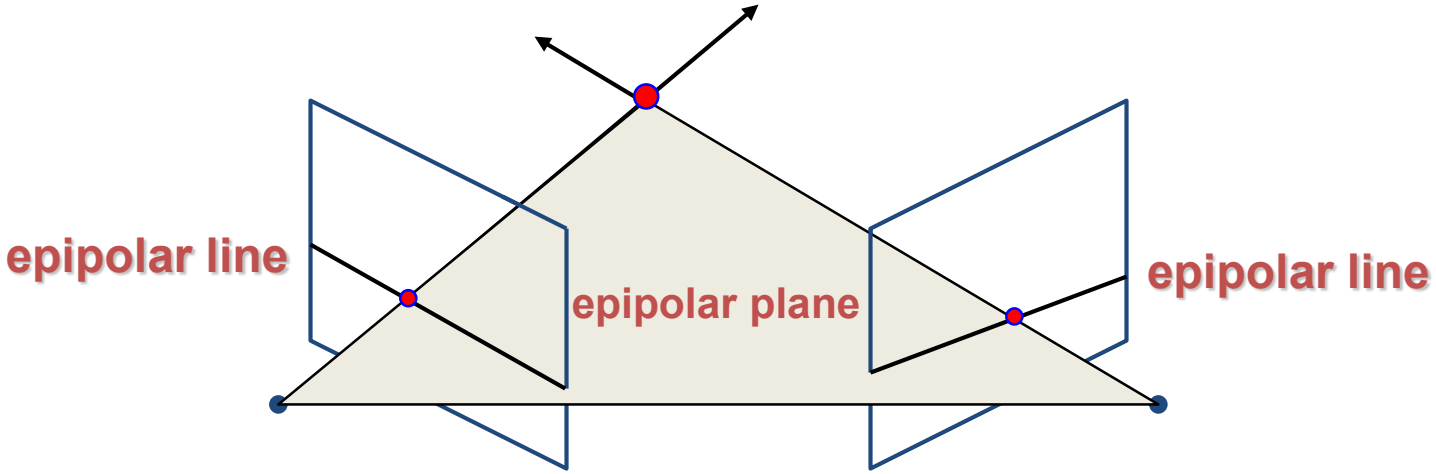


$$disparity = x - x' = \frac{B \cdot f}{Z}$$

Disparity is inversely proportional to depth!

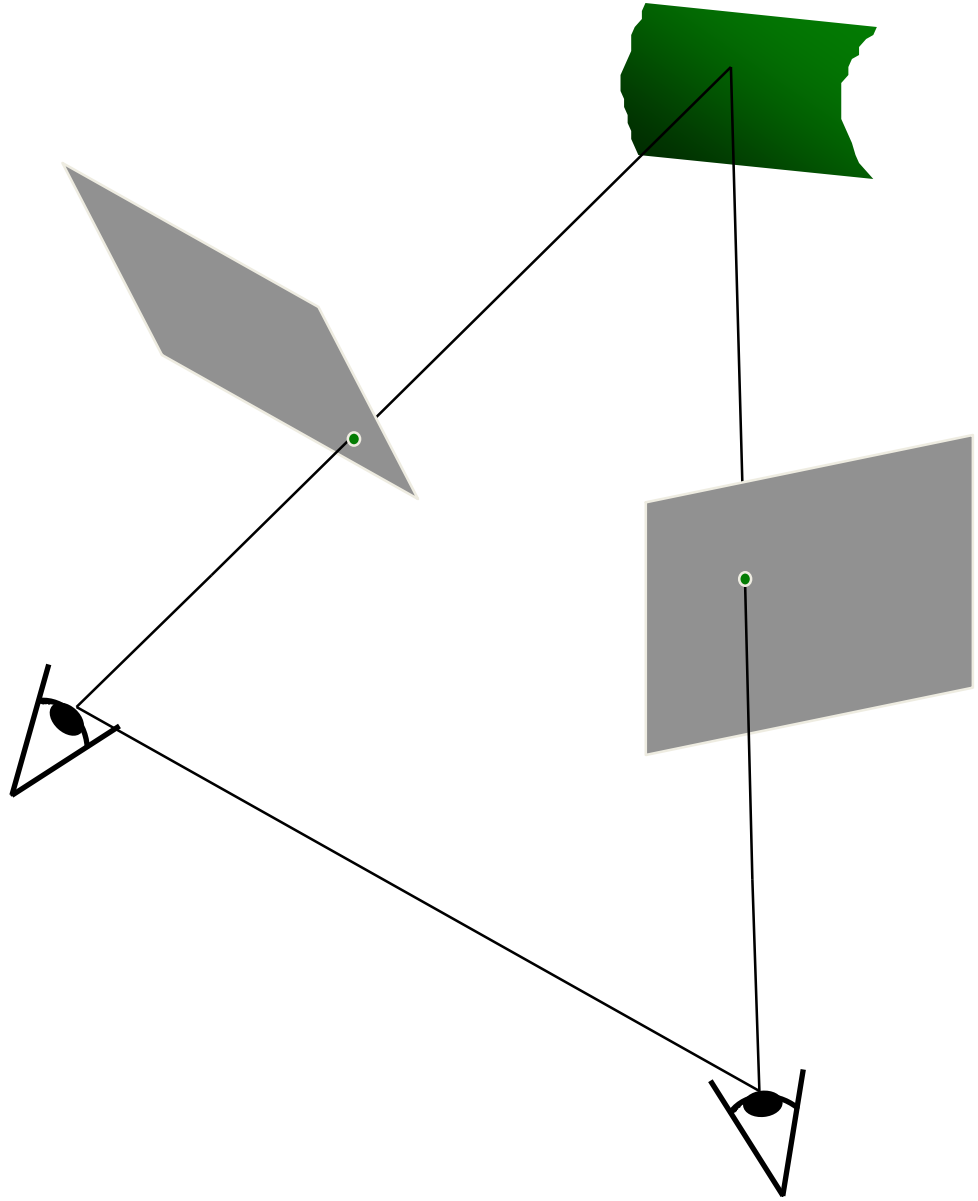
Stereo Correspondence

- Determine Pixel Correspondence
  - Pairs of points that correspond to same scene point

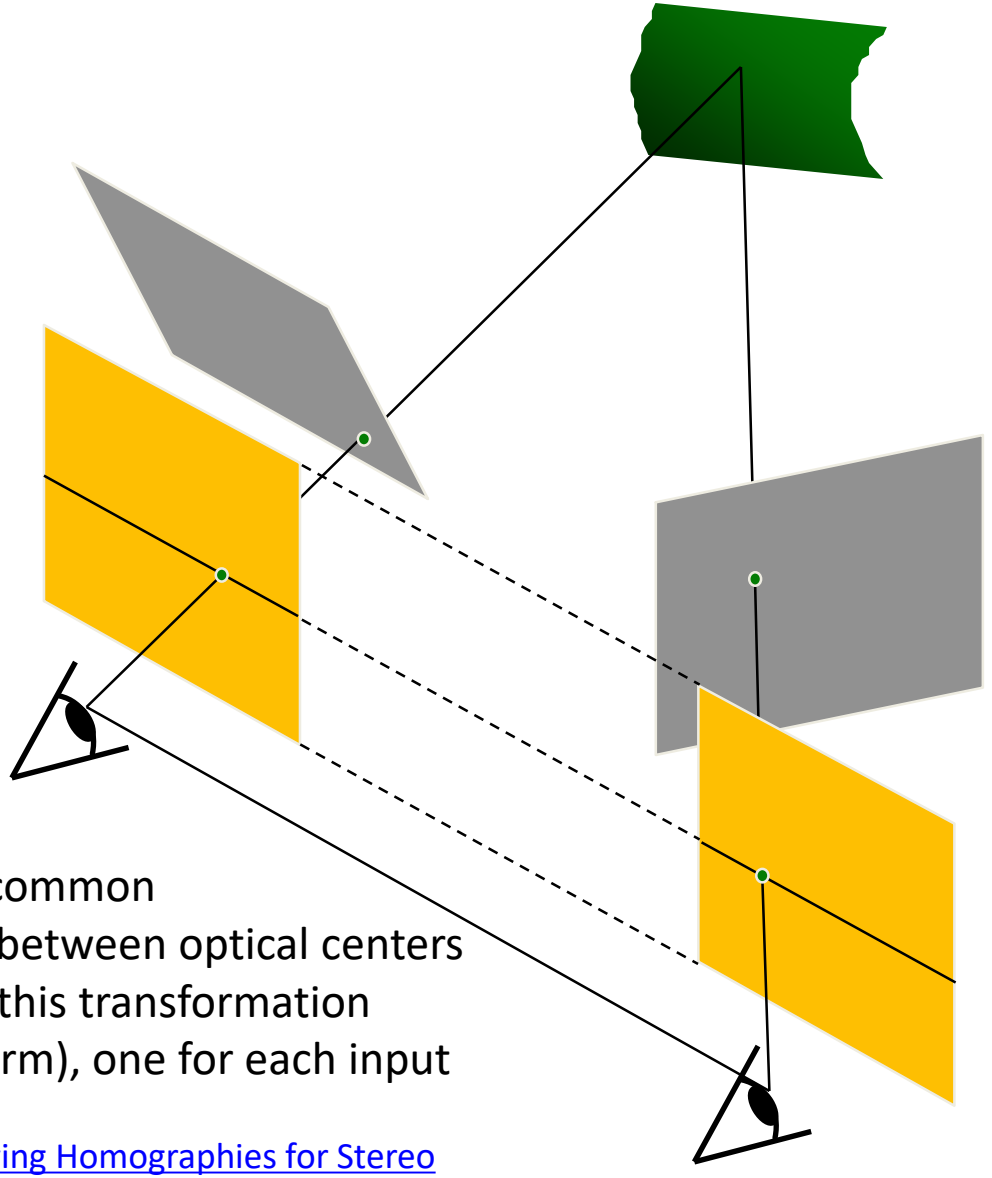


- Epipolar Constraint
  - Reduces correspondence problem to 1D search along *conjugate epipolar lines*
  - Java demo: http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html

Stereo image rectification



Stereo image rectification



- reproject image planes onto a common
- plane parallel to the line between optical 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.

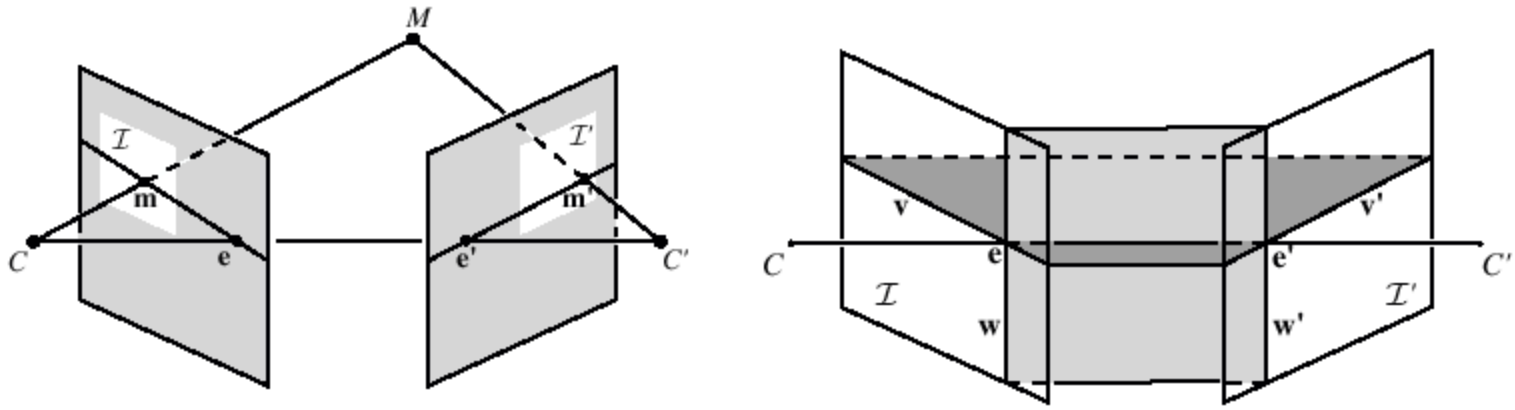
Stereo: epipolar geometry

- for *two* images (or images with collinear camera centers), can find epipolar lines
- epipolar lines are the projection of the *pencil* of planes passing through the centers

- **Rectification:** warping the input images (perspective transformation) so that epipolar lines are horizontal

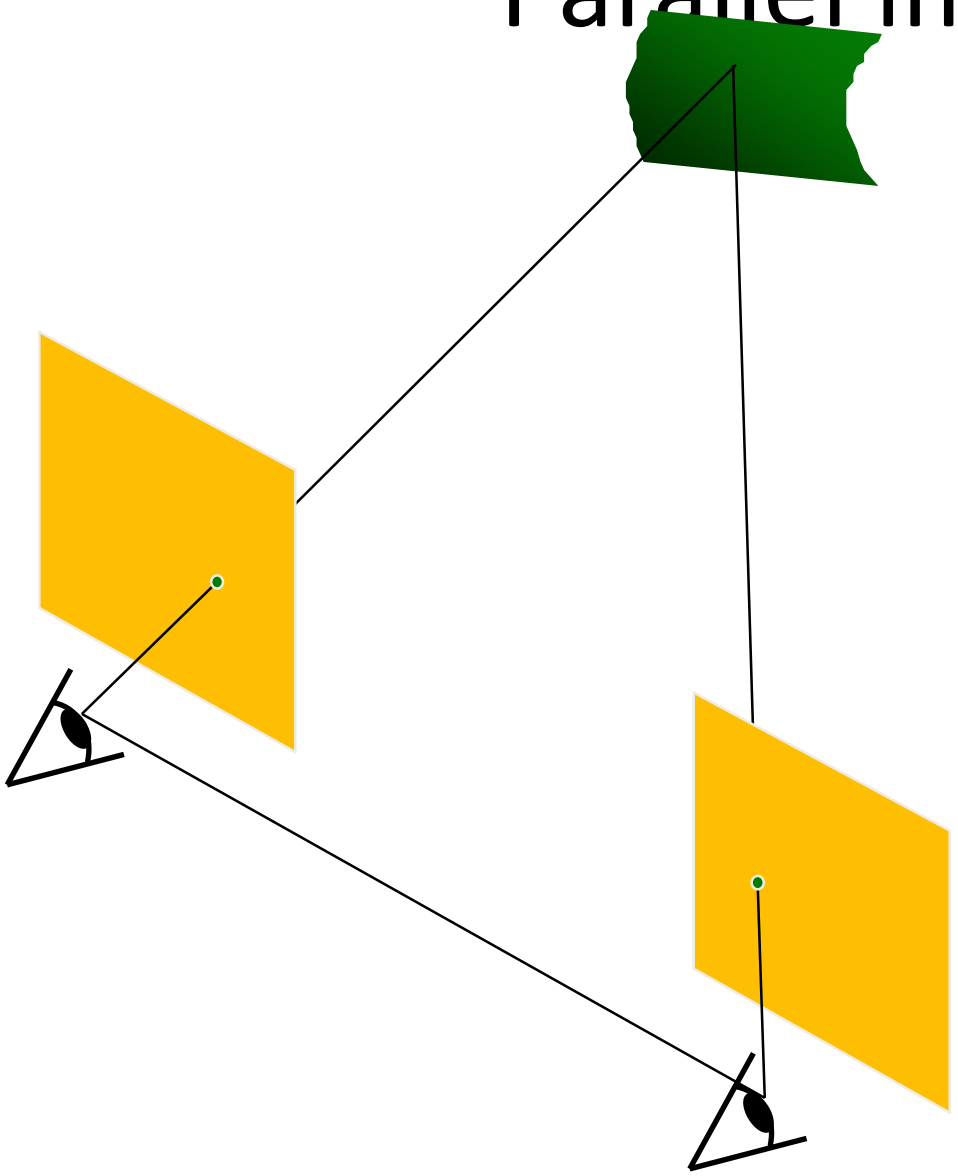
Rectification

- Project each image onto same plane, which is parallel to the epipole
- Resample lines (and shear/stretch) to place lines in correspondence, and minimize distortion



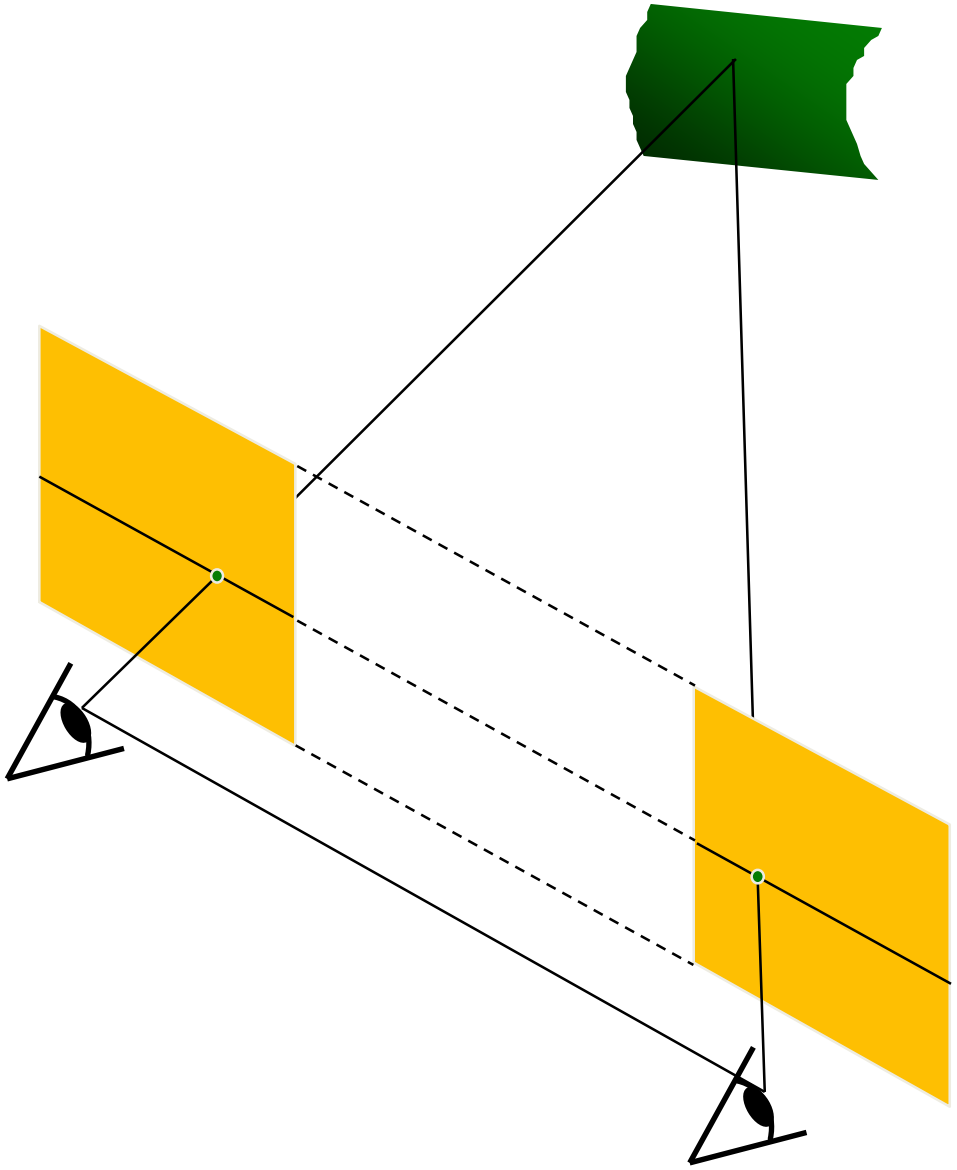
- [Loop and Zhang, CVPR 99]

Parallel images



- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same

Simplest Case: Parallel images



- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same
- Then, epipolar lines fall along the horizontal scan lines of the images

Binocular stereo

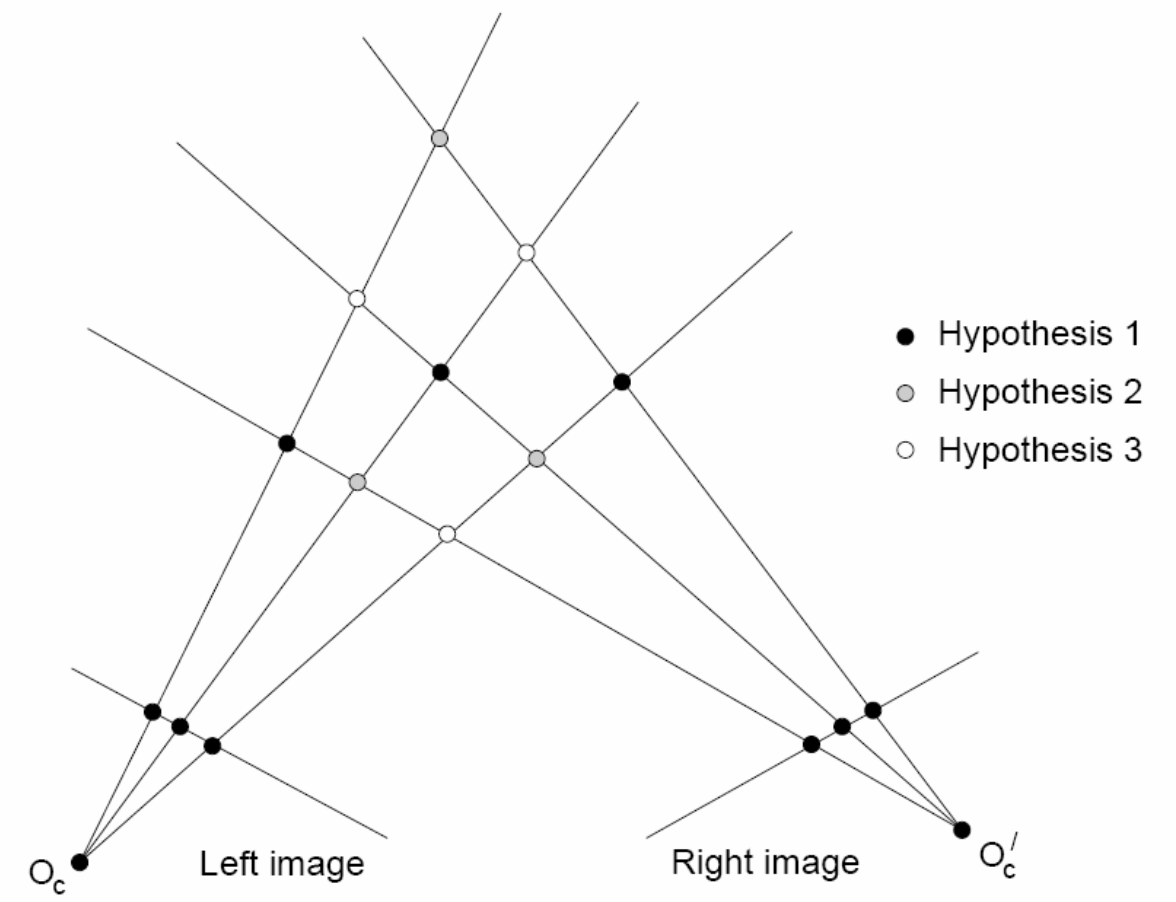
- Given a calibrated (rectified) binocular stereo pair, fuse it to produce a depth image

image 1

image 2

Dense depth map

Correspondence problem

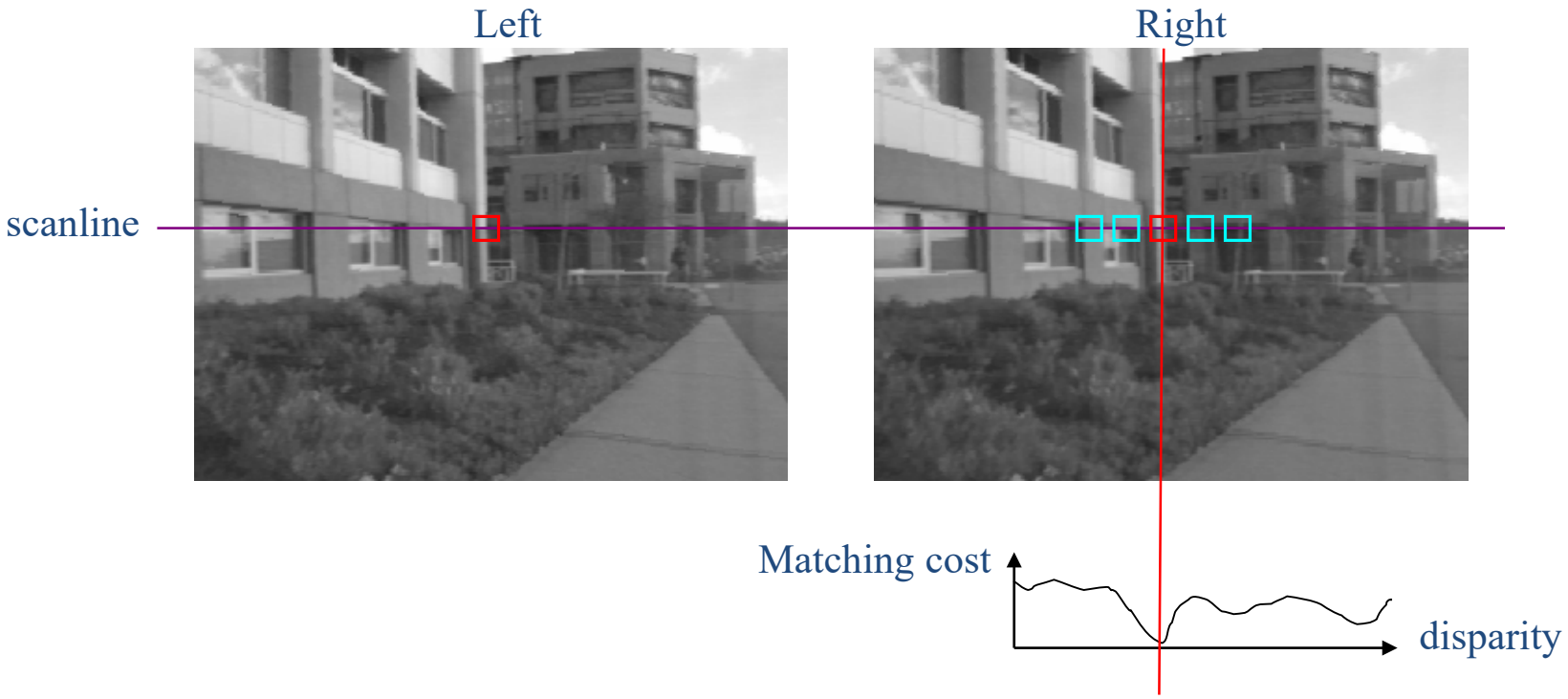


- Multiple matching hypotheses satisfy the epipolar constraint, but which one is correct?

Correspondence problem

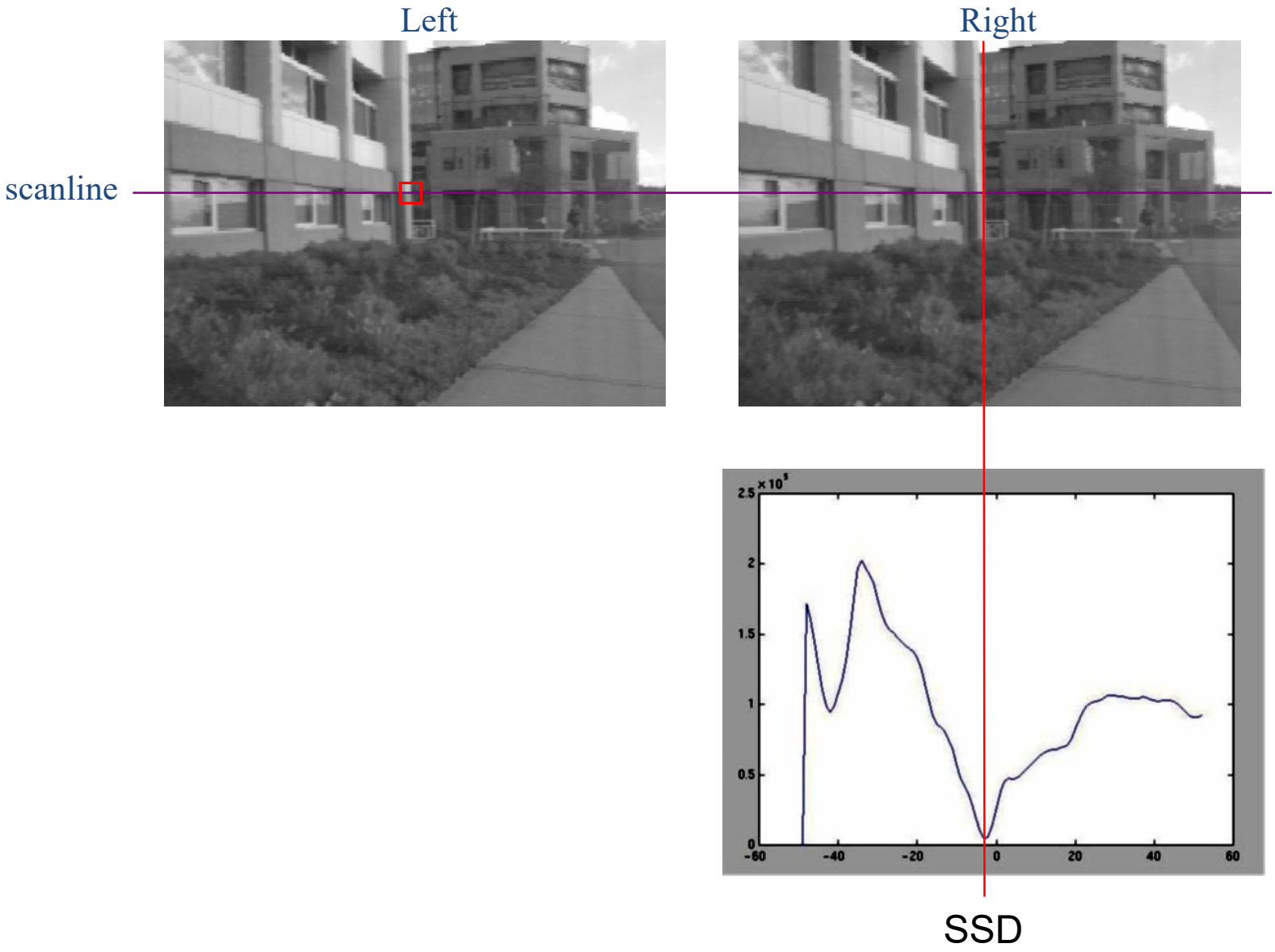
- Let's make some assumptions to simplify the matching problem
  - The baseline is relatively small (compared to the depth of scene points)
  - Then most scene points are visible in both views
  - Also, matching regions are similar in appearance

Correspondence search with similarity constraint

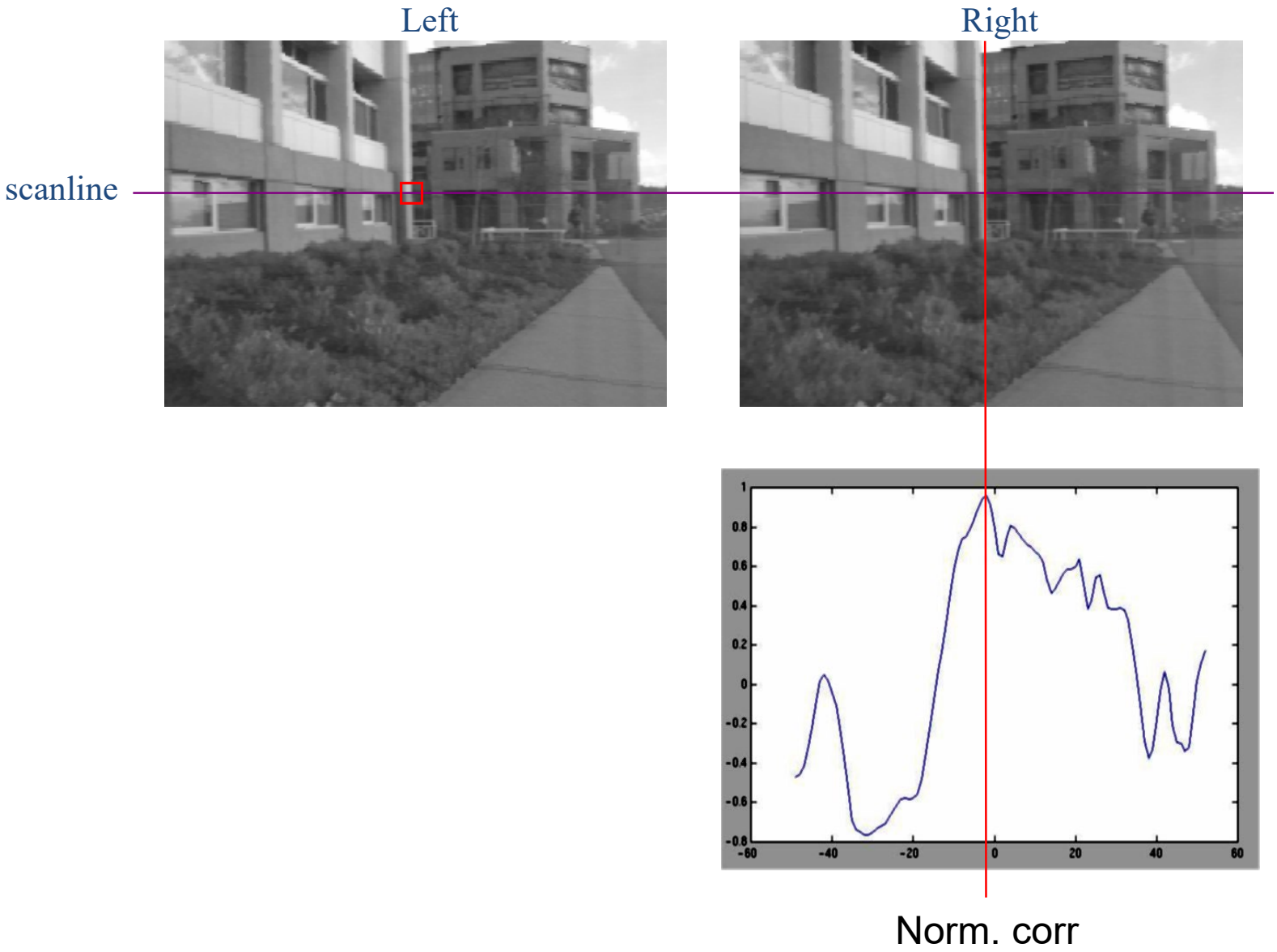


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search with similarity constraint

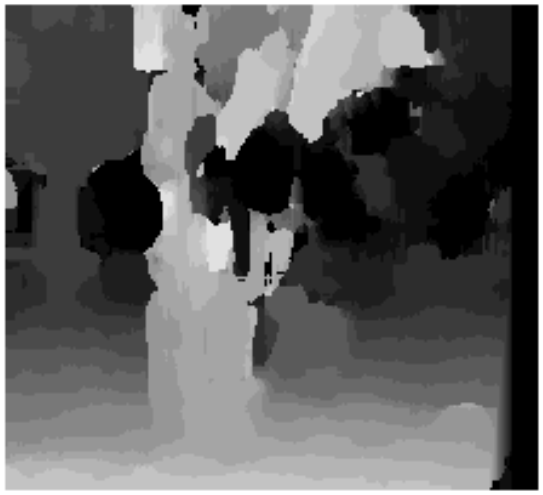


Correspondence search with similarity constraint



Effect of window size

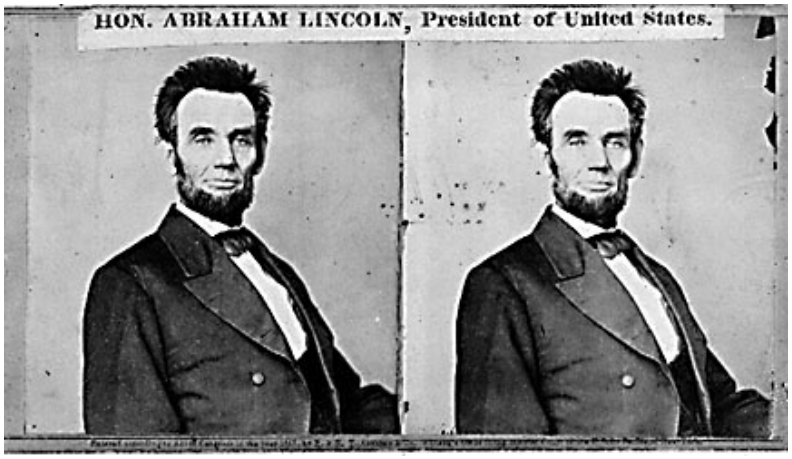
W = 3



W = 20

- Smaller window
  - + More detail
  - • More noise
- Larger window
  - + Smoother disparity maps
  - • Less detail

Limitations of similarity constraint



Textureless surfaces

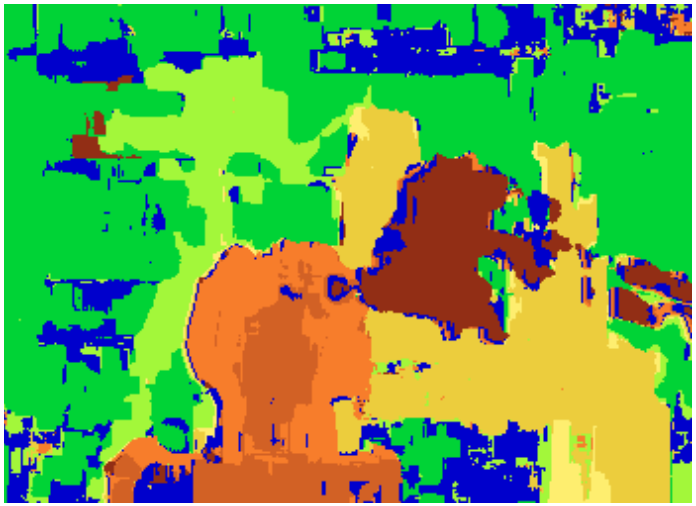
Occlusions, repetition

Non-Lambertian surfaces, specularities

Results with window search

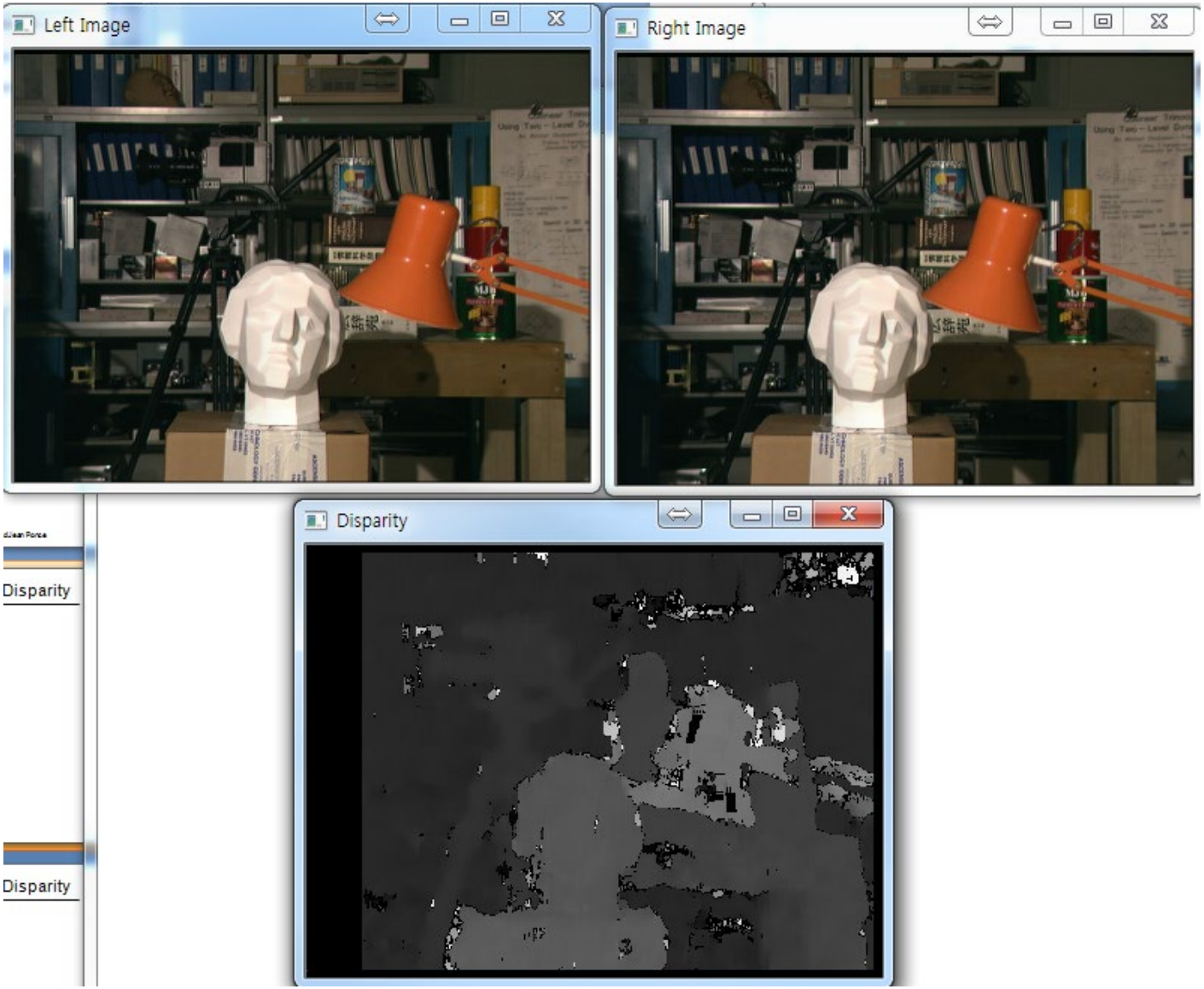
Data

Window-based matching



Ground truth

Sample code. Block matching based Disparity



http://vision.deis.unibo.it/~smatt/stereo_smp.html

Block matching based Disparity

```
1  import numpy as np
2  import cv2
3  from matplotlib import pyplot as plt
4
5  imgL = cv2.imread('scene1.row3.col1.ppm',0)
6  imgR = cv2.imread('scene1.row3.col2.ppm',0)
7
8  cv2.imshow('Left Image',imgL)
9  cv2.imshow('Right Image', imgR)
10
11 stereo = cv2.StereoBM_create(numDisparities=16, blockSize=15)
12 disparity = stereo.compute(imgL,imgR)
13 plt.imshow(disparity,'gray')
14 plt.colorbar()
15 plt.show()
16
17 cv2.waitKey(0)
18 cv2.destroyAllWindows()
19
```