Camera Geometry

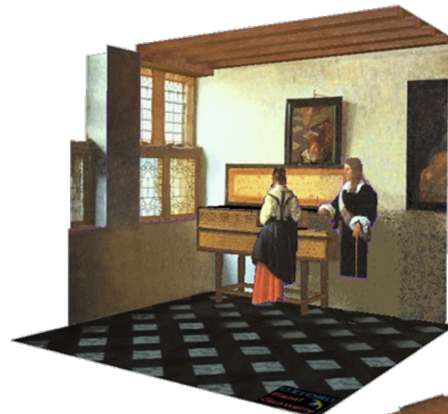
A. Mantegna, *Martyrdom of St. Christopher*, c. 1450

Overview

- Motivation: recovery of 3D structure
- Pinhole projection model
- Properties of projection
- Perspective projection matrix
- Orthographic projection

Given an image, can we recover 3D structure?

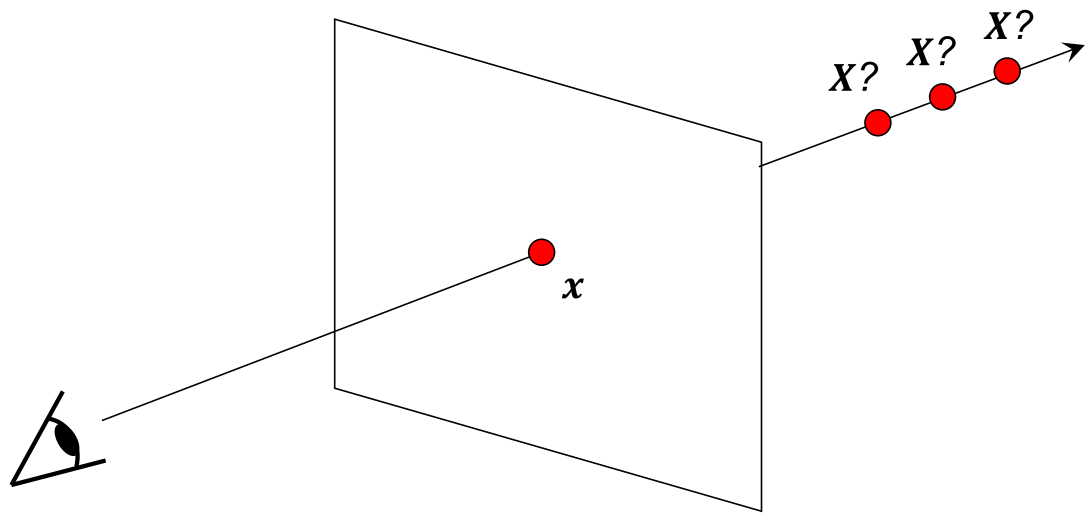
J. Vermeer, *Music Lesson*, 1662



A. Criminisi, M. Kemp, and A. Zisserman, [Bringing Pictorial Space to Life: computer techniques for the analysis of paintings](), *Proc. Computers and the History of Art*, 2002

Things aren’t always as they appear…

Single-view ambiguity



Single-view ambiguity

Rashad Alakbarov shadow sculptures

Anamorphic perspective

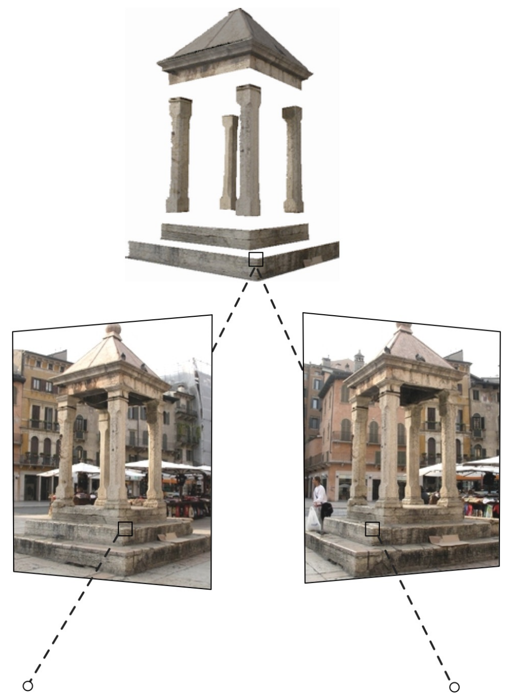
H. Holbein The Younger, *The Ambassadors*, 1533

https://en.wikipedia.org/wiki/Anamorphosis

Our goal: Recovery of 3D structure

- When certain assumptions hold, we can recover structure from a single view

- In general, we need *multi-view geometry*



[Image source](#)

- But first, we need to understand the geometry of a single camera…

More about horizons

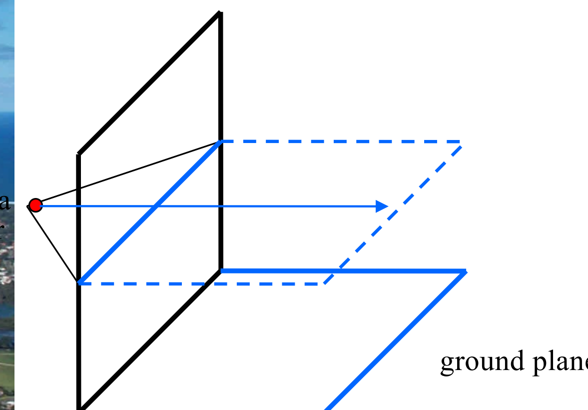
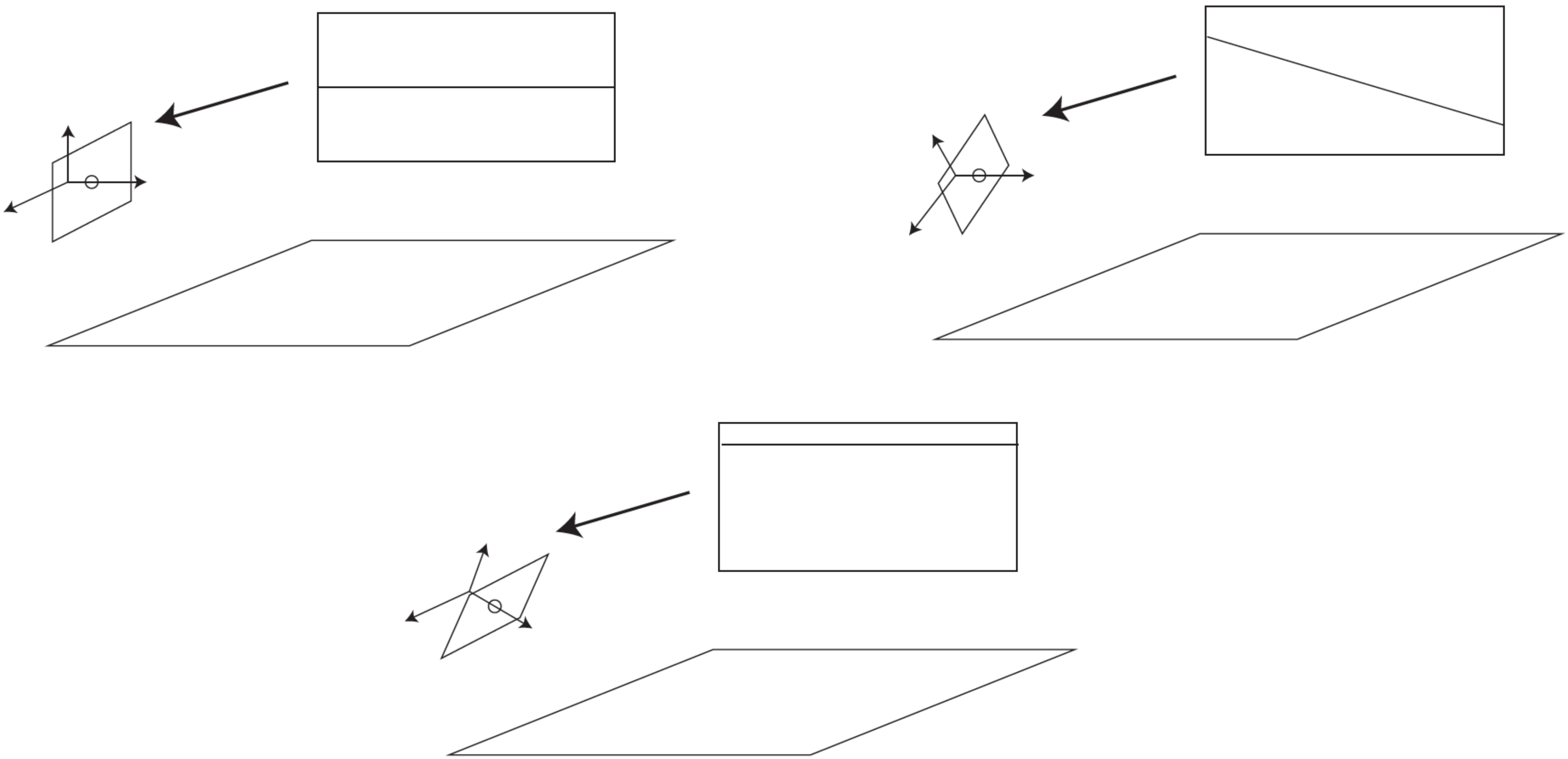


Image source: S. Seitz

Yet more about horizons



Projection of 3D shapes

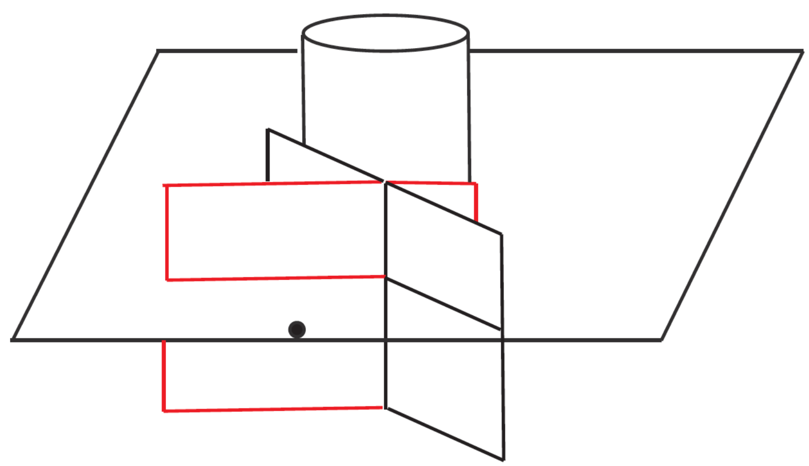
- What is are the relationships between the geometric properties of general 3D surfaces and their 2D projections?

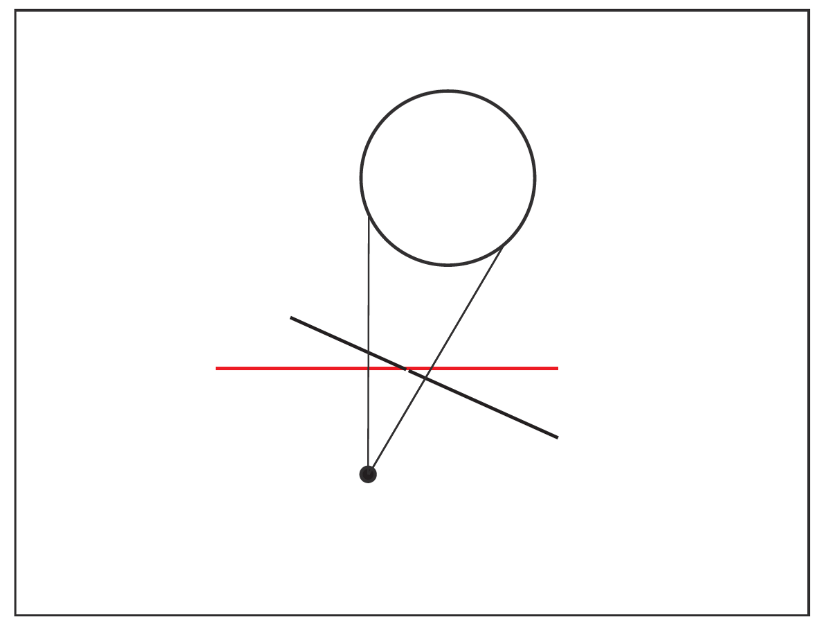
Barbara Hepworth sculpture

Camera rotation

Black camera is red camera, but rotated around the focal point

There is a one-one map between pixels in black image plane and red image plane (this is a homography, as we'll show later)

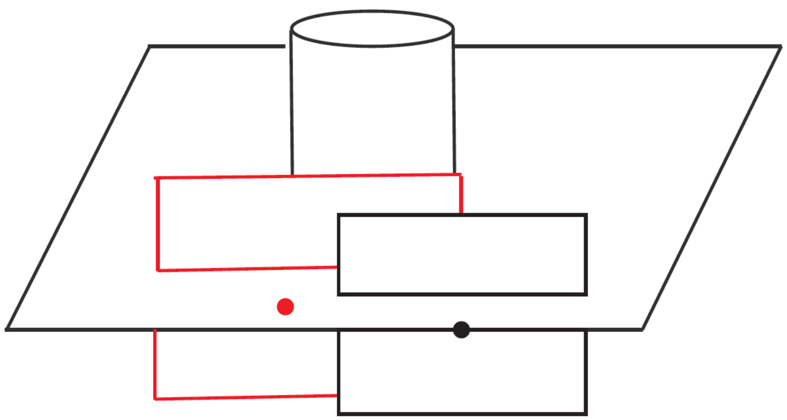


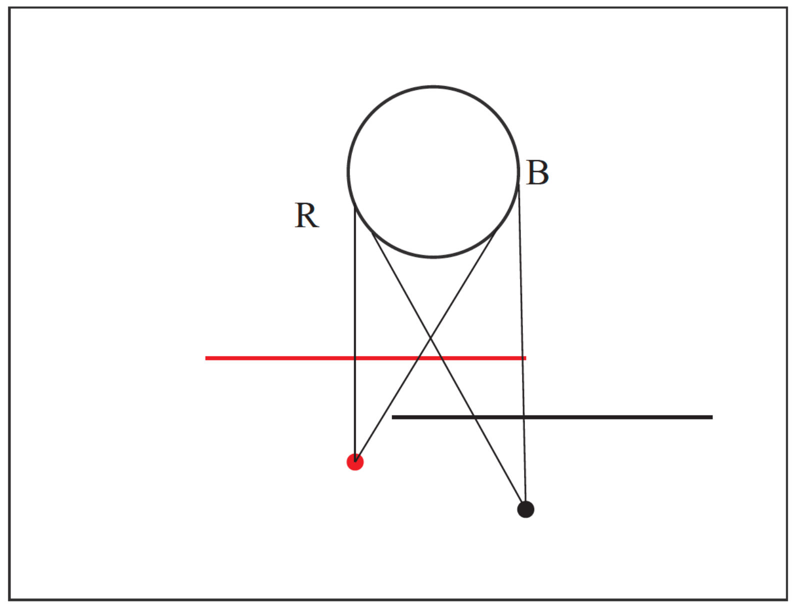


Camera translation

There is no longer a 1-1 map; some pixels in red camera cannot be seen in black, and vice versa

Here the red camera can see R and the black camera can't; the black camera can see B and the red camera can't.





Projection of 3D shapes

- What is the shape of the projection of a sphere?

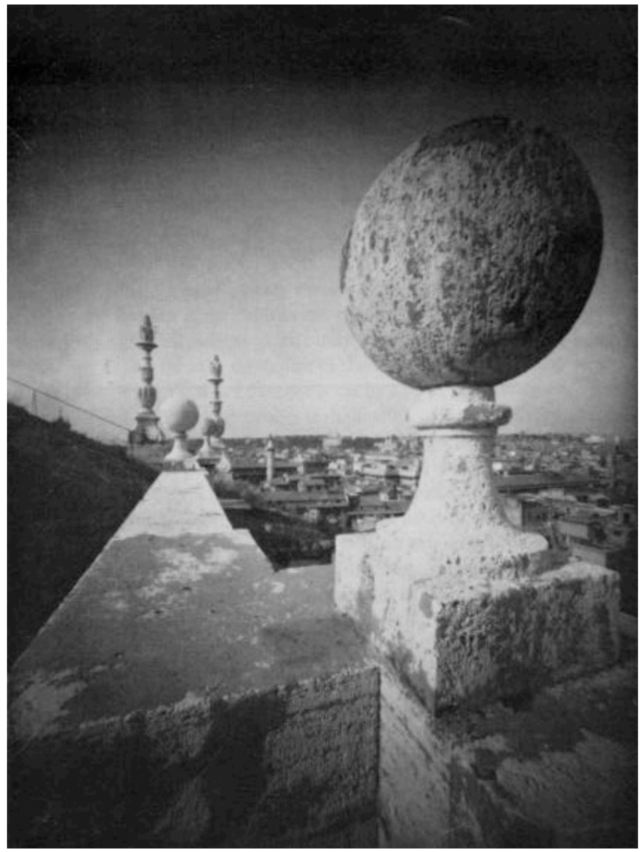
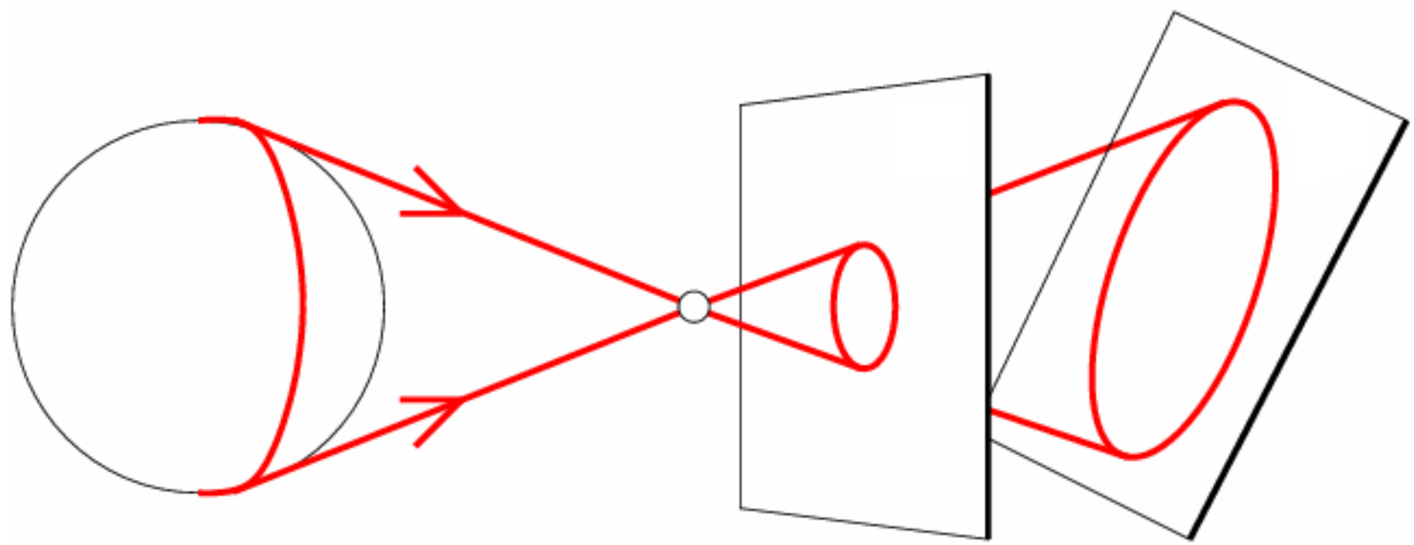


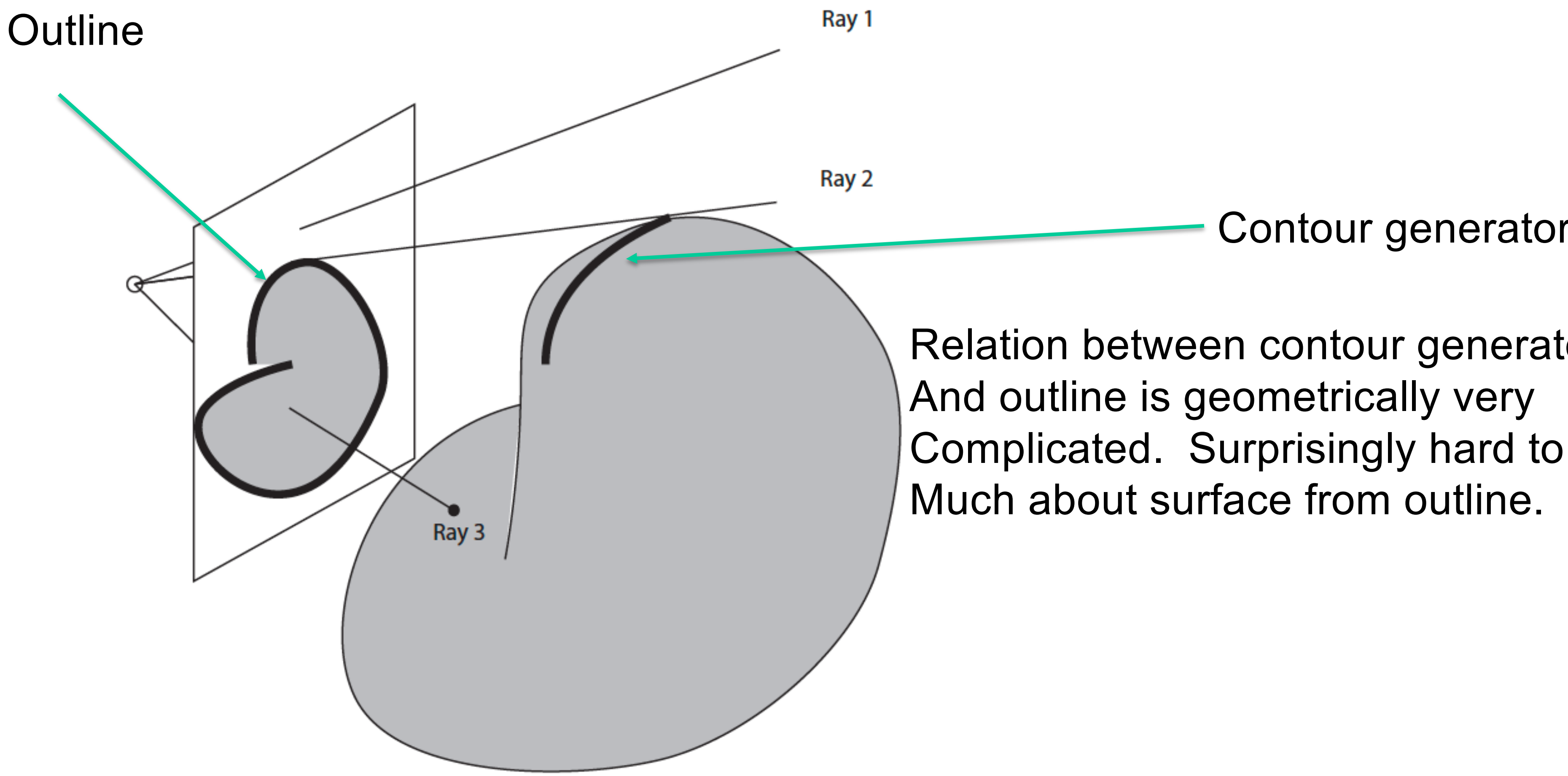
Image source: F. Durand

Projection of 3D shapes

- What is the shape of the projection of a sphere?

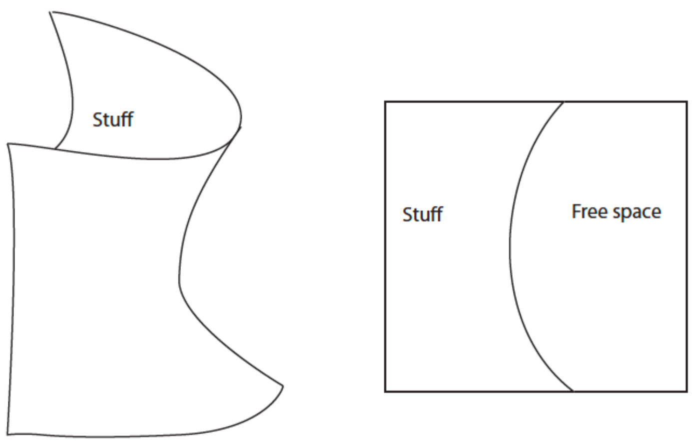


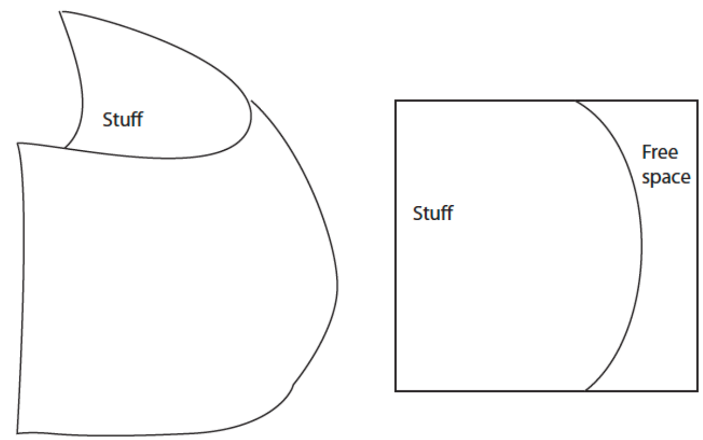
Projection of 3D shapes



Relation between contour generator
And outline is geometrically very
Complicated. Surprisingly hard to say
Much about surface from outline.

Projection of 3D shapes





J. Koenderink. [What does the occluding contour tell us about solid shape?]() Perception 13 (321-330), 1984

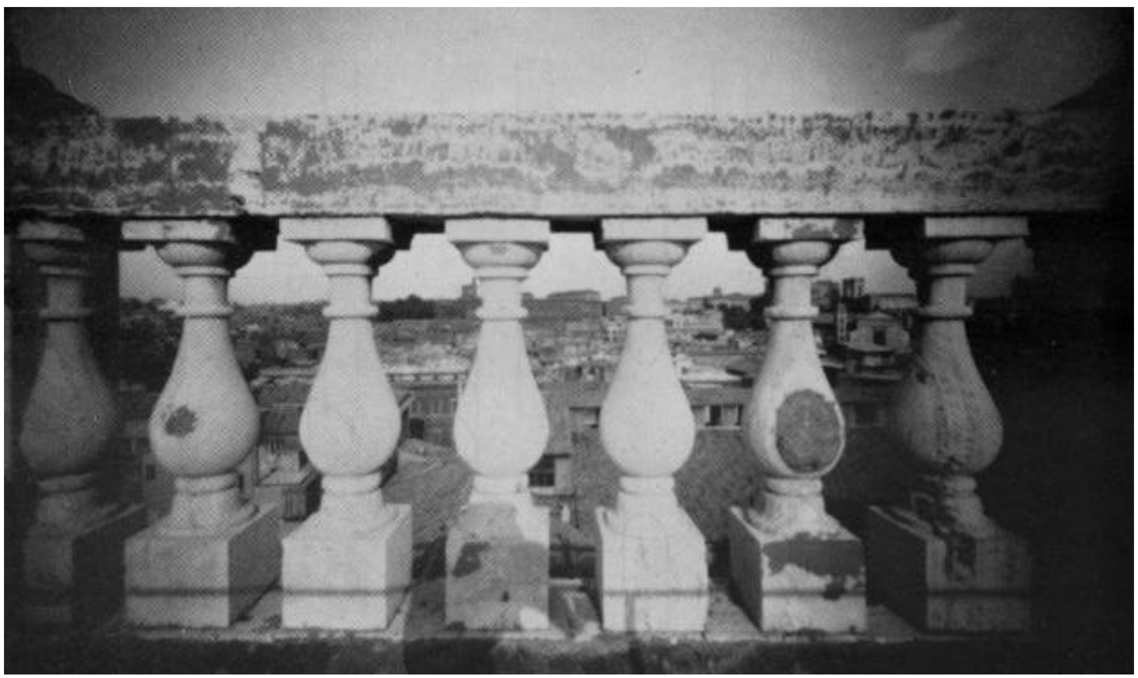
Projection of 3D shapes

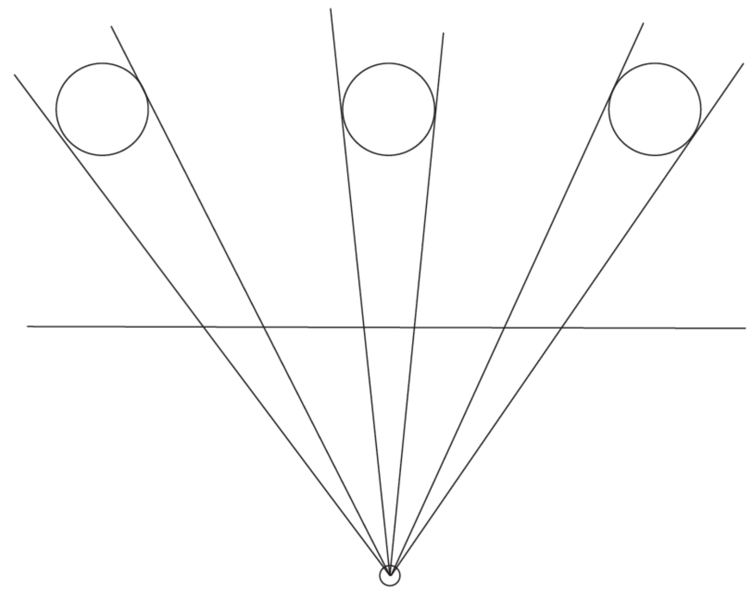
Figure 4. Details from Dürer's "Samson killing the lion". (Bartsch #2; the print dates from 1498.)

J. Koenderink. What does the occluding contour tell us about solid shape? Perception 13 (321-330), 1984

Projection of 3D shapes

- Are the widths of the projected columns equal?
  - The exterior columns are wider
  - This is not an optical illusion, and is not due to lens flaws
  - Phenomenon pointed out by Leonardo Da Vinci





Source: F. Durand

Perspective distortion: People

Overview

- Motivation: recovery of 3D structure
- Pinhole projection model
- Properties of projection
- Perspective projection matrix
- Orthographic projection

Homogeneous coordinates

- Add an extra coordinate and use an equivalence relation
- for 2D
  - equivalence relation
    k*(X,Y,Z) is the same as (X,Y,Z)
- for 3D
  - equivalence relation
    k*(X,Y,Z,T) is the same as (X,Y,Z,T)
- “Ordinary” or “non-homogeneous” coordinates
  - properly called affine coordinates
  - in 3D, affine -> homogeneous

$$(x, y, z) \to k * (x, y, z, 1)$$

  - in 3D, homogeneous to affine

$$(X, Y, Z, T) \to \left(\frac{X}{T}, \frac{Y}{T}, \frac{Z}{T}\right)$$

Homogenous coordinates

- Notice (0, 0, 0, 0) is meaningless (HC's for 3D)
  - also (0, 0, 0) in 2D
- Basic notion
  - Possible to represent points "at infinity" by careful use of zero
  - Where parallel lines intersect
    - eg

$$(tX, tY, tZ, 1) \quad \text{and} \quad (tX + a, tY + b, tZ + c, 1)$$

$$\text{intersect at} \quad (X, Y, Z, 0)$$

  - Where parallel planes intersect (etc)
- Can write the action of a perspective camera as a matrix

Homogeneous coordinates

Example: 23.1 *Lines on the affine plane*

Lines on the affine plane form one important example of homogeneous coordinates. A line is the set of points (x, y) where $ax+by+c=0$. We can use the coordinates (a, b, c) to represent a line. If $(d, e, f) = \lambda(a, b, c)$ for $\lambda \neq 0$ (which is the same as $(d, e, f) \equiv (a, b, c)$), then (d, e, f) and (a, b, c) represent the same line. This means the coordinates we are using for lines are homogeneous coordinates, and the family of lines in the affine plane is a projective plane. Notice that encoding lines using affine coordinates must leave out some lines. For example, if we insist on using $(u, v, 1) = (a/c, b/c, 1)$ to represent lines, the corresponding equation of the line would be $ux + vy + 1 = 0$. But no such line can pass through the origin – our representation has left out every line through the origin.

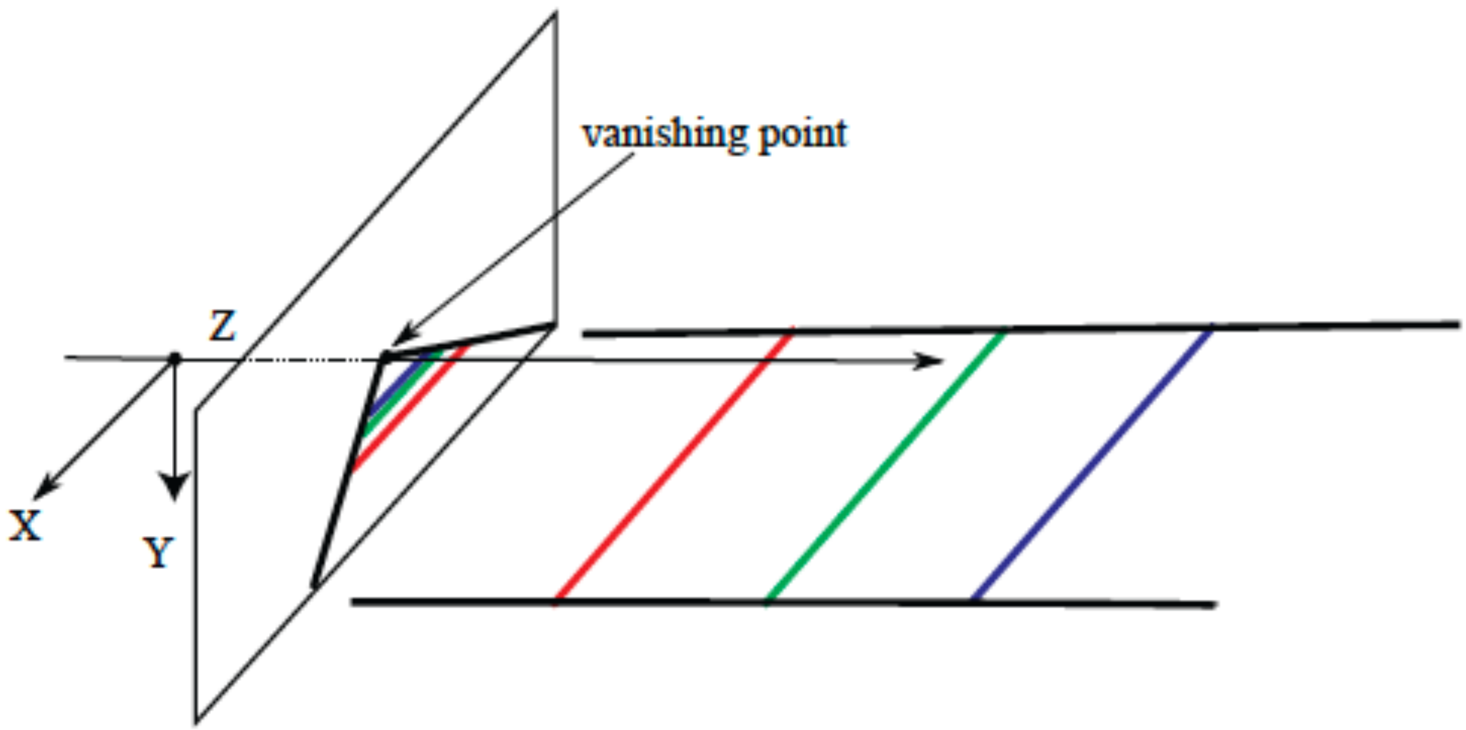
Homogeneous coordinates for a line

23.1.2 The projective line

In homogenous coordinates, we represent a point on a 1D space with two coordinates, so (X_1, X_2) (by convention, homogeneous coordinates are written with capital letters). Two sets of homogeneous coordinates (U_1, U_2) and (V_1, V_2) represent different points if there is no $\lambda \neq 0$ such that $\lambda(U_1, U_2) = (V_1, V_2)$. The set of all distinct points is known as a *projective line*. You should think of the projective line as an ordinary line (an *affine line*) with an "extra point". Every point on an affine line has a corresponding point on a projective line. A point on an affine line is given by a single coordinate x. This point can be identified with the point on a projective line given by $(X_1, X_2) = \lambda(x, 1)$ (for $\lambda \neq 0$) in homogeneous coordinates. The extra point has coordinates $(X_1, 0)$. These are the homogeneous coordinates of a single point (check this), but this point would be "at infinity" on the affine line.

There isn't anything special about the point on the projective line given by $(X_1, 0)$. You can see this by identifying the point x on the affine line with $(X_1, X_2) = \lambda(1, x)$ (for $\lambda \neq 0$). Now $(X_1, 0)$ is a point like any other, and $(0, X_2)$ is "at infinity". A little work establishes that there is a 1-1 mapping between the projective line and a circle (exercises).

You can see the point at infinity

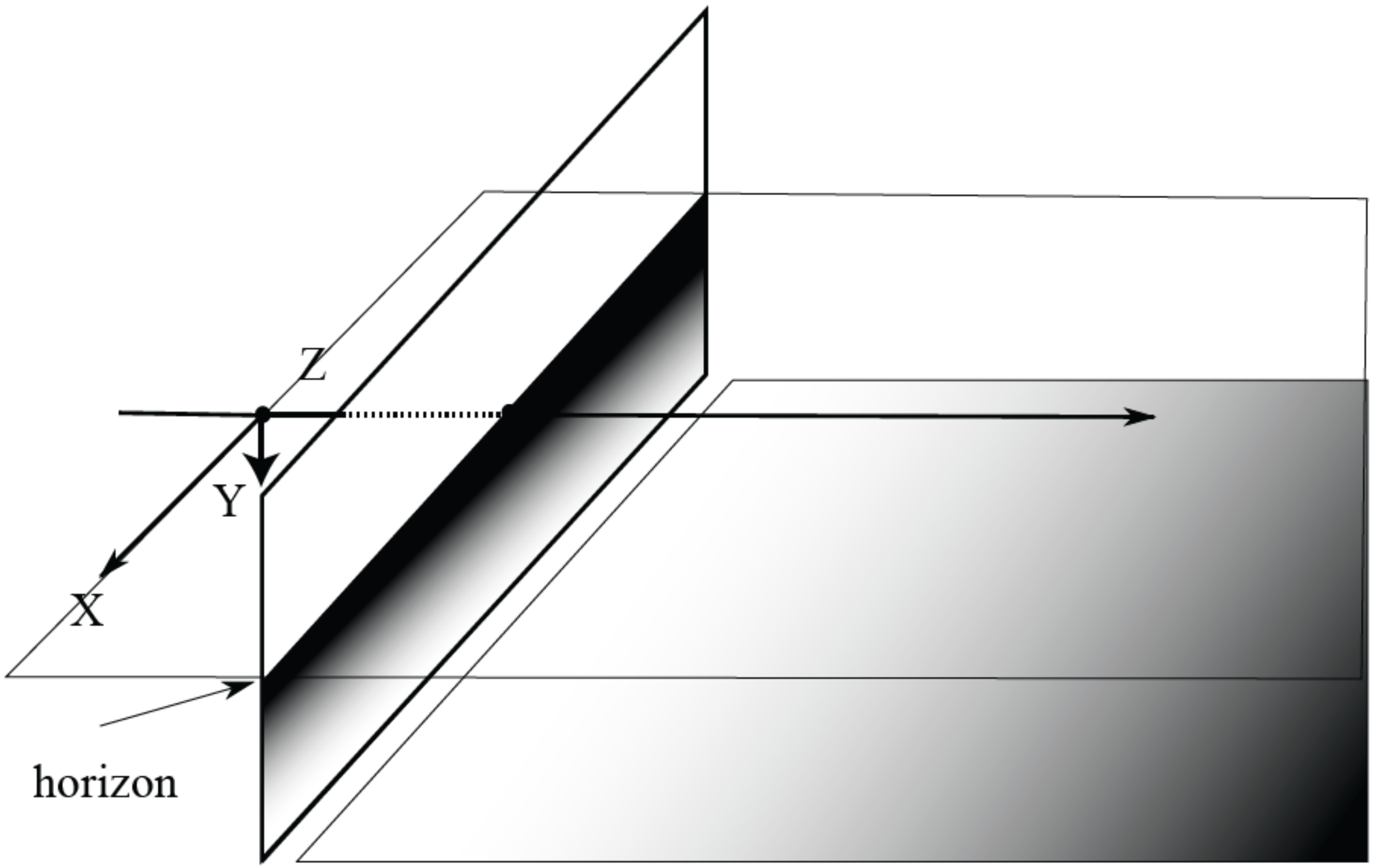


Homogeneous coordinates for the plane

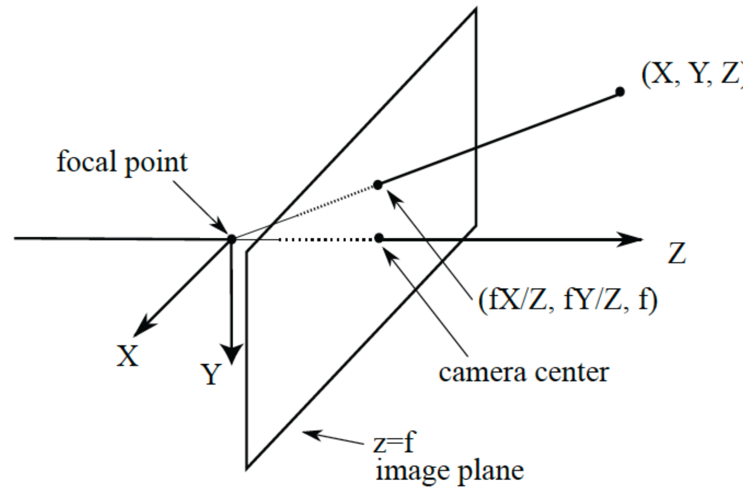
23.1.3 The projective plane

The space represented by three homogeneous coordinates is known as a *projective plane*. You can map an *affine plane* (the usual plane, with coordinates x, y) to a projective plane by writing $(X_1, X_2, X_3) = (x, y, 1)$. Notice that there are points on the projective plane — the points where $X_3 = 0$ — that are missing. These points form a projective line (check this!). This line is often referred to as the *line at infinity*.

You can see the line at infinity, too!



The camera matrix



- Turn previous expression into HC's
  - HC's for 3D point are (X,Y,Z,T)
  - HC's for point in image are (U,V,W)

In Camera coordinate system:

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

The camera matrix

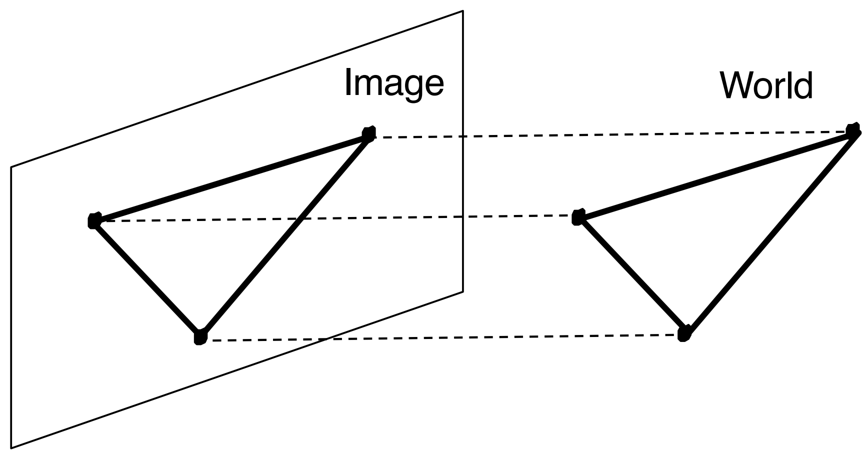
$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\mathcal{C}_p} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Notice: focal point = **f** = (0, 0, 0, 1)^T has the property:

$$\mathcal{C}_p \mathbf{f} = \mathbf{0}$$

Orthographic projection

- Special case of perspective projection
  - Distance from center of projection to image plane is infinite
  - Also called “parallel projection”



Slide by Steve Seitz

Orthographic projection

- Special case of perspective projection
  - Distance from center of projection to image plane is infinite
  - Also called “parallel projection”

Orthographic projection

19.1.3 Scaled Orthographic Projection and Orthographic Projection

Under some circumstances, perspective projection can be simplified. Assume the camera views a set of points which are close to one another compared with the distance to the camera. Write $\mathbf{X}_i = (X_i, Y_i, Z_i)$ for the i'th point, and assume that $Z_i = Z(1+\epsilon_i)$, where ϵ_i is quite small. In this case, the distance to the set of points is much larger than the *relief* of the points, which is the distance from nearest to furthest point. The i'th point projects to $(fX_i/Z_i, fY_i/Z_i)$, which is approximately $(f(X_i/Z)(1-\epsilon_i), f(Y_i/Z)(1-\epsilon_i))$. Ignoring ϵ_i because it is small, we have the projection model

$$(X, Y, Z) \to (f/Z)(X, Y) = s(X, Y).$$

Camera matrix for orthographic projection

- Almost never encounter orthographic projection

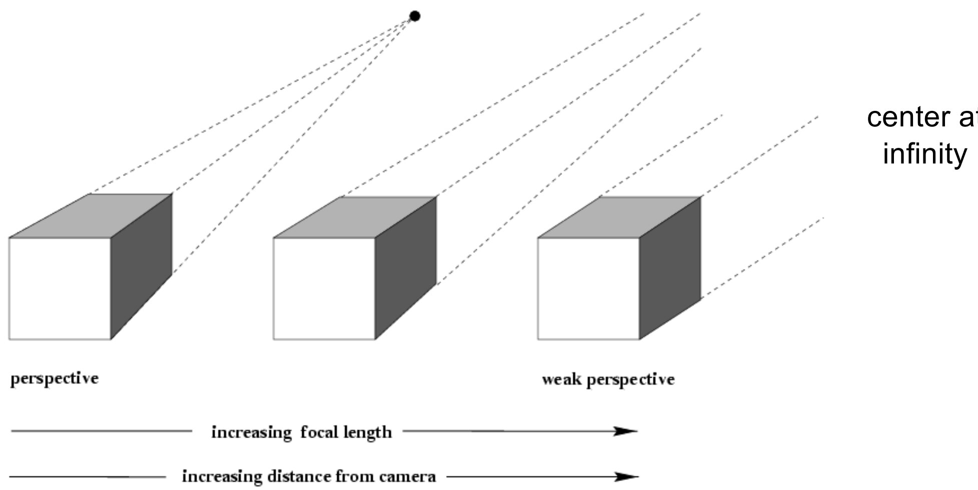
Remember this: *Scaled orthographic projection maps*

$$(X, Y, Z) \to s(X, Y)$$

where s is some scale. The model applies when the distance to the points being viewed is much greater than their relief. Many views of people have this property.

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \mathcal{C} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathcal{W} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Approximating an orthographic camera



Source: Hartley & Zisserman

Outline

- Camera calibration using vanishing points
- Measurements from a single image
  - Measuring height above the ground plane
  - Measuring within planes

Using a ruler

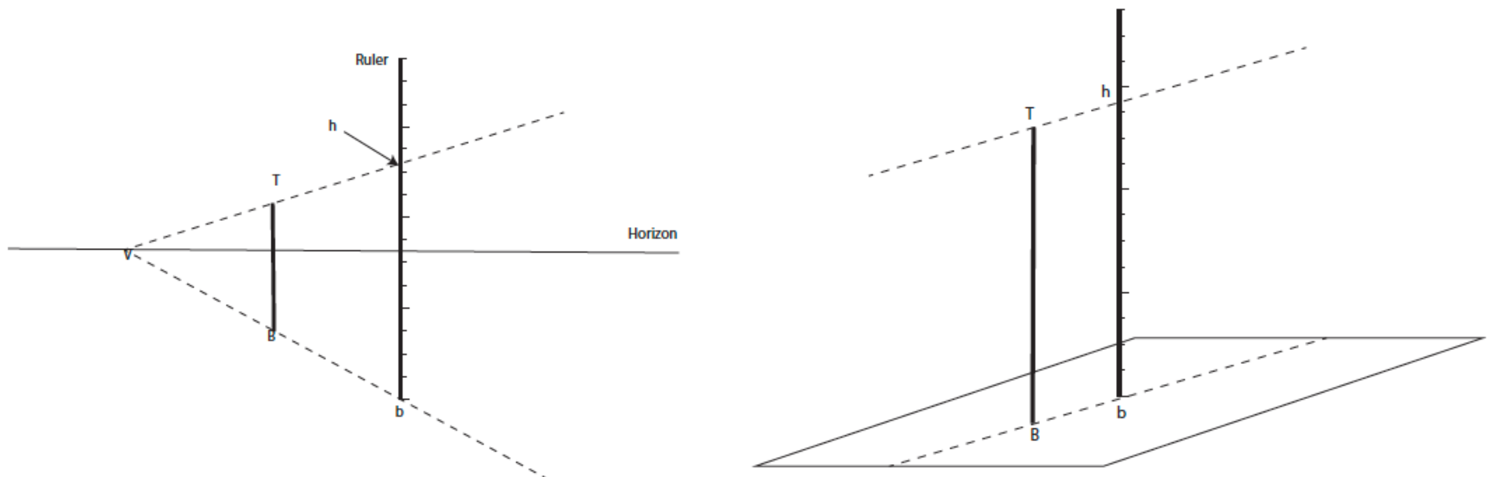


FIGURE 24.3: **Left**, *an image of a ruler and an object, which just happen to be standing perpendicular to a ground plane. In an* uncalibrated *image like this, we can measure the height of the object. Construct the line bB, and intersect that with the horizon to get the point V. The line from the top of the object T to the true height of the object on the ruler (h) is parallel in 3D to bB. In turn, the line Th must intersect the horizon at V. So if you construct VT, it will intersect the ruler at h yielding the height of the object.* **Right** *shows a 3D view; the line Th must be parallel to bB, and so in the image these two lines intersect at the horizon.*

Using a ruler

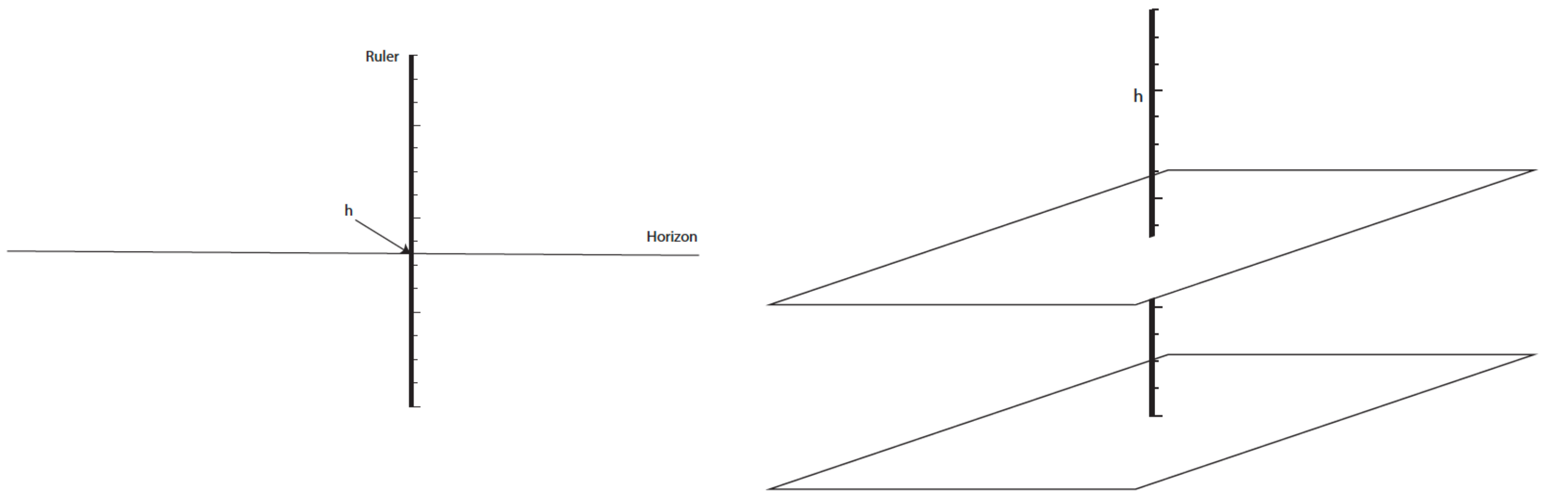


FIGURE 24.4: **Left**, *an image of a ruler which just happens to be standing perpendicular to a ground plane. In an* uncalibrated *image like this, we can measure the height of the camera focal point above the ground plane. The plane through the focal point parallel to the ground plan (and so the same height above the ground plane as the focal point) must form the horizon, so the intersection between horizon and ruler yields the height of the focal point.* **Right** *shows a 3D view; the bottom plane is the ground plane, and the top plane is the plane through the focal point parallel to the ground plane.*

Working without a ruler is harder than might seem

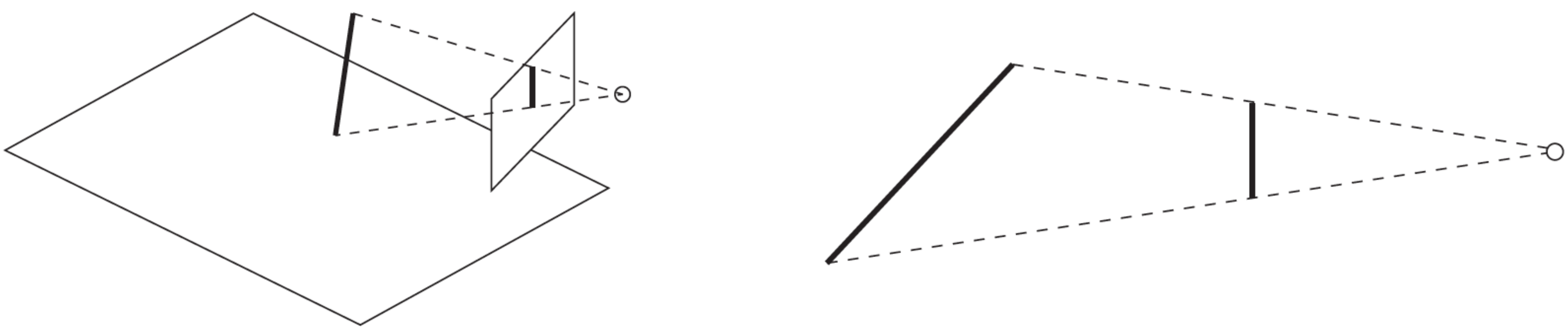


FIGURE 24.5: **Left**, *a perspective camera views a reference object perpendicular to a ground plane. This produces a line segment in the image plane.* **Right** *shows the reference object and the line segment in the image plane.*

Working without a ruler is harder than might seem

Parametrize the reference line segment in 3D

using affine coordinates to get $\mathbf{p}+t\mathbf{d}$, where $\mathbf{d}$ is a unit vector (so a step of 1 in t is a step of length 1 along the reference segment). Write c_{ij} for the i, j'th component of the 3×4 camera matrix. Then the homogeneous coordinates for the image line will be

$$\left(\begin{array}{c} (c_{11}p_1 + c_{12}p_2 + c_{13}p_3 + c_{14}) + t(c_{11}d_1 + c_{12}d_2 + c_{13}d_3 + c_{14}) \\ (c_{21}p_1 + c_{22}p_2 + c_{23}p_3 + c_{24}) + t(c_{21}d_1 + c_{22}d_2 + c_{23}d_3 + c_{24}) \\ (c_{31}p_1 + c_{32}p_2 + c_{33}p_3 + c_{34}) + t(c_{31}d_1 + c_{32}d_2 + c_{33}d_3 + c_{34}) \end{array} \right) = \left(\begin{array}{c} a+bt \\ c+dt \\ e+ft \end{array} \right).$$

SInce we know the image is a line, we can ignore one of these three homogeneous coordinates, so the transformation is a projective transformation. Now on the 3D

Working without a ruler is harder than might seem

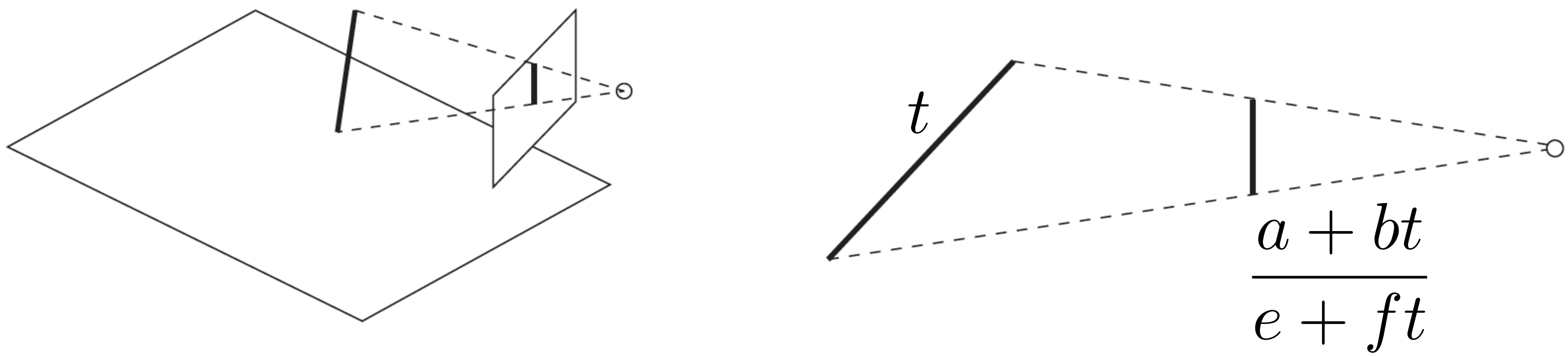


FIGURE 24.5: **Left**, *a perspective camera views a reference object perpendicular to a ground plane. This produces a line segment in the image plane.* **Right** *shows the reference object and the line segment in the image plane.*

Working without a ruler is harder than might seem

Now on the 3D reference line segment, the points $t = 0$ and $t = 1$ are the same distance apart as the points $t = 1$ and $t = 2$. But in the image line, using affine coordinates, these points are

$$\frac{a}{c}, \frac{a+b}{c+d}, \frac{a+2b}{c+2d}$$

which are not, in general, evenly spaced (check this with, for example, $a = 0$, $b = 1$, $c = 1$, $d = 1$).

The Cross-ratio

A clever trick from projective geometry allows us to use a reference object to measure heights. Write $\mathbf{P}_1, \ldots, \mathbf{P}_4$ for the coordinates of four points on a projective line, written in homogeneous coordinates. Write $\mathcal{M}$ for a projective transformation of the line to itself (so a 2×2 matrix with non-zero determinant. Finally, write

$$d(\mathbf{P}_i, \mathbf{P}_j) = \det\left([\mathbf{P}_i \mathbf{P}_j]\right).$$

Notice that

$$\det\left([\mathcal{M}\mathbf{P}_i \mathcal{M}\mathbf{P}_j)]\right) = \det\left(\mathcal{M}[\mathbf{P}_i \mathbf{P}_j]\right) = \det\left(\mathcal{M}\right) \det\left([\mathbf{P}_i \mathbf{P}_j]\right)$$

The Cross-ratio

which means that

$$\frac{d(\mathbf{P}_1, \mathbf{P}_2)d(\mathbf{P}_3, \mathbf{P}_4)}{d(\mathbf{P}_1, \mathbf{P}_3)d(\mathbf{P}_2, \mathbf{P}_4)}$$

is a *projective invariant* — computing the value of this *cross ratio* using $\mathbf{P}_1, \ldots, \mathbf{P}_4$ or using $\mathcal{M}\mathbf{P}_1, \ldots, \mathcal{M}\mathbf{P}_4$ will yield the same number, as long as $\mathcal{M}$ is a projective transformation.

The Cross-ratio

Now check that the cross-ratio of the four points $(0, 1)$, $(a, 1)$, $(b, 1)$ and $(1, 0)$ is a/b (notice the last point is the point at infinity). We can use this observation to measure height relative to a reference object. Using the notation of Figure 24.6, we construct the line Bb from the base of the object to the base of the reference object. Produce this line to intersect the horizon at V. Now construct VT, which intersects the reference object at h. In 3D, the line VT is parallel to the ground plane, so that the point h in 3D is the same height above the ground plane as the point T in 3D. The vanishing point for the vertical lines (the object and the reference object) is at infinity in this image, *so we know where it lies on line bt.* Write P for this vanishing point, r for the height of the reference object and o for the height of the object. Then we have

$$\frac{d(b,h)d(t,P)}{d(b,t),b(h,P)} = \frac{r}{o}$$

but we *know* the height of the reference object and we can measure the cross ratio, so we can recover o.

The Cross-ratio

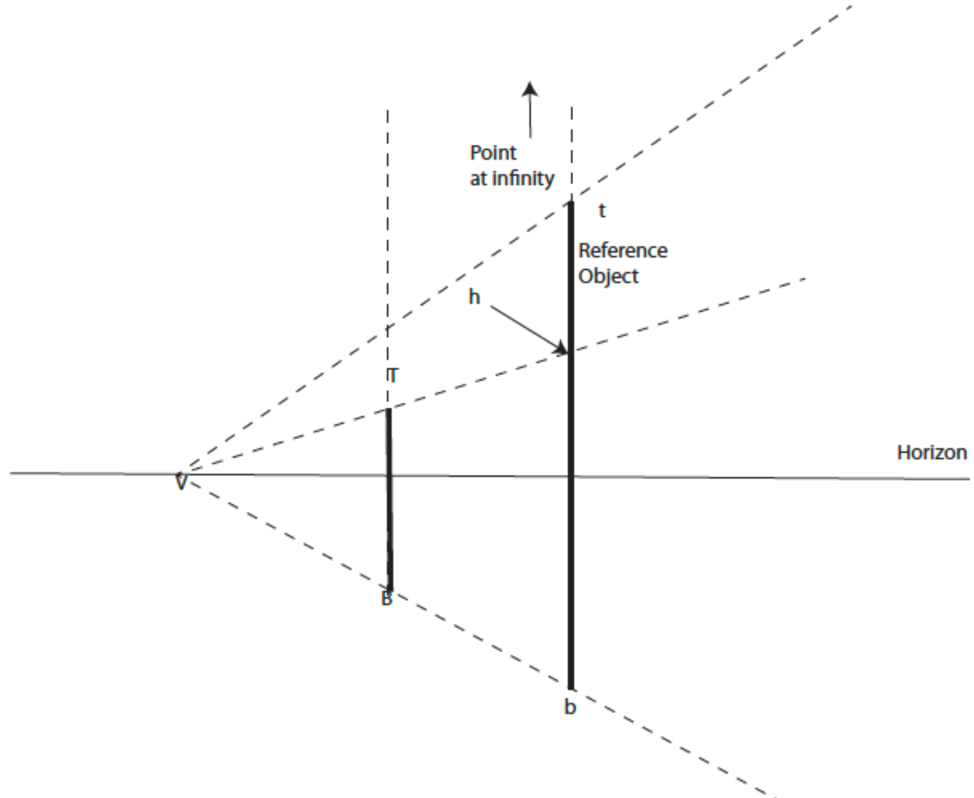


FIGURE 24.6: *A perspective camera views a reference object and another object perpendicular to a ground plane. This produces a line segment in the image plane. Constructing appropriate lines in the figure and taking a cross ratio yields the height of the object.*

The Cross-ratio

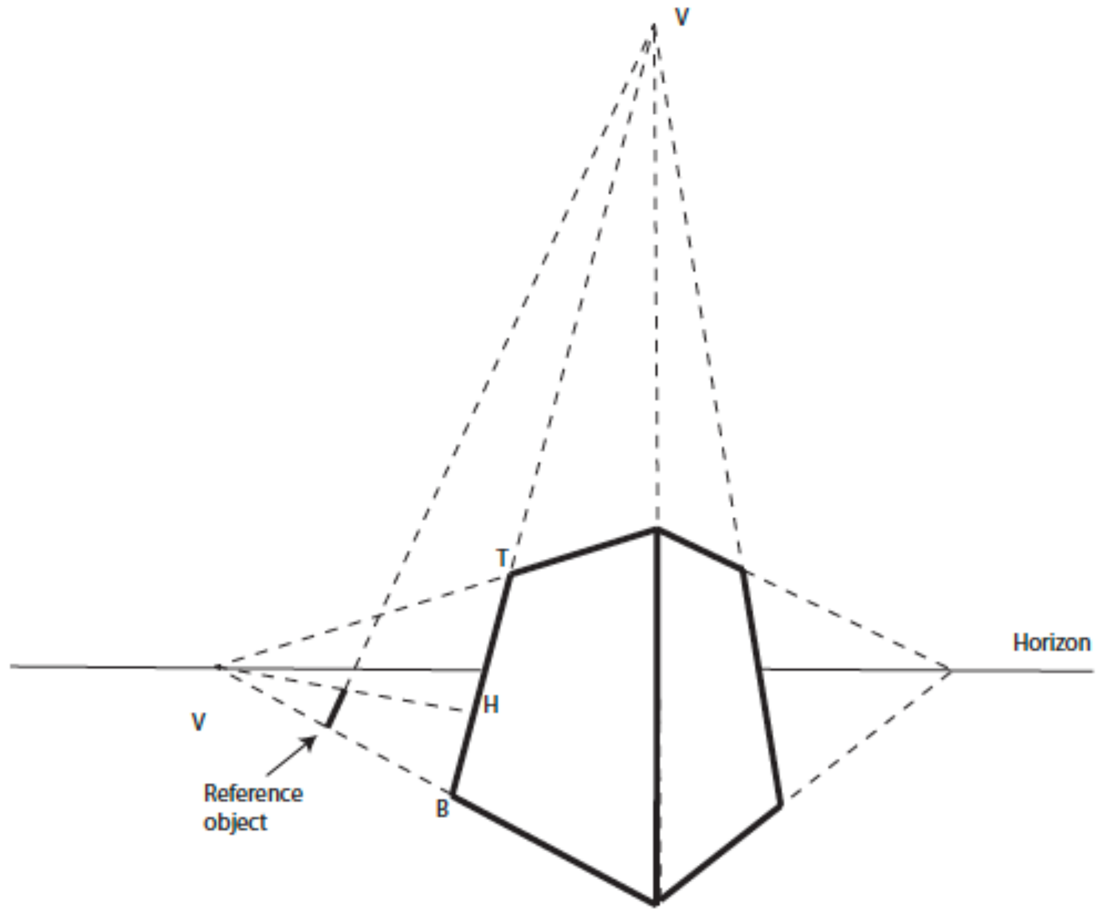
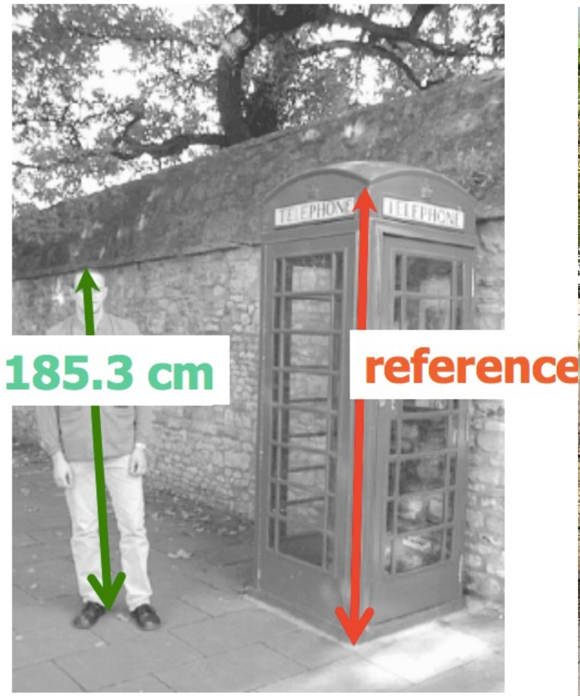


FIGURE 24.7: *A building and a person viewed in a more extreme perspective view than that of 24.6. The person has known height, and can act as reference object. The same construction as in that figure yields the height of the building relative to the person.*

Single-view measurement examples



That booth is still there! (Oxford, September 2022)

A. Criminisi, I. Reid, and A. Zisserman, Single View Metrology, IJCV 2000
Figure from UPenn CIS580 slides