Computer Vision (7CCSMCVI / 6CCS3COV)

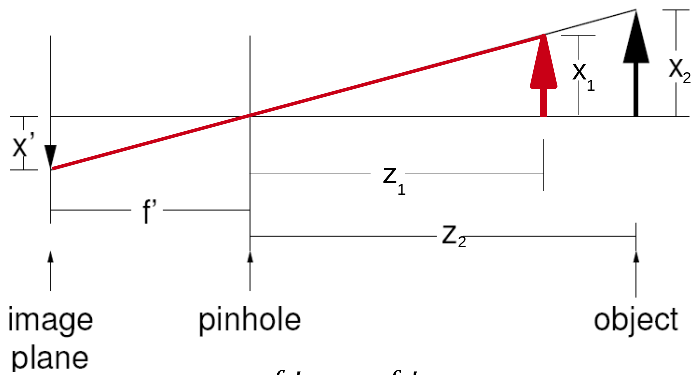
Recap

- **Image formation**
- **Low-level vision**
- **Mid-level vision**
  - grouping and segmentation (finding matching elements within an image)
  - correspondence problem (finding matching elements across images)
    - for all locations or selected interest points
    - comparing image intensities or descriptors
    - finding matches by search
    - determining similarity between elements
    - dealing with false matches by model fitting
  - Stereo and Depth ← Today
- **High-level vision**

Today

- Stereo vision
  - – stereo camera geometry
    - » coplanar cameras (simple case)
    - » non-coplanar cameras (complex case)
  - – disparity measurement
    - » calculating depth
  - – correspondence
    - » stereo constraints used to solve the correspondence problem
- Other cues to depth
  - – Binocular
  - – Oculomotor
  - – Monocular
  - – Motion

Why is stereo vision important?

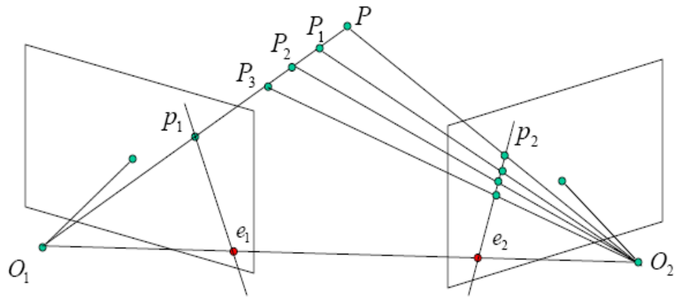


$$x' = \frac{f'}{Z_1} X_1 = \frac{f'}{Z_2} X_2$$

A camera projects 3D points onto a 2D plane

3D points on the same line-of-sight have the same 2D image location (i.e. imaging results in depth information loss)

Why is stereo vision important?



Depth information can be recovered using two images and a knowledge of geometry.

e.g. all points P, P_1, P_2, and P_3 project to the same location in the left image, but to different locations in the right image.

The right image allows us to measure how far each of these points are from the left camera.

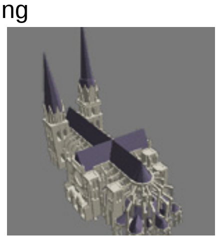
Why is stereo vision important?

This is useful for:

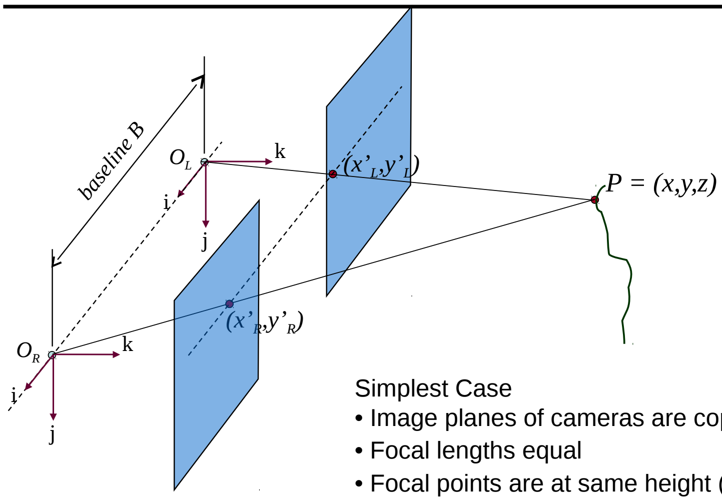
Path planning / collision avoidance (car / robot)


virtual advertising

Not stereo, but same methods can be used for 3D model building

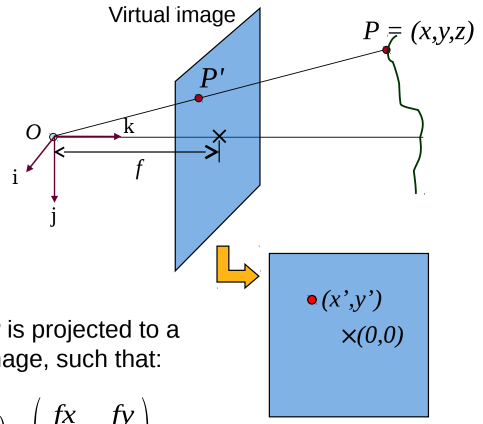
 +  +  → 

Stereo: coplanar cameras



Simplest Case
- Image planes of cameras are coplanar
- Focal lengths equal
- Focal points are at same height (i.e. x-axes collinear)
- Intersection of optical axes at infinity (i.e. z-axes parallel)

Image formation reminder



3D scene point *P* is projected to a point *P'* on the image, such that:

$$P' = (x', y') = \left(\frac{fx}{z}, \frac{fy}{z} \right)$$

Assuming that the image centre is *(0,0)* [see lecture 2]

Stereo: coplanar cameras

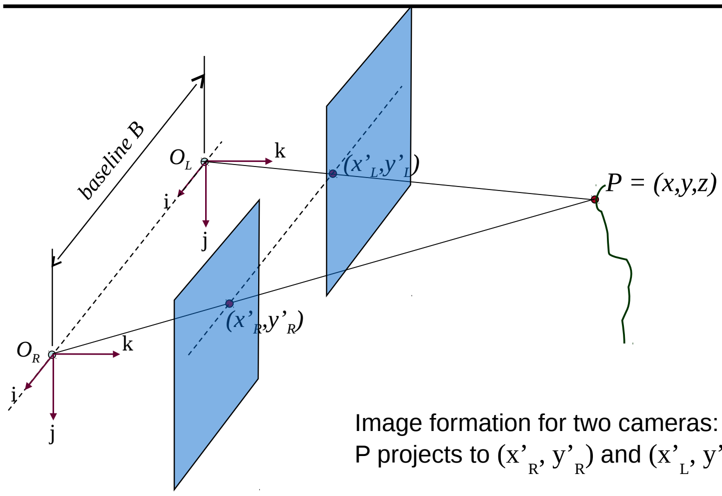
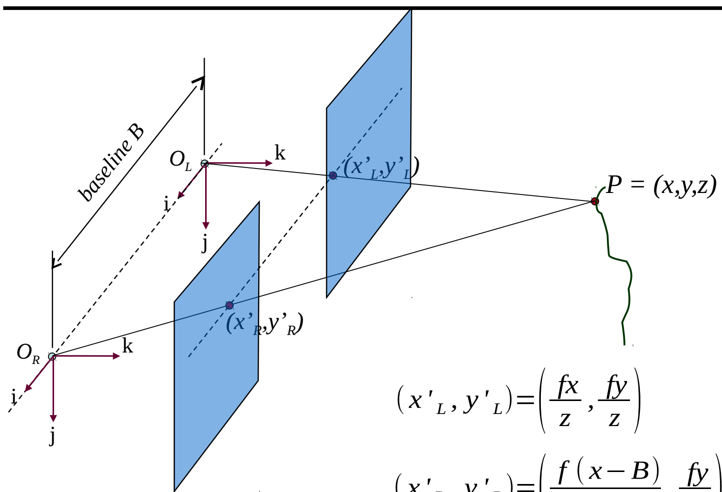


Image formation for two cameras: P projects to (x'_R, y'_R) and (x'_L, y'_L)

Note: Because x-axes of cameras are collinear, $y'_L = y'_R$

Stereo: coplanar cameras

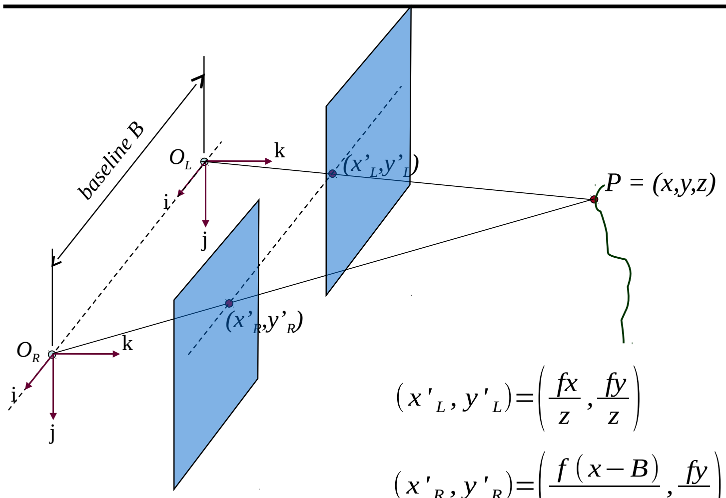


$$(x'_L, y'_L) = \left(\frac{fx}{z}, \frac{fy}{z} \right)$$

$$(x'_R, y'_R) = \left(\frac{f(x-B)}{z}, \frac{fy}{z} \right)$$

Using the coordinate system of the left camera (since $x_R = x_L - B$)

Stereo: coplanar cameras



$$(x'_L, y'_L) = \left(\frac{fx}{z}, \frac{fy}{z}\right)$$

$$(x'_R, y'_R) = \left(\frac{f(x-B)}{z}, \frac{fy}{z}\right)$$

Disparity, $d = x'_L - x'_R = \frac{fx}{z} - \frac{f(x-B)}{z} = \frac{fB}{z}$ ⇒ $z = f\frac{B}{d}$

Disparity

Depth is inversely proportional to disparity.

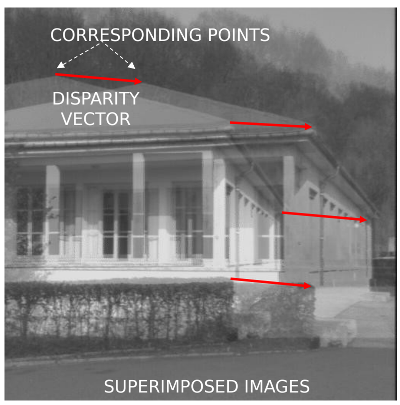
$$z = f\frac{B}{d}$$

If the baseline distance is known, and we can measure the disparity, then we can calculate the depth of a point.

Even without the baseline, we can know the relative depths of points from their relative disparities.

Disparity

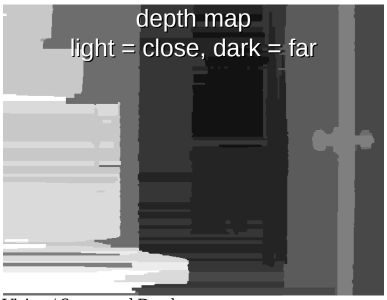
The difference vector of the image coordinates of two corresponding points.



NOTE

- the disparity, d, of a point is a 2D vector.
- disparity is measured in pixels and can be positive or negative
- a pair of stereo images defines a field of disparity vectors (a disparity map)
- For coplanar cameras disparity is horizontal only

Disparity / depth map: example



Disparity / depth map: example

left image

right image

False shallow region caused by false matches

depth map
light = close, dark = far

The stereo correspondence problem

To measure disparity, it is necessary to find corresponding points in the stereo pair of images.

To solve the stereo correspondence problem, we can use:

- **Correlation-based methods**
  yield dense disparity maps: a disparity value at each pixel.
- **Feature-based methods**
  yield sparse disparity maps: a disparity value at interest points only.

The stereo correspondence problem

To measure disparity, it is necessary to find corresponding points in the stereo pair of images.

Basic requirements to be able to solve the correspondence problem:

1. Most scene points visible in both images
2. Corresponding image regions appear "similar"

These assumptions hold if:

- The distance of the 3D point from the cameras is much larger than the baseline: z >> B

The stereo correspondence problem

To measure disparity, it is necessary to find corresponding points in the stereo pair of images.

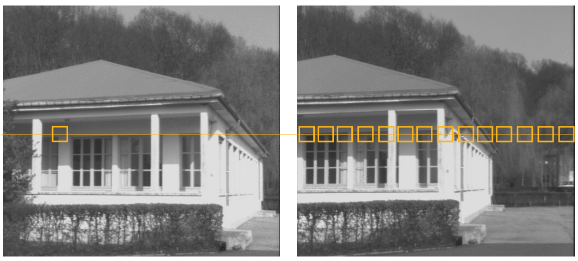
As we saw in the previous lecture, solving the correspondence problem is not easy.

However, we can use knowledge about the stereo camera system to help find a solution...

Stereo Constraints on Correspondence

Epipolar constraint

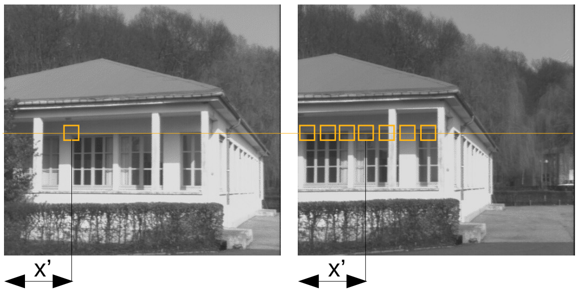
For coplanar cameras, $y'_L = y'_R$ so 2D search can be reduced to a 1D search along the "epipolar" line (= the corresponding row of pixels for coplanar cameras).



Stereo Constraints on Correspondence

Maximum disparity

Length of search region depends on the maximum expected disparity, often predictable geometrically ($d_{max} = fB/z_{min}$).

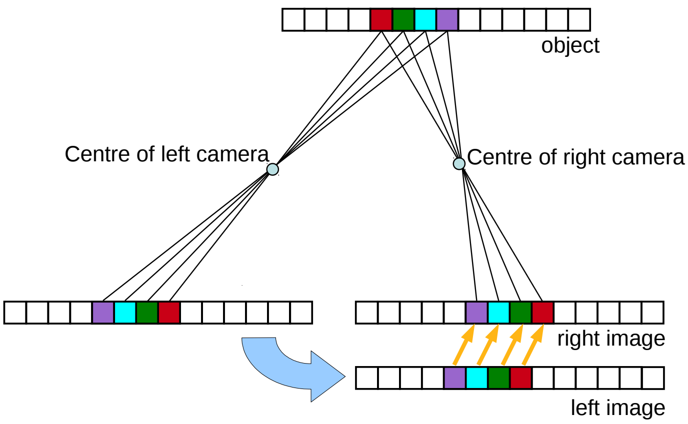


For each point (x', y') in the left image, search for its corresponding point between (x'-d_{max},y') and (x'+d_{max},y') in the right image.

Stereo Constraints on Correspondence

Continuity

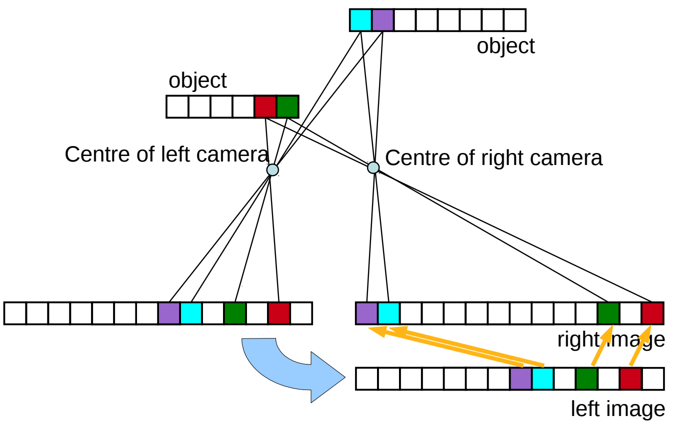
Neighbouring points should have similar disparities, because the environment is made of continuous surfaces over which depth varies smoothly.



Stereo Constraints on Correspondence

Continuity

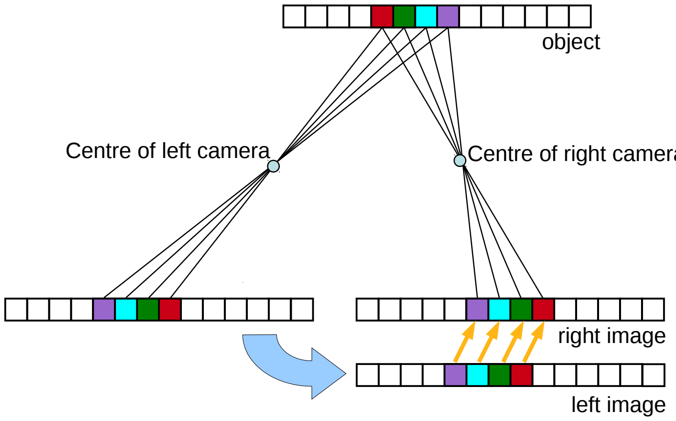
The exception is at discontinuities where depth can change suddenly.



Stereo Constraints on Correspondence

Uniqueness

A location in one image may only match a single location in the another image.

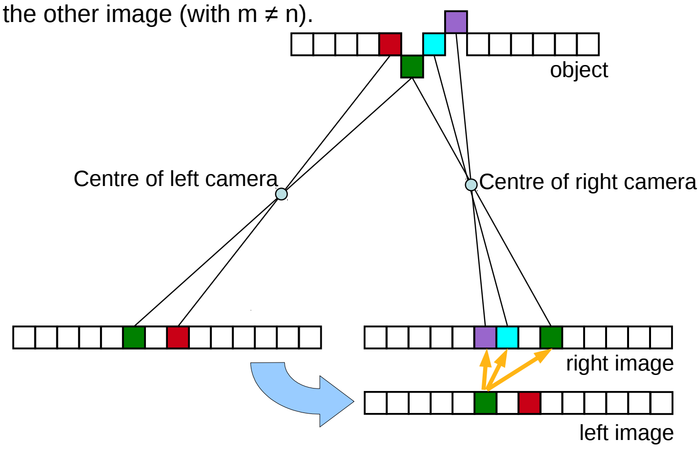


Stereo Constraints on Correspondence

Uniqueness

The exception is when a surface lies along a line-of-sight for one camera (in this case one location may match many locations).

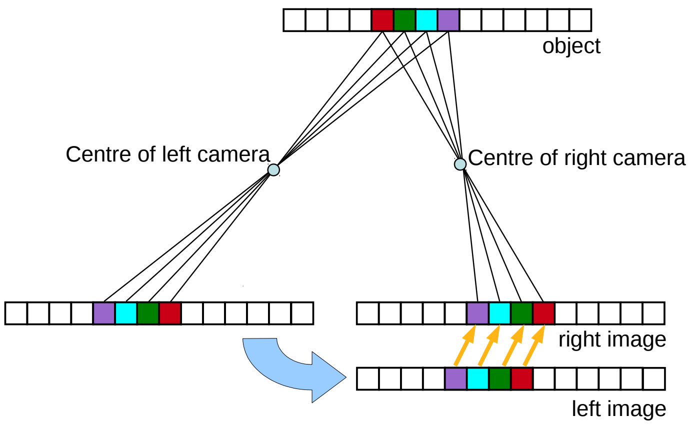
If fact any inclined surface may project to n pixels in one image and m pixels in the other image (with m ≠ n).



Stereo Constraints on Correspondence

Ordering

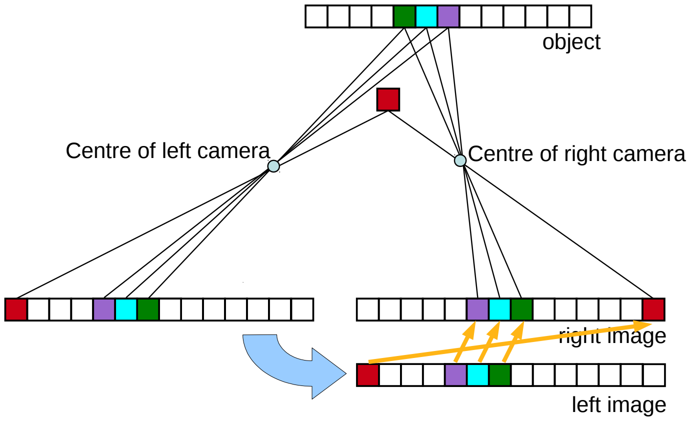
Matching points along corresponding epipolar lines should be in the same order.



Stereo Constraints on Correspondence

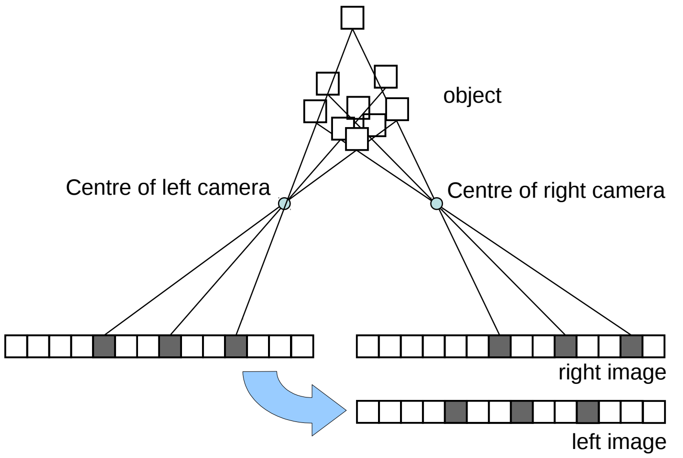
Ordering

The exception is if a foreground object occludes a surface.



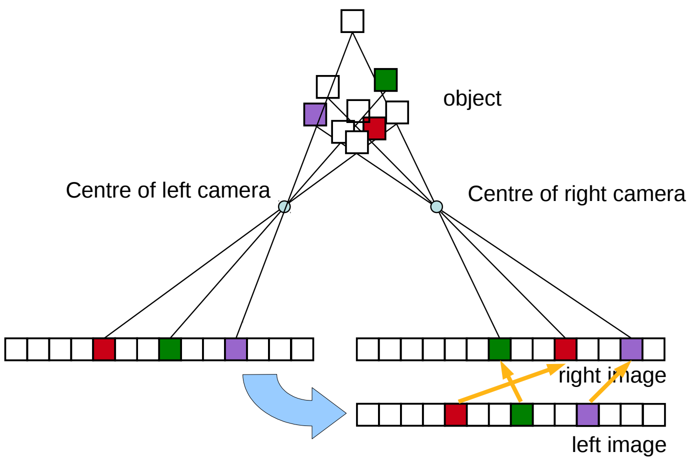
Correspondence problems

Correspondence is fundamentally ambiguous, i.e. there are many possible solutions.



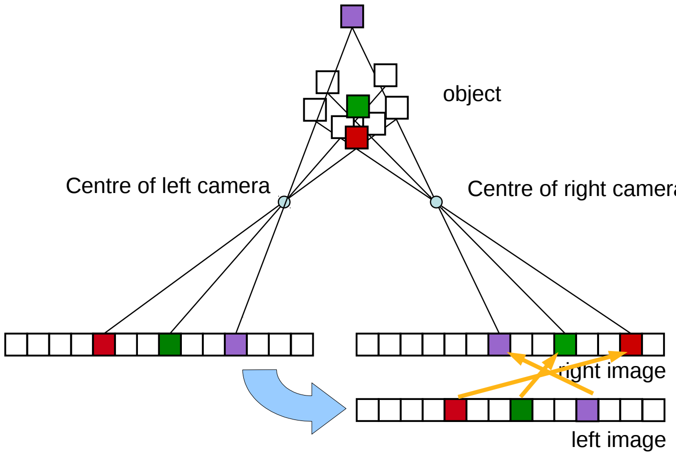
Correspondence problems

Correspondence is fundamentally ambiguous, i.e. there are many possible solutions.



Correspondence problems

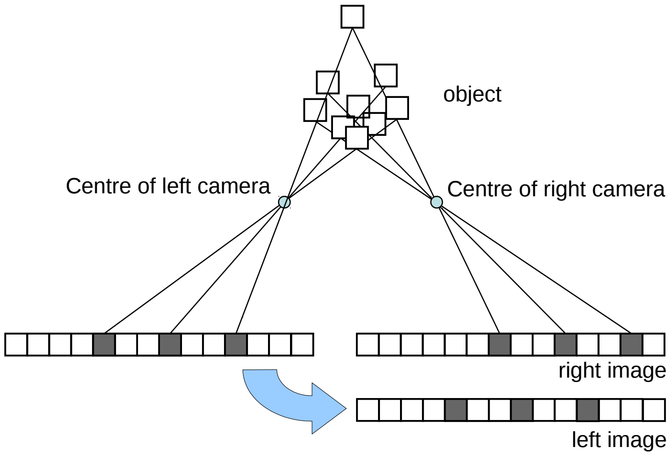
Correspondence is fundamentally ambiguous, i.e. there are many possible solutions.



Correspondence problems

Correspondence is fundamentally ambiguous, i.e. there are many possible solutions.

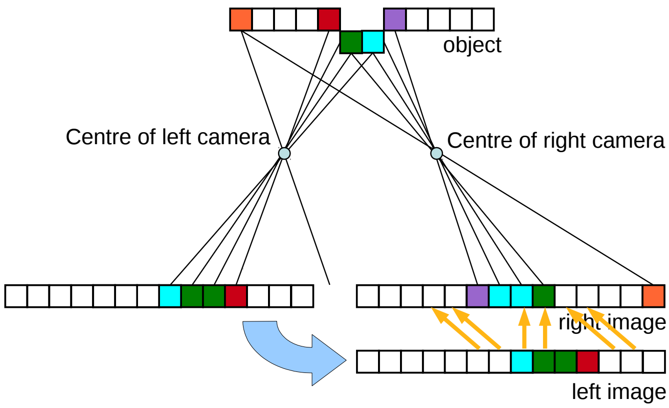
We are trying to use imperfect constraints to narrow down these many potential solutions to the correct one.



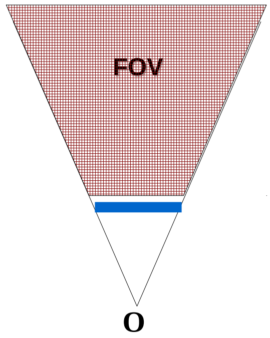
Correspondence problems

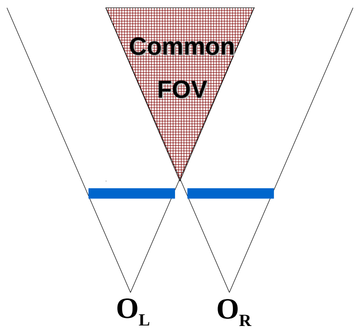
Some points in each image will have no corresponding points in the other image:

1. due to occlusion (e.g.)
2. the cameras might have different fields of view (e.g.)



Field of view

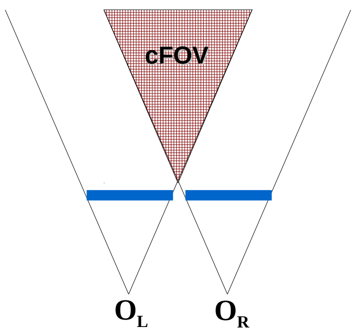


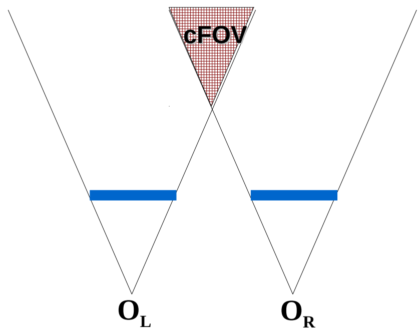


One camera can only take an image of those locations within its field of view (FOV)

A stereo pair of cameras can find depth for those locations within the common FOV of the two cameras

Stereo: coplanar cameras



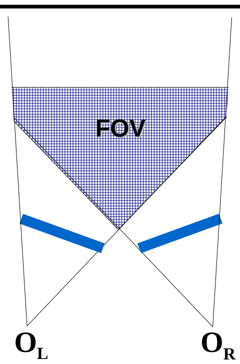


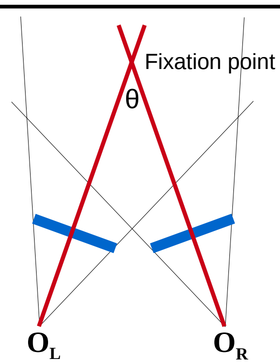
Short baseline
- large common FOV
- large depth error
(changes in depth cause only small changes in disparity)

$$z = f\,\frac{B}{d}$$

Long baseline
- small common FOV
- small depth error
(changes in depth cause larger changes in disparity)

Stereo: non-coplanar cameras





Intersecting optical axes
- large common FOV
- small depth error

Convergence angle θ (“vergence”)

Stereo: non-coplanar cameras

Disparity measured using angles instead of distance

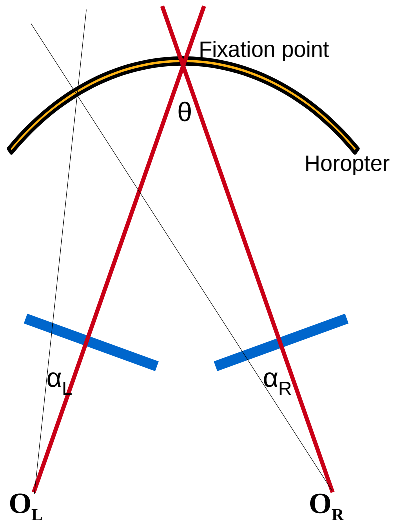
Disparity = $\alpha_L - \alpha_R$

Disparity = zero at fixation point

Disparity = zero at all points on a curve where rays at equal angles intersect.

Curve with zero-disparity called the "horopter"

Location of horopter depends on vergence angle.



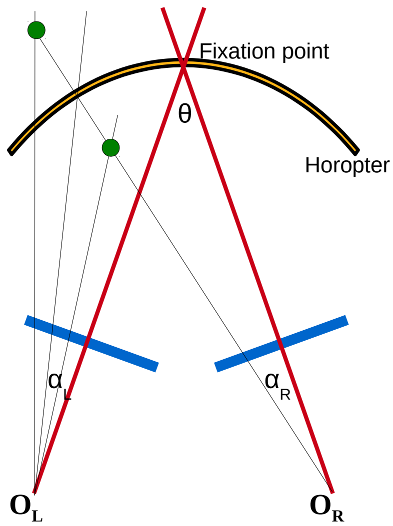
Stereo: non-coplanar cameras

Magnitude of disparity increases with the distance of objects from the horopter

$(\alpha_L - \alpha_R) > 0$: outside of the horopter

$(\alpha_L - \alpha_R) < 0$: inside the horopter

Need to be consistent with signs of angles, and order of subtraction.

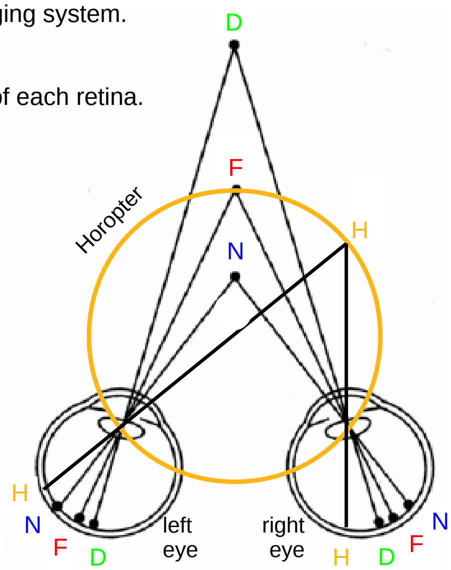


Stereo: non-coplanar eyes

Humans employ a non-coplanar imaging system.

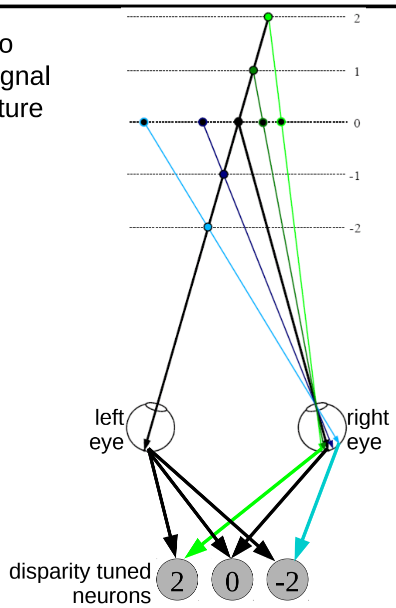
When point F is fixated:

- The images of F fall on the foveas of each retina.
- The images of points on the horopter (e.g. H) fall an equal distance from the foveas of each retina.
- The images of points nearer than the horopter (e.g. N) are displaced outwards ("crossed" disparties)
- The images of points more distant than the horopter (e.g. D) are displaced inwards ("uncrossed disparities)
- The further the point from the horopter, the greater the displacement.



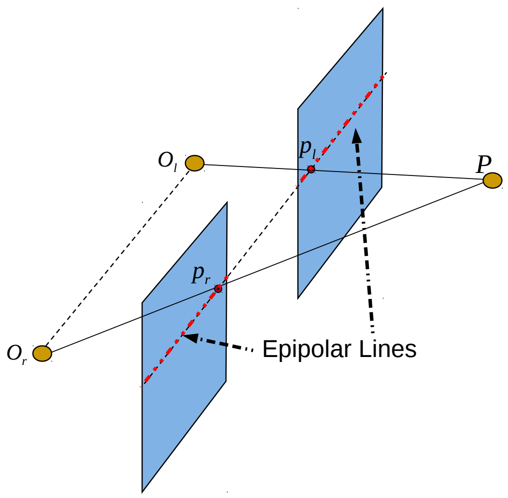
Stereo: non-coplanar eyes

Some cortical neurons are tuned to retinal disparity, and hence, can signal the depth of image points [see lecture 4].



Epipolar geometry

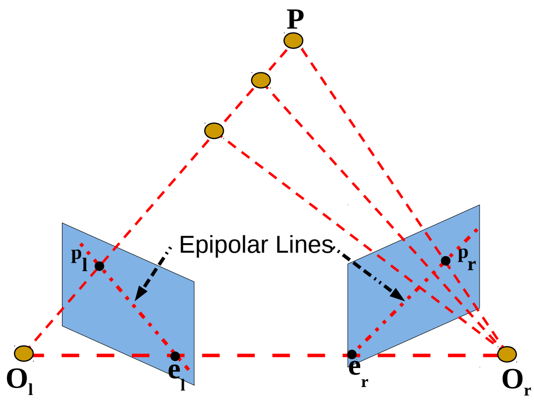
For coplanar cameras, corresponding points in the two images are on corresponding rows of each image.



Epipolar geometry

For non-coplanar cameras, corresponding points still occur on lines, but these are no longer horizontal lines.

A point in the image plane of the left camera corresponds to a line in the image plane of the right camera (and vice versa).



These lines which correspond to points in the other image are called epipolar lines.

2D search for correspondence can be reduced to a 1D search along the "epipolar" line

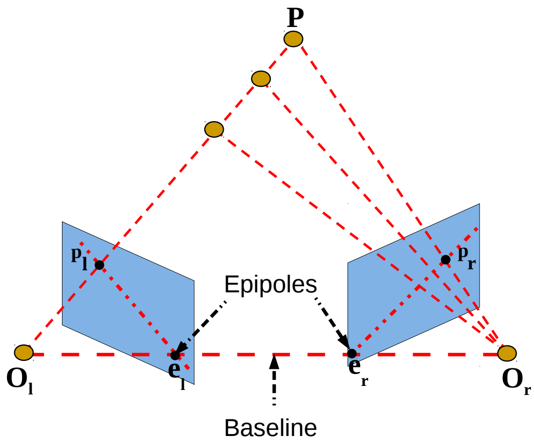
Epipolar geometry: terminology

Baseline: the line through the camera projection centres O_l, O_r

Epipole: projection of the optic centre of one camera on the image plane of the other camera.

The right (left) epipole e_r (e_l) is the image of the left (right) camera projection centre O_l (O_r) in the right (left) image plane.

Equivalently, the epipoles are the intersection of the baseline with each image plane.



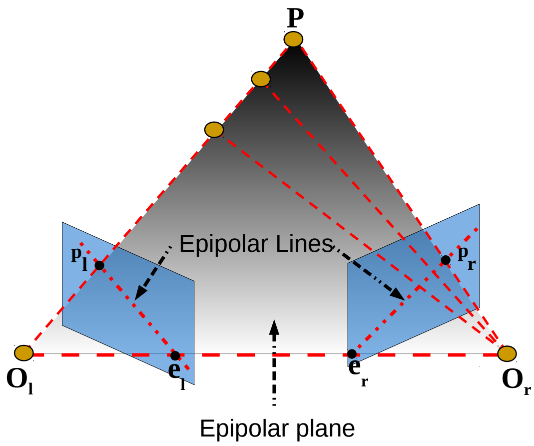
Epipolar geometry: terminology

Epipolar plane: a plane going through a particular 3D point and the optic centres of both cameras.

Epipolar lines: intersection of the epipolar plane (for a particular 3D point) and each image plane.

All epipolar lines in the left image go through e_l and all epipolar lines in the right image go through e_r.

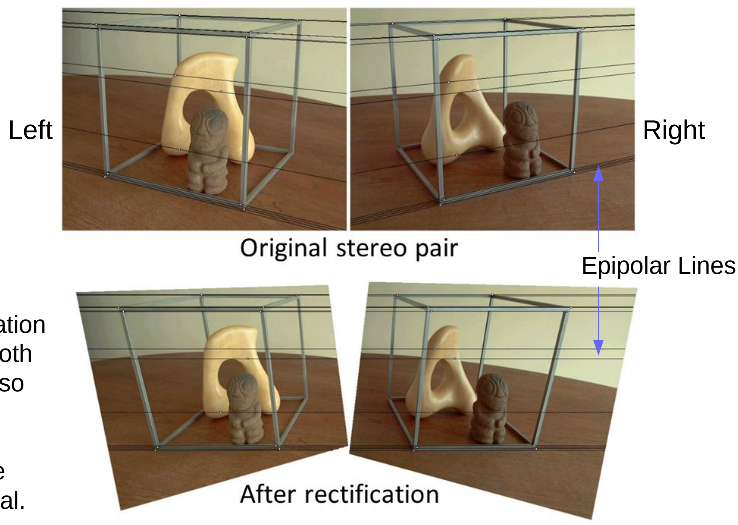
Conjugated epipolar lines: the epipolar lines generated by the same 3D point on the left and right image planes.



Epipolar Constraint: Corresponding points must lie on conjugated epipolar lines.

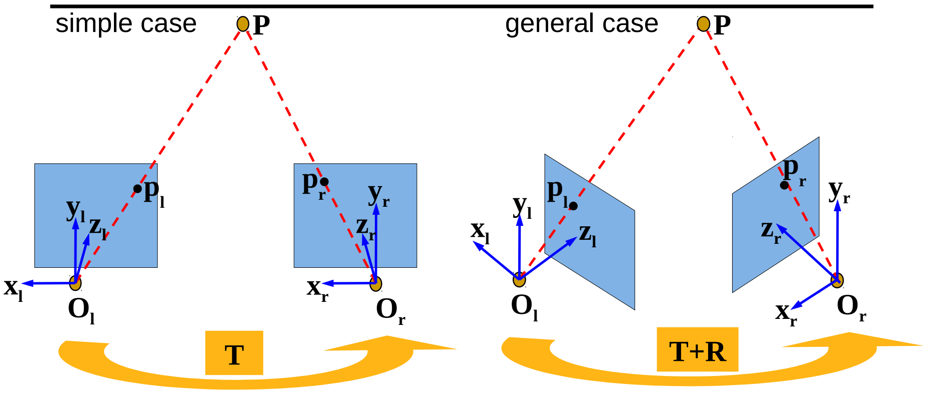
Rectification

Epipolar lines can be made parallel to the rows of the image via a transform called rectification.



Rectification warps both images so that all epipolar lines are horizontal.

Stereo geometry: summary



coplanar image planes (cameras translated along x-axis).

epipolar lines are horizontal scan lines

disparity measured in pixels

disparity inversely proportional to depth

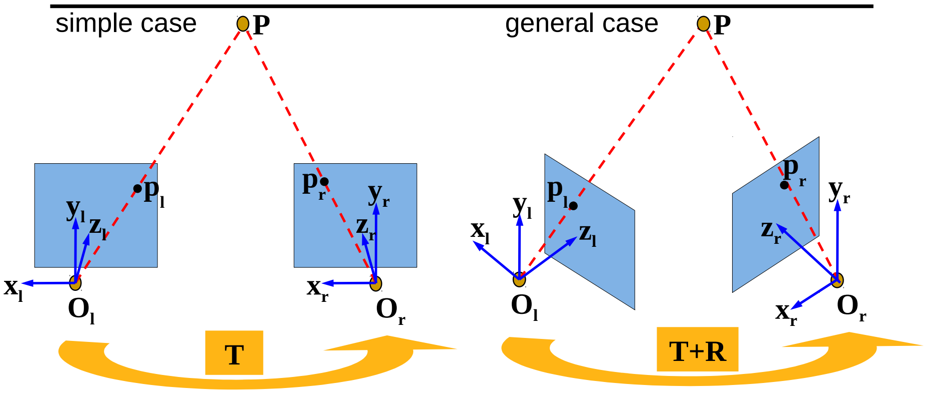
noncoplanar image planes (cameras related by a translation and rotation)

epipolar lines are lines radiating from the epipolar point

disparity measured using angles

disparity increases with distance from horopter

Stereo geometry: summary



depth = fB/d

depth found by, either:

1. triangulation (i.e. determining intersection of corresponding viewlines)

2. rectification of images and then treating as simple case

Both require knowledge of intrinsic and extrinsic camera parameters.

Depth

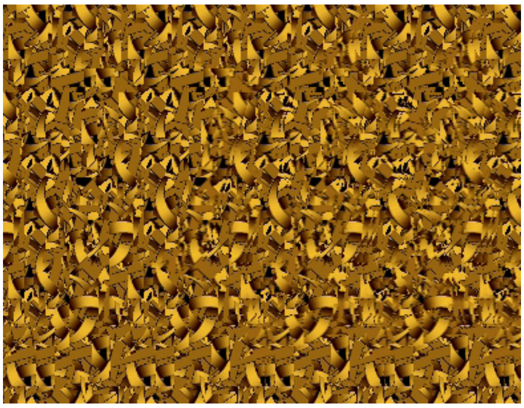
Recovery of depth information is important for:

- Controlling movement
  - mobile robot path planning
  - grasping an object with a robot arm
- 3D reconstruction
  - generating CAD models
  - virtual reality
- Object Recognition
  - depth (between objects) provides a cue for segmentation
  - depth (within an object) provides information about the shape of an object

Stereo is only one source of information about depth...

Cues to depth: Binocular cues

Disparity: difference in location of corresponding points in a stereo pair of images.



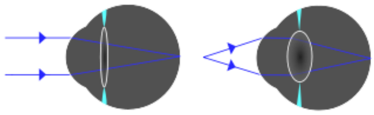
an autostereogram

Stereo vision is important (try catching a ball with one eye closed), and is sufficient (random dot stereograms and autostereograms demonstrate this as all other cues to depth have been removed).

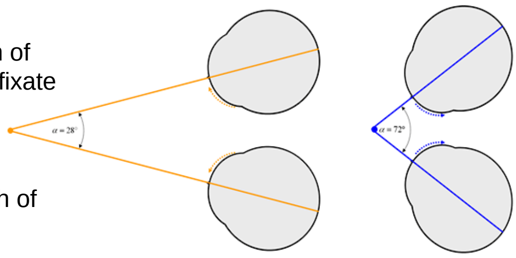
However, disparity is only one of many depth cues.

Cues to depth: Oculomotor cues

Accommodation: the shape of the lens (and hence its refractive power) can be modified by muscles in the eye to bring objects at different depths into focus. Hence, the state of these muscles provides information about the depth of the object being observed.



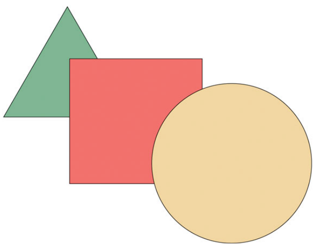
Convergence: the rotation of eyes/cameras can vary to fixate objects at different depths. Hence, the angle of convergence provides information about the depth of the object being fixated.



Cues to depth: Monocular cues

i.e. cues that provide depth information with one eye closed

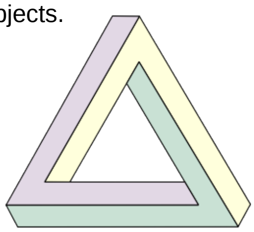
Interposition: when the boundary of one object is interrupted by the presence of another object, we use this pattern of occlusion to determine the relative depth of the objects. The near object is perceived as “interposed” between the far object and the observer.



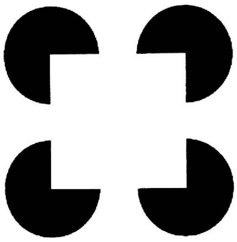
Cues to depth: Monocular cues

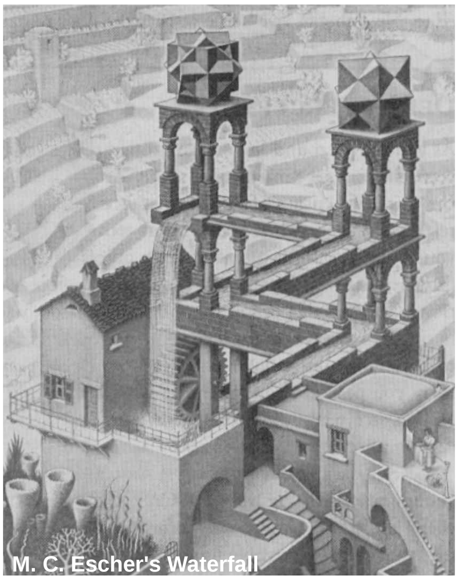
Interposition: manipulating interposition depth cues can produce impossible objects, or illusory objects.

Penrose triangle



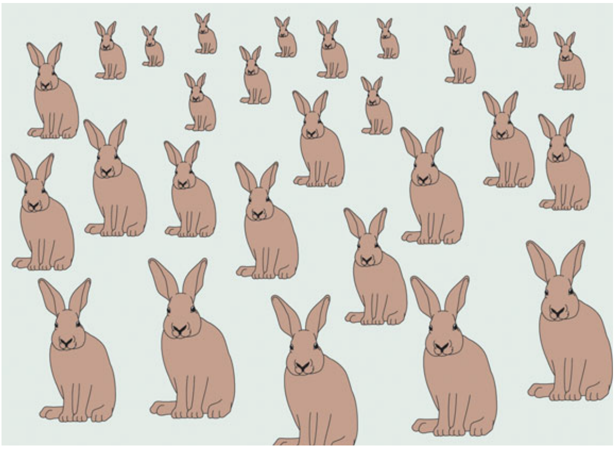
kanizsa square





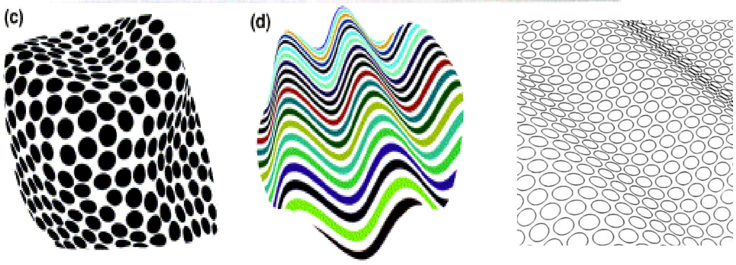
Cues to depth: Monocular cues

Size familiarity: distant objects necessarily produce a smaller image than nearby objects of the same size. The larger of two identical objects tends to be perceived as closer that the smaller one.



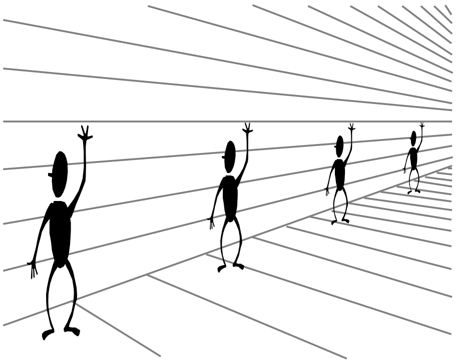
Cues to depth: Monocular cues

Texture gradients: for uniformly textured surfaces, the texture elements get smaller and more closely spaced with distance.



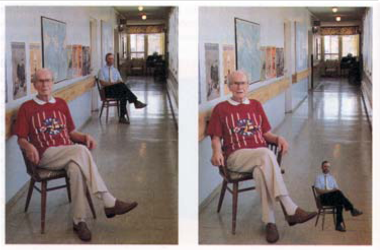
Cues to depth: Monocular cues

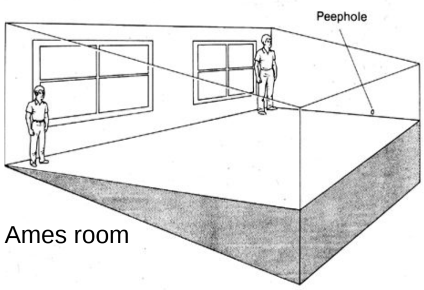
Linear perspective: the property of parallel lines converging at infinity provides a cue to depth (and size).



Cues to depth: Monocular cues

Linear perspective: manipulating perspective can produce unusual perceptions (overcoming size familiarity).





Cues to depth: Monocular cues

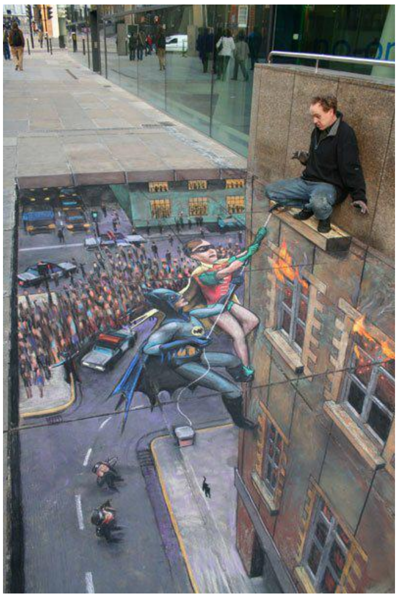
Linear perspective: manipulating perspective can produce unusual perceptions (overcoming prior beliefs about gravity!).

Cues to depth: Monocular cues

Linear perspective: manipulating perspective can produce a strong perception of 3D structure, e.g. Trompe L'oeil art.

Cues to depth: Monocular cues

Linear perspective: manipulating perspective can produce a strong perception of 3D structure, e.g. Trompe L'oeil art.

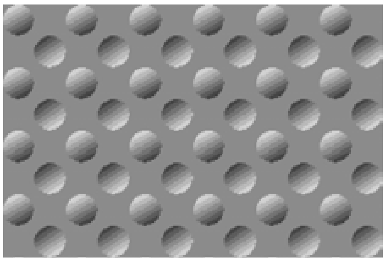


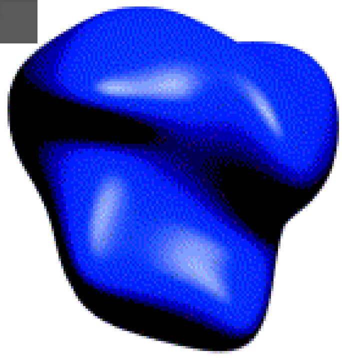
Cues to depth: Monocular cues

Aerial perspective: due to the scattering of light by particles in the atmosphere, distant objects look fuzzier and have lower luminance contrast and colour saturation.

Cues to depth: Monocular cues

Shading: the distribution of light and shadow on objects provides a cue for depth (the brain assumes, usually, that light comes from above).





Cues to depth: Monocular cues

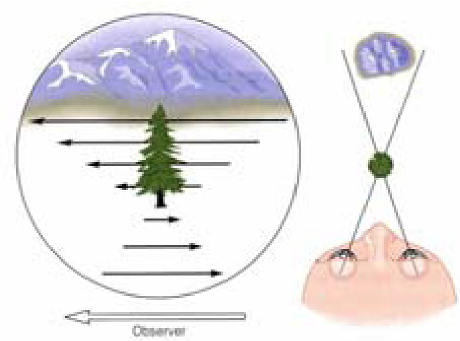
Shading: the distribution of light and shadow on objects provides a cue for depth (the brain assumes, usually, that light comes from above).

Hollow face illusion

shading fails to recover depth when pitted against familiarity

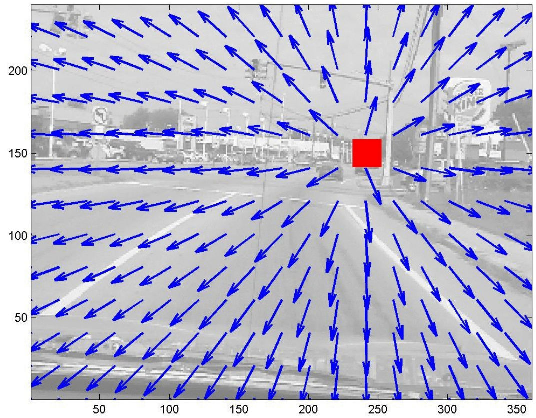
Cues to depth: Motion induced cues

Motion parallax: speed and direction of image motion induced by movement of the camera/eye varies with depth. Objects closer than the fixation point appear to move in a direction opposite to the observer, while objects further away appear to move in the same direction.



Cues to depth: Motion induced cues

Optic Flow: As a camera moves forward or backward, the pattern of stimulation across the entire visual field changes, producing a pattern of expanding or contracting “optic flow”. Points closer to the camera move more quickly across the image plane

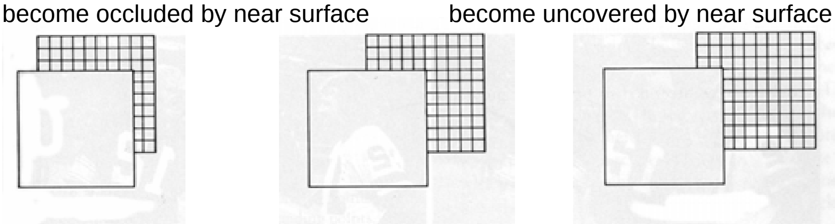


Cues to depth: Motion induced cues

Accretion and deletion : parts of an object can appear or disappear when an observer moves relative to two surfaces that are at different depths.

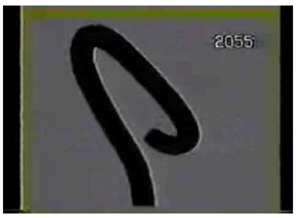
Motion to left causes far surface to become occluded by near surface

Motion to right causes far surface to become uncovered by near surface



Cues to depth: Motion induced cues

Structure from motion (kinetic depth): movement of an object or of the camera can induce the perception of 3D structure.



Depth is conveyed solely by the variation in velocity and direction of each dot: no depth is seen in static image.

Summary: Stereo Vision

Aim: to infer depth from 2 or more images taken from different viewpoints. Two sub-problems:

1. Correspondence. Determining which points in the left and right images are projections of the same point in the scene. Constraints:

- epipolar constraint (corresponding points lie along the epipolar line)
- maximum disparity (extent of search reduced by knowledge of baseline and minimum depth)
- continuity (disparity varies smoothly assuming continuous surfaces)
- uniqueness (one-to-one matches assuming surfaces are approximately perpendicular to camera and no occlusion)
- ordering (points occur in same order in each image assuming no occlusion)

Summary: Stereo Vision

2. Reconstruction. Given the correspondence between points, and camera parameters, calculate the depth (z).

Depth related to disparity (d).

For coplanar cameras: $d = x'_L - x'_R$

$$z = f\frac{B}{d}$$

For non coplanar cameras: $d = \alpha_L - \alpha_R$

$d > 0$ outside of the horopter

$d < 0$ inside the horopter

Summary: Depth

Binocular
- Disparity

Oculomotor
- Accommodation (shape of lens in eye)
- Convergence (angle of rotation of cameras/eyes)

Monocular
- Interposition (occlusion of one object/part of object by another)
- Size familiarity (the smaller the object the greater its depth)
- Texture gradients (texture become smaller and denser with depth)
- Linear perspective (convergence of lines at vanishing points)
- Aerial perspective (reduction in contrast and saturation with depth)
- Shading (light and shadow)

Motion
- Motion parallax (depth related to image motion caused by camera motion)
- Optic Flow (depth related to speed of image expansion/contraction)
- Accretion and deletion (changes in occlusion due to camera movement)
- Structure from motion (depth perception due to object motion)