universität innsbruck

iis

Lecture 7. Stereopsis I

703142. Computer Vision

Assoz.Prof. Antonio Rodríguez-Sánchez, PhD.

Most slides thanks to Richard Wildes and Antonio Torralba

Tentative schedule

Date	Topic
06-03-2021	Introduction.
13-03-2021	Review: Image Formation. Introduction to OpenCV.
20-03-2021	Review: Image processing.
27-03-2021	Review: Feature extraction.
17-04-2021	Motion I: Optical Flow. OpenCV.
24-04-2021	Motion II: Spatiotemporal filters. Spatiotemporal analysis.
08-05-2021	Stereopsis I: Correspondence. Epipolar geometry.
15-05-2021	Stereopsis II: RANSAC. Reconstruction.
22-05-2021	Object Recognition I: Categories. Photo editing.
05-06-2021	Object Recognition II: Objects, faces, instances. Viola and Jones
12-06-2021	Invited talk: Neuroscience in stereo and motion.
19-06-2021	Object Recognition III: Deep Learning. CNNs in Pytorch. Exam questions.
26-06-2021	FINAL EXAM.

Outline

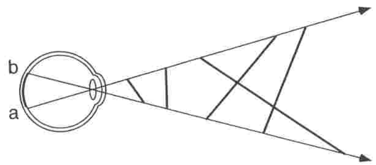
- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

Outline

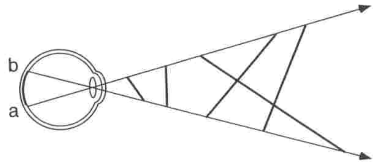
- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

Introduction

Introduction

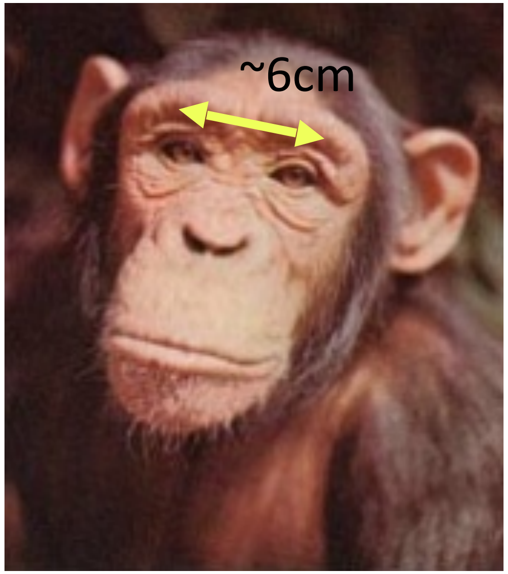


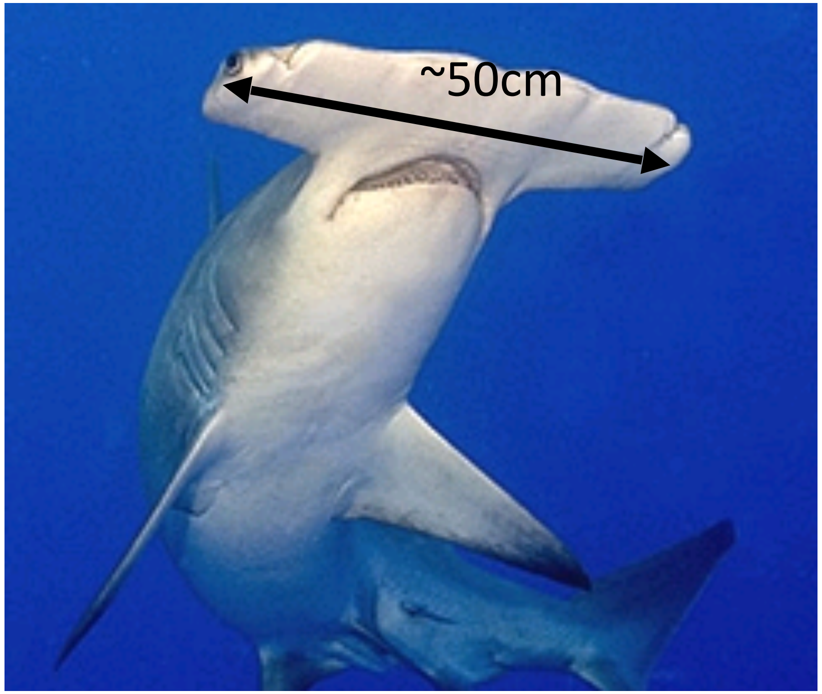
Introduction



Introduction

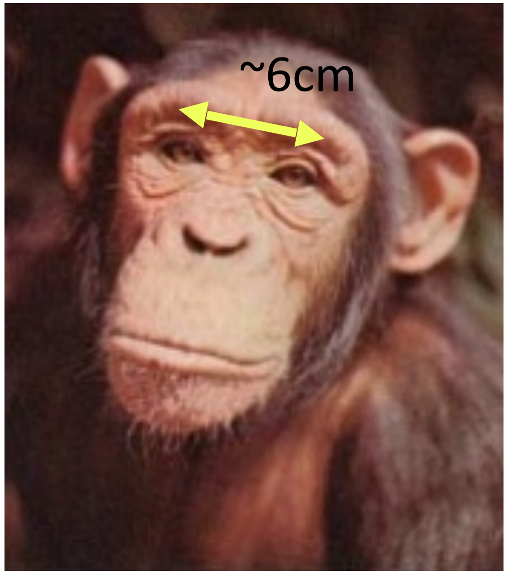
- Stereo vision in biological systems

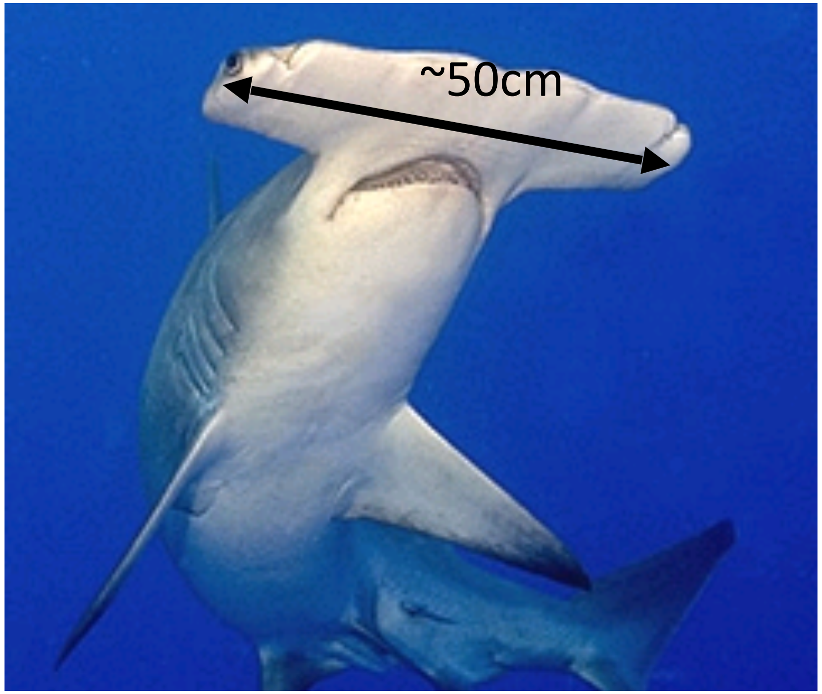




Introduction

- Stereo vision in biological systems





- After 10 meters disparity is quite small and depth from stereo is unreliable…

Introduction

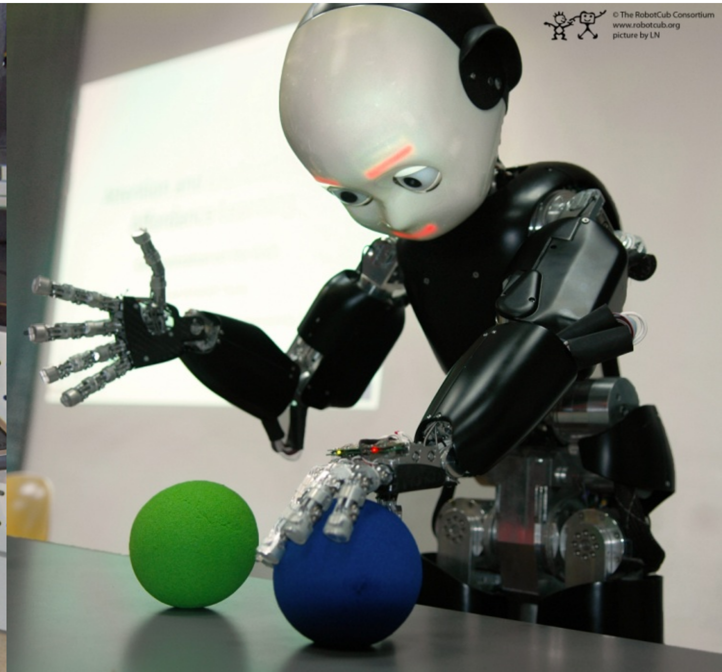
- Stereo cameras

Introduction

- Stereo computer vision cameras (e.g. Bumblebee)

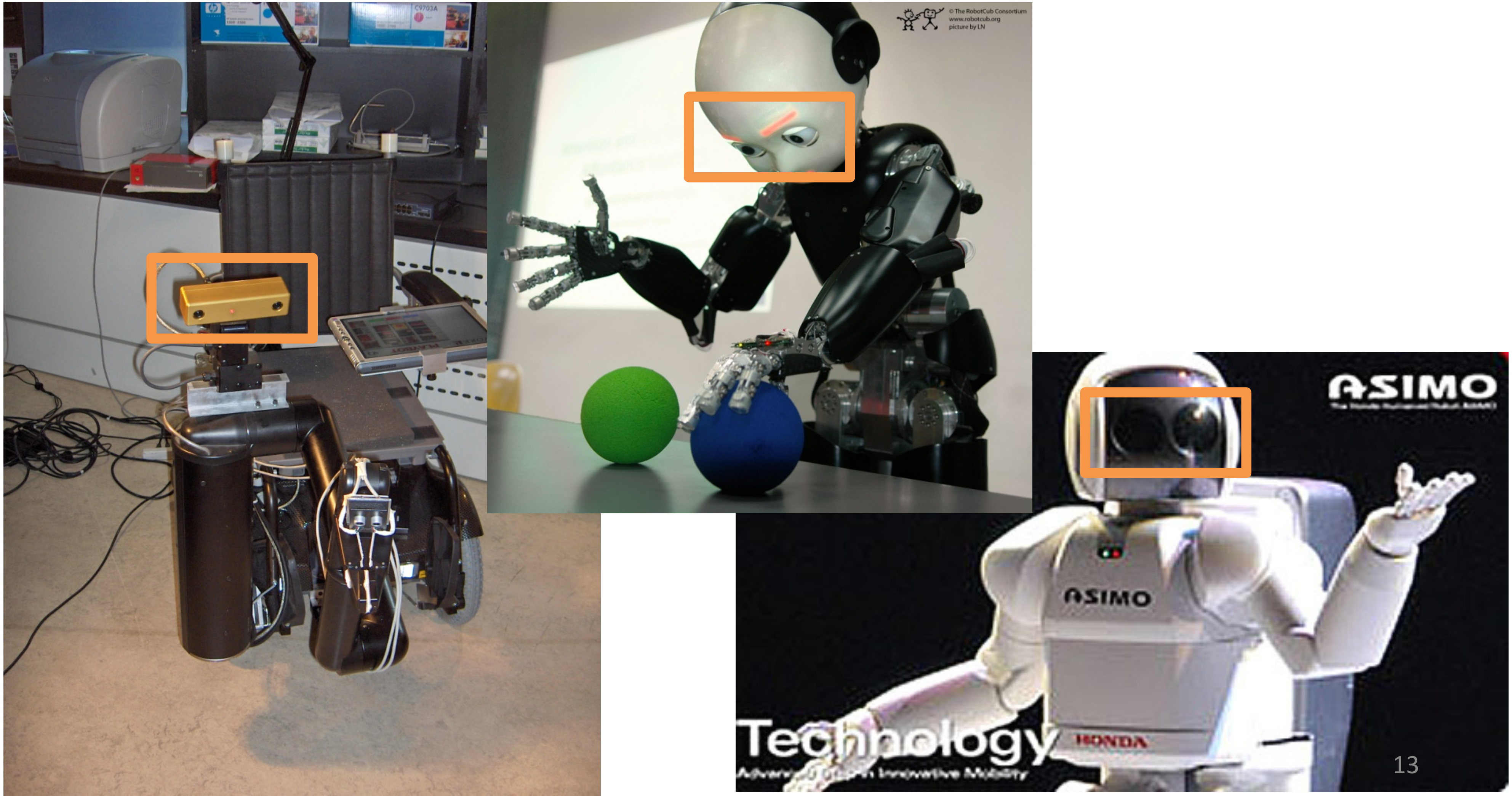
Introduction

- Stereo computer vision systems



Introduction

- Stereo computer vision systems



Curiosity

HONDA

Evolution of Asimo

ASIMO
http://www.honda.co.jp/ASIMO/

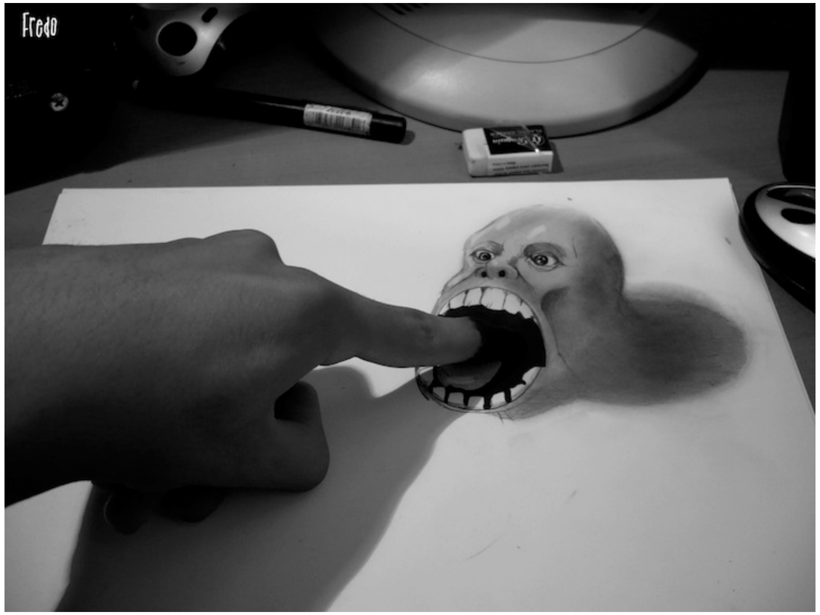
Outline

- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

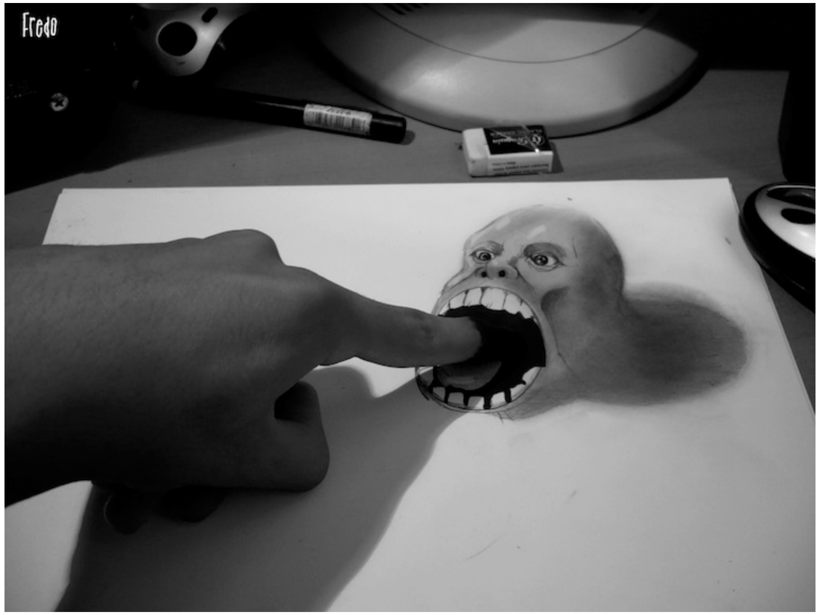
3D shapes from 2D images

- Shape-from-X
  - Various sources of image derived information can support the inference of three dimensional shape.
  - The term shape-from-X, with X being bound to some particular image-based information source, is used to refer collectively to such methods for shape recovery from images.
  - In some cases X involves multiple images
    - Binocular stereo
    - Motion parallax
    - Focus
  - In some cases X requires only a single image.
  - Visual artists exploit the human ability to perform shape-from-X to depict 3D via 2D renderings.

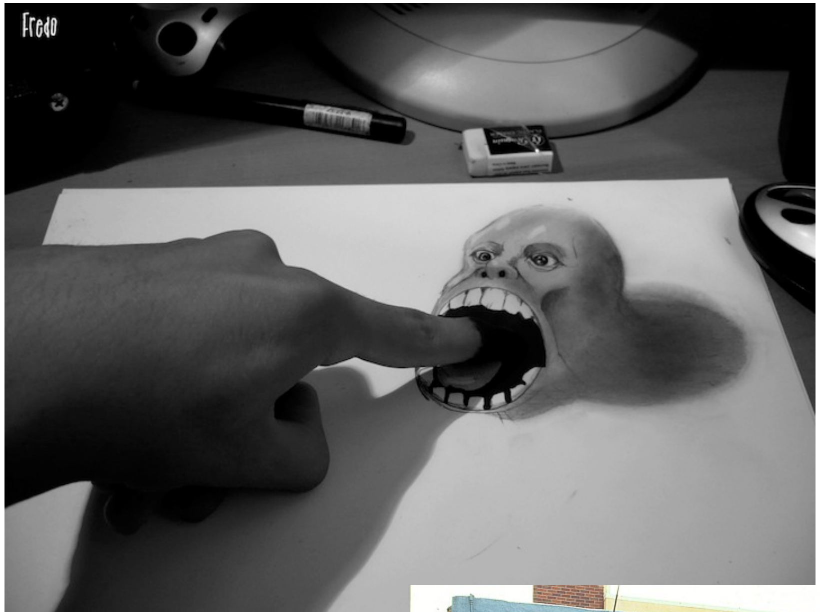
3D shapes from 2D images



3D shapes from 2D images



3D shapes from 2D images



3D shapes from 2D images

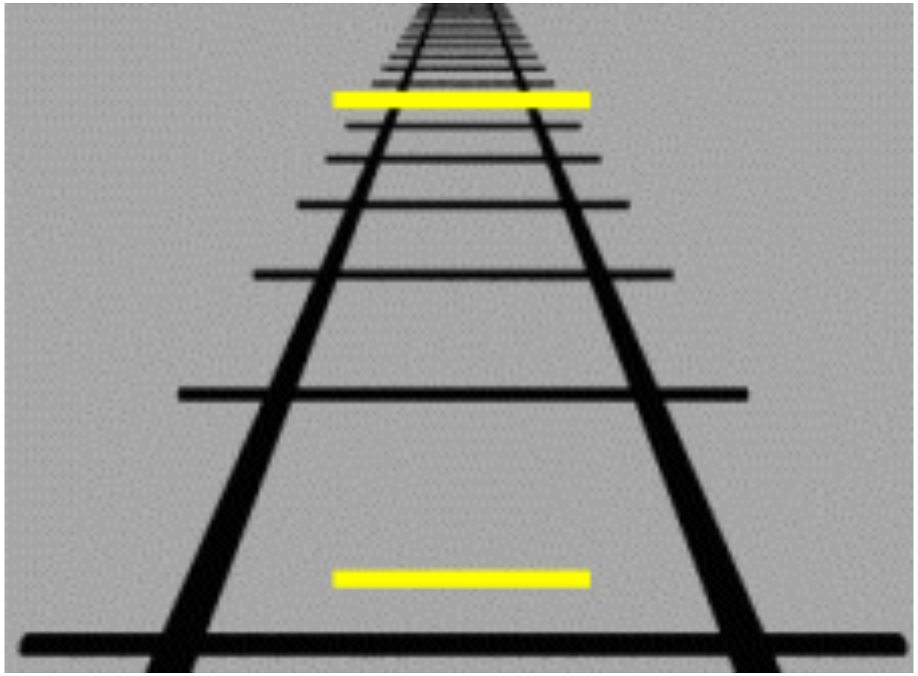
- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

- Simple image cues
  - Perspective
    - Based on the apparent convergence of parallel lines to common vanishing points with increasing distance from the observer.
    - Characteristic of the visual field rather than the visual world. It approximates how we see (the retinal image) rather than what we see, the objects in the world.
      (Gibson: “Perspective order”)

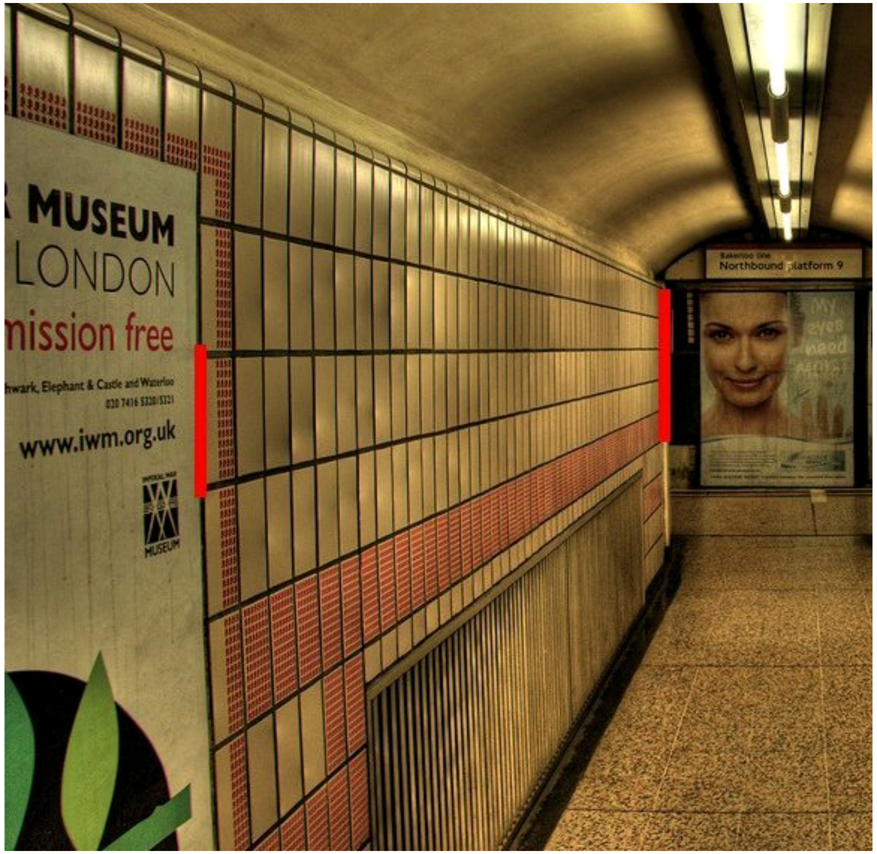


Ponzo’s illusion

Source: Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Perspective
    - Based on the apparent convergence of parallel lines to common vanishing points with increasing distance from the observer.
    - Characteristic of the visual field rather than the visual world. It approximates how we see (the retinal image) rather than what we see, the objects in the world.
      (Gibson: "Perspective order")

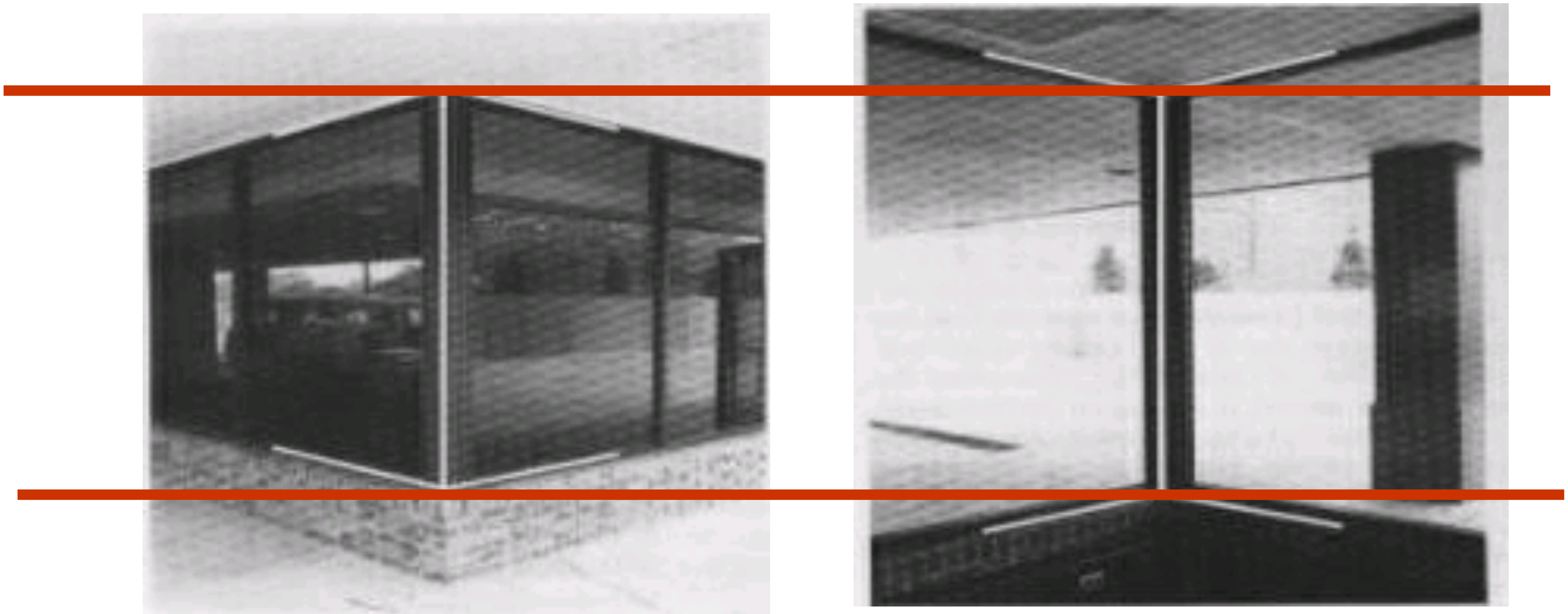


Walt Anthony (2006)

Source: Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Perspective



Muller-Lyer (1889)

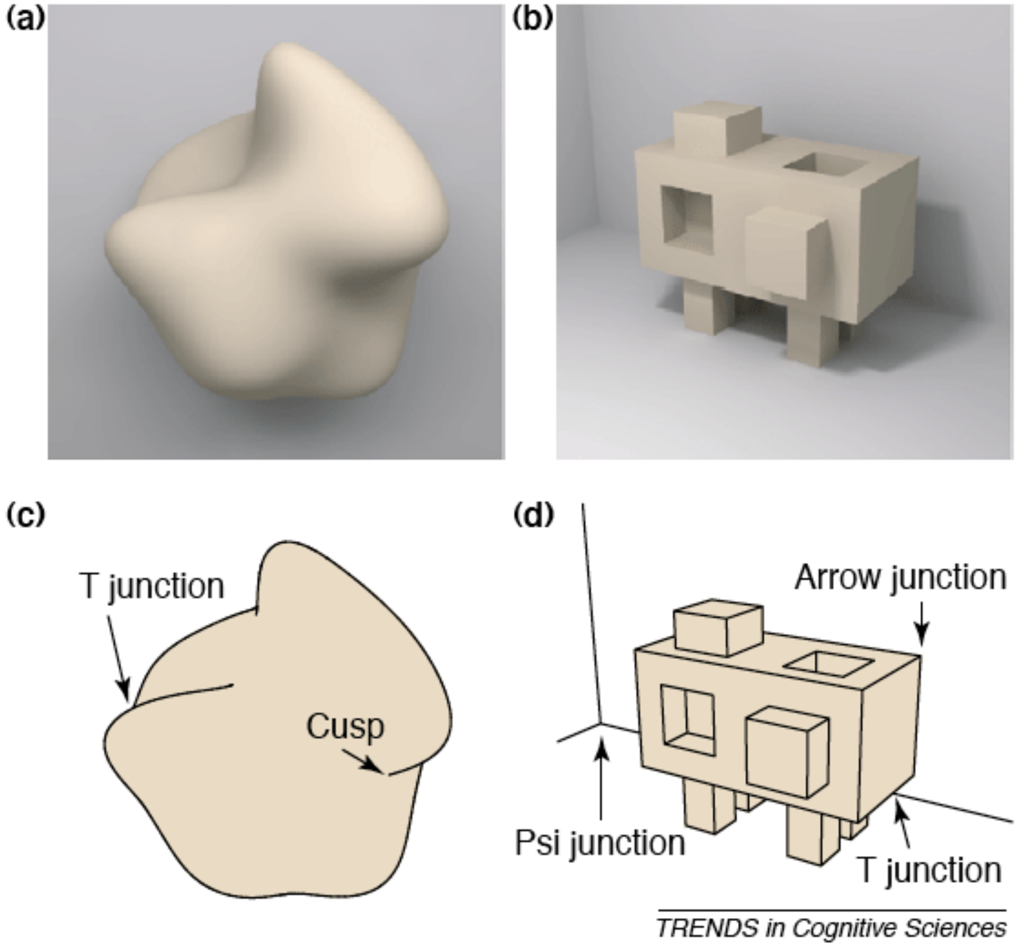
Source: Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

- Simple image cues
  - Contour
    - Outline of a silhouette

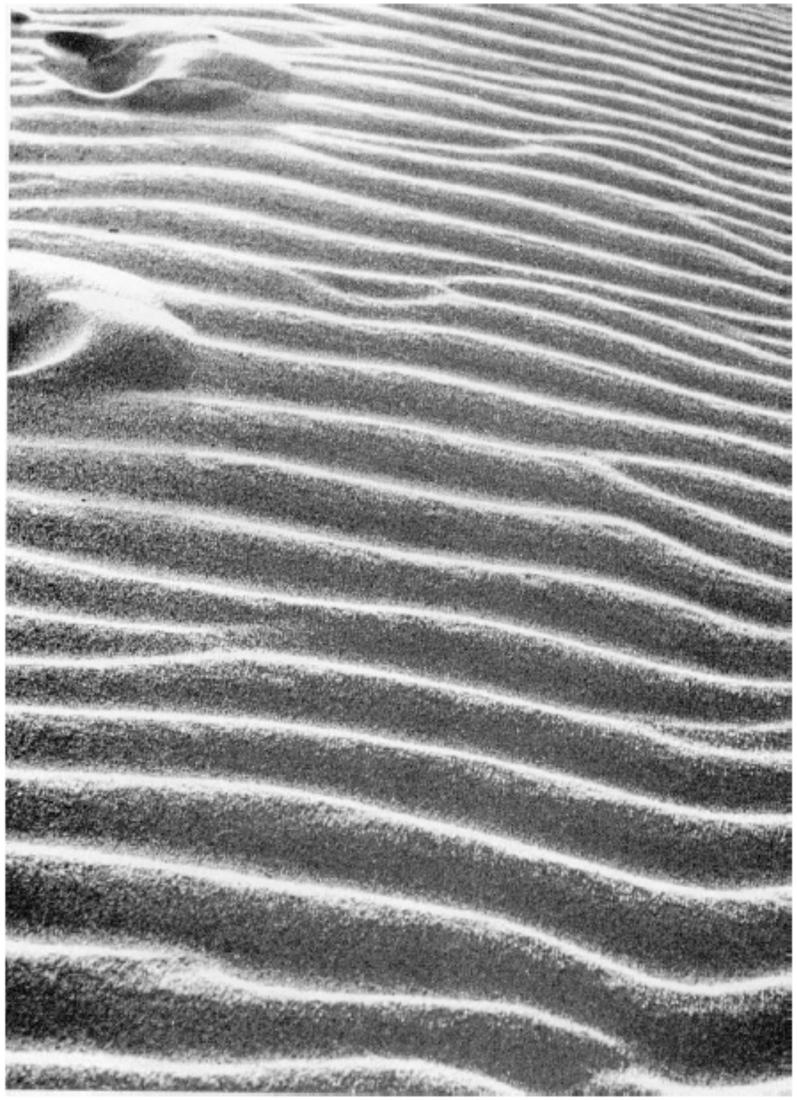


3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

- Simple image cues
  - Texture
    - Variations on surface



3D shapes from 2D images

- Simple image cues
  - Texture
    - Variations on surface

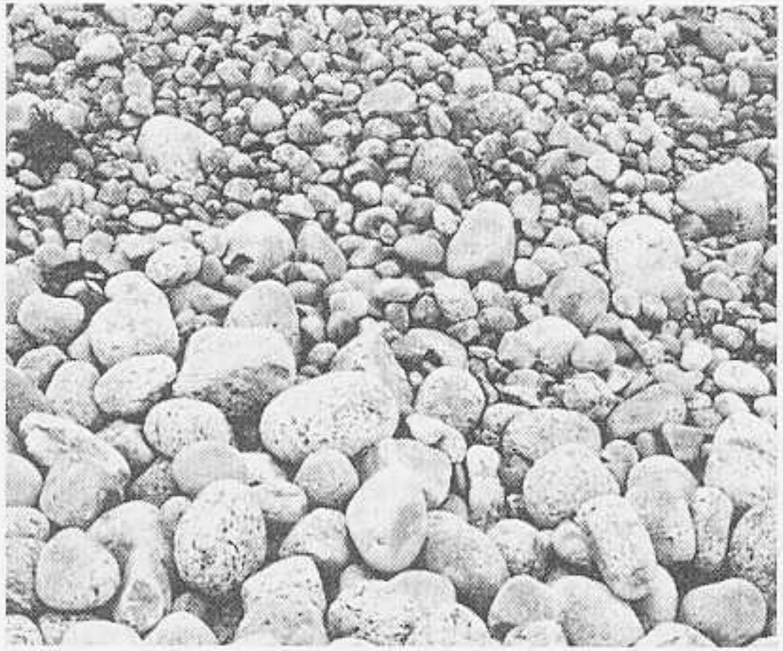


FIGURE 8.27
Texture gradients provide information about depth. (Frank Siteman/Stock, Boston.)
© Frank Sitman/Stock Boston

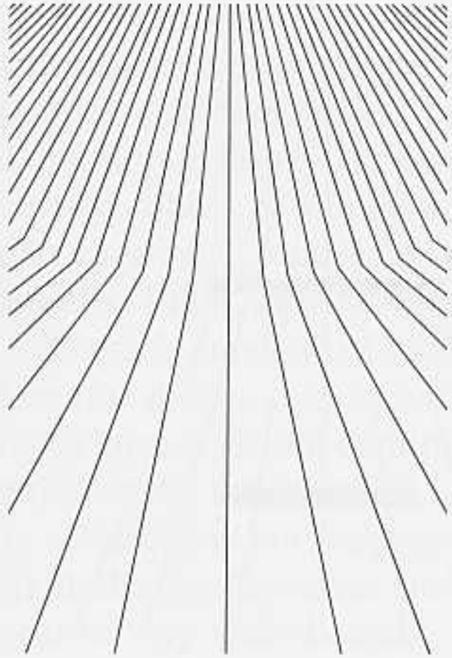


FIGURE 8.28
Texture discontinuity signals the pre
corner.

3D shapes from 2D images

- Simple image cues
  - Texture
    - Variations on surface

3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

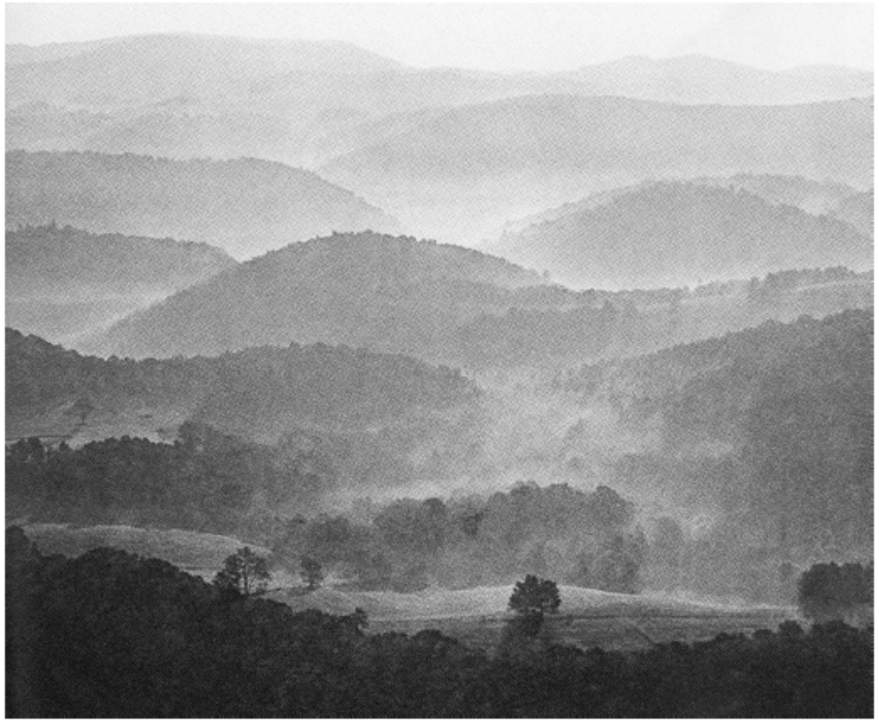
3D shapes from 2D images

- Simple image cues
  - Aerial perspective
    - Based on the effect of air on the color and visual acuity of objects at various distances from the observer.
    - Consequences:
      - Distant objects have lower contrast
      - Distant objects appear bluer

Source: Richard Wildes and Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Aerial perspective
    - Based on the effect of air on the color and visual acuity of objects at various distances from the observer.
    - Consequences:
      - Distant objects have lower contrast
      - Distant objects appear bluer



Source: Richard Wildes and Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Aerial perspective
    - Based on the effect of air on the color and visual acuity of objects at various distances from the observer.
    - Consequences:
      - Distant objects have lower contrast
      - Distant objects appear bluer

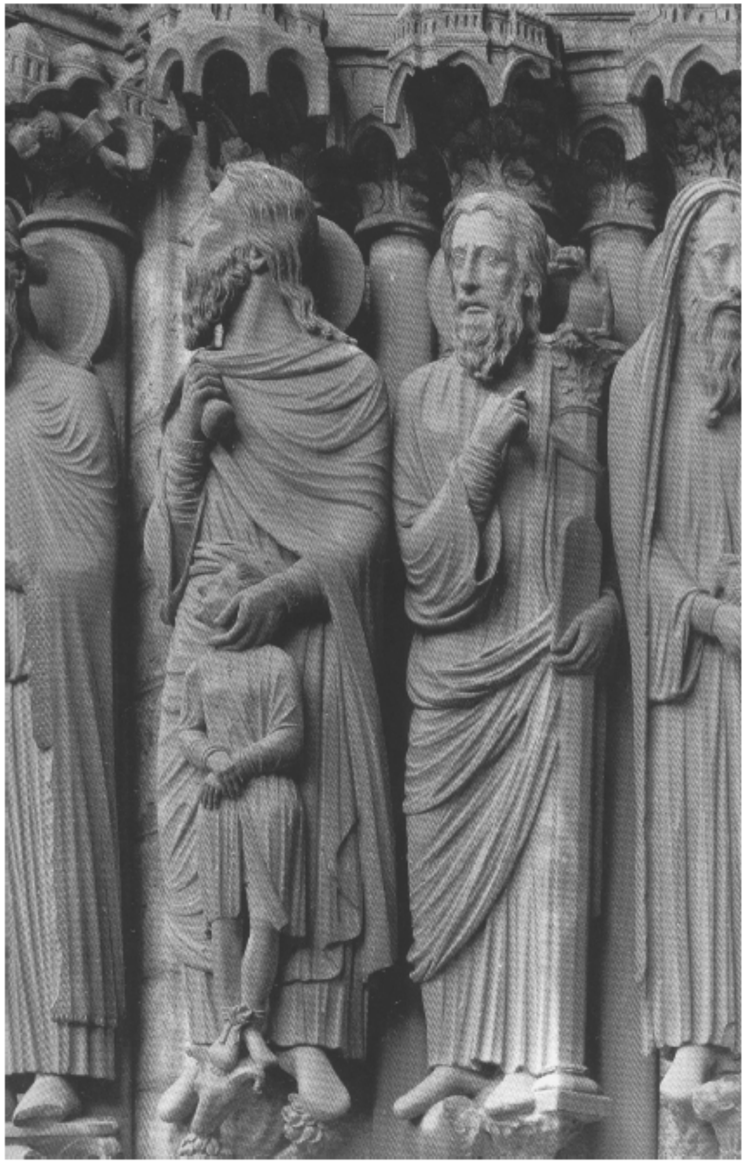
Source: Antonio Torralba slides

3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

- Simple image cues
  - Shading



3D shapes from 2D images

- Simple image cues
  - Perspective
  - Contour
  - Texture
  - Aerial perspective
  - Shading
  - Familiar size

3D shapes from 2D images

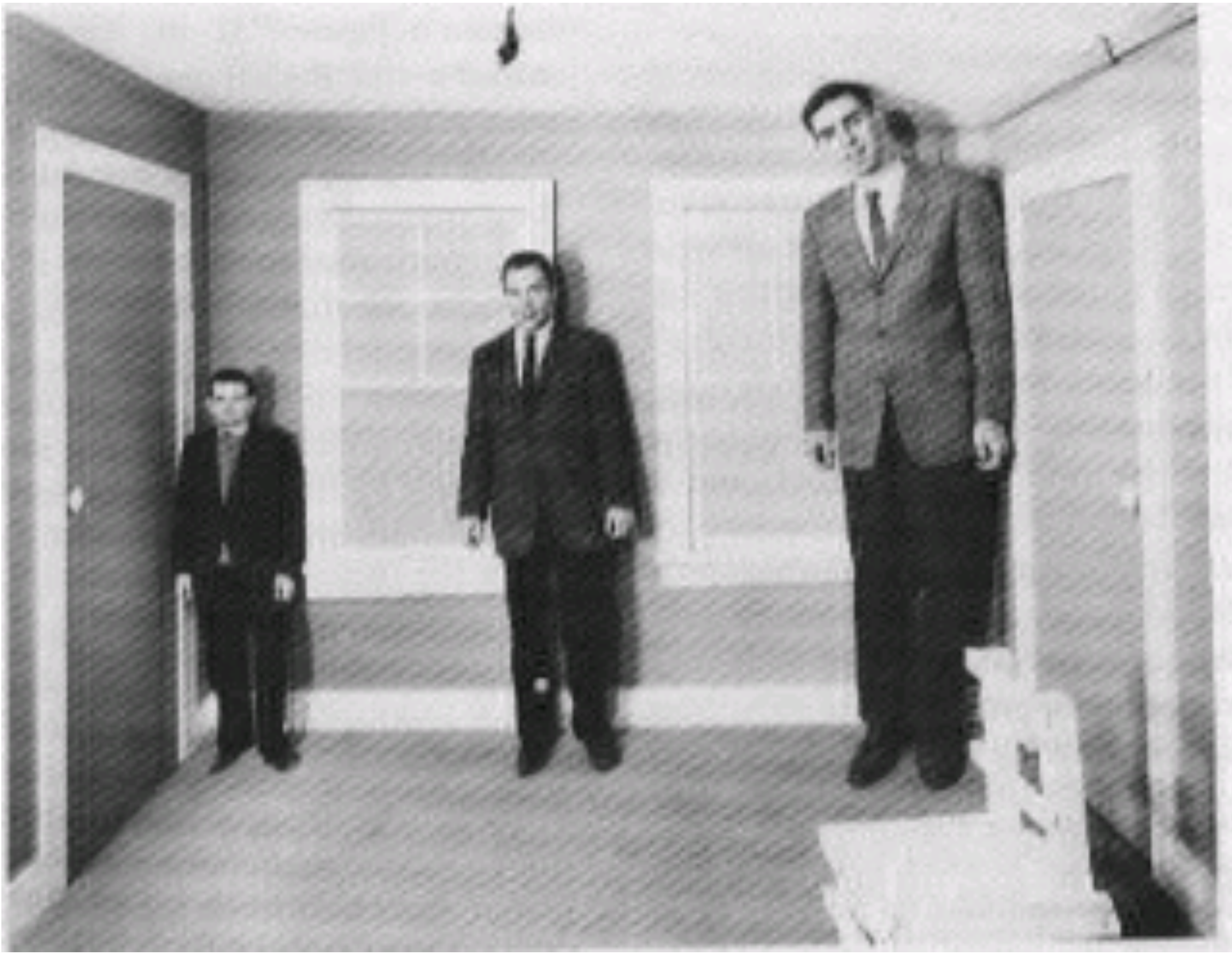
- Simple image cues
  - Familiar size
    - Apparent reduction in size of objects at a greater distance from the observer
    - Size perspective is thought to be conditional, requiring knowledge of the objects.
    - But, material textures also get smaller with distance, so possibly, no need of perceptual learning?

3D shapes from 2 images

- Simple image cues
  - Familiar size
    - Apparent reduction in size of objects at a greater distance from the observer
    - Size perspective is thought to be conditional, requiring knowledge of the objects.
    - But, material textures also get smaller with distance, so possibly, no need of perceptual learning?

Curiosity

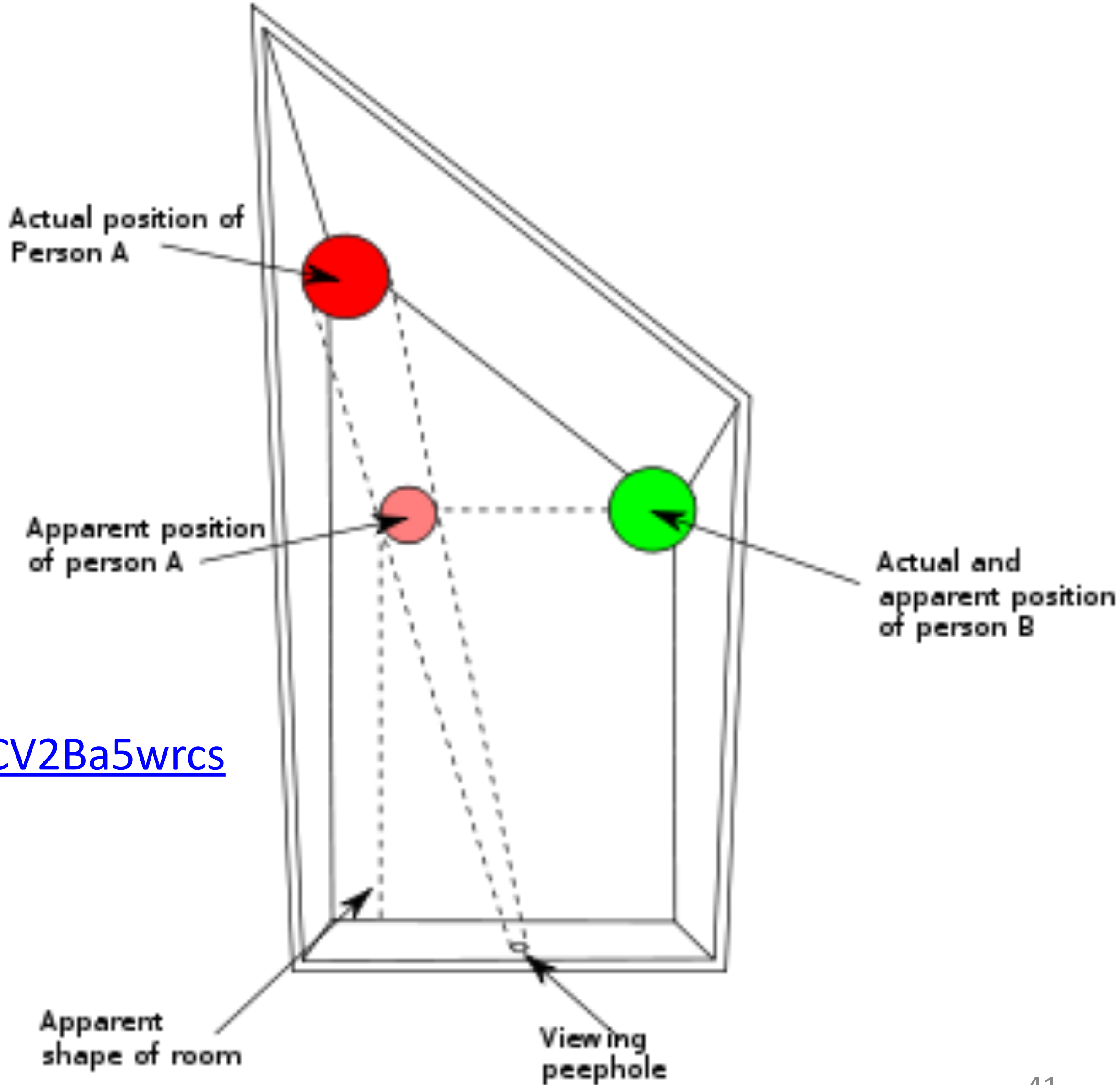
- Perspective vs familiar size



Curiosity

THE AMES ROOM

In the Ames room illusion, two people standing in a room appear to be of dramatically different sizes, even though they are the same size.



http://www.youtube.com/watch?v=hCV2Ba5wrcs

Outline

- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

Stereo vision

- Definition
  - The inference of information about the 3D structure of a scene from two or more images is referred to as **stereo vision**.

Stereo vision

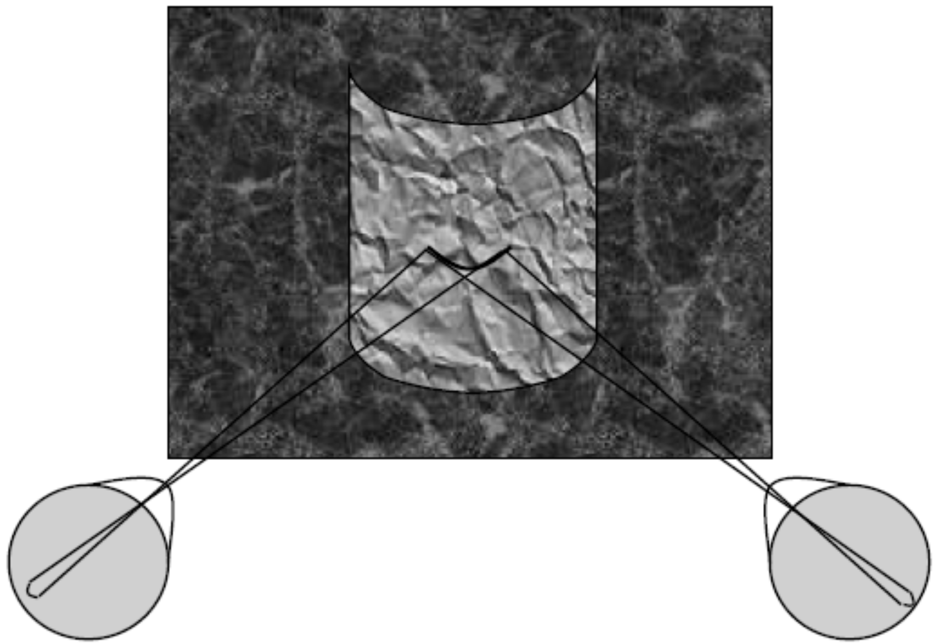
- Definition
  - The inference of information about the 3D structure of a scene from two or more images is referred to as **stereo vision**.
- Information of interest may take the form of
  - Distance measurements (range map)
  - Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities

Stereo vision

- Minimally, two or more images are required with spatial displacement of the scene and/or sensor.
- For the case of two spatially displaced sensors we speak of binocular stereo.
  - We will concentrate on this case.

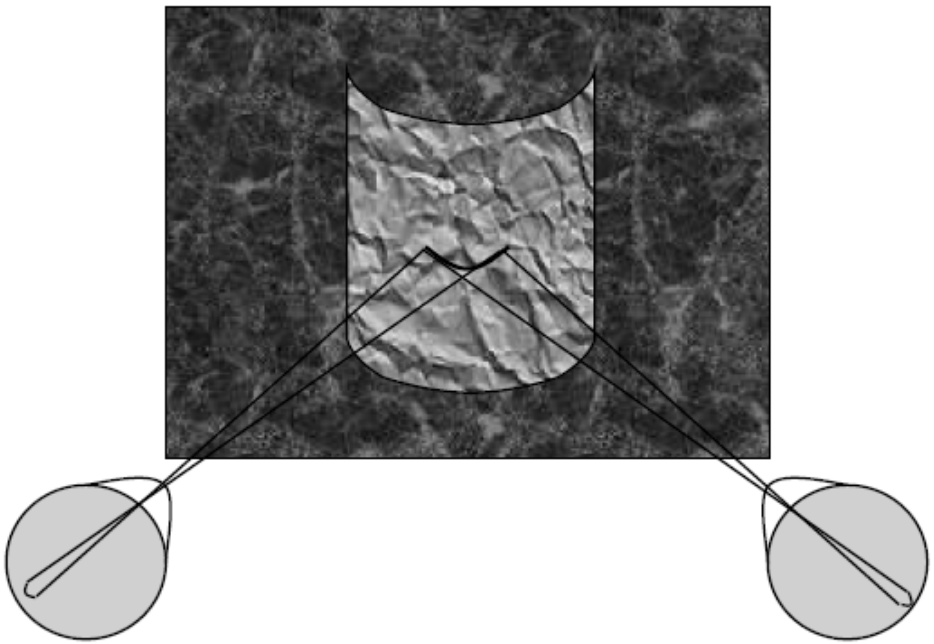
Stereo vision

- Minimally, two or more images are required with spatial displacement of the scene and/or sensor.
- For the case of two spatially displaced sensors we speak of binocular stereo.
  - We will concentrate on this case.



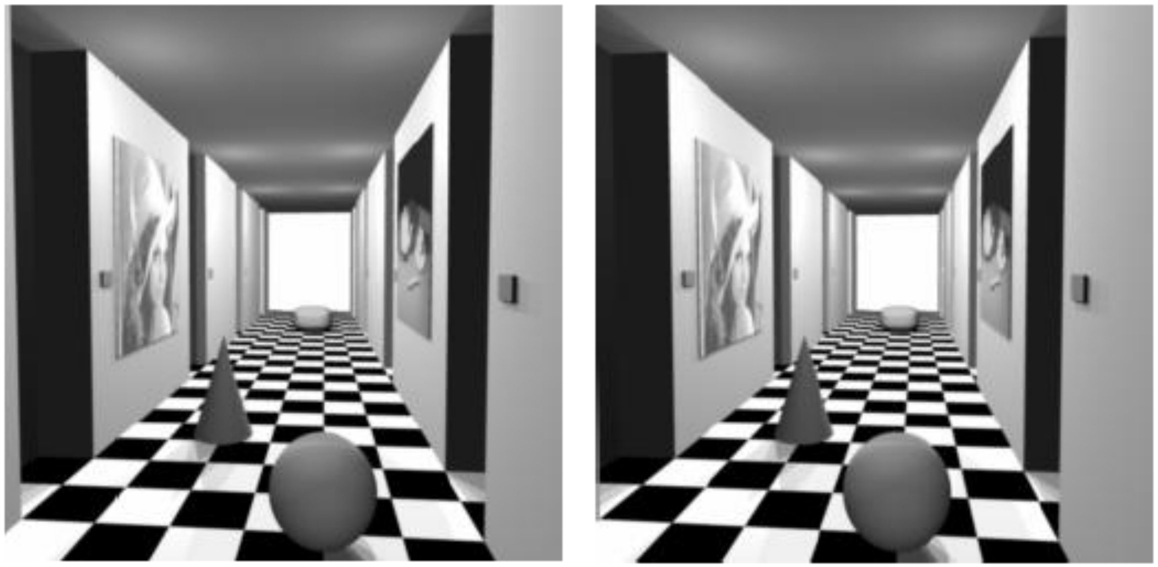
Stereo vision

- Owing to the geometry of the situation
  - 3D scene points will project to different locations in a pair of spatially displaced optical sensors
  - From this difference in location we recover the 3D information.



Stereo vision

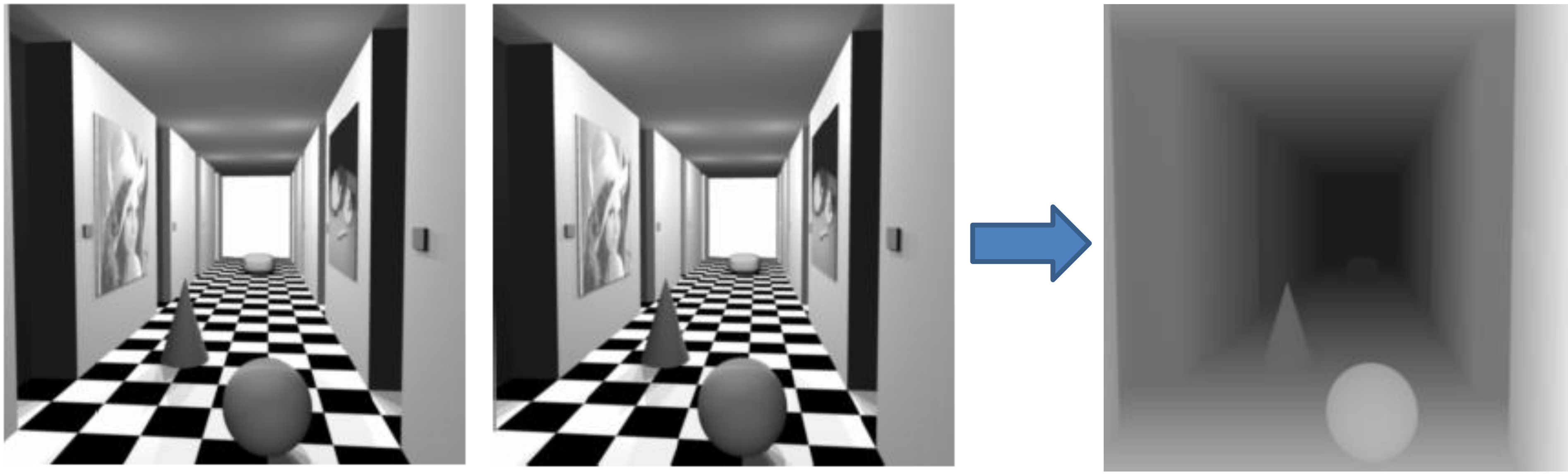
- Owing to the geometry of the situation
  - 3D scene points will project to different locations in a pair of spatially displaced optical sensors
  - From this difference in location we recover the 3D information.



Stereo pair

Stereo vision

- Owing to the geometry of the situation
  - 3D scene points will project to different locations in a pair of spatially displaced optical sensors
  - From this difference in location we recover the 3D information.

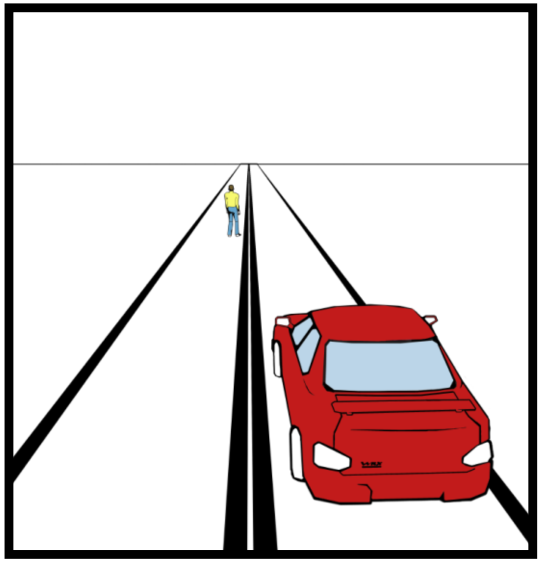


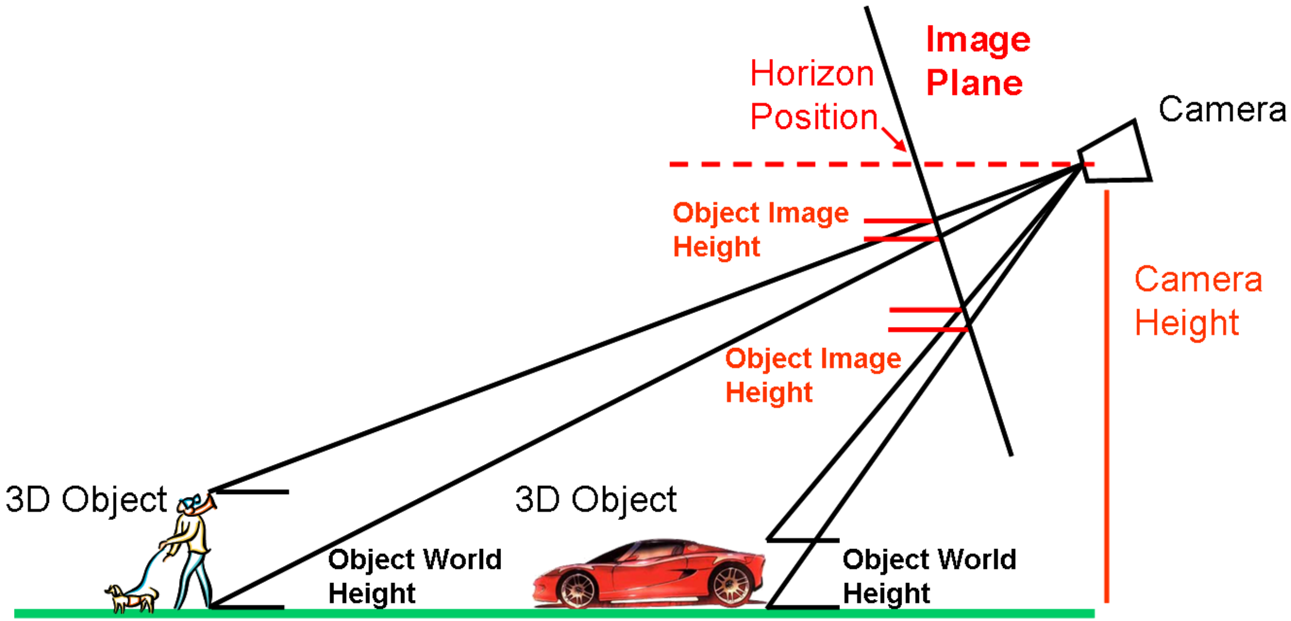
Stereo pair

Range map

Stereo vision

- Owing to the geometry of the situation
  - 3D scene points will project to different locations in a pair of spatially displaced optical sensors
  - From this difference in location we recover the 3D information.





Source: Pictures by Derek Hoeim

The two problems of stereo

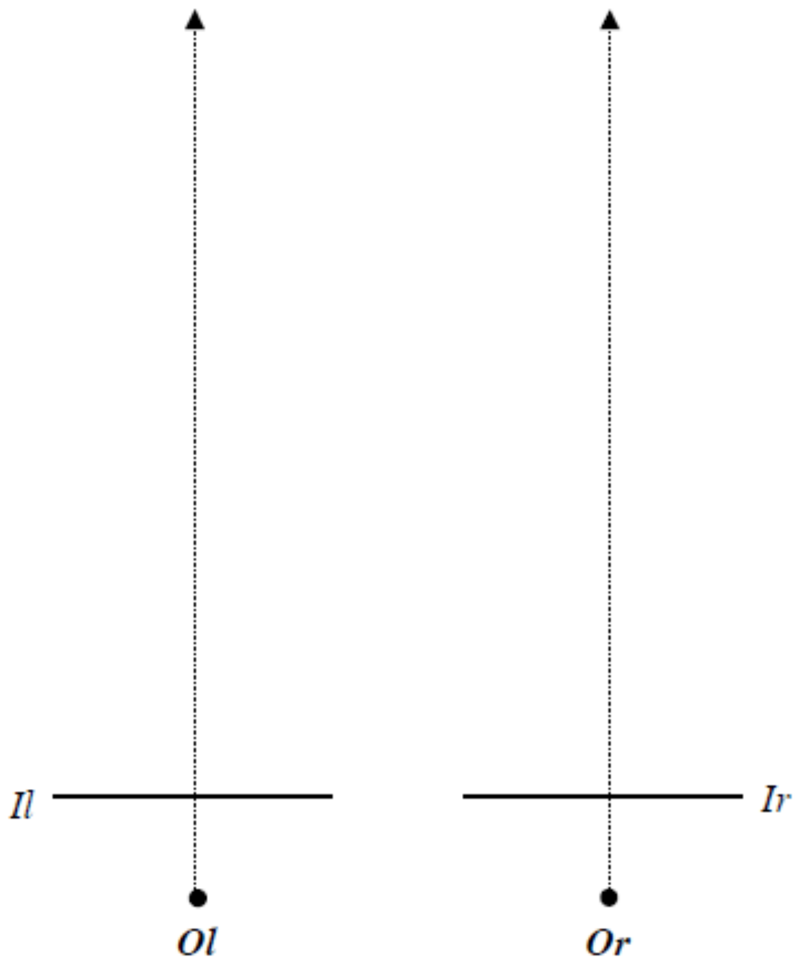
1. Correspondence
   - Which parts of the left and right images are projections of the same element in the 3D scene.
   - Which image parts should not be matched as they are not visible in the other image.
   - We require an analysis and algorithm to establish correspondences between all points that are visible in both images.
     - E.g. RANSAC

The two problems of stereo

2. Reconstruction
   - Let the difference in position of matched elements between the two views be called disparity.
   - The disparities of all the image points form the disparity map.
   - If the geometry of the stereo system is known (intrinsic and extrinsic camera parameters), then the disparity map can be converted to a 3D map of the imaged scene.
   - We require an analysis and algorithm that allows us to reconstruct the 3D scene from the matched binocular elements.

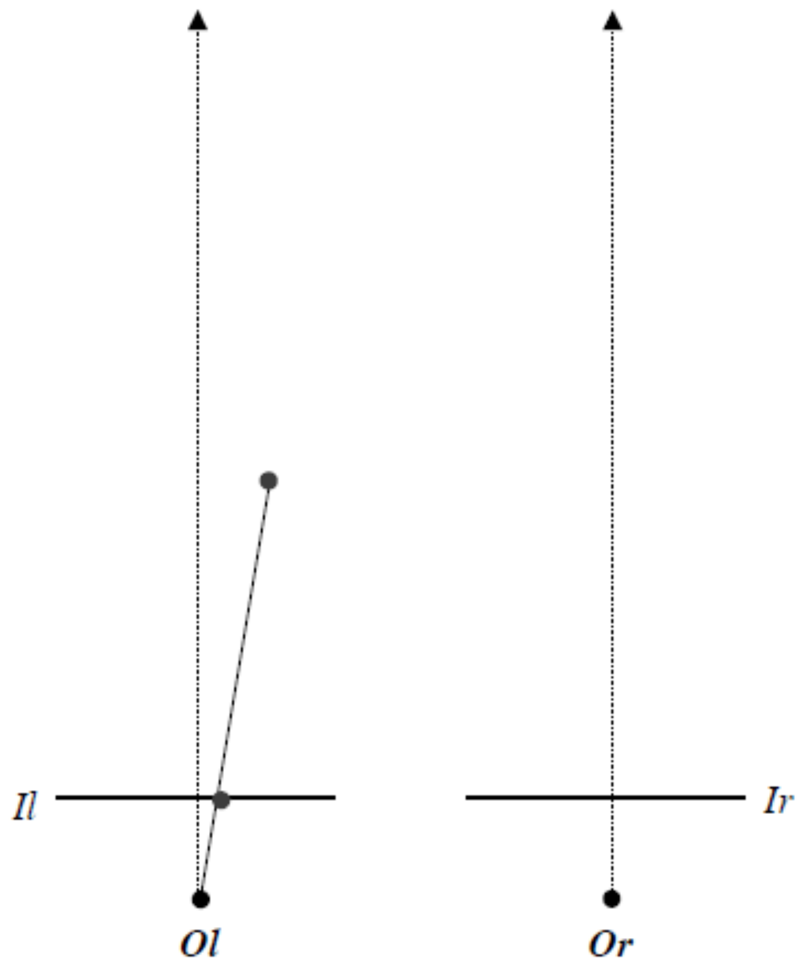
A simple stereo system

- Geometric model
  - Consider the top-down view of two pinhole cameras
  - The left and right images are coplanar, let
    - ***Il*** and ***Ir*** be the left and right images, respectively.
    - ***Ol*** and ***Or*** be the left and right centres of projection, respectively.
  - Take the optical axes as parallel
    - The fixation point, the intersection of the two optical axes, is at infinity



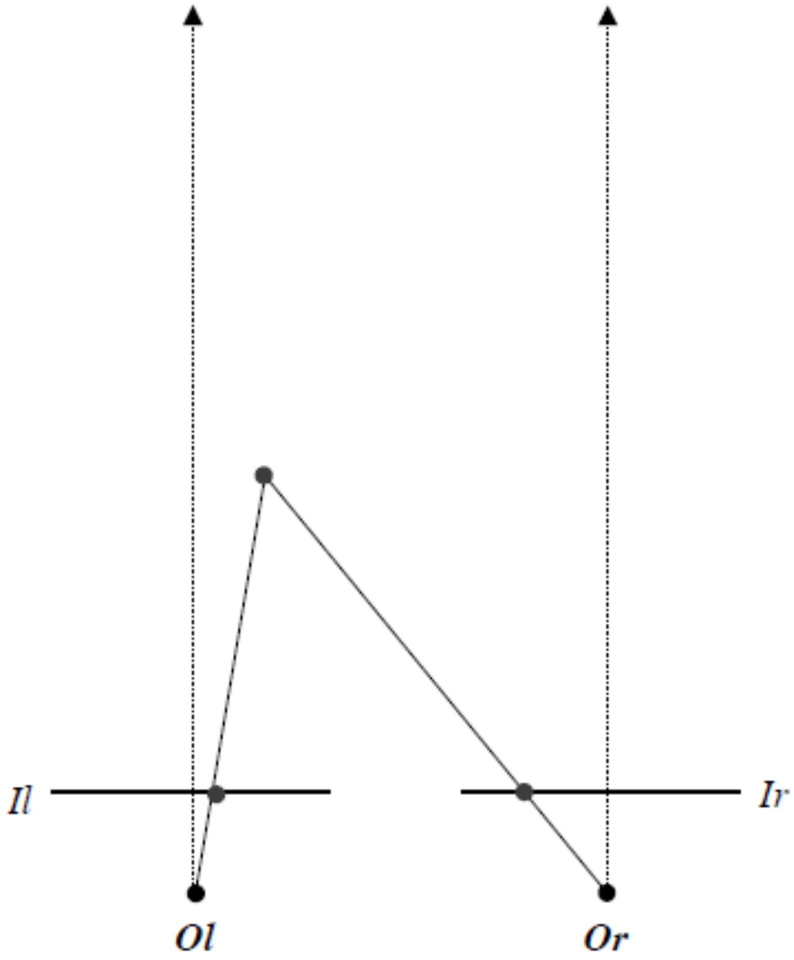
A simple stereo system

- Geometric model
  - **Recovery of position in space**
    - Position in space is determined via the intersection of rays
      - Defined by the centres of projection
      - And the left and right images of a point of concern.



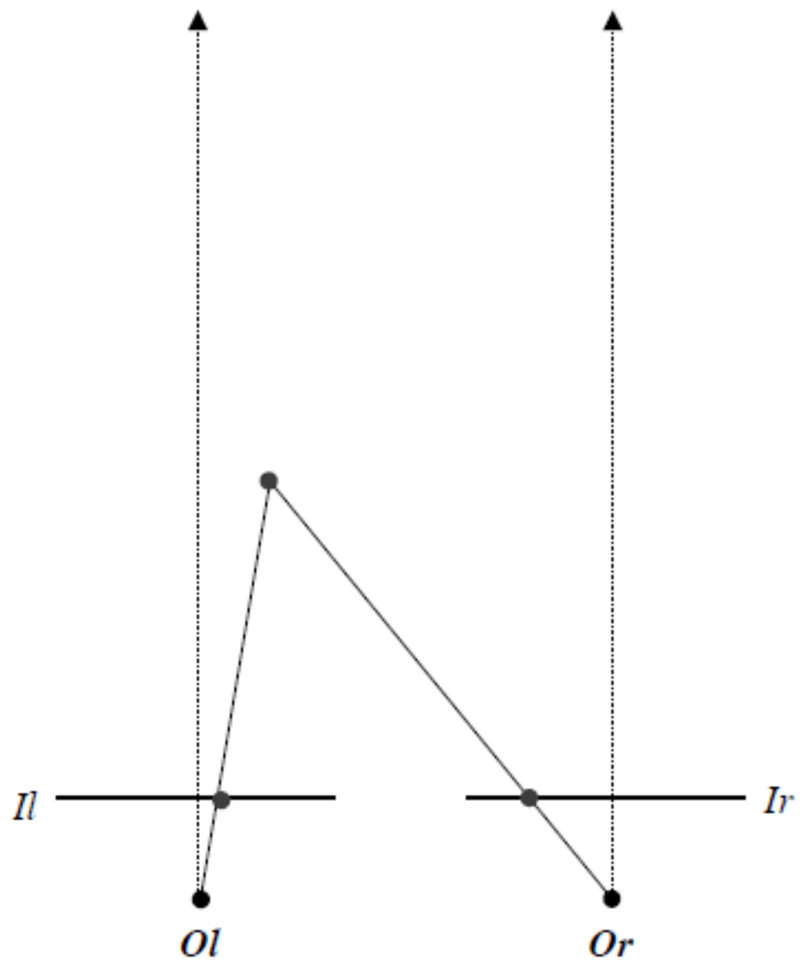
A simple stereo system

- Geometric model
  - **Recovery of position in space**
    - Position in space is determined via the intersection of rays
      - Defined by the centres of projection
      - And the left and right images of a point of concern.



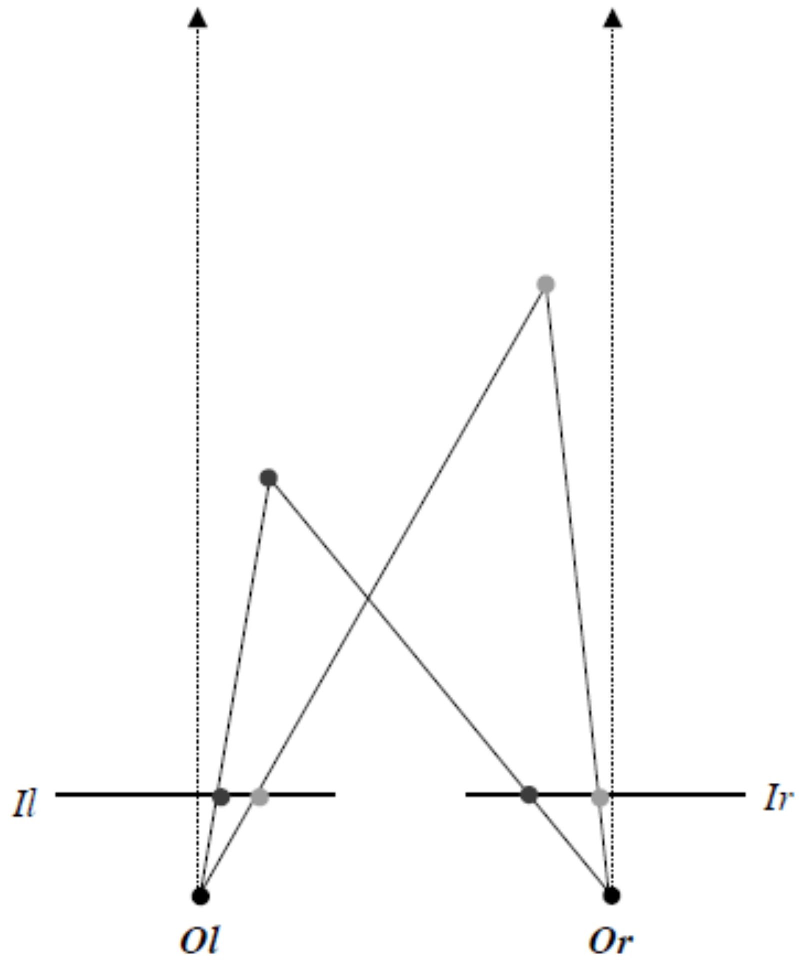
A simple stereo system

- Geometric model
  - **Recovery of position in space**
    - Position in space is determined via the intersection of rays
      - Defined by the centres of projection
      - And the left and right images of a point of concern.
    - A process known as triangulation



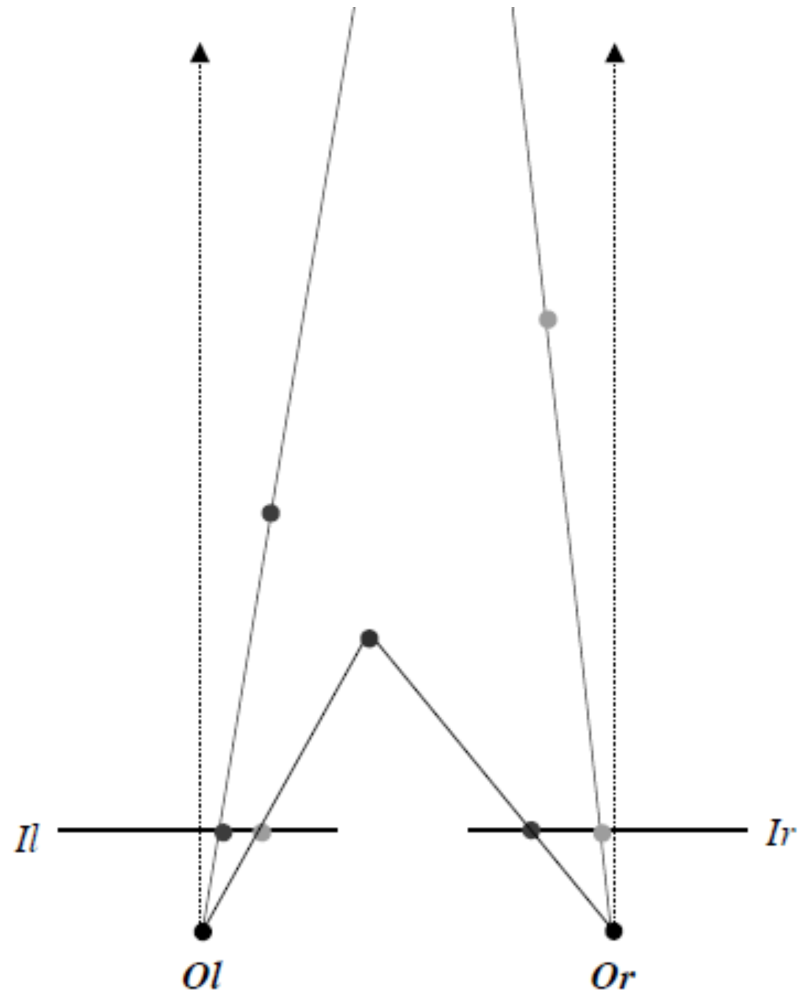
A simple stereo system

- Geometric model
  - **Recovery of position in space**
    - Position in space is determined via the intersection of rays
      - Defined by the centres of projection
      - And the left and right images of a point of concern.
    - A process known as triangulation



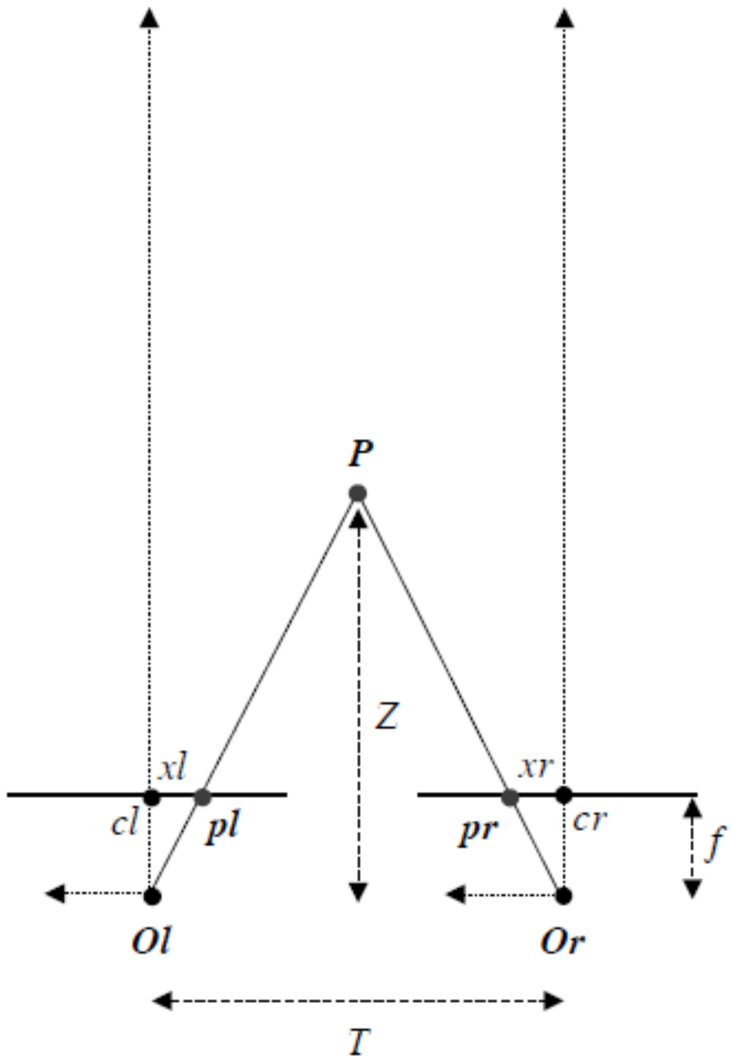
A simple stereo system

- Geometric model
  - **Recovery of position in space**
    - Position in space is determined via the intersection of rays
      - Defined by the centres of projection
      - And the left and right images of a point of concern.
    - A process known as triangulation
    - Triangulation depends critically on **correspondence**



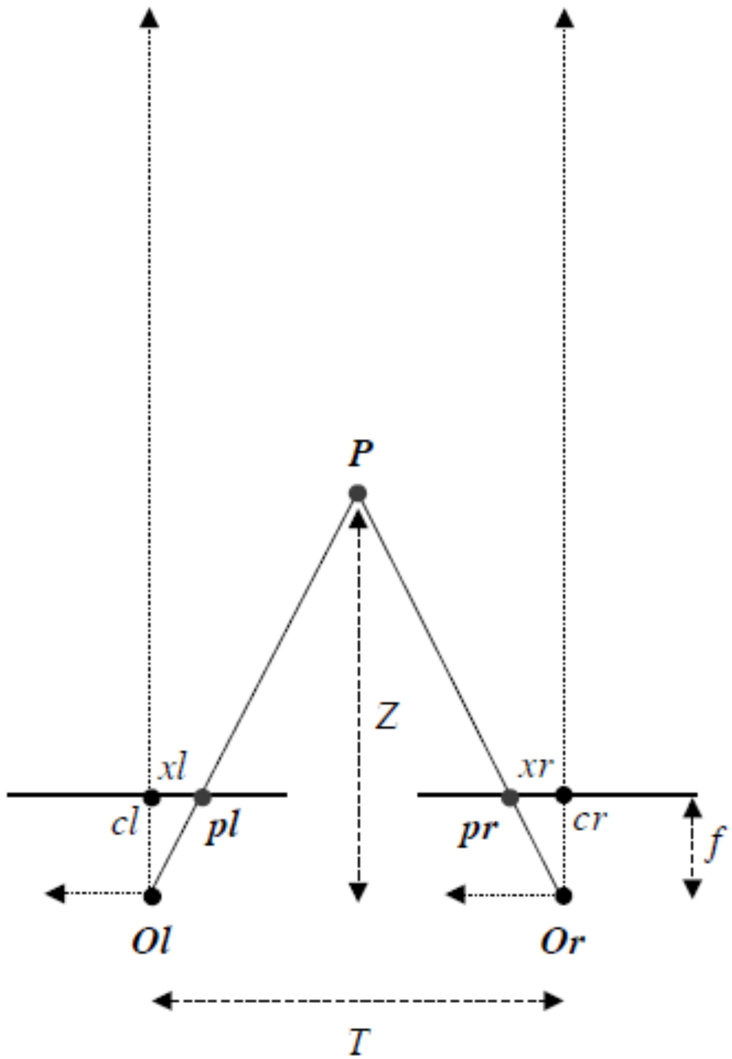
A simple stereo system

- Equations of triangulation
  - Consider a point ***P*** and its projections ***pl*** and ***pr***.



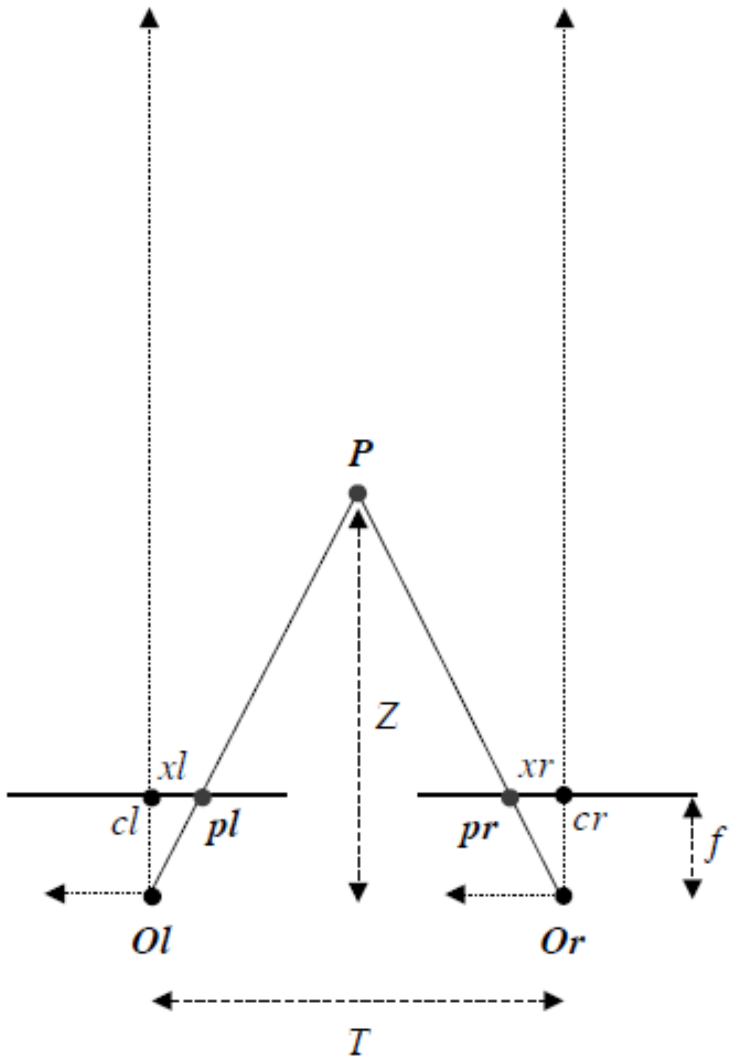
A simple stereo system

- Equations of triangulation
  - Consider a point *P* and its projections *pl* and *pr*.
  - *T* be the distance between the centres of projection, the baseline.



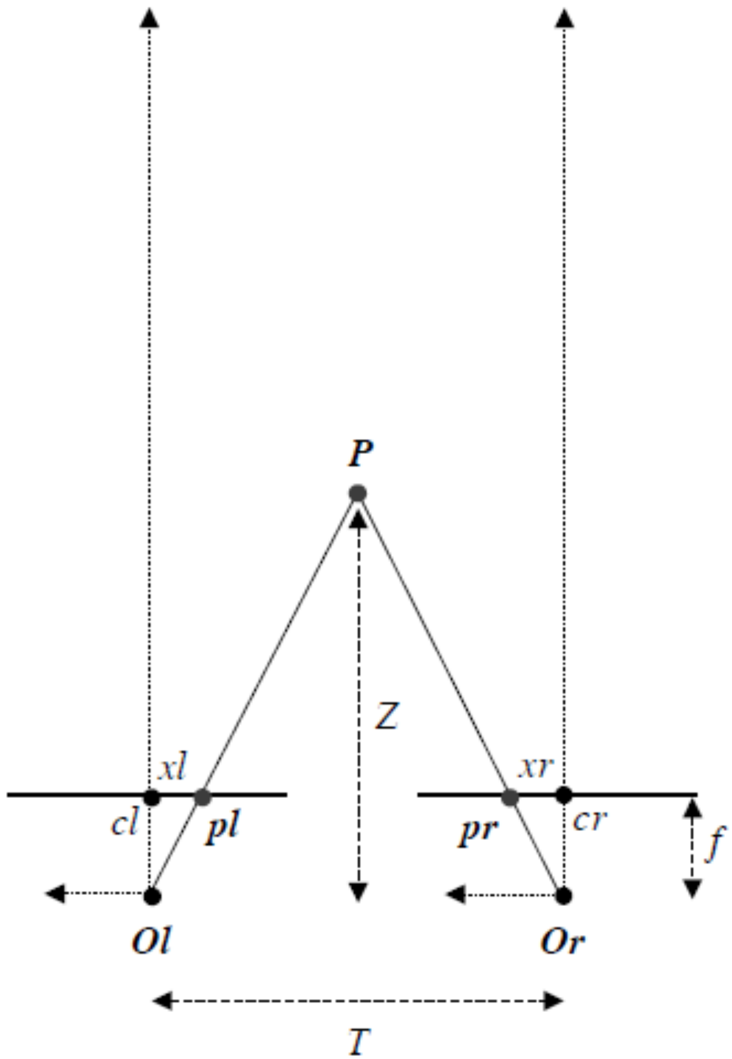
A simple stereo system

- Equations of triangulation
  - Consider a point ***P*** and its projections ***pl*** and ***pr***.
  - ***T*** be the distance between the centres of projection, the baseline.
  - ***cl*** and ***cr*** be the centre points of the left and right images, respectively



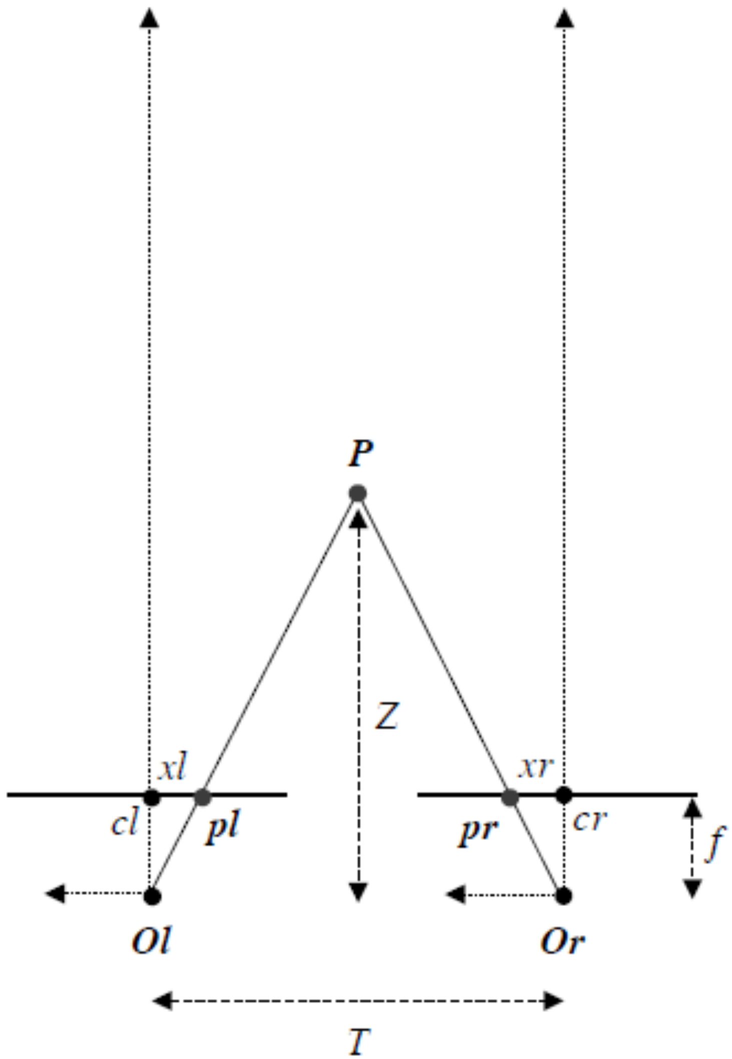
A simple stereo system

- Equations of triangulation
  - Consider a point ***P*** and its projections ***pl*** and ***pr***.
  - ***T*** be the distance between the centres of projection, the baseline.
  - ***cl*** and ***cr*** be the centre points of the left and right images, respectively
  - ***xl*** and ***xr*** be the coordinates of ***pl*** and ***pr***, respectively.



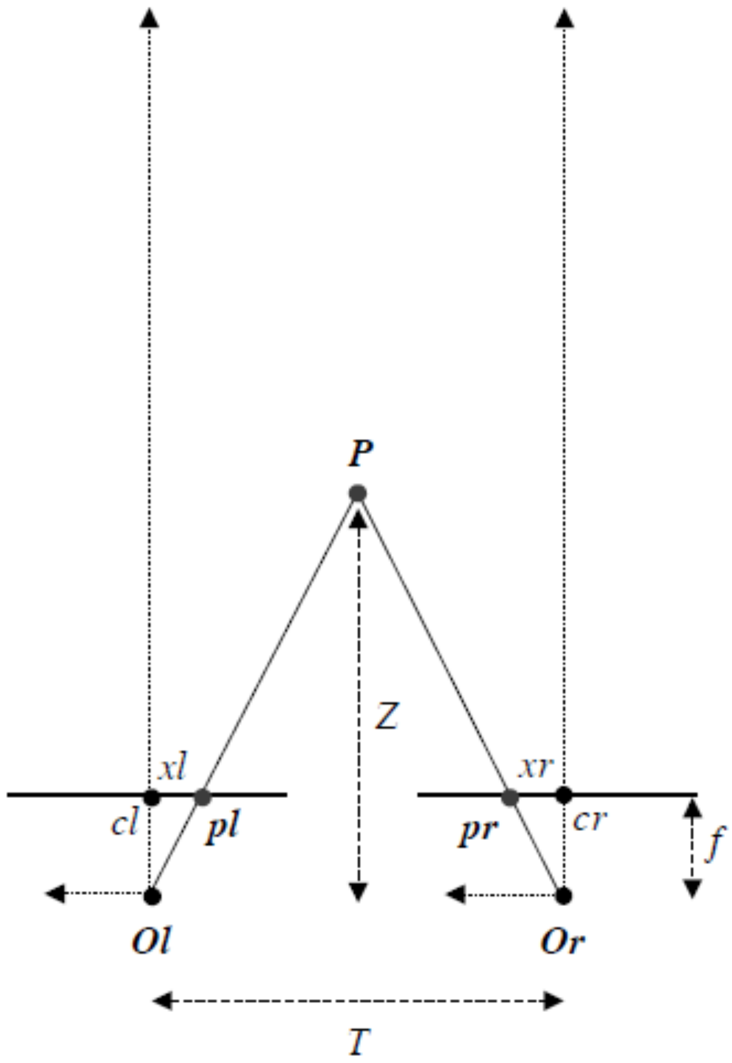
A simple stereo system

- Equations of triangulation
  - Consider a point *P* and its projections *pl* and *pr*.
  - *T* be the distance between the centres of projection, the baseline.
  - *cl* and *cr* be the centre points of the left and right images, respectively
  - *xl* and *xr* be the coordinates of *pl* and *pr*, respectively.
  - *f* be the common focal length of the two cameras



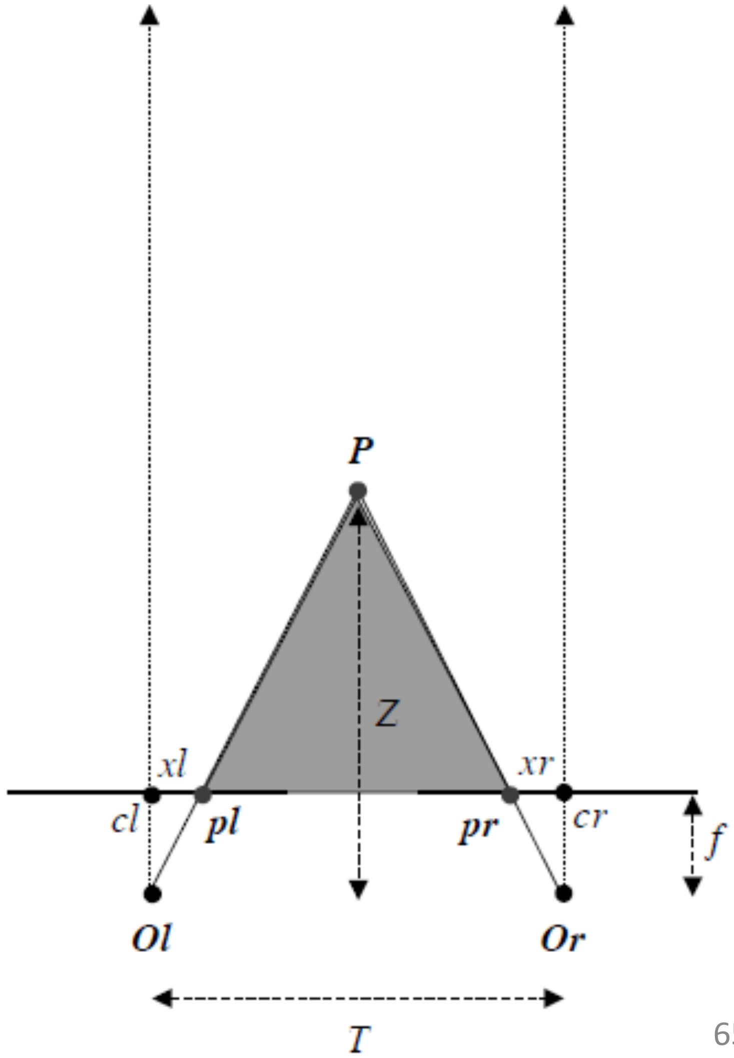
A simple stereo system

- Equations of triangulation
  - Consider a point *P* and its projections *pl* and *pr*.
  - *T* be the distance between the centres of projection, the baseline.
  - *cl* and *cr* be the centre points of the left and right images, respectively
  - *xl* and *xr* be the coordinates of *pl* and *pr*, respectively.
  - *f* be the common focal length of the two cameras
  - *Z* be the distance of *P* from the baseline



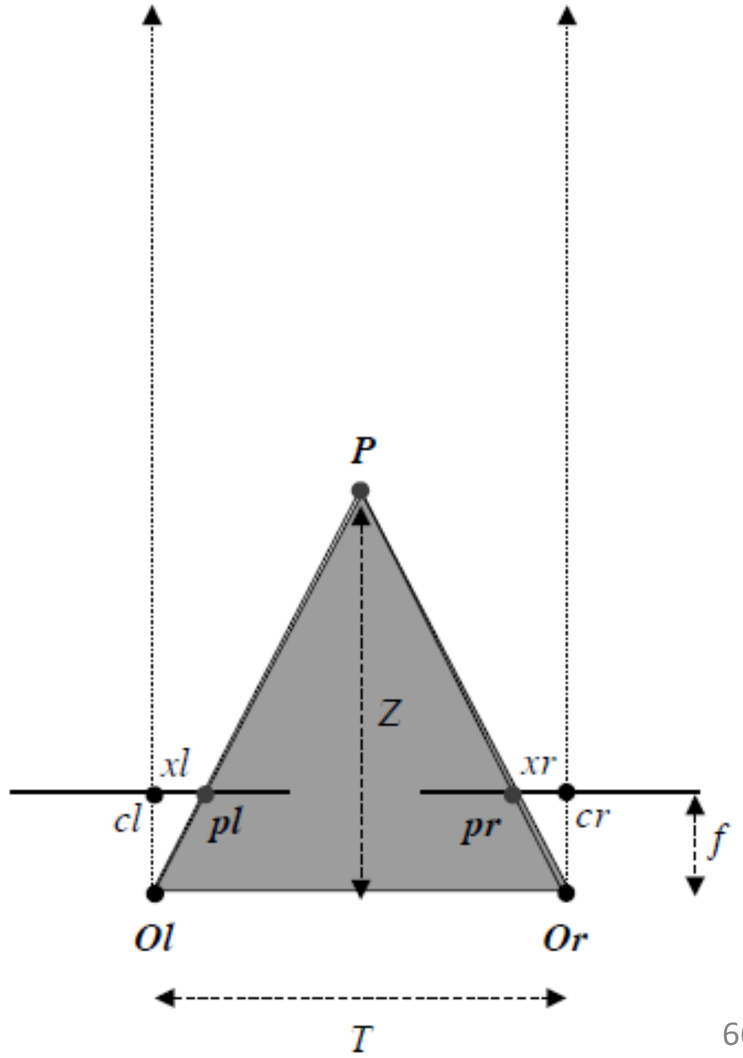
A simple stereo system

- Equations of triangulation
  - Similar triangles ***(pl,P,pr)*** and ***(Ol,P, Or)***



A simple stereo system

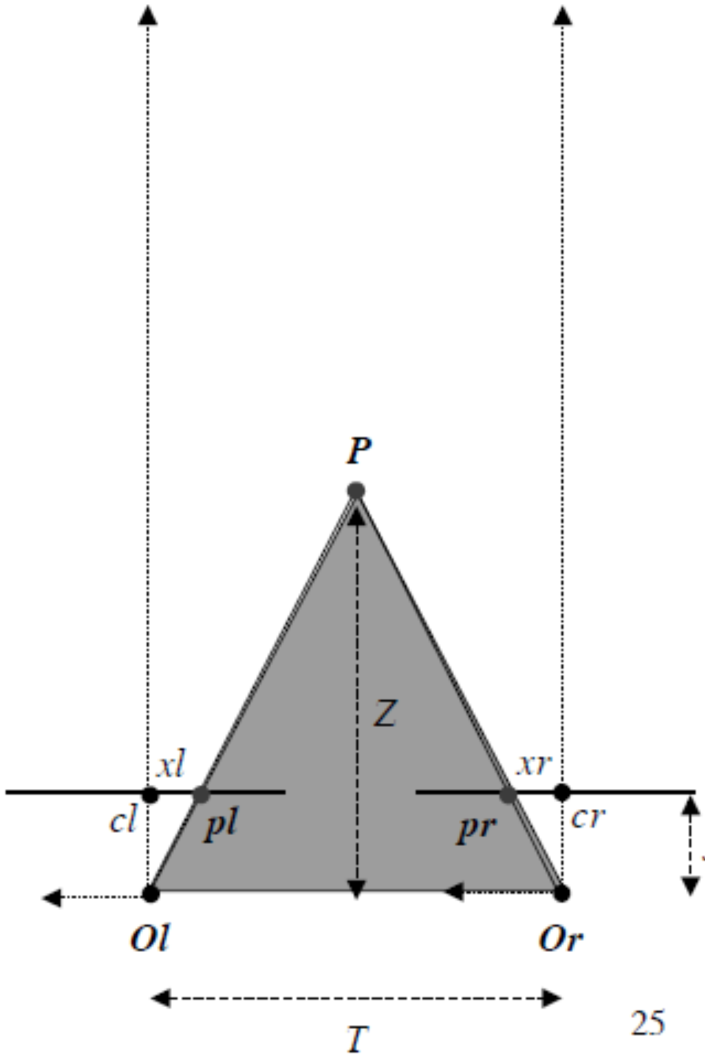
- Equations of triangulation
  - Similar triangles ***(pl,P,pr)*** and ***(Ol,P, Or)***



A simple stereo system

- Equations of triangulation
  - Similar triangles ***(pl,P,pr)*** and ***(Ol,P, Or)***
  - We can write

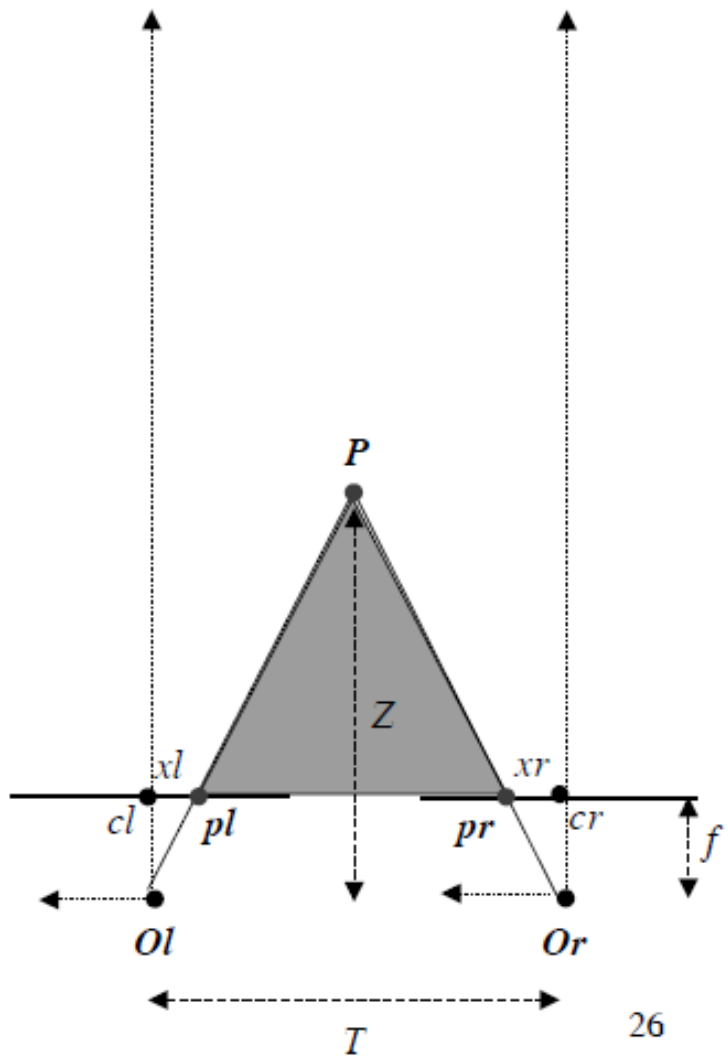
$$\frac{T + xl - xr}{Z - f} = \boxed{\frac{T}{Z}}$$



A simple stereo system

- Equations of triangulation
  - Similar triangles ***(pl,P,pr)*** and ***(Ol,P, Or)***
  - We can write

$$\boxed{\frac{T + xl - xr}{Z - f}} = \frac{T}{Z}$$



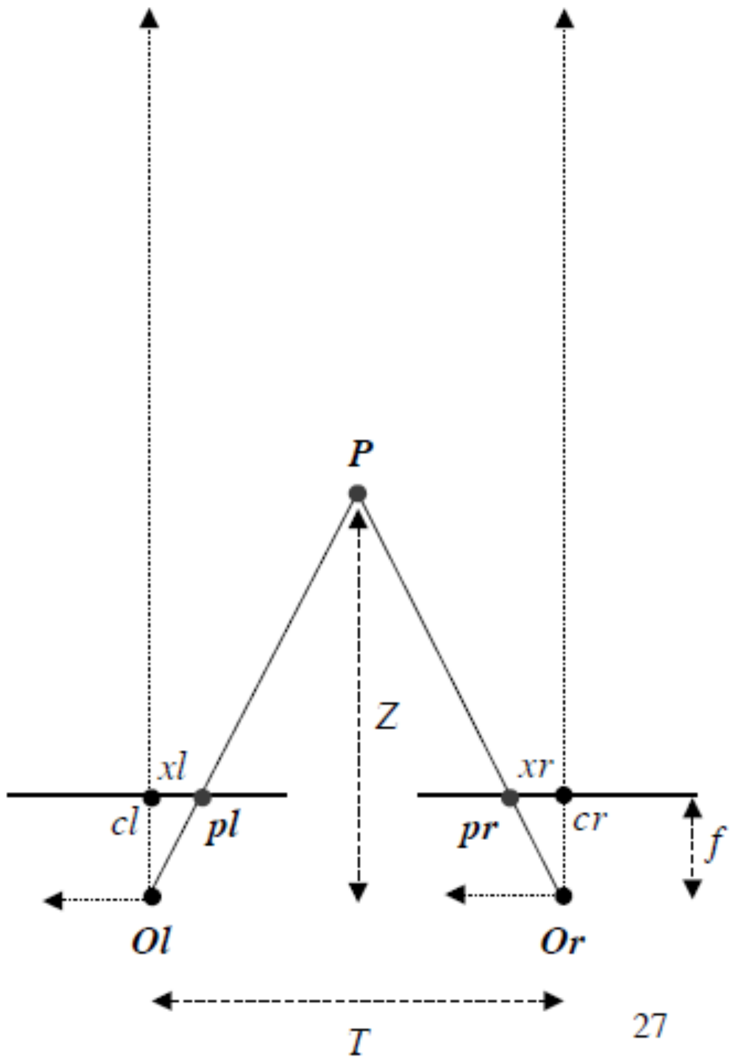
A simple stereo system

- Equations of triangulation
  - Similar triangles *(pl,P,pr)* and *(Ol,P, Or)*
  - We can write

$$\frac{T + xl - xr}{Z - f} = \frac{T}{Z}$$

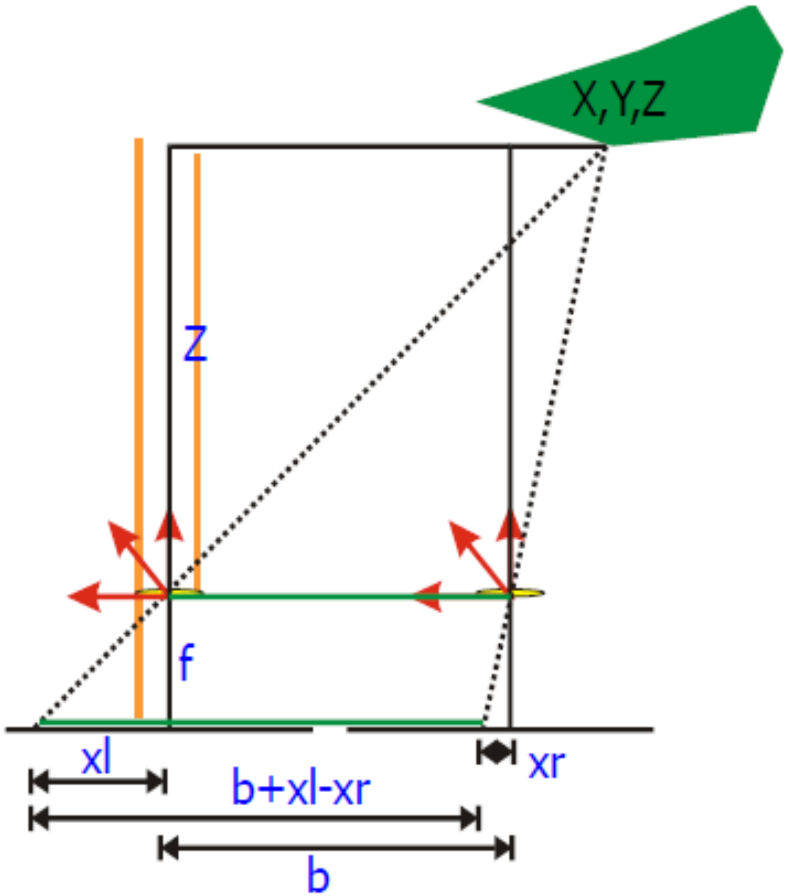
  - Letting $d = xr - xl$ be the **disparity**, we solve for Z as

$$Z = f\frac{T}{d}$$



A simple stereo system

A slightly different diagram to achieve the same result



Anwendung der Strahlensätze
(wobei x_l-x_r= Disparität d):

$$\boxed{\frac{Z}{Z+f}} = \boxed{\frac{b}{b+x_l-x_r}}$$

$$\Leftrightarrow Z(b+x_l-x_r) = b(Z+f)$$

$$\Leftrightarrow Z(x_l-x_r) = bf$$

$$\Leftrightarrow Z = \frac{bf}{x_l-x_r} = \frac{bf}{d}$$

Source: University of Magdeburg

Parameters of a stereo system

- Intrinsic parameters
  - For our simple model we have ***f***, ***cl*** and ***cr***.
  - More generally, all of the intrinsic parameters of the two camera systems are of interest.

Parameters of a stereo system

- Intrinsic parameters
  - For our simple model we have f, cl and cr.
  - More generally, all of the intrinsic parameters of the two camera systems are of interest.
- Extrinsic parameters
  - For our simple model we have T.
  - More generally, all of the extrinsic parameters (translation and rotation) that relate the two camera systems are of interest.

Parameters of a stereo system

- Remarks
  - To perform Euclidean reconstruction all of these parameters must be known.
    - A need for accurate calibration.
  - Interesting information can be recovered with only partial (or no calibration).
  - In the parallel optical axis model, disparity can only decrease with distance to objects.
    - More generally, disparity decreases with distance from the fixation point, the convergence of the optical axes.

Outline

- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

Correspondence

- Assumptions
  - Most scene points are visible from both viewpoints
  - Corresponding image regions are similar in appearance
  - Reasonably true for stereo systems where fixation distance >> baseline.
  - …but false in general

Correspondence

- Assumptions
  - Most scene points are visible from both viewpoints
  - Corresponding image regions are similar in appearance
  - Reasonably true for stereo systems where fixation distance >> baseline.
  - …but false in general
- Cast correspondence as search
  - Which image elements are to be matched?
  - What similarity measure to adopt?
  - Postpone issue that not all points have correspondences.

Correspondence

- Assumptions
  - Most scene points are visible from both viewpoints
  - Corresponding image regions are similar in appearance
  - Reasonably true for stereo systems where fixation distance >> baseline.
  - …but false in general
- Cast correspondence as search
  - Which image elements are to be matched?
  - What similarity measure to adopt?
  - Postpone issue that not all points have correspondences.
- Consider two classes of correspondence method
  - Area-based
  - Feature-based

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

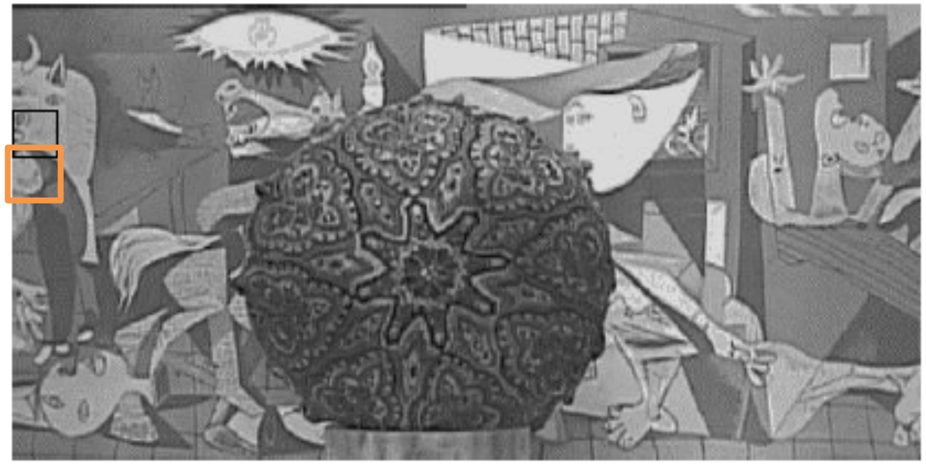
- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

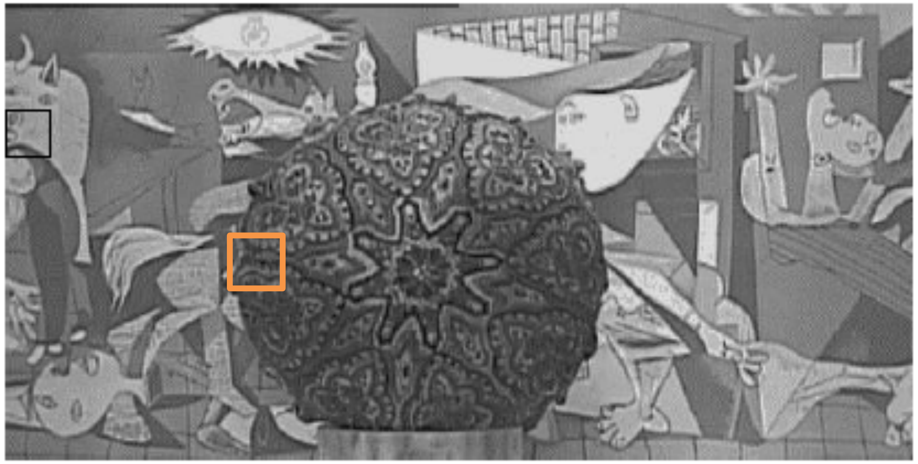


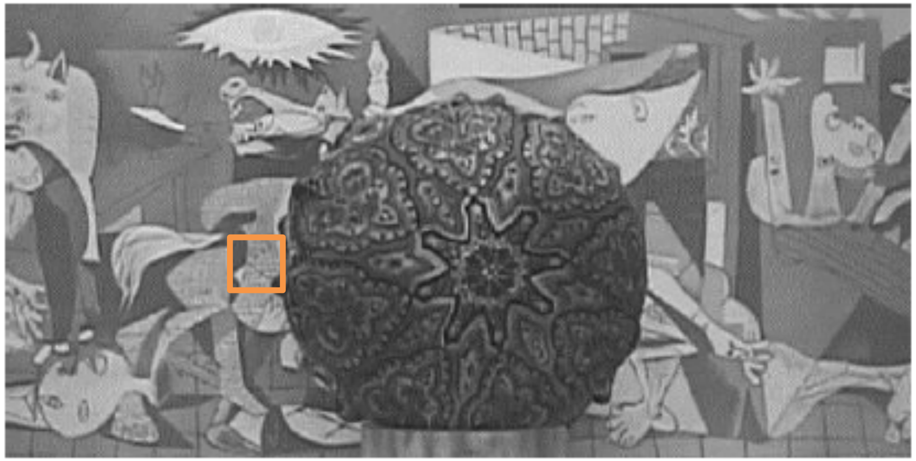
Correspondence

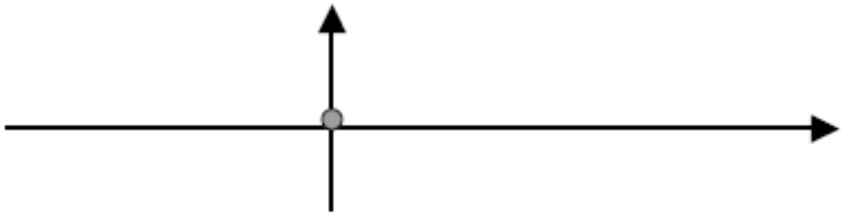
- Area-based
  - Exploit all available information
  - Elements to be matched
    - Image windows
    - Typically of fixed size
    - Spatially overlapping

Correspondence

- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.

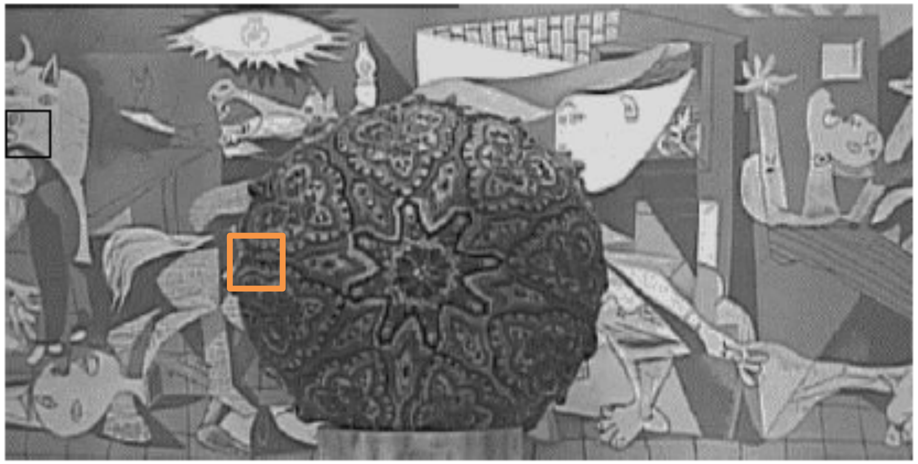


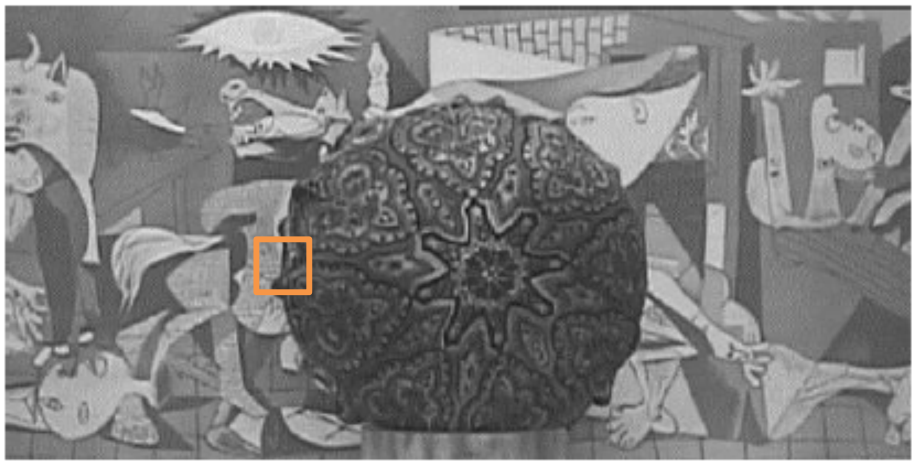


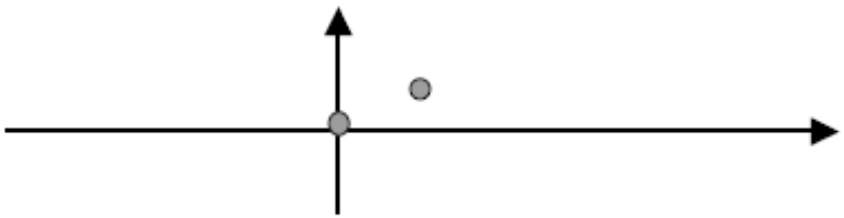


Correspondence

- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.

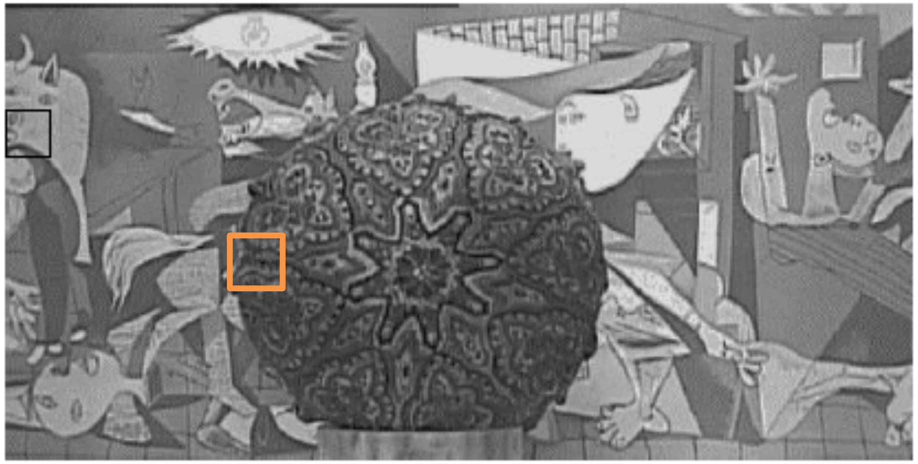


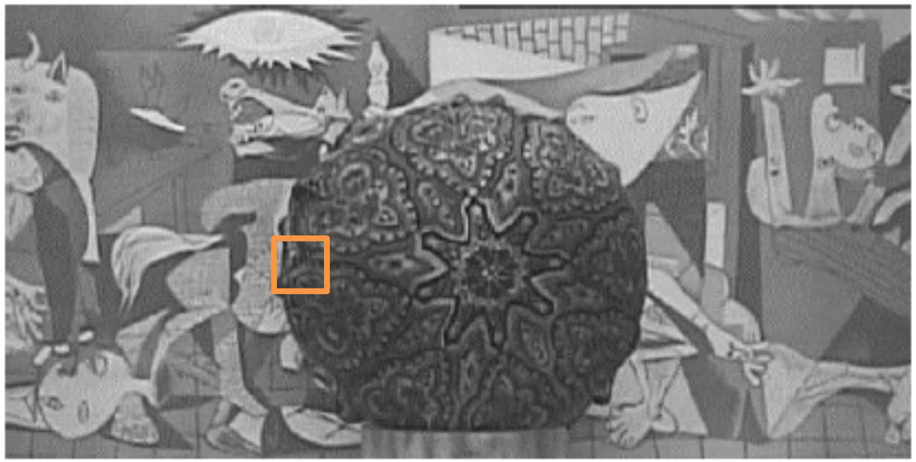


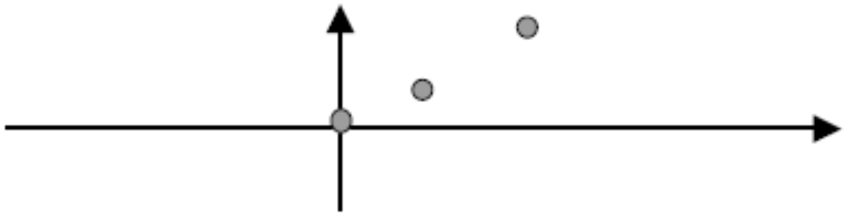


Correspondence

- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.

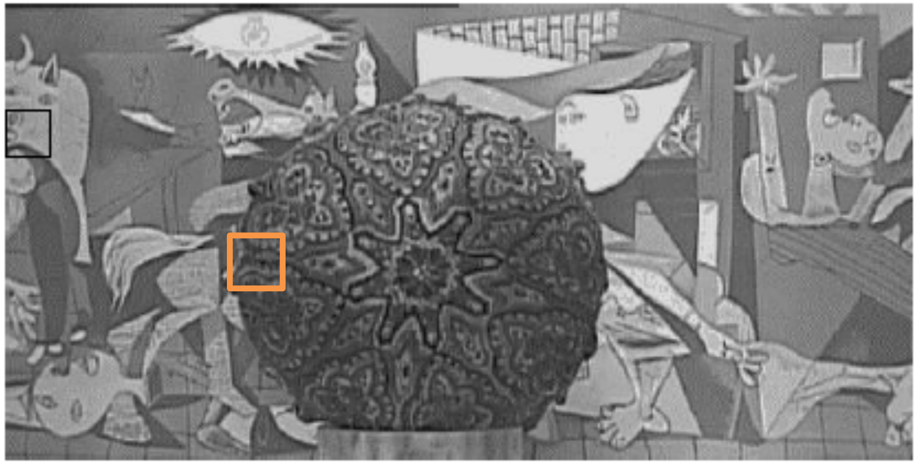


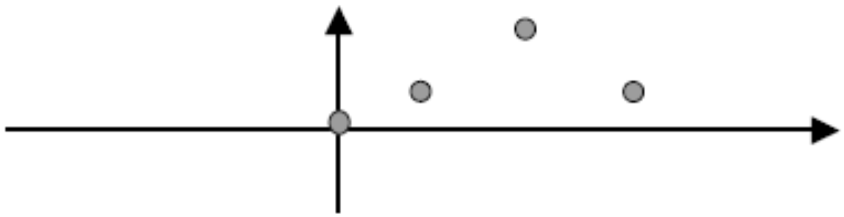




Correspondence

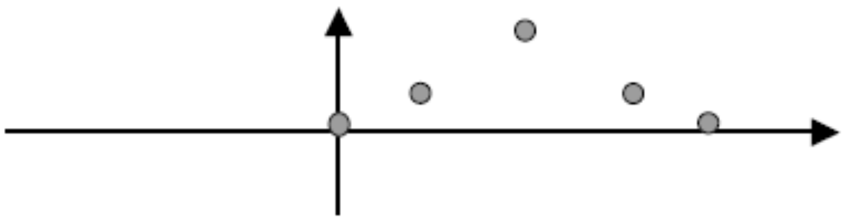
- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.

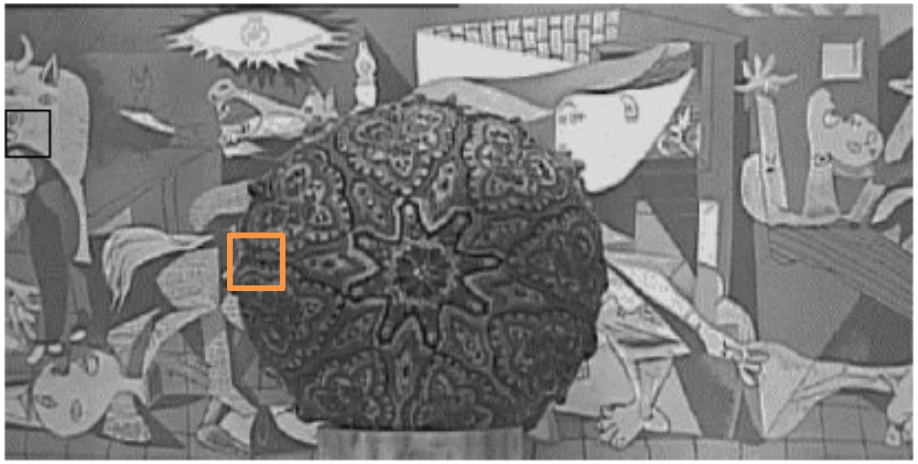




Correspondence

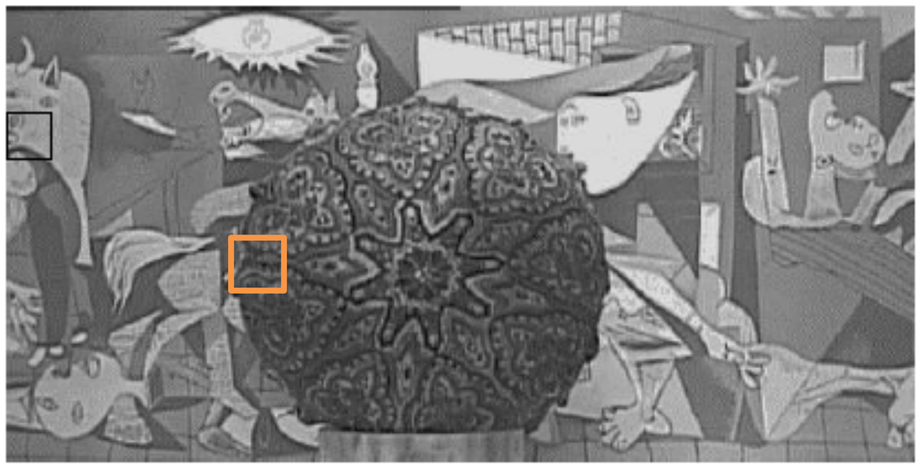
- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.

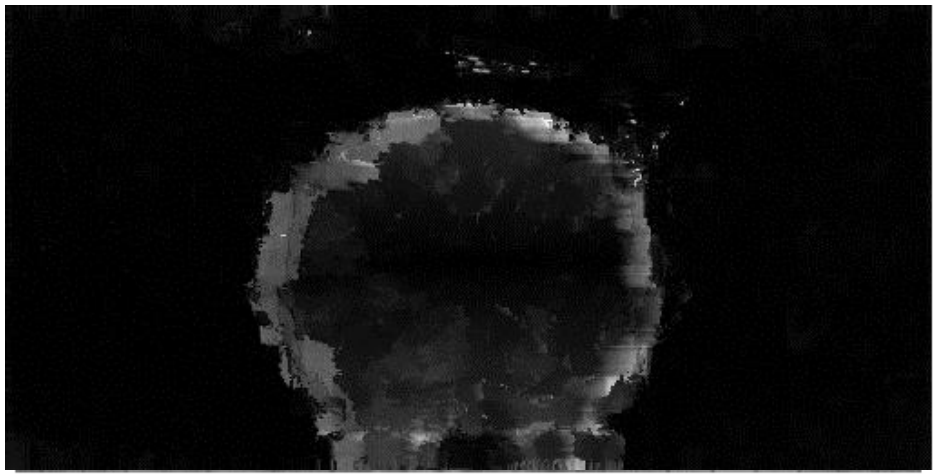


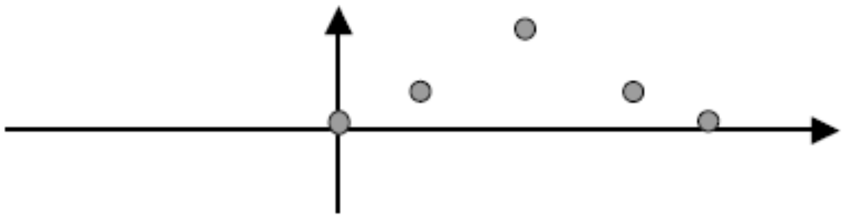


Correspondence

- Area-based
  - Similarity measure
    - An integrated pixel difference over windows in the two images.
    - Corresponding element is that which maximizes the measure over some search region.







Correspondence

Algorithm

- **Input:** Stereo pair of images Il and Ir
- **Output:** Array of disparities (disparity map), one value for each pixel of Il.
- **Notation:** Let
  - $\boldsymbol{pl}$ and $\boldsymbol{pr}$ be pixels in the left and right images, resp.
  - $2W+1$ the width (in pixels) of the match window,
  - $R(\boldsymbol{pl})$ the search range in the right image associated with $\boldsymbol{pl}$
  - $m(u,v)$ a function of two pixel values, u and v.
- For each pixel $\boldsymbol{pl}=(i,j)$ in the left image

1. For each displacement $\boldsymbol{d} = (d1,d2)$ of $R(\boldsymbol{pl})$ calculate

$$c(\boldsymbol{d}) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} m[Il(i+k,\, j+l), Ir(i+k-d1,\, j+l-d2)]$$

2. The disparity of $\boldsymbol{pl}$ is the vector $\boldsymbol{d}$ that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Correspondence

Algorithm

- **Input:** Stereo pair of images Il and Ir
- **Output:** Array of disparities (disparity map), one value for each pixel of Il.
- **Notation:** Let
  - $\boldsymbol{pl}$ and $\boldsymbol{pr}$ be pixels in the left and right images, resp.
  - $2W+1$ the width (in pixels) of the match window,
  - $R(\boldsymbol{pl})$ the search range in the right image associated with $\boldsymbol{pl}$
  - $m(u,v)$ a function of two pixel values, u and v.
- For each pixel $\boldsymbol{pl}=(i,j)$ in the left image

1. For each displacement $\boldsymbol{d} = (d1,d2)$ of $R(\boldsymbol{pl})$ calculate

$$c(\boldsymbol{d}) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} m[Il(i+k,\, j+l),\, Ir(i+k-d1,\, j+l-d2)]$$

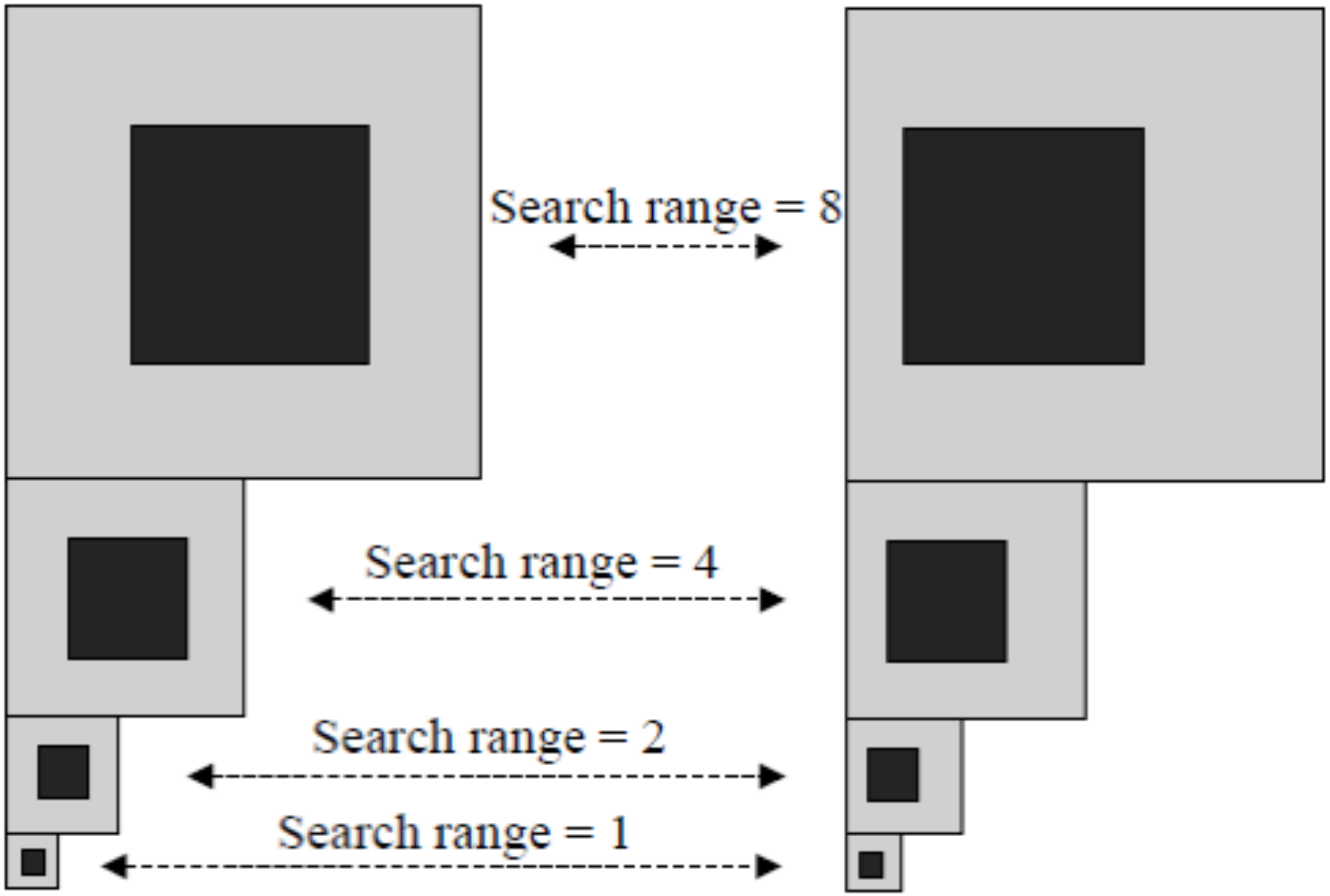
2. The disparity of $\boldsymbol{pl}$ is the vector $\boldsymbol{d}$ that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Remarks

- Search region can be centred about (0,0).
- Size of window depends on knowledge Re. Spatial scale of stable image features.
- Oriented-bandpass image representation can systematically expose image structure for matching.
- Coarse-to-fine (pyramid) refinement can support large search ranges with modest expense.

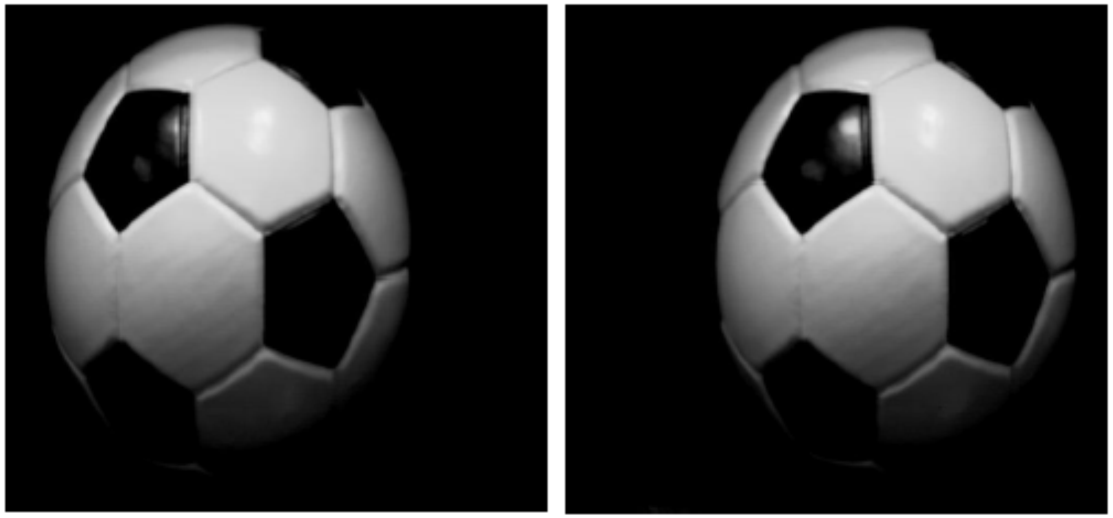
Correspondence

- Coarse to fine: Benefits



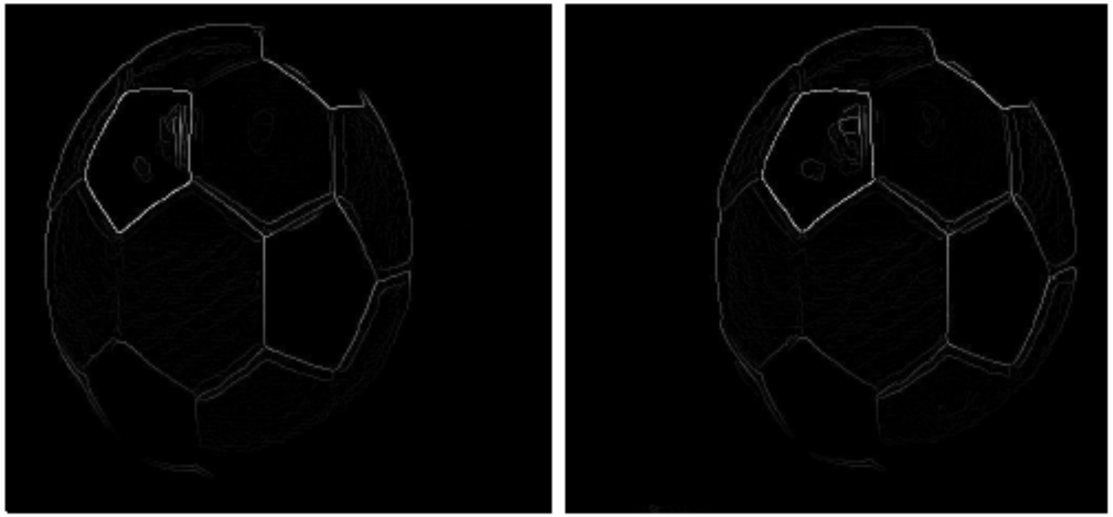
Correspondence

- Feature-based
  - Not all features are created equal.
- Elements to be matched
  - Sparse set of extracted features.
    - Edges
    - Corners…
  - Numerical and/or symbolic descriptors
    - Feature length
    - Feature orientation
    - Average contrast…



Correspondence

- Feature-based
  - Not all features are created equal.
- Similarity measure
  - Sparse set of extracted features.
  - Distance between feature descriptors.
  - Corresponding element is that which minimizes distance between feature distance.



Correspondence

Algorithm

- **Input:** Stereo pair of images Il and Ir and associated sets of feature descriptors
- **Output:** List of feature correspondences and (possibly sparse) disparity map.
- **Notation:** Let
  - fl and fr be left and right image feature descriptors, respectively
  - $R(fl)$ be the search range in the right image associated with left-image feature descriptor fl
  - $d(fl,fr)$ be the disparity between features fl and fr.
- For each fl pixel in the left image set

1. Compare the similarity measure between fl and each image feature in $R(fl)$.
2. Select the right-image feature that maximizes the similarity measure.
3. Save the correspondence and the disparity of fl

Correspondence

Algorithm

- **Input:** Stereo pair of images Il and Ir and associated sets of feature descriptors
- **Output:** List of feature correspondences and (possibly sparse) disparity map.
- **Notation:** Let
  - $\boldsymbol{fl}$ and $\boldsymbol{fr}$ be left and right image feature descriptors, respectively
  - $R(\boldsymbol{fl})$ be the search range in the right image associated with left-image feature descriptor $\boldsymbol{fl}$
  - $d(\boldsymbol{fl},\boldsymbol{fr})$ be the disparity between features $\boldsymbol{fl}$ and $\boldsymbol{fr}$.
- For each $\boldsymbol{fl}$ pixel in the left image set

1. Compare the similarity measure between $\boldsymbol{fl}$ and each image feature in $R(\boldsymbol{fl})$.
2. Select the right-image feature that maximizes the similarity measure.
3. Save the correspondence and the disparity of $\boldsymbol{fl}$

Representative similarity measure

- Inverse of weighted average of distances between feature descriptors
- For example, let
  - len_l and len_r be feature lengths in left and right images, resp.
  - con_l and con_l be feature contrast in left and right images, resp.
- Then the similarity measure would be

$$S = \frac{1}{w_l(len_l - len_r)^2 + w_c(con_l - con_r)^2}$$

Correspondence

Algorithm

- **Input:** Stereo pair of images Il and Ir and associated sets of feature descriptors
- **Output:** List of feature correspondences and (possibly sparse) disparity map.
- **Notation:** Let
  - fl and fr be left and right image feature descriptors, respectively
  - $R(fl)$ be the search range in the right image associated with left-image feature descriptor fl
  - $d(fl,fr)$ be the disparity between features fl and fr.
- For each fl pixel in the left image set

1. Compare the similarity measure between fl and each image feature in $R(fl)$.
2. Select the right-image feature that maximizes the similarity measure.
3. Save the correspondence and the disparity of fl

Remarks

- Starting search and search range can be set similarly to area-based methods, about (0,0)
- Coarse-to-fine processing can also be employed to good advantage
  - Initially extract features from coarse resolution imagery,
  - Perform matching
  - Increase resolution and repeat

Correspondence

- Final remarks
  - Area-based
    - Easier to implement
    - Provide dense disparity maps
    - Require reasonably textured images to drive local match measure
    - Sensitive to viewpoint and illumination changes between images

Correspondence

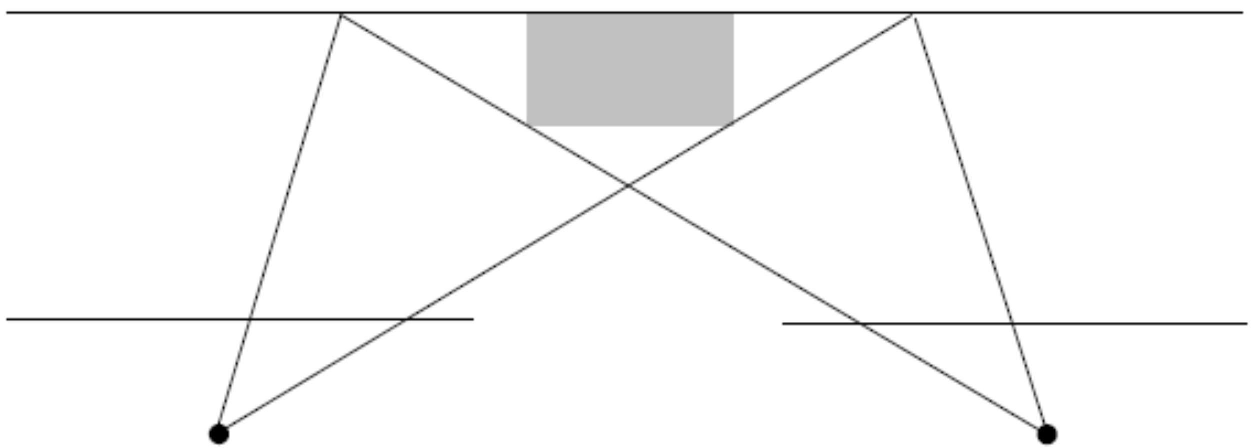
- Final remarks
  - Area-based
    - Easier to implement
    - Provide dense disparity maps
    - Require reasonably textured images to drive local match measure
    - Sensitive to viewpoint and illumination changes between images
  - Feature-based
    - Most suitable when a priori knowledge suggests appropriate feature sets
    - Although only sparse disparity is produced, can be suitable for many applications
    - Well chosen features can be more robust to viewpoint and illumination variations.

Correspondence

- Unmatchable points
  - Both methods can be stymied in attempting to match points that appear in only one of the two views.
    - Due to half occlusion
    - Due to noise

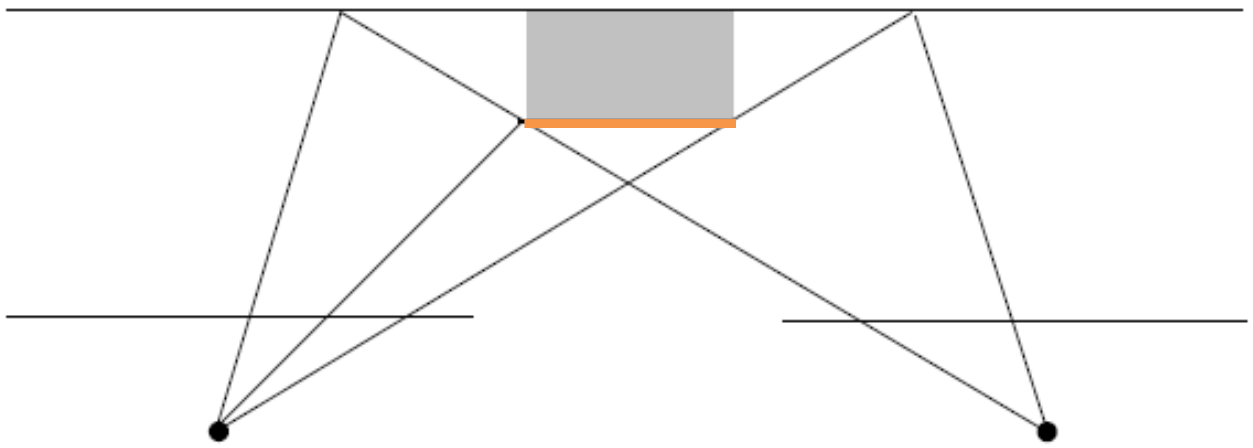
Correspondence

- Unmatchable points
  - Both methods can be stymied in attempting to match points that appear in only one of the two views.
    - Due to half occlusion
    - Due to noise



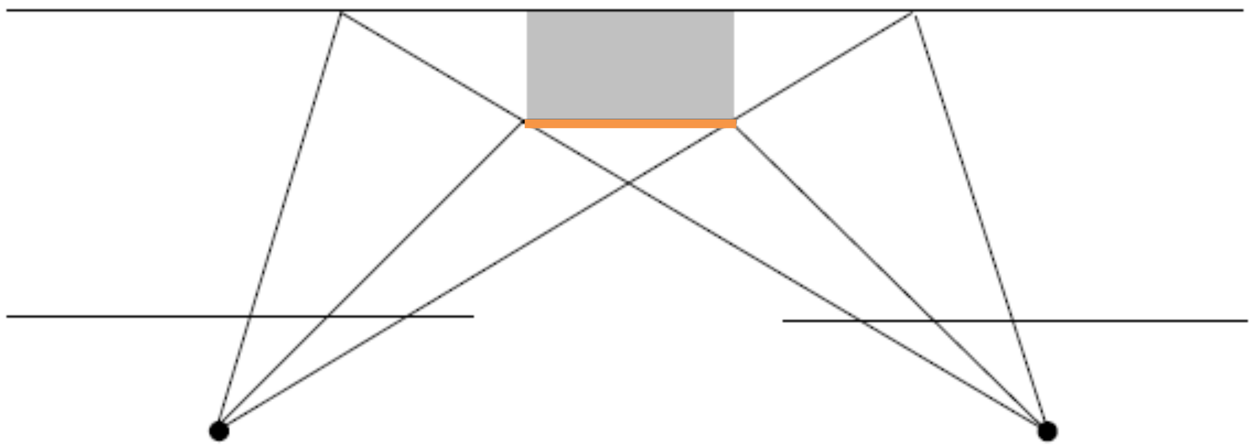
Correspondence

- Unmatchable points
  - Both methods can be stymied in attempting to match points that appear in only one of the two views.
    - Due to half occlusion
    - Due to noise



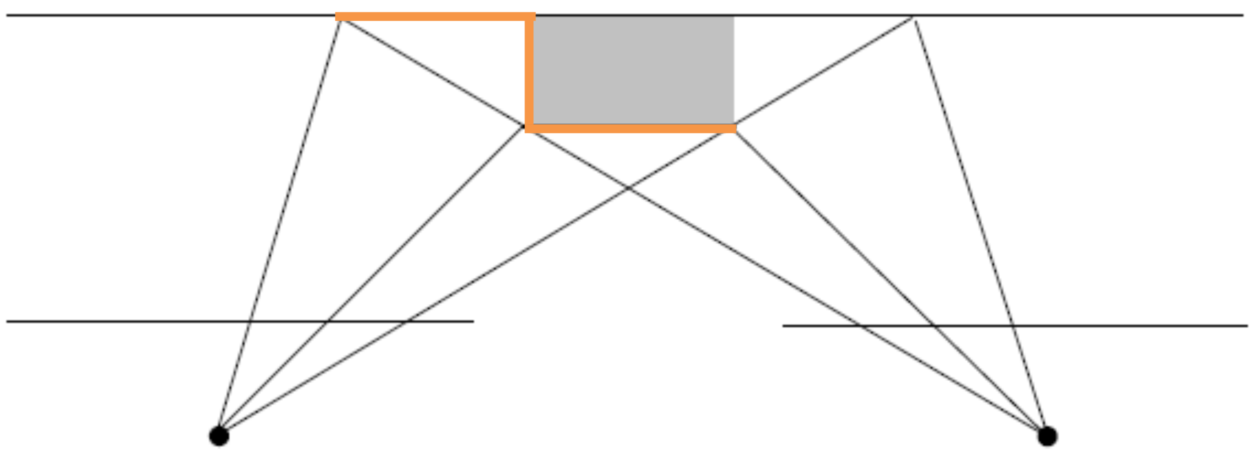
Correspondence

- Unmatchable points
  - Both methods can be stymied in attempting to match points that appear in only one of the two views.
    - Due to half occlusion
    - Due to noise



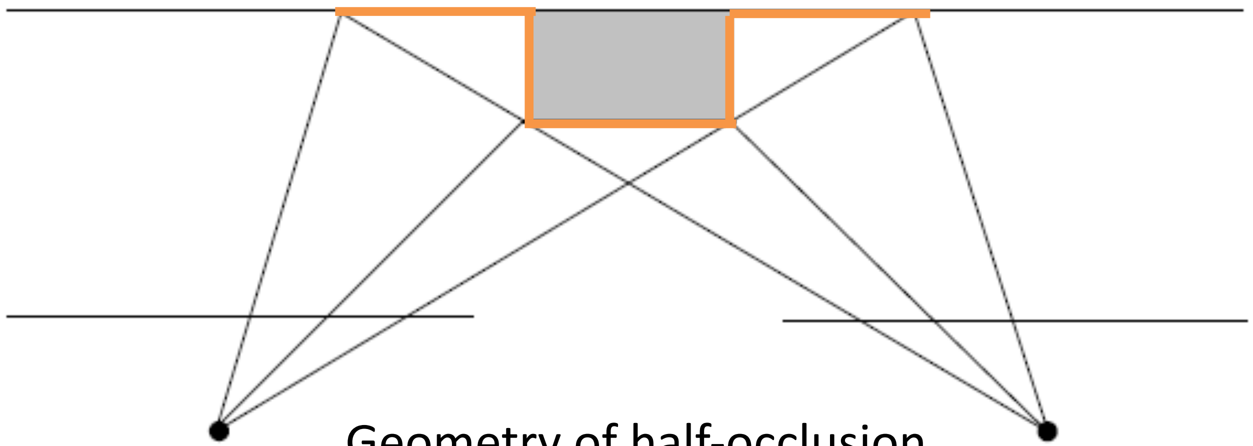
Correspondence

- Unmatchable points
  - Both methods can be stymied in attempting to match points that appear in only one of the two views.
    - Due to half occlusion
    - Due to noise



Correspondence

- Unmatchable points
  - Various techniques are available to help diagnose such situations
    - Left-right checking looks for consistent matches left-to-right and right-to-left
    - Epipolar constraint limits match region so it becomes less likely that false matches are encountered.

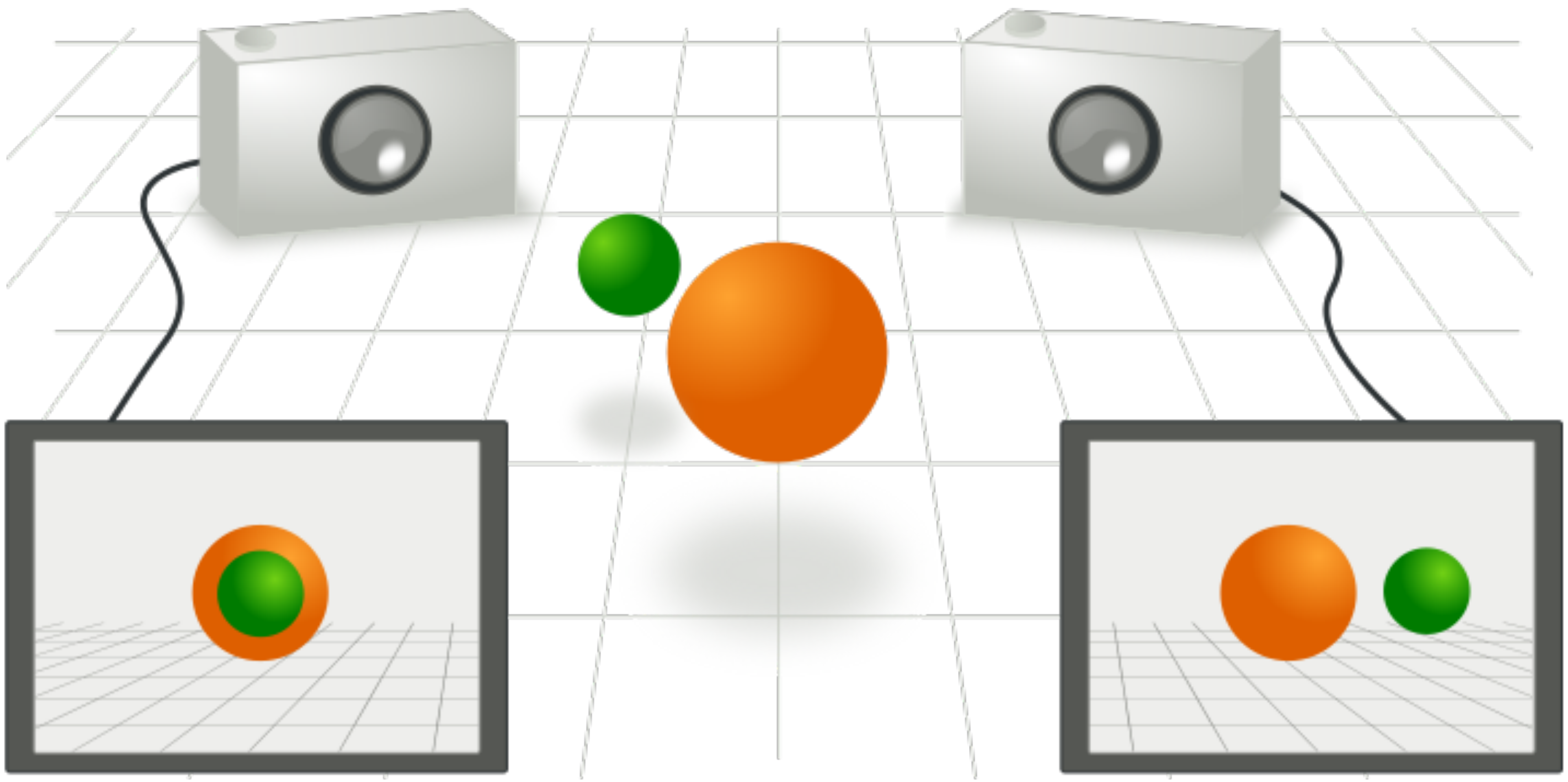


Geometry of half-occlusion

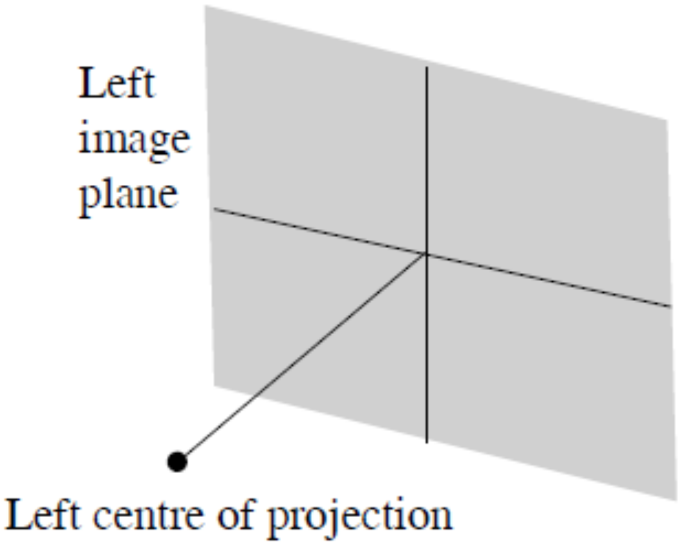
Outline

- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction

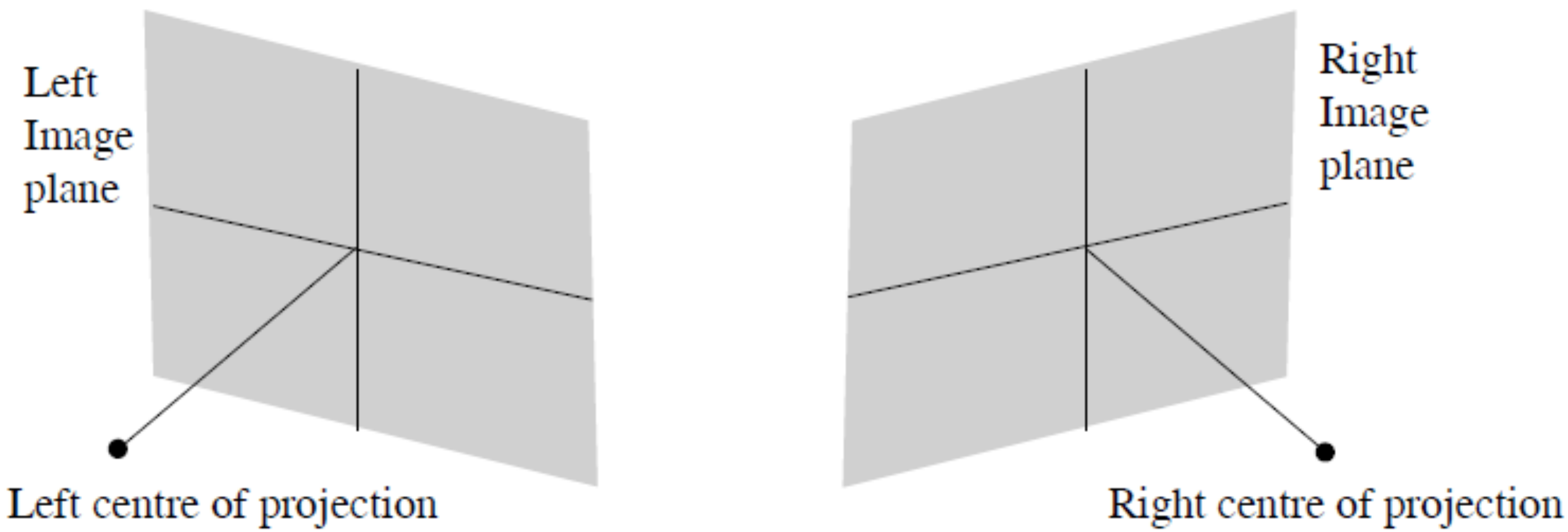
Epipolar geometry



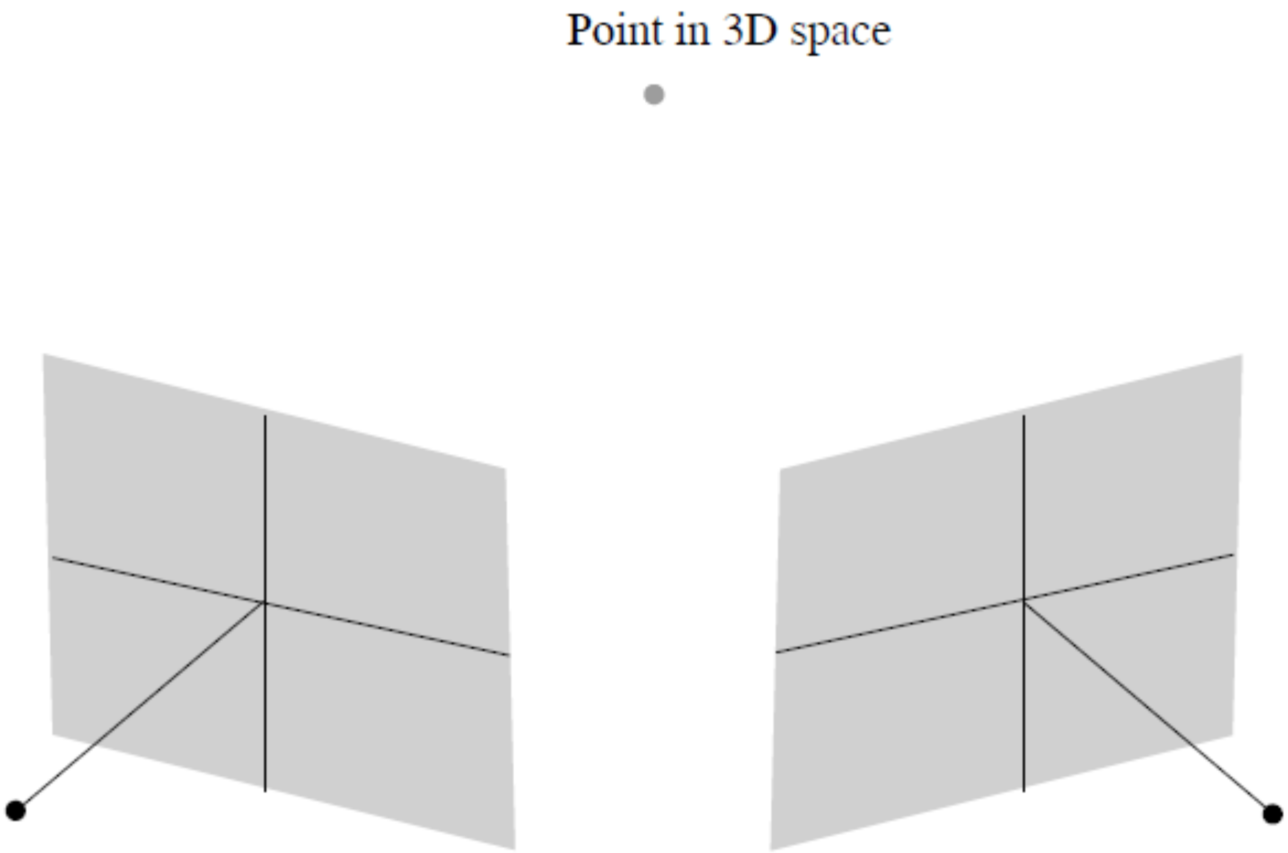
Epipolar geometry



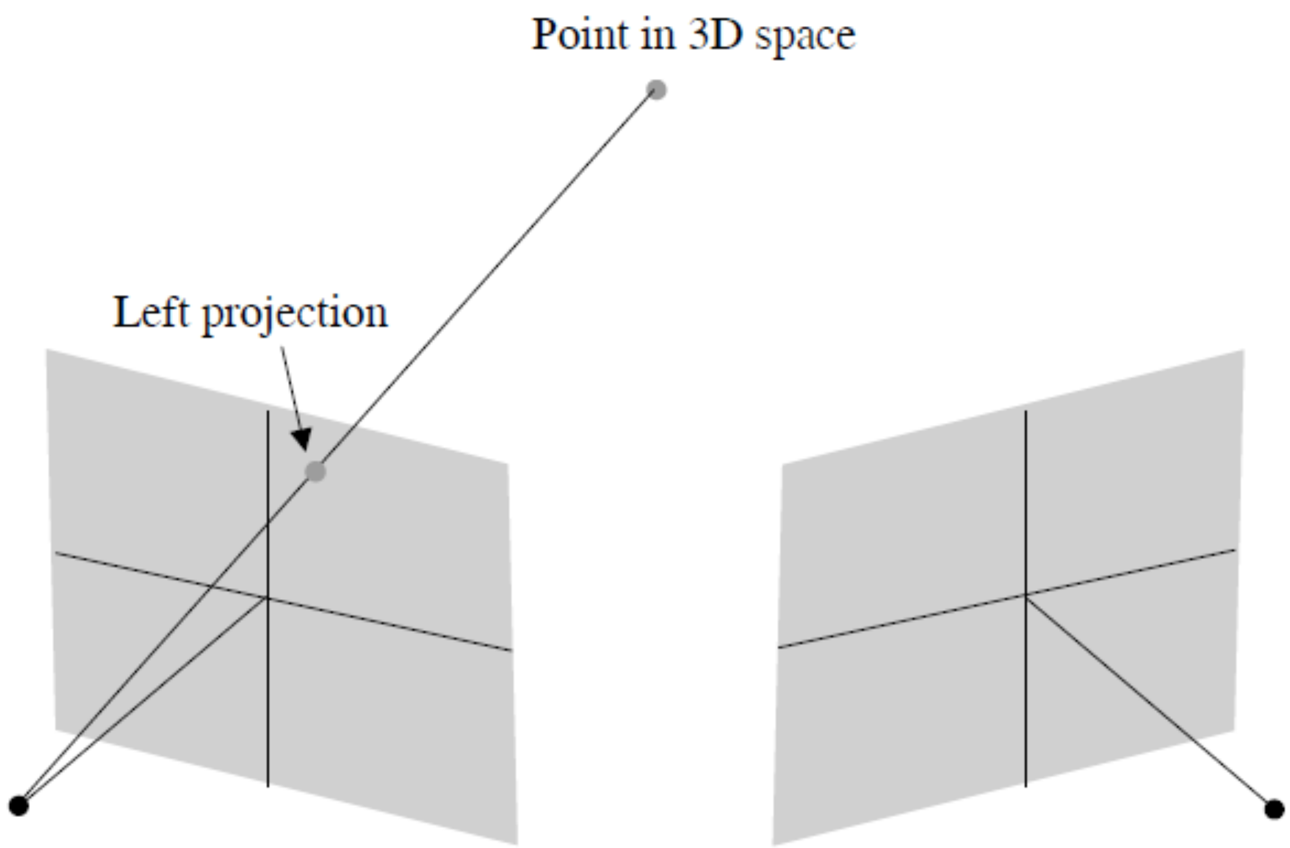
Epipolar geometry



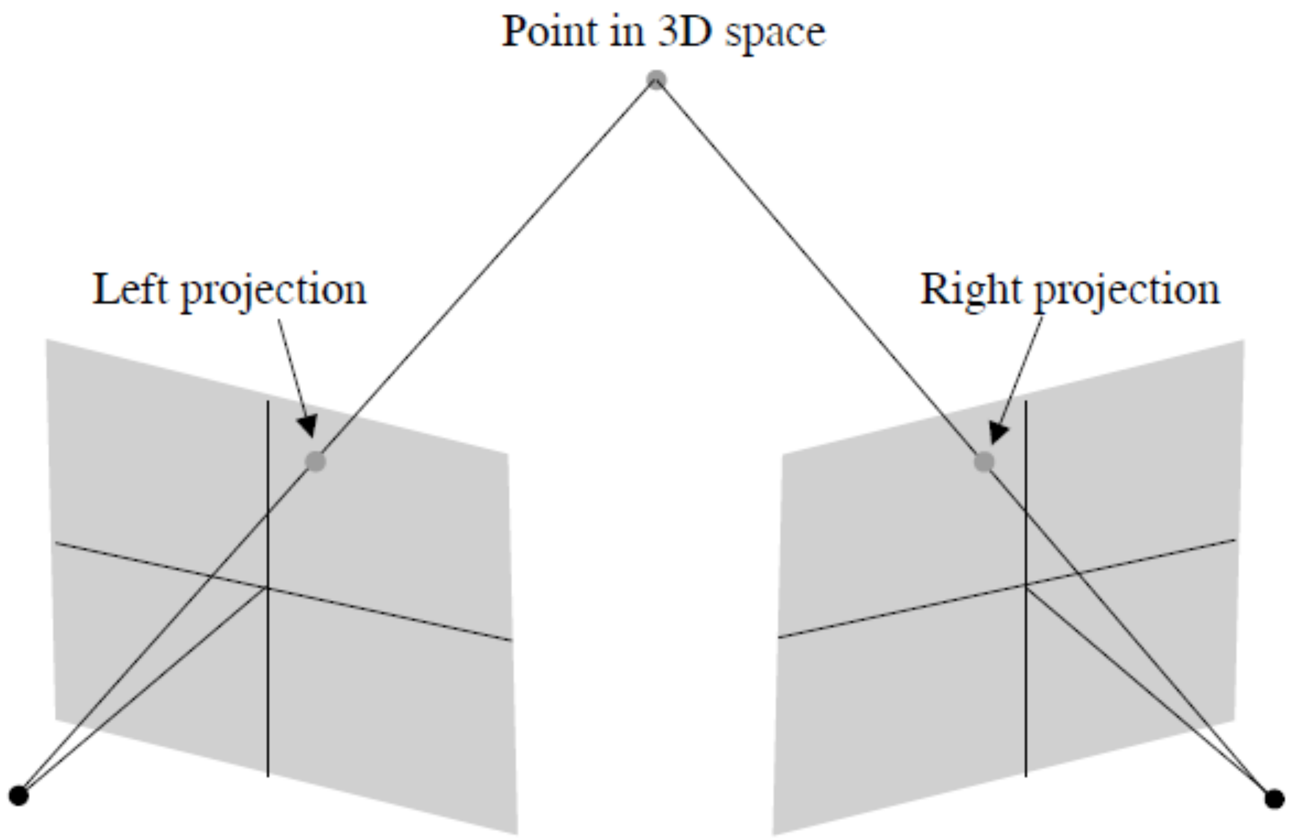
Epipolar geometry



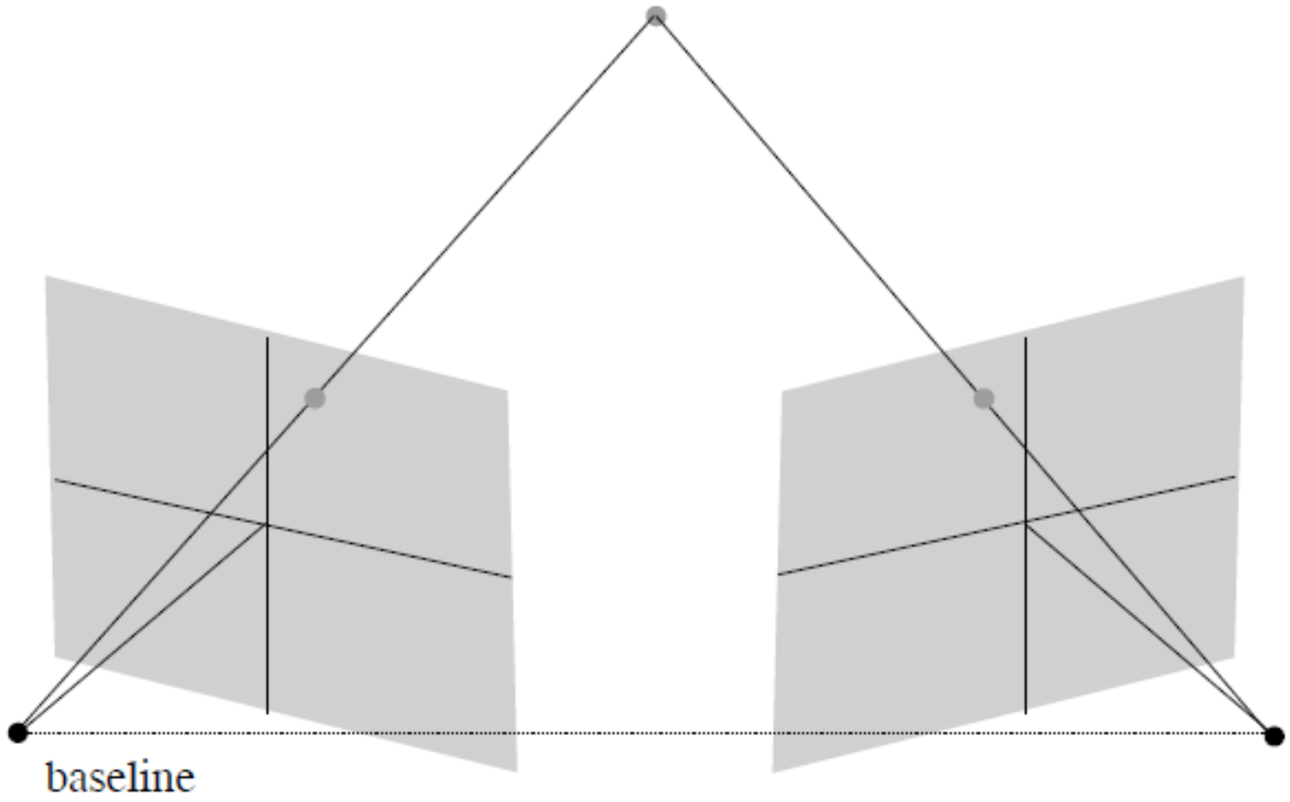
Epipolar geometry



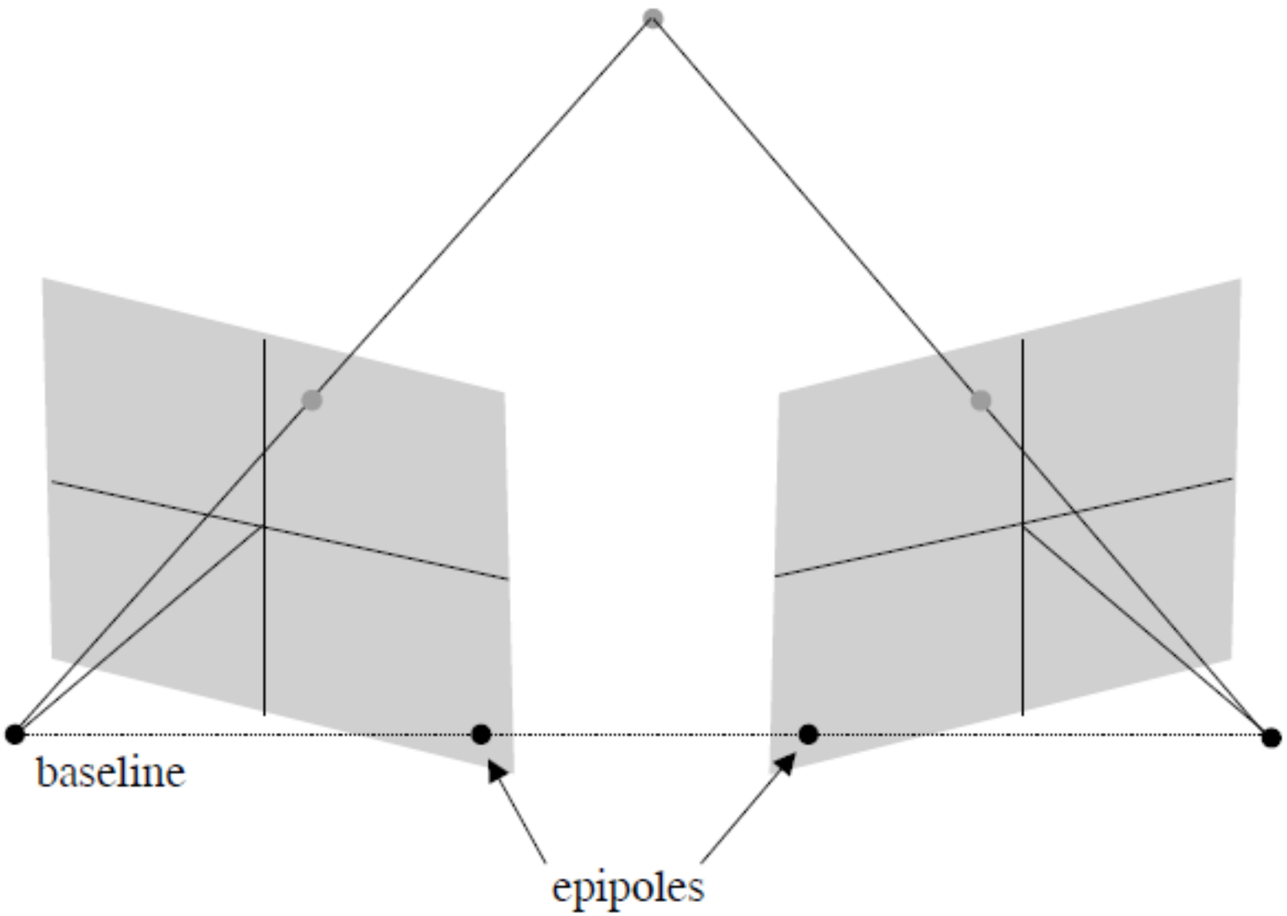
Epipolar geometry



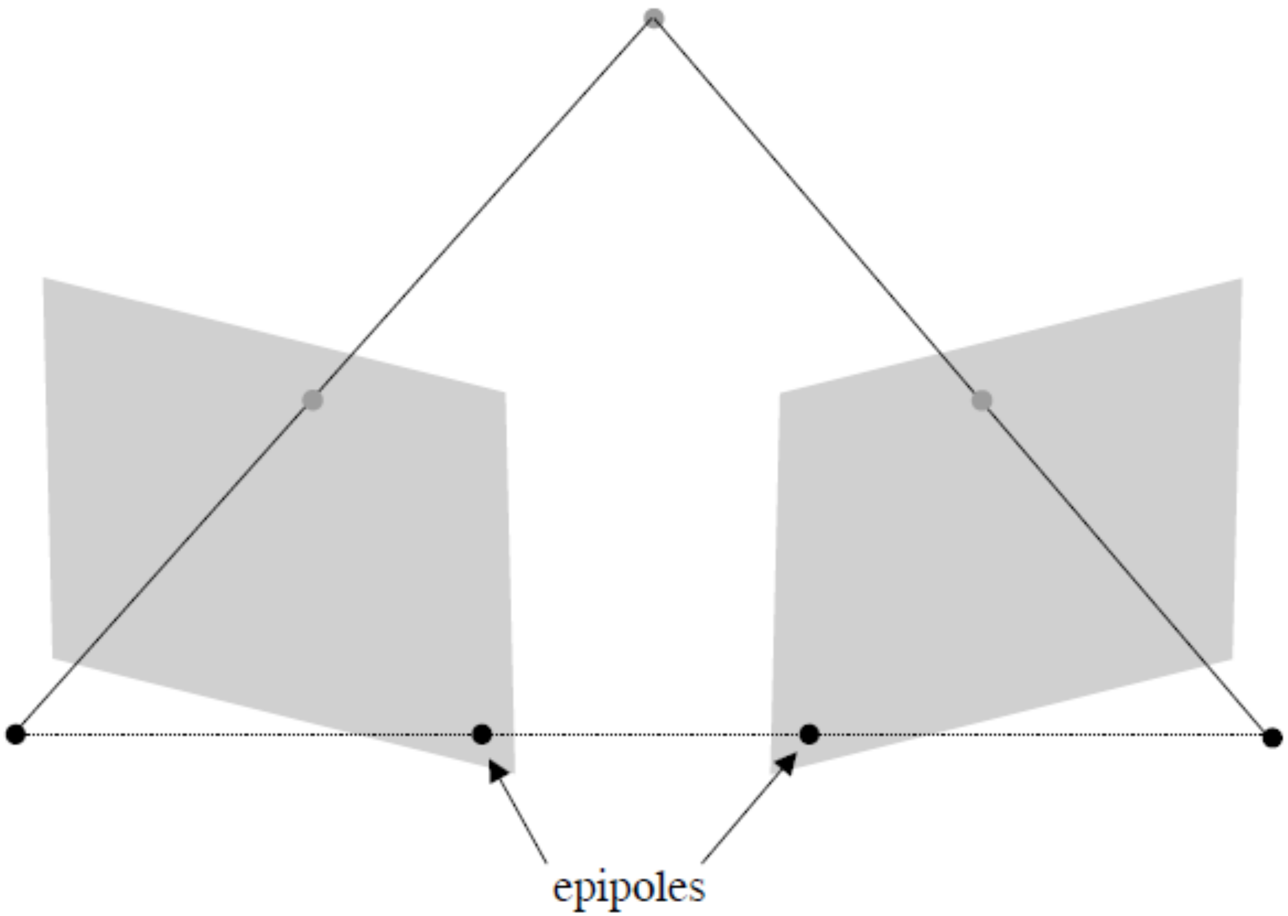
Epipolar geometry



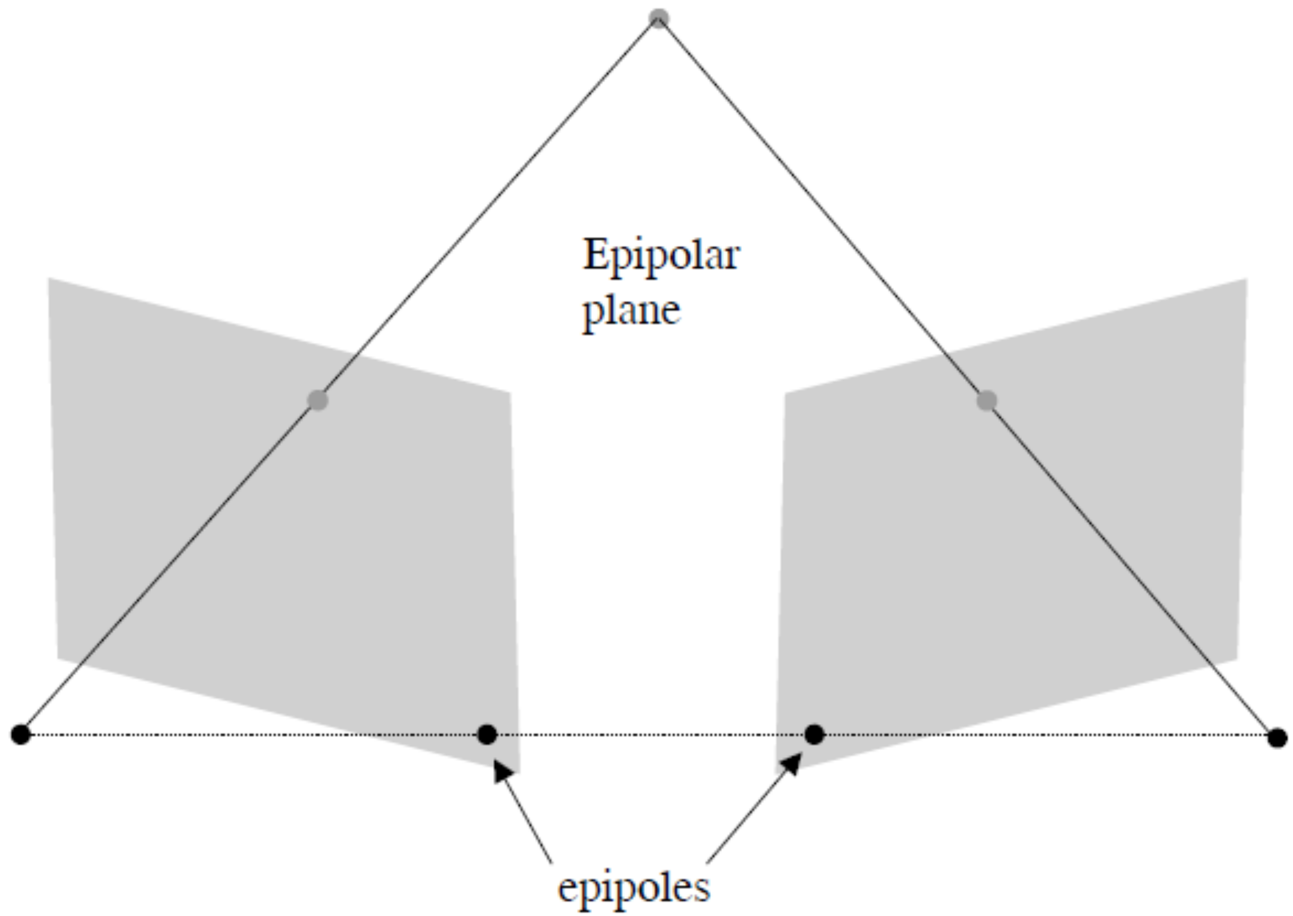
Epipolar geometry



Epipolar geometry

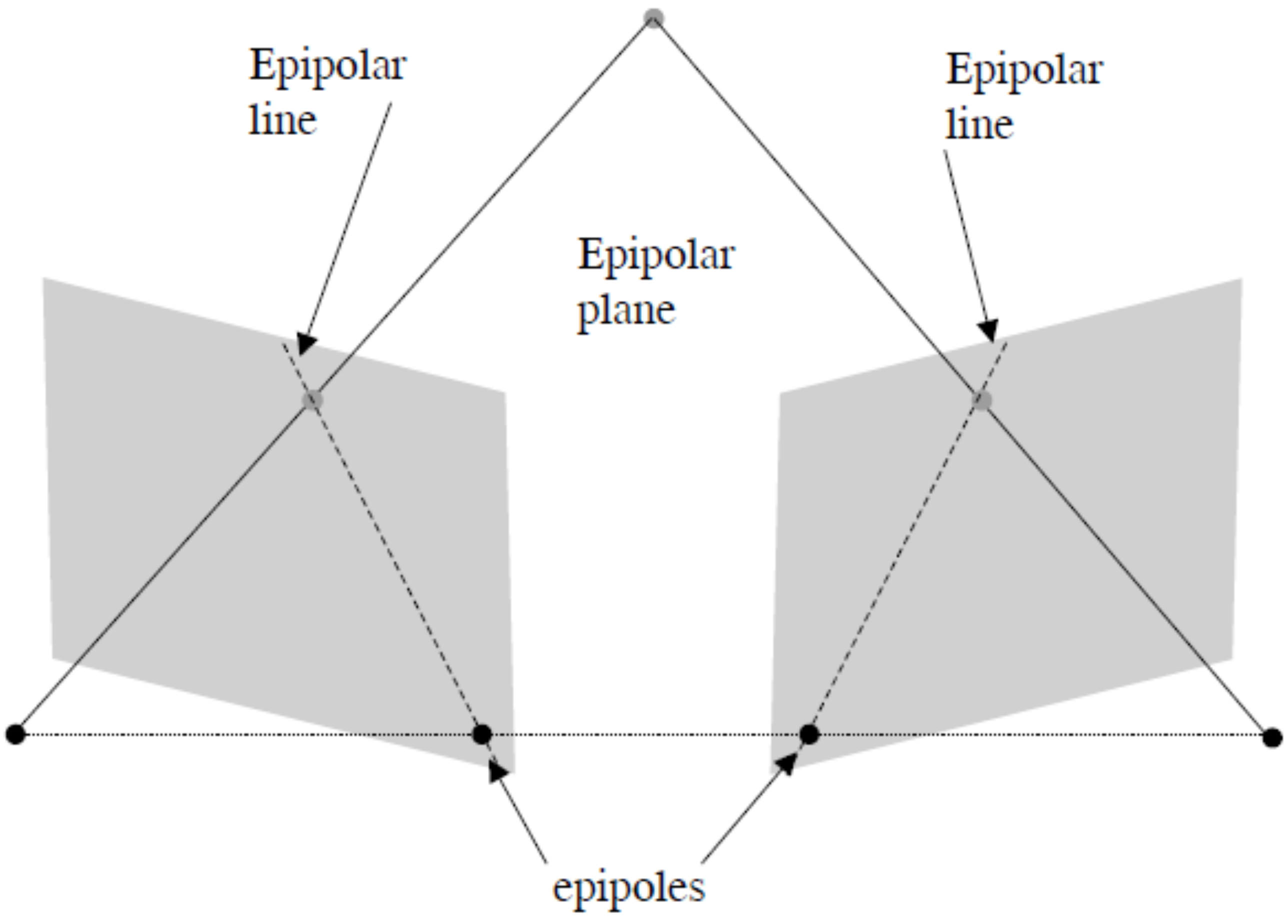


Epipolar geometry



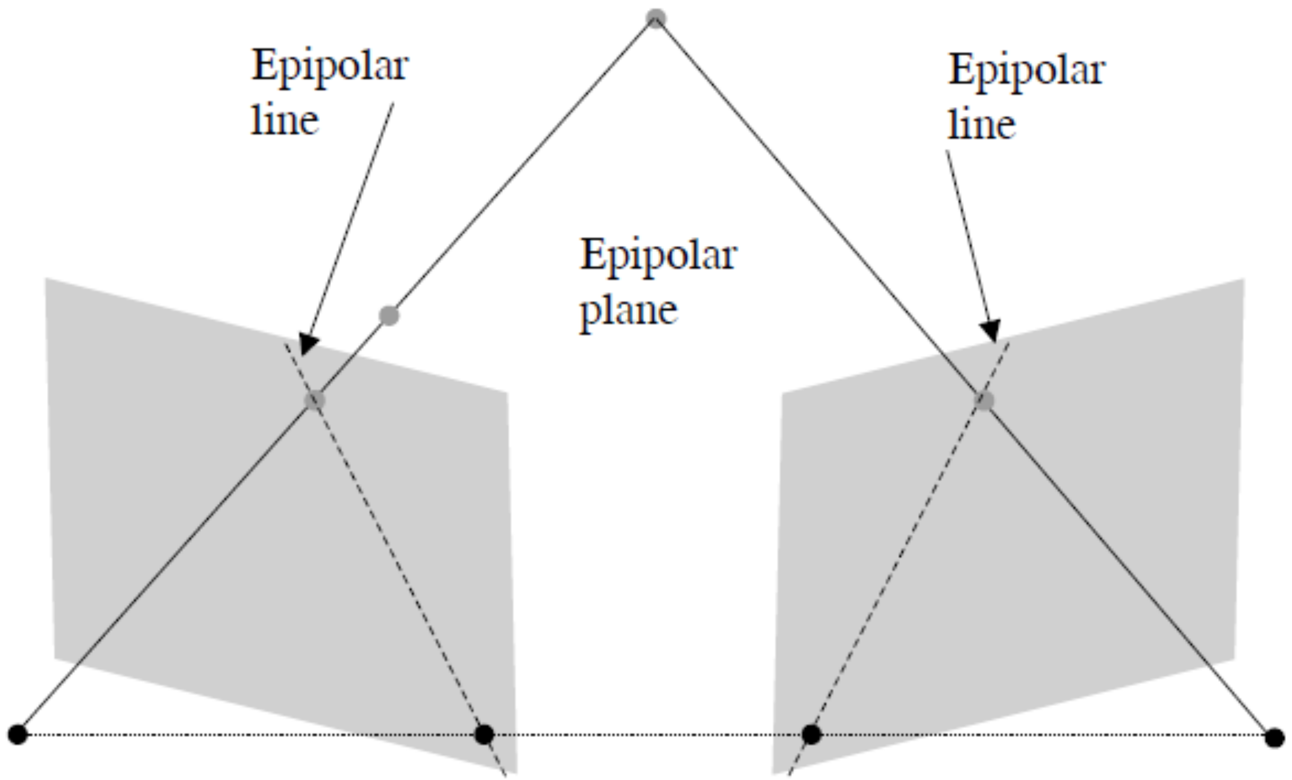
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



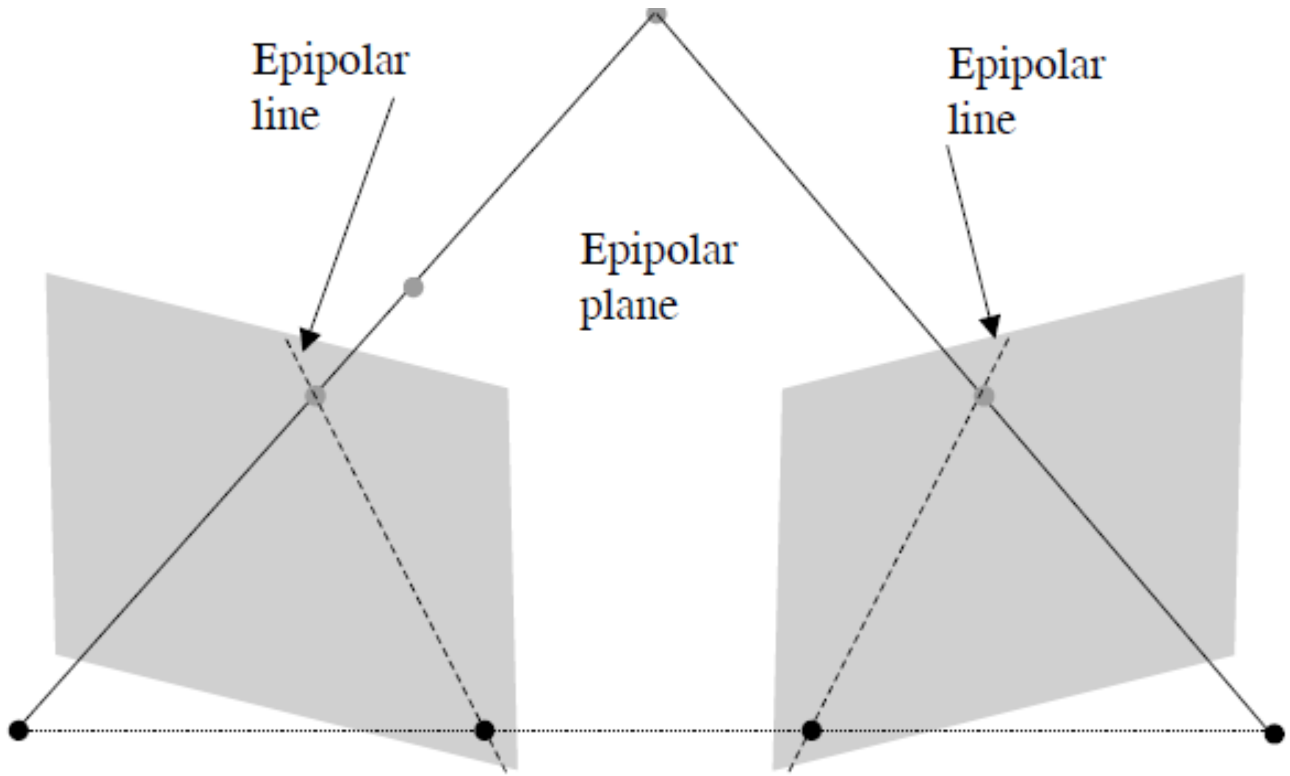
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



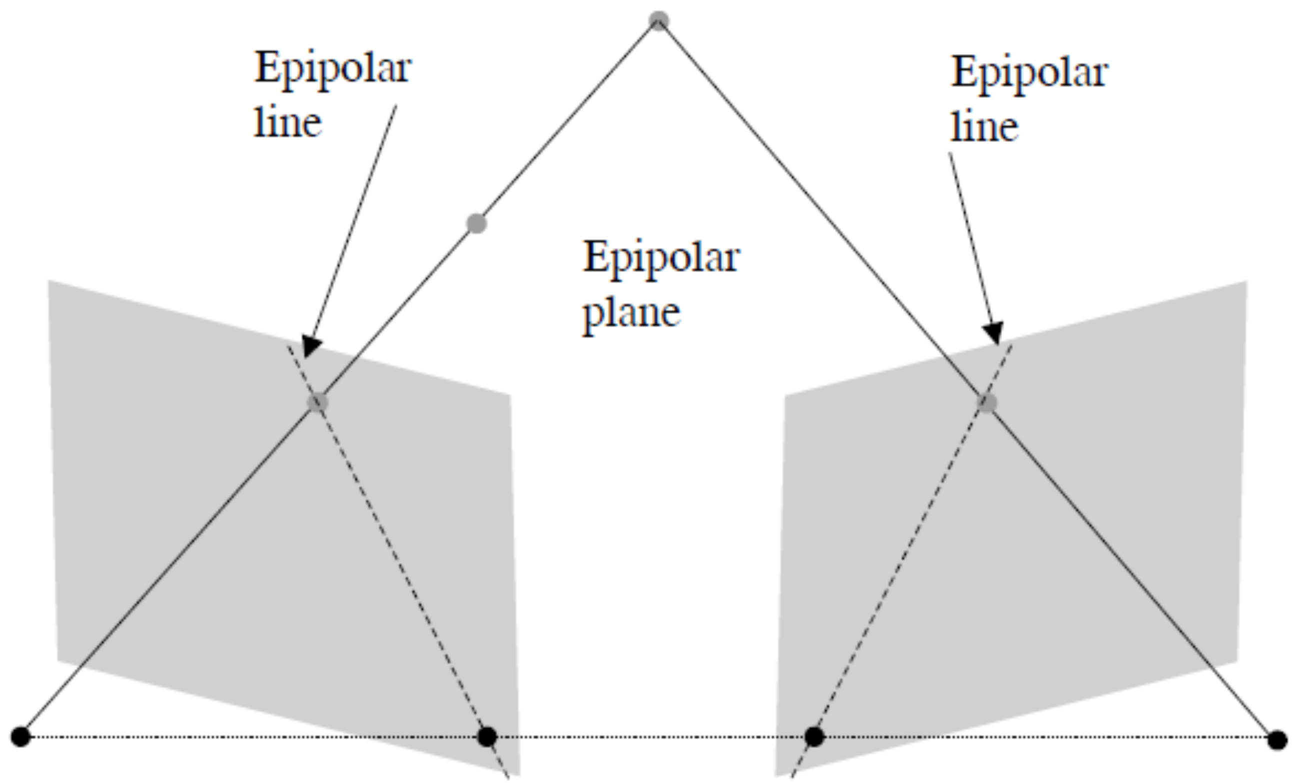
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



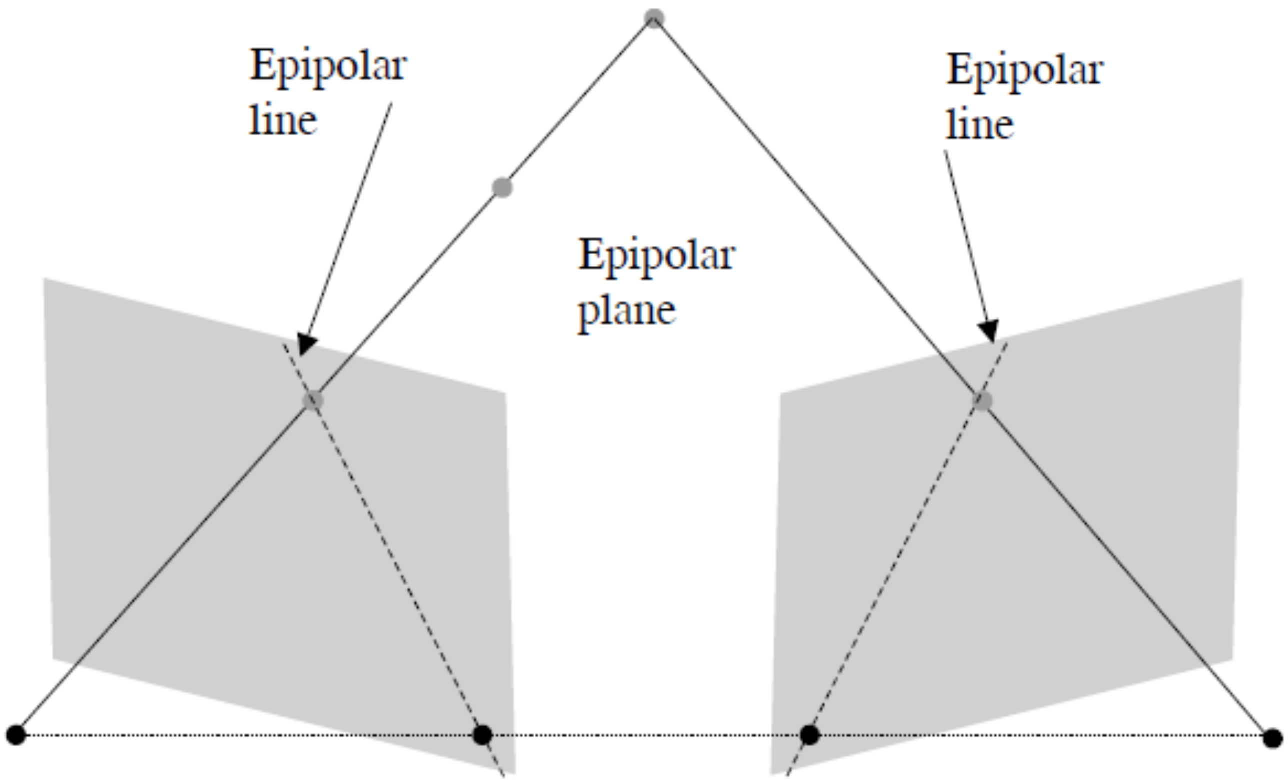
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



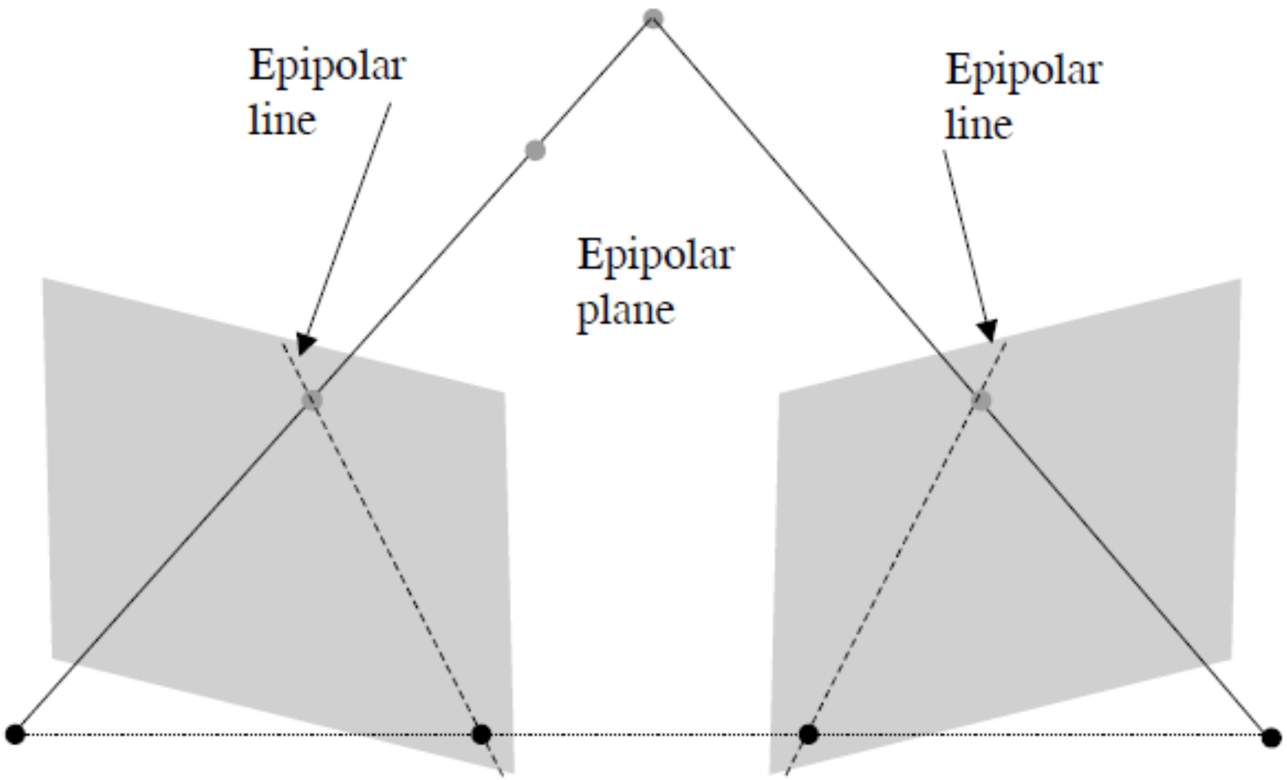
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



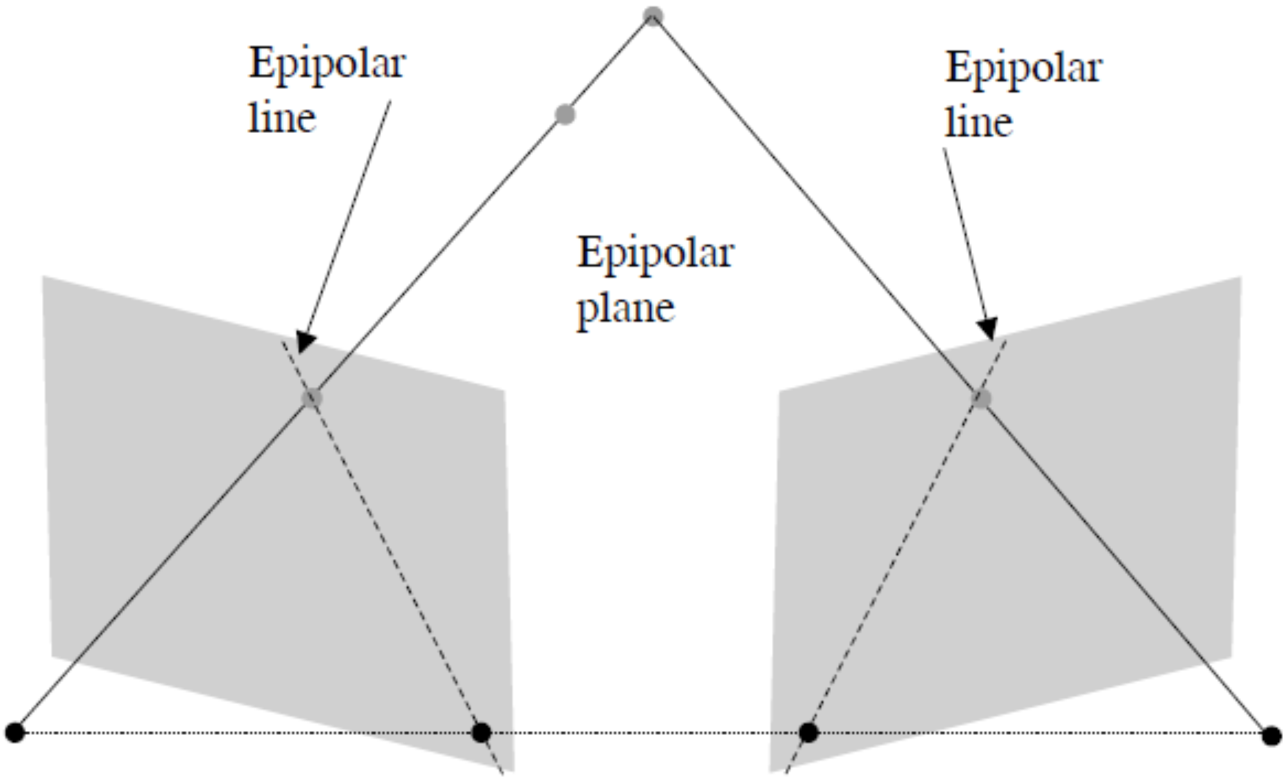
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



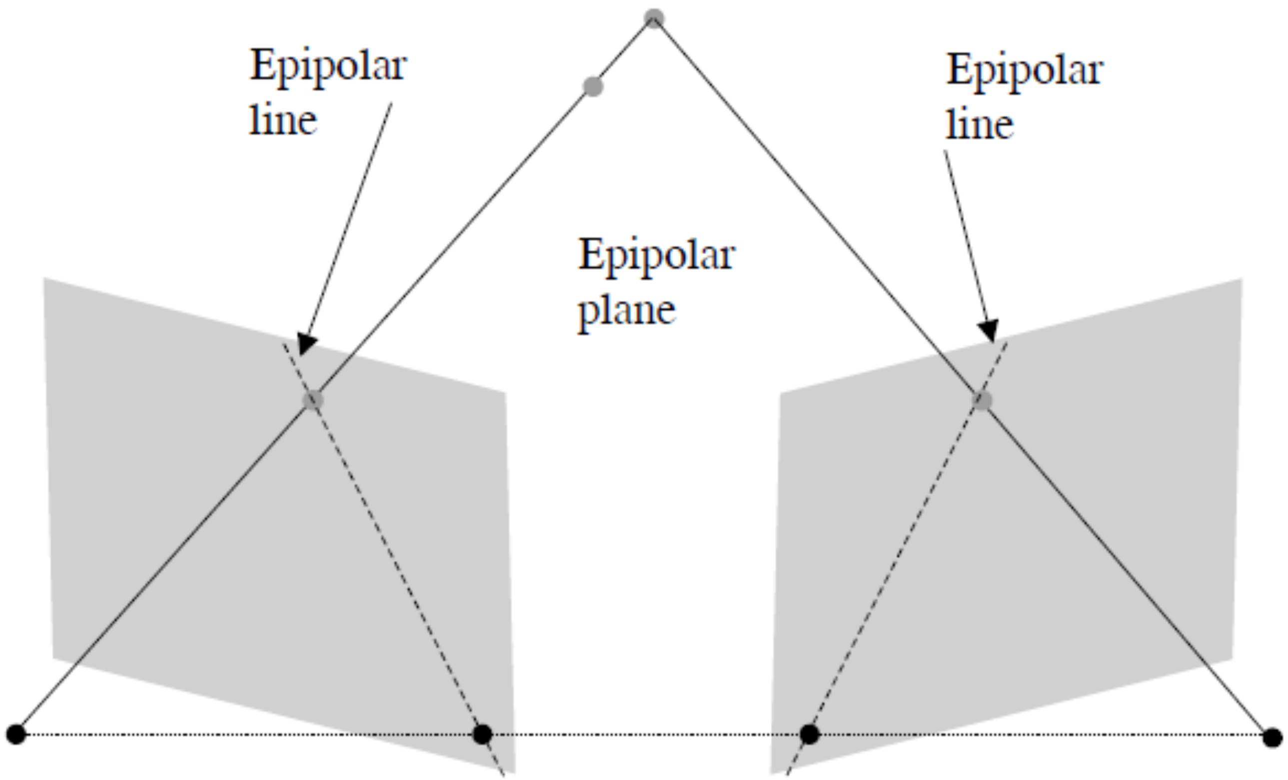
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



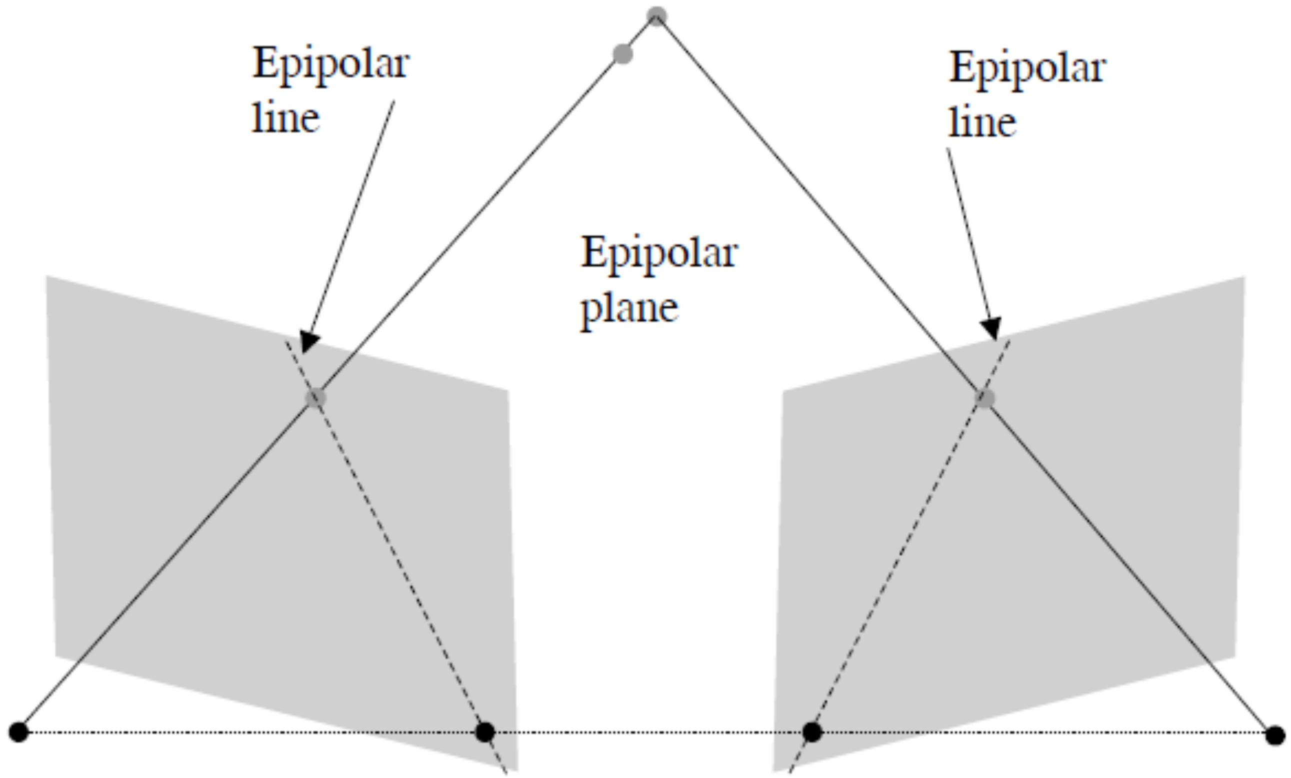
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



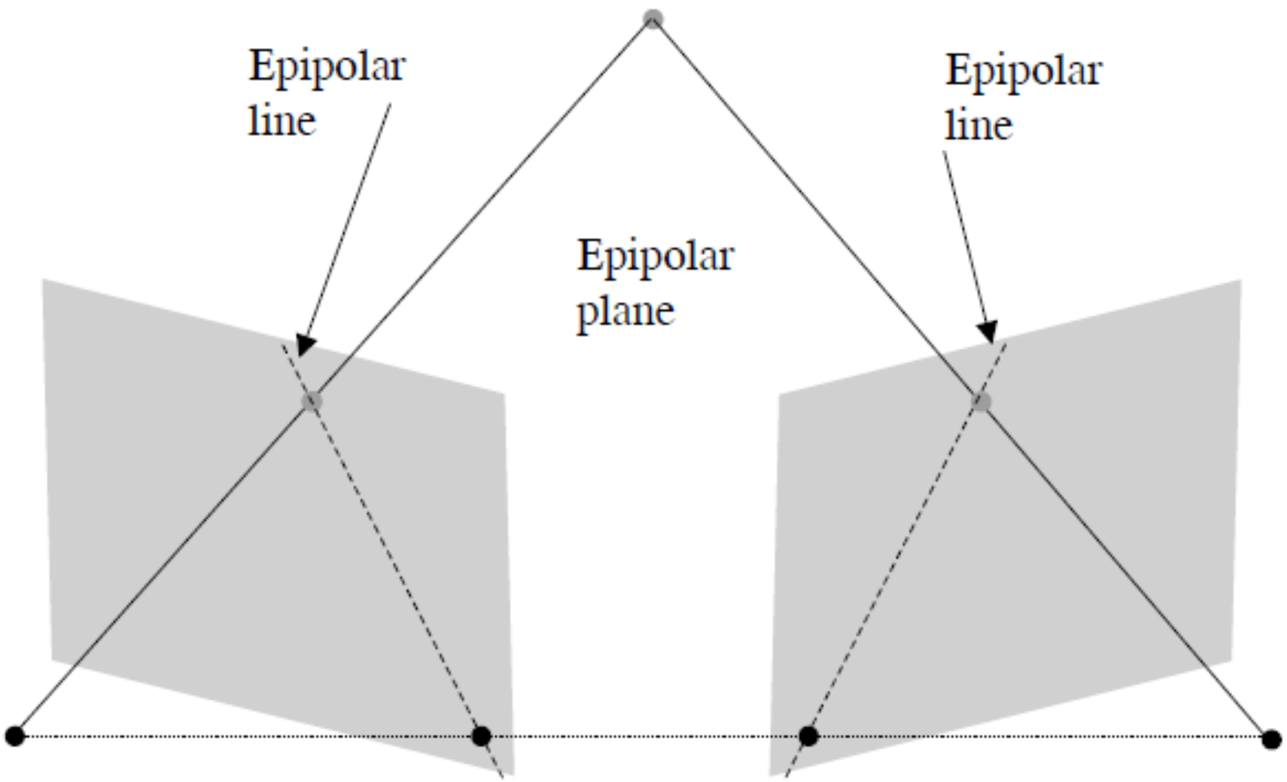
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



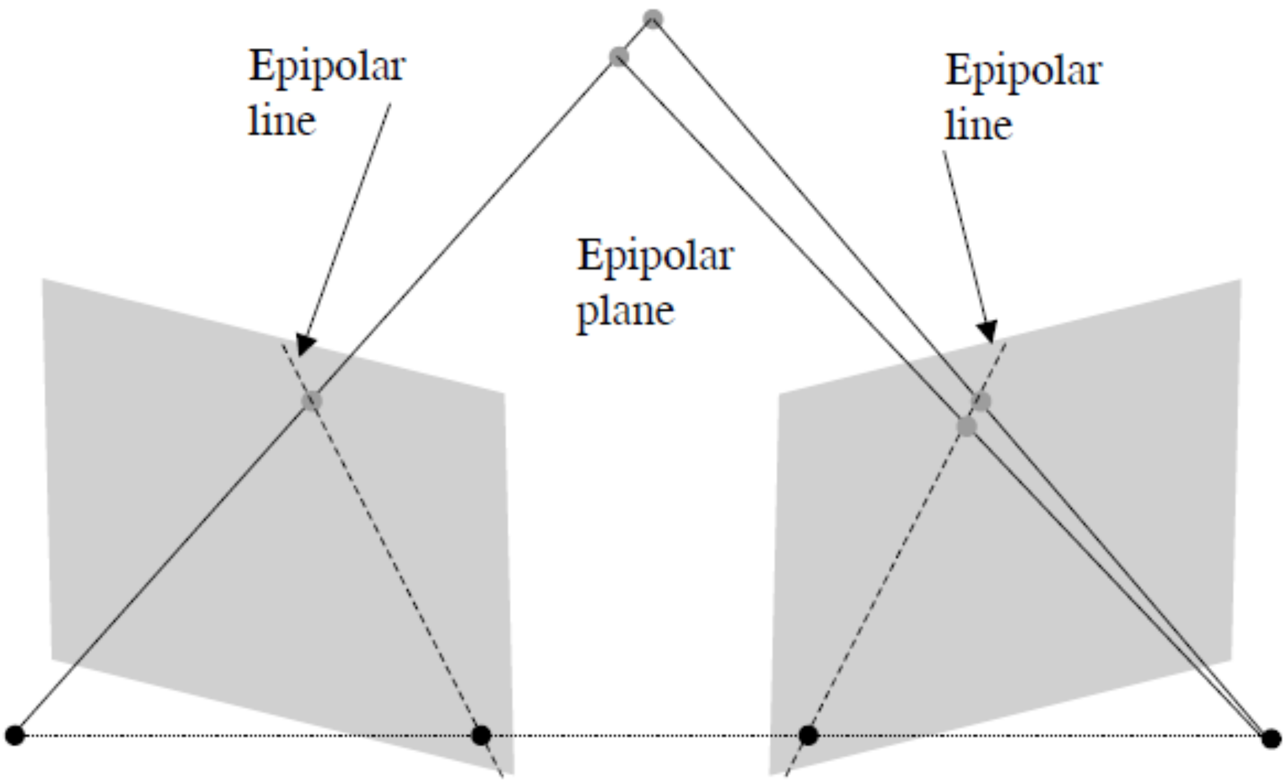
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



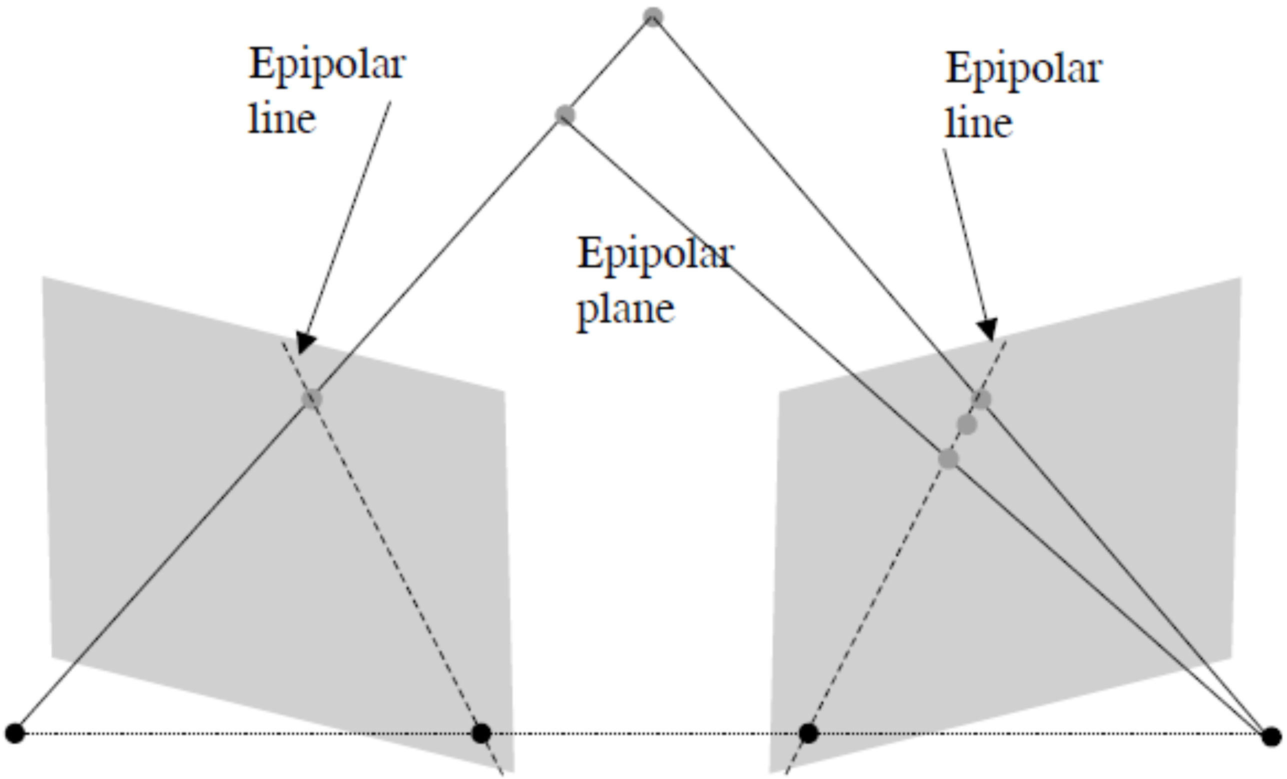
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  – There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  – The lines where the plane intersects the images are called conjugate epipolar lines



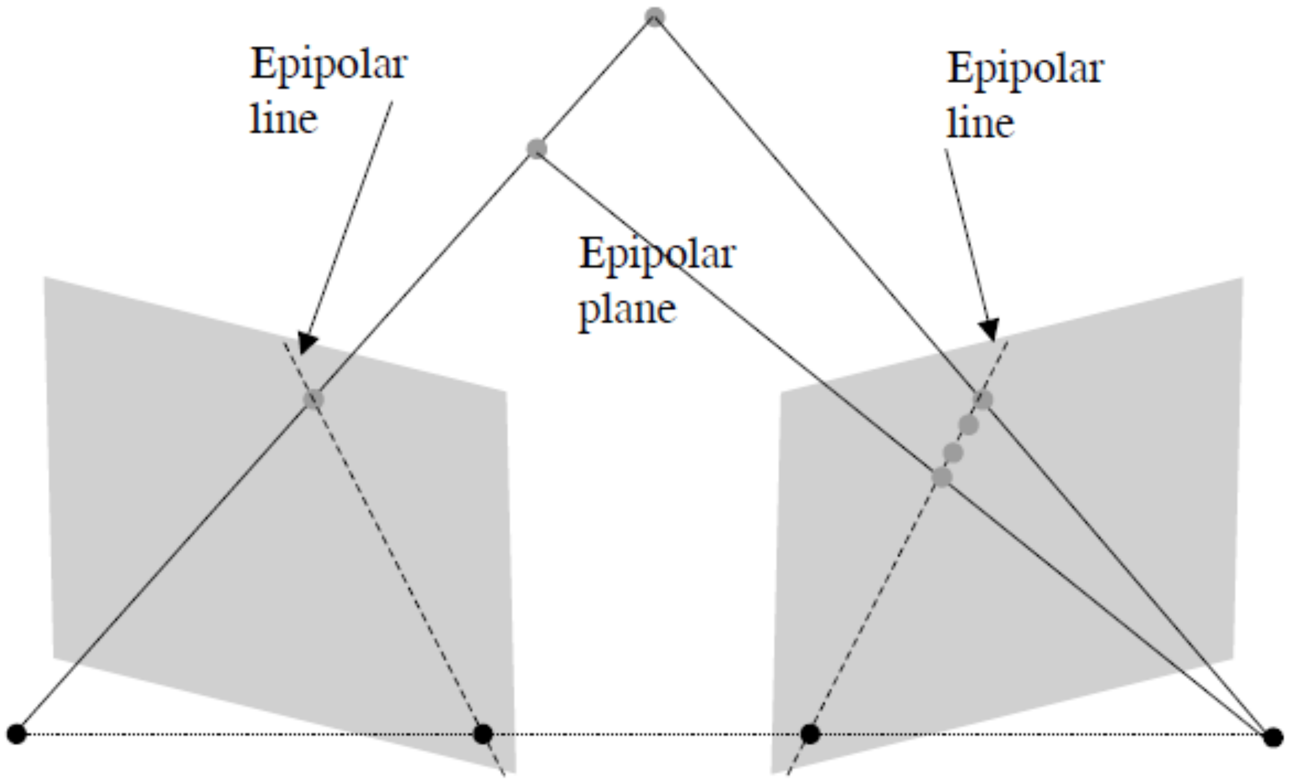
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



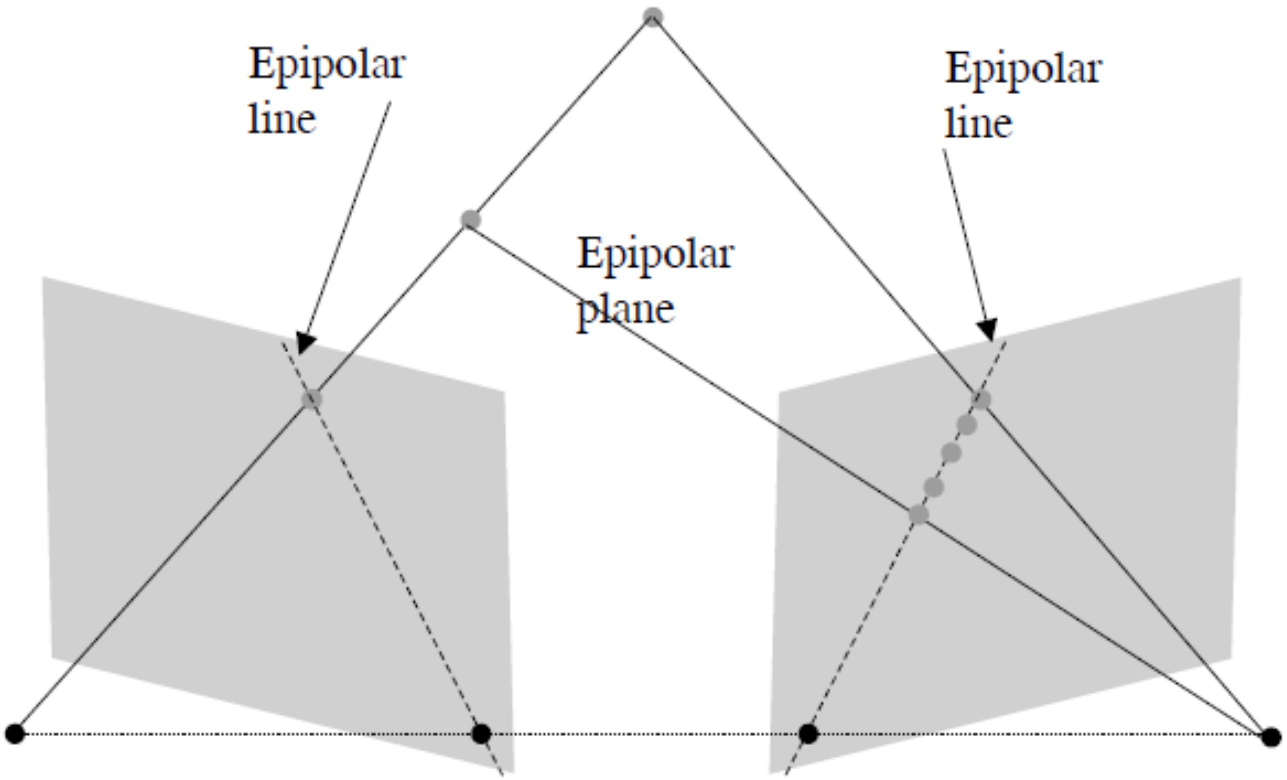
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



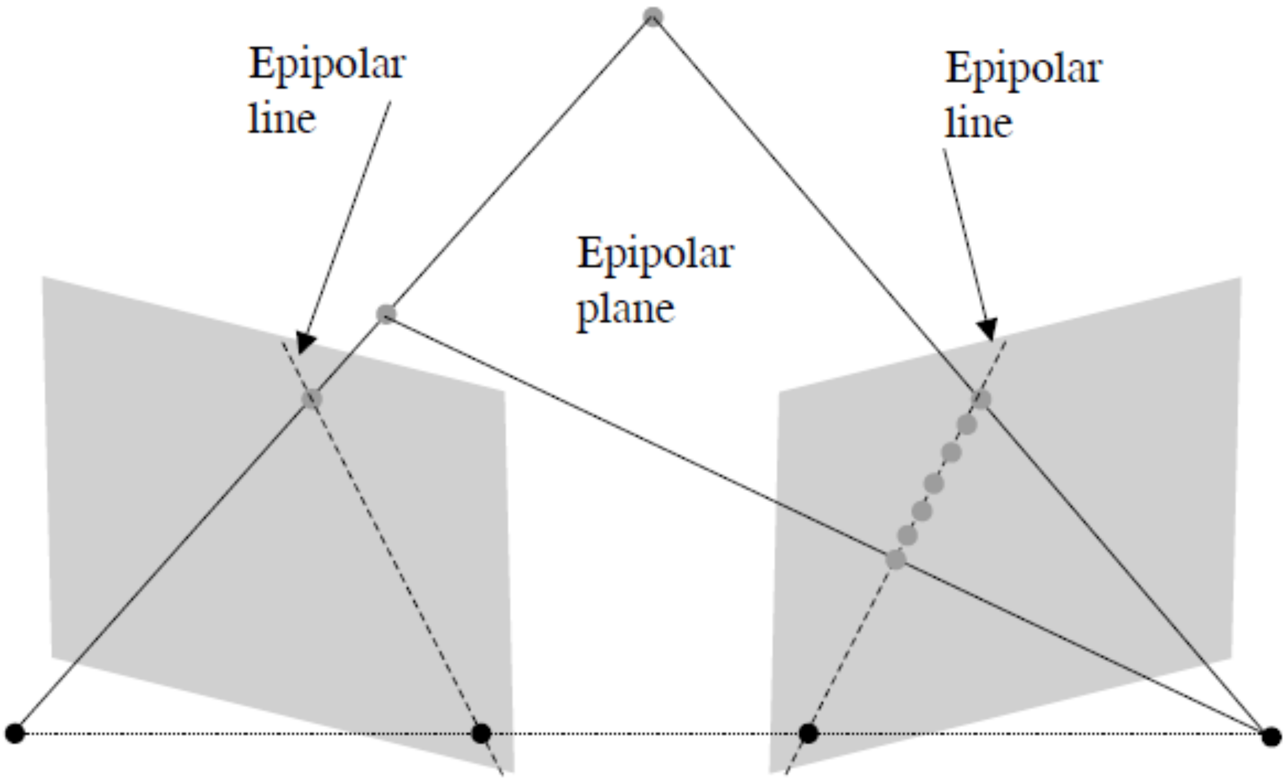
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



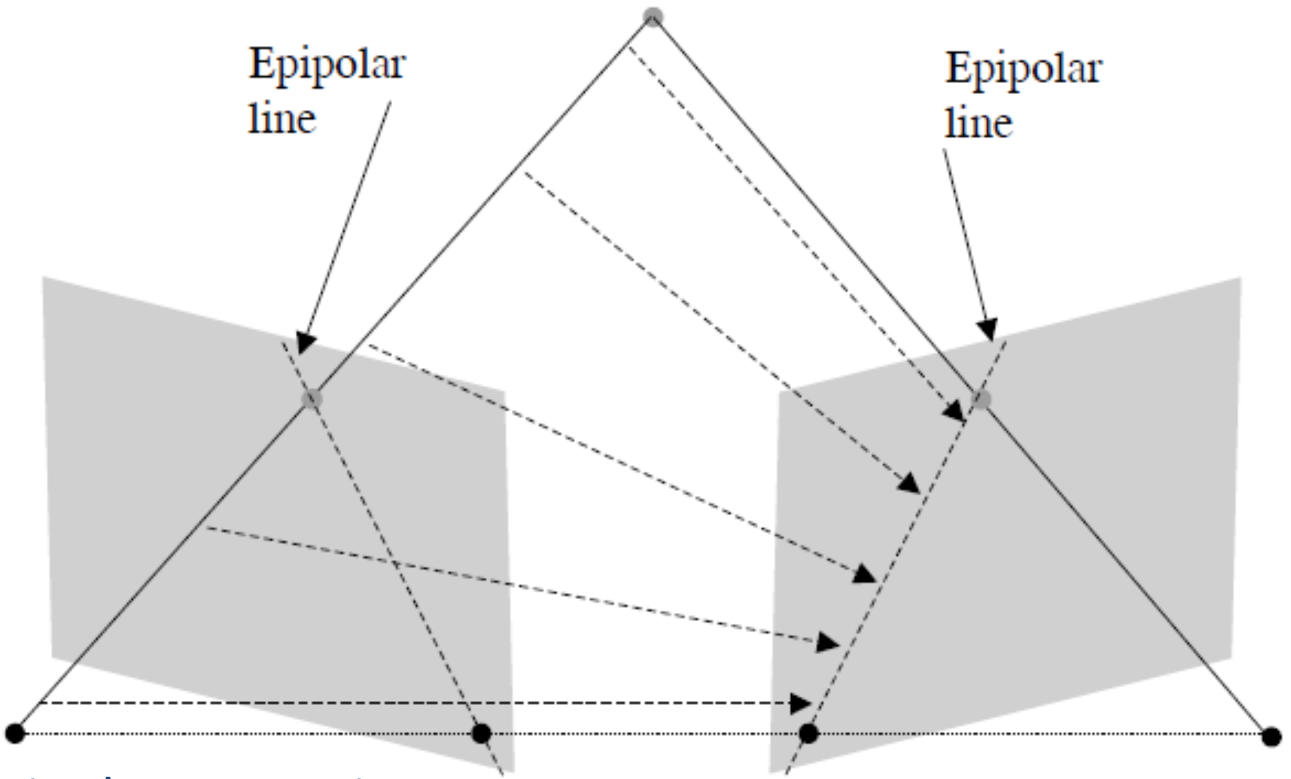
Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



Epipolar geometry

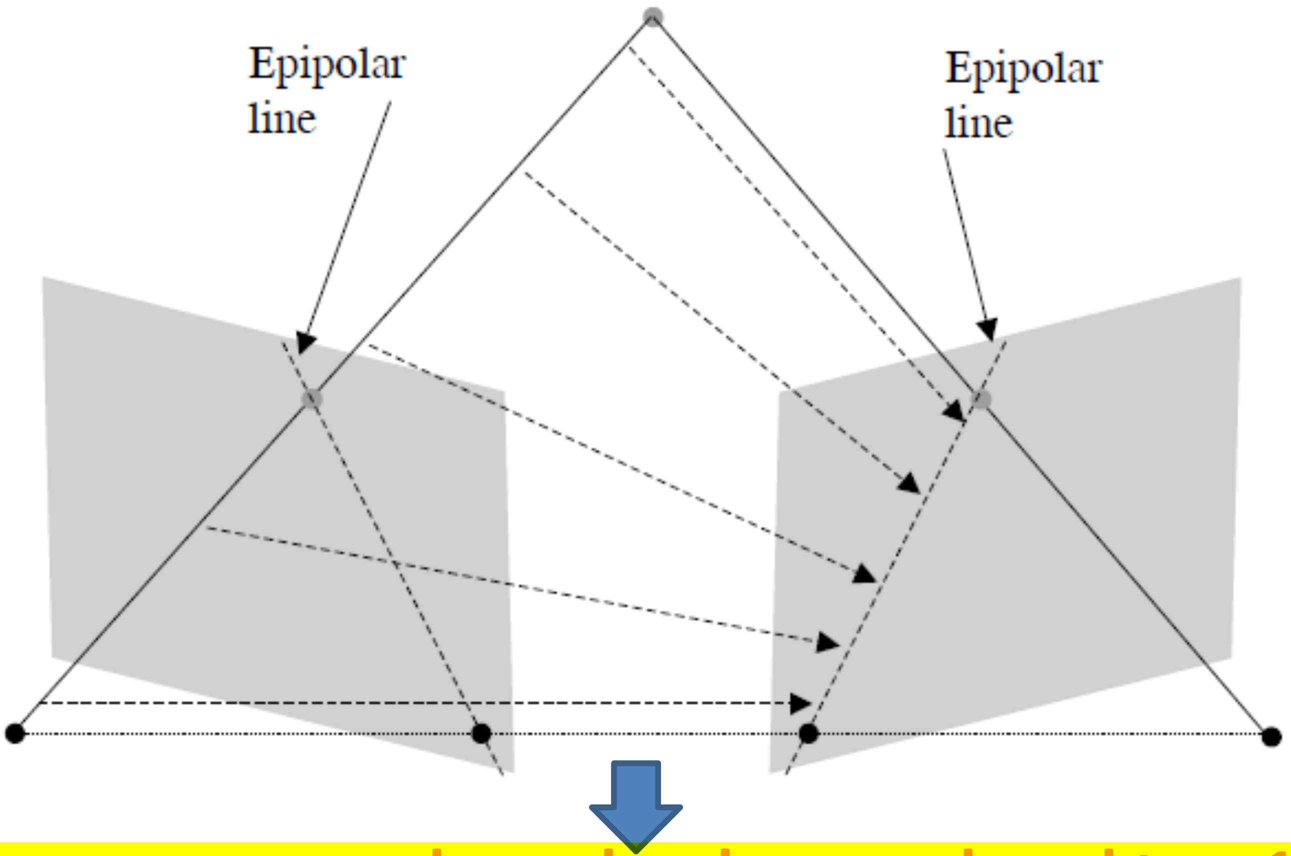
- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



- The epipolar constraint
  - The corresponding points must lie on conjugated epipolar lines

Epipolar geometry

- Given a stereo pair of cameras and a point in 3D space
  - There is a plane that goes through the point and the centres of projection of the cameras
    - We call this plane the epipolar plane
  - The lines where the plane intersects the images are called conjugate epipolar lines



Stereo correspondence has been reduced to a 1D search !

Epipolar geometry

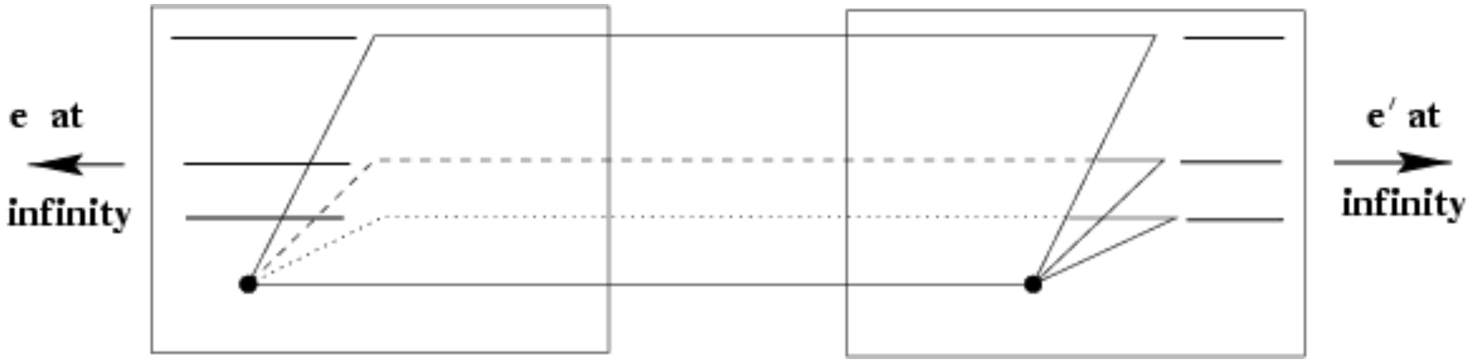
Stereo correspondence has been reduced to a 1D search !

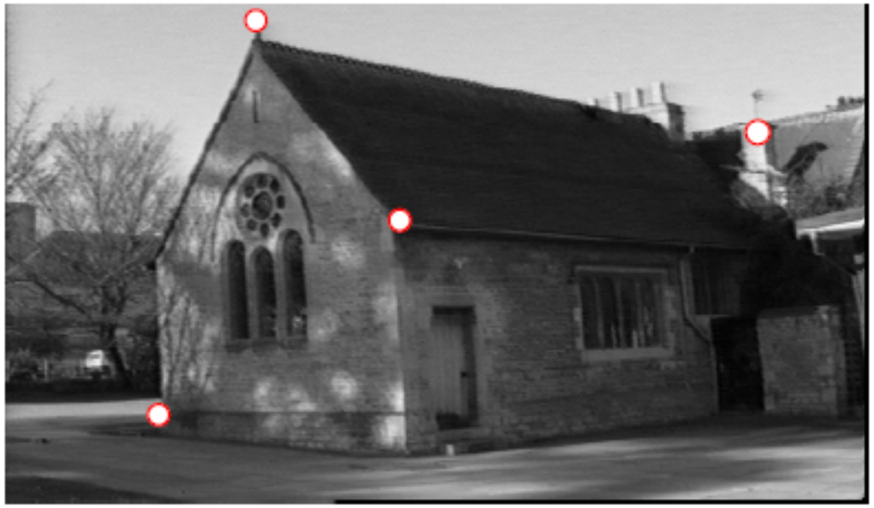
Epipolar geometry

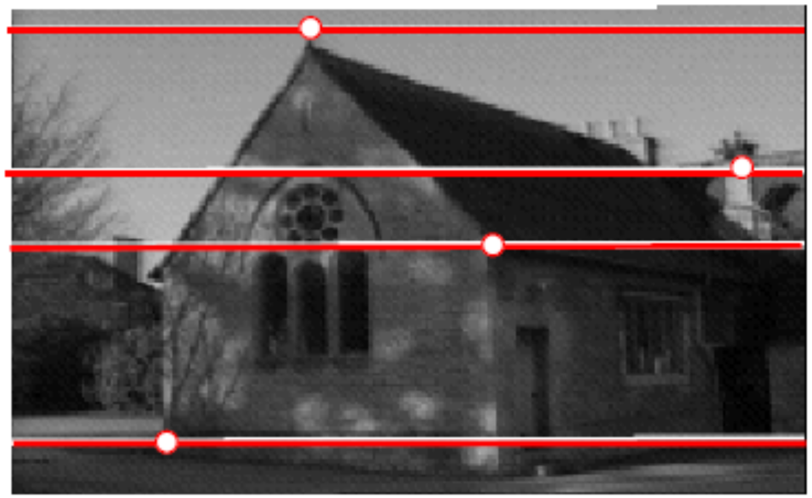
Stereo correspondence has been reduced to a 1D search !

Slide credit: Kristen Grauman

Epipolar geometry



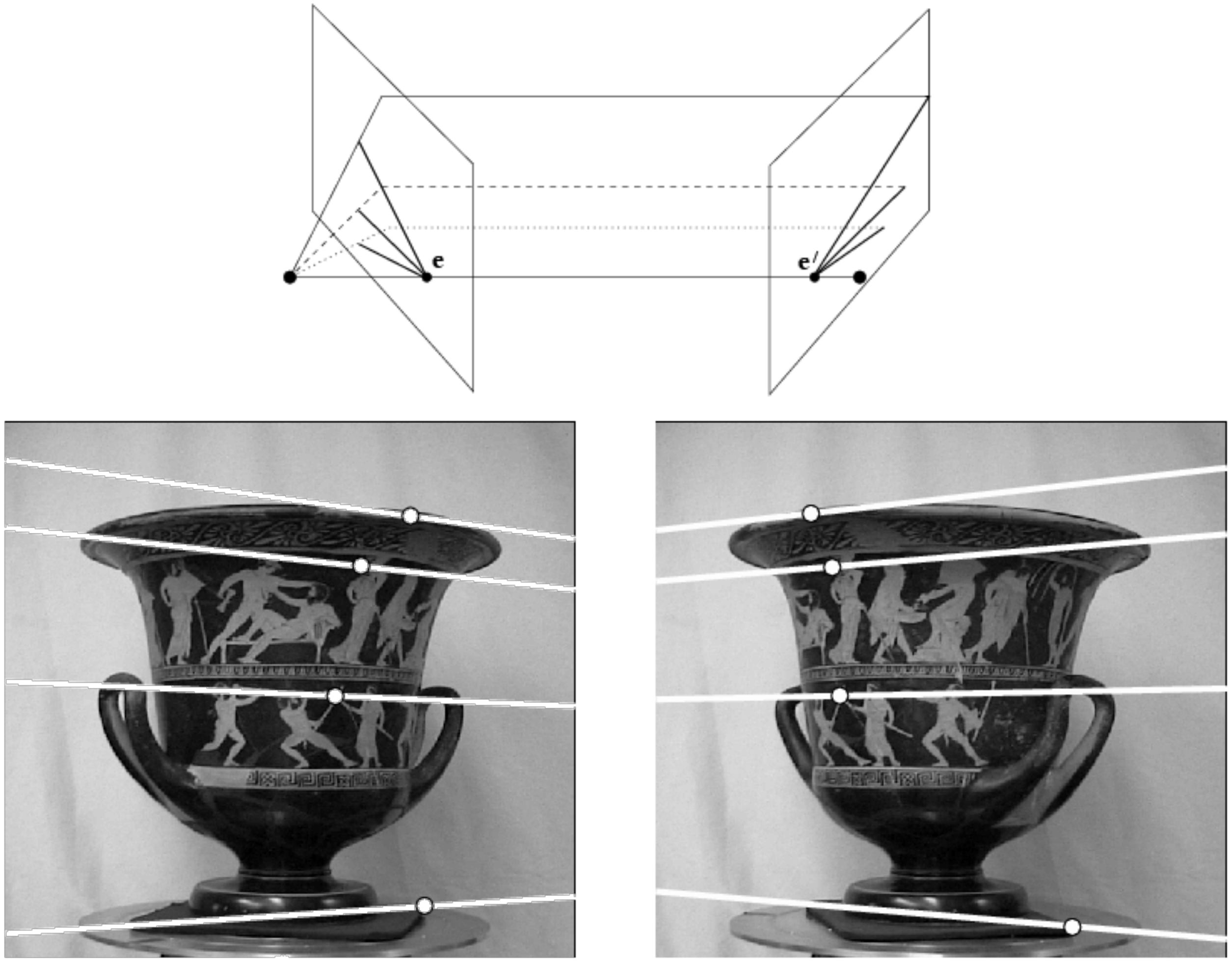




Stereo correspondence has been reduced to a 1D search !

Source: Antonio Torralba slides

Epipolar geometry

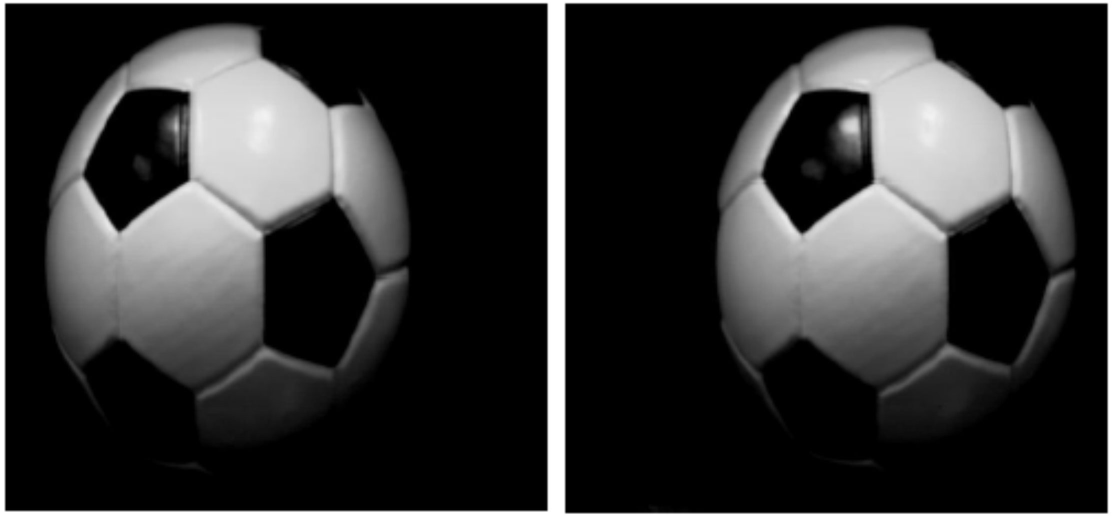


Stereo correspondence has been reduced to a 1D search !

Slide credit: Kristen Grauman

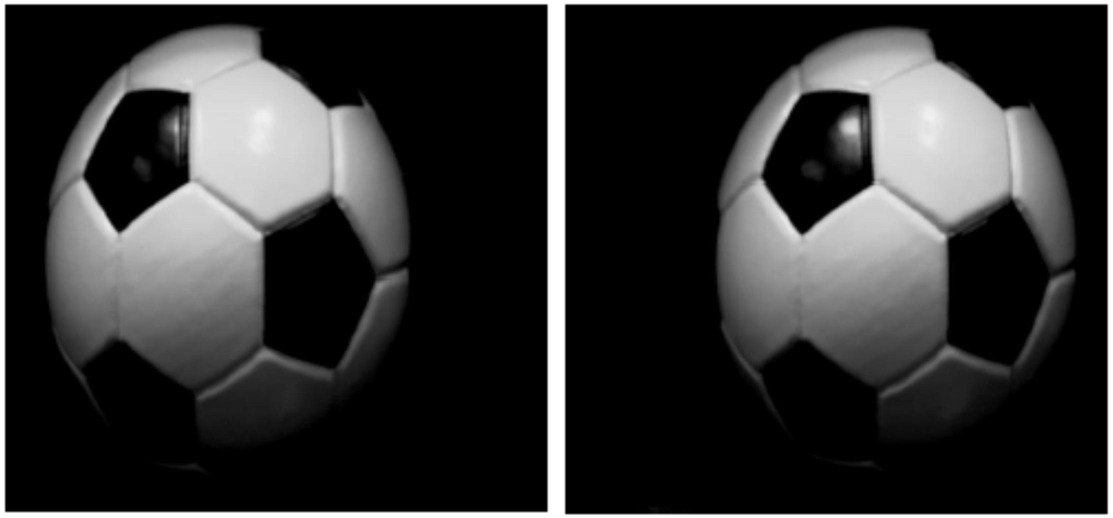
Epipolar geometry

- Recovery
  - Given a stereo pair of images, how do we calculate the epipolar geometry.
  - Once in hand, binocular correspondence is simplified to 1D search.



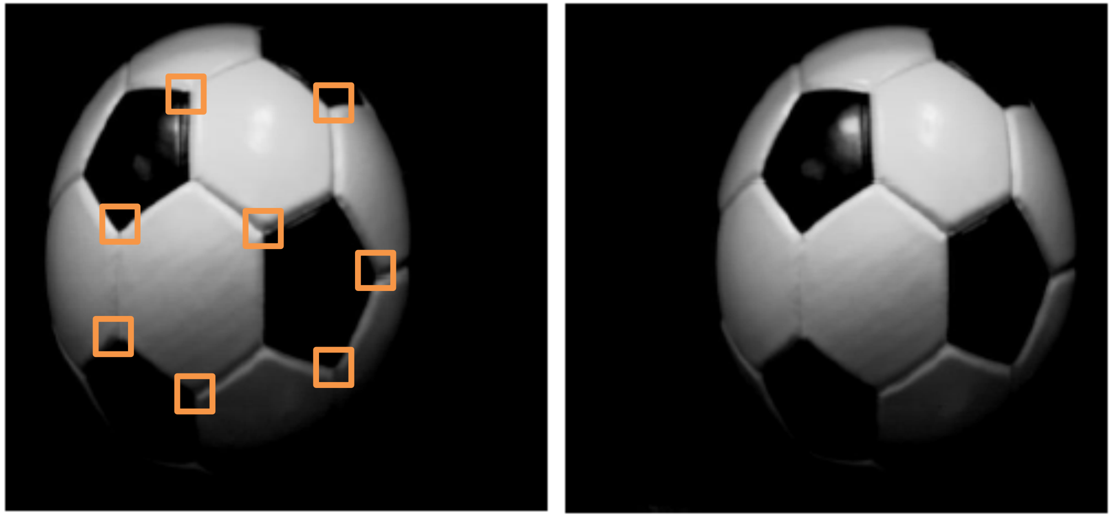
Epipolar geometry

- Two approaches
  1. Calculate camera-to-camera transformation
     - 8 point algorithm
  2. Calculate camera-to-world transformations
     - camera calibration



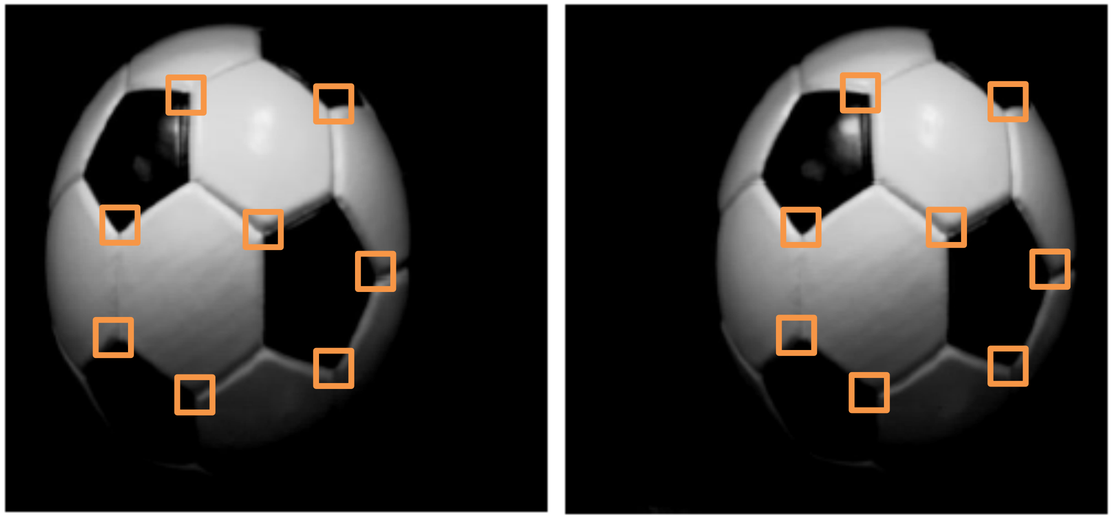
Epipolar geometry

- 8 point algorithm
  - Extract well localized distinctive features on each image.
    - E.g., corners



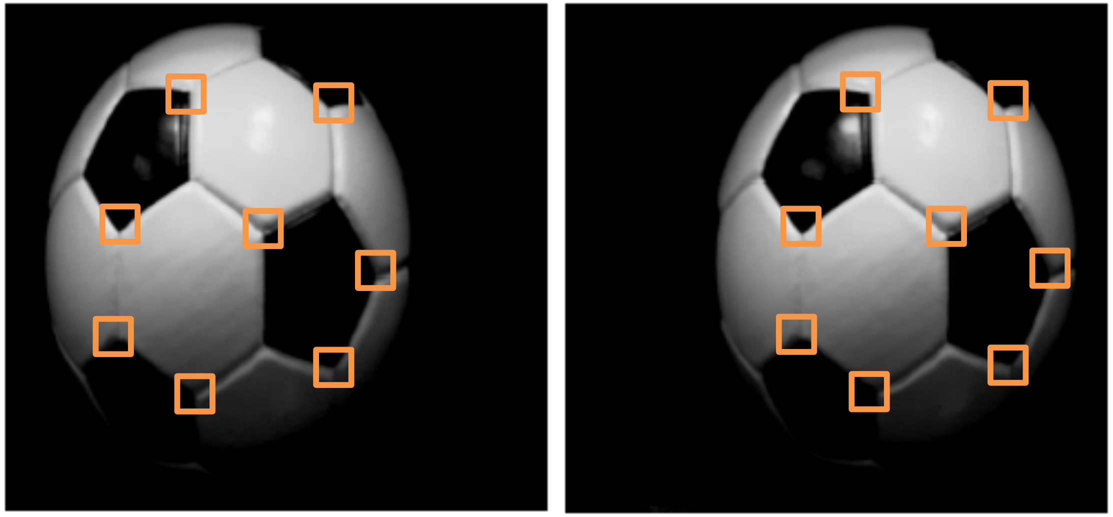
Epipolar geometry

- 8 point algorithm
  - Extract well localized distinctive features on each image.
    - E.g., corners
  - Establish 8 (or more) precise correspondences between features across the image pair.
    - As general 2D search and match problem.



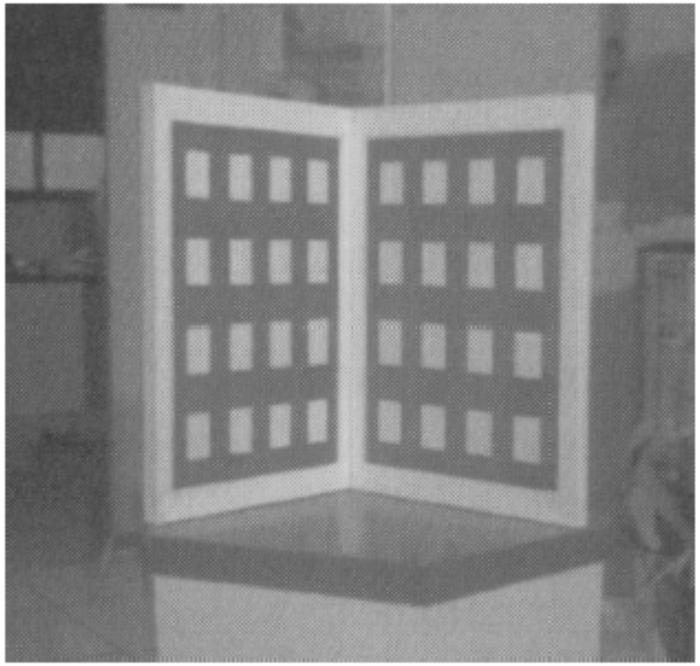
Epipolar geometry

- 8 point algorithm
  - Extract well localized distinctive features on each image.
    - E.g., corners
  - Establish 8 (or more) precise correspondences between features across the image pair.
    - As general 2D search and match problem.
  - This allows for the parameters of a geometric transformation to be estimated that unambiguously defines the epipolar geometry.
    - Transformation intimately related to relative intrinsic and extrinsic camera parameters of the system.



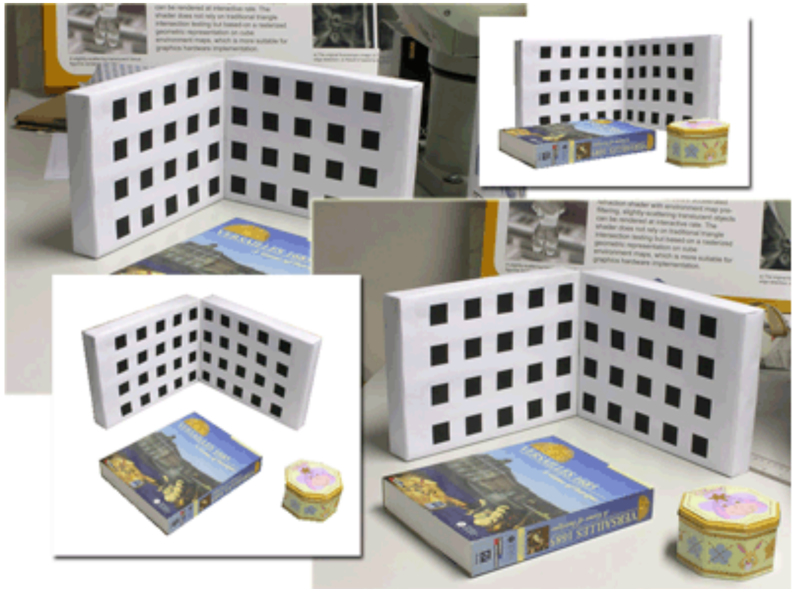
Epipolar geometry

- Camera calibration
  - Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely measured in 3D position.



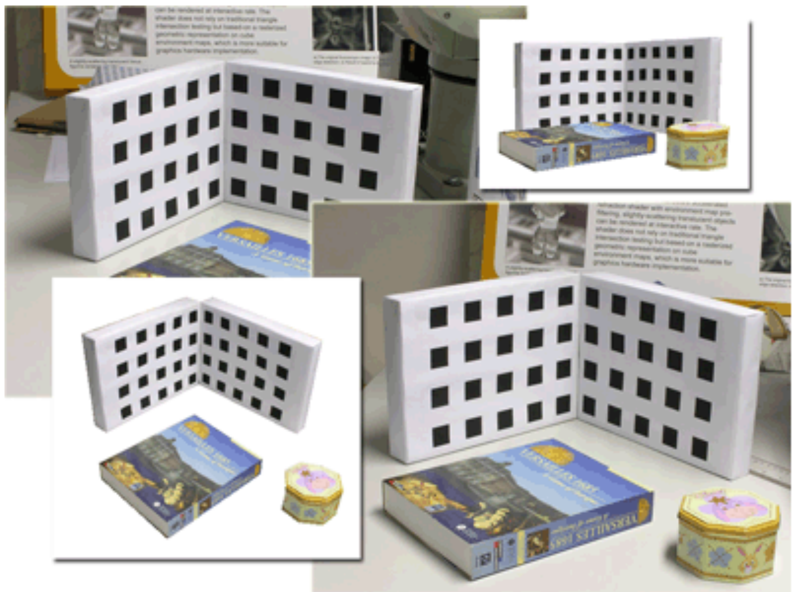
Epipolar geometry

- Camera calibration
  - Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely measured in 3D position.
  - Capture images of the calibration pattern with both cameras.



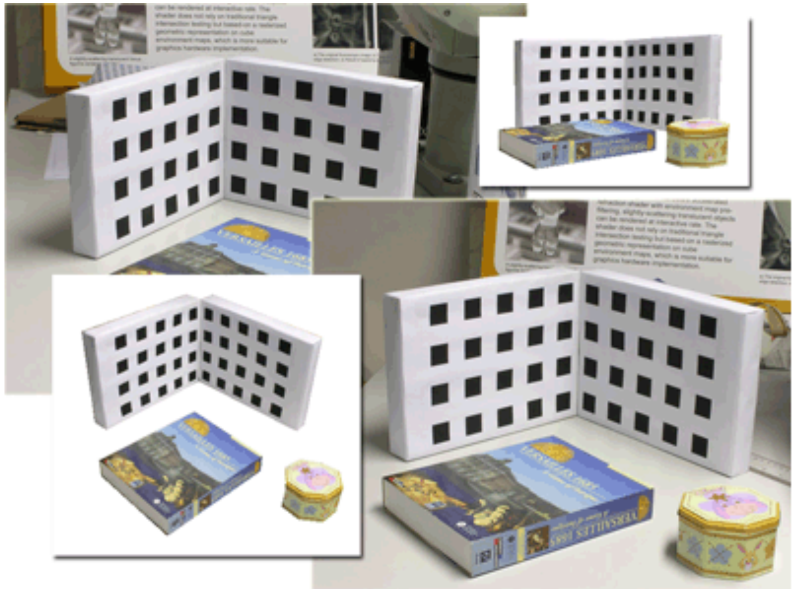
Epipolar geometry

- Camera calibration
  - Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely measured in 3D position.
  - Capture images of the calibration pattern with both cameras.
  - Precisely extract corresponding features between each image and the 3D calibration pattern.
    - Since this only need be done once (occasionally), can be done with human intervention.



Epipolar geometry

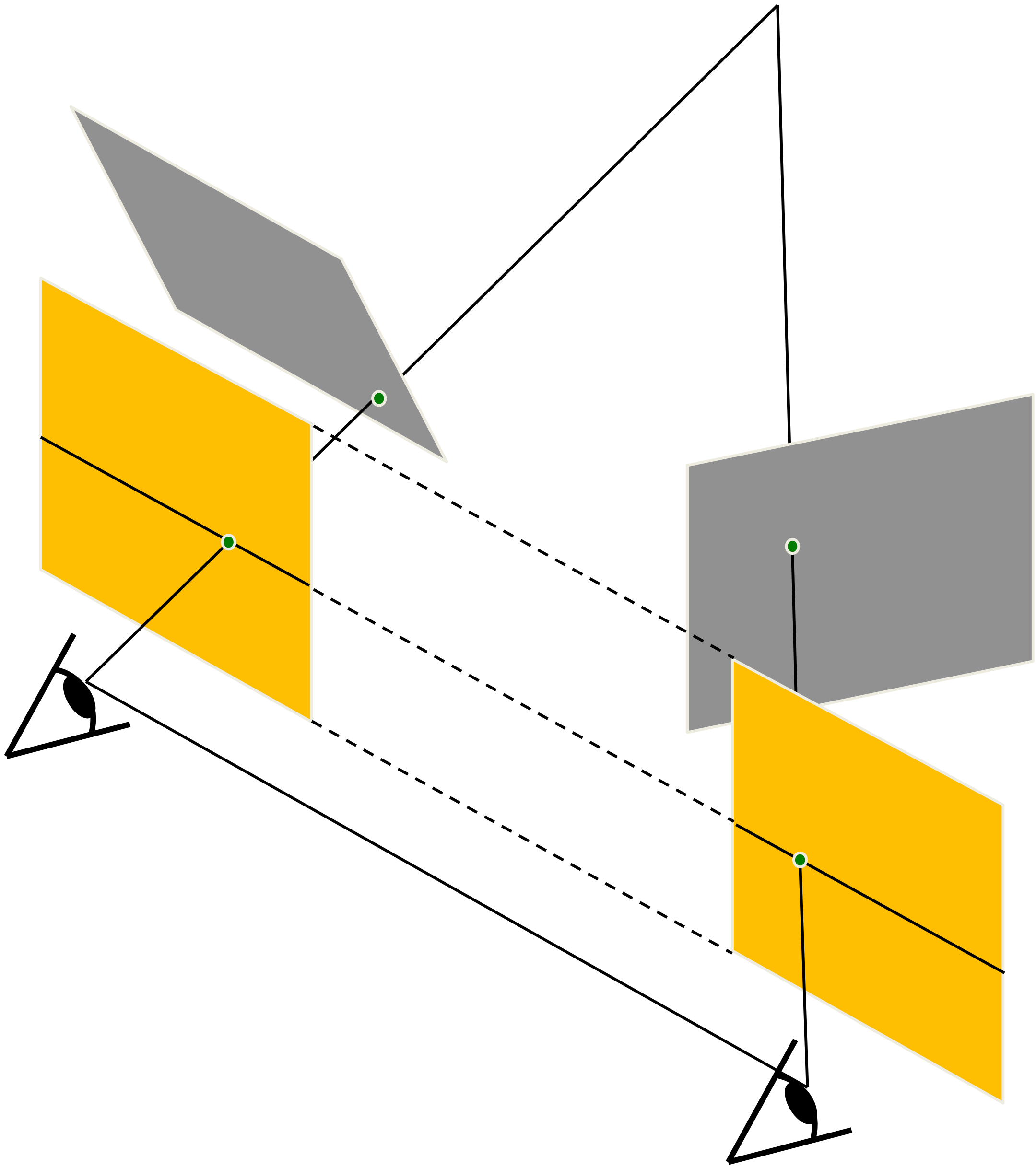
- Camera calibration
  - Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely measured in 3D position.
  - Capture images of the calibration pattern with both cameras.
  - Precisely extract corresponding features between each image and the 3D calibration pattern.
    - Since this only need be done once (occasionally), can be done with human intervention.
  - This allows for exact recovery of the intrinsic and extrinsic camera parameters
    - For each camera separately.



Epipolar geometry

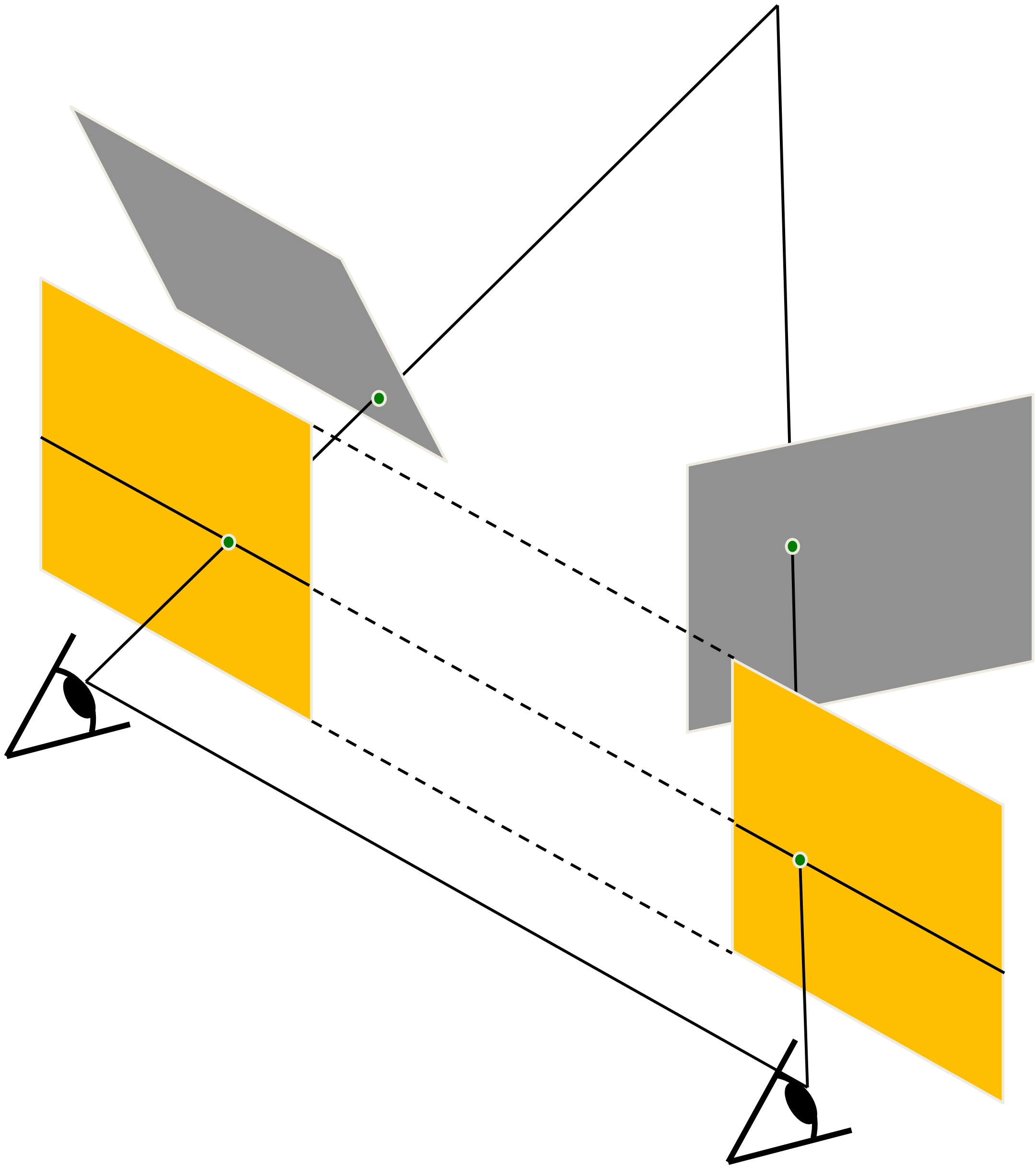
- Camera calibration
  - Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely measured in 3D position.
  - Capture images of the calibration pattern with both cameras.
  - Precisely extract corresponding features between each image and the 3D calibration pattern.
    - Since this only need be done once (occasionally), can be done with human intervention.
  - This allows for exact recovery of the intrinsic and extrinsic camera parameters
    - For each camera separately.
  - The relative camera geometry is then straightforward to recover
    - Which yields the epipolar geometry.

Stereo image rectification



Slide credit: Kristen Grauman

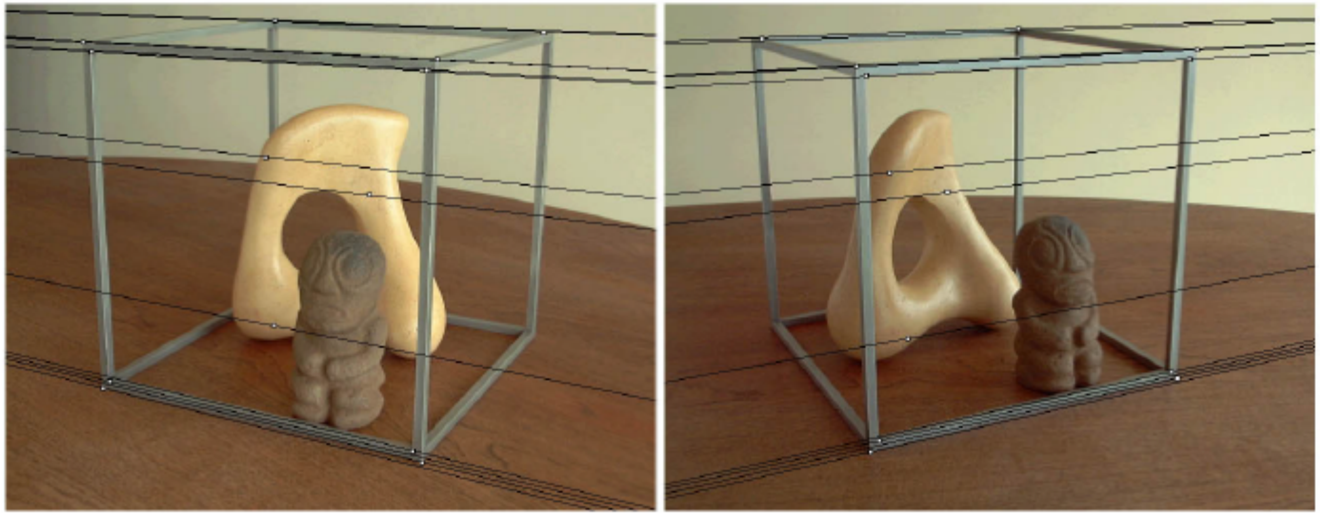
Stereo image rectification



- Reproject image planes onto a common plane parallel to the line between optical centers
- Pixel motion is horizontal after this transformation
- Two homographies (3x3 transforms), one for each input image reprojection

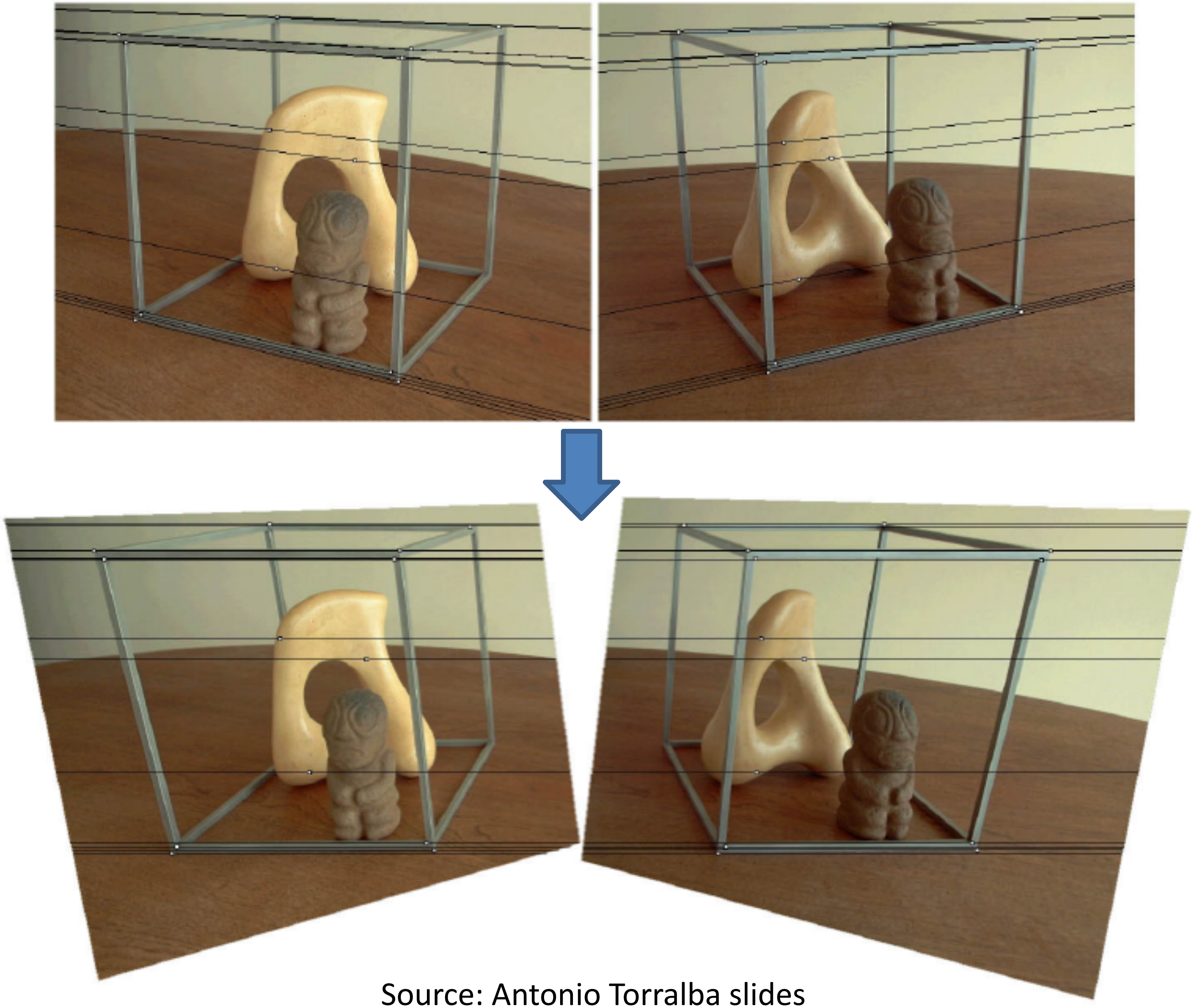
Slide credit: Kristen Grauman

Stereo image rectification



Source: Antonio Torralba slides

Stereo image rectification



Source: Antonio Torralba slides

Summary

- Introduction
- 3D shapes from 2D images
- Stereo vision
- Correspondence
- Epipolar geometry
- The fundamental matrix
- The essential matrix
- RANSAC
- 3D reconstruction