EECS 4422/5323 Computer Vision

Unit 5: Stereopsis

Outline

- Introduction
- The correspondence problem
- Epipolar geometry
- 3D reconstruction
- Empirical examples
- Summary

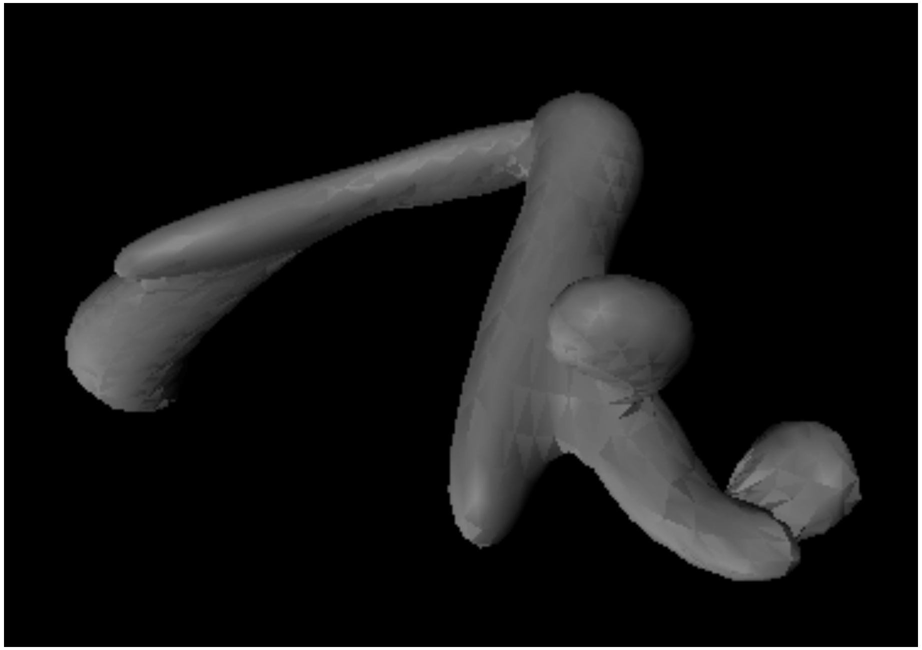
Introduction: 3D shape from 2D images

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.

Single image cues

- Perspective
- Contour
- Texture
- Aerial perspective
- Shading



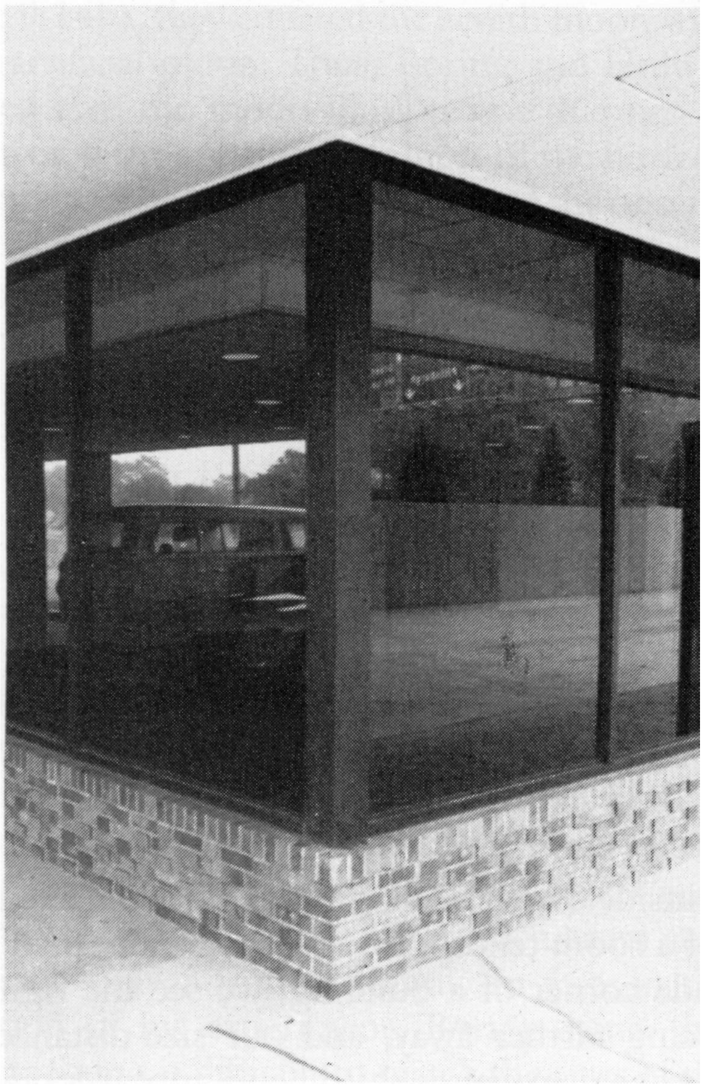
Introduction: 3D shape from 2D images

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.

Single image cues

- Perspective
- Contour
- Texture
- Aerial perspective
- Shading



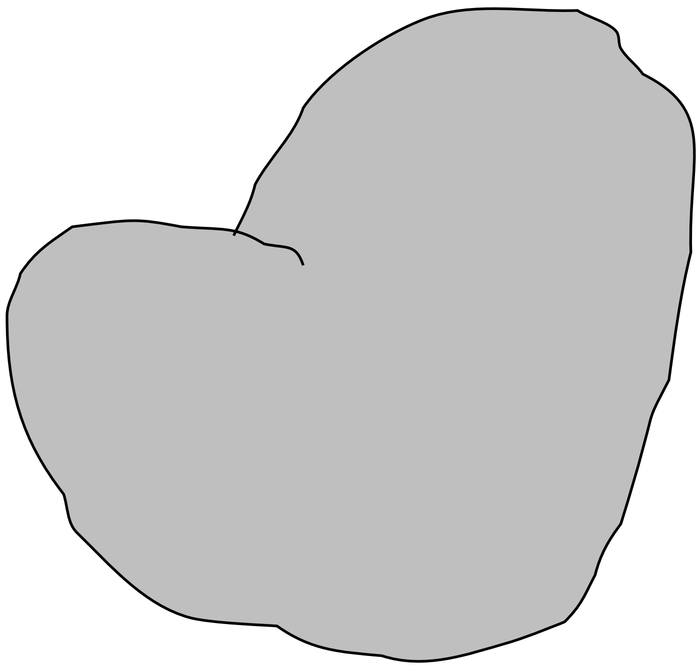
Introduction: 3D shape from 2D images

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.

Single image cues

- Perspective
- Contour
- Texture
- Aerial perspective
- Shading



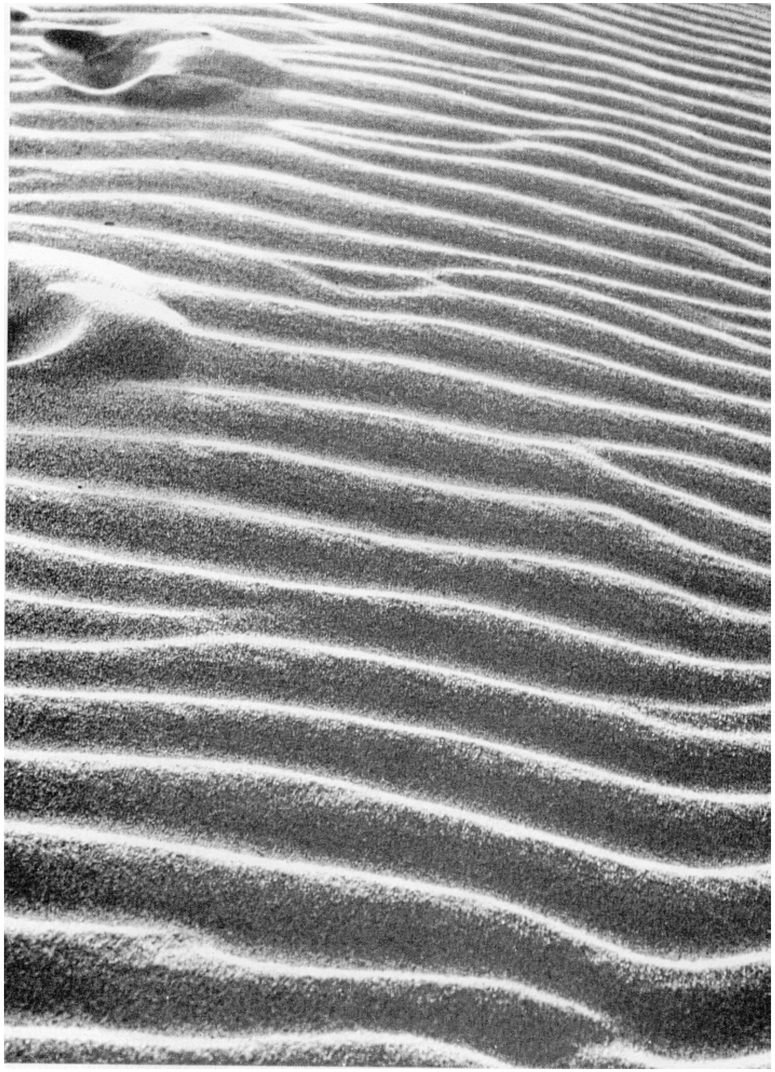
Introduction: 3D shape from 2D images

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.

Single image cues

- Perspective
- Contour
- Texture
- Aerial perspective
- Shading



Introduction: 3D shape 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.

Single image cues

- Perspective
- Contour
- Texture
- Aerial perspective
- Shading

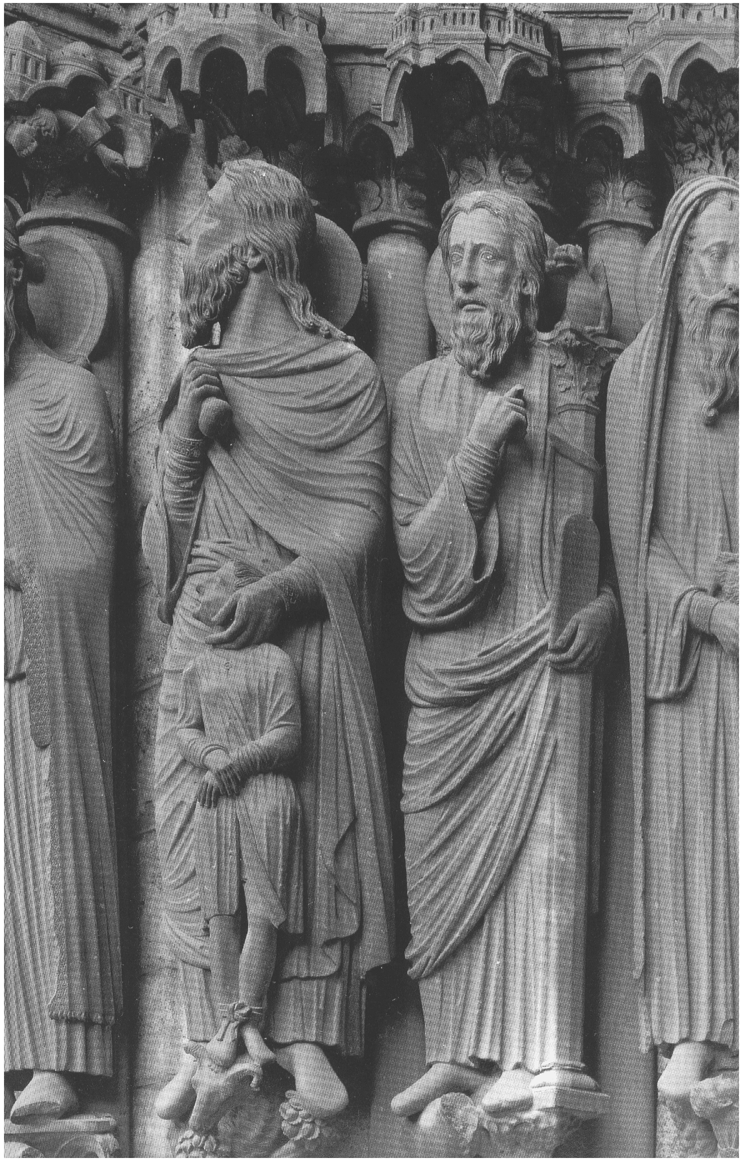
Introduction: 3D shape from 2D images

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.

Single image cues

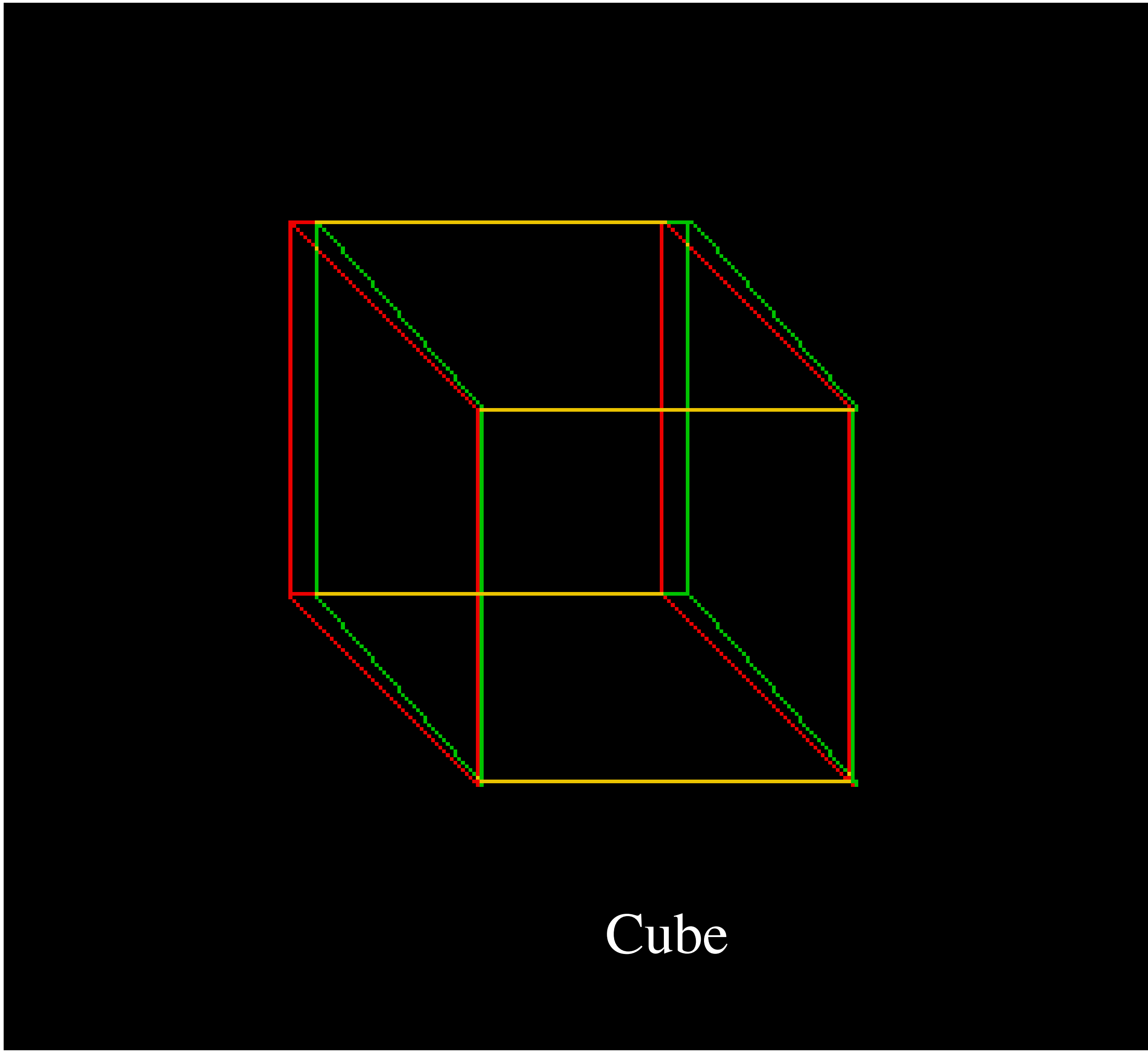
- Perspective
- Contour
- Texture
- Aerial perspective
- Shading



Introduction: Motivation

Definition

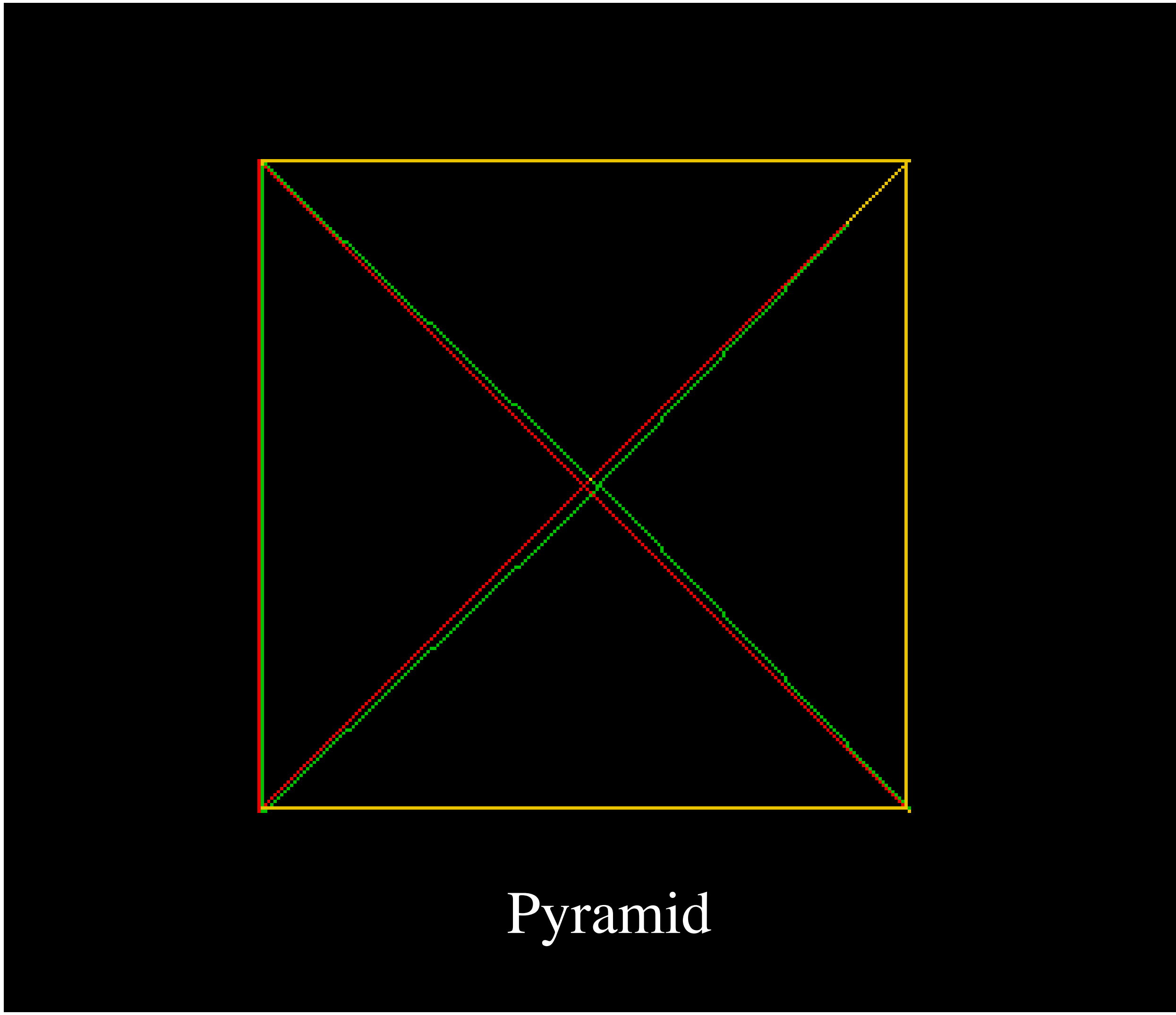
- The inference of information about the 3D structure of a scene from two or more images is referred to as stereo vision.
- Information of interest may take the form of
  - Distance measurements
  - Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities
- Minimally, two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



Introduction: Motivation

Definition

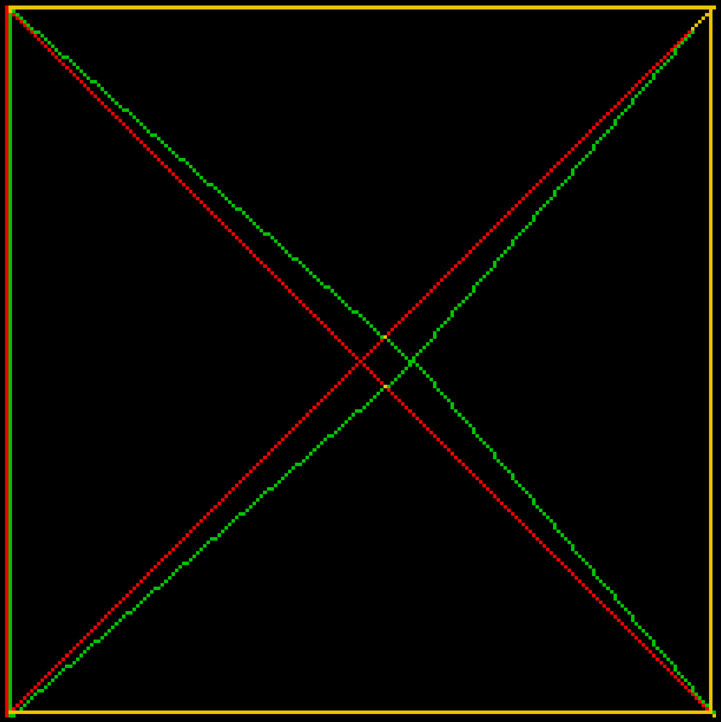
- The inference of information about the 3D structure of a scene from two or more images is referred to as stereo vision.
- Information of interest may take the form of
  - Distance measurements Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities
- Minimally, two or more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



Introduction: Motivation

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 Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities
- 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.
- 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.

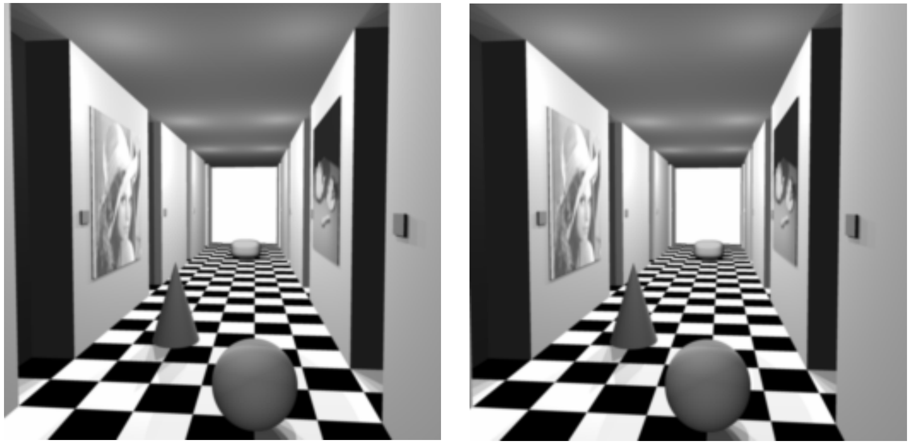


Pyramid (lots of depth)

Introduction: Motivation

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
- 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.
- 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.

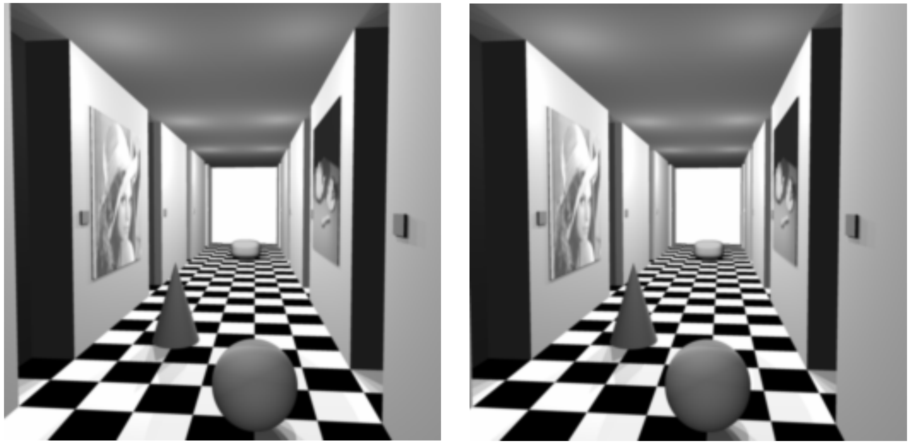


Stereo pair

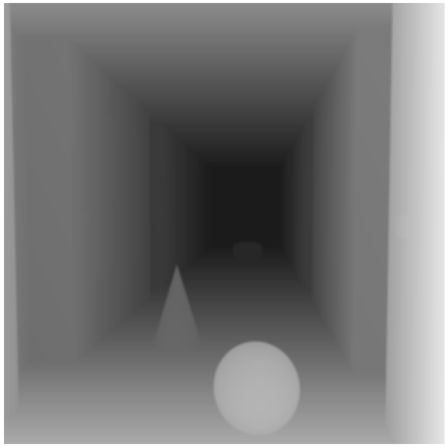
Introduction: Motivation

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 (range map)
  - Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities
- 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.
- 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

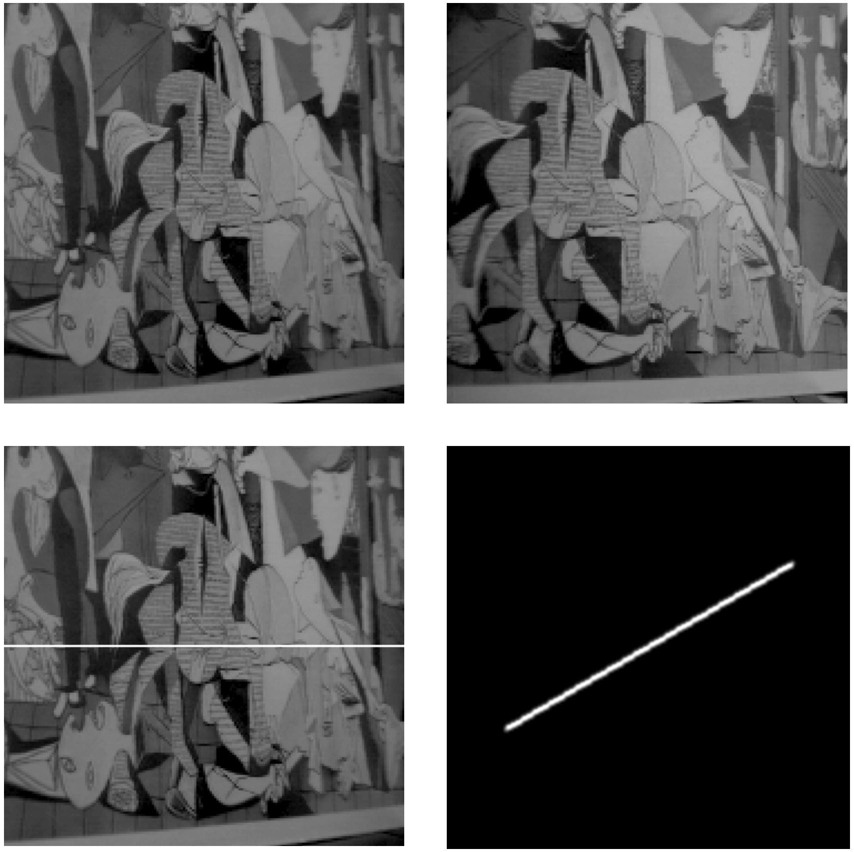


Range map

Introduction: Motivation

Definition

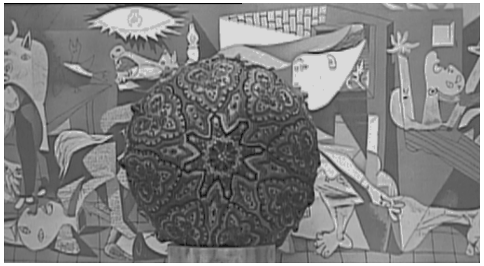
- The inference of information about the 3D structure of a scene from two or more images is referred to as stereo vision.
- Information of interest may take the form of
  - Distance measurements (range map)
  - Surface orientation
  - Surface curvature
  - Arrangement of surface discontinuities
- Minimally,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

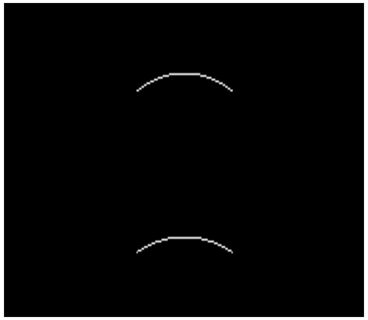


Introduction: Motivation

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
- 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.
- 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.

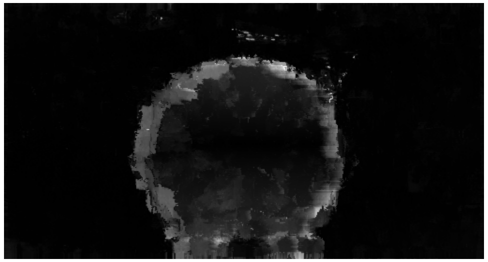




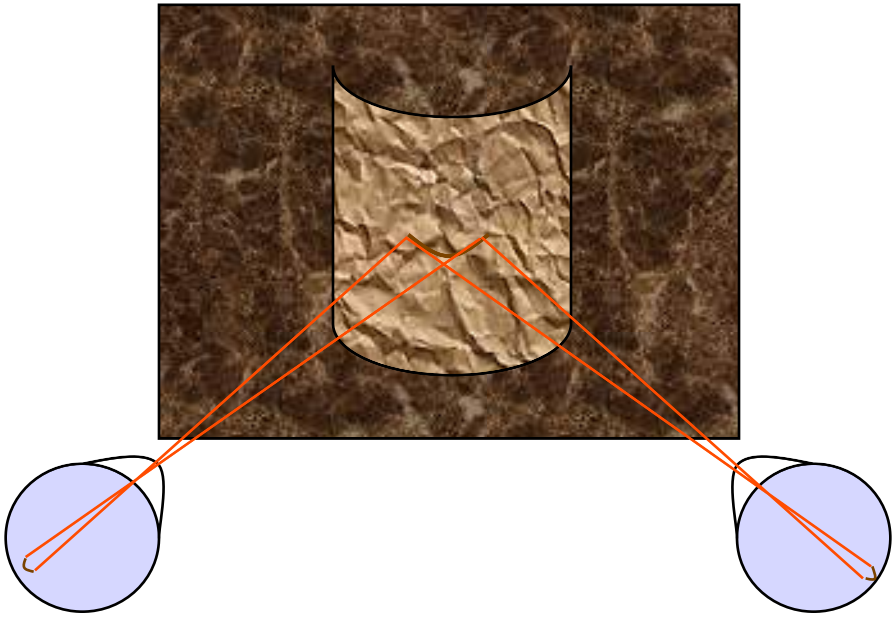
Introduction: Motivation

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
- 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.
- 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.



Introduction: Motivation



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
- 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.
- 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.

Introduction: The two problems of stereo

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.

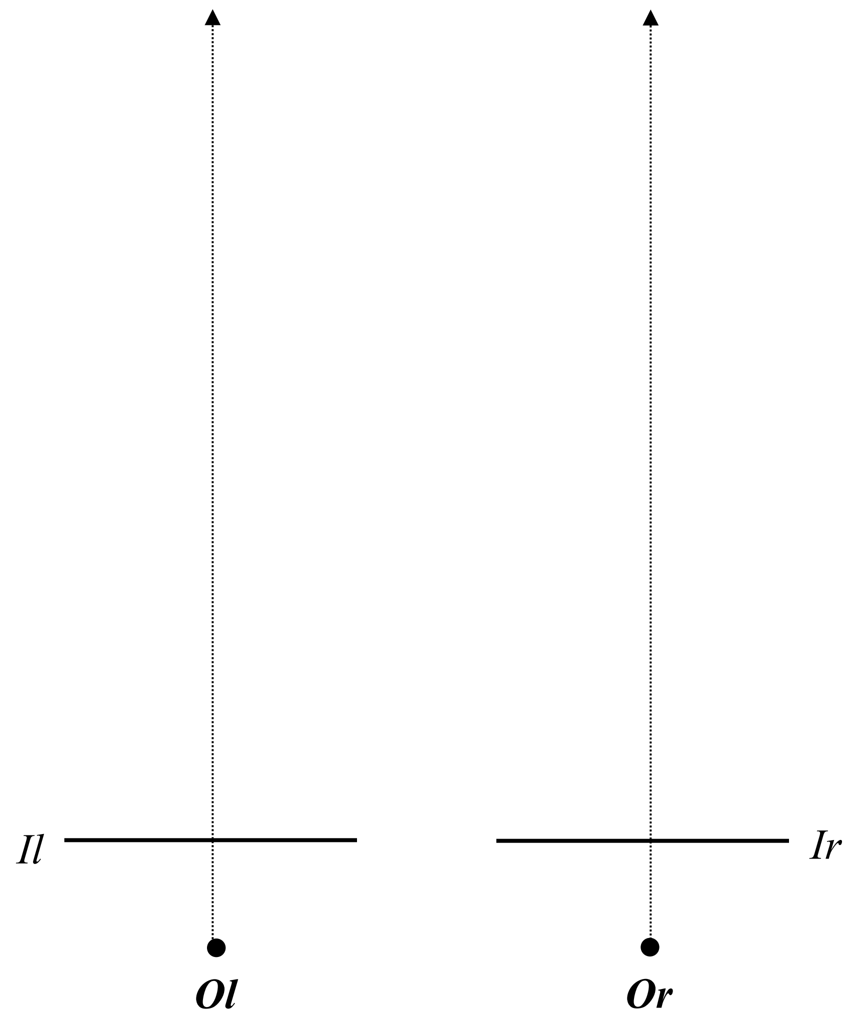
Reconstruction

- 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 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.

Introduction: 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



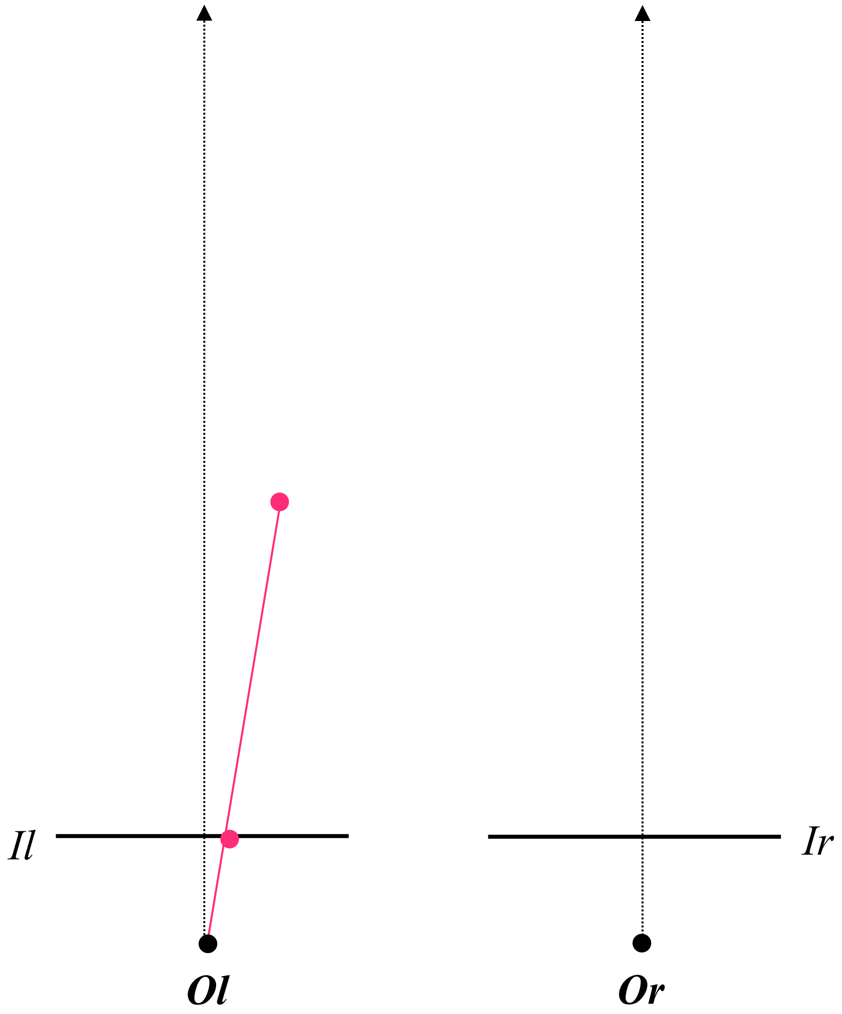
Introduction: A simple stereo system

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

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.



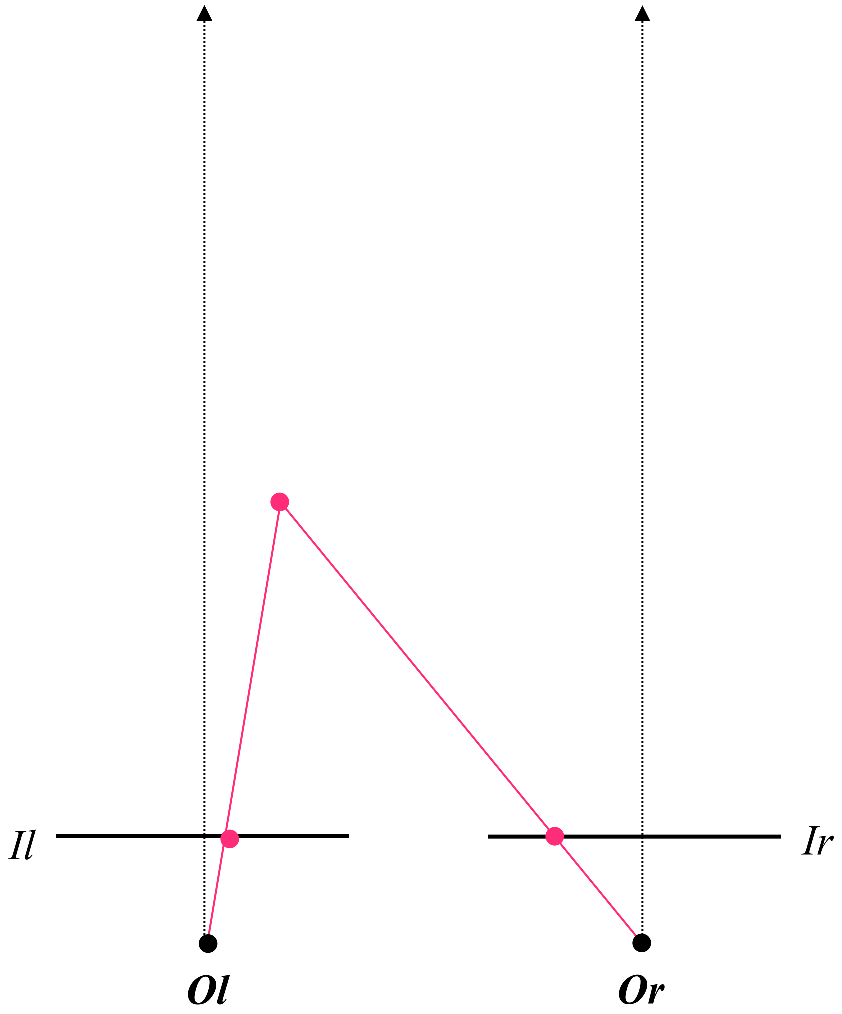
Introduction: A simple stereo system

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

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.



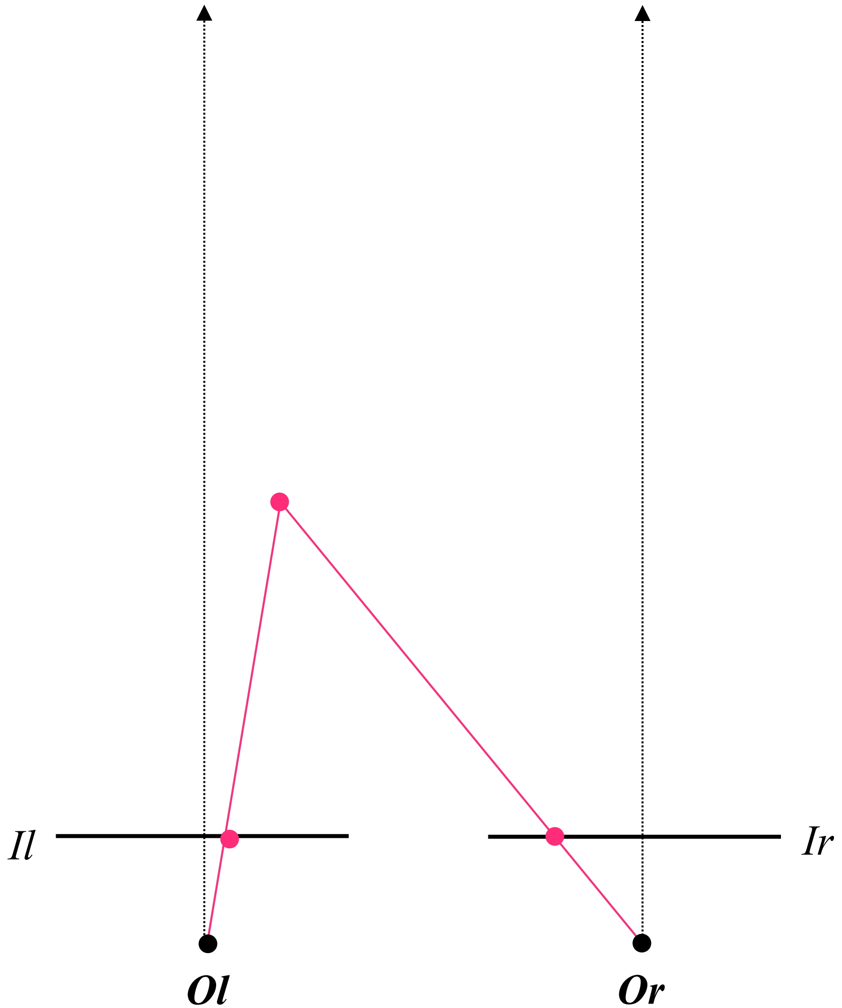
Introduction: A simple stereo system

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

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.



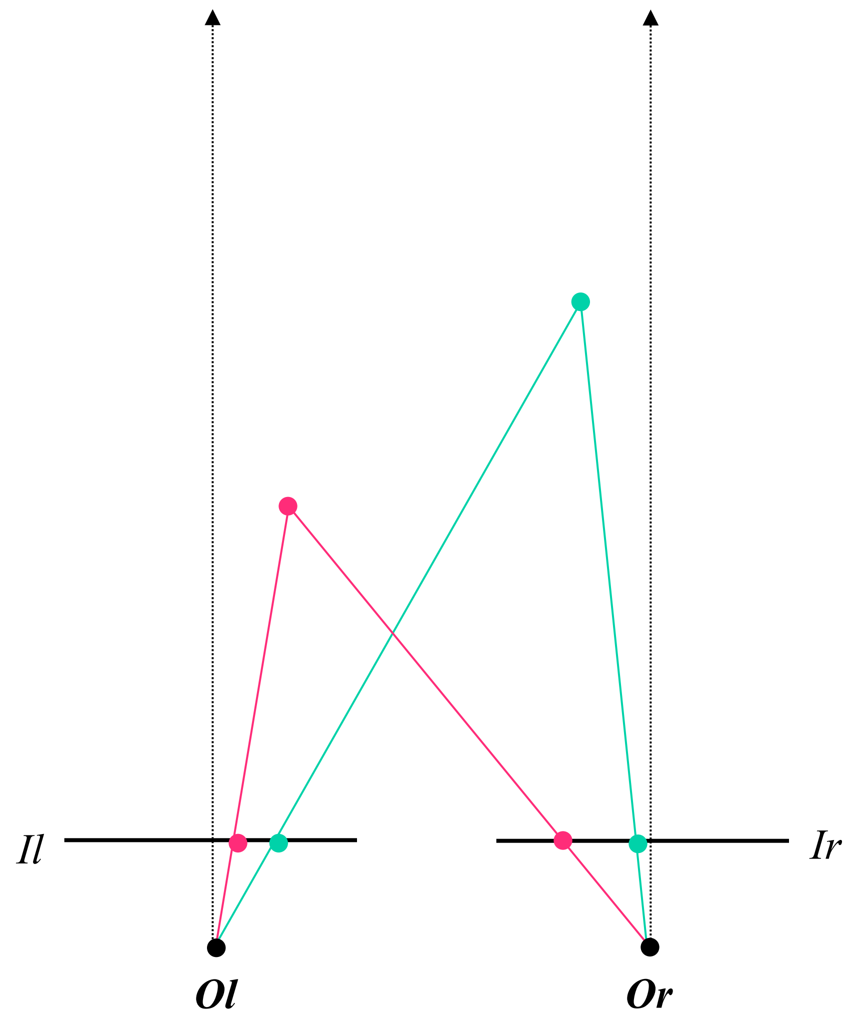
Introduction: A simple stereo system

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

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.



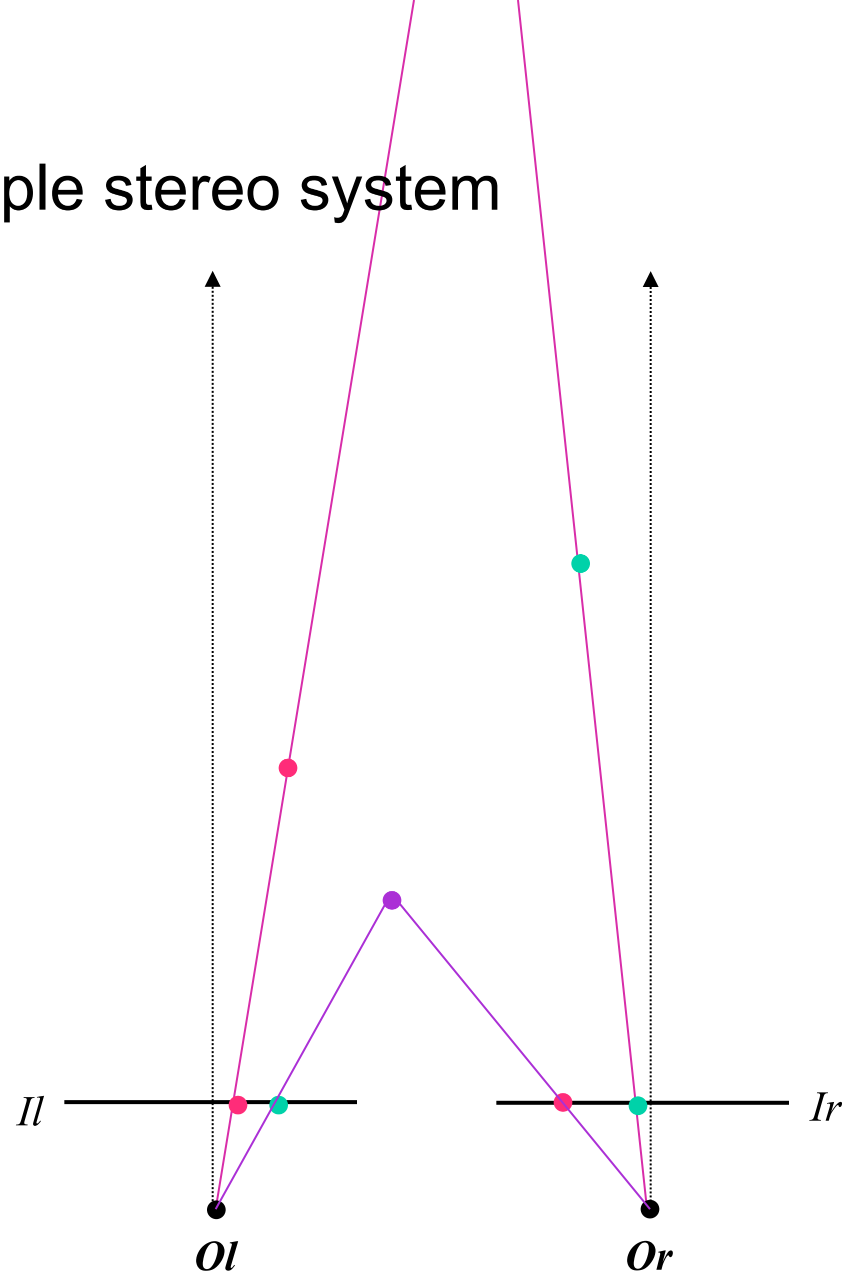
Introduction: A simple stereo system

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

Recovery of position in space

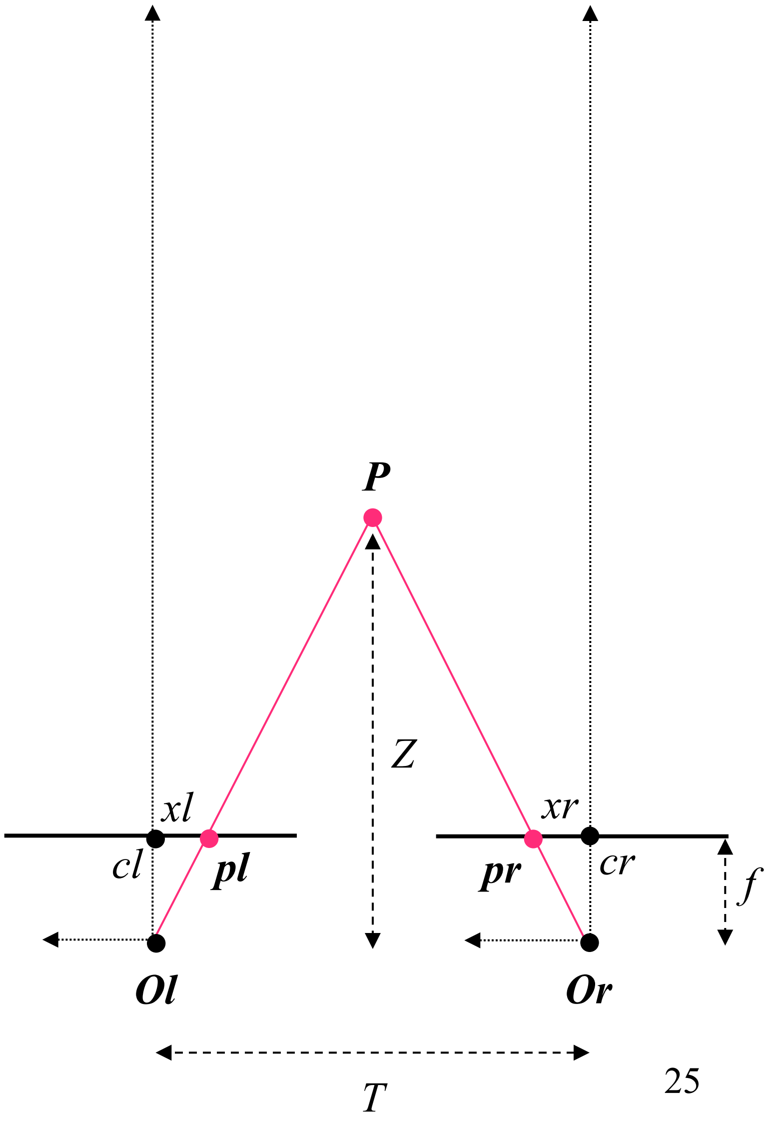
- Position in space is determined via the intersection of rays
  - Defined by the centres of projection
  - And the left and right images of a point of concern.
- A process known as triangulation.
- Triangulation depends critically on correspondence.



Introduction: A simple stereo system

Equations of triangulation

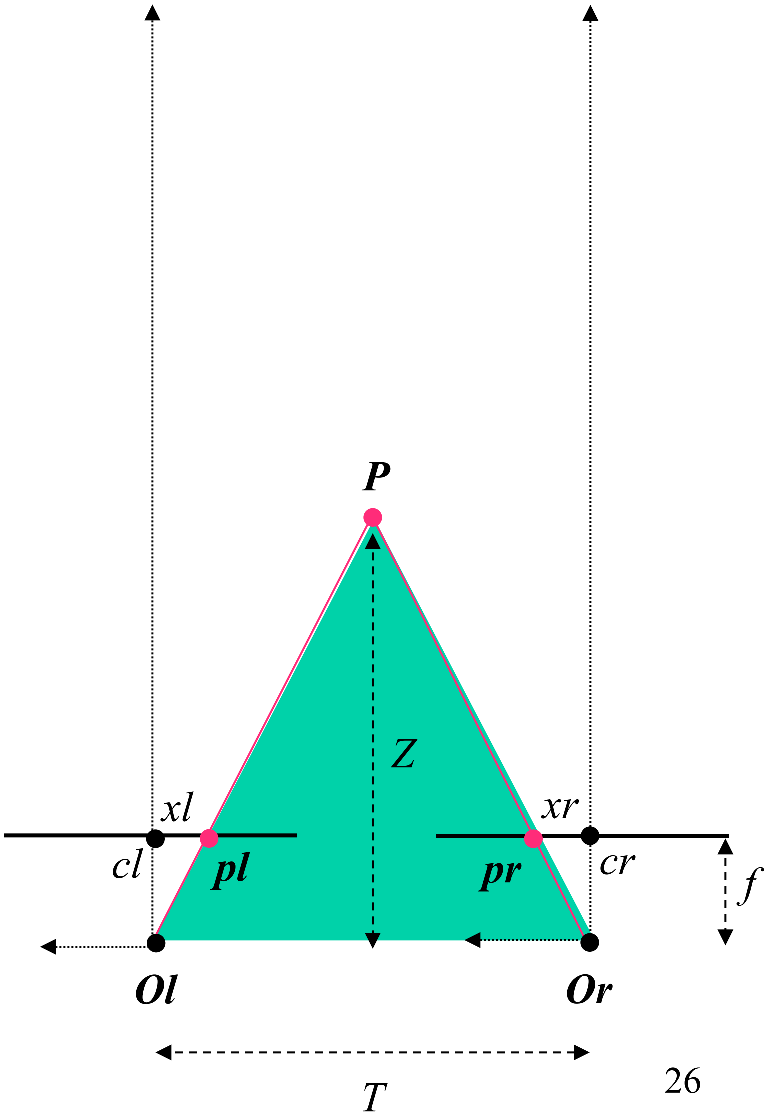
- Consider a point ***P*** and its projections ***pl*** and ***pr***.
- Let
  - T be the distance between the centres of projection, the baseline.
  - cl and cr be the centre points of the left and right images, respectively
  - xl and xr be the coordinates of ***pl*** and ***pr***, respectively.
  -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



Introduction: A simple stereo system

Equations of triangulation

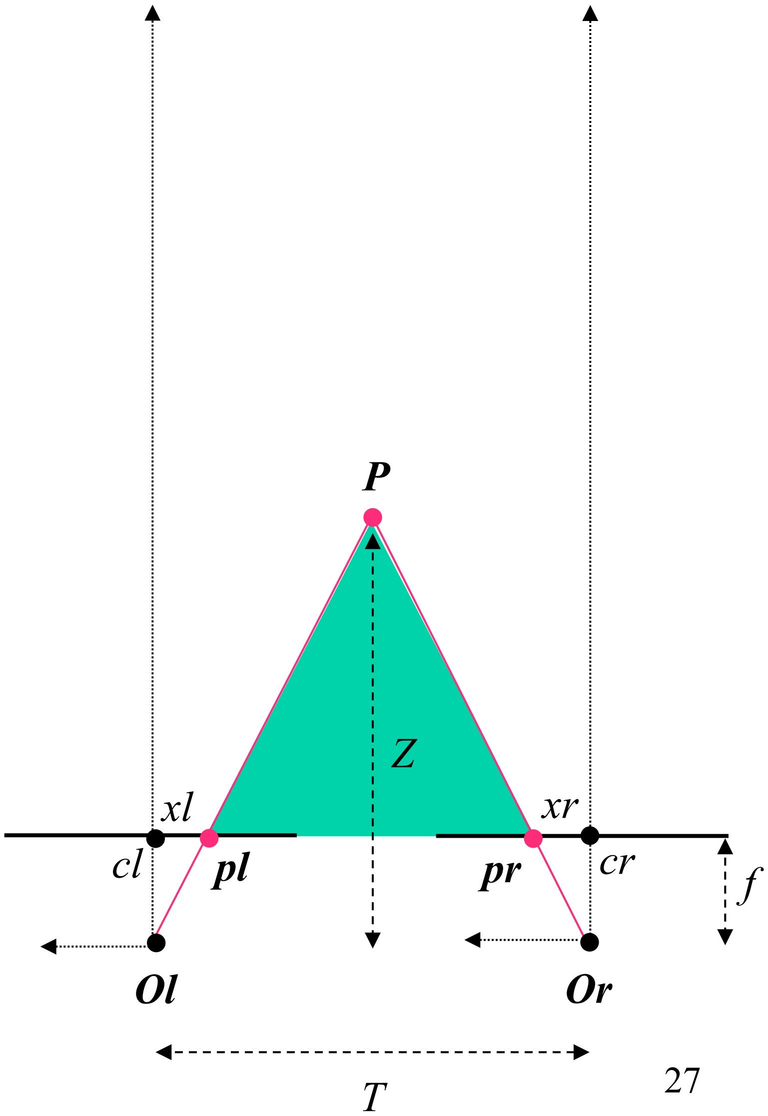
- Consider a point ***P*** and its projections ***pl*** and ***pr***.
- Let
  - *T* be the distance between the centres of projection, the baseline.
  - *cl* and *cr* be the centre points of the left and right images, respectively
  - *xl* and *xr* be the coordinates of ***pl*** and ***pr***, respectively.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



Introduction: A simple stereo system

Equations of triangulation

- Consider a point P and its projections pl and pr.
- Let
  - 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
- Similar triangles (pl,P,pr) and(Ol,P,Or)



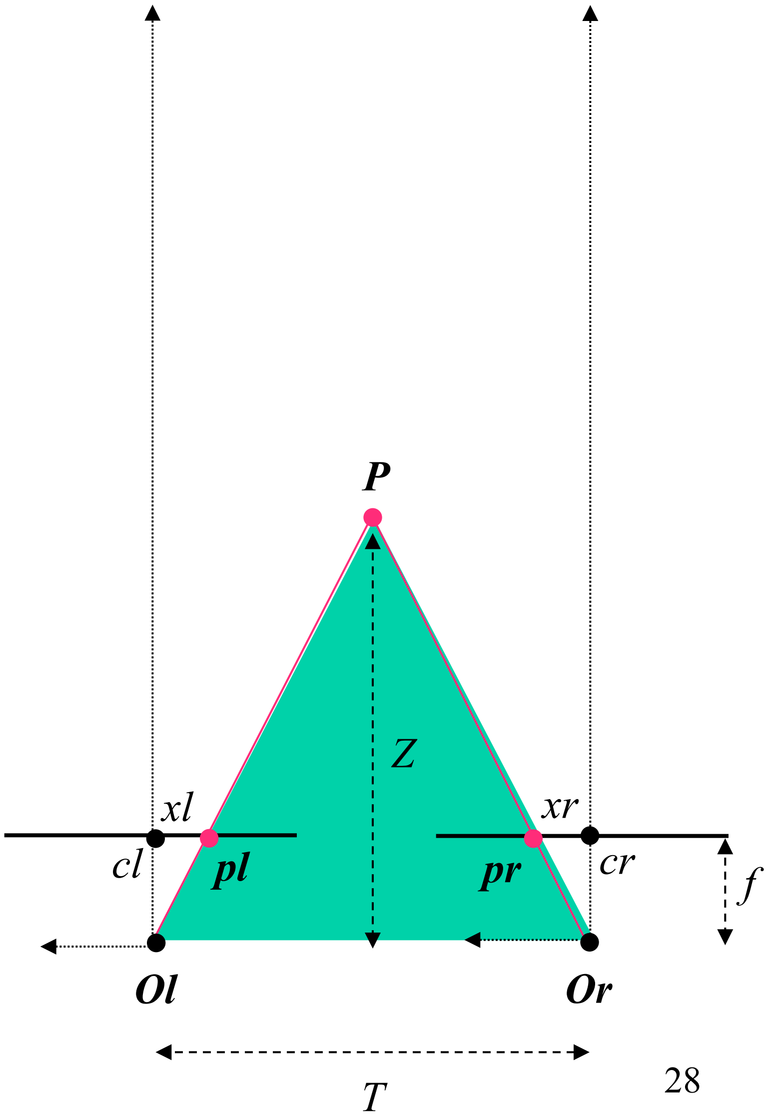
Introduction: A simple stereo system

Equations of triangulation

- Consider a point ***P*** and its projections ***pl*** and ***pr***.
- Let
  - 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
- Similar triangles (***pl,P,pr***) and(***Ol,P,Or***)

  Allow us to write

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



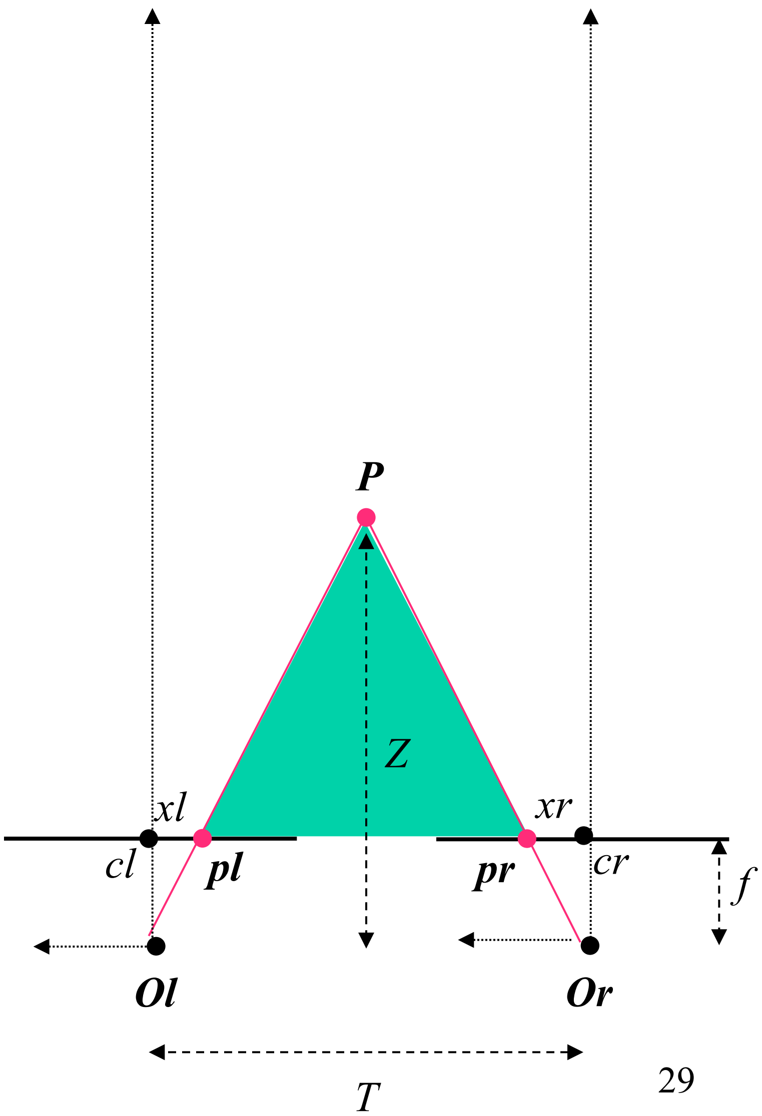
Introduction: A simple stereo system

Equations of triangulation

- Consider a point ***P*** and its projections ***pl*** and ***pr***.
- Let
  - 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
- Similar triangles (***pl,P,pr***) and(***Ol,P,Or***)

  Allow us to write

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



Introduction: A simple stereo system

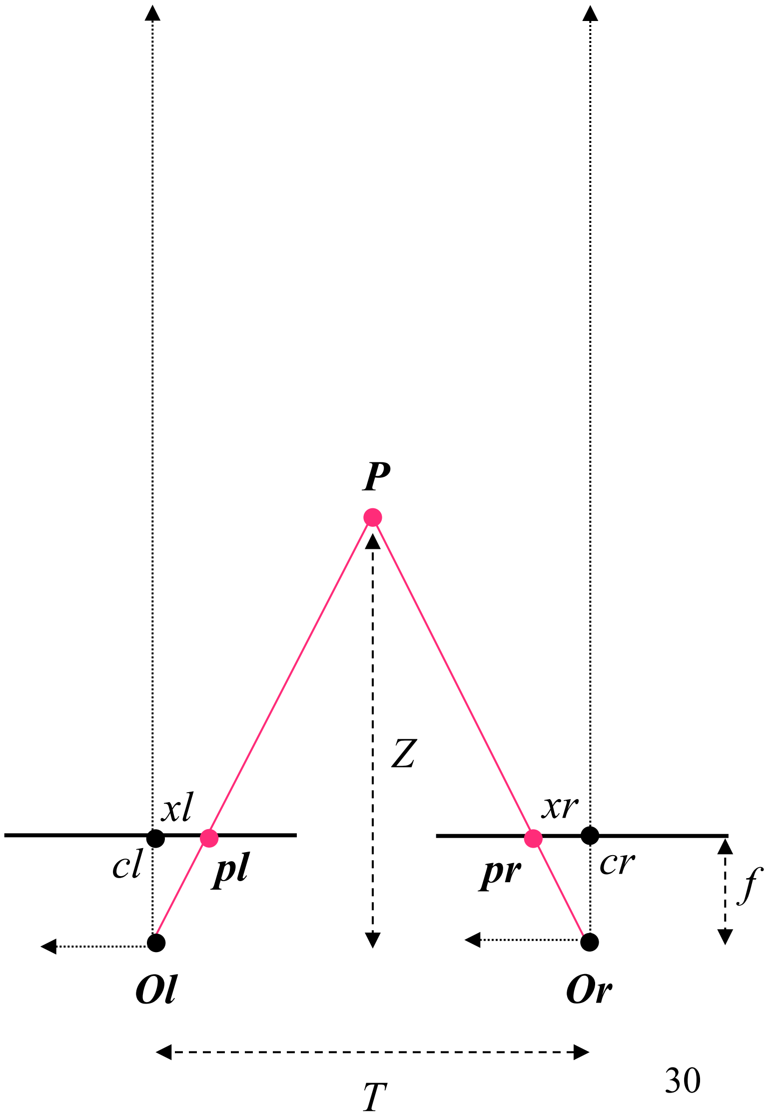
Equations of triangulation

- Consider a point ***P*** and its projections ***pl*** and ***pr***.
- Let
  - 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
- Similar triangles (***pl,P,pr***) and(***Ol,P,Or***)
  Allow us to write

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

- Letting $d = xr\text{-}xl$ be the disparity, we solve for Z as

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



Introduction: The 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.
- Note: In the terminology of photogrammetry we speak of interior orientation.

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.
- Note: In the terminology of photogrammetry, the geometry relating one camera to another is called relative orientation; a separate parameterization (absolute or exterior orientation) would relate the camera (pair) to the world.

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.
  - That is, disparity decreases as we move toward infinity, the effective convergence of the optical axes.
  - More generally, disparity magnitude decreases with closeness to the fixation point, the convergence of the optical axes.

Outline

- Introduction
- **The correspondence problem**
- Epipolar geometry
- 3-D reconstruction
- Empirical examples
- Summary

Correspondence: Basics

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

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

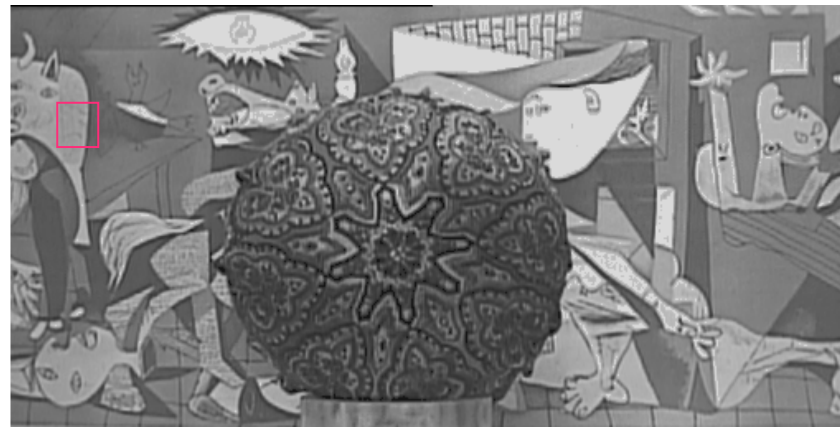
Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.



Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Correspondence: Area-based

Motivation

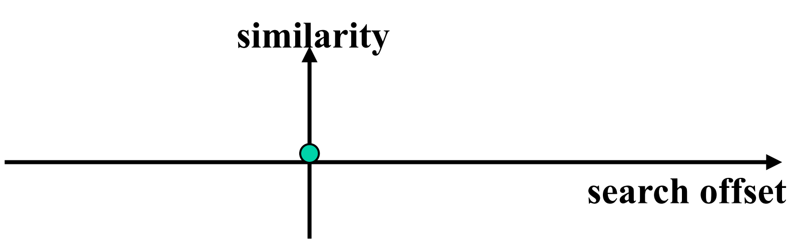
- Exploit all available information

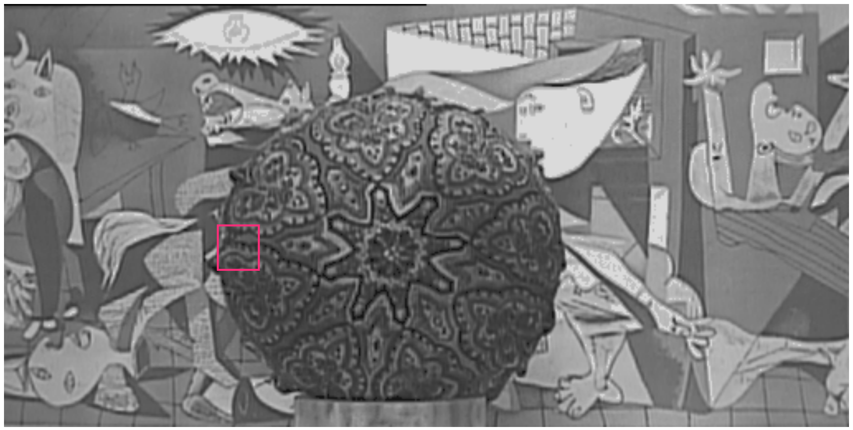
Elements to be matched

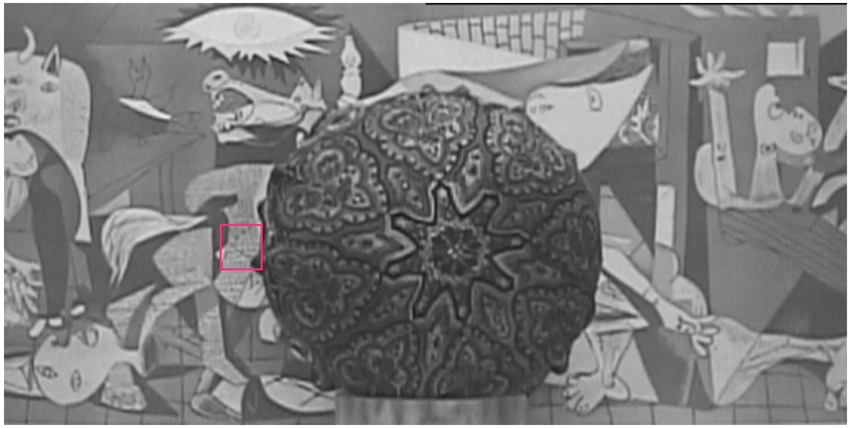
- Image windows
- Typically of fixed size
- Spatially overlapping.

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

Motivation

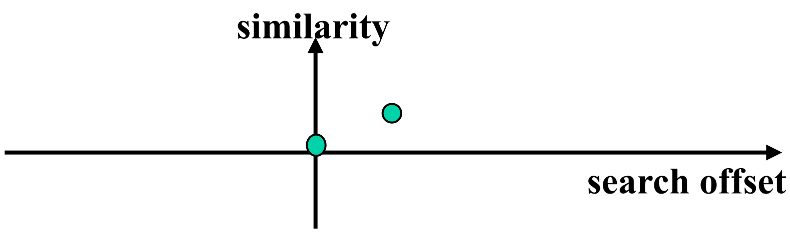
- Exploit all available information

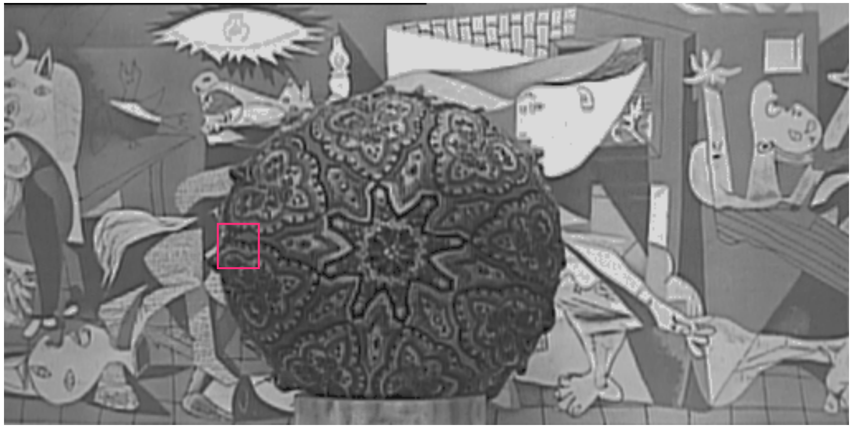
Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

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

Motivation

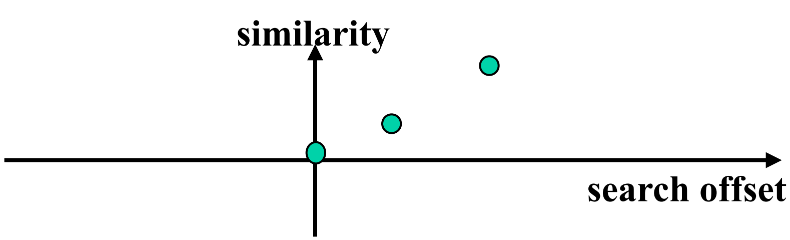
- Exploit all available information

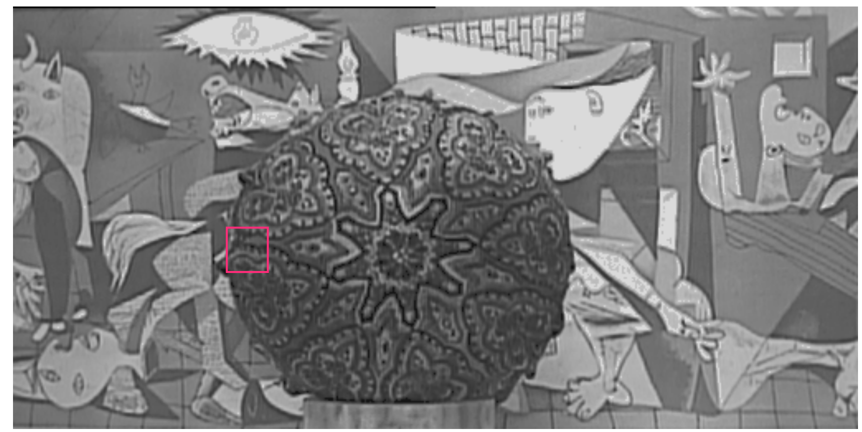
Elements to be matched

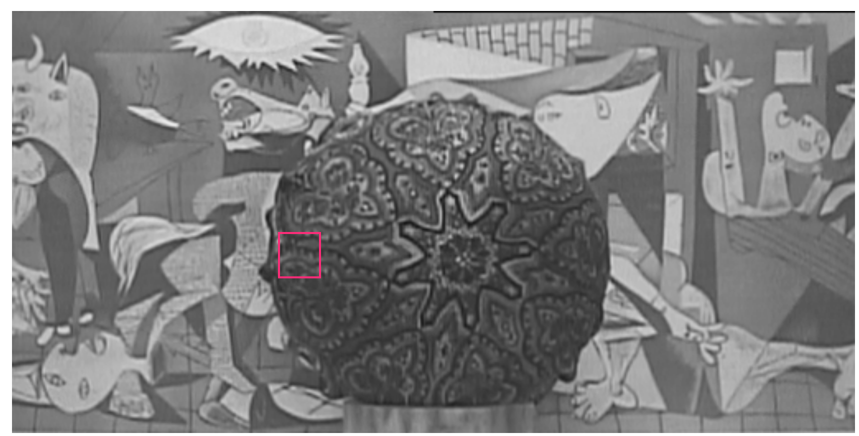
- Image windows
- Typically of fixed size
- Spatially overlapping.

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

Motivation

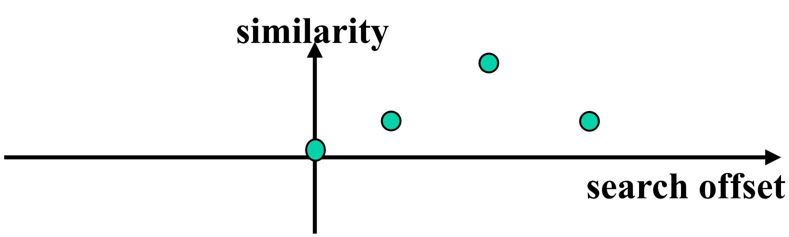
- Exploit all available information

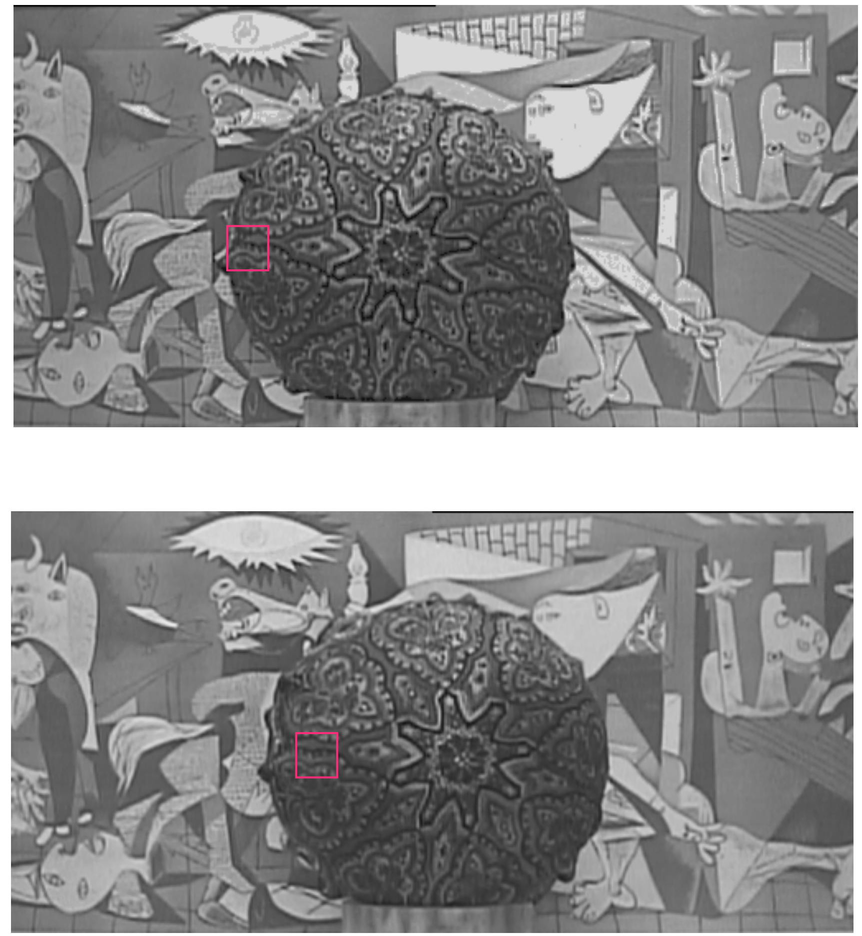
Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

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

Motivation

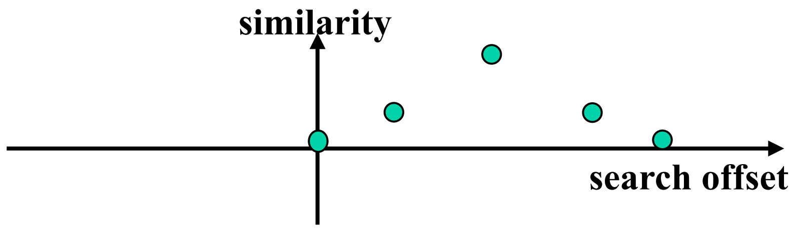
- Exploit all available information

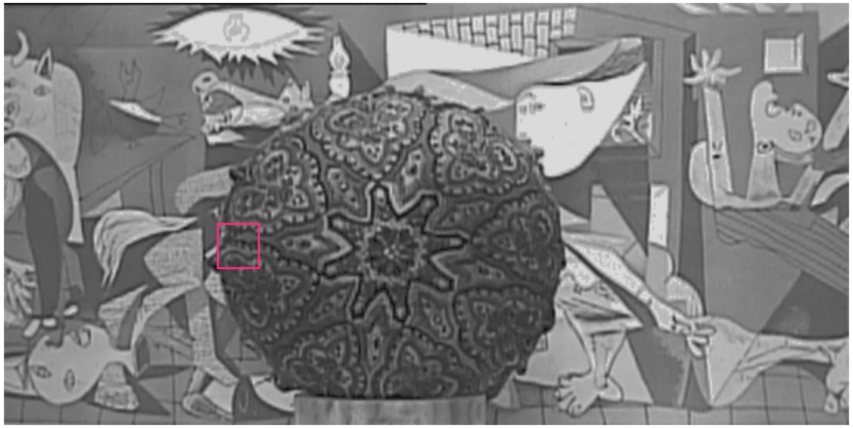
Elements to be matched

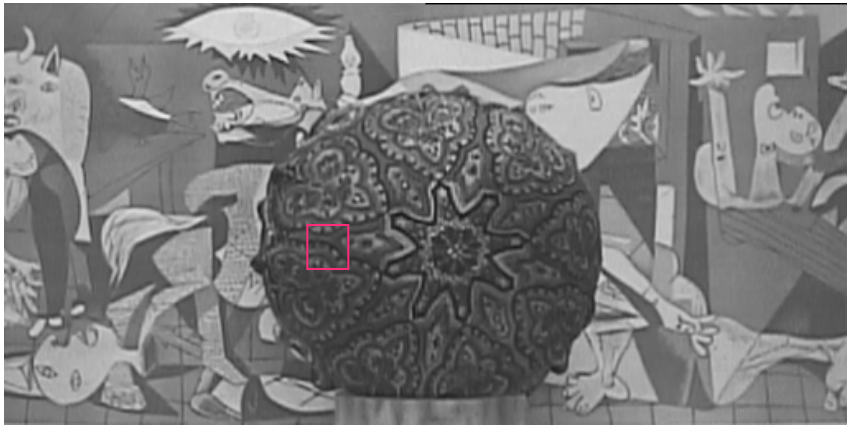
- Image windows
- Typically of fixed size
- Spatially overlapping.

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

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Similarity measure

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

Formalization: For disparity d

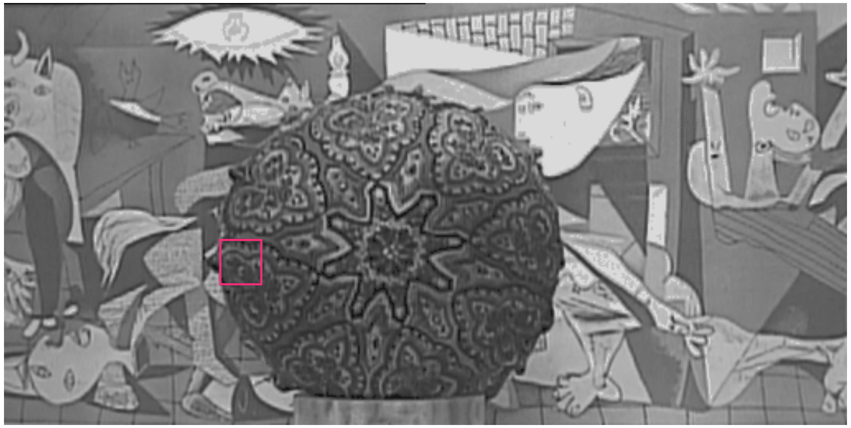
- *Local matching* seeks:

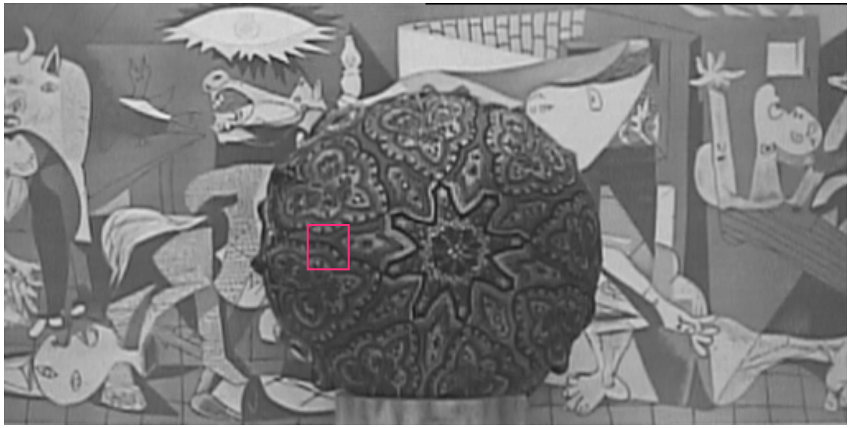
  For each *(x,y)*

$$\max_{d}[match(d)]$$

- *Global matching* seeks:

$$\max_{d(x,y)}\{\sum_{(x,y)} match[d(x,y)]\}$$





Correspondence: Area-based

Motivation

- Exploit all available information

Elements to be matched

- Image windows
- Typically of fixed size
- Spatially overlapping.

Similarity measure

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

Formalization: For disparity d

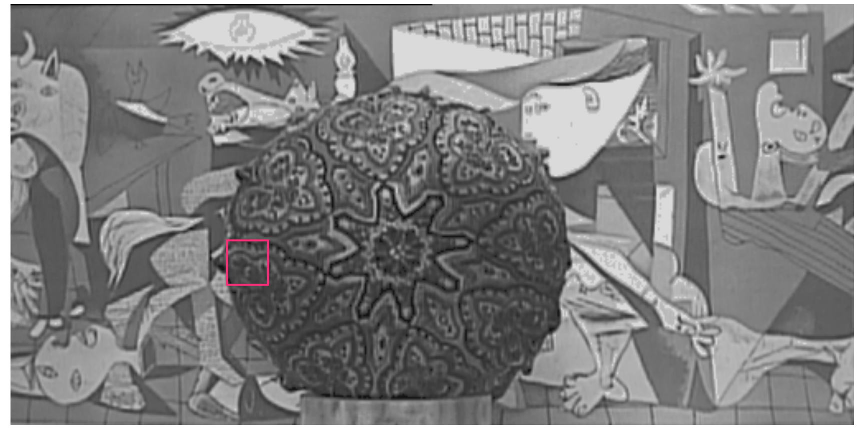
- *Local matching* seeks:

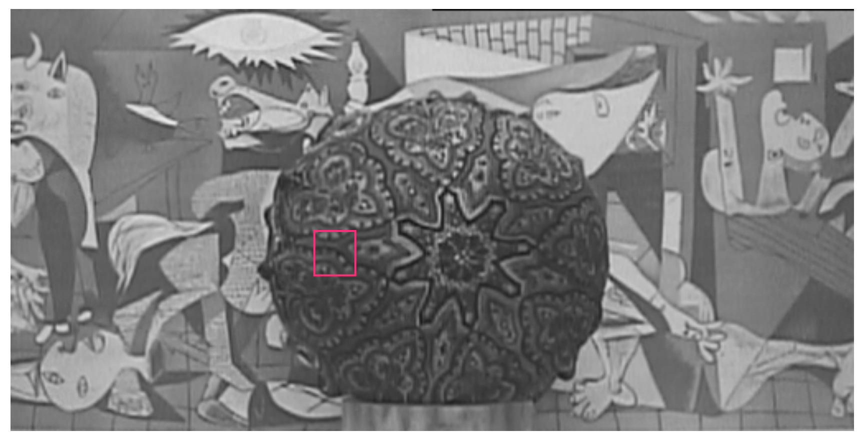
  For each *(x,y)*

  $$\max_d [match(d)]$$

- *Global matching* seeks:

  $$\max_{d(x,y)} \{ \sum_{(x,y)} match[d(x,y)] \}$$





Correspondence: Area-based

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: Area-based

Algorithm

- **Input:** Stereo pair of images *Il* and *Ir*
- **Output:** Array of disparities (disparity map), one value for each pixel of *Il*.
- **Notation:** Let
  - ***pl*** and ***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 ***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 ***pl*** is the vector ***d*** that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Remarks: We have

- (i,j) as a particular image location in the left image about which the match window currently is defined.

Correspondence: Area-based

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: We have

- (i,j) as a particular image location in the left image about which the match window currently is defined.
- (k,l), $-W<=k,l<=W$ defining the domain of the window centred about (i,j).

Correspondence: Area-based

Algorithm

- **Input:** Stereo pair of images Il and Ir
- **Output:** Array of disparities (disparity map), one value for each pixel of Il.
- **Notation:** Let
  - ***pl*** and ***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 ***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 ***pl*** is the vector ***d*** that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Remarks: We have

- (i,j) as a particular image location in the left image about which the match window currently is defined.
- (k,l), $-W<=k,l<=W$ defining the domain of the window centred about (i,j).
- $(d1,d2)$ as the shift in the right image at which we currently evaluate the match.

Correspondence: Area-based

Algorithm

- **Input:** Stereo pair of images Il and Ir
- **Output:** Array of disparities (disparity map), one value for each pixel of Il.
- **Notation:** Let
  - ***pl*** and ***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 ***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 ***pl*** is the vector ***d*** that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Correspondence: Area-based

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 minimizes (maximizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Standard choices for m(u,v)

- The squared difference $m(u,v)= (u-v)^2$ yields the Sum of Squared Differences (SSD).
- The product $m(u,v)=uv$ yields cross-correlation.

Correspondence: Area-based

Match measures: Sum of squared differences (SSD)

- Let $m(u,v) = (u-v)$^2

$$c(\boldsymbol{d}) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} m[Il(i+k, j+l), Ir(i+k-d1, j+l-d2)]$$
$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} [Il(i+k, j+l) - Ir(i+k-d1, j+l-d2)]^2$$

- We quantify the goodness of match by
  1. Taking the pixelwise difference of brightnesses between the two images, Il and Ir, within the match window defined by W.
  2. Squaring the difference (because we only care about the magnitude of the discrepency).
  3. Summing over the window.
- In the end, better matches are defined as having smaller SSDs.

Correspondence: Area-based

Match measures: From SSD to correlation

- Let us expand the square inside the summation

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

$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} [Il(i+k, j+l) - Ir(i+k-d1, j+l-d2)]^2$$

$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} [Il(i+k, j+l)^2 - 2Il(i+k, j+l)Ir(i+k-d1, j+l-d2) + Ir(i+k-d1, j+l-d2)^2]$$

- Because they do not depend on the interaction of the two images, we neglect the first and last terms inside the summations and restrict consideration to

$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} [-2Il(i+k, j+l)Ir(i+k-d1, j+l-d2)]$$

- Apparently the portions of the images within the match windows are most similar when

$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} Il(i+k, j+l)Ir(i+k-d1, j+l-d2)$$

is maximized.

Correspondence: Area-based

Match measures: Correlation

- Let $m(u,v) = (uv)$

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

$$= \sum_{k=-W}^{W} \sum_{l=-W}^{W} Il(i+k, j+l) Ir(i+k-d1, j+l-d2)$$

- We quantify the goodness of match by
  1. Taking the pixelwise product between the two images, Il and Ir, within the match window defined by W.
  2. Summing over the window.
- In the end, better matches are defined as having larger correlations.

Correspondence: Area-based

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 minimizes (maximizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

Standard choices for m(u,v)

- The squared difference $m(u,v)= (u-v)^2$ yields the Sum of Squared Differences (SSD).
- The product $m(u,v)=uv$ yields cross-correlation.
- Remark: Although more computationally expensive, SSD can be superior due to its being less biased by large or small image intensity values.

Correspondence: Area-based

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

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

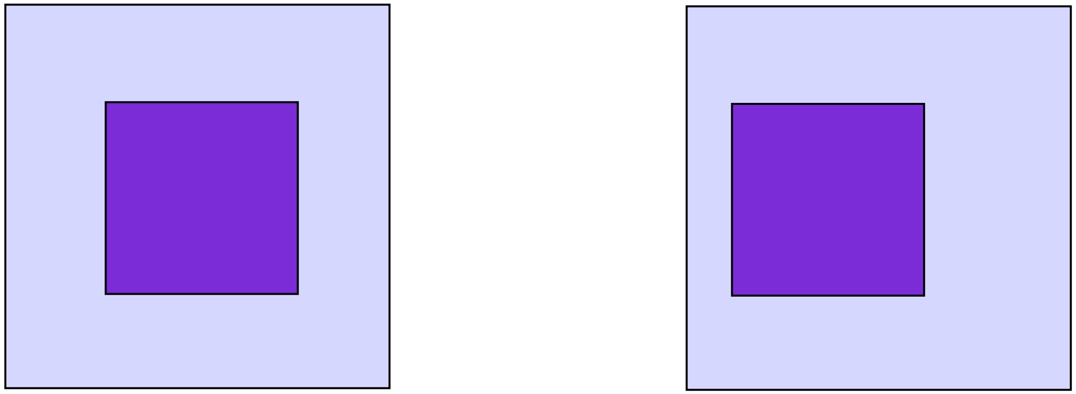
2. The disparity of $\boldsymbol{pl}$ is the vector $\boldsymbol{d}$ that maximizes (minimizes) $c(\boldsymbol{d})$ over $R(\boldsymbol{pl})$.

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.

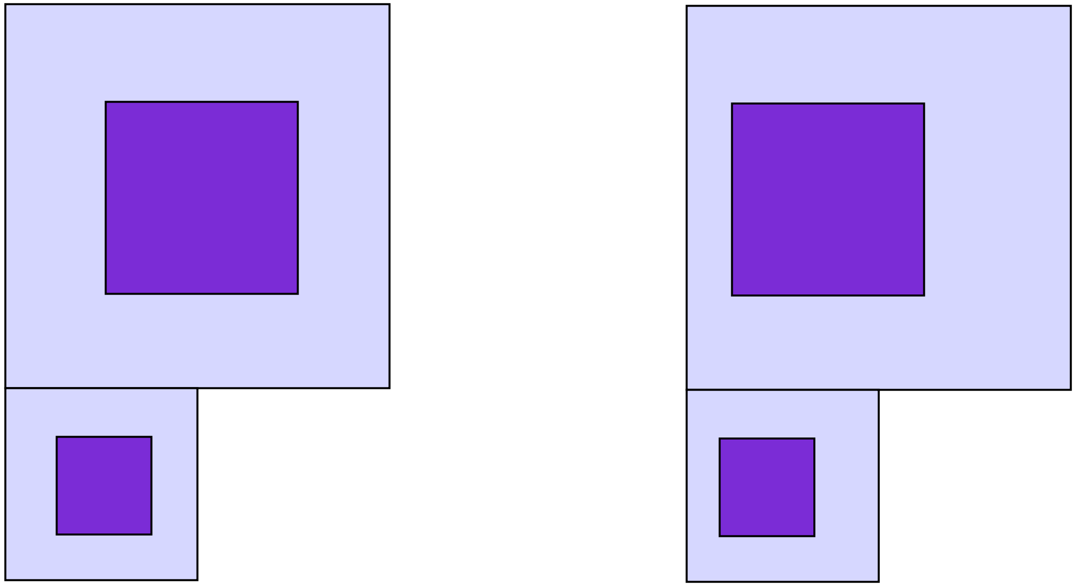
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



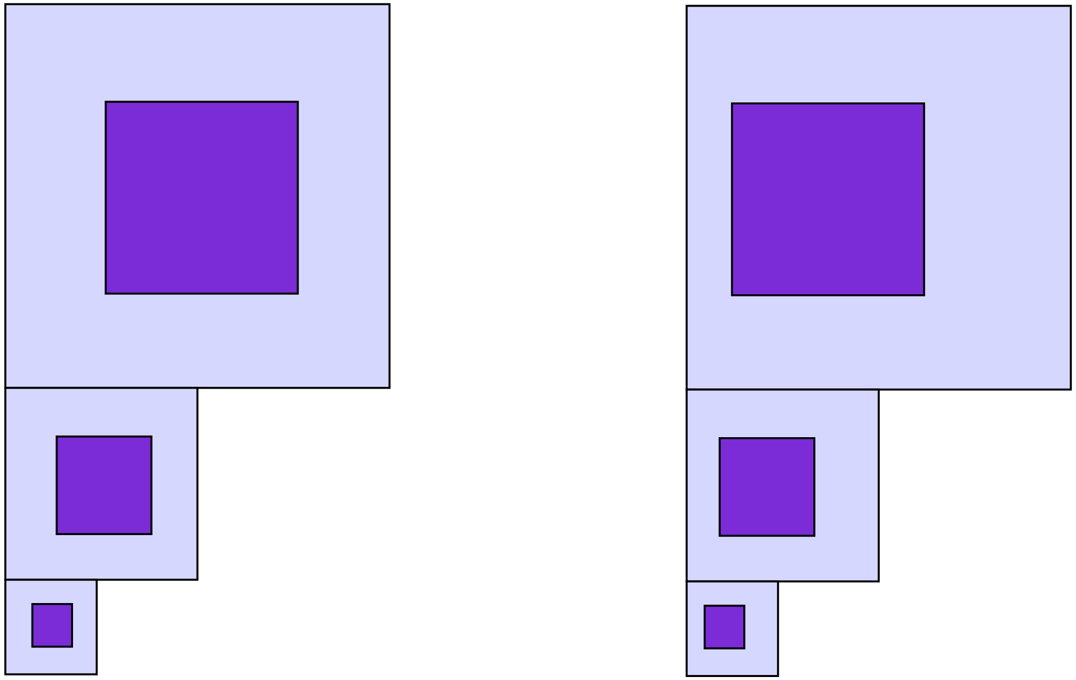
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



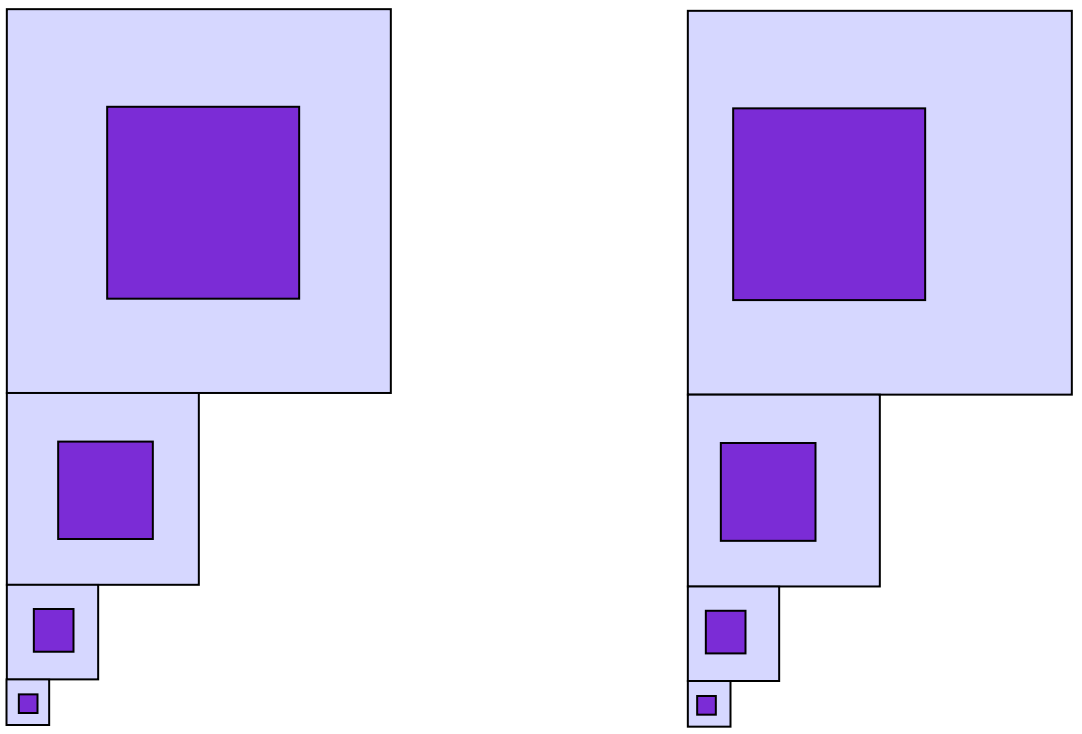
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



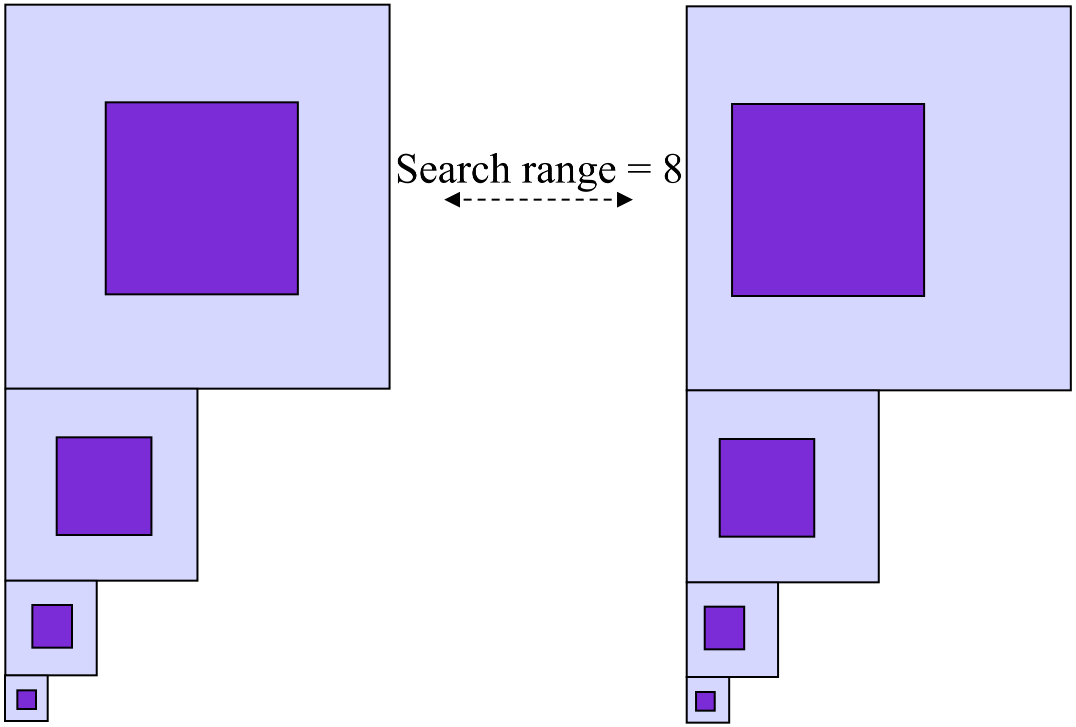
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



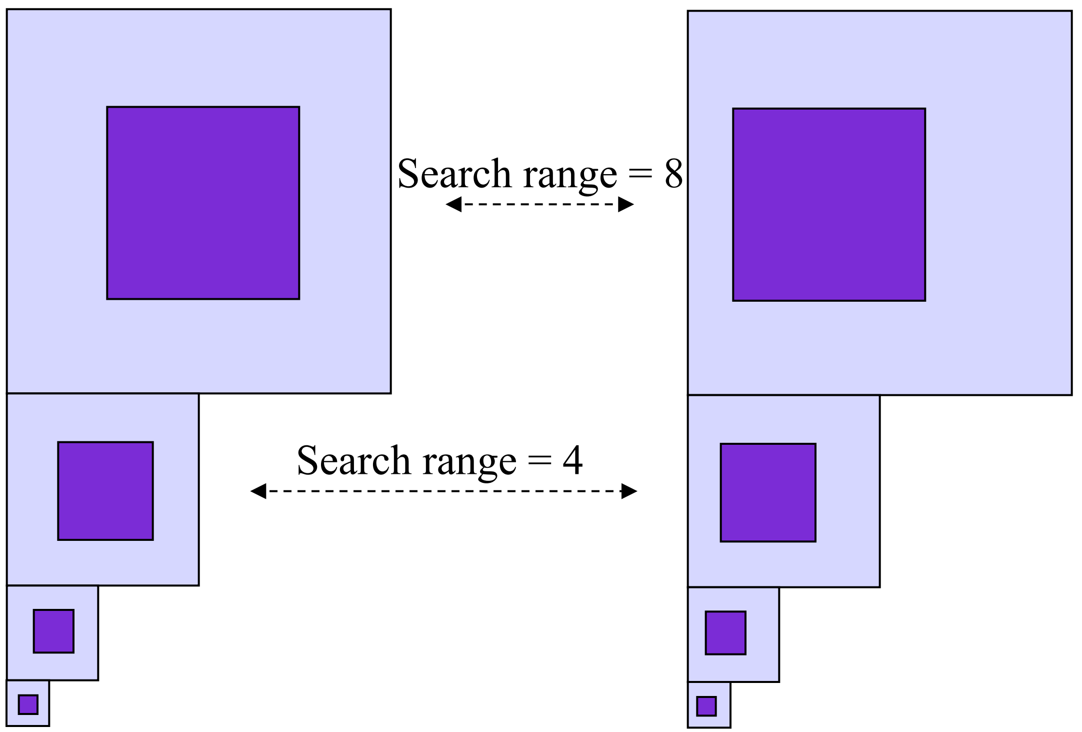
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



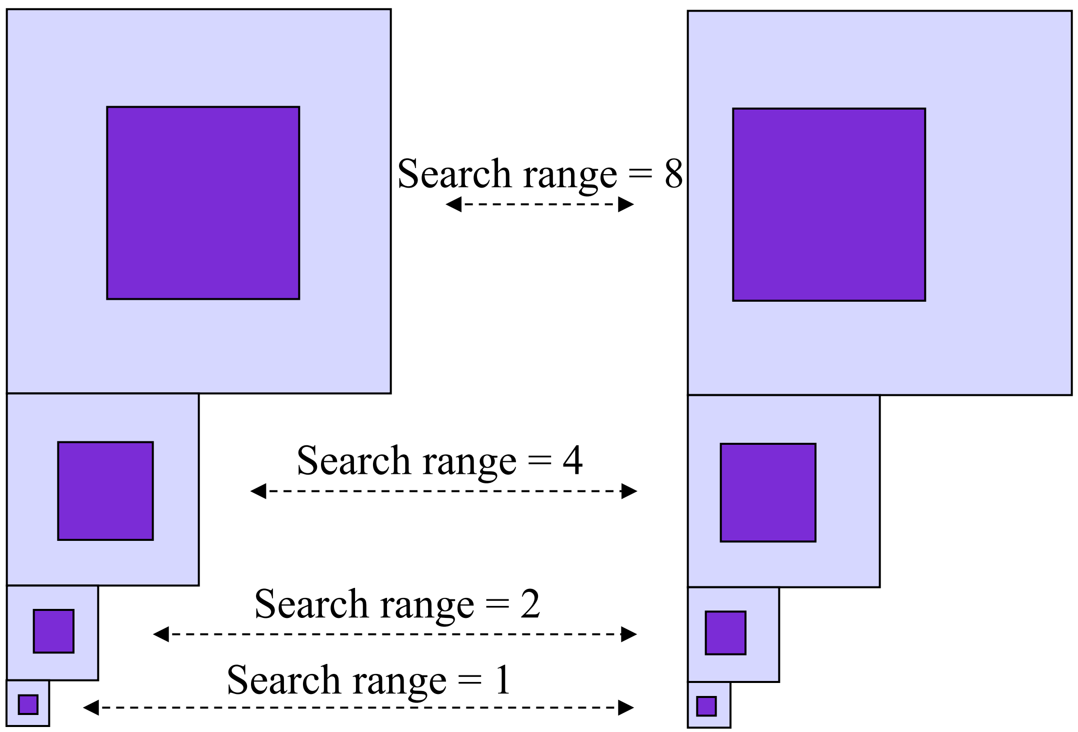
Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence

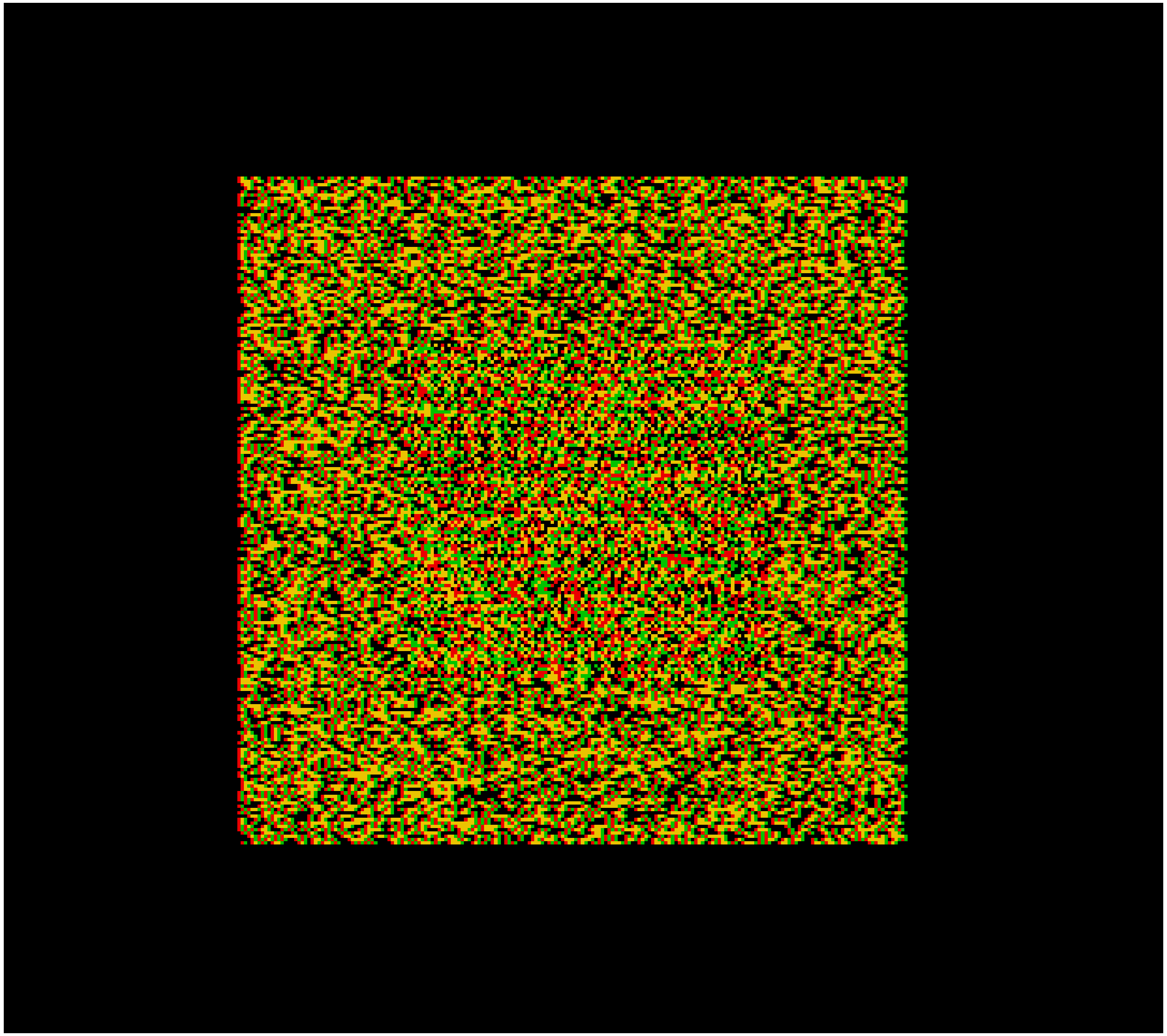


Correspondence: Area-based

Benefit of coarse-to-fine stereo correspondence



Correspondence: Random-dot stereogram



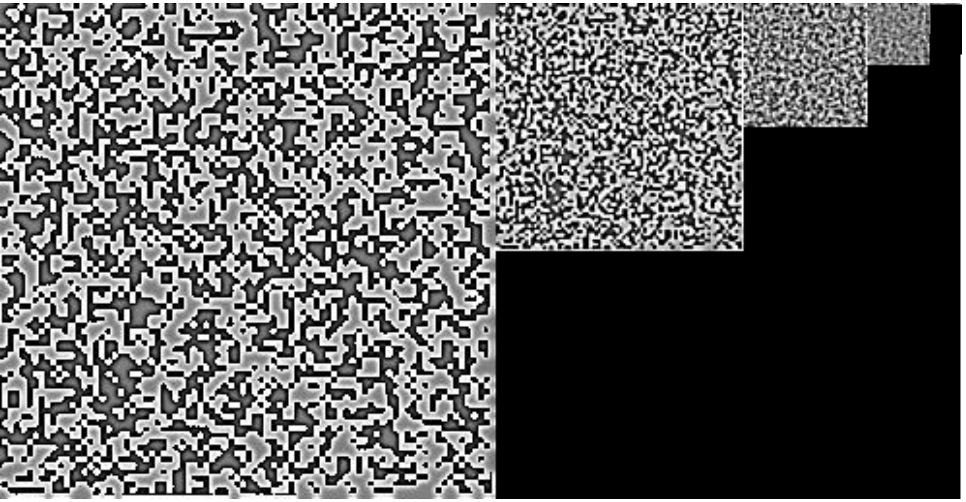
Correspondence: Random-dot stereogram

Correspondence: Random-dot stereogram

Correspondence: Random-dot stereogram

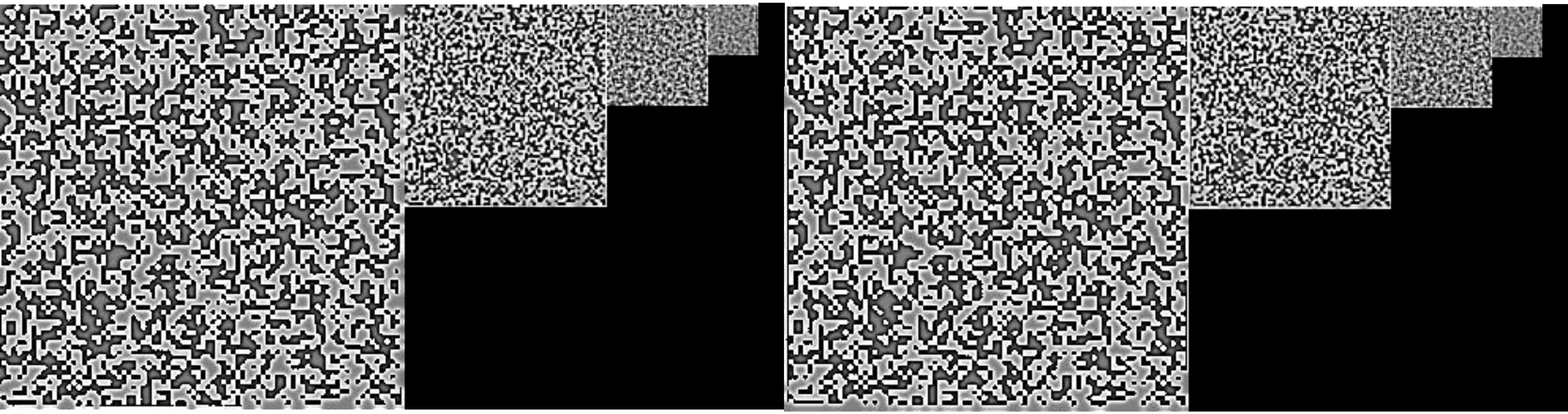
Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram



Multiresolution (Laplacian) pyramid representations

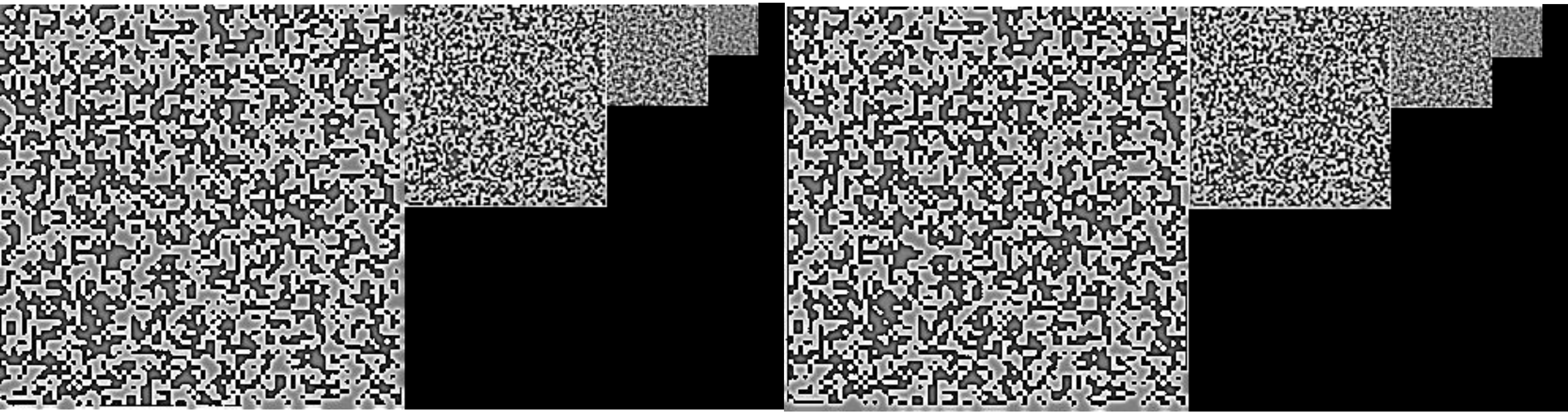
Correspondence: Random-dot stereogram



Multiresolution (Laplacian) pyramid representations

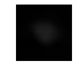
Correspondence: Random-dot stereogram

Estimated (multiresolution) disparity

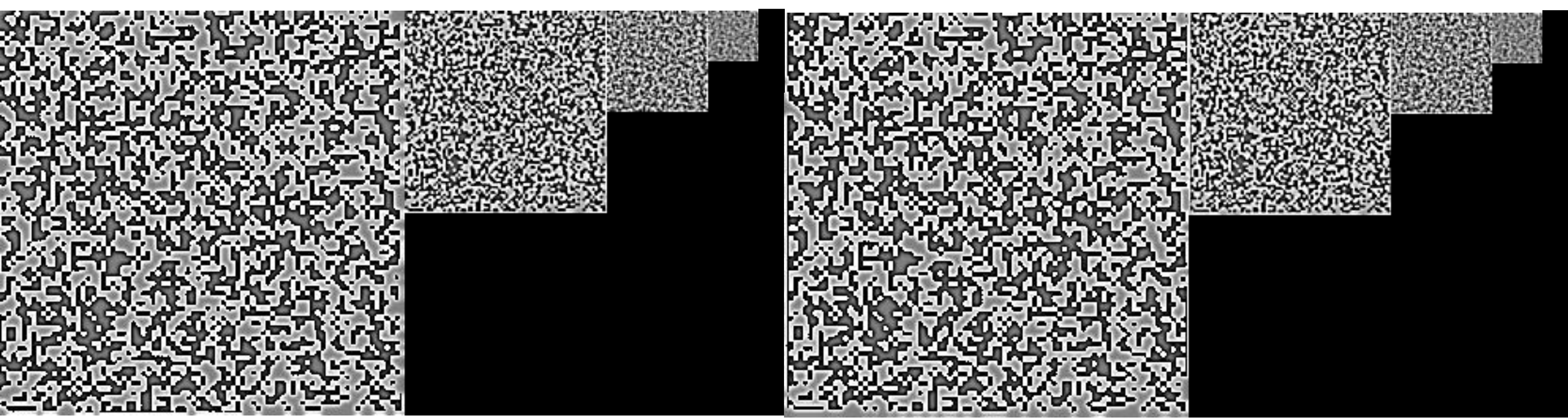


Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram

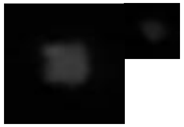


Estimated (multiresolution) disparity

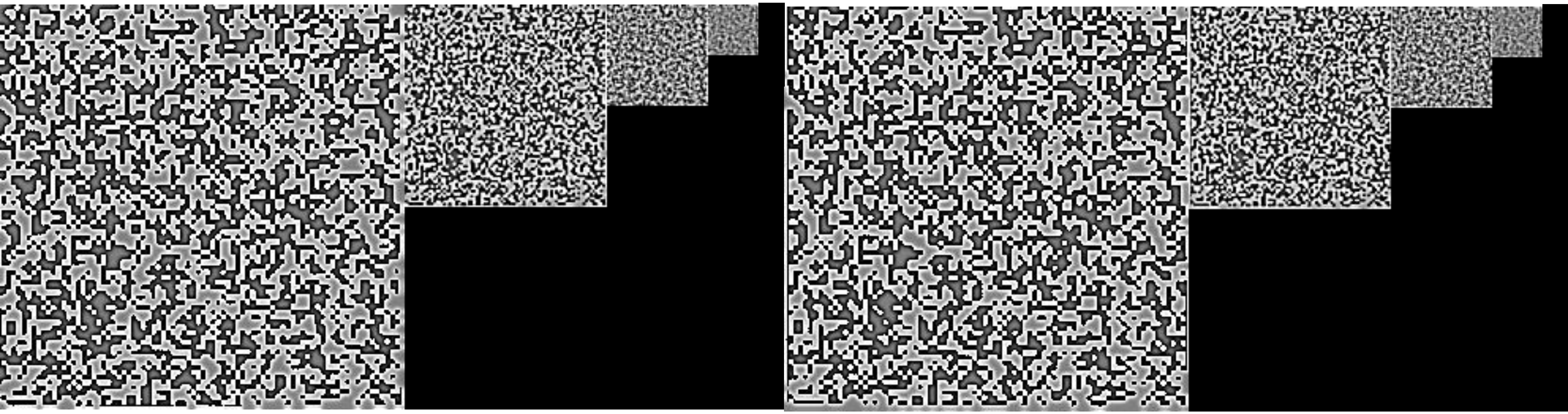


Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram

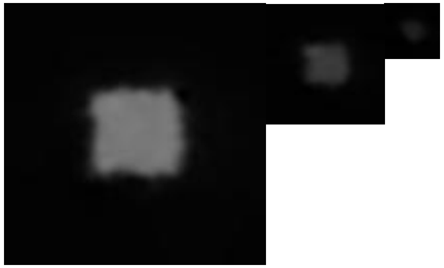


Estimated (multiresolution) disparity

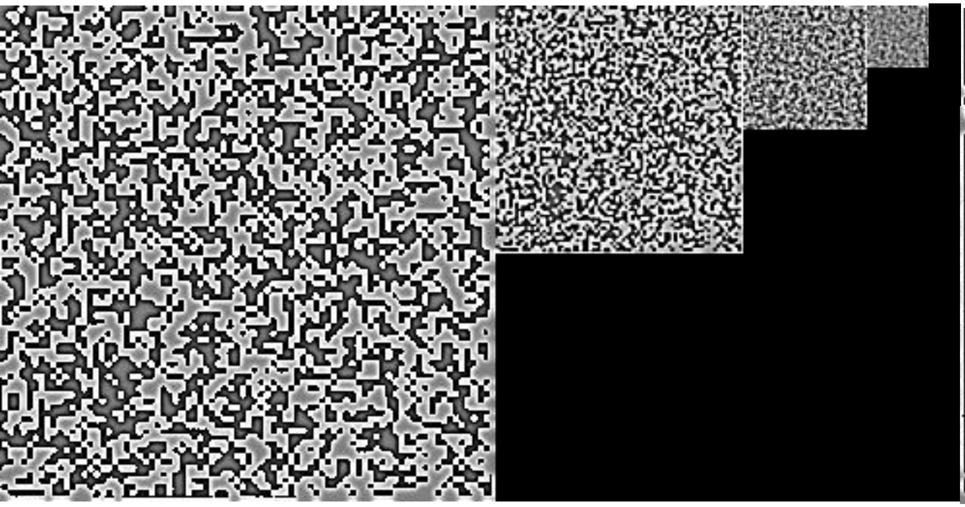


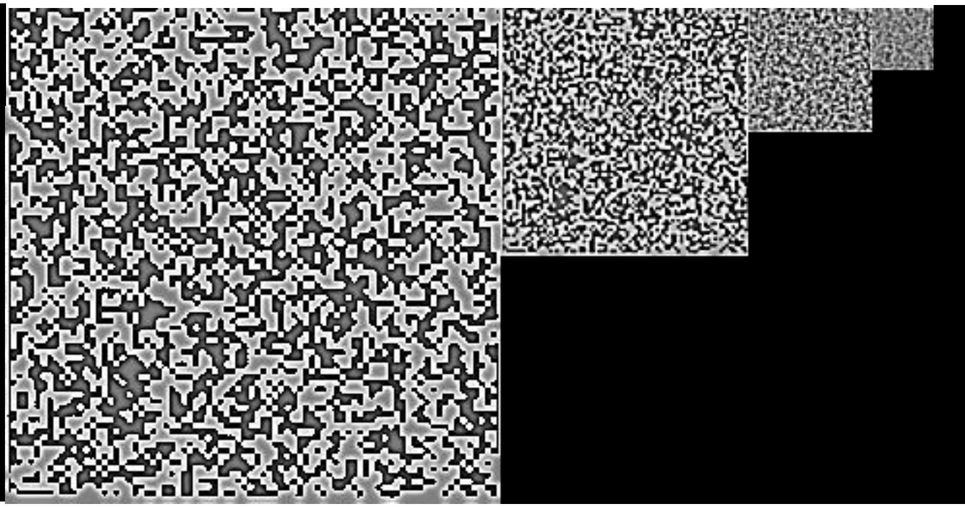
Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram



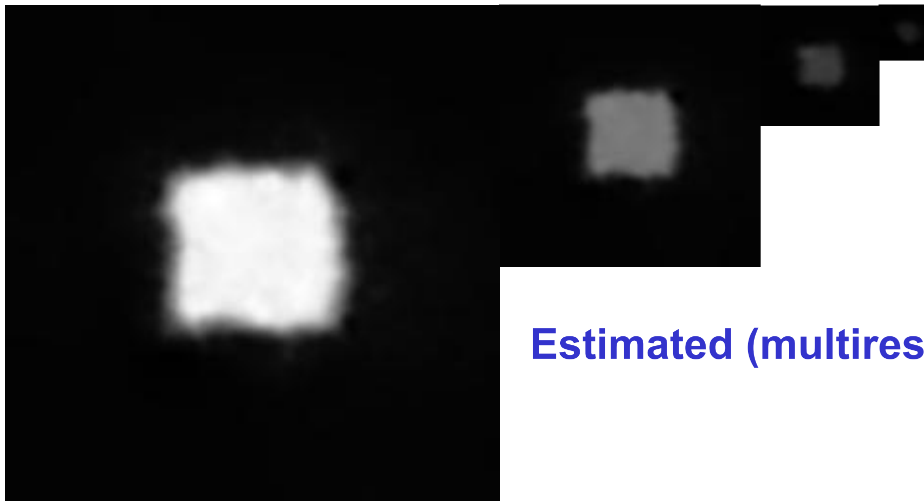
Estimated (multiresolution) disparity



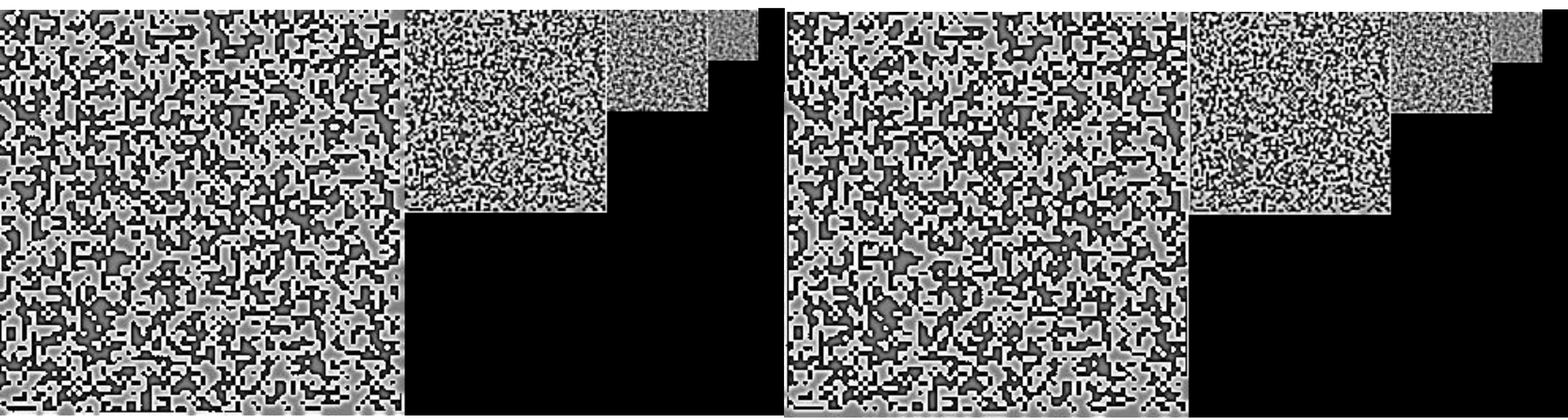


Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram



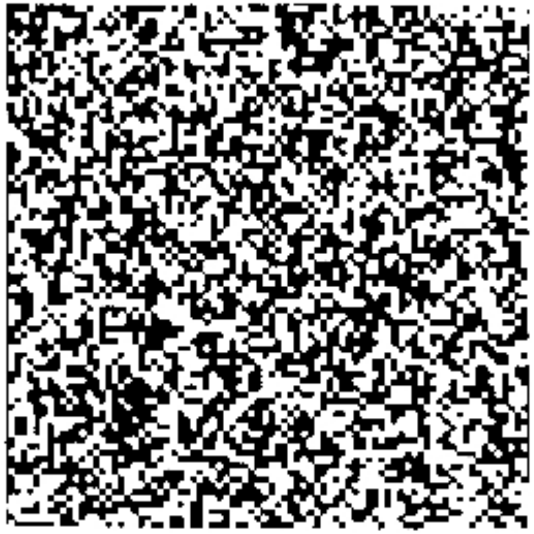
Estimated (multiresolution) disparity



Multiresolution (Laplacian) pyramid representations

Correspondence: Random-dot stereogram

Input:
Stereo pair



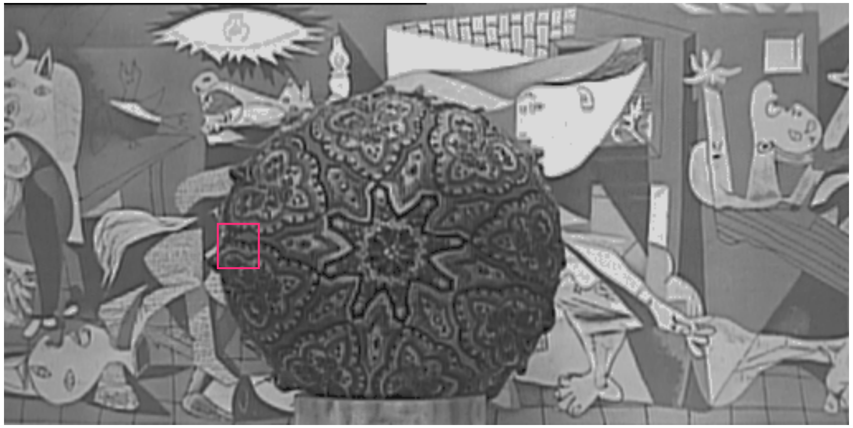
Output:
Disparity map

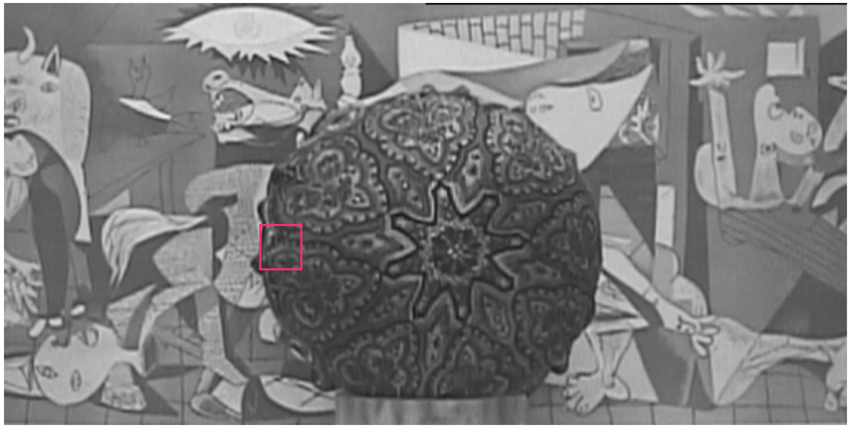
Correspondence: Area-based (recap.)

Motivation
- Exploit all available information

Elements to be matched
- Image windows
- Typically of fixed size
- Spatially overlapping.

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





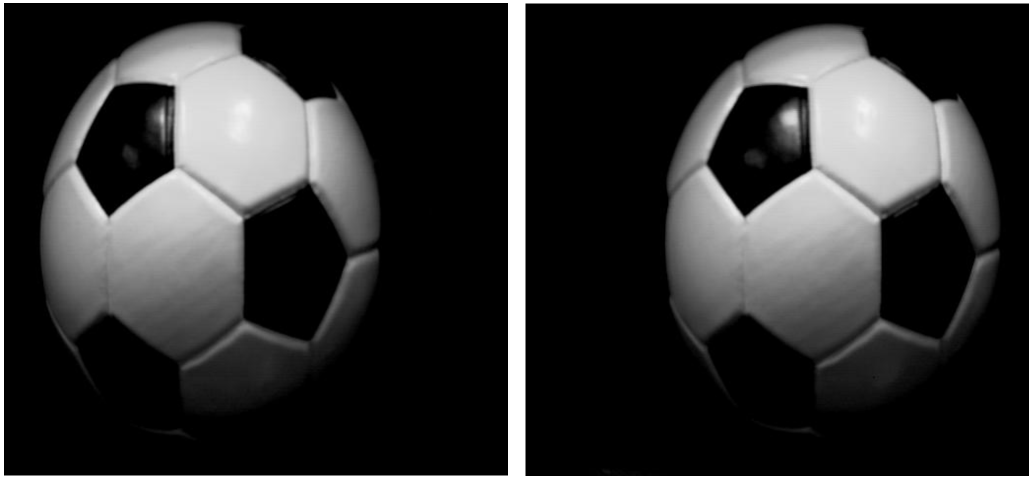
Correspondence: Feature-based

Motivation

- 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…



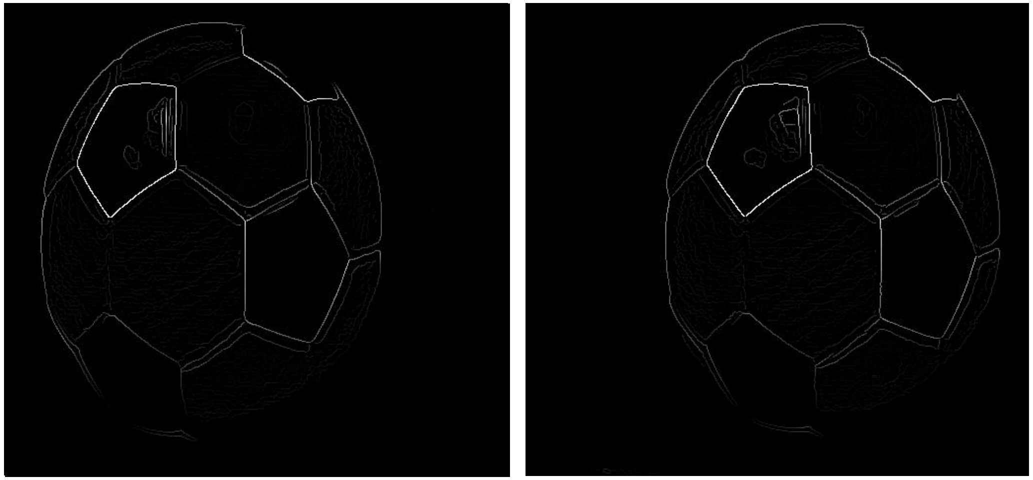
Correspondence: Feature-based

Motivation

- 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

Motivation

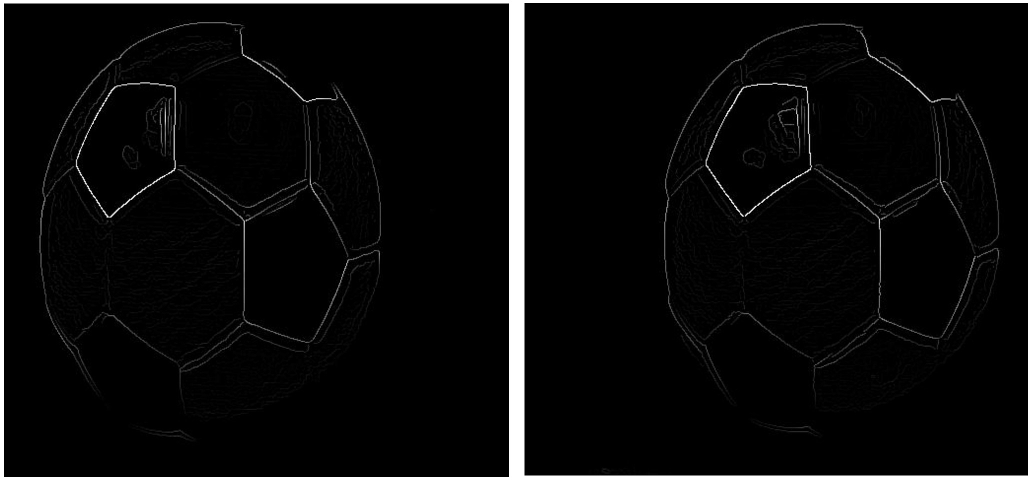
- Not all features are created equal.

Elements to be matched

- Sparse set of extracted features.
- Numerical and/or symbolic descriptors associated with features.

Similarity measure

- Distance between feature descriptors.
- Corresponding element is that which minimizes distance between feature descriptors.



Correspondence: Feature-based

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 $\boldsymbol{fl}$
  - $\boldsymbol{d(fl,fr)}$ 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 $\boldsymbol{fl}$

Correspondence: Feature-based

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
  - $R(\boldsymbol{fl})$ be the search range in the right image associated with left-image feature descriptor $\boldsymbol{fl}$
  - $\boldsymbol{d(fl,fr)}$ 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_r 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: Feature-based

Algorithm

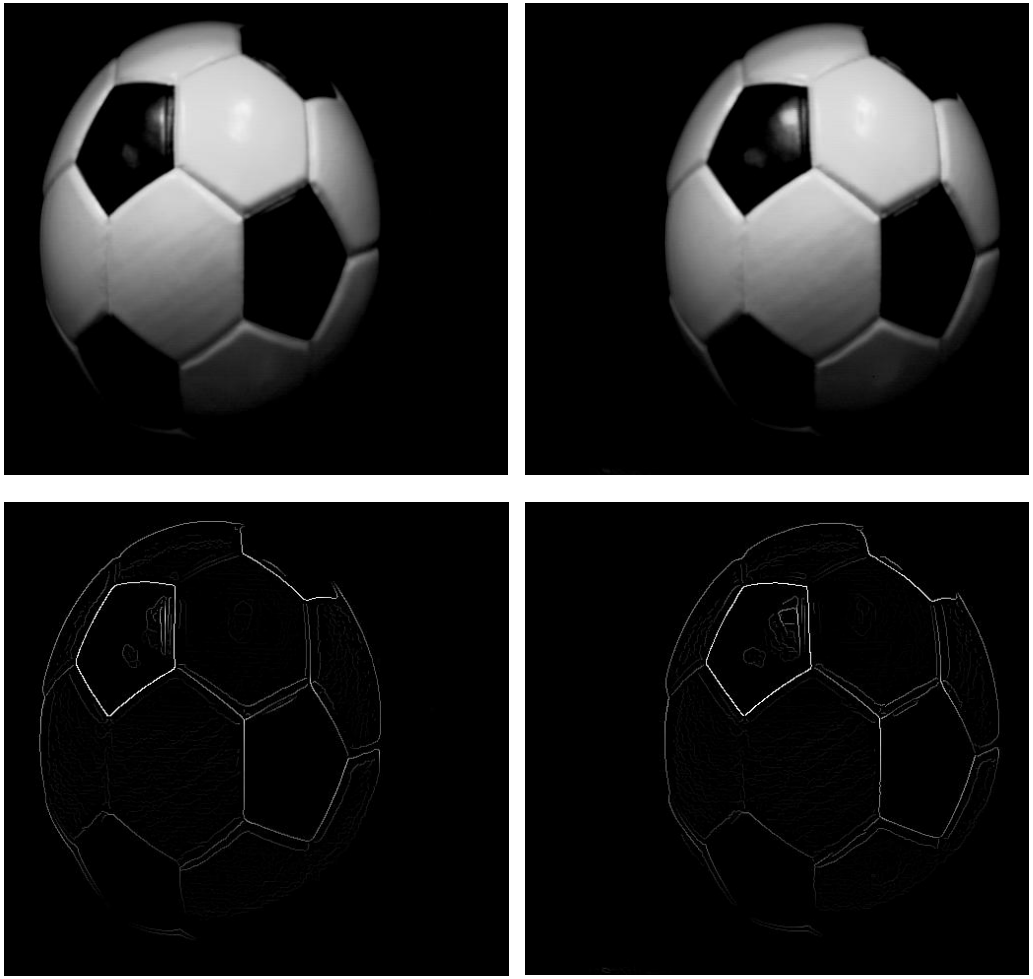
- **Input:** Stereo pair of images *Il* and *Ir* and associated sets of feature descriptors
- **Output:** List of feature correspondences and (possibly sparse) (disparity map).
- **Notation:** Let
  - *R(**fl**)* be the search range in the right image associated with left-image feature descriptor ***fl***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



Motivation

- Not all features are created equal.

Elements to be matched

- Sparse set of extracted features.
- Numerical and/or symbolic descriptors associated with features.

Similarity measure

- Distance between feature descriptors.
- Corresponding element is that which minimizes distance between feature descriptors.

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

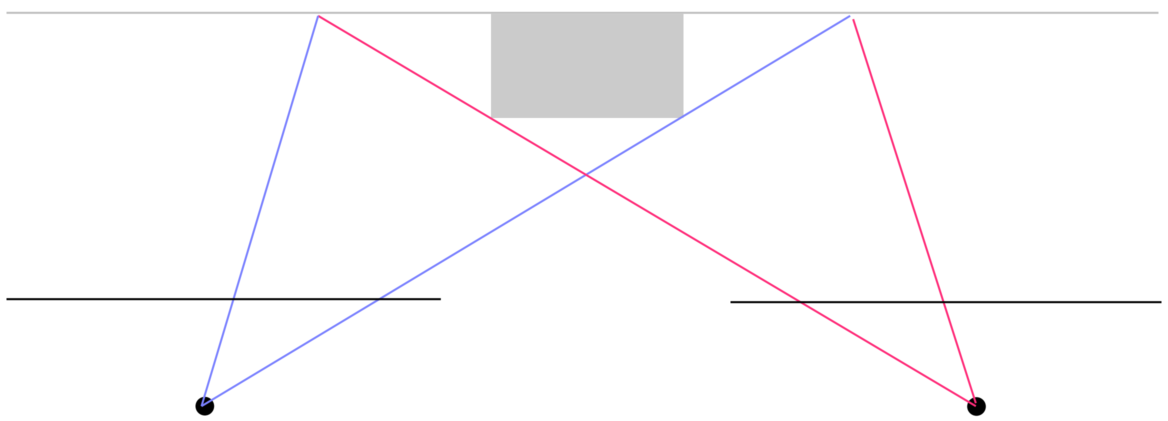
Feature-based

- 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: Final remarks

The issue of unmatchable points

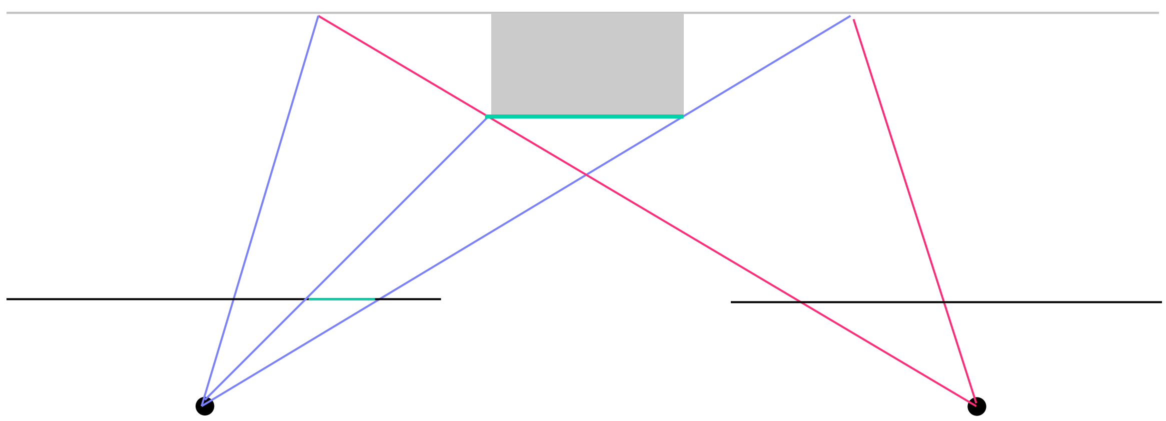
- Both methods can be stymied in attempting to match points that appear in only one of the two views.
  - Due to half occlusion
  - Due to noise



Correspondence: Final remarks

The issue of unmatchable points

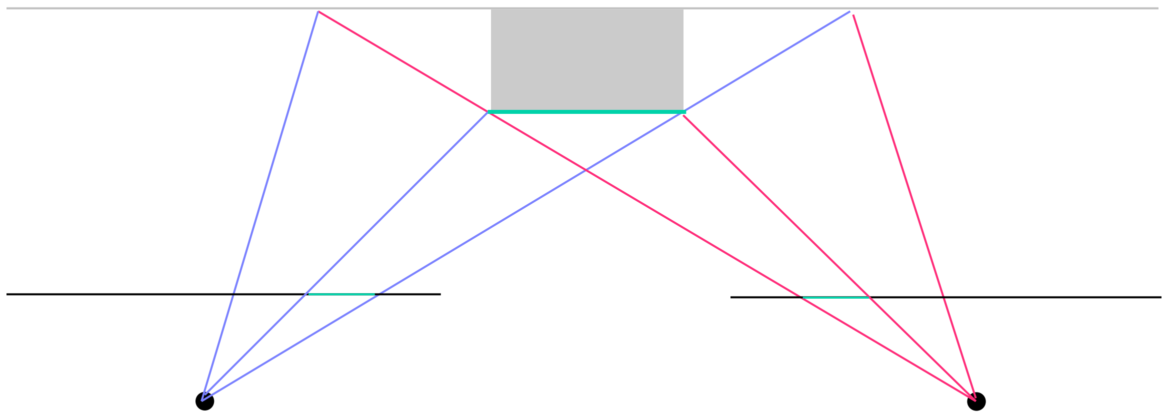
- Both methods can be stymied in attempting to match points that appear in only one of the two views.
  - Due to half occlusion
  - Due to noise



Correspondence: Final remarks

The issue of unmatchable points

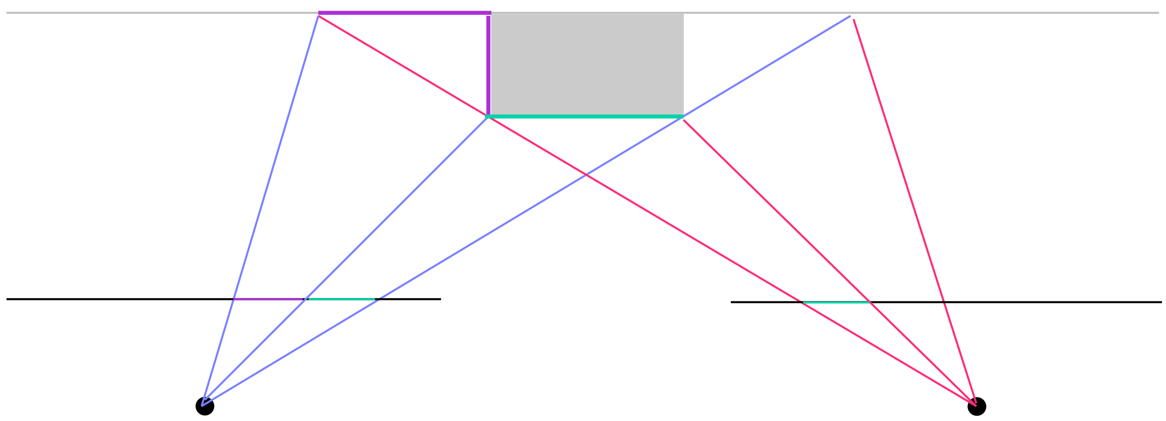
- Both methods can be stymied in attempting to match points that appear in only one of the two views.
  - Due to half occlusion
  - Due to noise



Correspondence: Final remarks

The issue of unmatchable points

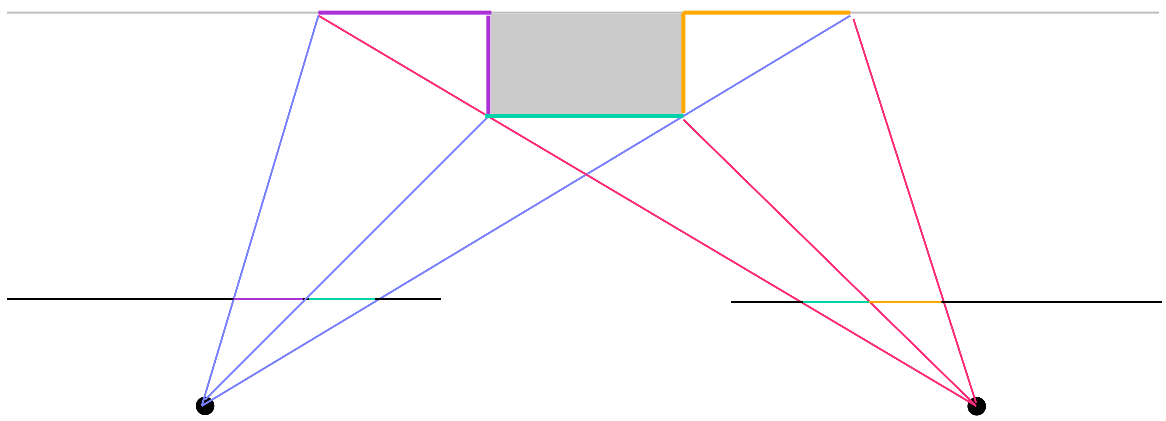
- Both methods can be stymied in attempting to match points that appear in only one of the two views.
  - Due to half occlusion
  - Due to noise



Correspondence: Final remarks

The issue of 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

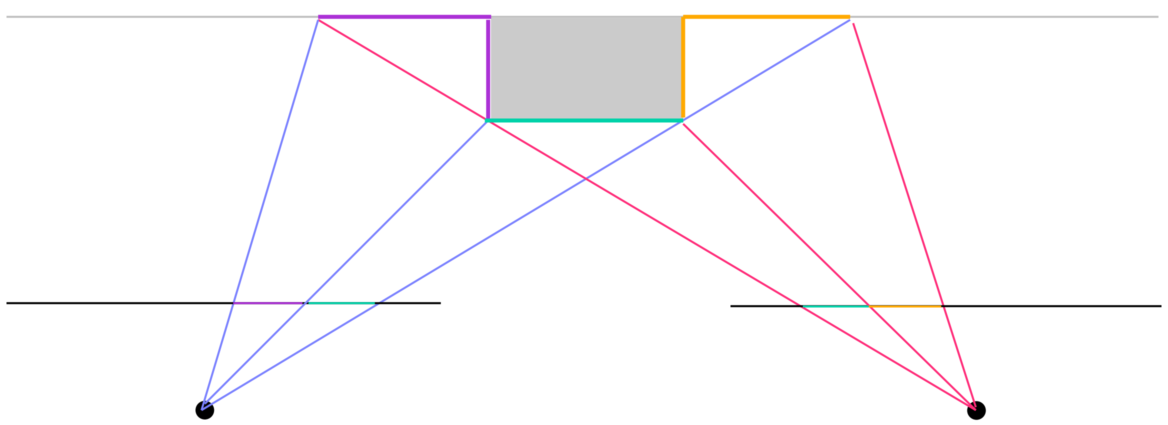


Geometry of half-occlusion

Correspondence: Final remarks

The issue of 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
- 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.

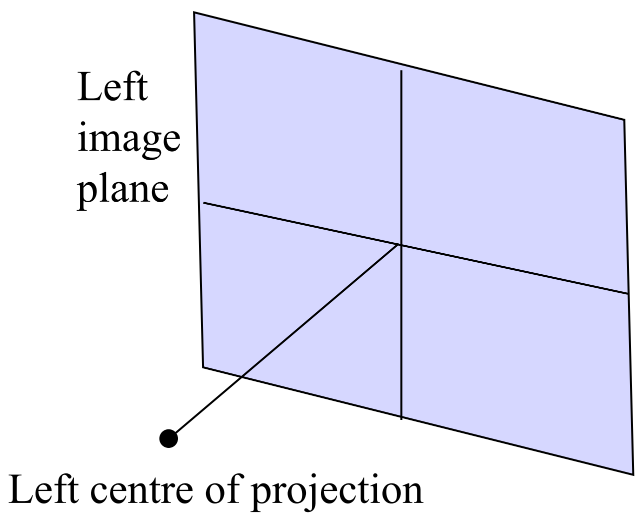


Geometry of half-occlusion

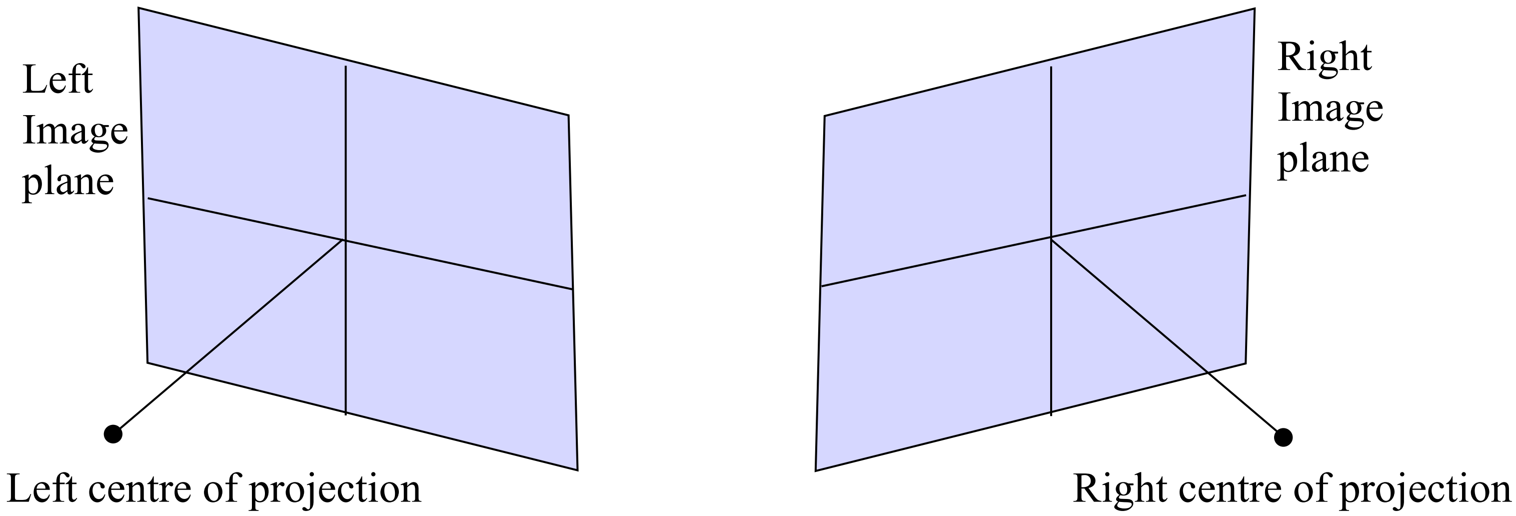
Outline

- Introduction
- The correspondence problem
- Epipolar geometry
- 3-D reconstruction
- Empirical examples
- Summary

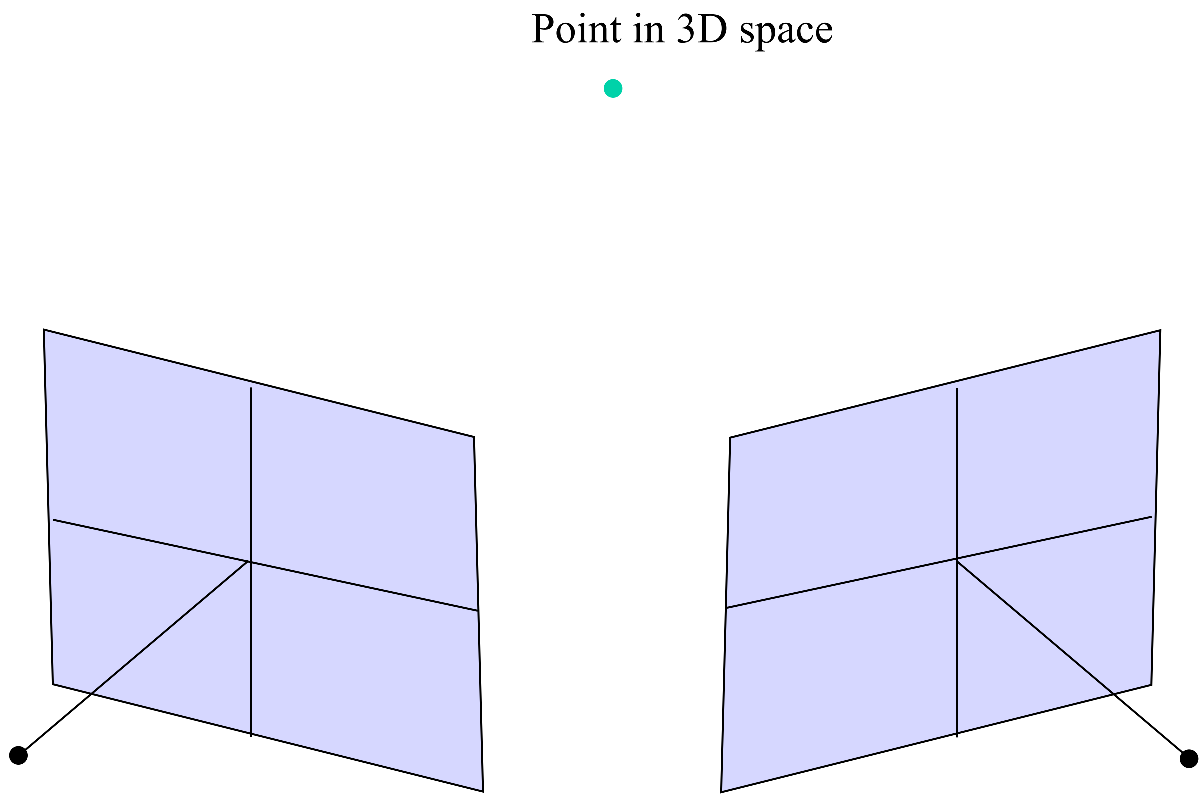
Epipolar geometry: Pictorial explanation



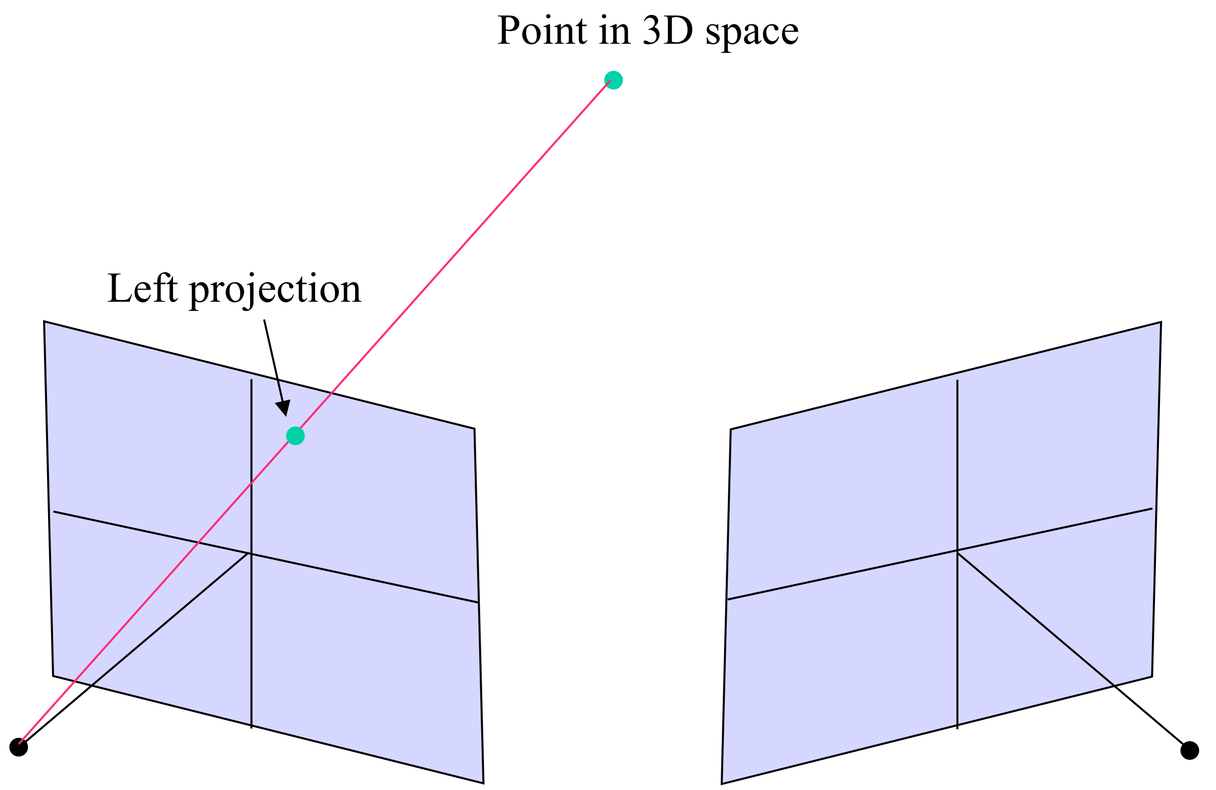
Epipolar geometry: Pictorial explanation



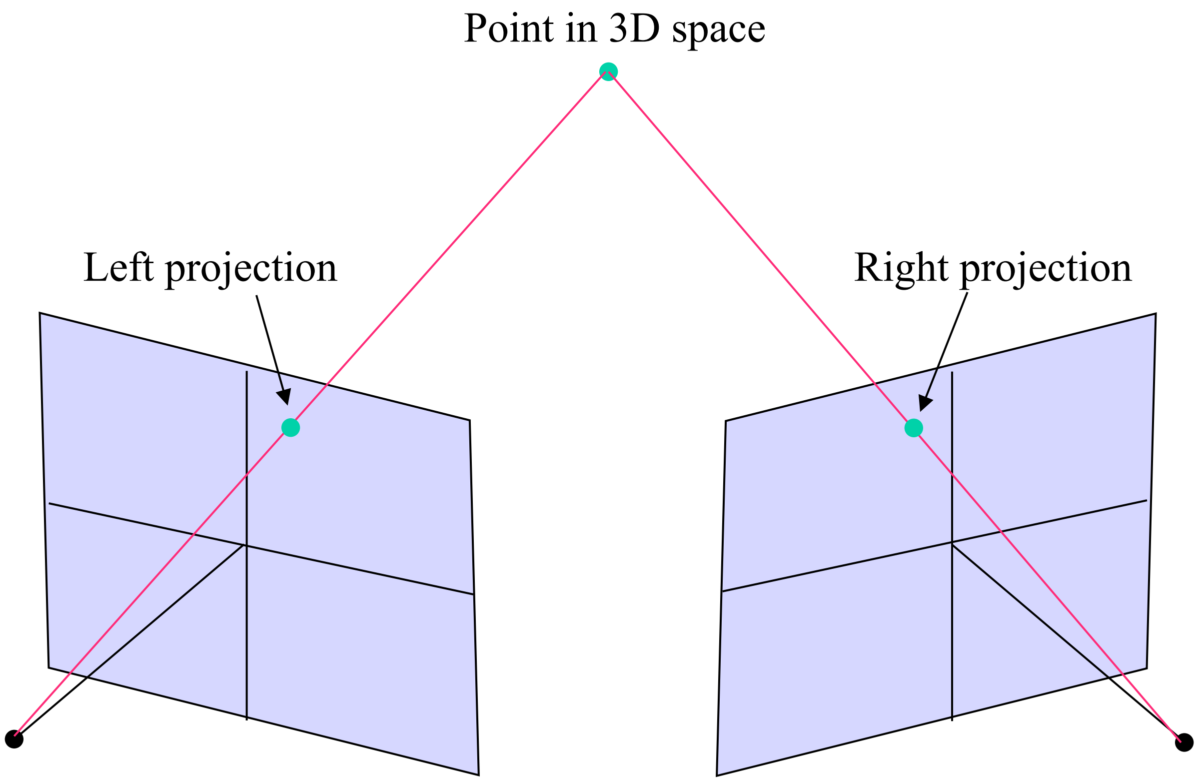
Epipolar geometry: Pictorial explanation



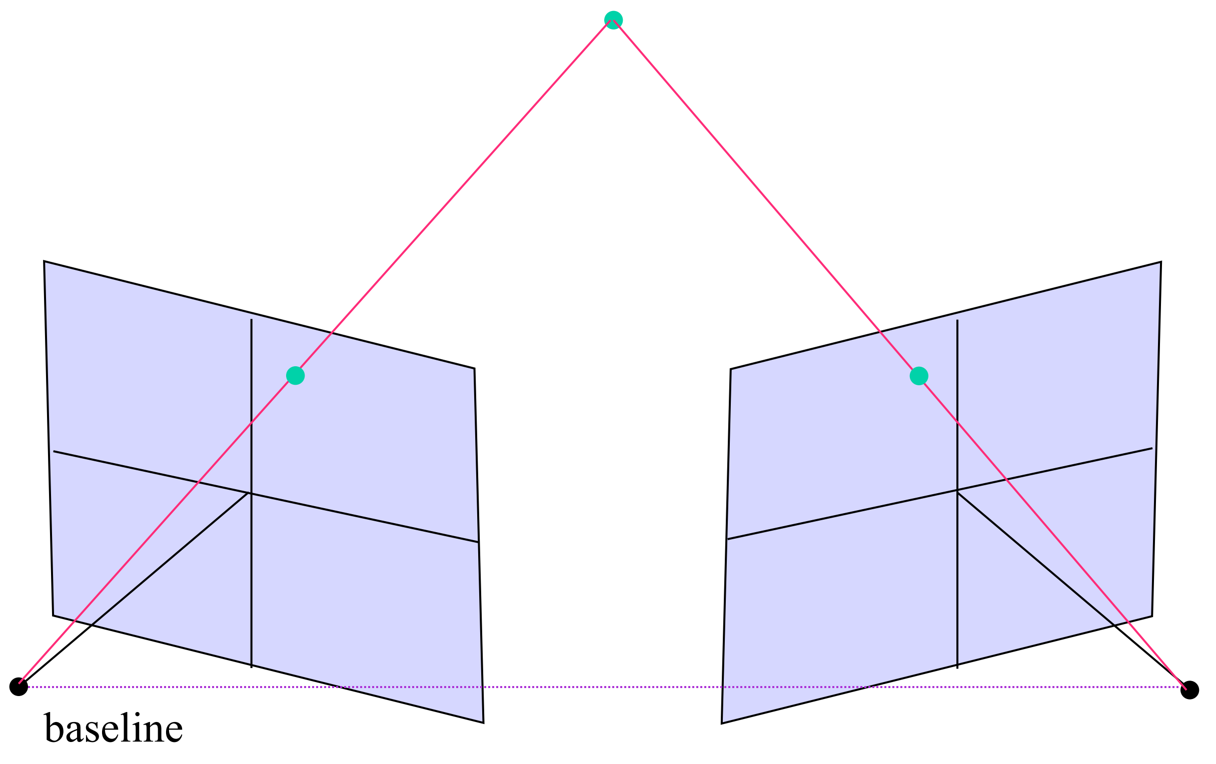
Epipolar geometry: Pictorial explanation



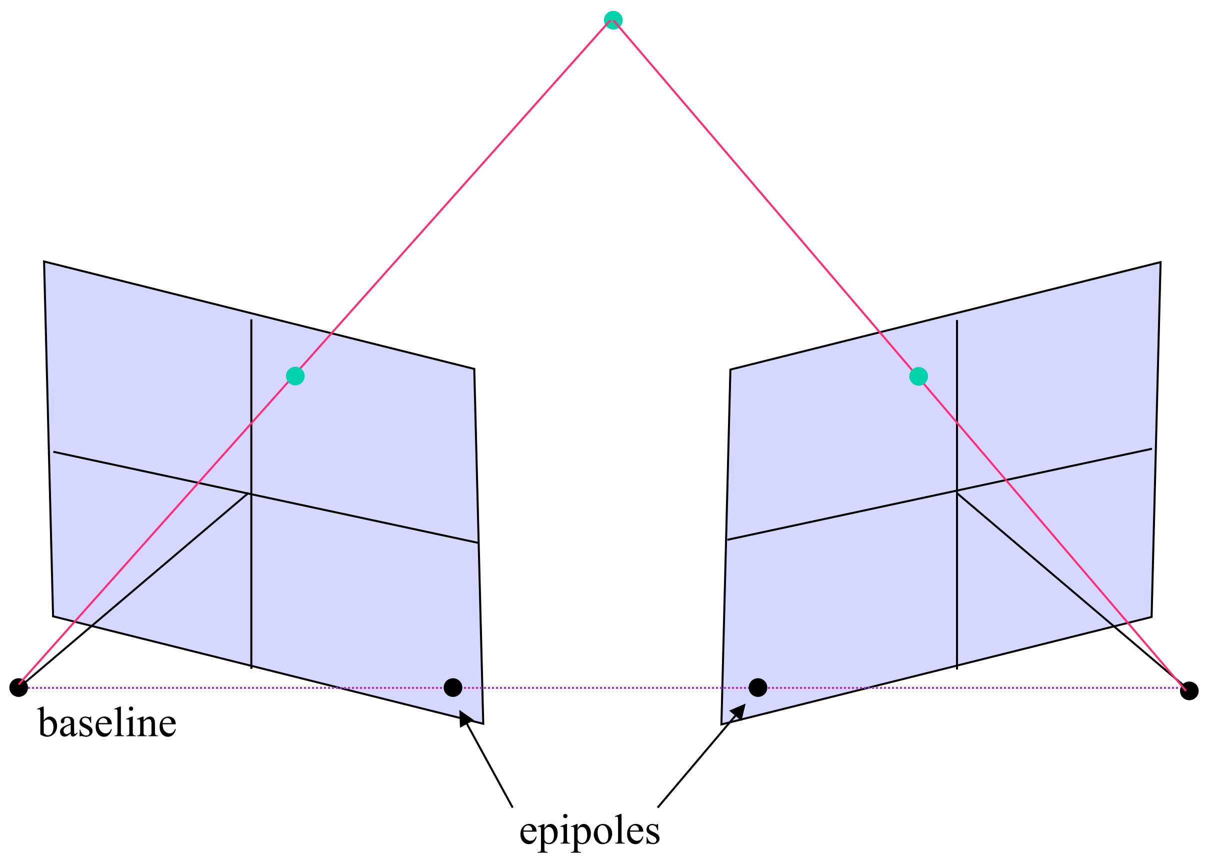
Epipolar geometry: Pictorial explanation



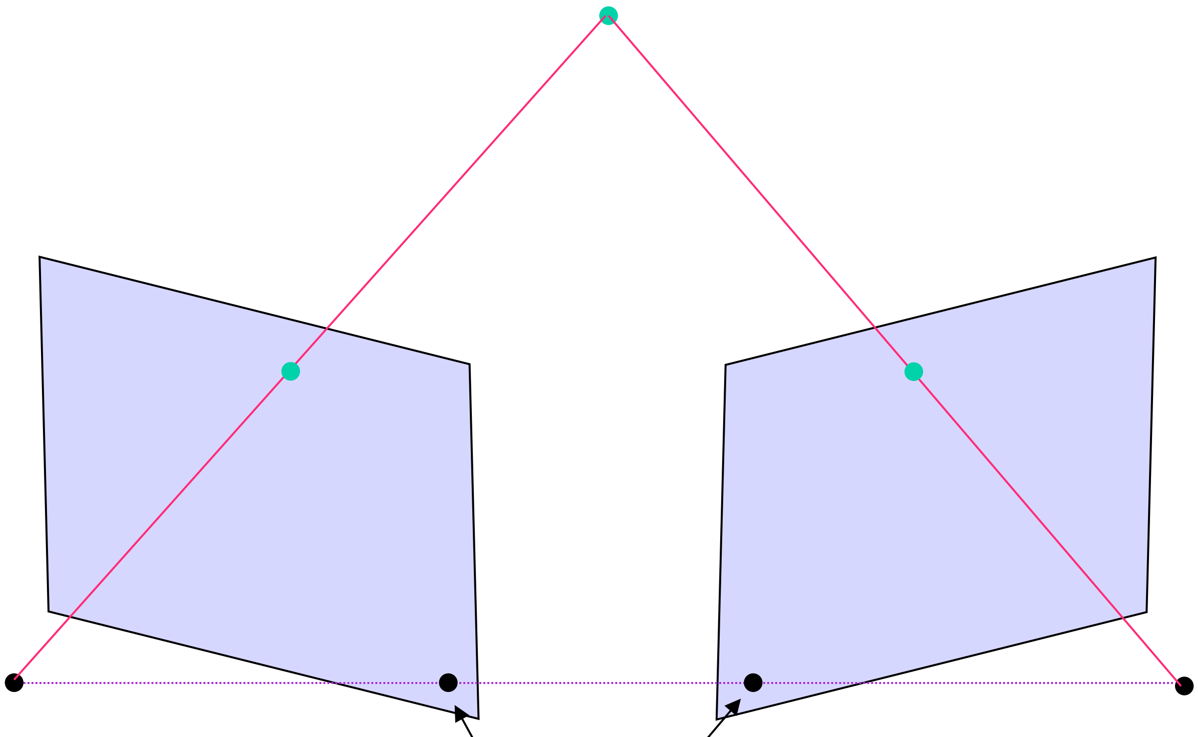
Epipolar geometry: Pictorial explanation



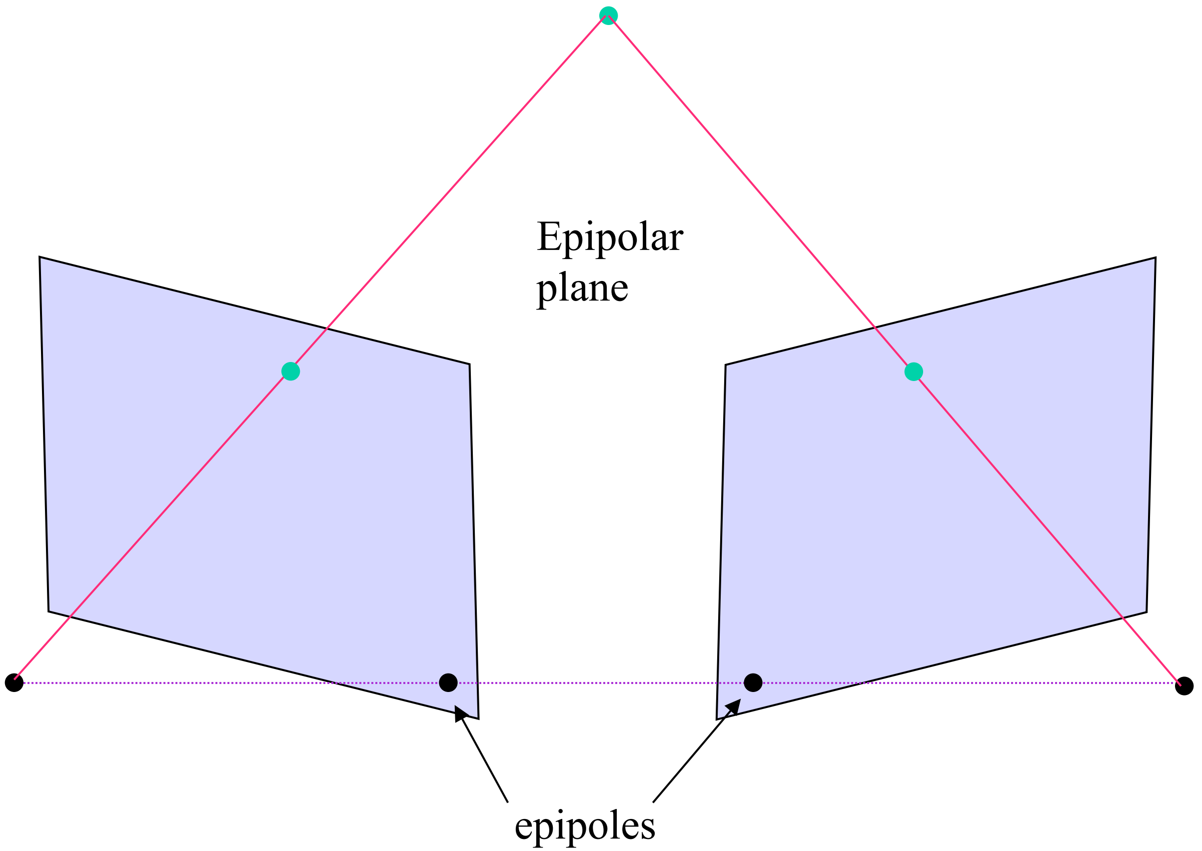
Epipolar geometry: Pictorial explanation



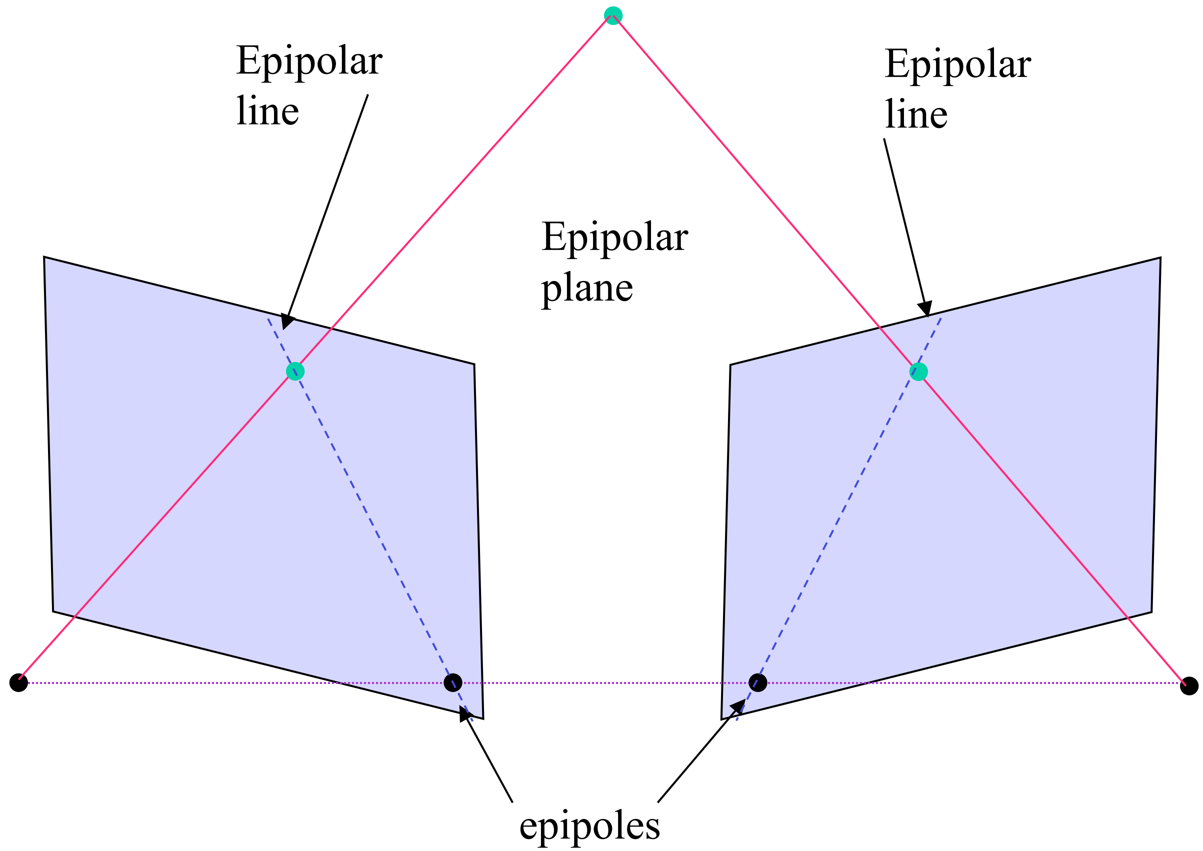
Epipolar geometry: Pictorial explanation



Epipolar geometry: Pictorial explanation



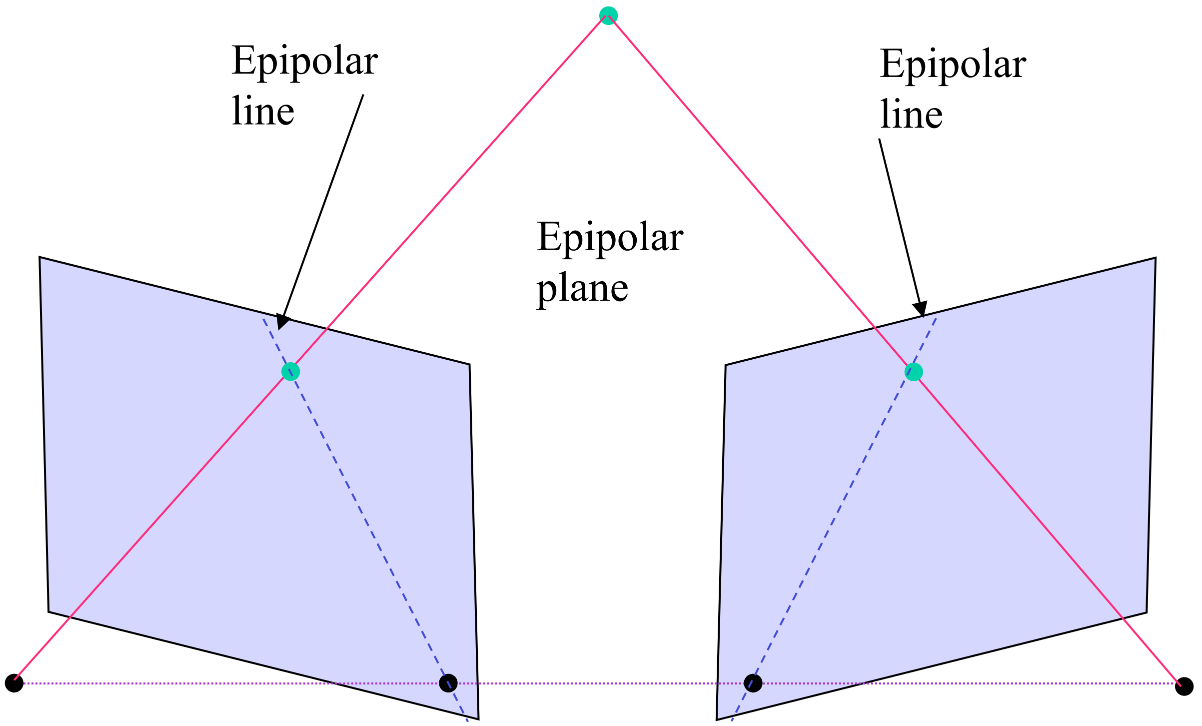
Epipolar geometry: Pictorial explanation



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

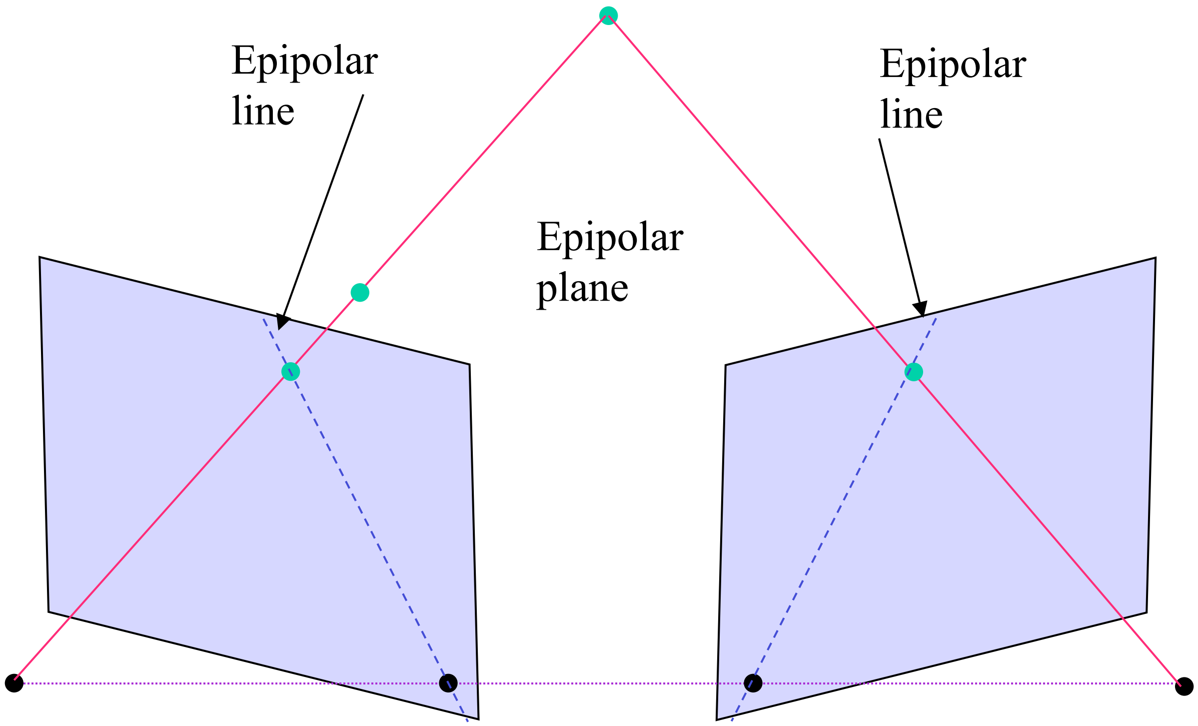
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

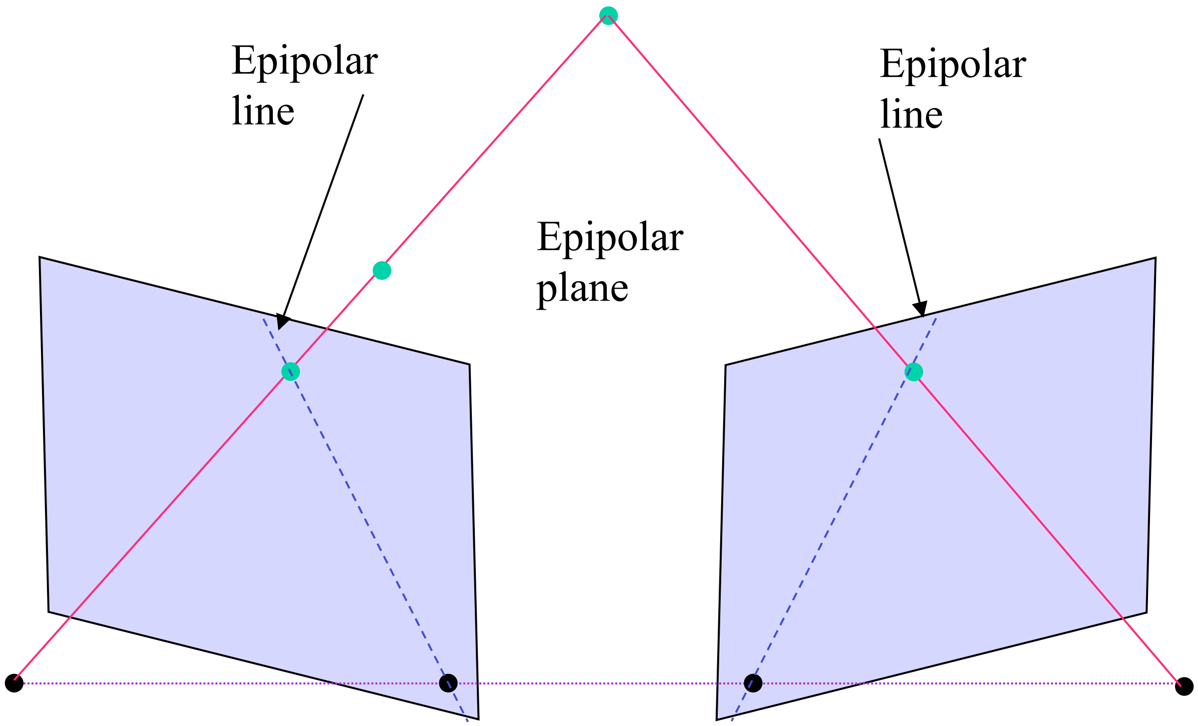
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

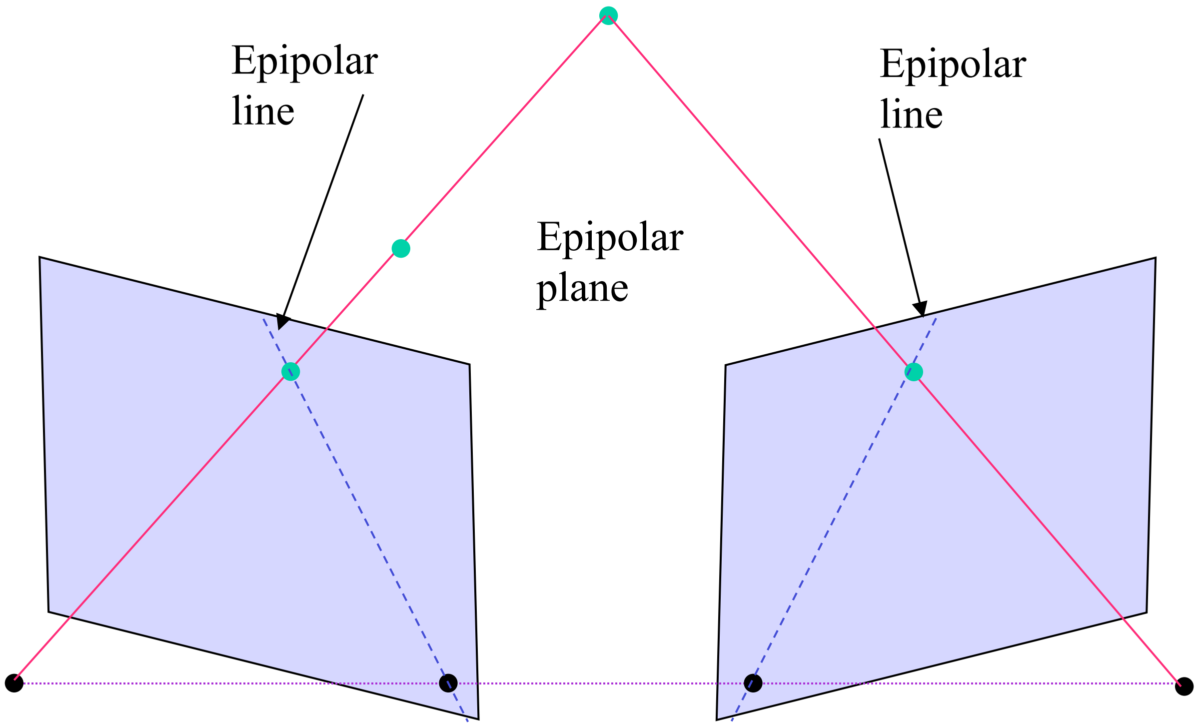
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

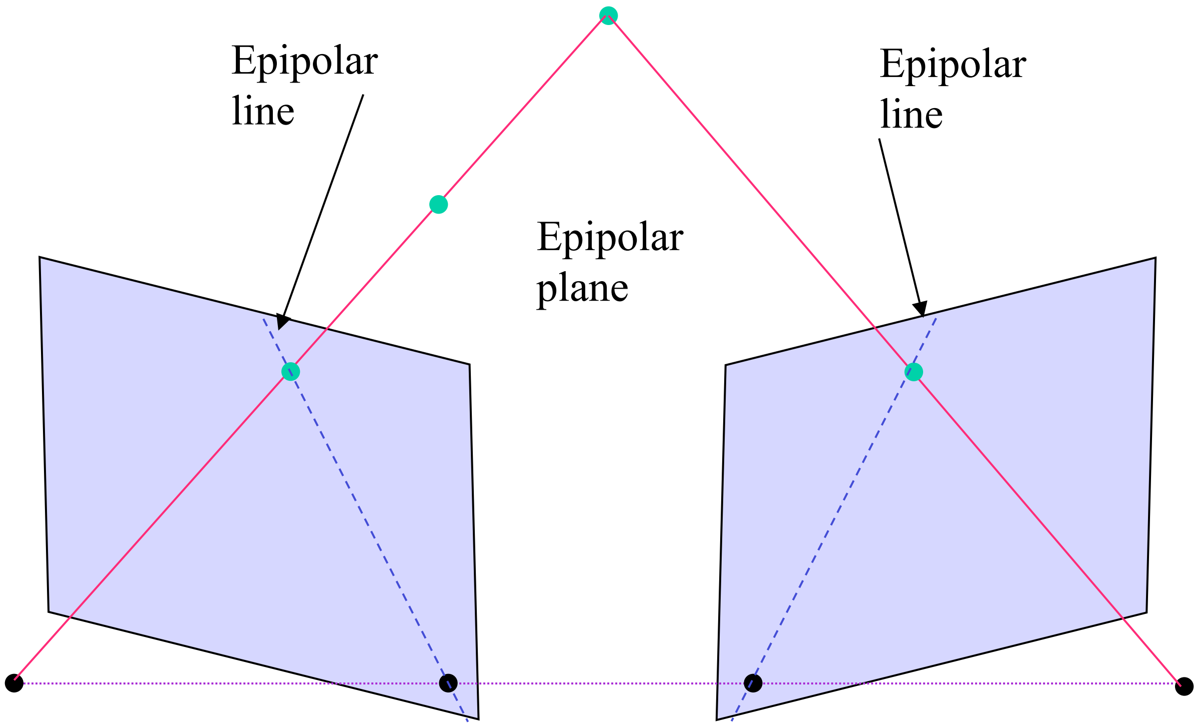
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

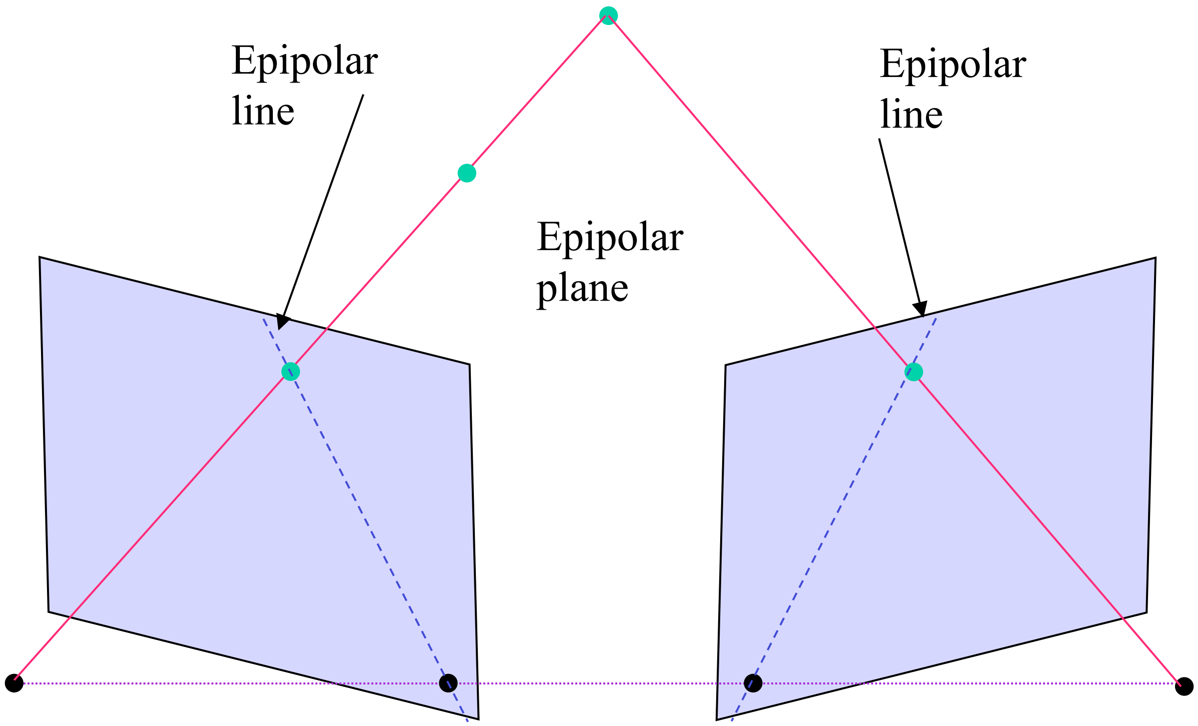
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

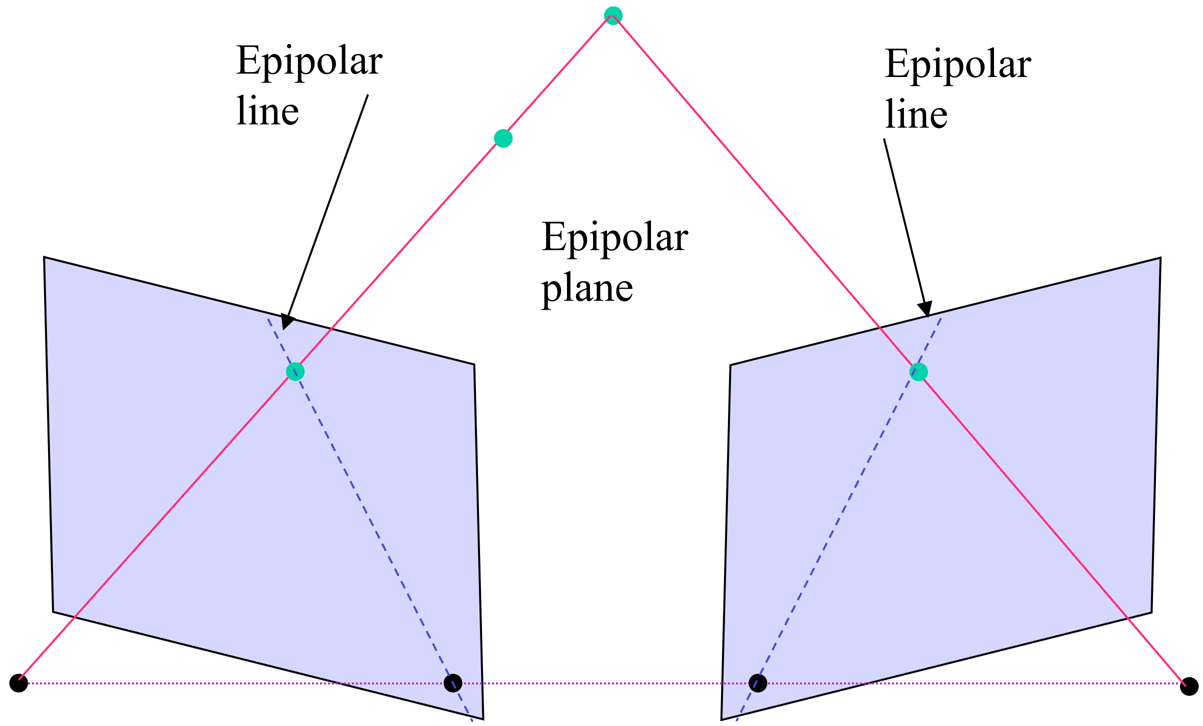
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

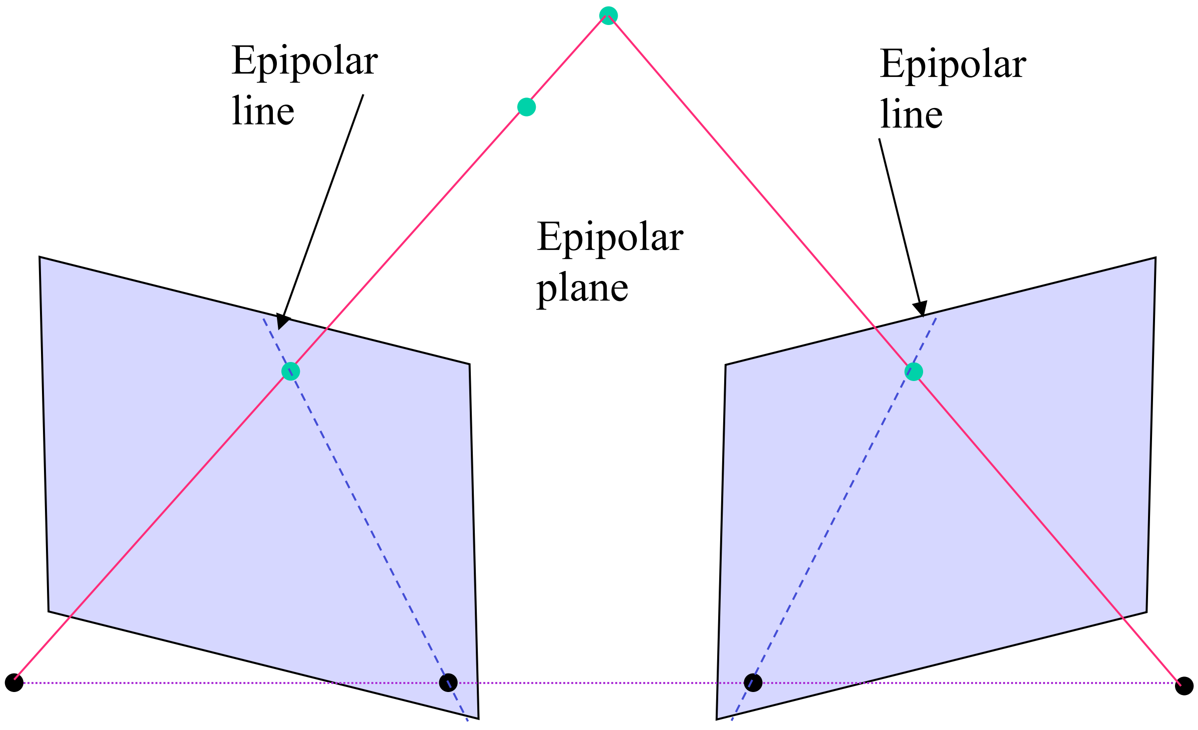
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

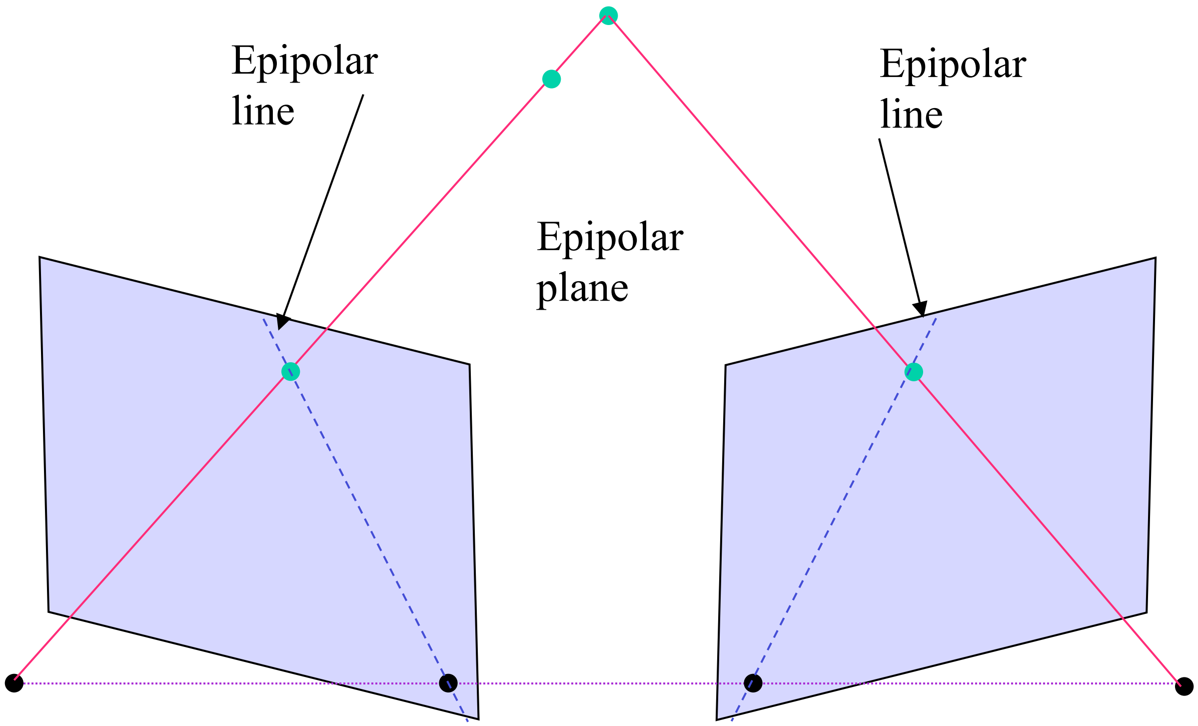
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

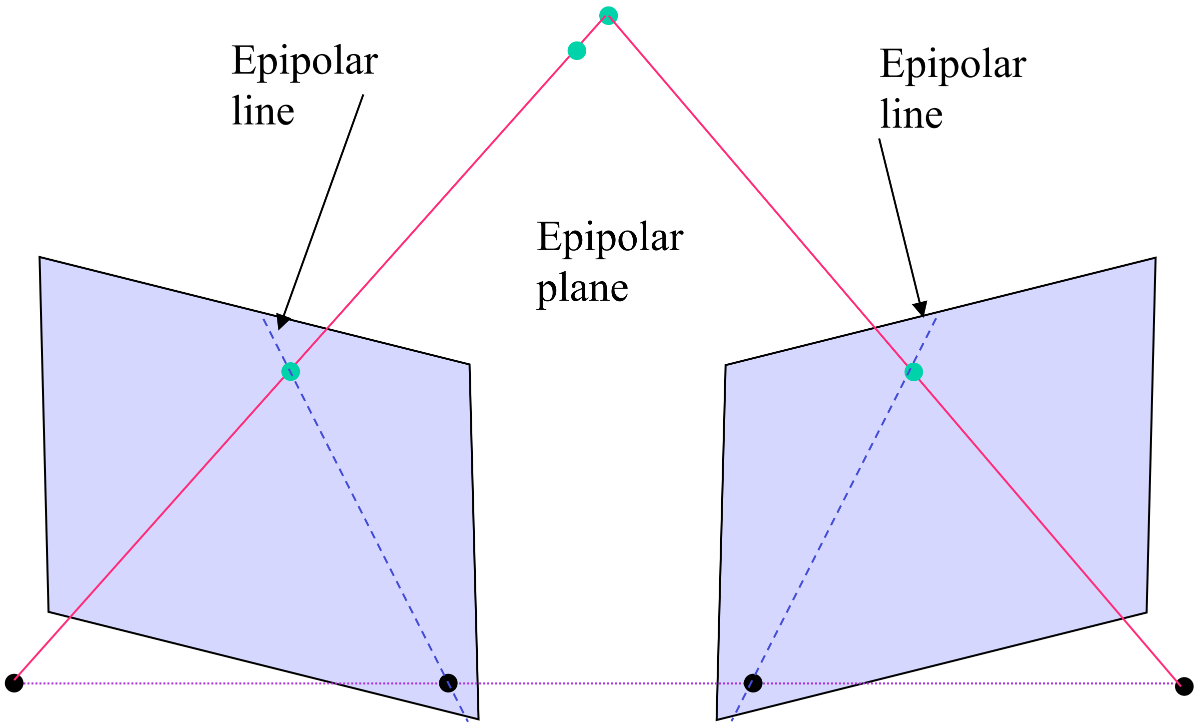
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

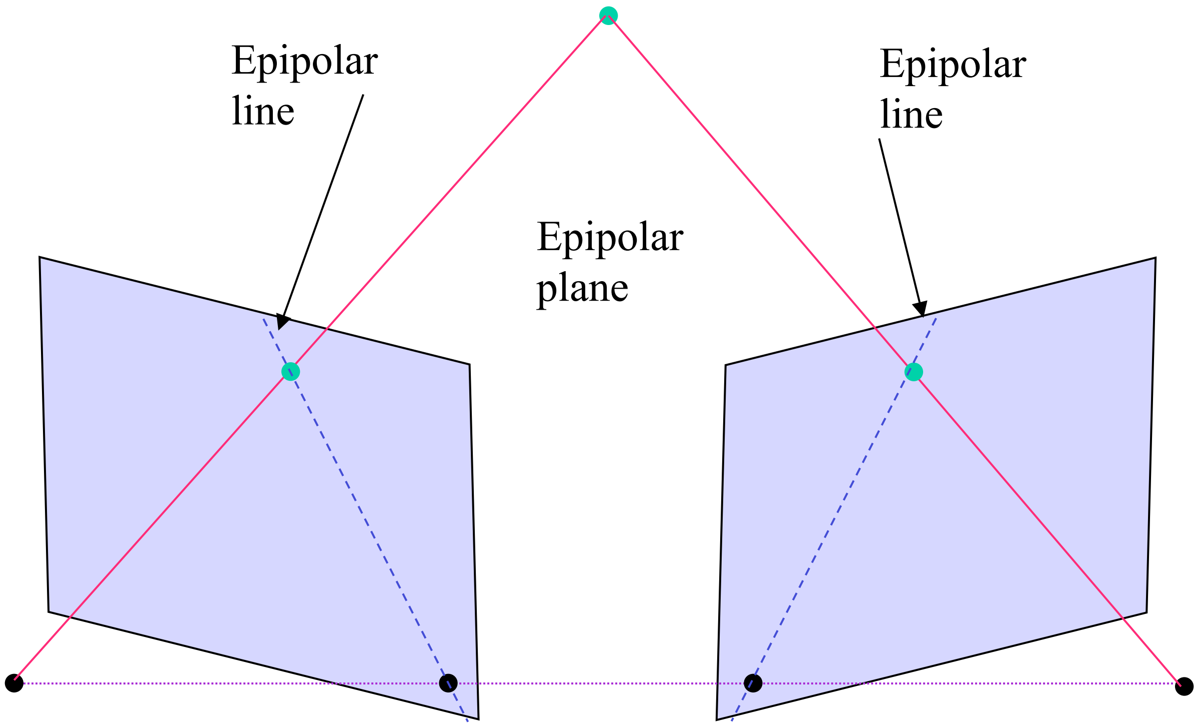
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

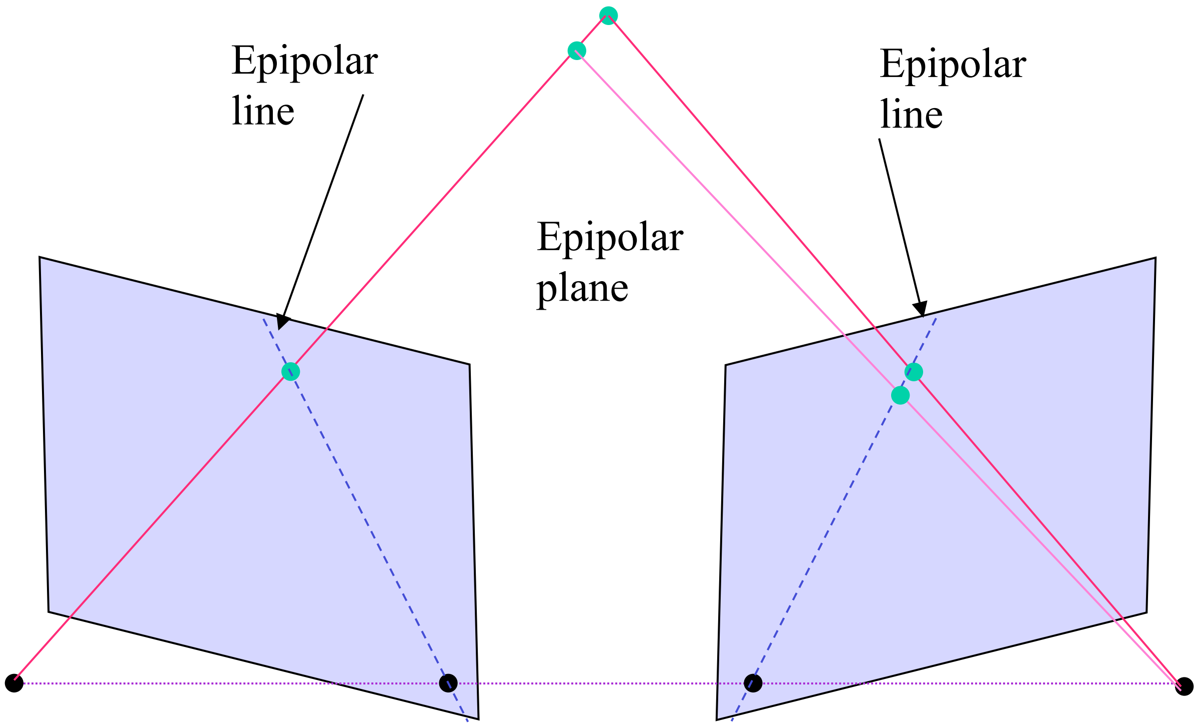
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

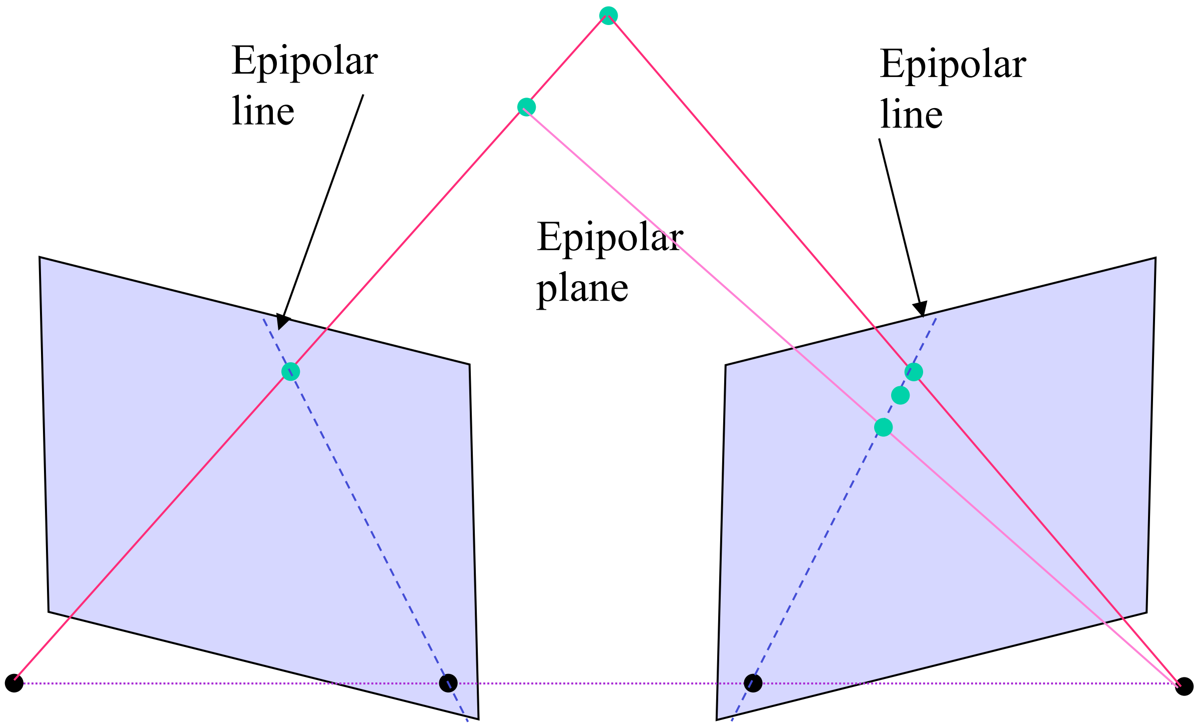
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

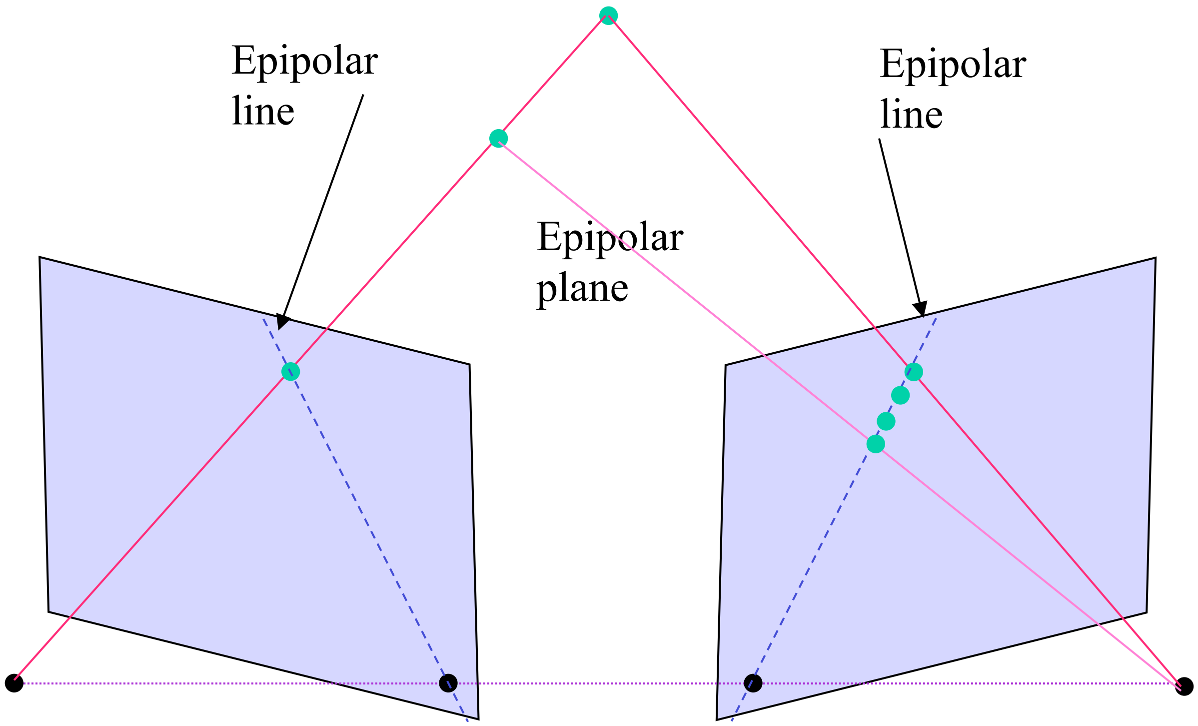
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

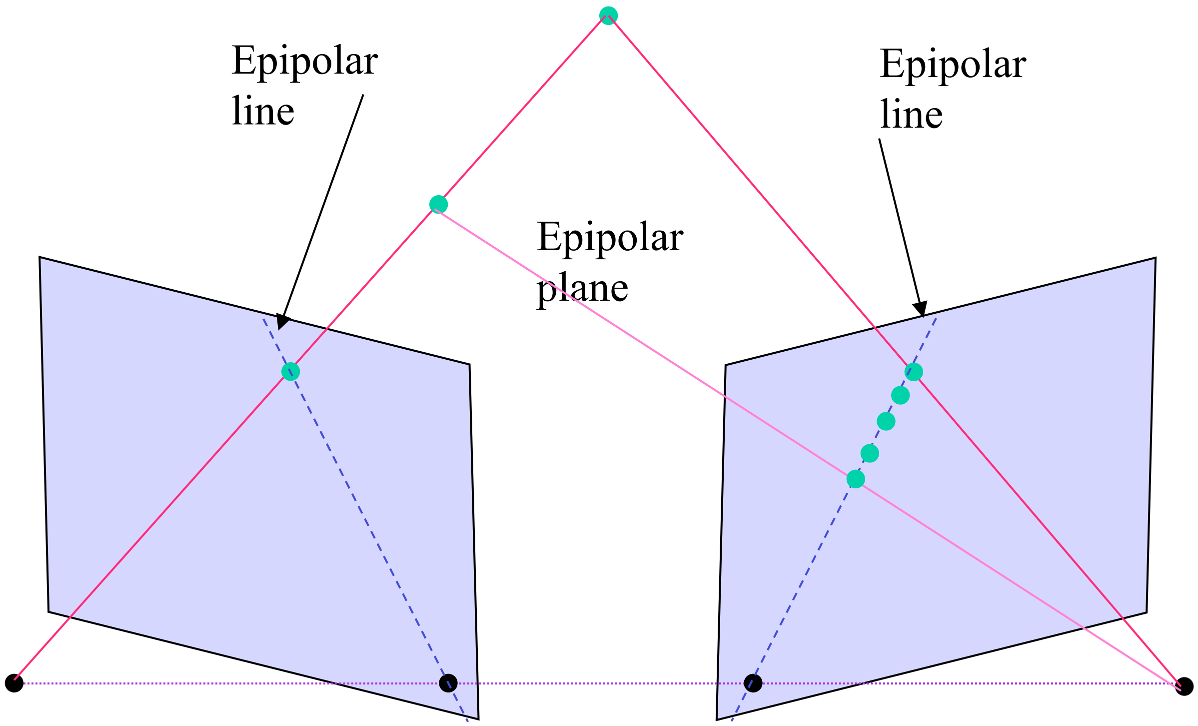
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

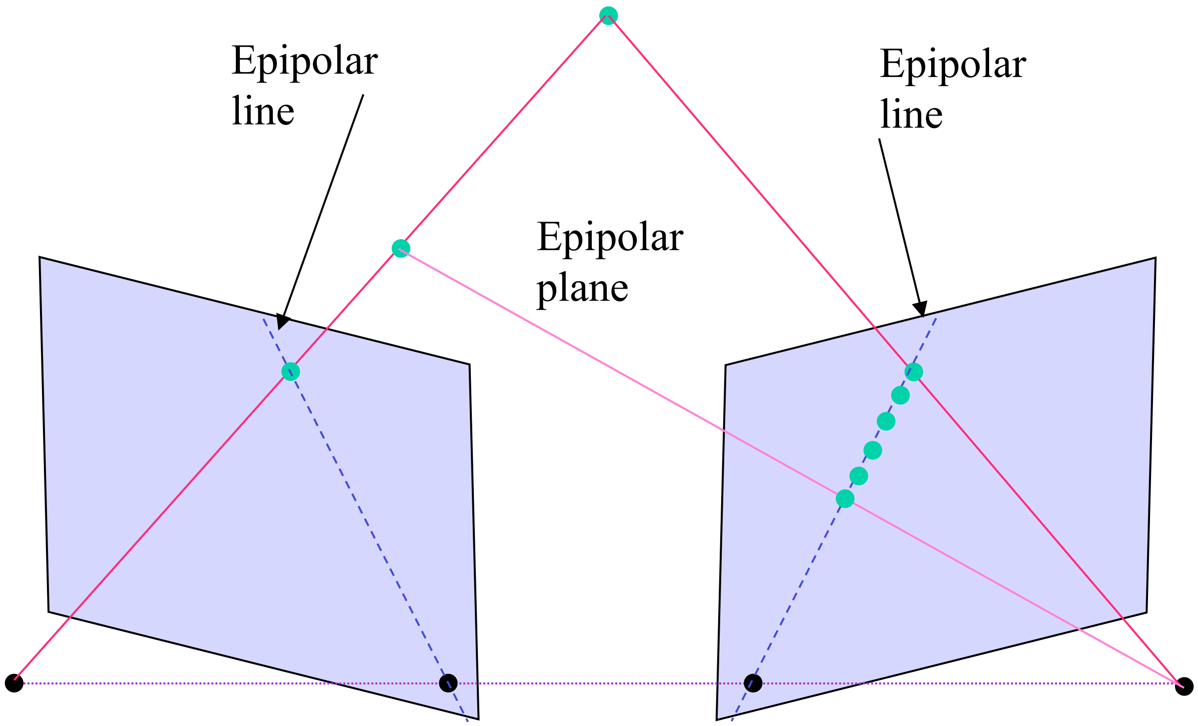
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

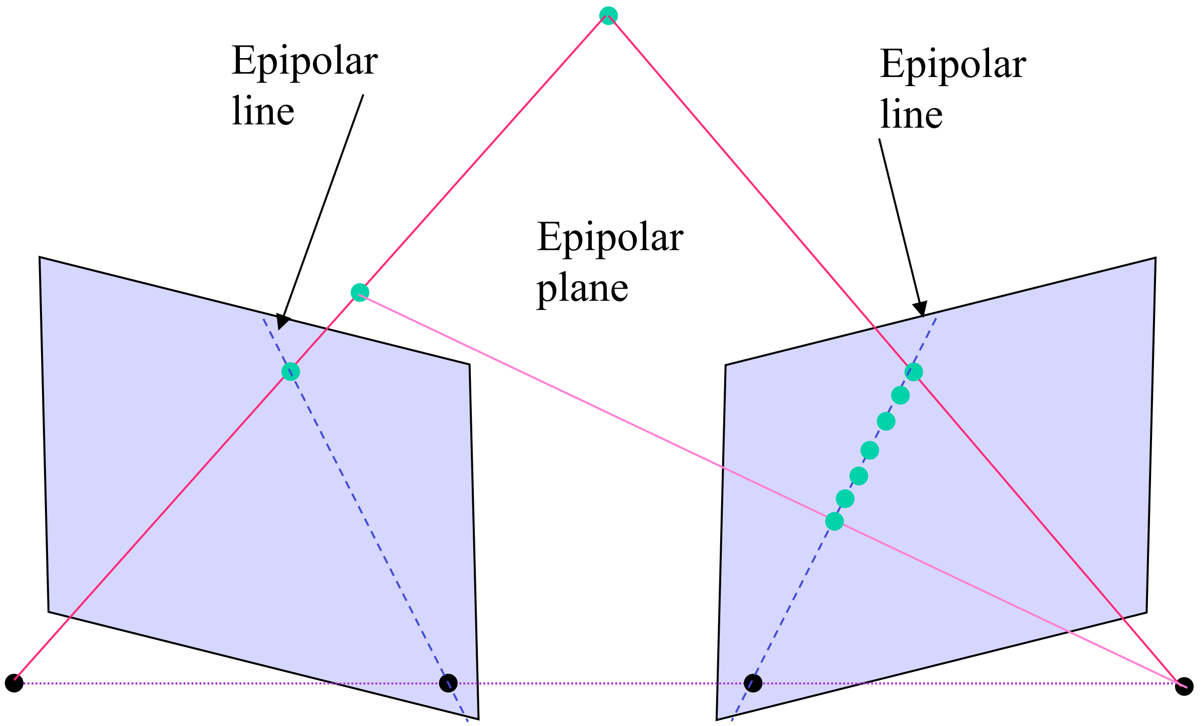
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

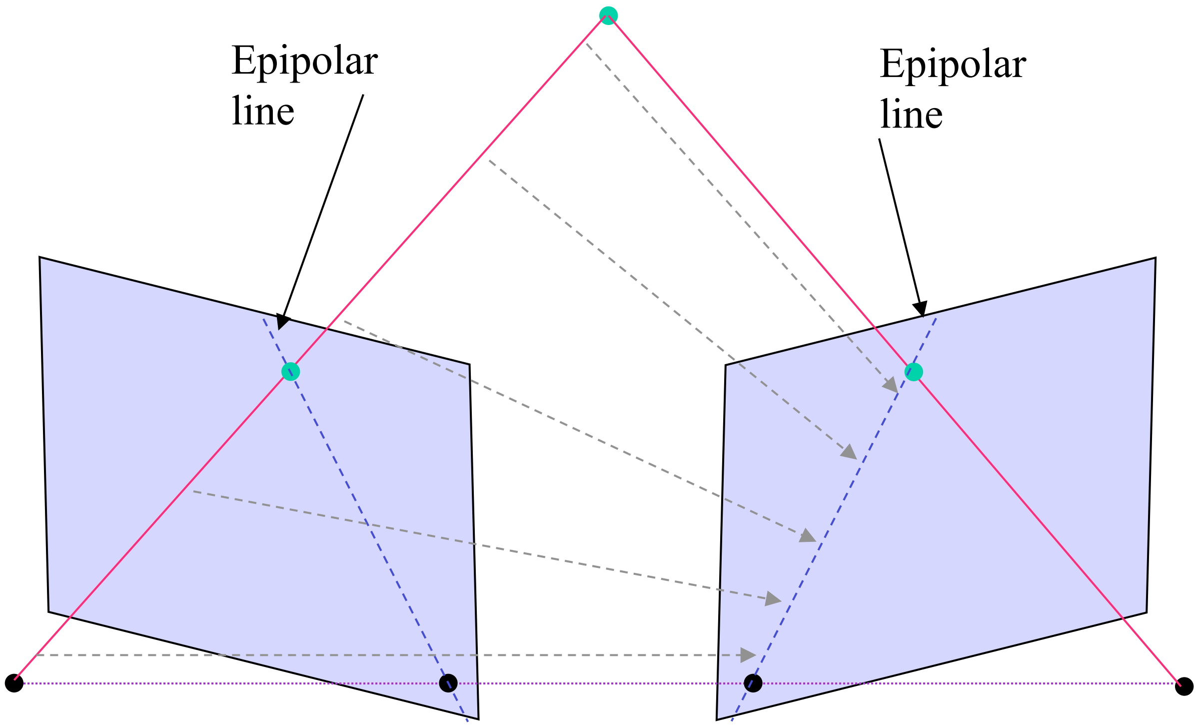
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

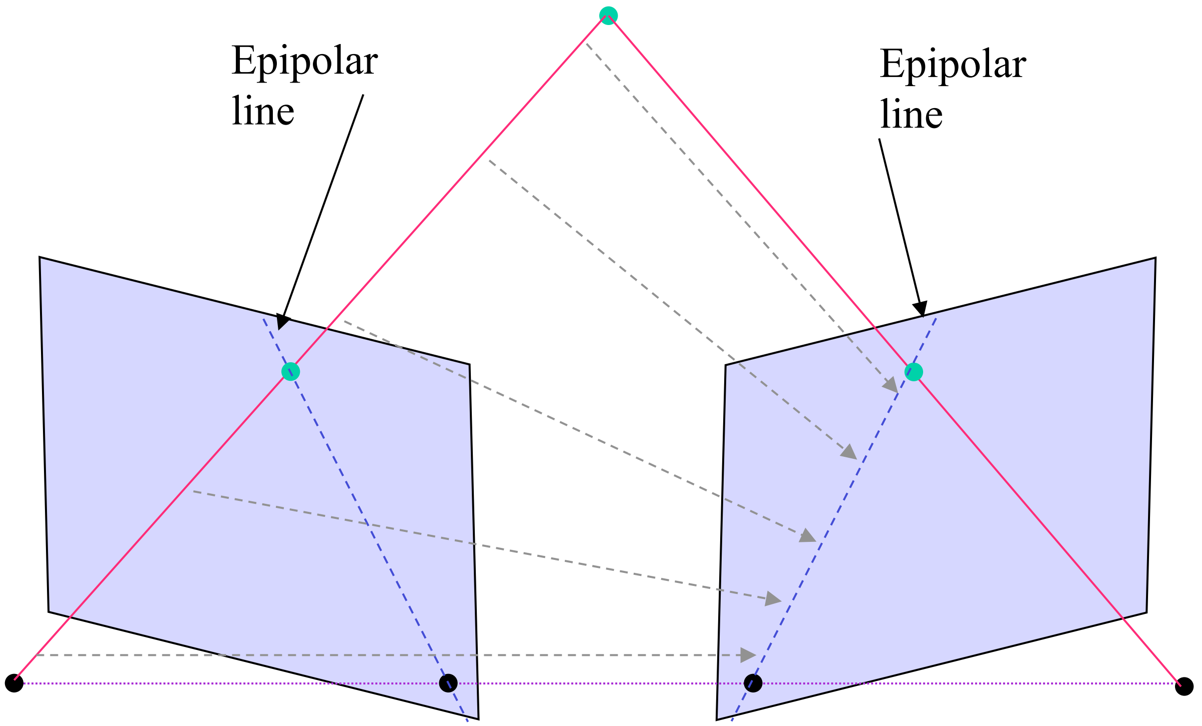
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



Epipolar geometry: Pictorial explanation

Given a stereo pair of cameras and a point in 3D space

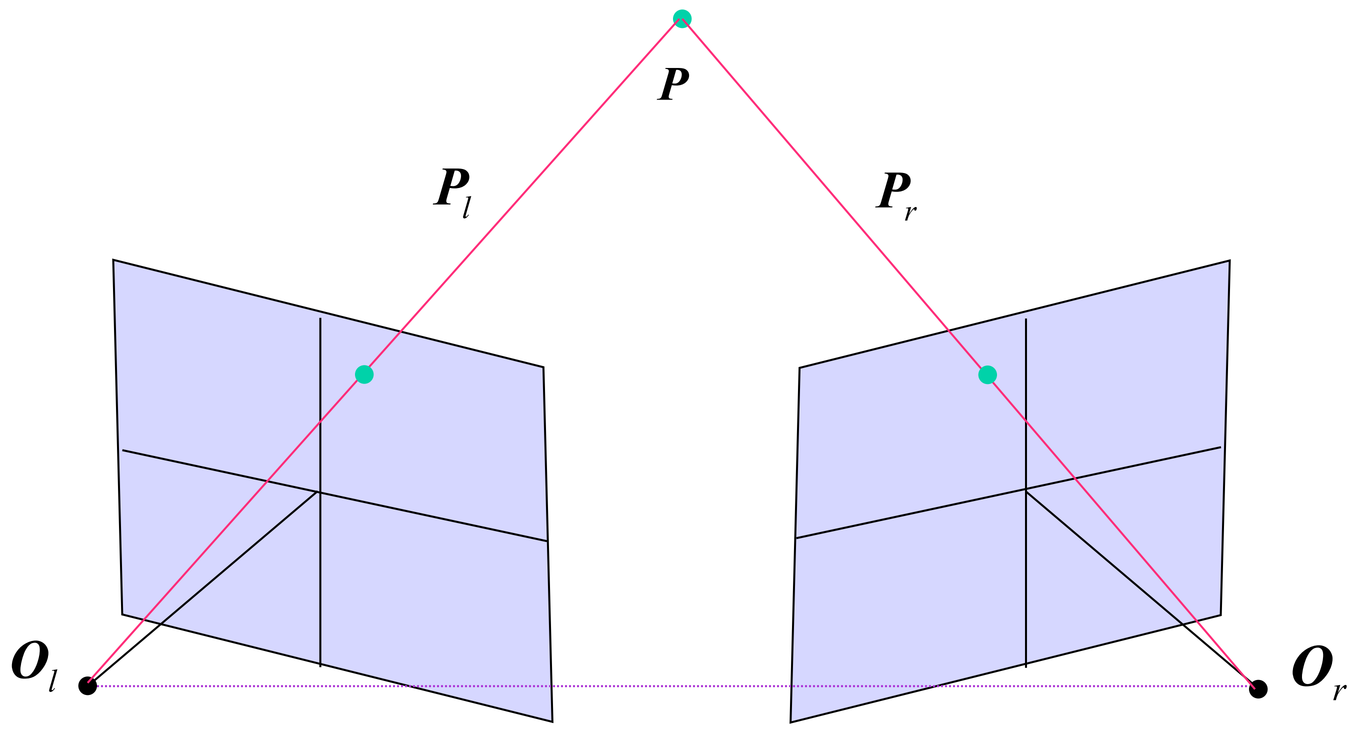
- There is plane that goes through the point and the centres of projection of the cameras
  - We call this plane the epipolar plane
- The lines where the plane intersects the the images are called conjugate epipolar lines



The epipolar constraint

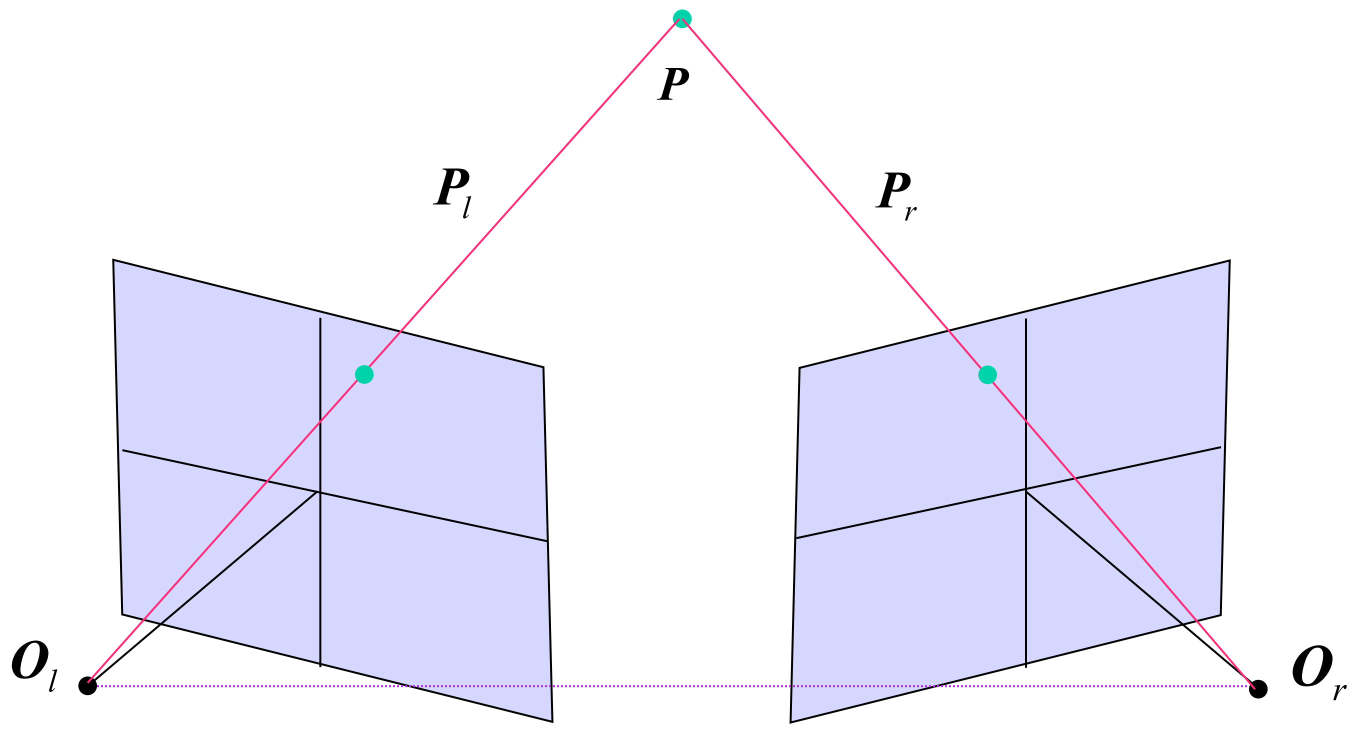
- Corresponding points must lie on conjugated epipolar lines.
- Stereo correspondence has been reduced to a 1D search!

Epipolar geometry: Analytic explanation



Equation of the epipolar plane

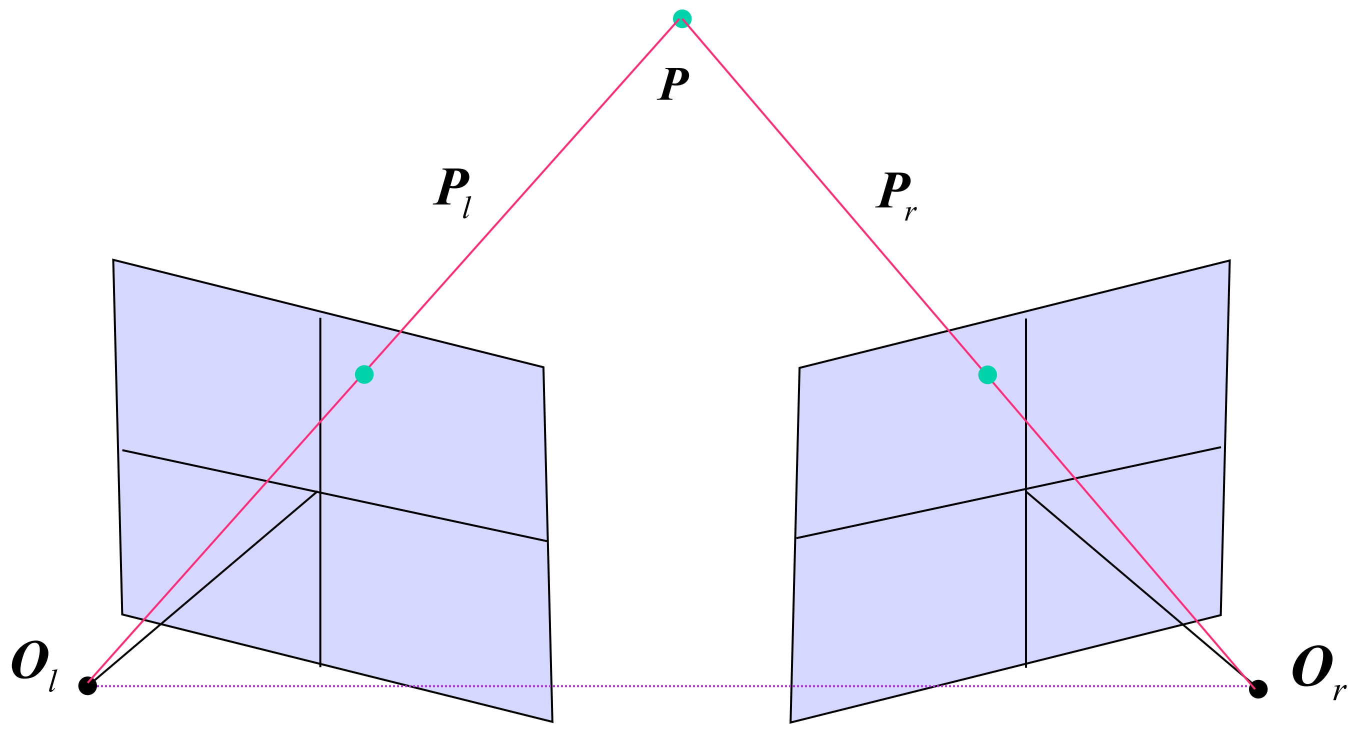
Epipolar geometry: Analytic explanation



Equation of the epipolar plane

- Let $T = O_r - O_l$ define the translation vector that shifts the left centre of projection, O_l, to the right centre of projection, O_r.

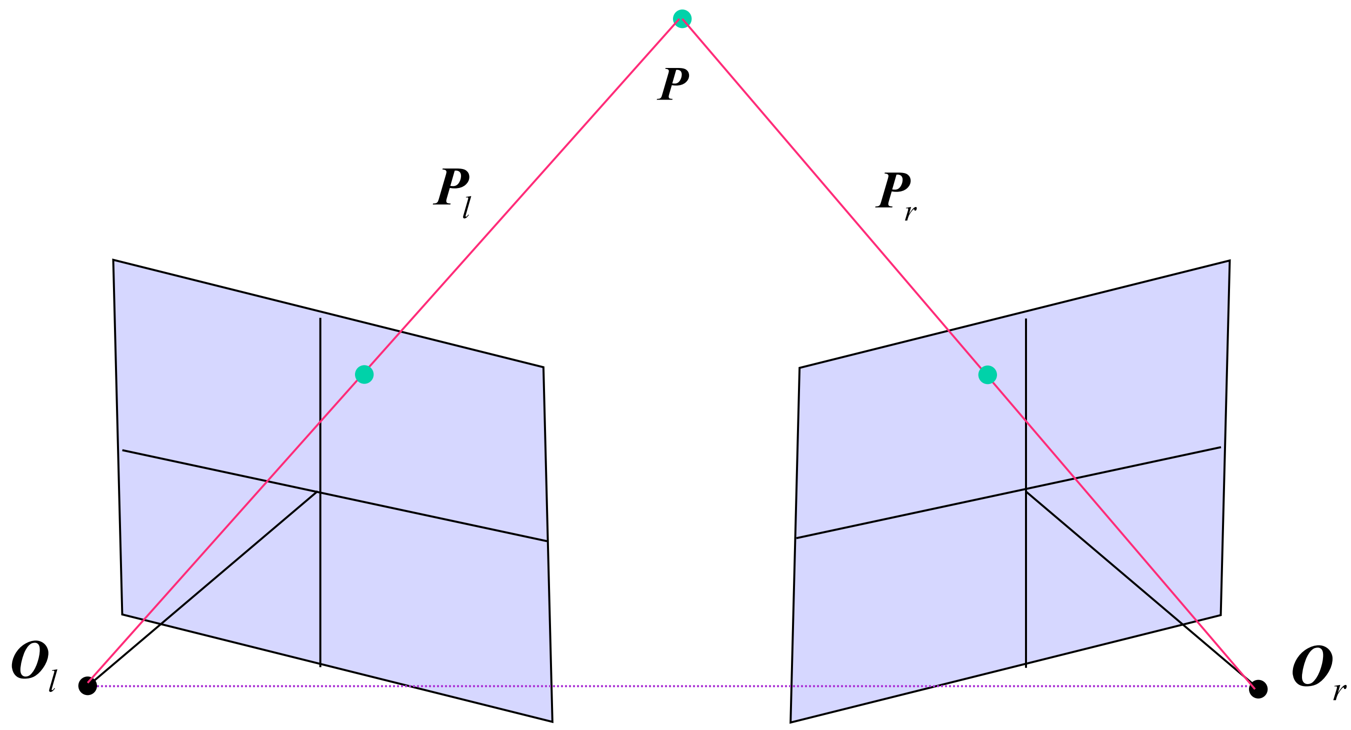
Epipolar geometry: Analytic explanation



Equation of the epipolar plane

- Let $\boldsymbol{T} = \boldsymbol{O}_r - \boldsymbol{O}_l$ define the translation vector that shifts the left centre of projection, $\boldsymbol{O}_l$, to the right centre of projection, $\boldsymbol{O}_r$.
- By definition of the cross product, $\boldsymbol{T} \times \boldsymbol{P}_l$ defines a normal to the plane defined by $\boldsymbol{T}$ and the coordinate of the point of regard in the left system, $\boldsymbol{P}_l$.

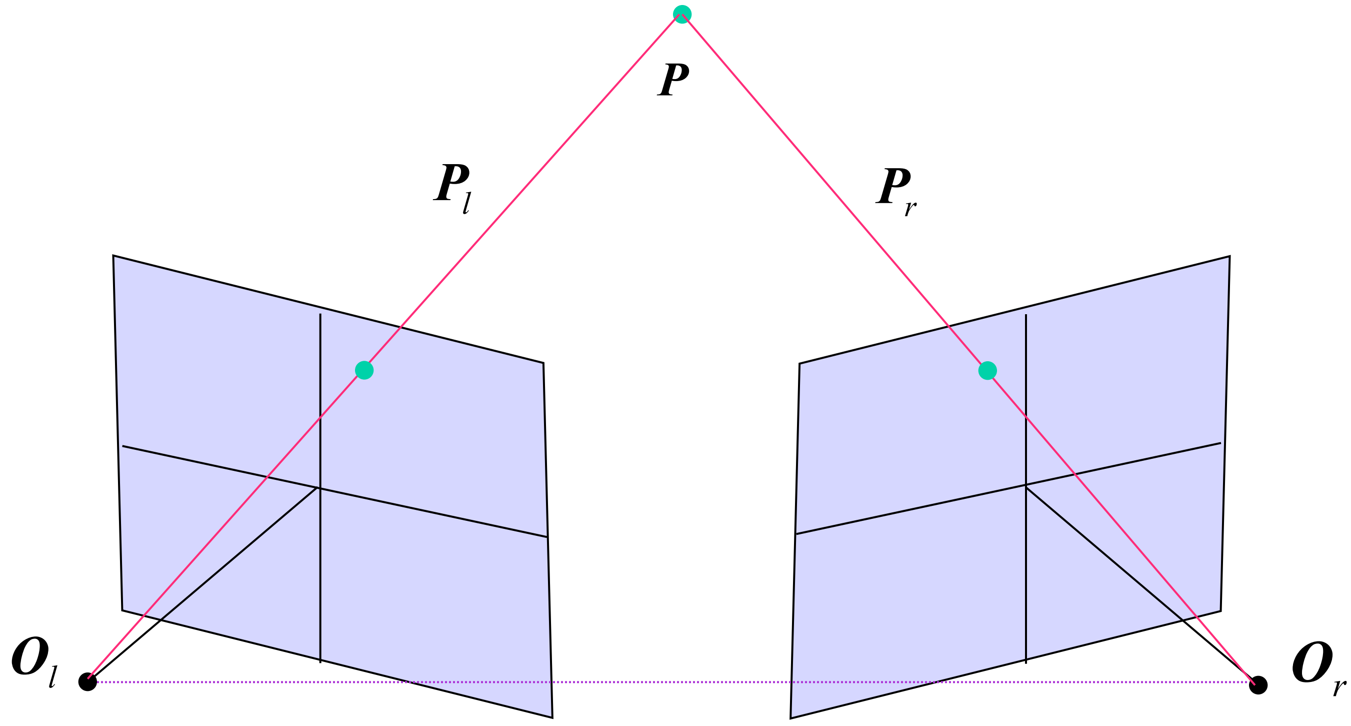
Epipolar geometry: Analytic explanation



Equation of the epipolar plane

- Let $T = O_r - O_l$ define the translation vector that shifts the left centre of projection, O_l, to the right centre of projection, O_r.
- By definition of the cross product, $T \times P_l$ defines a normal to the plane defined by T and the coordinate of the point of regard in the left system, P_l.
- For any other vector in the epipolar plane, its projection on the normal must be 0.
  - One such vector is given by $P_l - T$

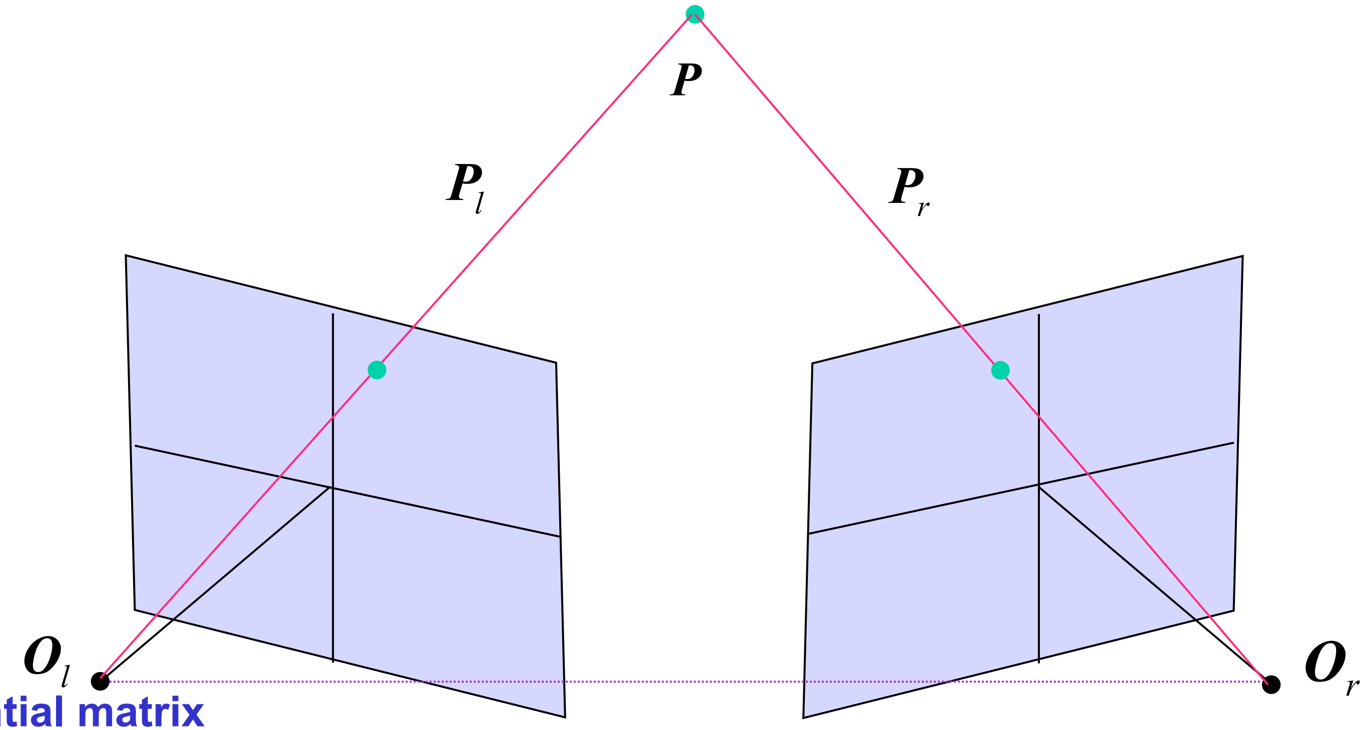
Epipolar geometry: Analytic explanation



Equation of the epipolar plane

- Let $\boldsymbol{T} = \boldsymbol{O}_r - \boldsymbol{O}_l$ define the translation vector that shifts the left centre of projection, $\boldsymbol{O}_l$, to the right centre of projection, $\boldsymbol{O}_r$.
- By definition of the cross product, $\boldsymbol{T} \times \boldsymbol{P}_l$ defines a normal to the plane defined by $\boldsymbol{T}$ and the coordinate of the point of regard in the left system, $\boldsymbol{P}_l$.
- For any other vector in the epipolar plane, its projection on the normal must be 0.
  - One such vector is given by $\boldsymbol{P}_l - \boldsymbol{T}$
- Combining all these observations allows us to write the equation of the epipolar plane in terms of the coplanarity condition $\left(\boldsymbol{P}_l - \boldsymbol{T}\right)^{\mathsf{T}} \boldsymbol{T} \times \boldsymbol{P}_l = 0$

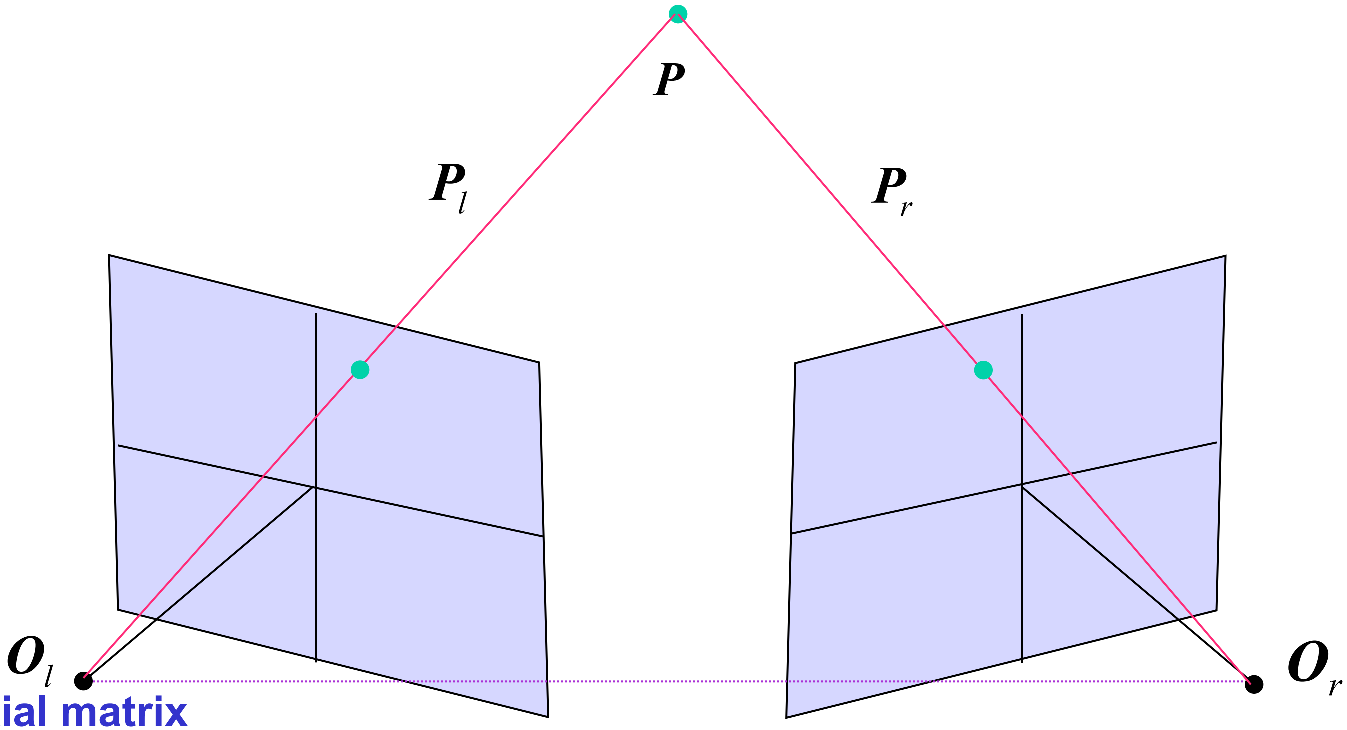
Epipolar geometry: Analytic explanation



The essential matrix

- Let
  - $\boldsymbol{T} = \boldsymbol{O}_r - \boldsymbol{O}_l$ define the translation vector that shifts the left centre of projection, $\boldsymbol{O}_l$, to the right centre of projection, $\boldsymbol{O}_r$.
  - **R** define the rotation matrix that aligns directions of the coordinate axes in the left and right coordinate systems
  - $\boldsymbol{P}_r = \mathbf{R}(\boldsymbol{P}_l - \boldsymbol{T})$ define the transformation between the coordinates of $\boldsymbol{P}$ from the left to right coordinate systems.

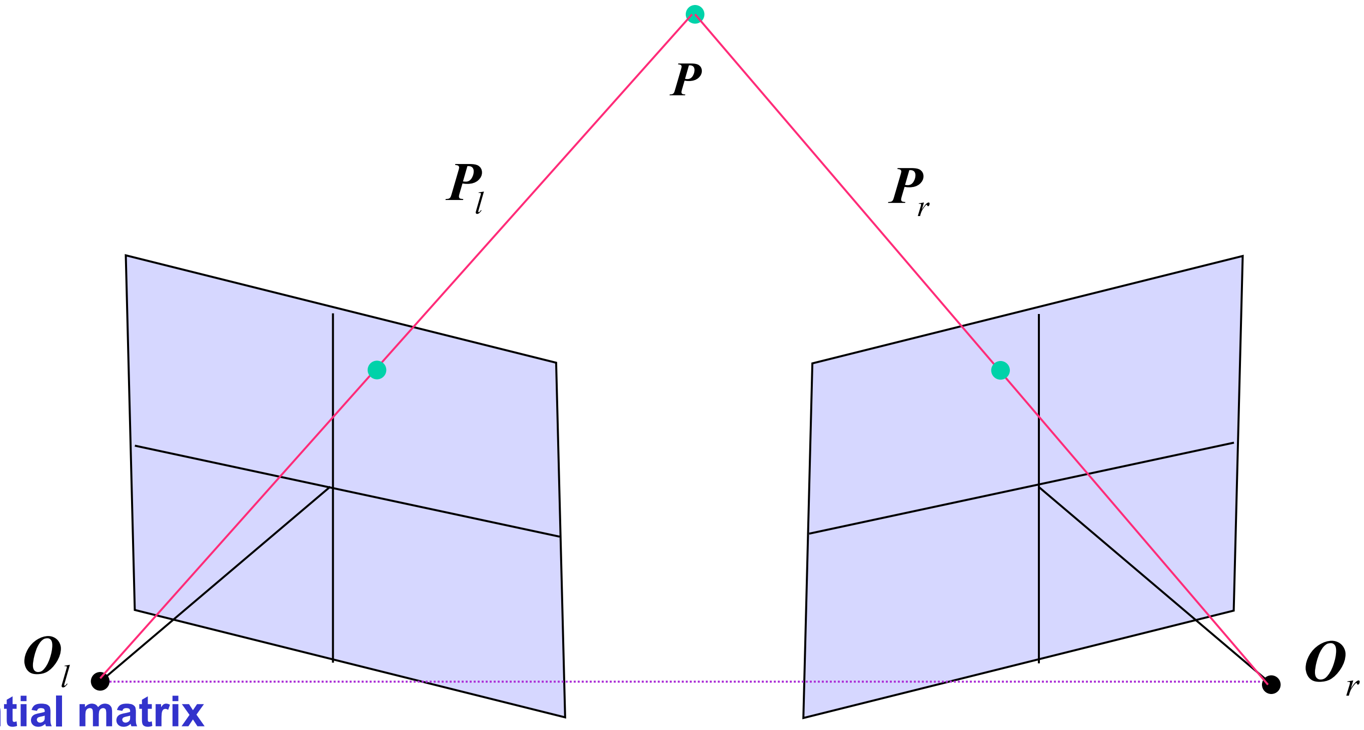
Epipolar geometry: Analytic explanation



The essential matrix

- Let
  - $\boldsymbol{T} = \boldsymbol{O}_r - \boldsymbol{O}_l$ define the translation vector that shifts the left centre of projection, $\boldsymbol{O}_l$, to the right centre of projection, $\boldsymbol{O}_r$.
  - **R** define the rotation matrix that aligns directions of the coordinate axes in the left and right coordinate systems
  - $\boldsymbol{P}_r = \mathbf{R}(\boldsymbol{P}_l - \boldsymbol{T})$ define the transformation between the coordinates of $\boldsymbol{P}$ from the left to right coordinate systems.
- We can now rewrite the coplanarity condition $(\boldsymbol{P}_l - \boldsymbol{T})^{\mathsf{T}} \boldsymbol{T} \times \boldsymbol{P}_l = 0$ as $(\mathbf{R}^{\mathsf{T}} \boldsymbol{P}_r)^{\mathsf{T}} \boldsymbol{T} \times \boldsymbol{P}_l = 0$

Epipolar geometry: Analytic explanation



The essential matrix

- Let
  - $T = O_r - O_l$ define the translation vector that shifts the left centre of projection, O_l, to the right centre of projection, O_r.
  - **R** define the rotation matrix that aligns directions of the coordinate axes in the left and right coordinate systems
  - $P_r = \mathbf{R}(P_l - T)$ define the transformation between the coordinates of P from the left to right coordinate systems.
- We can now rewrite the coplanarity condition $(P_l - T)^{\mathsf{T}} T \times P_l = 0$ as $(\mathbf{R}^{\mathsf{T}} P_r)^{\mathsf{T}} T \times P_l = 0$
- Noting that $T \times P_l = \mathbf{S}P_l$ for skew symmetric **S**, we further rewrite the coplanarity condition as.

$$(\mathbf{R}^{\mathsf{T}} P_r)^{\mathsf{T}} \mathbf{S} P_l = 0$$

Epipolar geometry: Analytic explanation

Interlude: Representing cross-products in terms of skew symmetric matrices

- For 3 x 1 $\boldsymbol{T} = (t_x, t_y, t_z)^{\mathsf{T}}$ and $\boldsymbol{P} = (X, Y, Z)^{\mathsf{T}}$ we define

$$\mathrm{T} \times \mathrm{P} = \begin{pmatrix} -t_z Y + t_y Z \\ t_z X - t_x Z \\ -t_y X + t_x Y \end{pmatrix}$$

- We notice that this calculation can be encapsulated in a matrix operation of the form

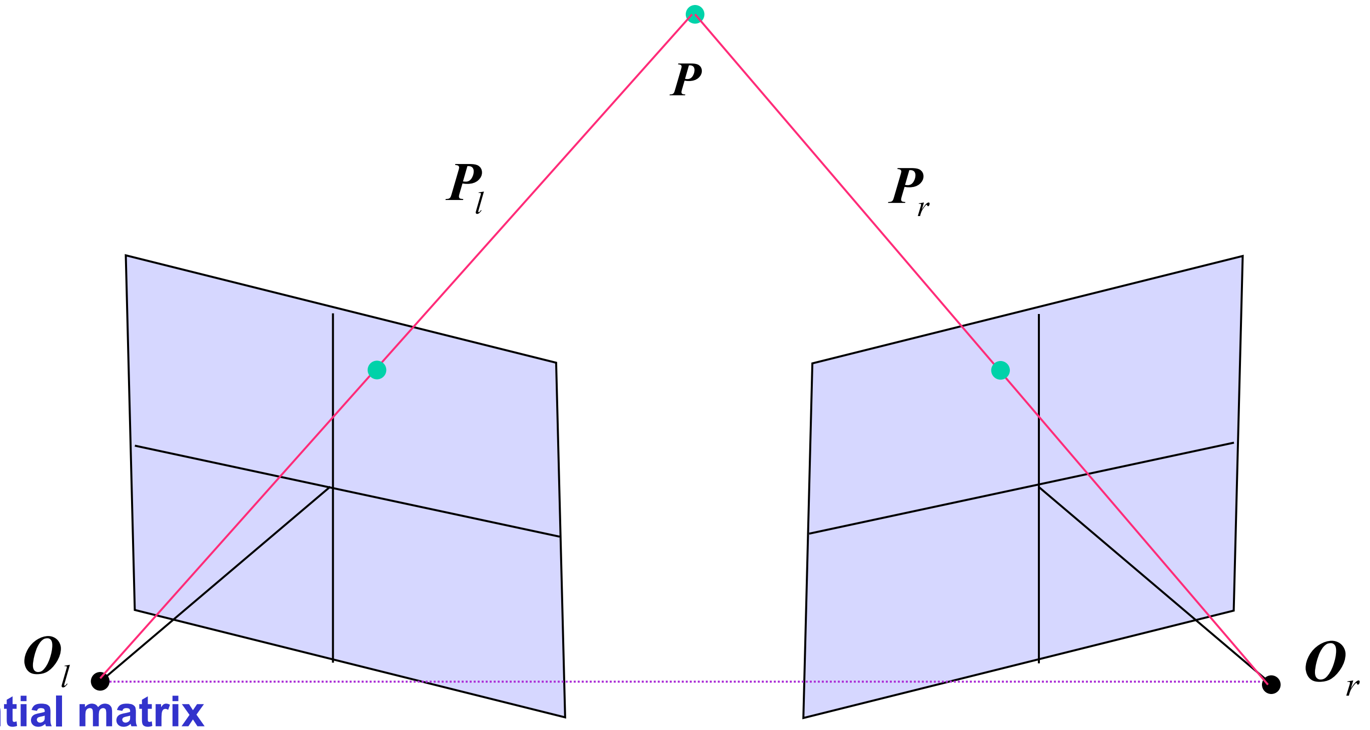
$$\begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

- Letting

$$\mathbf{S} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$$

- We have $\boldsymbol{T} \times \boldsymbol{P} = \mathbf{S}\,\boldsymbol{P}$
- Remark: We say that S is skew symmetric in that $\mathbf{S} = -\mathbf{S}^{\mathsf{T}}$

Epipolar geometry: Analytic explanation

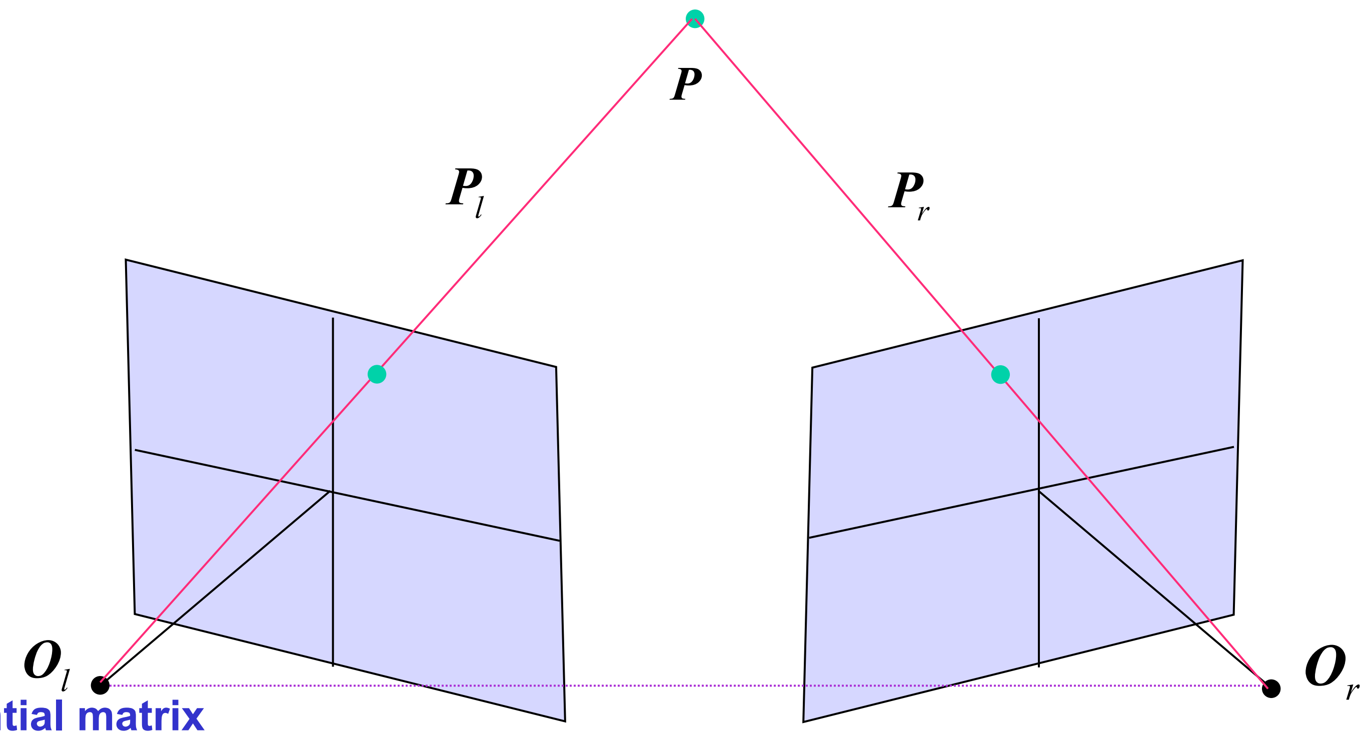


The essential matrix

- The coplanarity condition has been rewritten as

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l$$

Epipolar geometry: Analytic explanation



The essential matrix

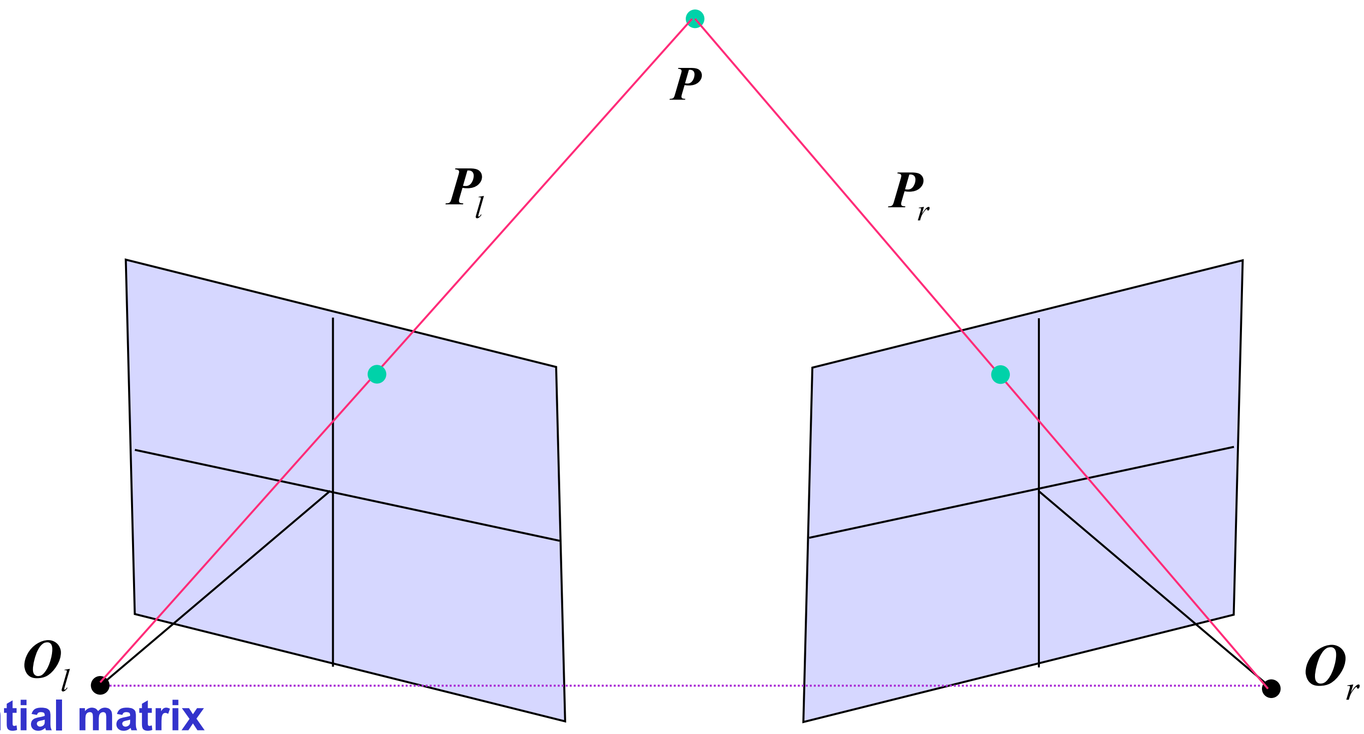
- The coplanarity condition has been rewritten as

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l$$

- Making use of rules of matrix manipulation we rearrange so that the left and right coordinates of P are on either side of an inner product of matrices **RS**

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l = P_r^{\mathsf{T}} \mathbf{R} \mathbf{S} P_l$$

Epipolar geometry: Analytic explanation



The essential matrix

- The coplanarity condition has been rewritten as

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l$$

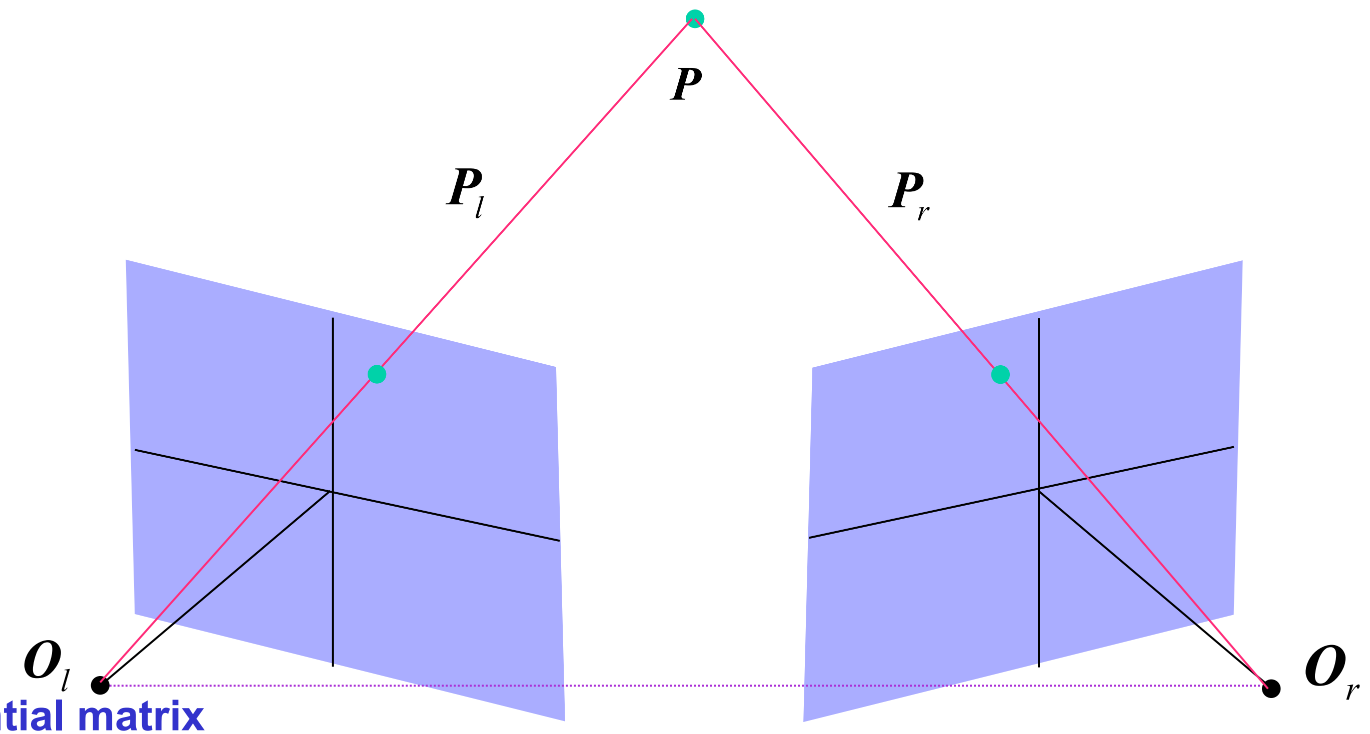
- Making use of rules of matrix manipulation we rearrange so that the left and right coordinates of P are on either side of an inner product of matrices **RS**

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l = P_r^{\mathsf{T}} \mathbf{R}\mathbf{S} P_l$$

- Letting **E**=**RS** we have

$$0 = \left(\mathbf{R}^{\mathsf{T}} P_r\right)^{\mathsf{T}} \mathbf{S} P_l = P_r^{\mathsf{T}} \mathbf{R}\mathbf{S} P_l = P_r^{\mathsf{T}} \mathbf{E} P_l$$

Epipolar geometry: Analytic explanation

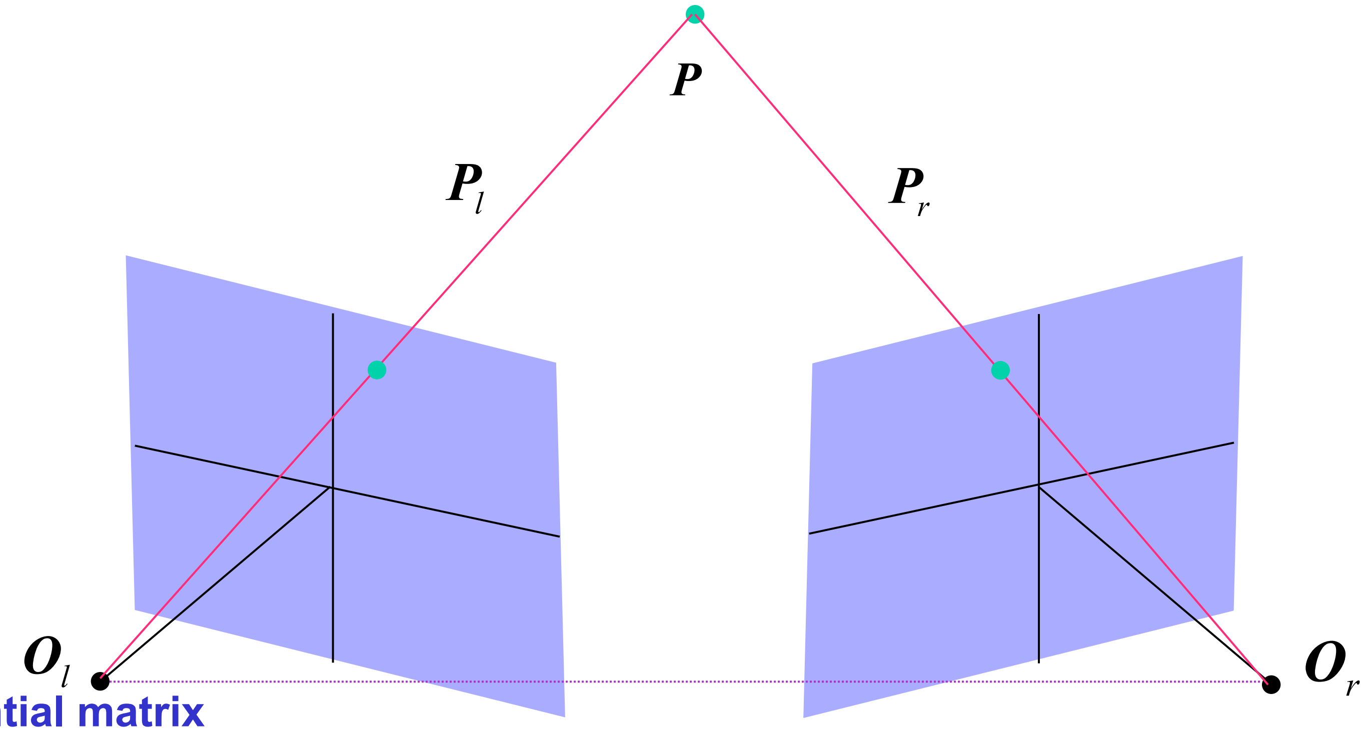


The essential matrix

- In terms of left and right coordinates of $\boldsymbol{P}$ we have reduced the coplanarity condition to

$$\boldsymbol{P}_r^{\mathsf{T}}\mathbf{E}\boldsymbol{P}_l = 0$$

Epipolar geometry: Analytic explanation



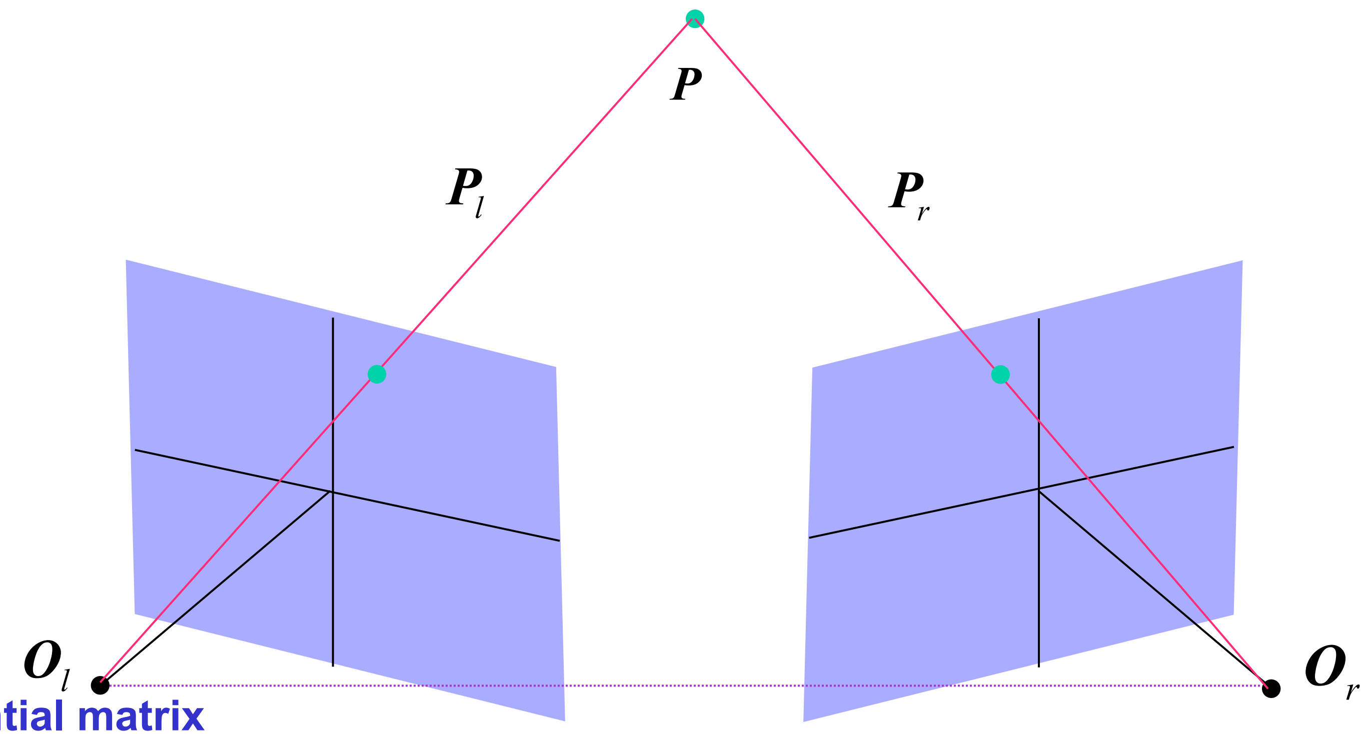
The essential matrix

- In terms of left and right coordinates of $\boldsymbol{P}$ we have reduced the coplanarity condition to

$$\boldsymbol{P}_r^{\mathsf{T}}\mathbf{E}\boldsymbol{P}_l = 0$$

- We complete our derivation by converting to camera coodinates in the image plane, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$

Epipolar geometry: Analytic explanation



The essential matrix

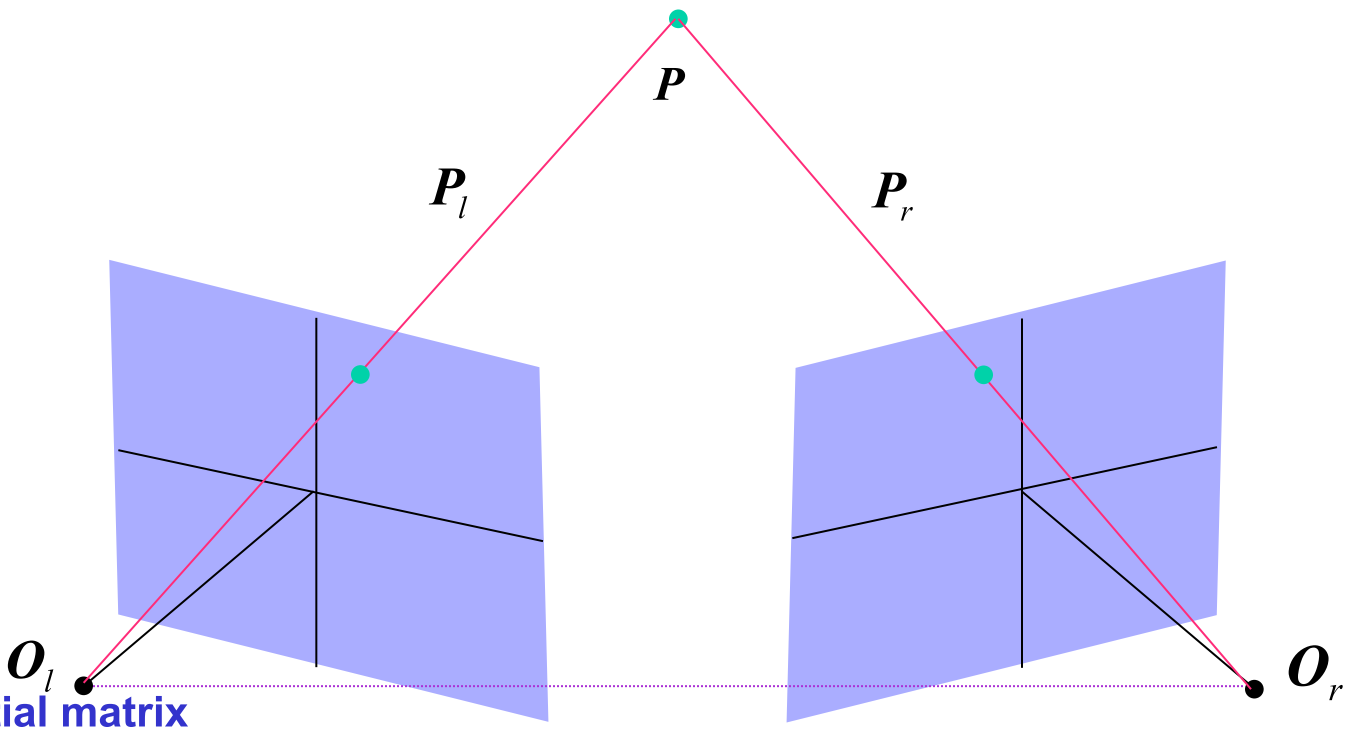
- In terms of left and right coordinates of $\boldsymbol{P}$ we have reduced the coplanarity condition to

$$\boldsymbol{P}_r^{\mathsf{T}}\mathbf{E}\boldsymbol{P}_l = 0$$

- We complete our derivation by converting to camera coodinates in the image plane, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$
- From perspective we recall that

$$\boldsymbol{p}_l = f\,\frac{\boldsymbol{P}_l}{Z_l}, \quad \boldsymbol{p}_r = f\,\frac{\boldsymbol{P}_r}{Z_r}$$

Epipolar geometry: Analytic explanation



The essential matrix

- In terms of left and right coordinates of $\boldsymbol{P}$ we have reduced the coplanarity condition to

$$\boldsymbol{P}_r^{\mathsf{T}}\mathbf{E}\boldsymbol{P}_l = 0$$

- We complete our derivation by converting to camera coodinates in the image plane, $\boldsymbol{p}_l, \boldsymbol{p}_r$
- From perspective we recall that

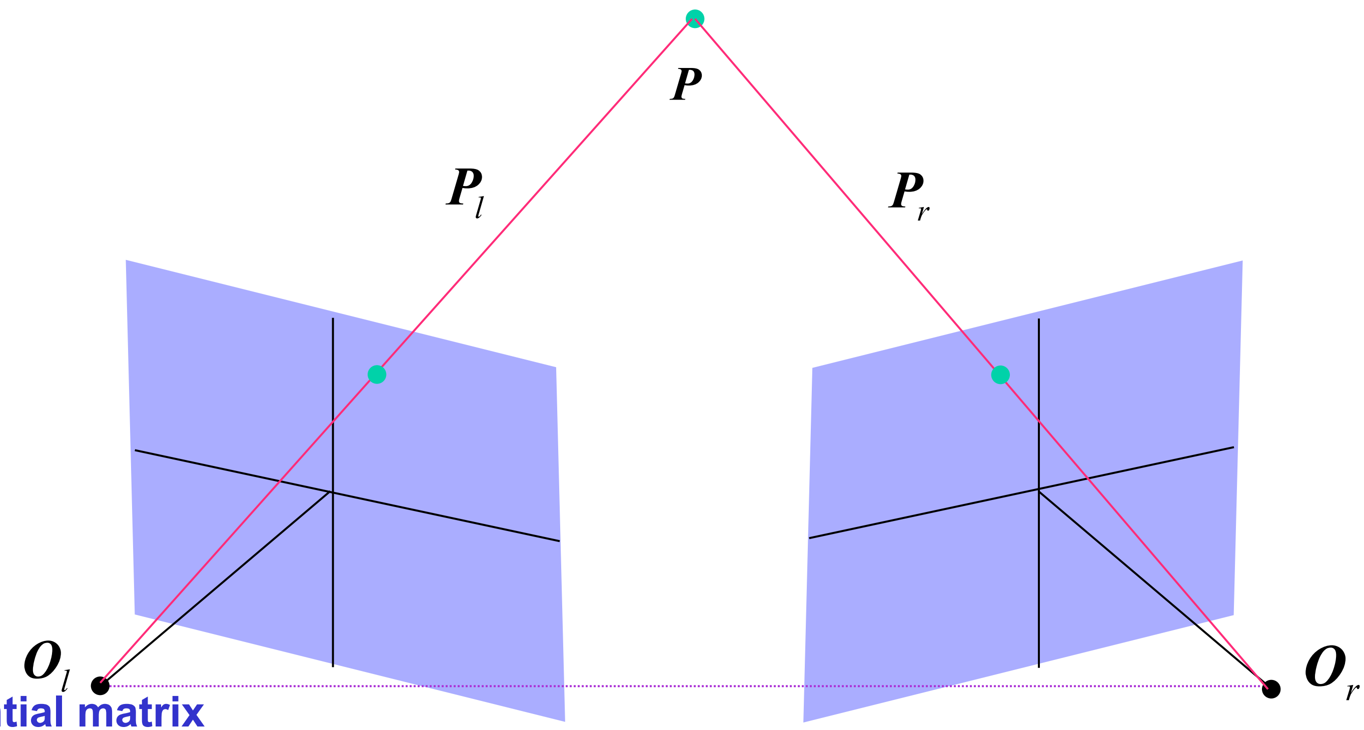
$$\boldsymbol{p}_l = f\frac{\boldsymbol{P}_l}{Z_l}, \quad \boldsymbol{p}_r = f\frac{\boldsymbol{P}_r}{Z_r}$$

- Substitution and division through by $f^2 / Z_l Z_r$ yields
-

$$\boldsymbol{p}_r^{\mathsf{T}}\mathbf{E}\boldsymbol{p}_l = 0$$

- Which expresses a fundamental constraint on any two image points, $\boldsymbol{p}_l, \boldsymbol{p}_r$, that are in binocular correspondence.

Epipolar geometry: Analytic explanation



The essential matrix

- In terms of left and right coordinates of P we have reduced the coplanarity condition to

$$P_r^{\mathsf{T}} \mathbf{E} P_l = 0$$

- We complete our derivation by converting to camera coodinates in the image plane, p_l, p_r
- From perspective we recall that

$$p_l = f \frac{P_l}{Z_l}, \quad p_r = f \frac{P_r}{Z_r}$$

- Substitution and division through by $f^2 / Z_l Z_r$ yields
- $$p_r^{\mathsf{T}} \mathbf{E} p_l = 0$$

- Which expresses a fundamental constraint on any two image points, p_l, p_r, that are in binocular correspondence. We refer to **E** as the essential matrix.

Epipolar geometry: Analytic explanation

The fundamental matrix

- Use of the essential matrix to relate corresponding image points in the left and right views allows us to write

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{E} \boldsymbol{p}_l = 0$$

assuming that we can measure image points in camera coordinates, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$, rather than pixel coordinates, $\tilde{\boldsymbol{p}}_l$, $\tilde{\boldsymbol{p}}_r$.

Epipolar geometry: Analytic explanation

The fundamental matrix

- Use of the essential matrix to relate corresponding image points in the left and right views as

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{E} \boldsymbol{p}_l = 0$$

assuming that we can measure image points in camera coordinates, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$, rather than pixel coordinates, $\tilde{\boldsymbol{p}}_l$, $\tilde{\boldsymbol{p}}_r$.

- Recall (from Unit 1) that camera and pixel coordinates are related via the matrices of intrinsic parameters.
- Let $\mathbf{M}_l, \mathbf{M}_r$ be the intrinsic camera parameter matrices for the left and right systems, resp. We have

$$\boldsymbol{p}_l = \mathbf{M}_l^{-1} \tilde{\boldsymbol{p}}_l, \quad \boldsymbol{p}_r = \mathbf{M}_r^{-1} \tilde{\boldsymbol{p}}_r$$

Epipolar geometry: Analytic explanation

The fundamental matrix

- Use of the essential matrix to relate corresponding image points in the left and right views as

$$\mathbf{p}_r^{\mathsf{T}} \mathbf{E} \mathbf{p}_l = 0$$

  assuming that we can measure image points in camera coordinates, $\mathbf{p}_l, \mathbf{p}_r$, rather than pixel coordinates, $\tilde{\mathbf{p}}_l, \tilde{\mathbf{p}}_r$.
- Recall (from Unit 1) that camera and pixel coordinates are related via the matrices of intrinsic parameters.
- Let $\mathbf{M}_l, \mathbf{M}_r$ be the intrinsic camera parameter matrices for the left and right systems, resp. We have

$$\mathbf{p}_l = \mathbf{M}_l^{-1} \tilde{\mathbf{p}}_l, \quad \mathbf{p}_r = \mathbf{M}_r^{-1} \tilde{\mathbf{p}}_r$$

- We can make use of these transformations from camera to pixel coordinates to rewrite our correspondence constraint equation as

$$\tilde{\mathbf{p}}_r^{\mathsf{T}} \mathbf{F} \tilde{\mathbf{p}}_l = 0$$

  where

$$\mathbf{F} = \mathbf{M}_r^{-\mathsf{T}} \mathbf{E} \mathbf{M}_l^{-1}$$

  is referred to as the fundamental matrix.

Epipolar geometry: Recovery

Problem statement

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

Two approaches

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

Epipolar geometry: Recovery

The 8 point algorithm

- For any pair of corresponding pixel coordinate-based features, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$, appeal to the epipolar constraint

$$\boldsymbol{p}_r^{\mathsf{T}}\mathbf{F}\boldsymbol{p}_l = 0$$

allows us to write one equation in the unknown components of the fundamental matrix, **F**.

Epipolar geometry: Recovery

The 8 point algorithm

- For any pair of corresponding pixel coordinate-based features, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$, appeal to the epipolar constraint

$$\boldsymbol{p}_r^{\mathrm{T}}\mathbf{F}\boldsymbol{p}_l = 0$$

allows us to write one equation in the unknown components of the fundamental matrix, **F**.

- In particular, letting

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}, \quad \boldsymbol{p}_l = \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix}, \quad \boldsymbol{p}_r = \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix}$$

we have

$$x_r x_l f_{11} + x_r y_l f_{12} + x_r f_{13} + y_r x_l f_{21} + y_r y_l f_{22} + y_r f_{23} + x_l f_{31} + y_l f_{32} + f_{33} = 0$$

Epipolar geometry: Recovery

The 8 point algorithm

- For any pair of corresponding pixel coordinate-based features, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$, appeal to the epipolar constraint

$$\boldsymbol{p}_r^{\mathsf{T}}\mathbf{F}\boldsymbol{p}_l = 0$$

allows us to write one equation in the unknown components of the fundamental matrix, **F**.

- In particular, letting

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}, \quad \boldsymbol{p}_l = \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix}, \quad \boldsymbol{p}_r = \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix}$$

we have

$$x_r x_l f_{11} + x_r y_l f_{12} + x_r f_{13} + y_r x_l f_{21} + y_r y_l f_{22} + y_r f_{23} + x_l f_{31} + y_l f_{32} + f_{33} = 0$$

- Similarly, for a set of n correspondences we can write

$$\mathbf{A}\mathrm{f} = \begin{pmatrix} x_{r1}x_{l1} & x_{r1}y_{l1} & x_{r1} & y_{r1}x_{l1} & y_{r1}y_{l1} & y_{r1} & x_{l1} & y_{l1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{rn}x_{ln} & & & & \cdots & & & & 1 \end{pmatrix} \mathrm{f} = 0$$

where

$$\boldsymbol{f} = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} & f_{33} \end{pmatrix}^{\mathsf{T}}$$

Epipolar geometry: Recovery

The 8 point algorithm

- Since the assembled system of equations

$$\mathbf{A}\,\boldsymbol{f} = 0$$

is homogenous, there are only 8 independent components in $\boldsymbol{f}$.

→ If we have 8 (or more) corresponding features, then we have 8 (or more) constraint equations in 8 unknowns.

Epipolar geometry: Recovery

The 8 point algorithm

- Since the assembled system of equations

  $$\mathbf{A}\boldsymbol{f} = 0$$

  is homogenous, there are only 8 independent components in $\boldsymbol{f}$.

  → If we have 8 (or more) corresponding features, then we have 8 (or more) constraint equations in 8 unknowns.

- Using the constraint equations we can solve for the components of **F** using any reasonable method for solving a system of homogenous linear equations.

Epipolar geometry: Recovery

The 8 point algorithm

- Since the assembled system of equations

$$\mathbf{A}\boldsymbol{f} = 0$$

  is homogenous, there are only 8 independent components in $\boldsymbol{f}$.

  → If we have 8 (or more) corresponding features, then we have 8 (or more) constraint equations in 8 unknowns.

- Using the constraint equations we can solve for the components of **F** using any reasonable method for solving a system of homogenous linear equations.
  - One such method appeals to the singular value decomposition, which minimizes $|\mathbf{A}\boldsymbol{f}|$ subject to the constraint that $|\boldsymbol{f}| = 1$.
  - In particular, the solution for $\boldsymbol{f}$ is given by the vector corresponding to the smallest singular value in the decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{T}$$

    i.e., as the last column of **V**.

Epipolar geometry: Recovery

Interlude: The singular value decomposition (SVD)

- Any m x n matrix **A** can be factored into

$$\mathbf{A} = \mathbf{UDV}^{\mathsf{T}} = (\text{orthogonal})(\text{diagonal})(\text{orthogonal})$$

- Remarks
  - The columns of **U** (m x m) are the eigenvectors of $\mathbf{AA}^{\mathsf{T}}$
  - The columns of **V** (n x n) are the eigenvectors of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$
  - The matrix **D** (m x n) has nonzero values that are the square roots of the nonzero eignevalues of both $\mathbf{AA}^{\mathsf{T}}$ and $\mathbf{A}^{\mathsf{T}}\mathbf{A}$. These diagonal values, σ_i , are ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$.
- Most numerical linear algebra packages provide support for SVD decomposition.

Epipolar geometry: Recovery

The 8 point algorithm

- Since the assembled system of equations

$$\mathbf{A}\boldsymbol{f} = 0$$

  is homogenous, there are only 8 independent components in $\boldsymbol{f}$.

  → If we have 8 (or more) corresponding features, then we have 8 (or more) constraint equations in 8 unknowns.

- Using the constraint equations we can solve for the components of **F** using any reasonable method for solving a system of homogenous linear equations.
  - One such method appeals to the singular value decomposition, which minimizes $|\mathbf{A}\boldsymbol{f}|$ subject to the constraint that $|\boldsymbol{f}| = 1$.
  - In particular, the solution for $\boldsymbol{f}$ is given by the vector corresponding to the smallest singular value in the decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{T}$$

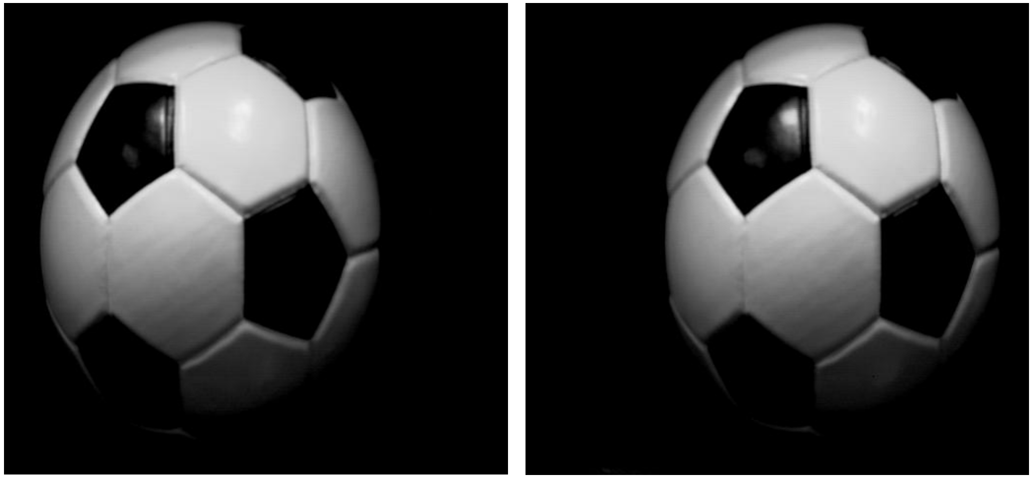
    i.e., as the last column of **V**.

- Remark: Given correspondences in terms of camera coordinates, we can perform similar calculations in terms of the essential matrix **E**.

Epipolar geometry: Recovery

8 point algorithm

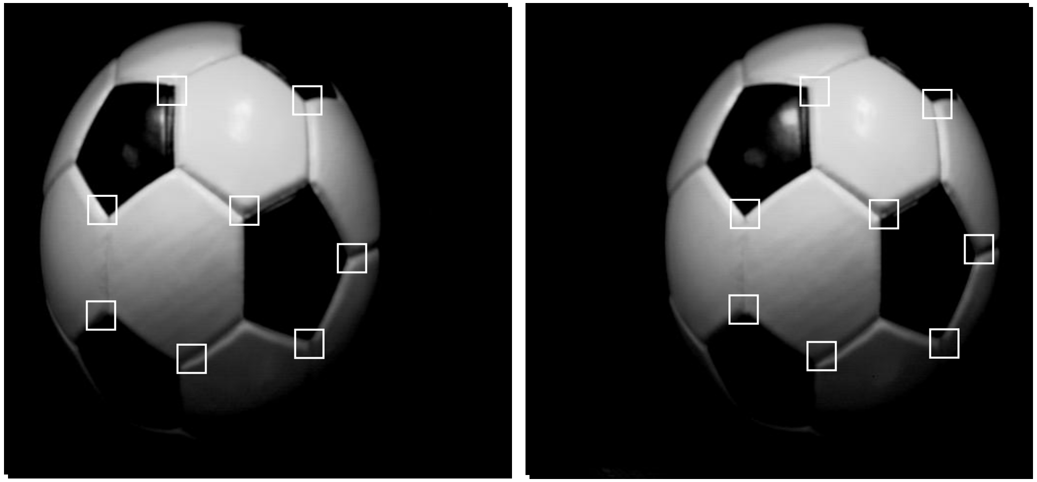
- To make the 8 point algorithm work, we must discover at least 8 corresponding features between a binocular image pair, …
- … without being able to avail ourselves to the epipolar constraint per se.



Epipolar geometry: Recovery

8 point algorithm

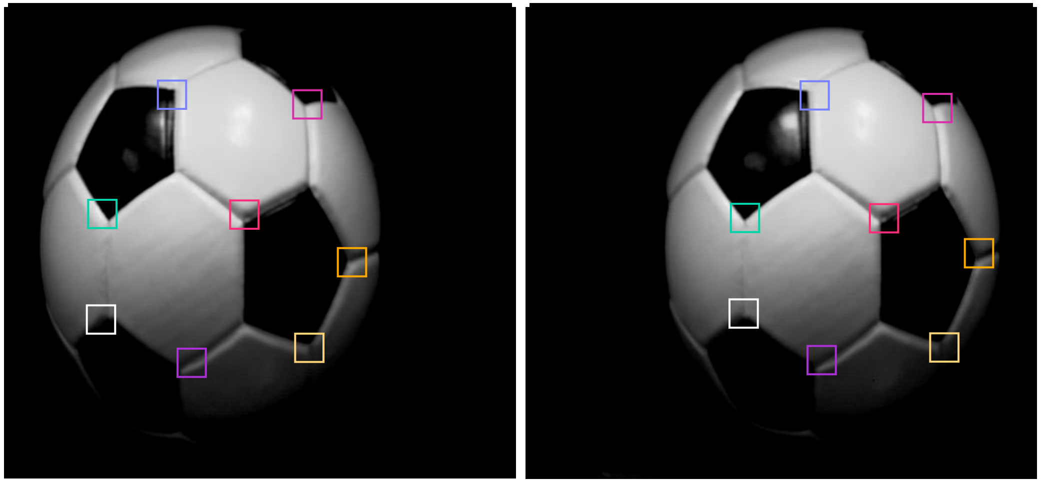
- To make the 8 point algorithm work, we must discover at least 8 corresponding features between a binocular image pair, …
- … without being able to avail ourselves to the epipolar constraint per se.
- Extract well localized distinctive features on each image.
  - e.g., corners



Epipolar geometry: Recovery

8 point algorithm

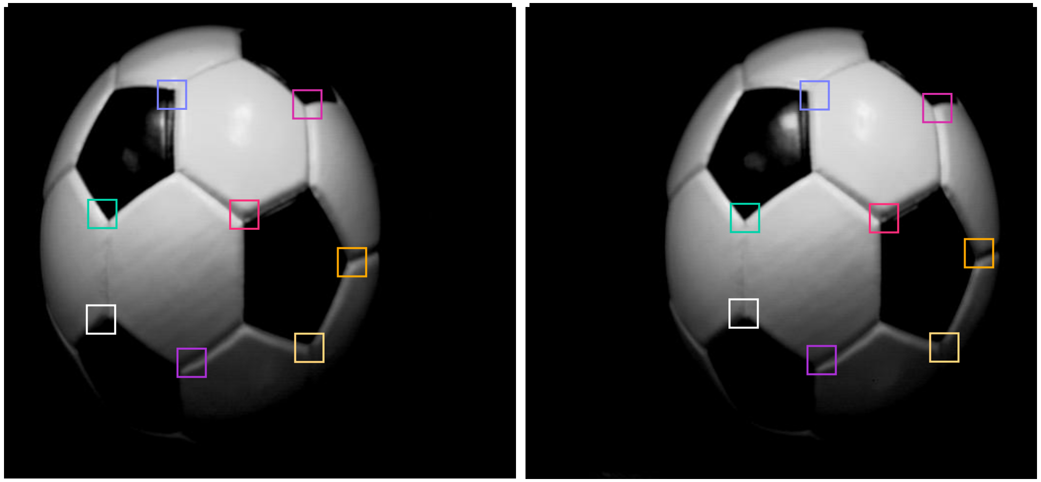
- To make the 8 point algorithm work, we must discover at least 8 corresponding features between a binocular image pair, …
- … without being able to avail ourselves to the epipolar constraint per se.
- Extract well localized distinctive features on each image.
  - e.g., corners
- Establish 8 (or more) precise correspondences between features across the image pair.
  - As general 2D search and match problem.



Epipolar geometry: Recovery

8 point algorithm

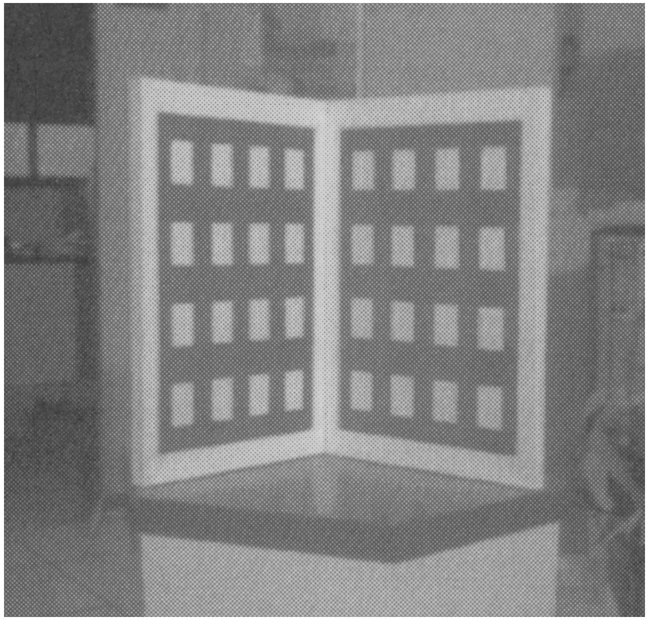
- To make the 8 point algorithm work, we must discover at least 8 corresponding features between a binocular image pair, …
- … without being able to avail ourselves to the epipolar constraint per se.
- Extract well localized distinctive features on each image.
  - e.g., corners
- Establish 8 (or more) precise correspondences between features across the image pair.
  - As general 2D search and match problem.
- Subsequently use the matches to form the system of linear constraint equations that are the basis of the 8 point algorithm.



Epipolar geometry: Recovery

Camera calibration

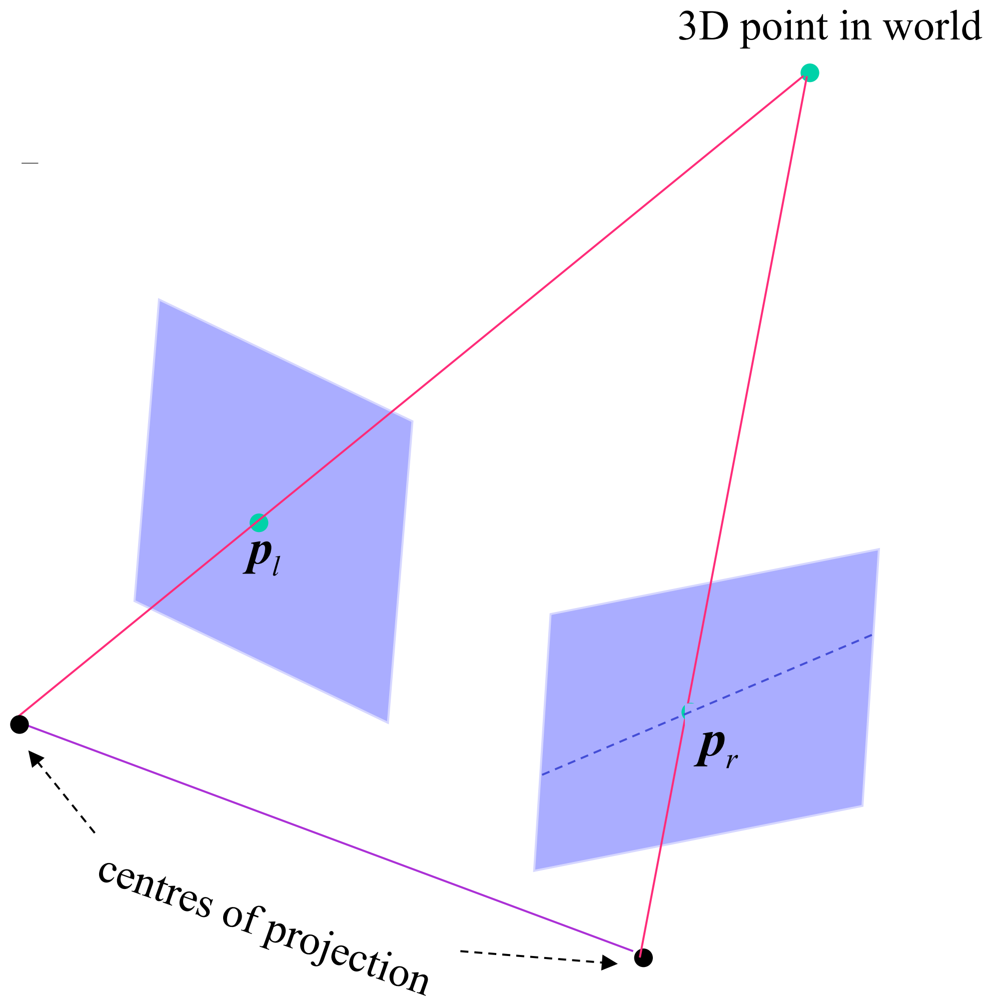
- Exploit an artificially constructed calibration pattern.
  - Designed to have features that can be precisely mensurated in 3D position.
- Capture images of the calibration pattern with both cameras.
- Precisely extract corresponding features between each image and the 3D calibration pattern.
  - Since 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.
- Remark: Camera calibration is covered in some detail in our textbook, chapter 6.



Epipolar geometry: Exploitation

Match constraint

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Use the recovered fundamental matrix **F** to constrain the search for matching points to a linear span..

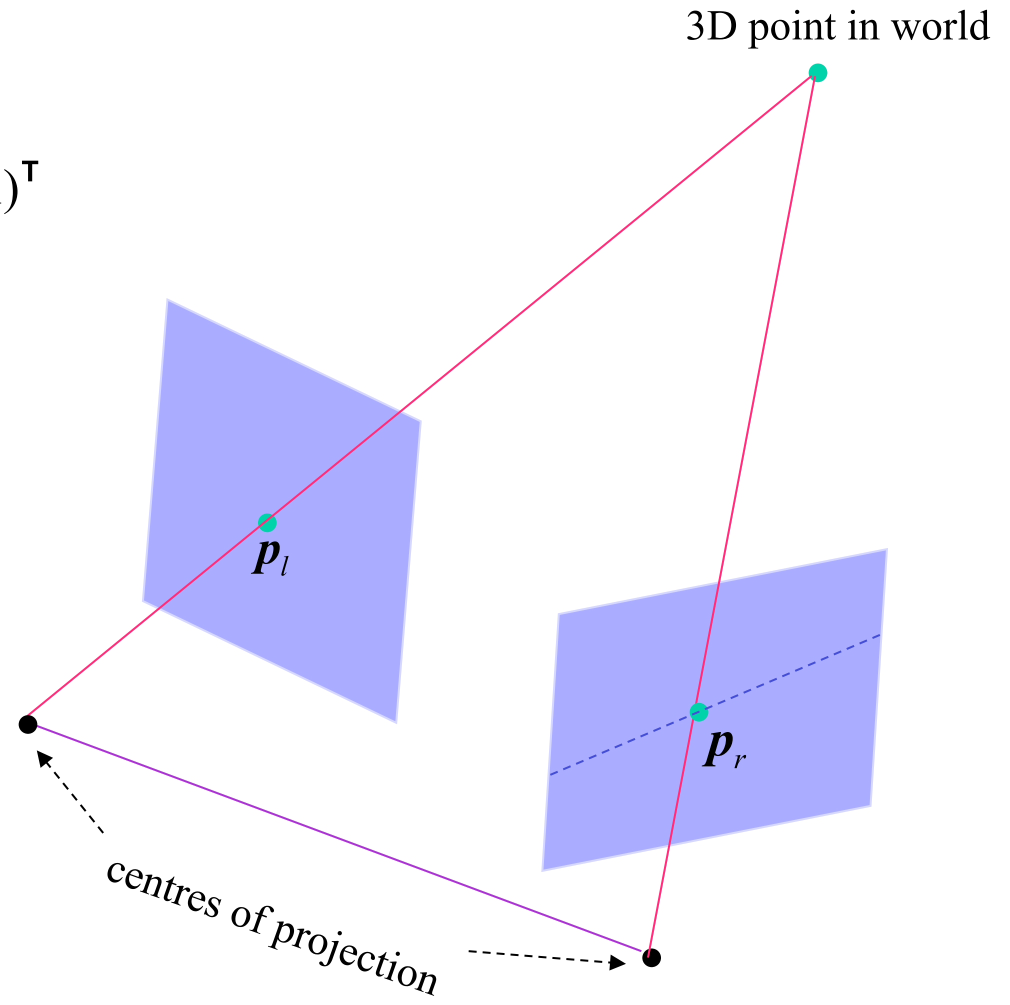


Epipolar geometry: Exploitation

Match constraint

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Use the recovered fundamental matrix **F** to constrain the search for matching points to a linear span..
- *Geometrically:* Given a point $\boldsymbol{p}_l = (x_l \quad y_l \quad 1)^{\mathsf{T}}$ in the left image, we can write

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{F} \boldsymbol{p}_l = 0$$



Epipolar geometry: Exploitation

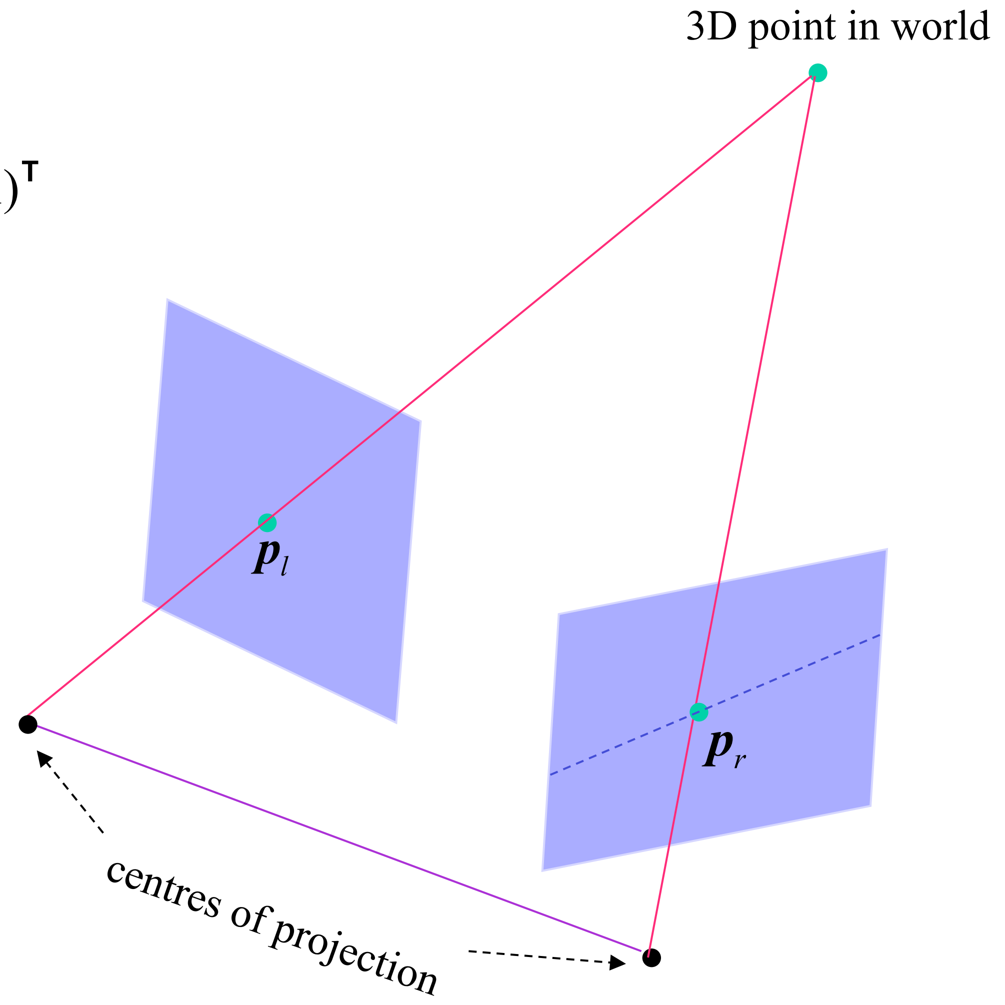
Match constraint

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Use the recovered fundamental matrix **F** to constrain the search for matching points to a linear span..
- *Geometrically:* Given a point $\boldsymbol{p}_l = (x_l \quad y_l \quad 1)^{\mathsf{T}}$ in the left image, we can write

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{F} \boldsymbol{p}_l = 0$$

letting

$$\mathbf{F}\boldsymbol{p}_l = \begin{pmatrix} x_l f_{11} + y_l f_{12} + f_{13} \\ x_l f_{21} + y_l f_{22} + f_{23} \\ x_l f_{31} + y_l f_{32} + f_{33} \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \boldsymbol{l}$$



Epipolar geometry: Exploitation

Match constraint

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Use the recovered fundamental matrix **F** to constrain the search for matching points to a linear span..
- *Geometrically:* Given a point $\boldsymbol{p}_l = (x_l \quad y_l \quad 1)^{\mathsf{T}}$ in the left image, we can write

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{F} \boldsymbol{p}_l = 0$$

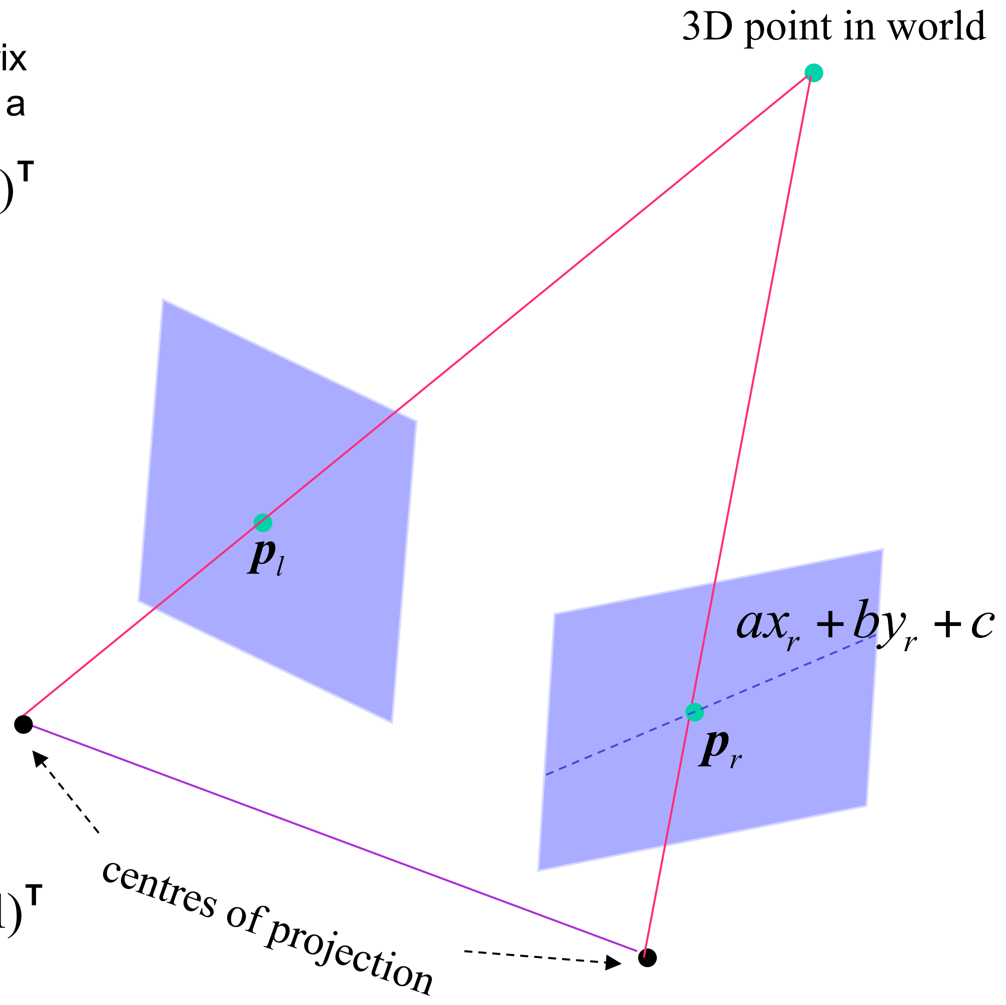
letting

$$\mathbf{F}\boldsymbol{p}_l = \begin{pmatrix} x_l f_{11} + y_l f_{12} + f_{13} \\ x_l f_{21} + y_l f_{22} + f_{23} \\ x_l f_{31} + y_l f_{32} + f_{33} \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \boldsymbol{l}$$

we write

$$\boldsymbol{p}_r^T \boldsymbol{l} = a x_r + b y_r + c = 0$$

which is the equation of a line in the right image on which the matching point $\boldsymbol{p}_r = (x_r \quad y_r \quad 1)^{\mathsf{T}}$ must lie.



Epipolar geometry: Exploitation

Match constraint

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Use the recovered fundamental matrix **F** to constrain the search for matching points to a linear span..
- *Geometrically:* Given a point $\boldsymbol{p}_l = (x_l \quad y_l \quad 1)^{\mathsf{T}}$ in the left image, we can write

$$\boldsymbol{p}_r^{\mathsf{T}} \mathbf{F} \boldsymbol{p}_l = 0$$

letting

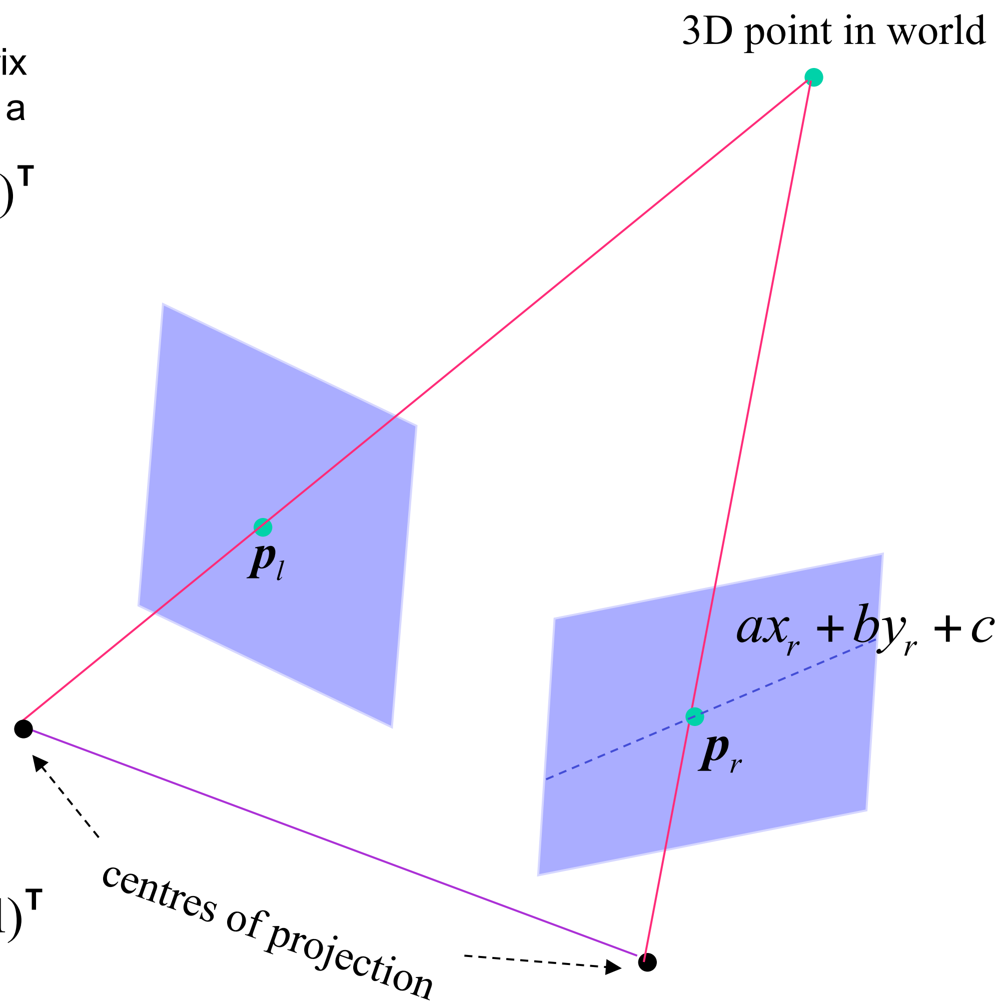
$$\mathbf{F}\boldsymbol{p}_l = \begin{pmatrix} x_l f_{11} + y_l f_{12} + f_{13} \\ x_l f_{21} + y_l f_{22} + f_{23} \\ x_l f_{31} + y_l f_{32} + f_{33} \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \boldsymbol{l}$$

we write

$$\boldsymbol{p}_r^T \boldsymbol{l} = a x_r + b y_r + c = 0$$

which is the equation of a line in the right image on which the matching point $\boldsymbol{p}_r = (x_r \quad y_r \quad 1)^{\mathsf{T}}$ must lie.

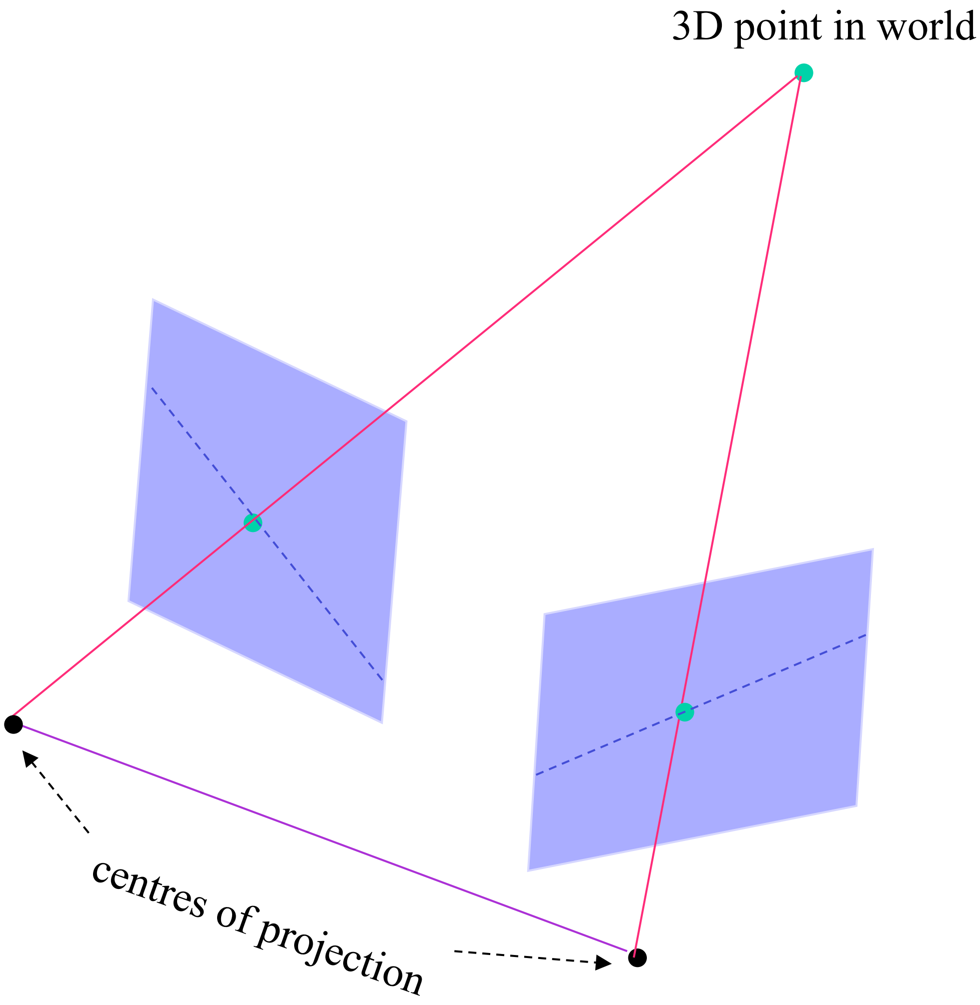
- *Benefit:* The having selected a feature in one image, correspondence need only search along a line in the other.



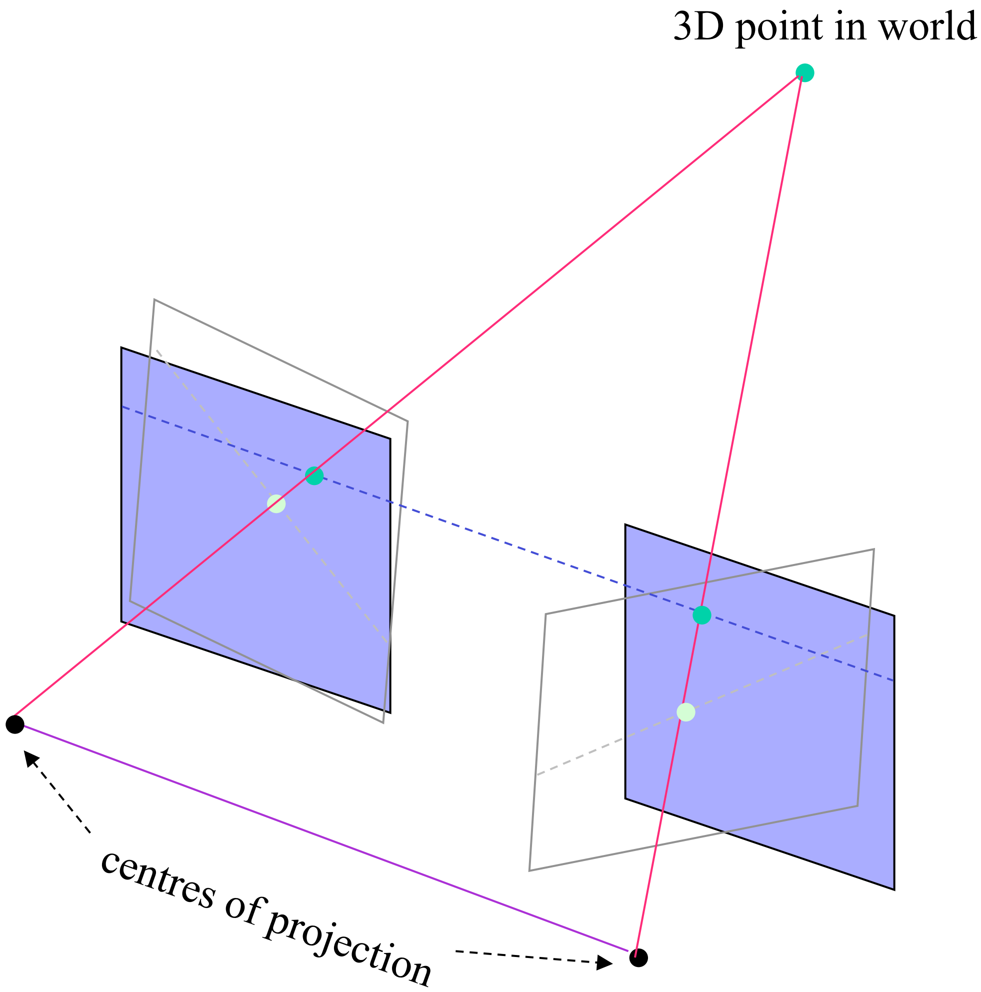
Epipolar geometry: Exploitation

Rectification

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Warp the images so that conjugate epipolar lines map to corresponding horizontal scan lines.
- *Geometrically:* The transformation amounts to a projective transformation of the images.
  - They are as if the original optical axes were parallel.
  - The simple stereo geometry that was introduced earlier.
- *Benefit:* The having selected a feature in one image, correspondence need only search along the corresponding horizontal raster in the other.



Epipolar geometry: Exploitation



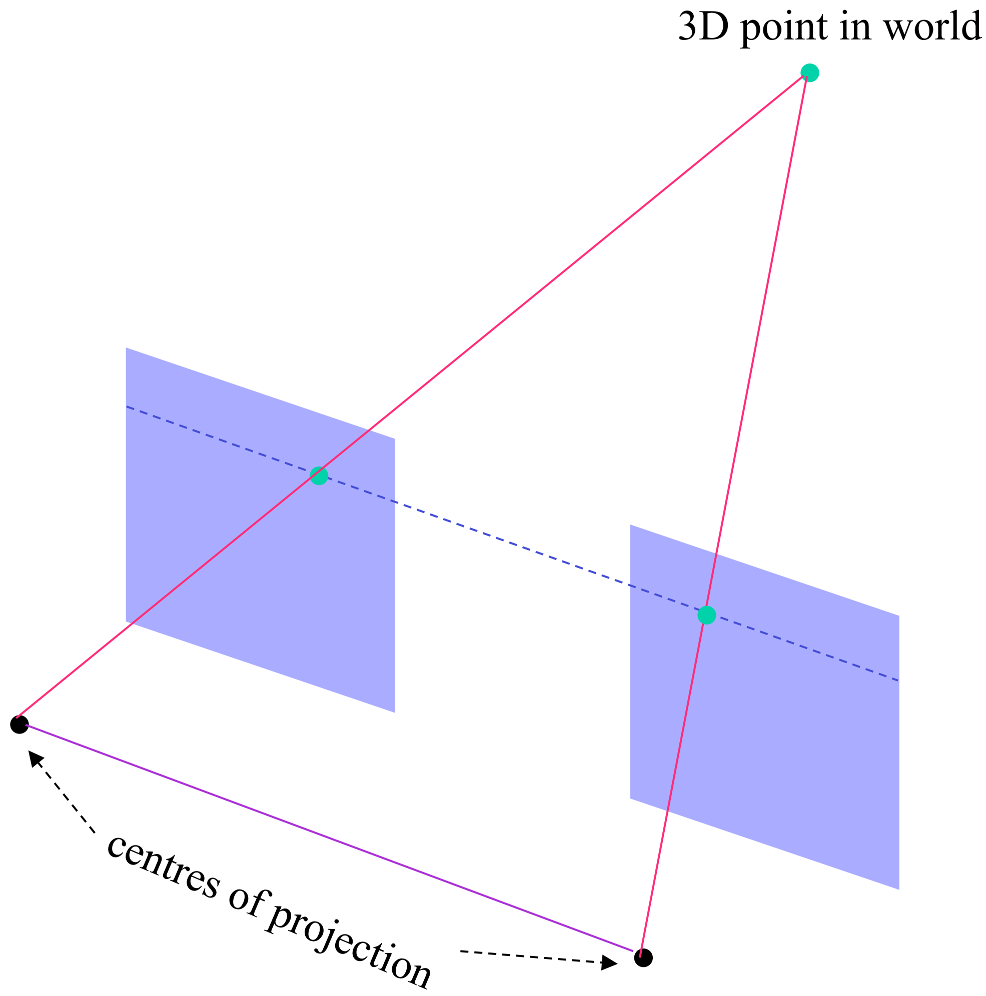
Rectification

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Warp the images so that conjugate epipolar lines map to corresponding horizontal scan lines.
- *Geometrically:* The transformation amounts to a projective transformation of the images.
  - They are as if the original optical axes were parallel.
  - The simple stereo geometry that was introduced earlier.
- *Benefit:* Having selected a feature in one image, correspondence need only search along the corresponding horizontal raster in the other.

Epipolar geometry: Exploitation

Rectification

- *Goal:* To allow the correspondence process best use of the recovered epipolar geometry.
- *Approach:* Warp the images so that conjugate epipolar lines map to corresponding horizontal scan lines.
- *Geometrically:* The transformation amounts to a projective transformation of the images.
  - They are as if the original optical axes were parallel.
  - The simple stereo geometry that was introduced earlier.
- *Benefit:* Having selected a feature in one image, correspondence need only search along the corresponding horizontal raster in the other.
  - Although, in practice we check a few nearby rasters as well, because life isn't perfect.



Epipolar geometry: Exploitation

Rectification analysis

- Assumptions (in both cameras)
  - The origin of the image reference frame is the principle point.
  - The focal length is f
  - The relative orientation between cameras is specified by rotation, **R**, and translation $\mathbf{T}$.
- What needs to be accomplished
  1. Rotate the left camera by **H** so that the epipole goes to infinity along the horizontal axis.
  2. Apply the same rotation to the right camera to recover the original geometry.
  3. Rotate the right camera by **R**
  4. Adjust the scale in both camera reference frames.

Epipolar geometry: Exploitation

Rectification analysis

- Assumptions (in both cameras)
  - The origin of the image reference frame is the principle point.
  - The focal length is f
  - The relative orientation between cameras is specified by rotation, **R**, and translation **T**.
- What needs to be accomplished
  1. Rotate the left camera by **H** so that the epipole goes to infinity along the horizontal axis.
  2. Apply the same rotation to the right camera to recover the original geometry.
  3. Rotate the right camera by **R**
  4. Adjust the scale in both camera reference frames.

Rectification formalization

- To specify **H** we need a triple of mutually orthogonal unit vectors, $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, so that

$$\mathbf{H} = \begin{pmatrix} \mathbf{e}_1^{\mathbf{T}} \\ \mathbf{e}_2^{\mathbf{T}} \\ \mathbf{e}_3^{\mathbf{T}} \end{pmatrix}$$

- Since the image centre is in the origin, the epipole is $T / \|T\|$, by definition and we want to map it to (1, 0, 0).
- Correspondingly, we take

$$\mathbf{e}_1 = T / \|T\|$$

$$\mathbf{e}_2 = (-t_y, t_x, 0)^{\mathbf{T}} / \sqrt{t_x^2 + t_y^2}$$

$$\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$$

Epipolar geometry: Exploitation

Rectification algorithm

1. Build **H**.
2. Let $\mathbf{R}_l = \mathbf{H}$ and $\mathbf{R}_r = \mathbf{RH}$
3. For each left camera point, $\boldsymbol{p}_l = (x, y, f)^{\mathsf{T}}$, compute

$$\mathbf{R}_l \boldsymbol{p}_l = (x', y', z')^{\mathsf{T}}$$

   and the coordinates of the corresponding rectified point as

$$\boldsymbol{p}_l' = \frac{f}{z'}(x', y', z')^{\mathsf{T}}$$

4. Repeat the previous step for all points in the right camera using $\mathbf{R}_r$

Epipolar geometry: Exploitation

Rectification algorithm

1. Build **H**.
2. Let $\mathbf{R}_l = \mathbf{H}$ and $\mathbf{R}_r = \mathbf{R}\mathbf{H}$
3. For each left camera point, $\boldsymbol{p}_l = (x, y, f)^{\mathsf{T}}$, compute

$$\mathbf{R}_l \boldsymbol{p}_l = (x', y', z')^{\mathsf{T}}$$

   and the coordinates of the corresponding rectified point as

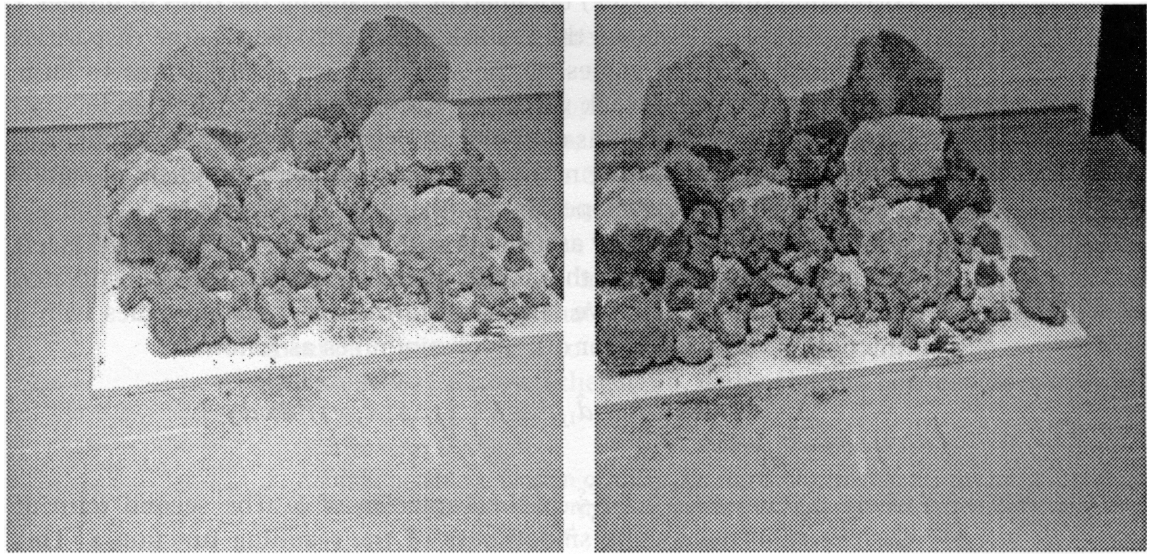
$$\boldsymbol{p}_l' = \frac{f}{z'}(x', y', z')^{\mathsf{T}}$$

4. Repeat the previous step for all points in the right camera using $\mathbf{R}_r$

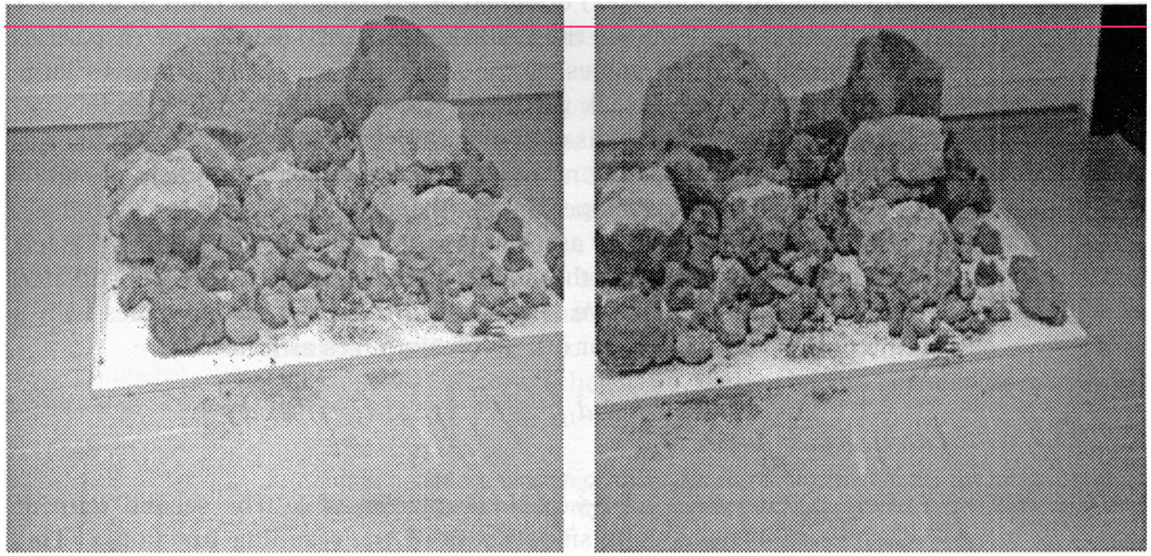
Remark

- A rectified image might not be in the same region of the image plane as the original image. To keep all points of the rectified images in regions of the same size as the originals the focal lengths can be adjusted.

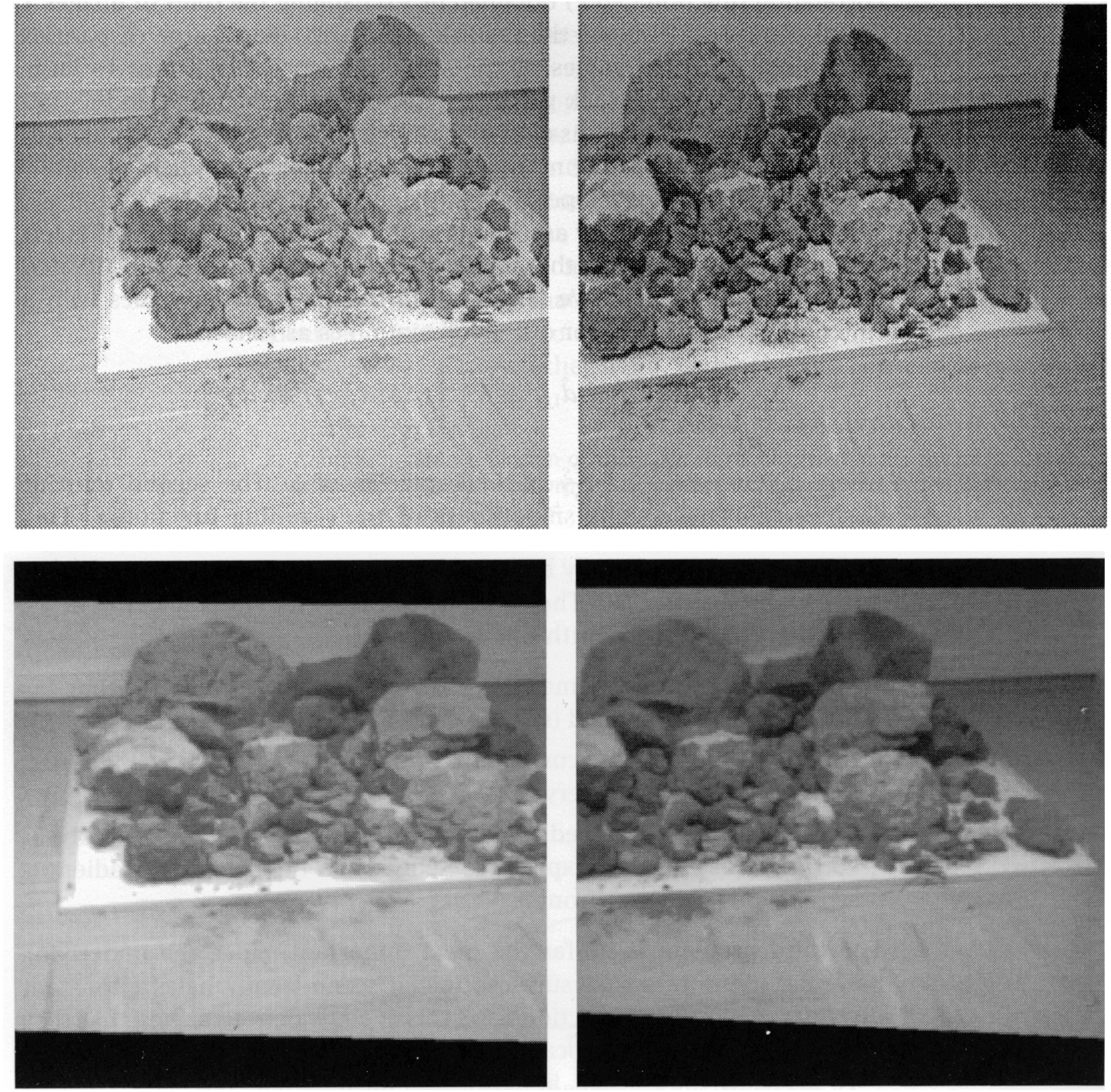
Epipolar geometry: Rectification example



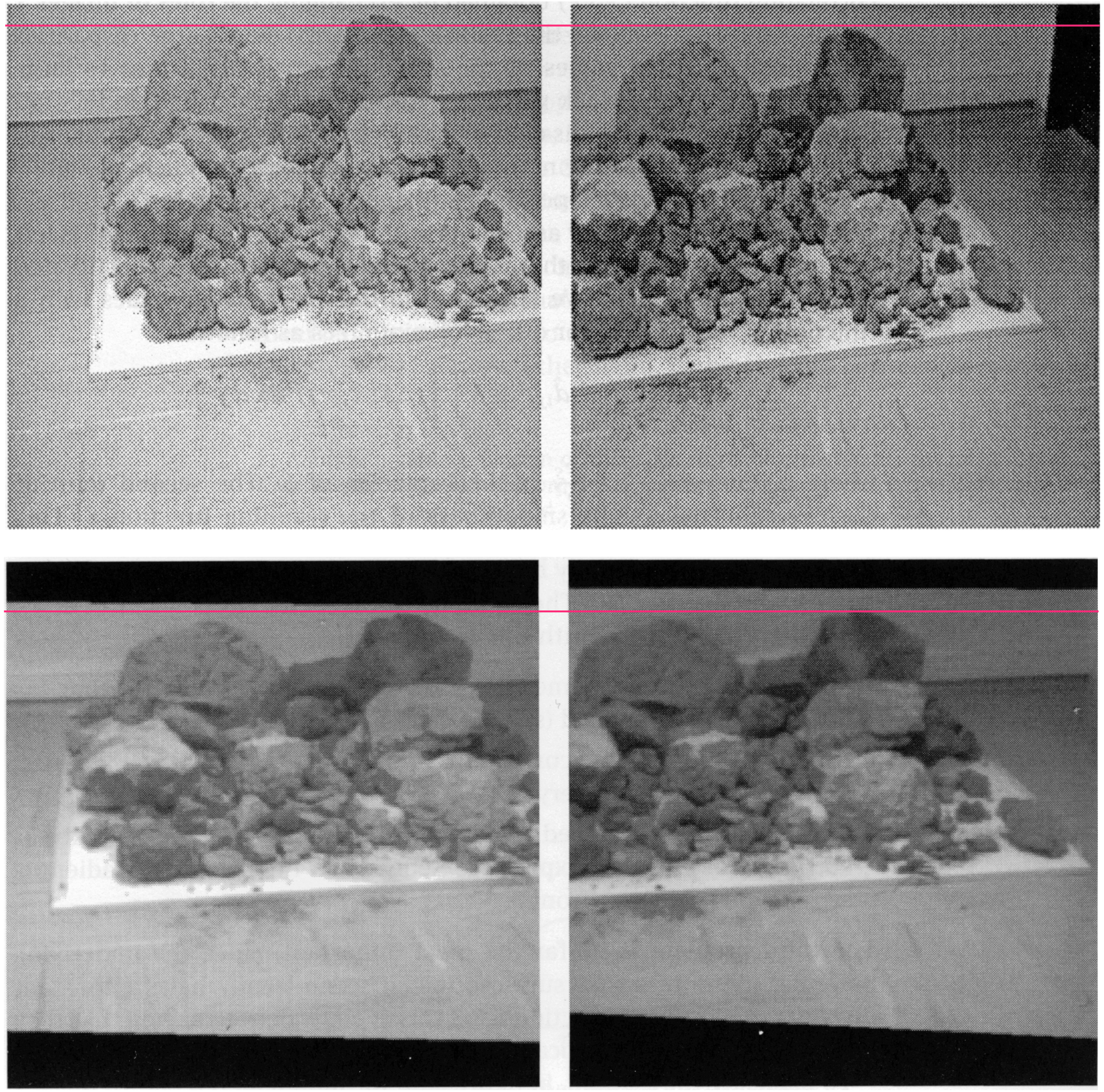
Epipolar geometry: Rectification example



Epipolar geometry: Rectification example



Epipolar geometry: Rectification example



Outline

- Introduction
- The correspondence problem
- Epipolar geometry
- **3-D reconstruction**
- Empirical examples
- Summary

3D reconstruction: Overview

Goal

- Given binocular stereo images and their correspondences.
- Recover the 3D geometry of the imaged scene.

What can be achieved

- Given intrinsic and extrinsic geometry of the cameras: Absolute Euclidean reconstruction.
- Given only intrinsic camera geometries: Reconstruction up to a scale factor.
- Given no information on camera geometries: Reconstruction up to a projective transformation.

3D reconstruction: Absolute Euclidean

Input

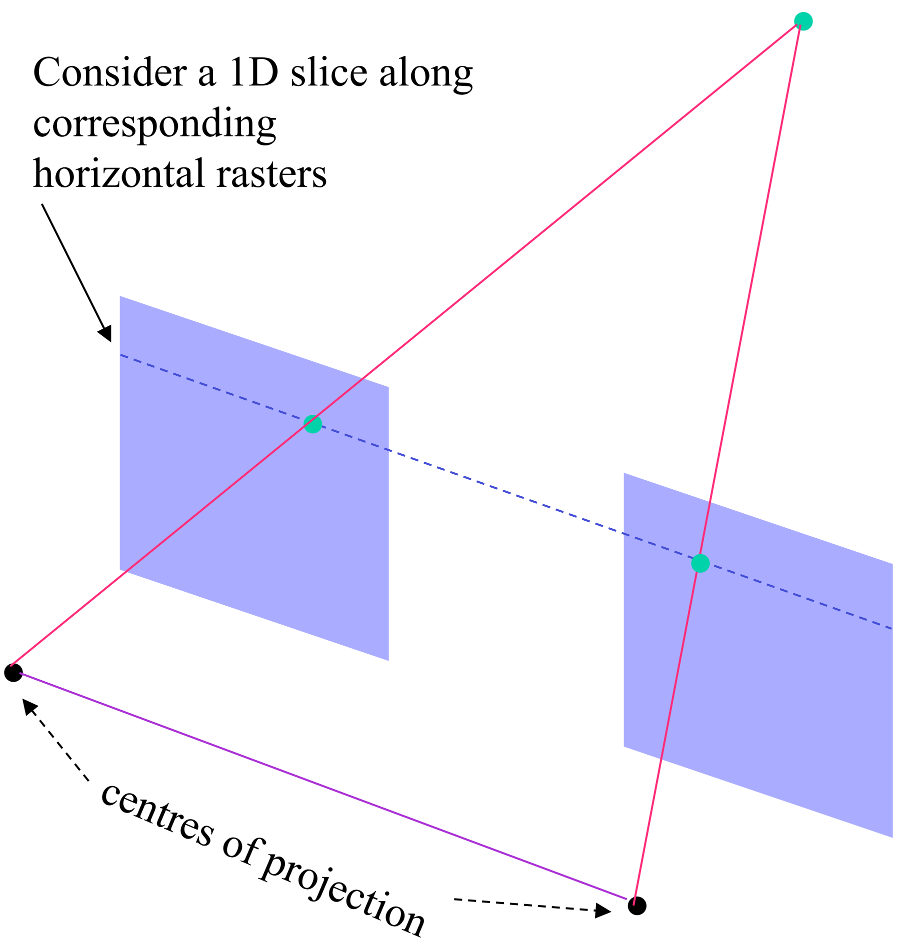
- A set of correspondences between binocular stereo images.
- Intrinsic and extrinsic camera geometry.

Output

- Euclidean distance to each matched point in the scene.
- Assuming rectified stereo pair we have simply

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

with disparity $d = xr\text{-}xl$, as calculated earlier in this unit



3D reconstruction: Absolute Euclidean

Input

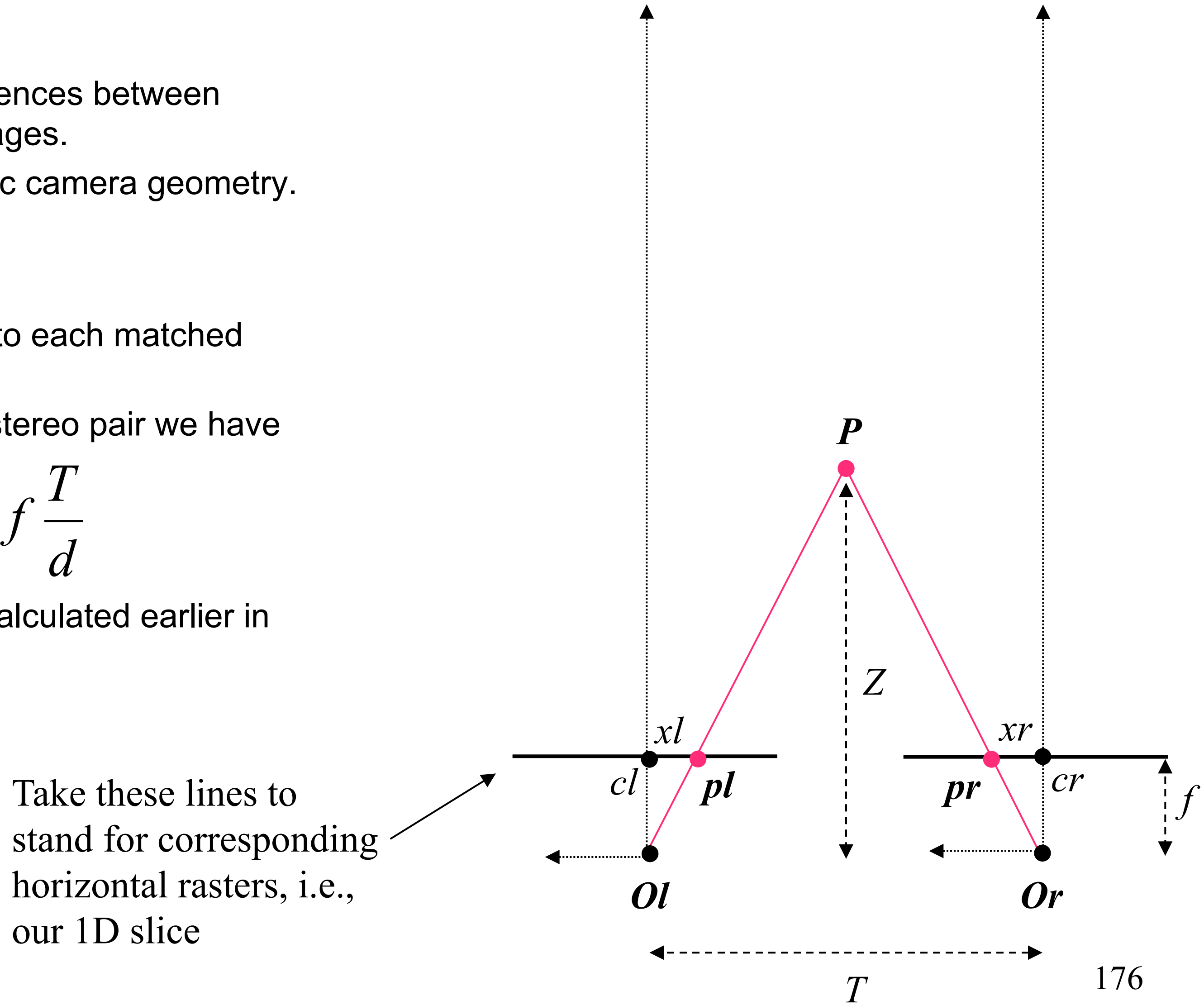
- A set of correspondences between binocular stereo images.
- Intrinsic and extrinsic camera geometry.

Output

- Euclidean distance to each matched point in the scene.
- Assuming rectified stereo pair we have simply

$$Z(\boldsymbol{P}) = f\,\frac{T}{d}$$

with disparity d as calculated earlier in this unit



3D reconstruction: Absolute Euclidean

Input

- A set of correspondences between binocular stereo images.
- Intrinsic and extrinsic camera geometry.

Output

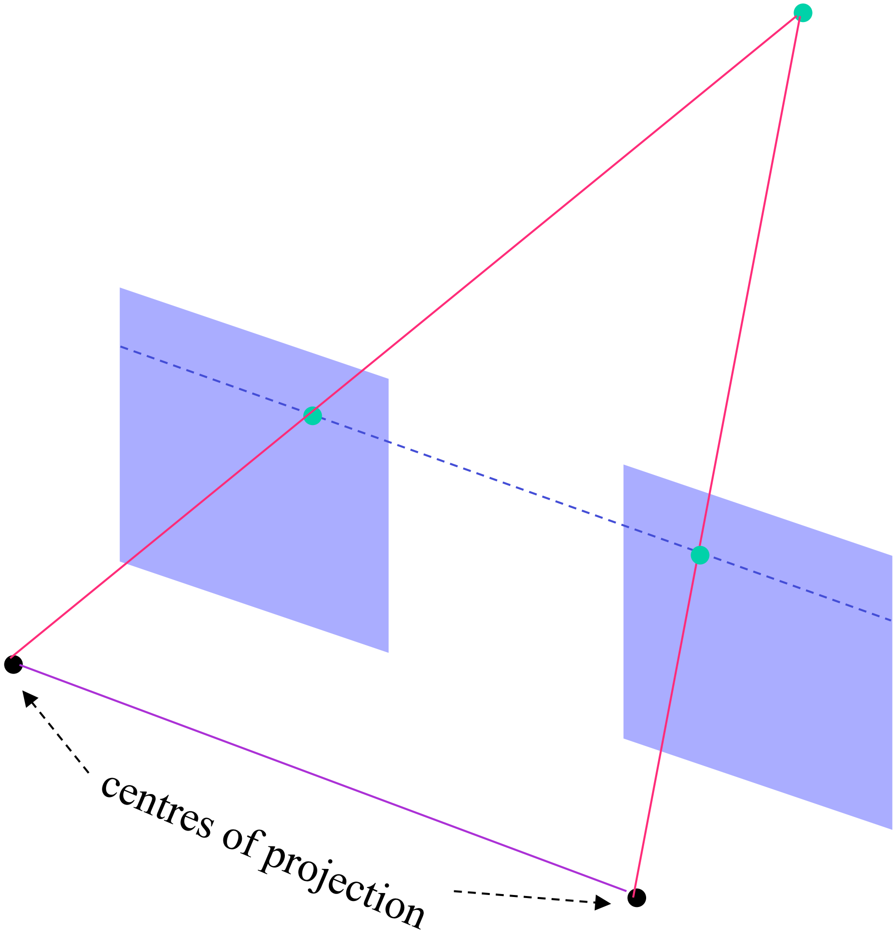
- Euclidean distance to each matched point in the scene.
- Assuming rectified stereo pair we have simply

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

with disparity d, as calculated earlier in this unit

Remark

- In practical situations we may need to do additional work.
- Matched points may fail to lie on exactly the same horizontal raster.
- Choose the desired 3D point estimate as the that of minimum distance between the rays from the centres of projection and the matched points.



3D reconstruction: Absolute Euclidean

Input

- A set of correspondences between binocular stereo images.
- Intrinsic and extrinsic camera geometry.

Output

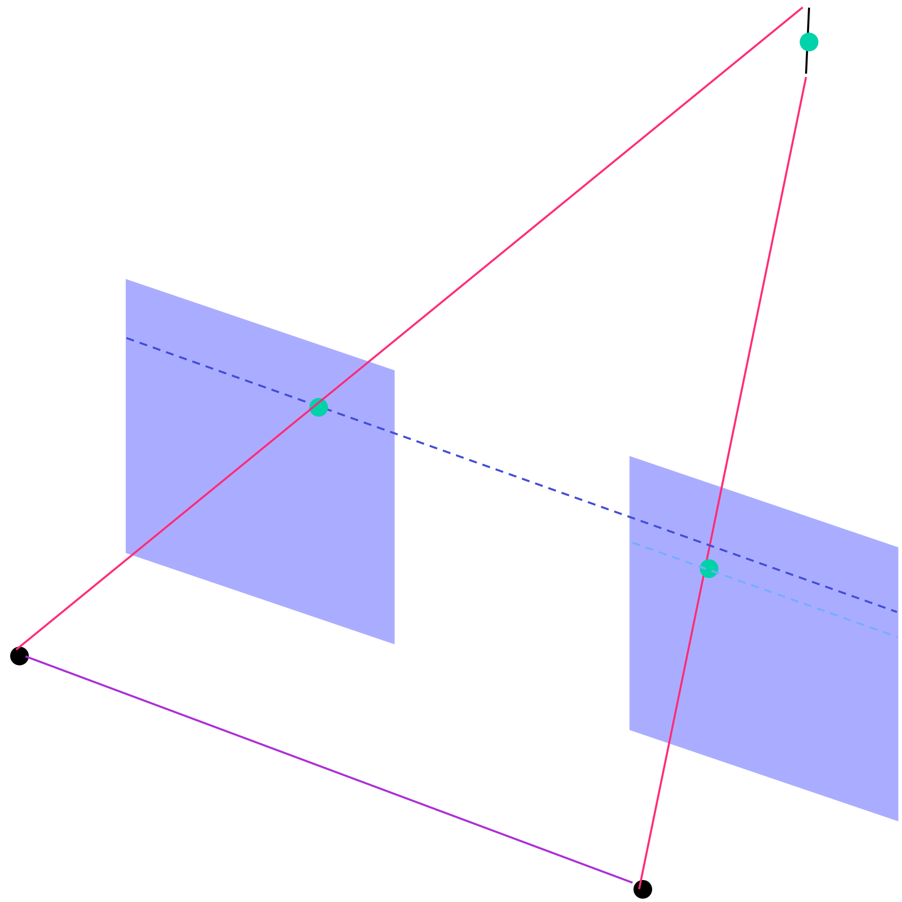
- Euclidean distance to each matched point in the scene.
- Assuming rectified stereo pair we have simply

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

with disparity d, as calculated earlier in this unit

Remark

- In practical situations we may need to do additional work.
- Matched points may fail to lie on exactly the same horizontal raster.
- Choose the desired 3D point estimate as the that of minimum distance between the rays from the centres of projection and the matched points.



3D reconstruction: Recap

Input

- Given binocular stereo images and their correspondences.
- Recover the 3D geometry of the imaged scene.

Output

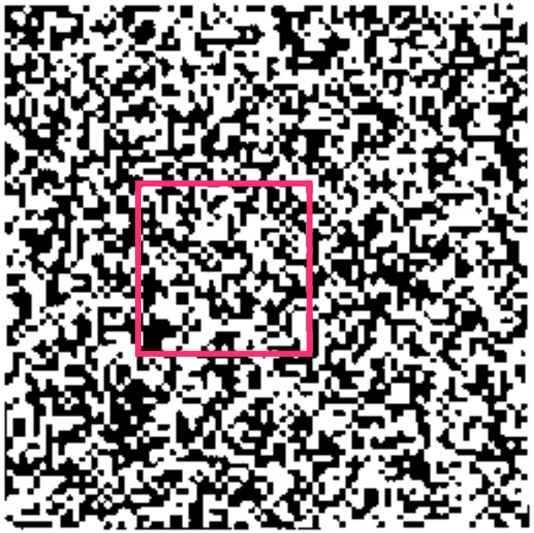
- 3D geometry of imaged scene
- Exactly what form this takes depends on how much is known about camera geometries
  - Given intrinsic and extrinsic geometry of the cameras: Absolute Euclidean reconstruction.
  - Given only intrinsic camera geometries: Reconstruction up to a scale factor.
  - Given no information on camera geometries: Reconstruction up to a projective transformation.
- While we have focused on recovery of range estimates (3D distance), other information is possible
  - 3D surface orientation
  - 3D surface curvature
  - 3D surface discontinuities

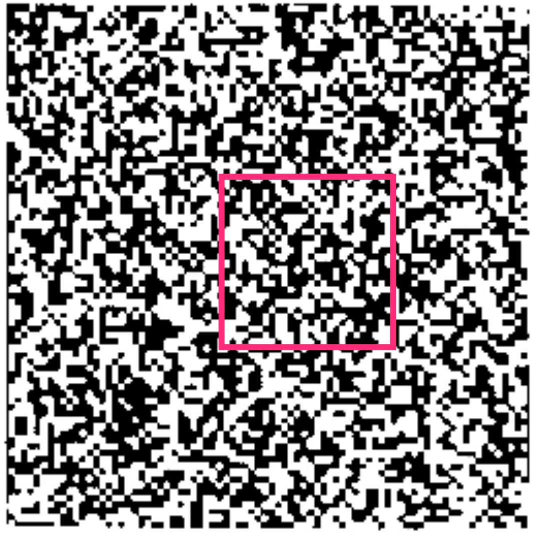
Outline

- Introduction
- The correspondence problem
- Epipolar geometry
- 3-D reconstruction
- **Empirical examples**
- Summary

Empirical examples: Random-dot stereogram

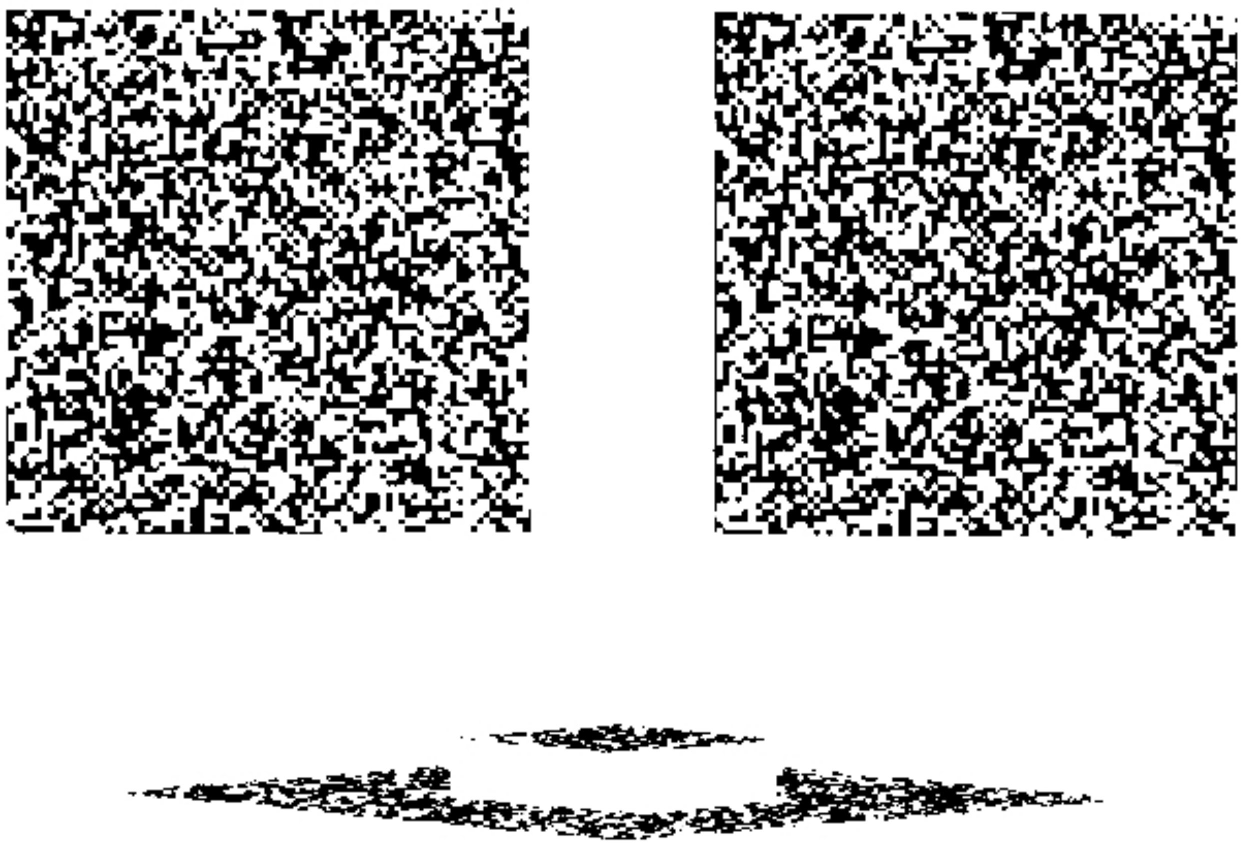
Empirical examples: Random-dot stereogram



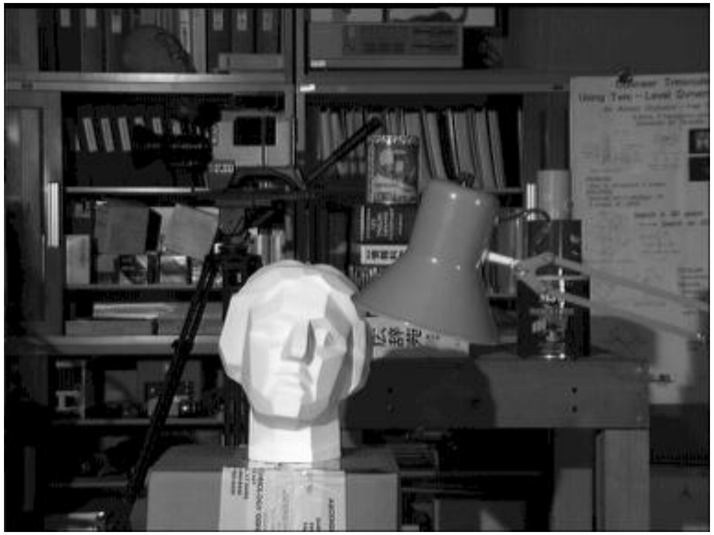


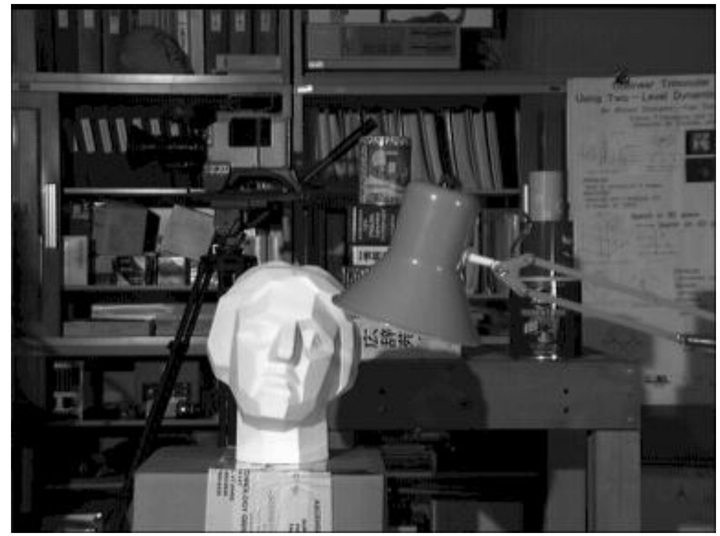
Empirical examples: Random-dot stereogram

Empirical examples: Random-dot stereogram

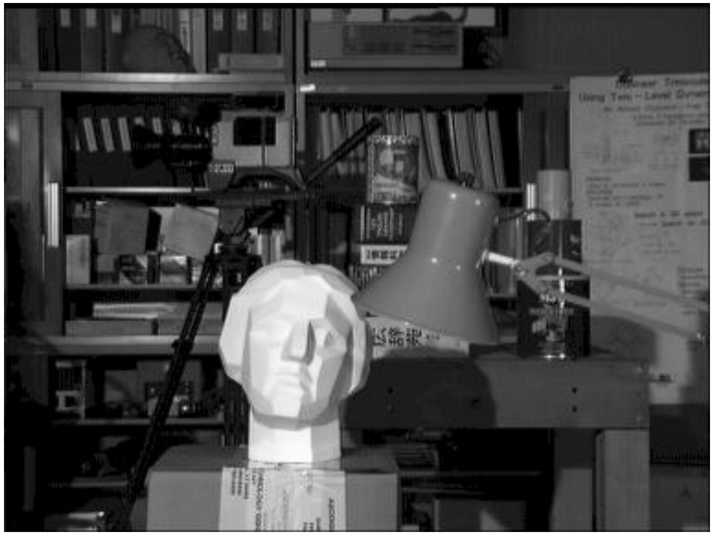


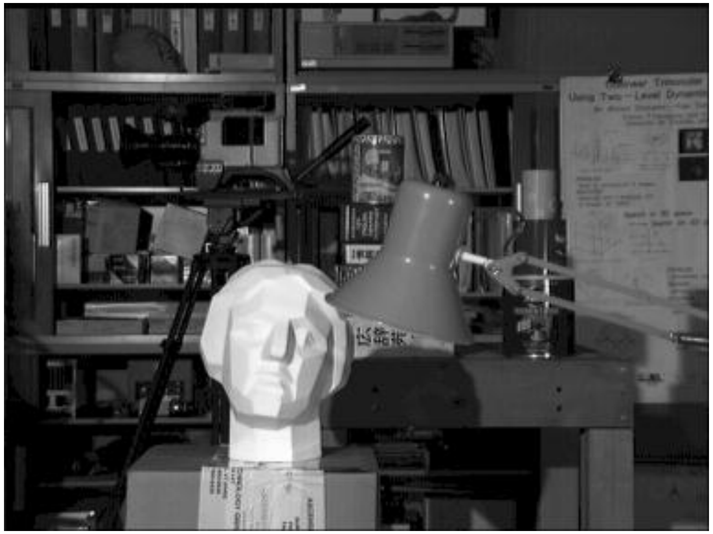
Empirical examples: Calibrated laboratory scene



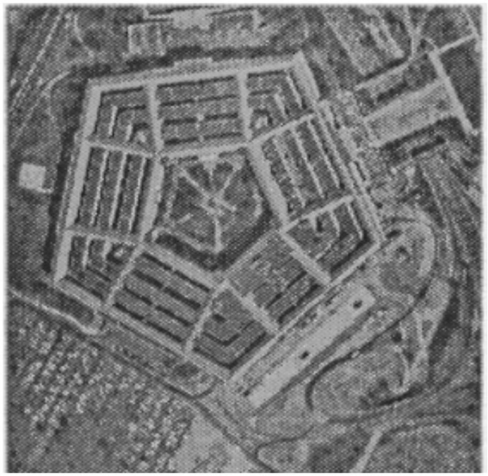


Empirical examples: Laboratory scene

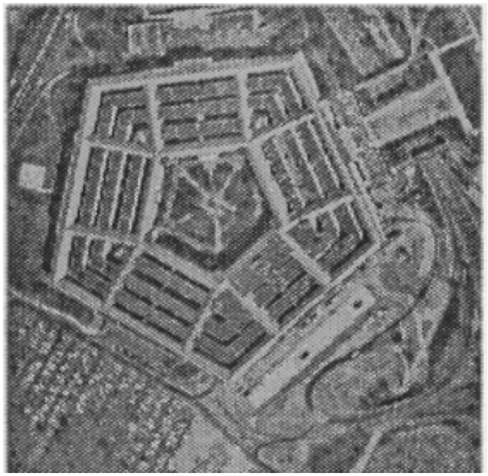
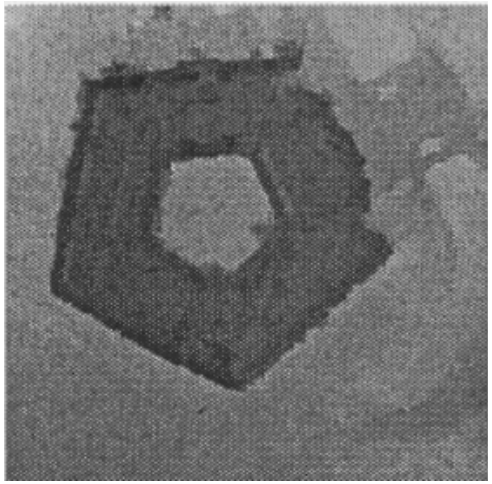




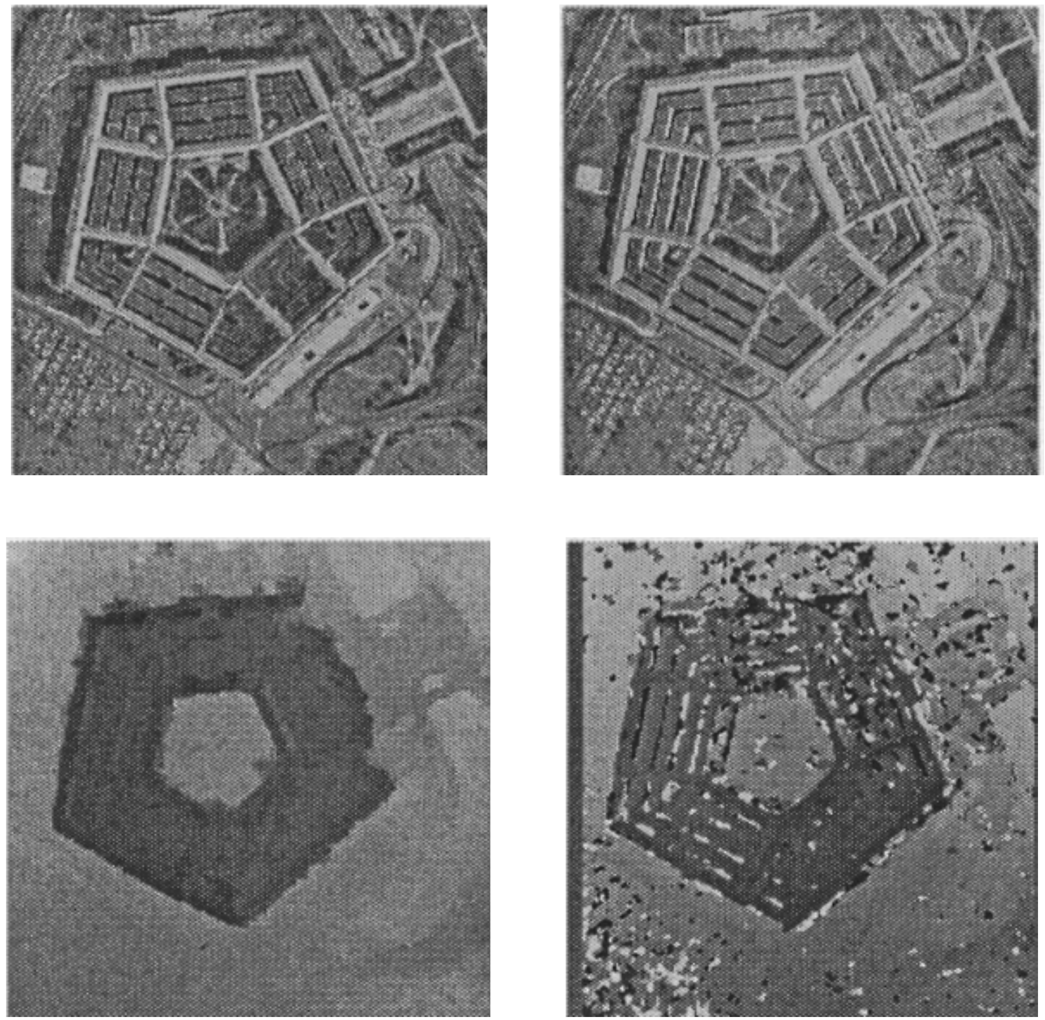
Empirical examples: Real-world scene



Empirical examples: Real-world scene

Empirical examples: Real-world scene

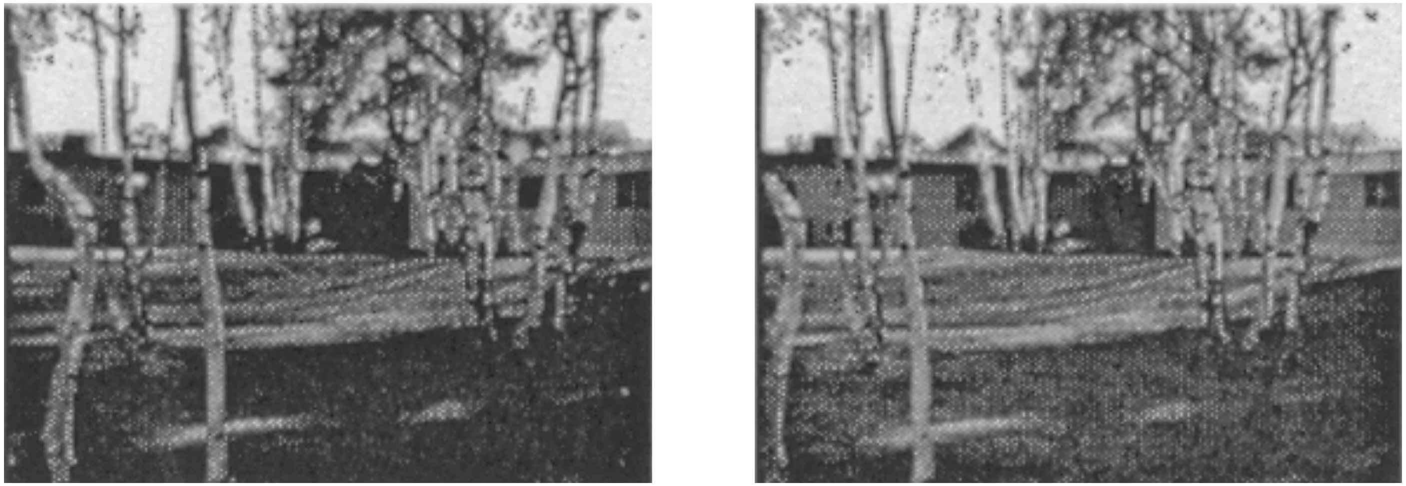


Empirical examples: Real-world scene

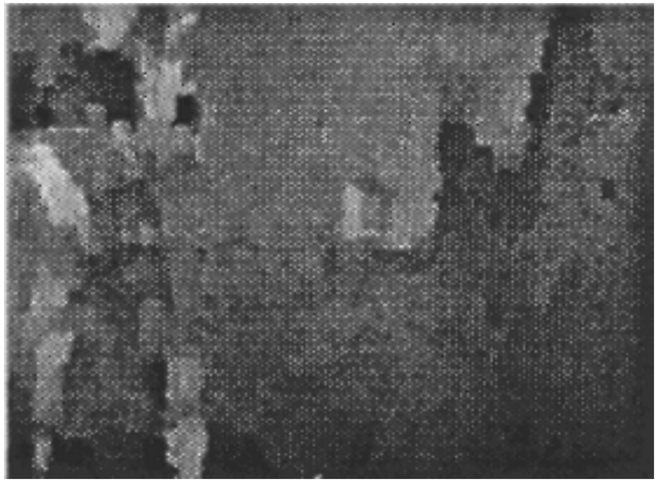
Empirical examples: Real-world scene



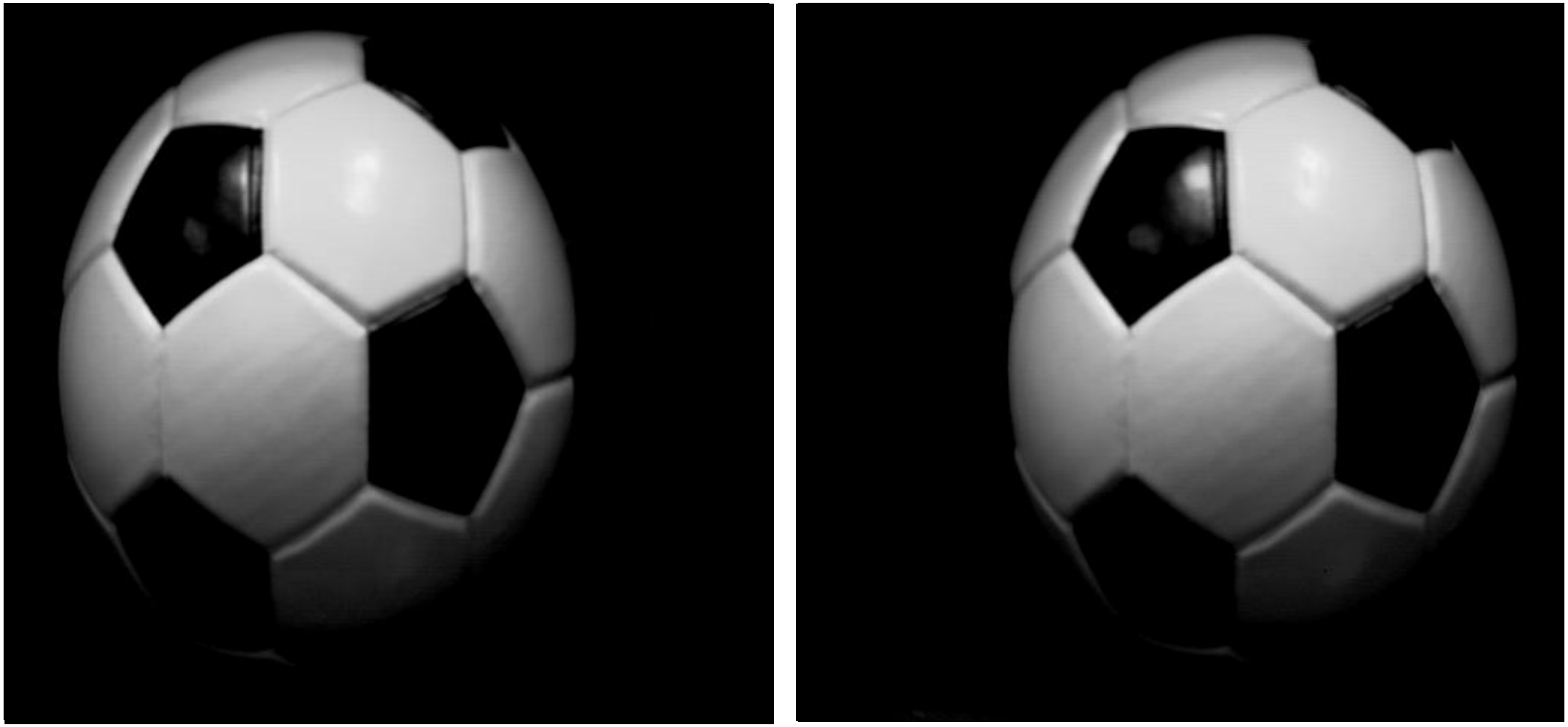
Empirical examples: Real-world scene



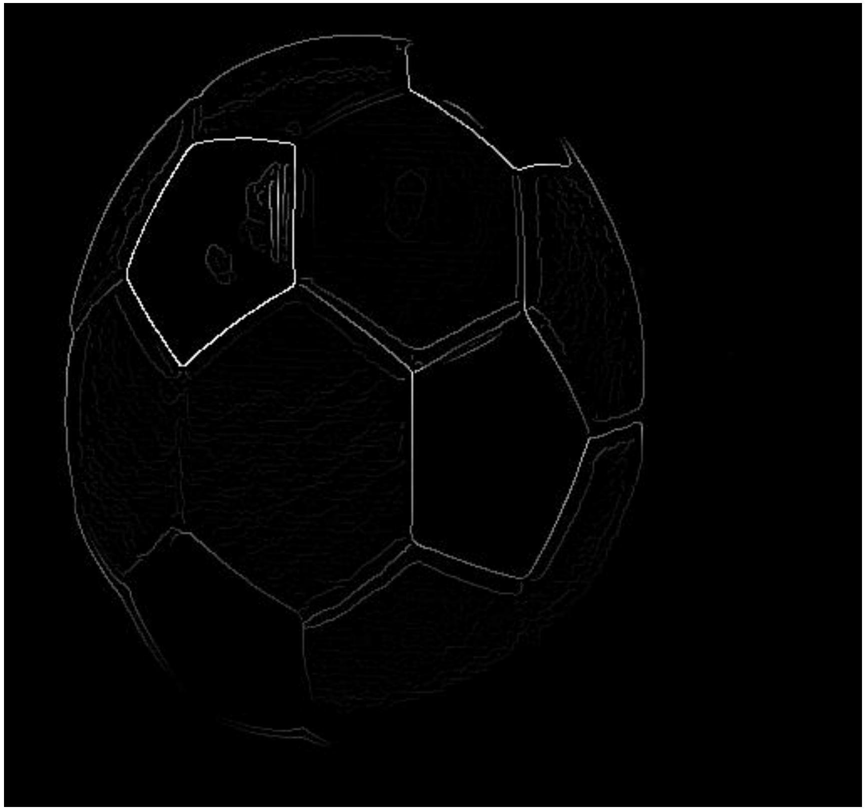
Empirical examples: Real-world scene

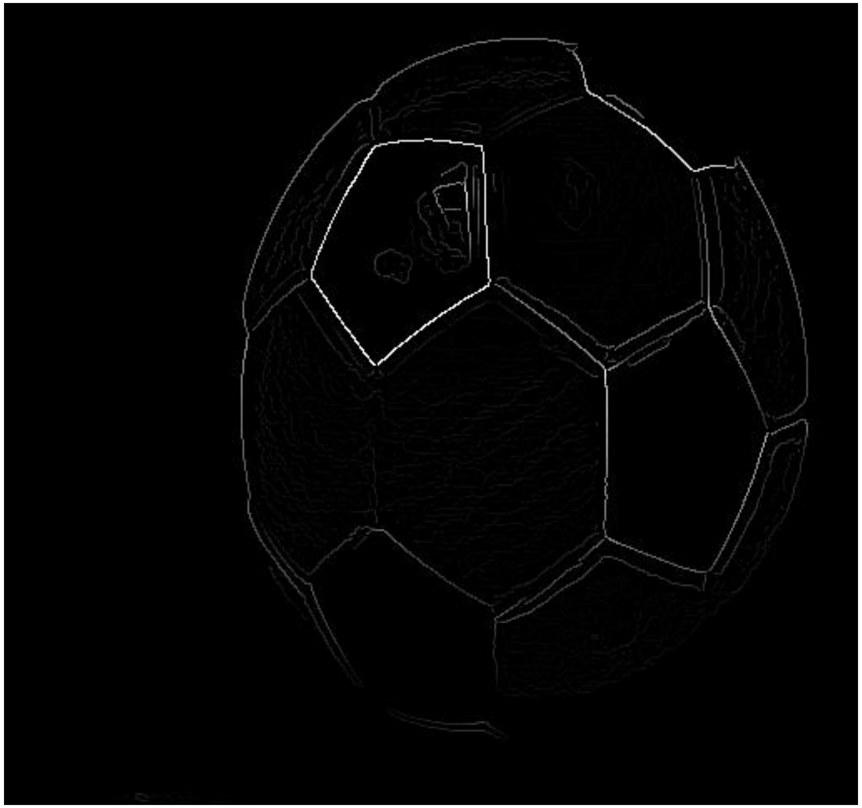


Empirical example: Feature-based

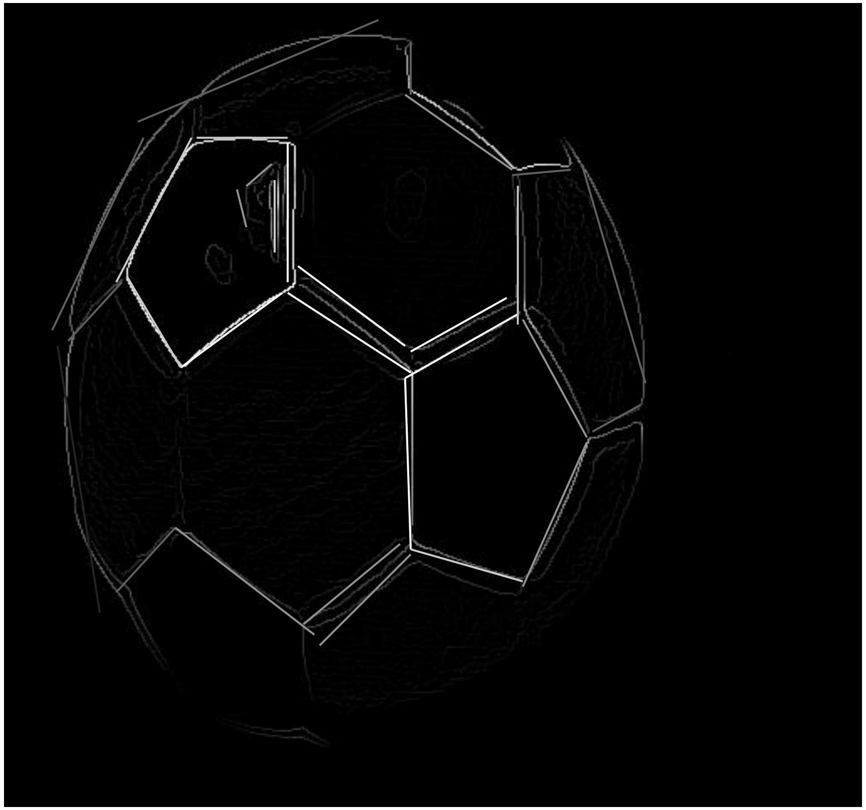


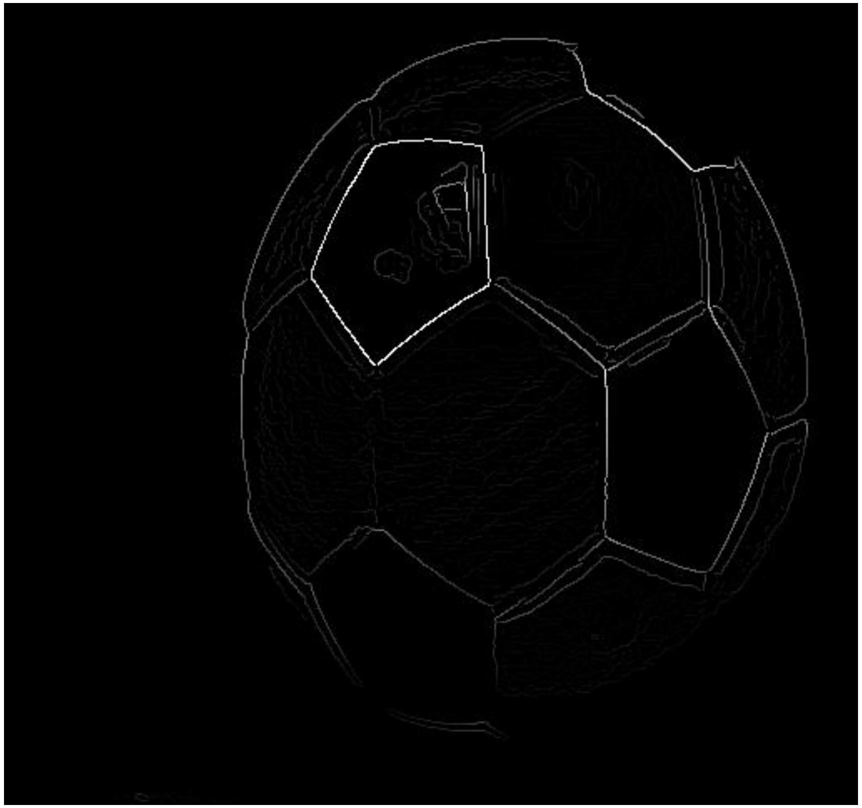
Empirical example: Feature-based



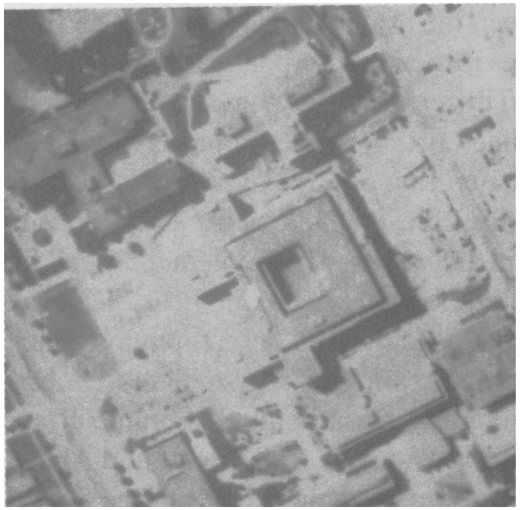


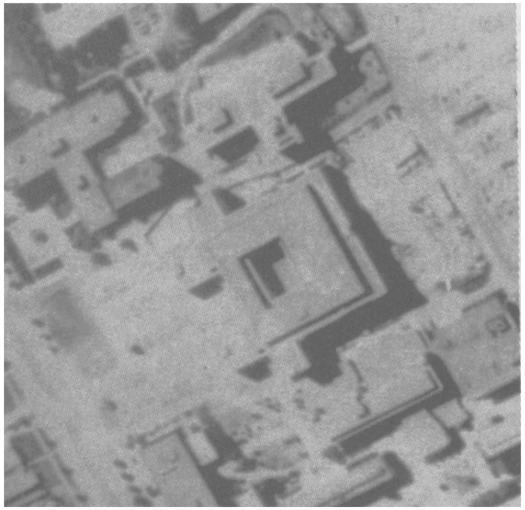
Empirical example: Feature-based



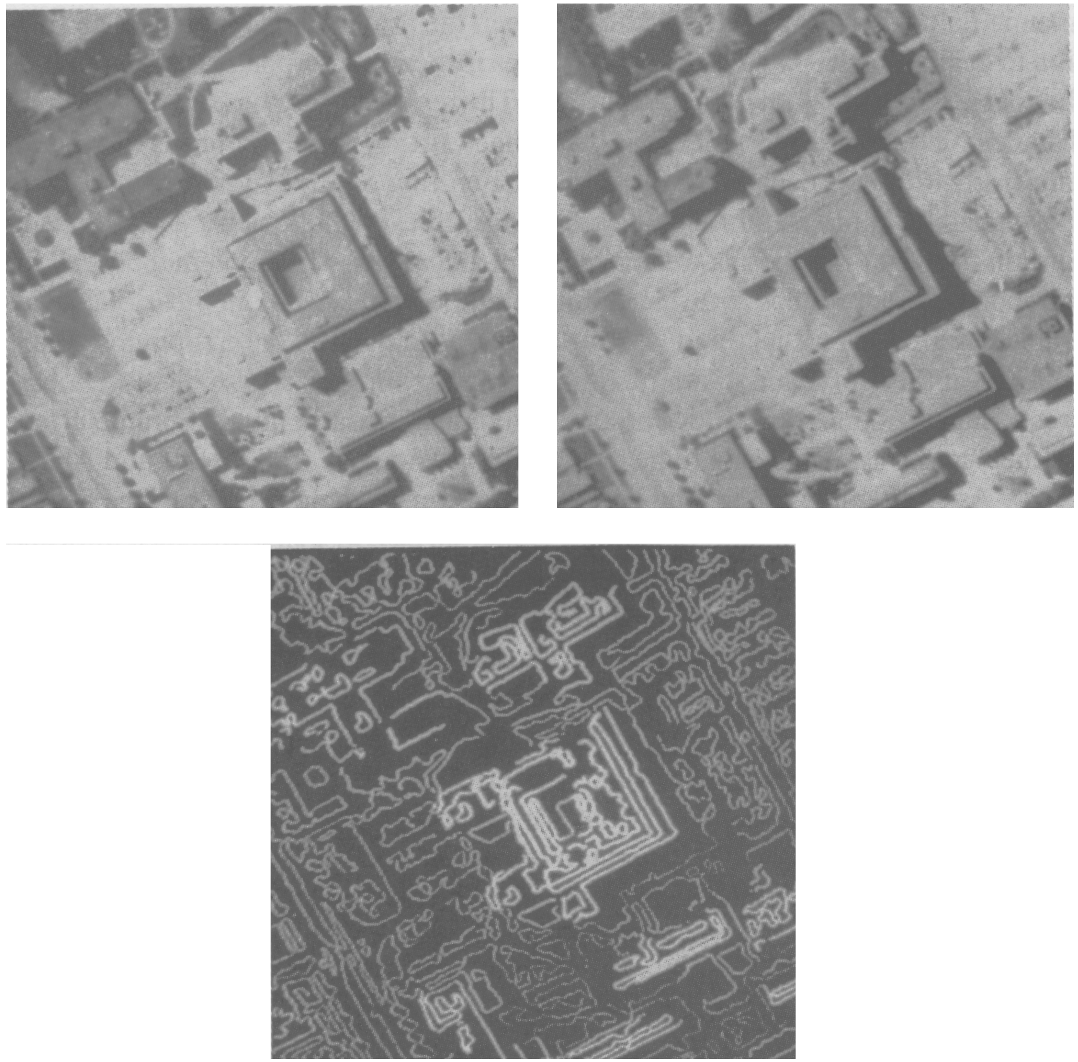


Empirical examples: Real-world scene

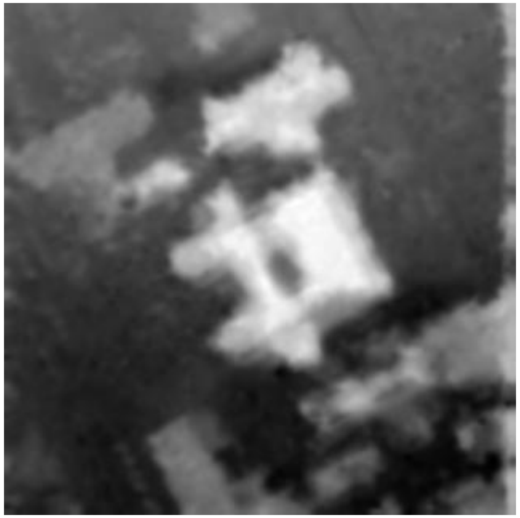




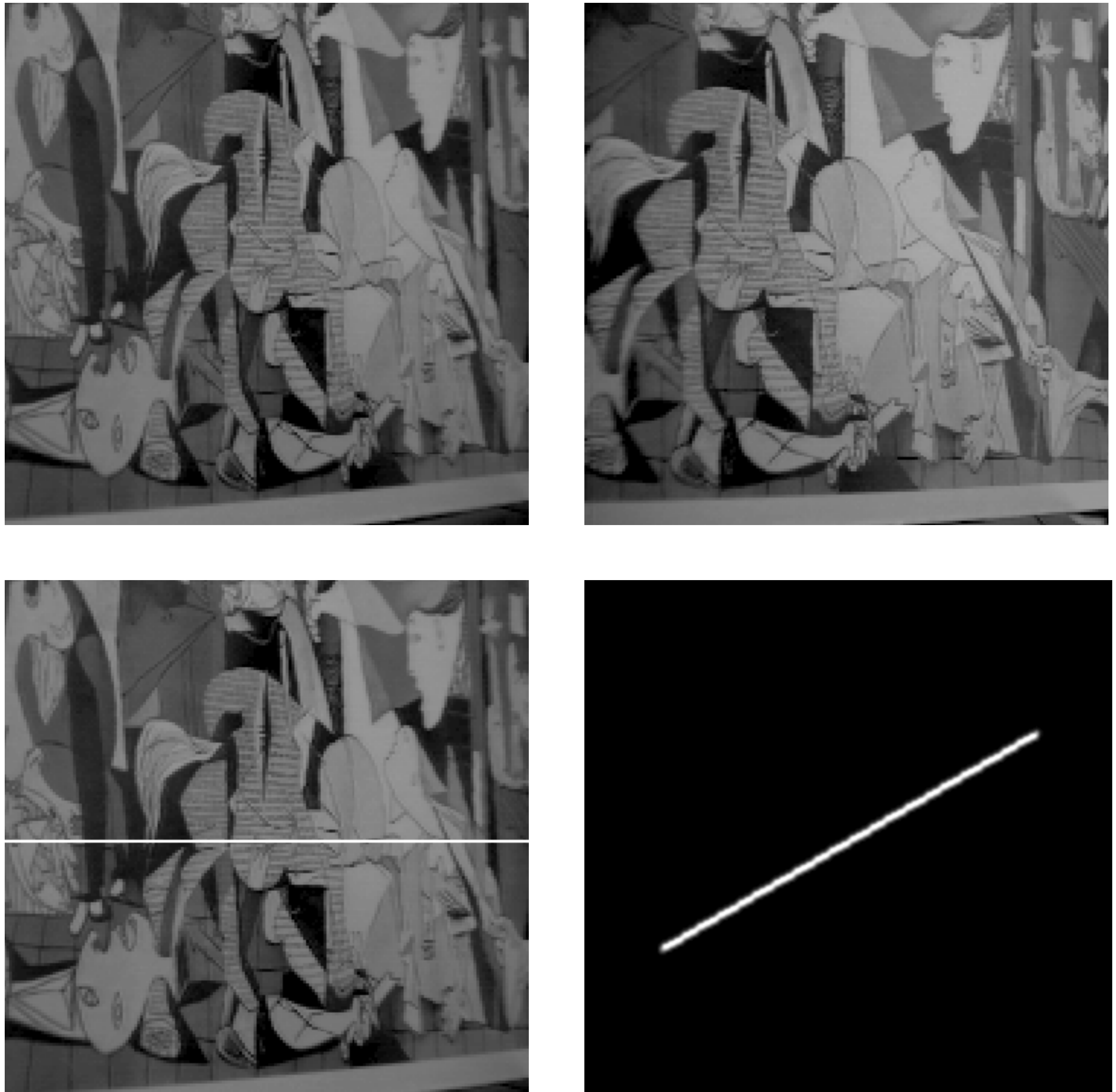
Empirical examples: Real-world scene



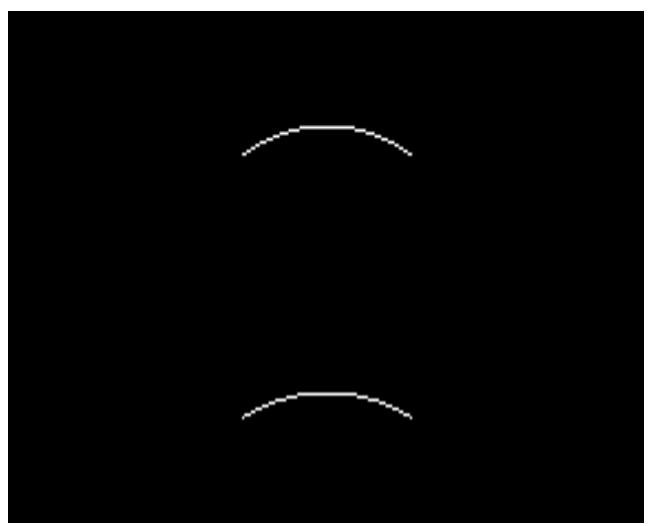
Empirical examples: Real-world scene



Empirical examples: Recovery of 3D orientation

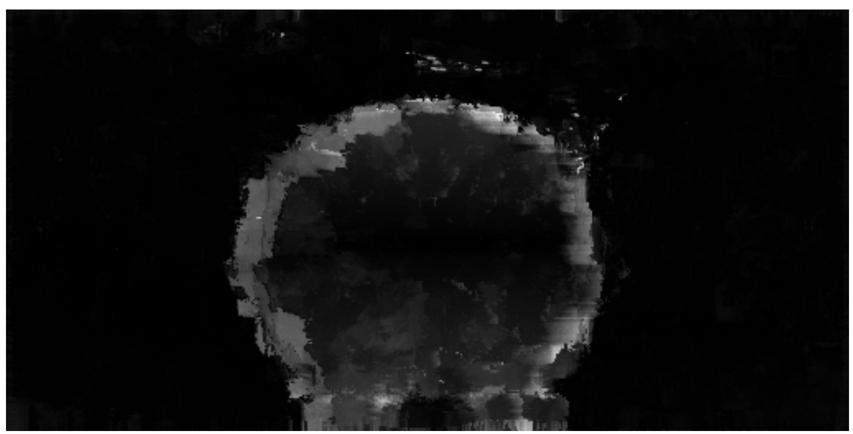


Empirical examples: Recovery of 3D curvature



Empirical examples: Recovery of 3D discontinuities

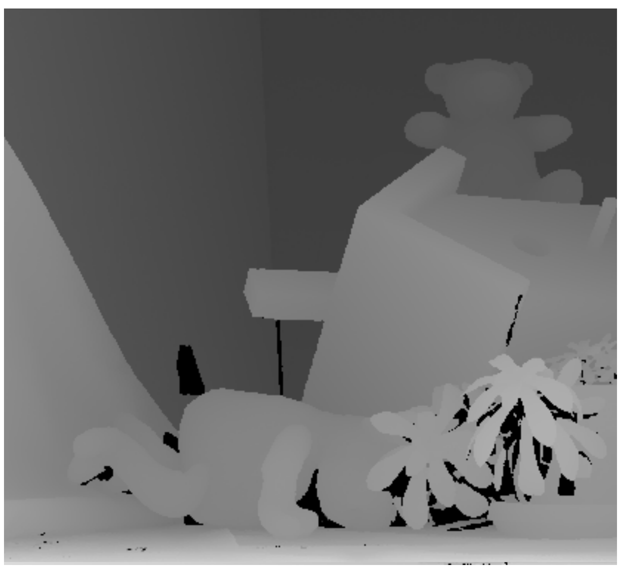
 



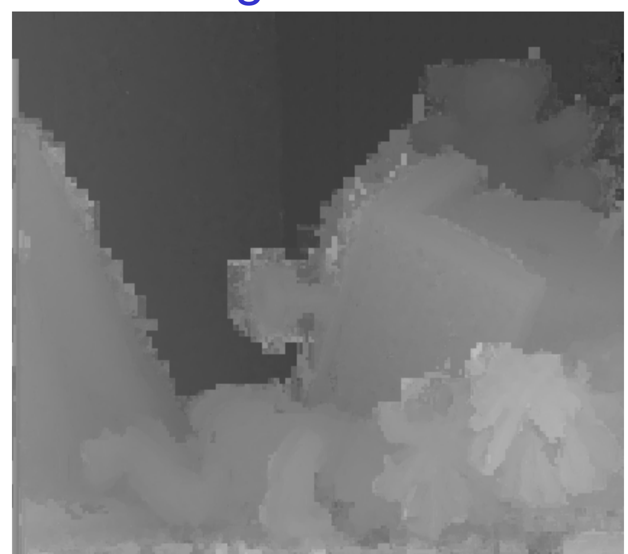
Empirical examples: Comparison to ground truth

left view

right view



ground truth disparity

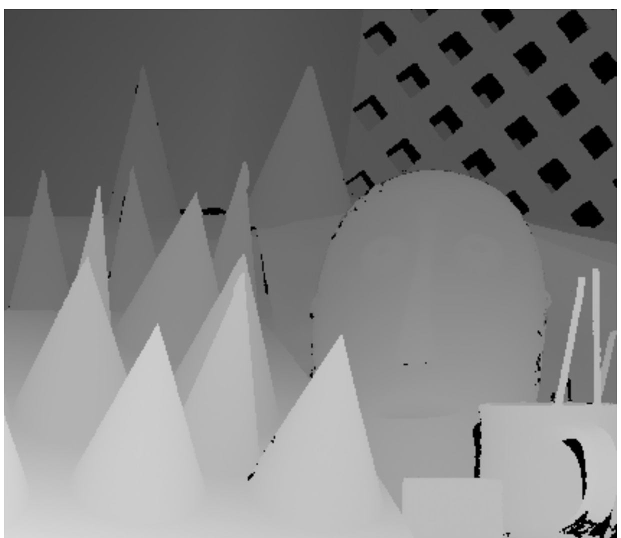


recovered disparity

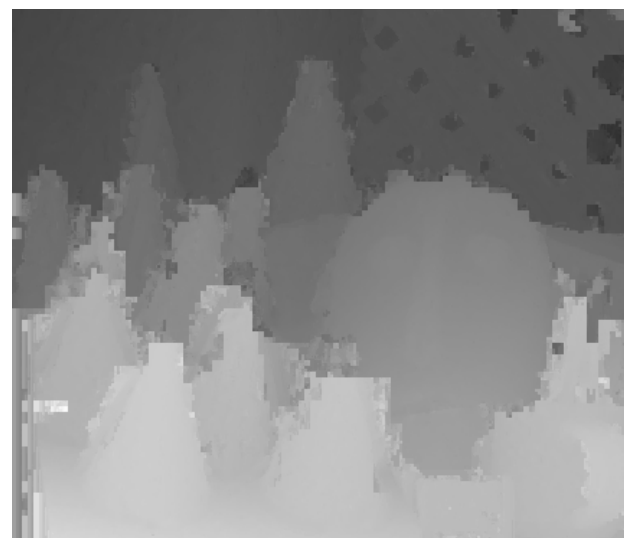
Empirical examples: Comparison to ground truth

left view

right view



ground truth disparity

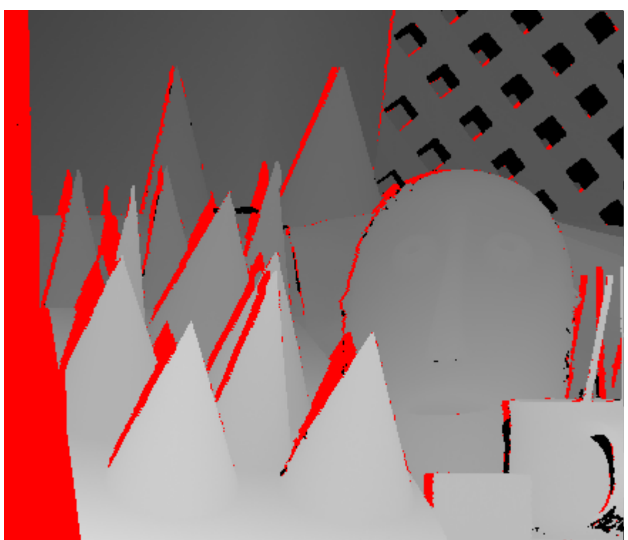


recovered disparity

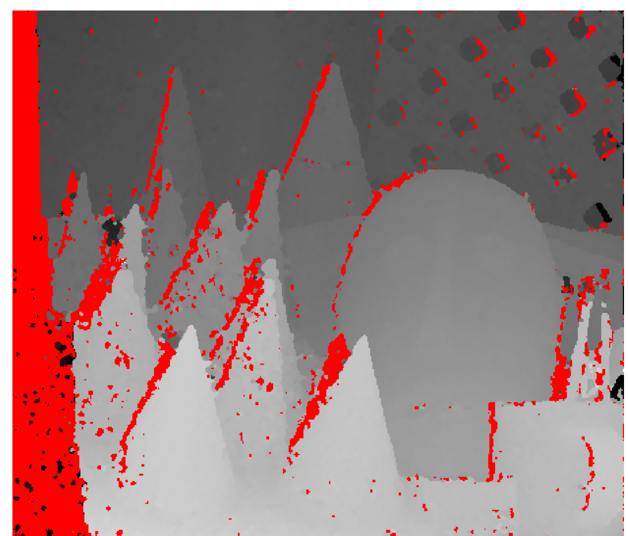
Empirical examples: Comparison to ground truth

left view

right view

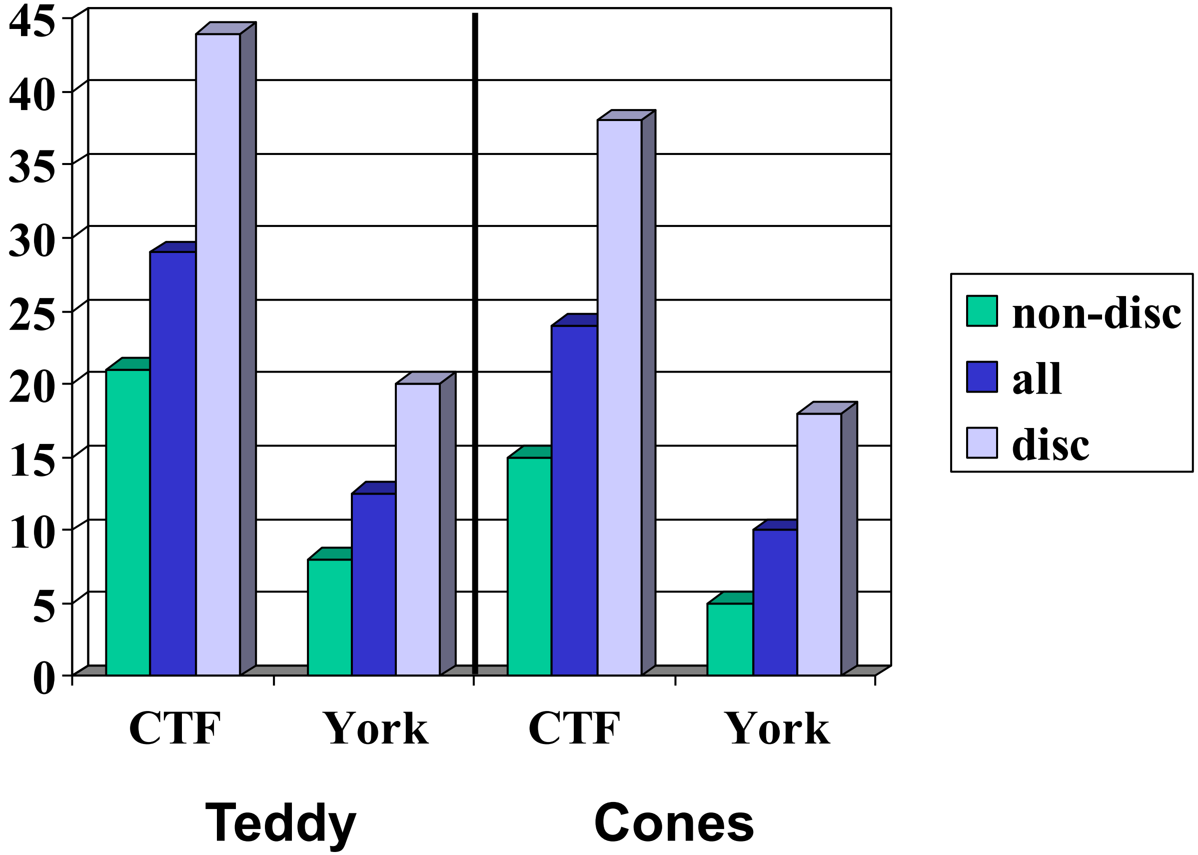


ground truth disparity



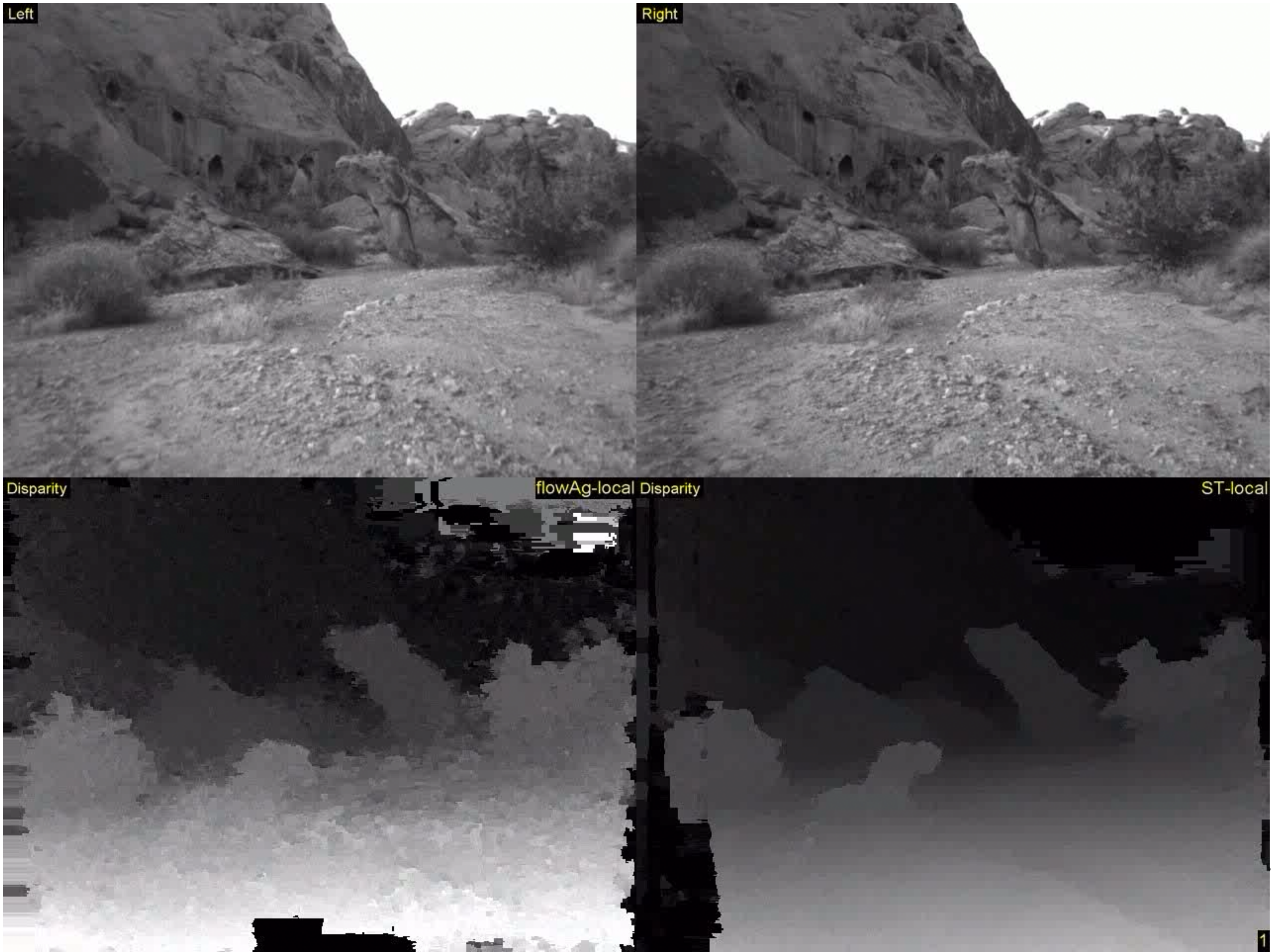
recovered disparity

Empirical examples: Comparison to ground truth



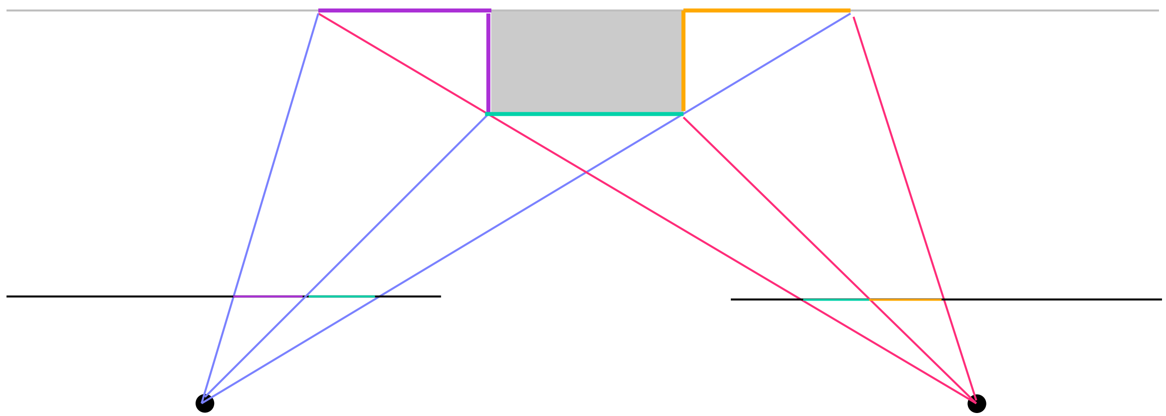
Error shown as percentage of points with greater than 1 pixel error.

Empirical examples: Stereo video



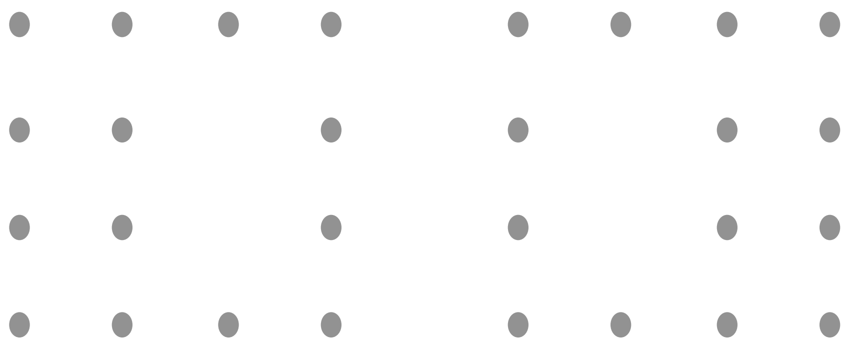
Empirical examples: A mystery

Recall the case of half-occlusion



Empirical examples: A mystery

A half-occlusion dot pattern stereogram: The only 3D cue is lack of correspondence

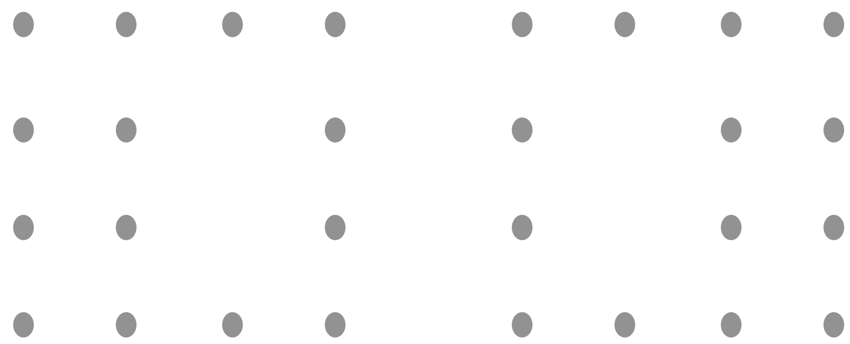


Status

- No extant computer vision algorithm can correctly infer 3D from such an impoverished input.

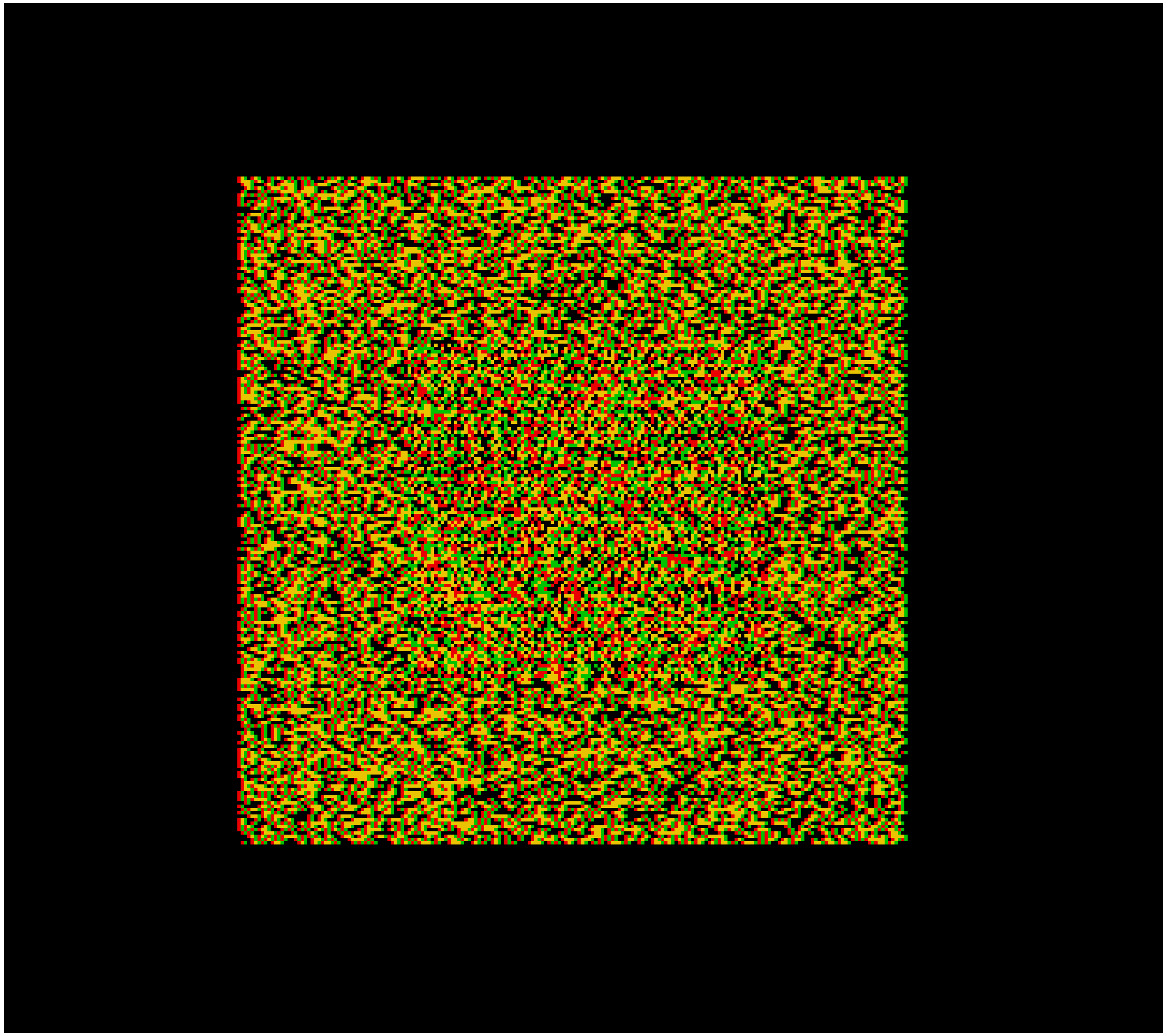
Empirical examples: A mystery

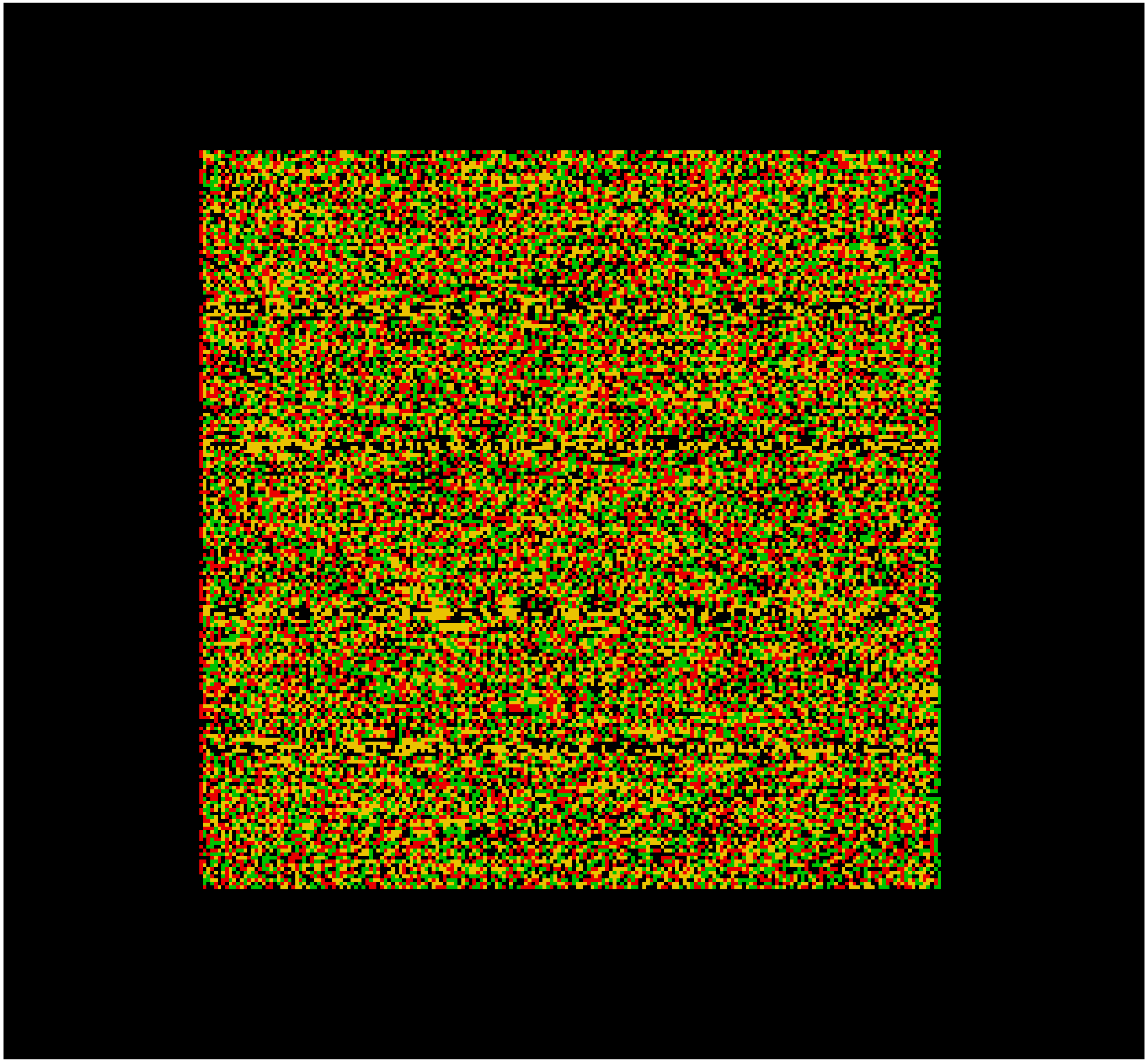
A half-occlusion dot pattern stereogram: The only 3D cue is lack of correspondence

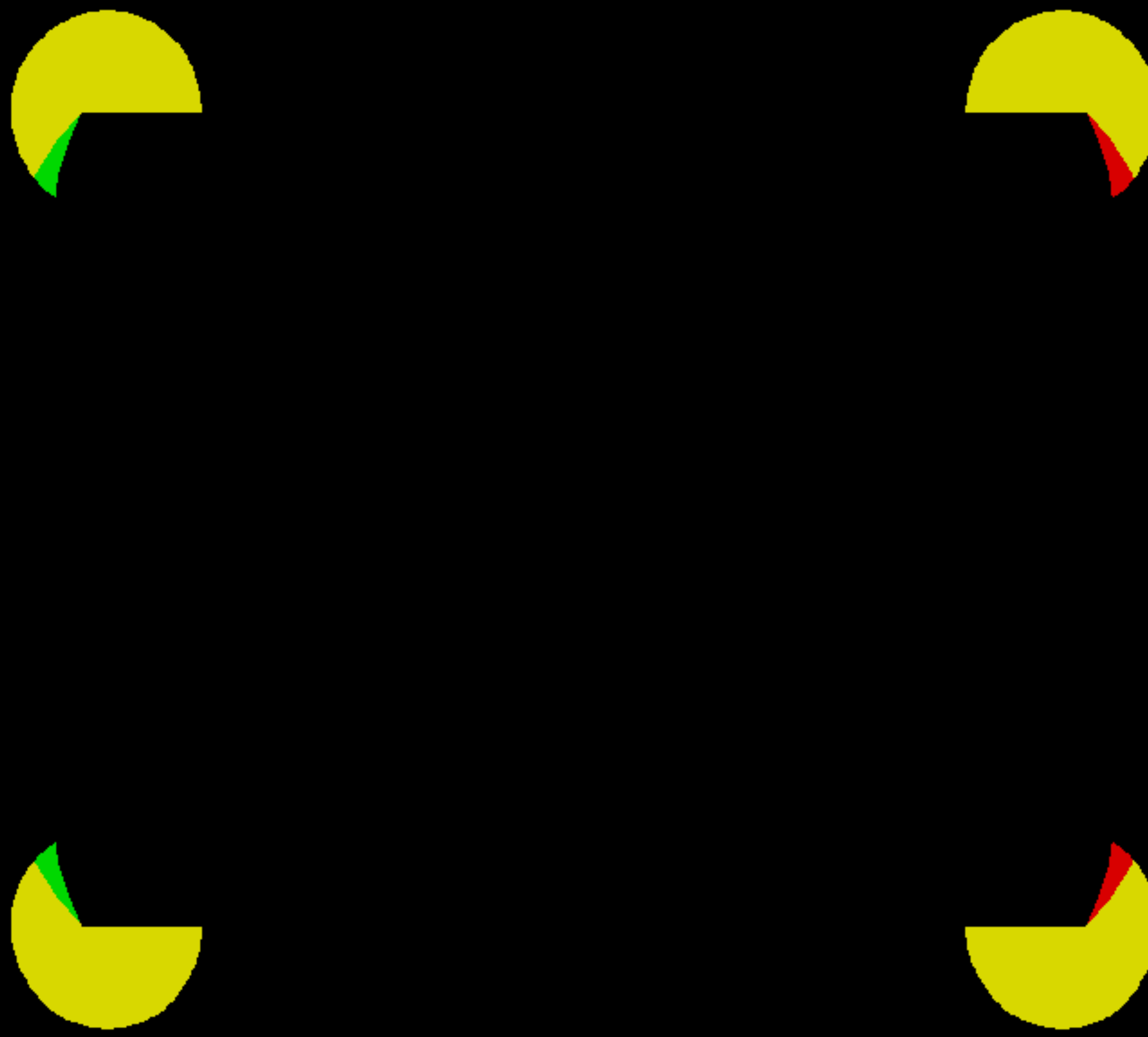


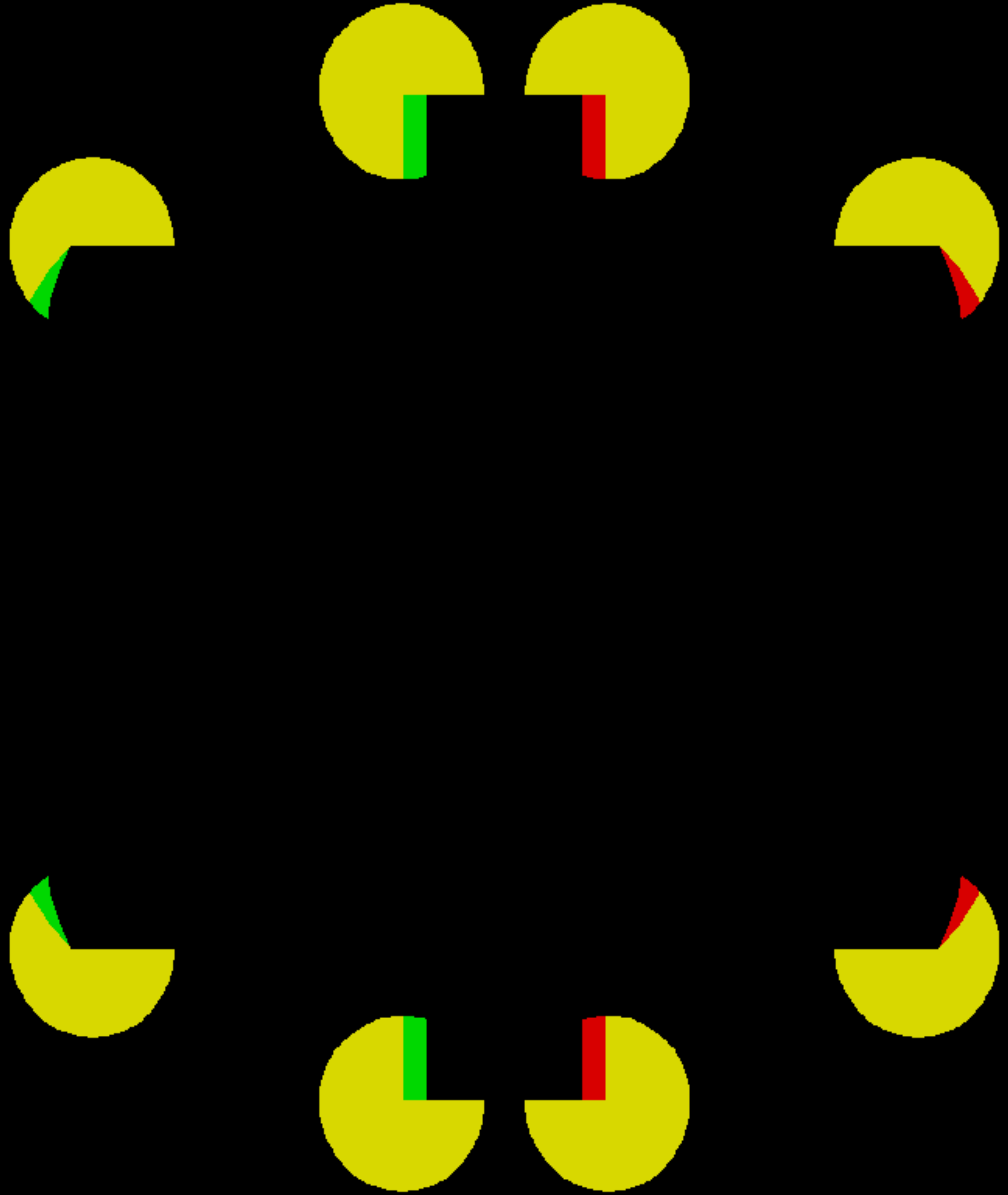
Status

- No extant computer vision algorithm can correctly infer 3D from such an impoverished input.
- …but humans can!









Summary

- The correspondence problem
- Epipolar geometry
- 3-D reconstruction
- Empirical examples