Introduction to binocular stereo vision

Kristian Kirk
Computer Vision and Media Technology Laboratory (CVMT)
Aalborg University (AAU)

What is binocular stereo vision?

- A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene

What is binocular stereo vision?

- A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene
- Used by humans and animals

What is binocular stereo vision?

- A way of getting depth (3-D) information about a scene from two 2-D views (images) of the scene
- Used by humans and animals
- Computational stereo vision
  - Programming machines to do stereo vision
  - Studied extensively in the past 25 years
  - Difficult; still being researched

Purpose of this talk:

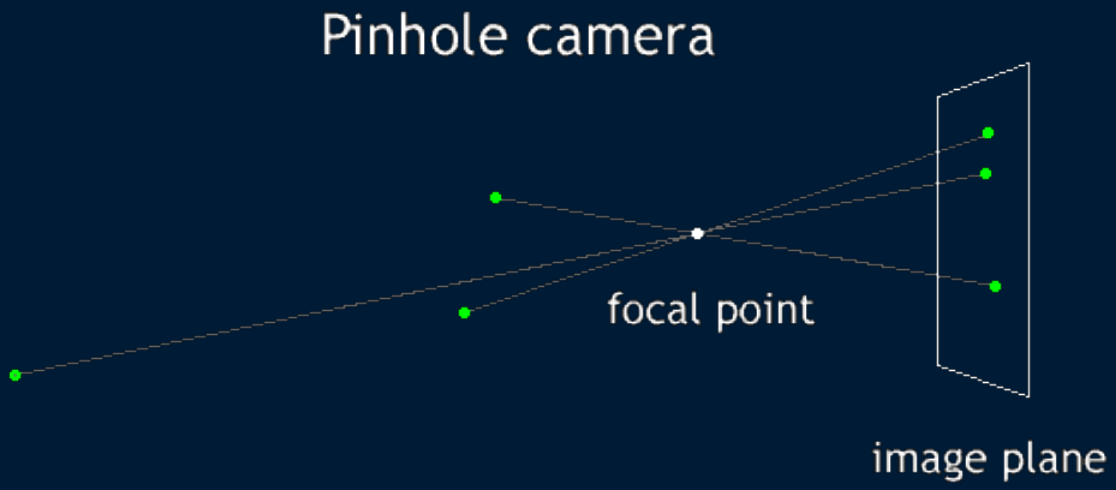
- An introduction to:
  - Basic principle of stereo vision
  - Computational stereo analysis
    - How does it work?
    - What is required?
    - Where are the difficulties?

Purpose of this talk:

- An introduction to:
  - Basic principle of stereo vision
  - Computational stereo analysis
    - How does it work?
    - What is required?
    - Where are the difficulties?

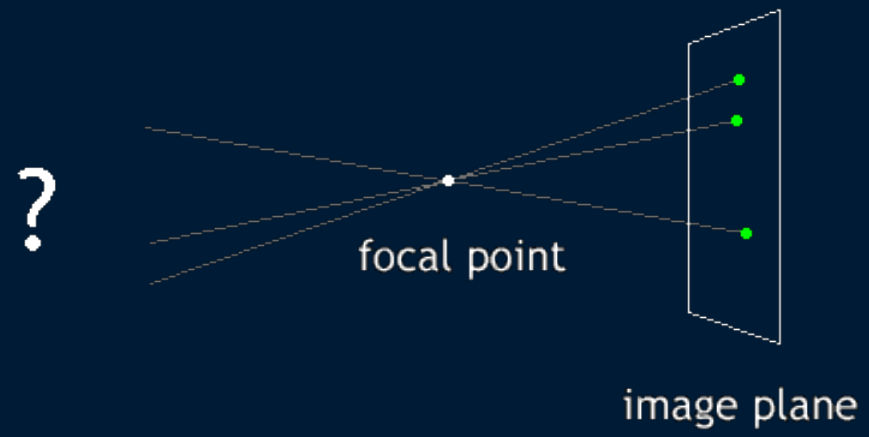
Fundamentals of stereo vision

- A camera model:
  - Models how 3-D scene points are transformed into 2-D image points
  - The pinhole camera: a simple linear model for perspective projection



Fundamentals of stereo vision

- The goal of stereo analysis:
  - The inverse process: From 2-D image coordinates to 3-D scene coordinates
  - Requires images from at least two views

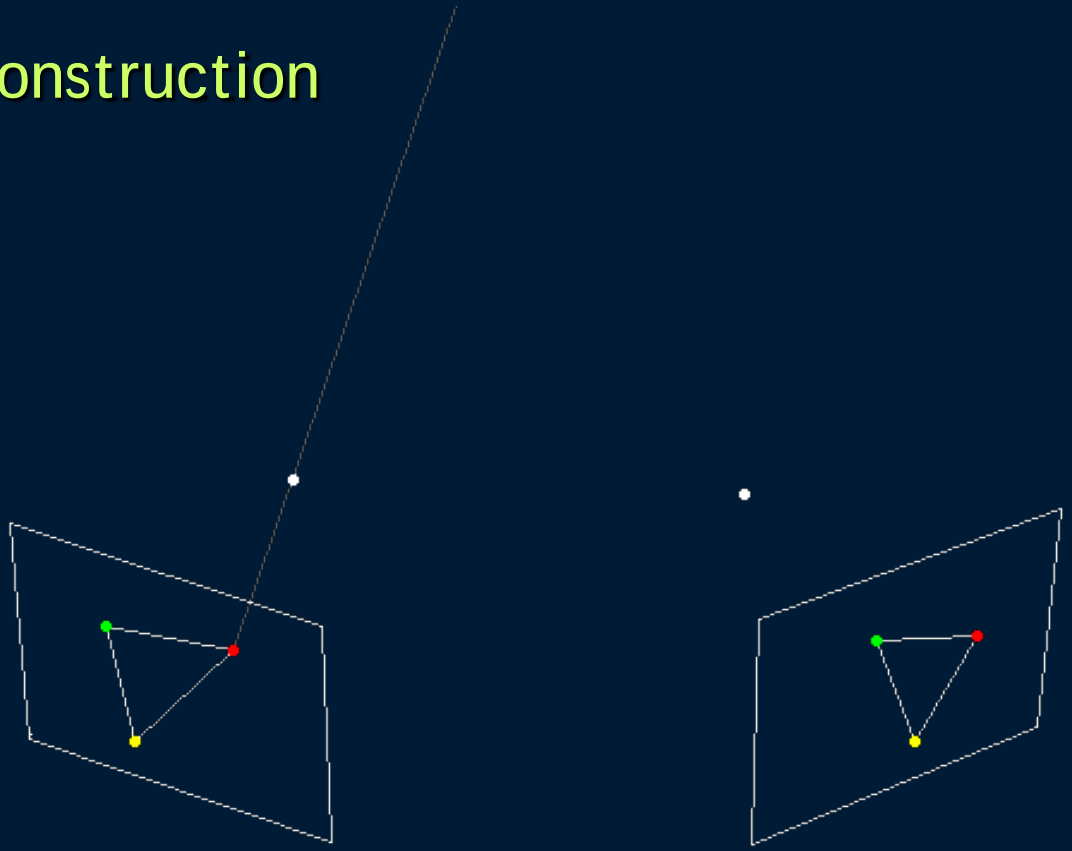


Fundamentals of stereo vision

- 3-D reconstruction

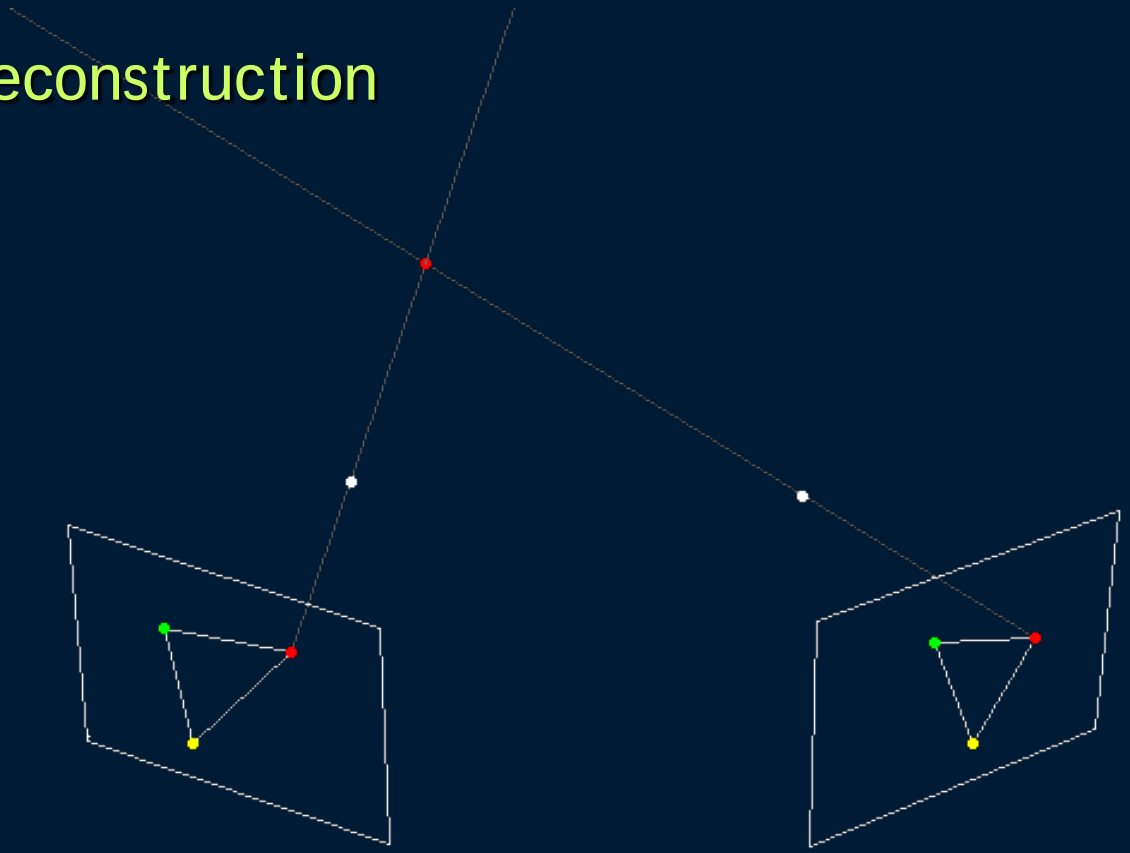
Fundamentals of stereo vision

- 3-D reconstruction



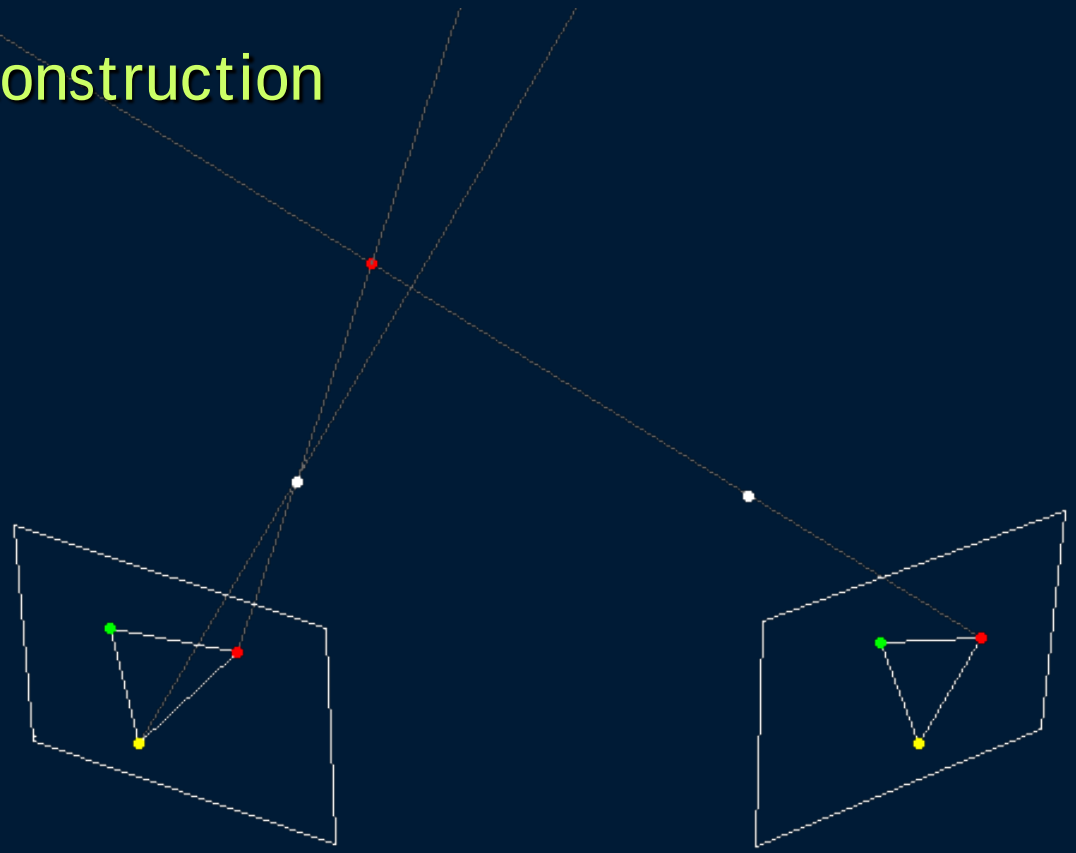
Fundamentals of stereo vision

- 3-D reconstruction



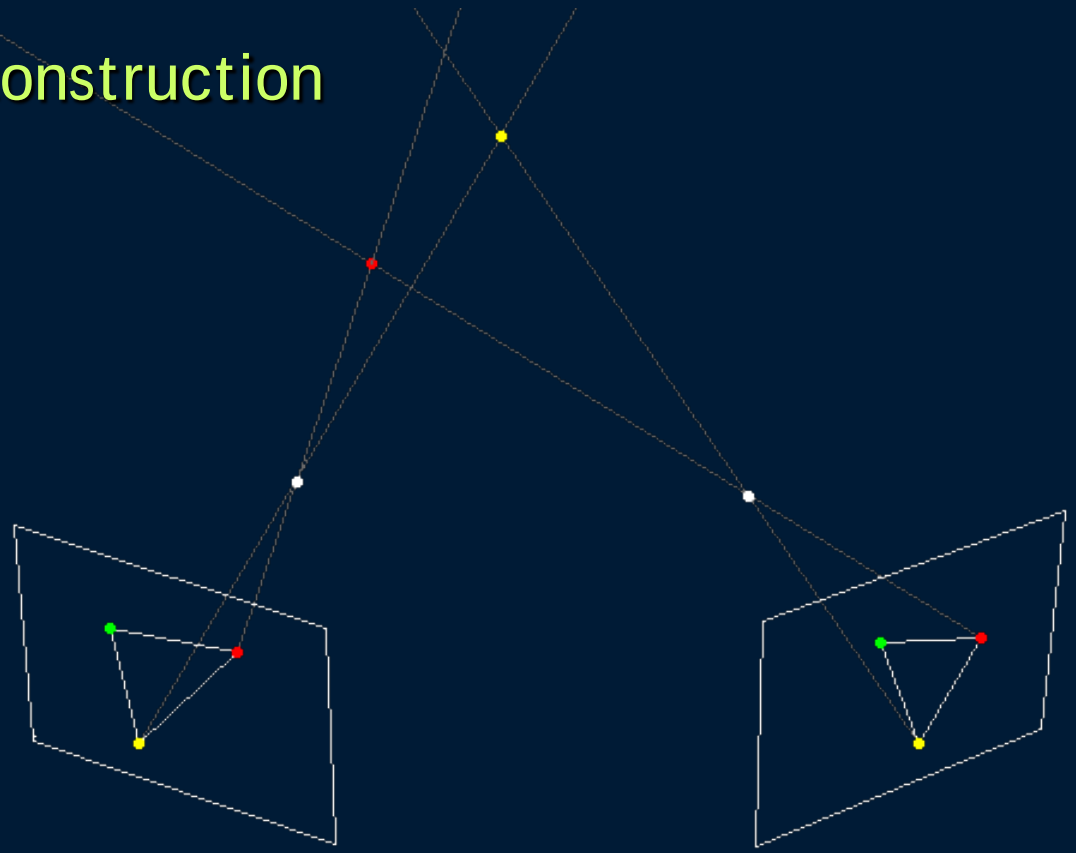
Fundamentals of stereo vision

- 3-D reconstruction



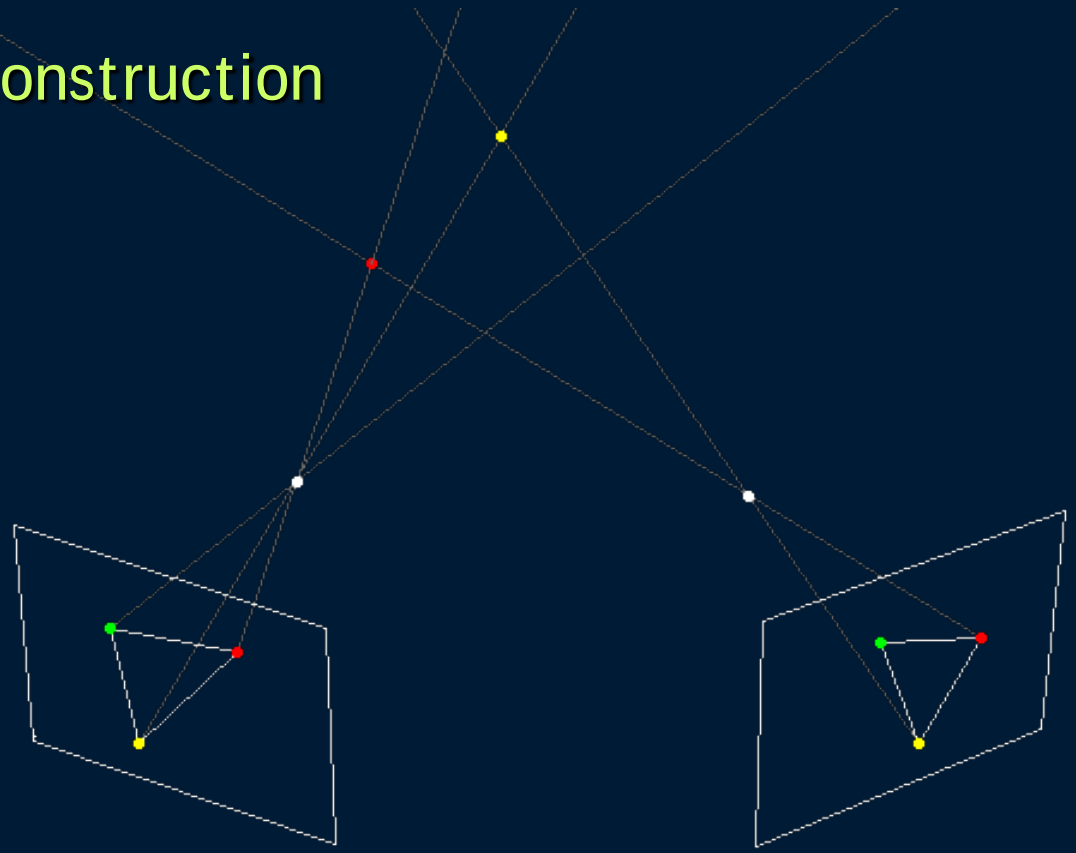
Fundamentals of stereo vision

- 3-D reconstruction



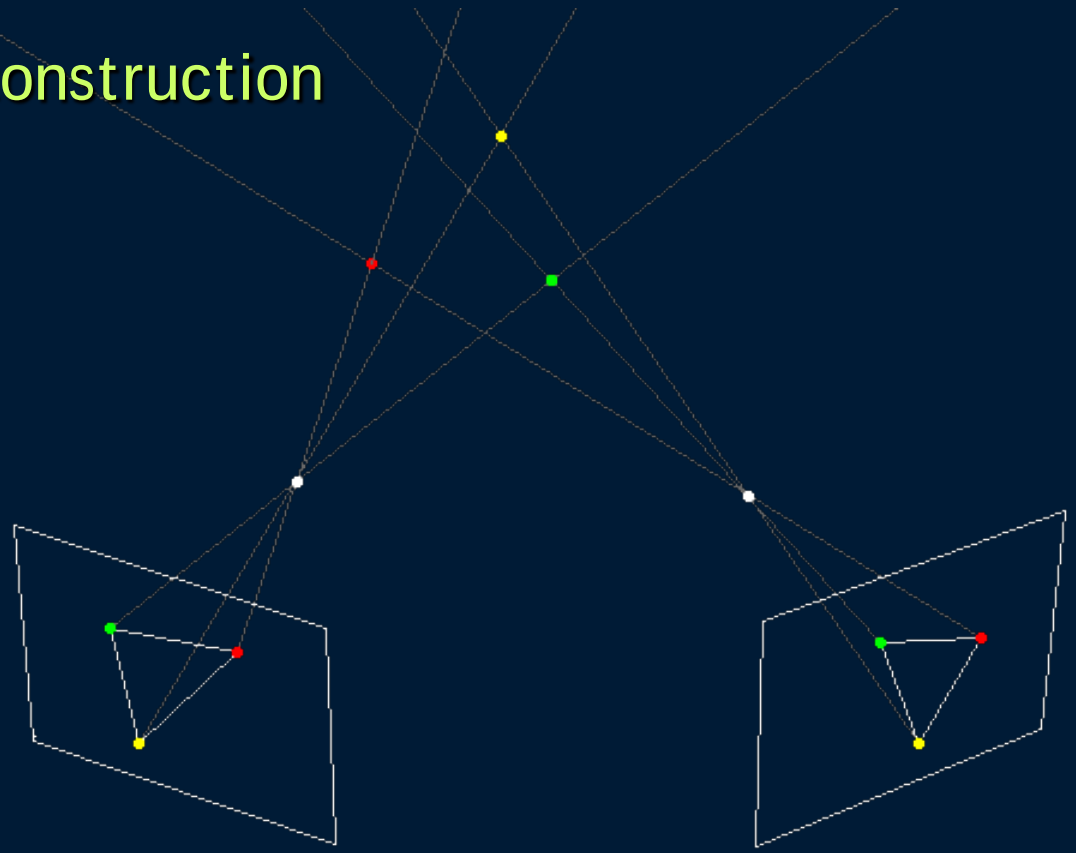
Fundamentals of stereo vision

- 3-D reconstruction



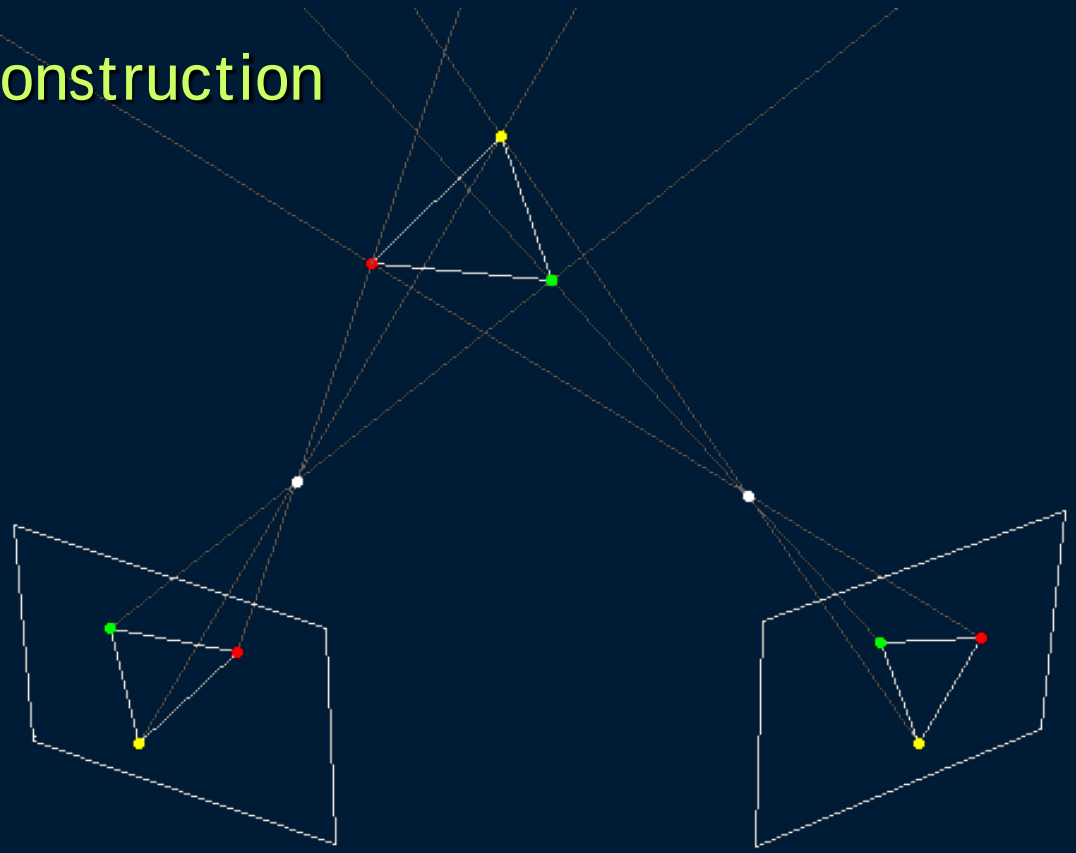
Fundamentals of stereo vision

- 3-D reconstruction



Fundamentals of stereo vision

- 3-D reconstruction

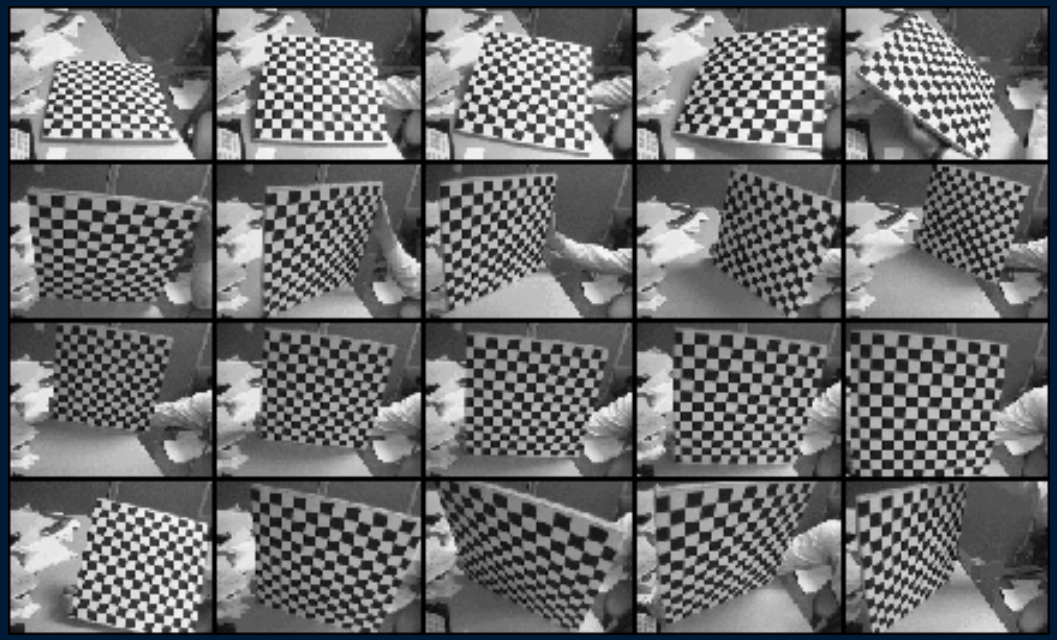


Prerequisites

- Camera model parameters must be known:
  - External parameters:
    - Positions, orientations
  - Internal parameters:
    - Focal length, image center, distortion, etc..

Prerequisites

- Camera calibration



Two subproblems

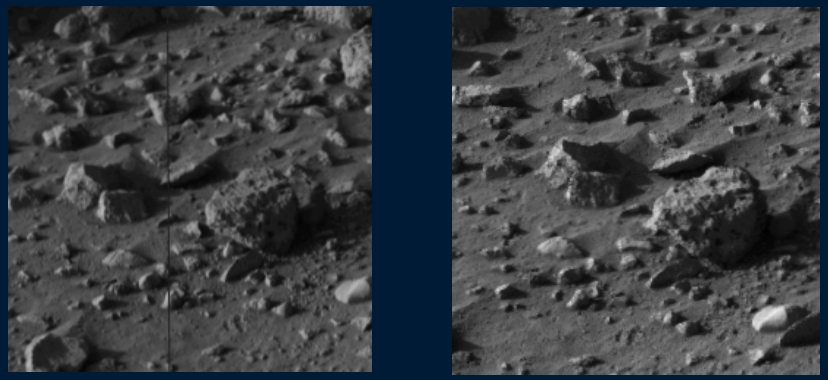
- Matching
  - Finding corresponding elements in the two images
- Reconstruction
  - Establishing 3-D coordinates from the 2-D image correspondences found during matching

Two subproblems

- Matching (hardest)
  - Finding corresponding elements in the two images
- Reconstruction
  - Establishing 3-D coordinates from the 2-D image correspondences found during matching

The matching problem

- Which image entities should be matched?
  - Two main approaches
    - Pixel/area-based (lower-level)
    - Feature-based (higher-level)



Matching challenges

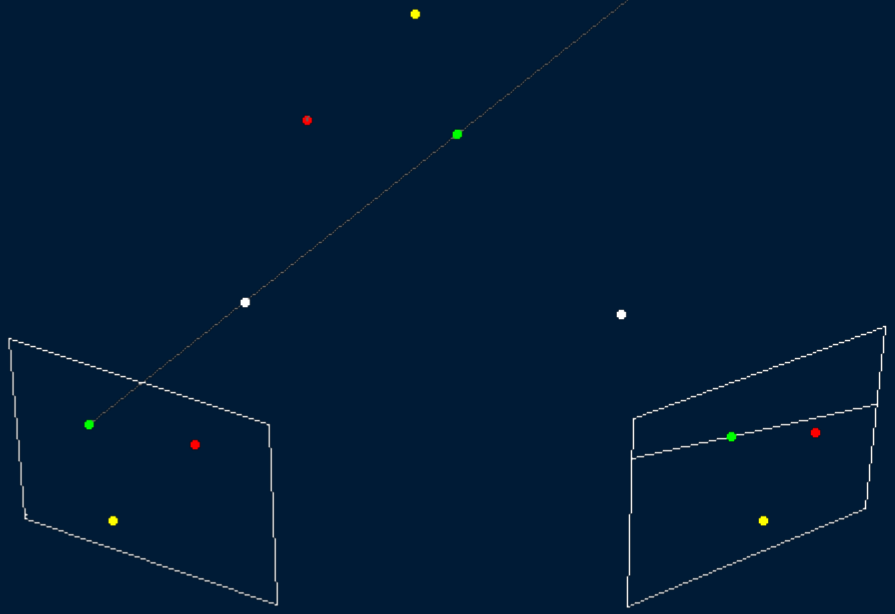
- Scene elements do not always look the same in the two images
  - Camera-related problems
    - Image noise, differing gain, contrast, etc..
  - Viewpoint-related problems:
    - Perspective distortions
    - Occlusions
    - Specular reflections

Choice of camera setup

- Baseline
  - distance between cameras (focal points)
- Trade-off
  - Small baseline: Matching easier
  - Large baseline: Depth precision better

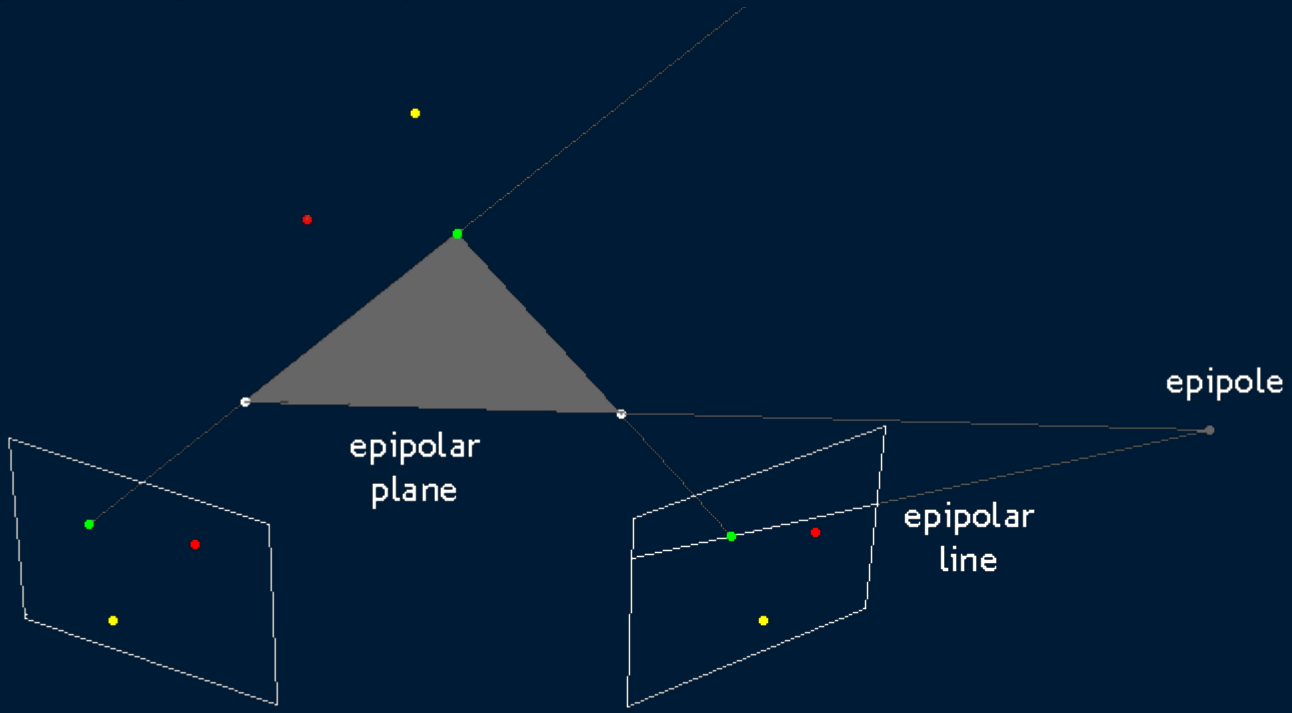
Matching clues

- Correspondance search is a 1-D problem
  - Matching point must lie on a line



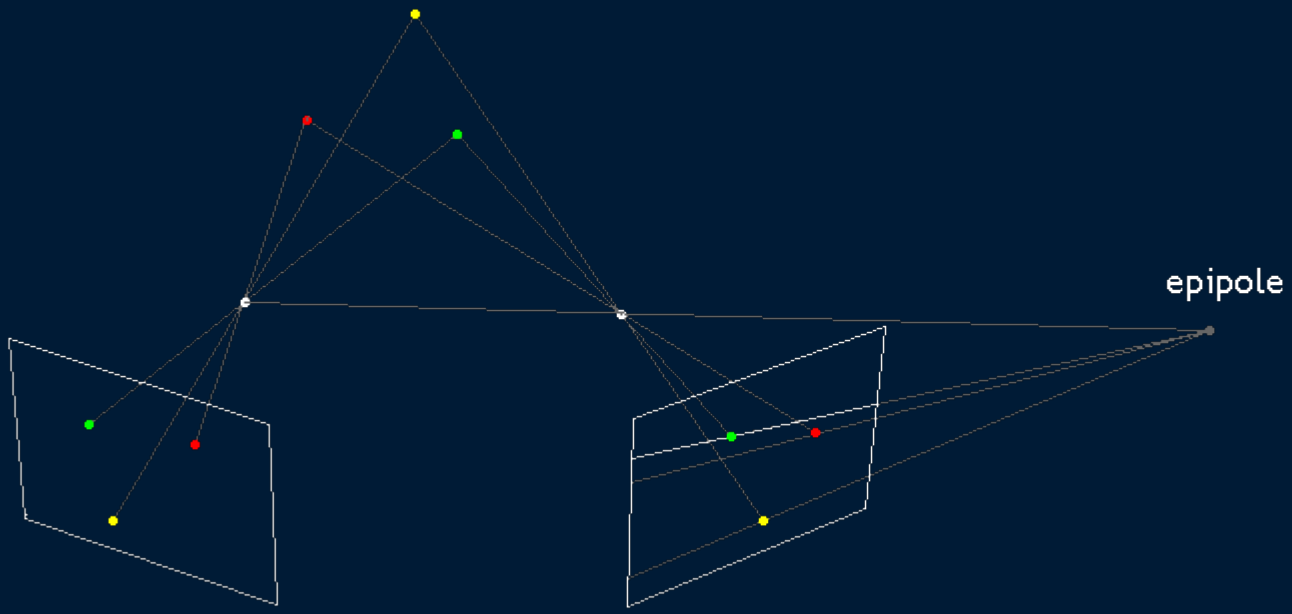
Matching clues

- Epipolar geometry



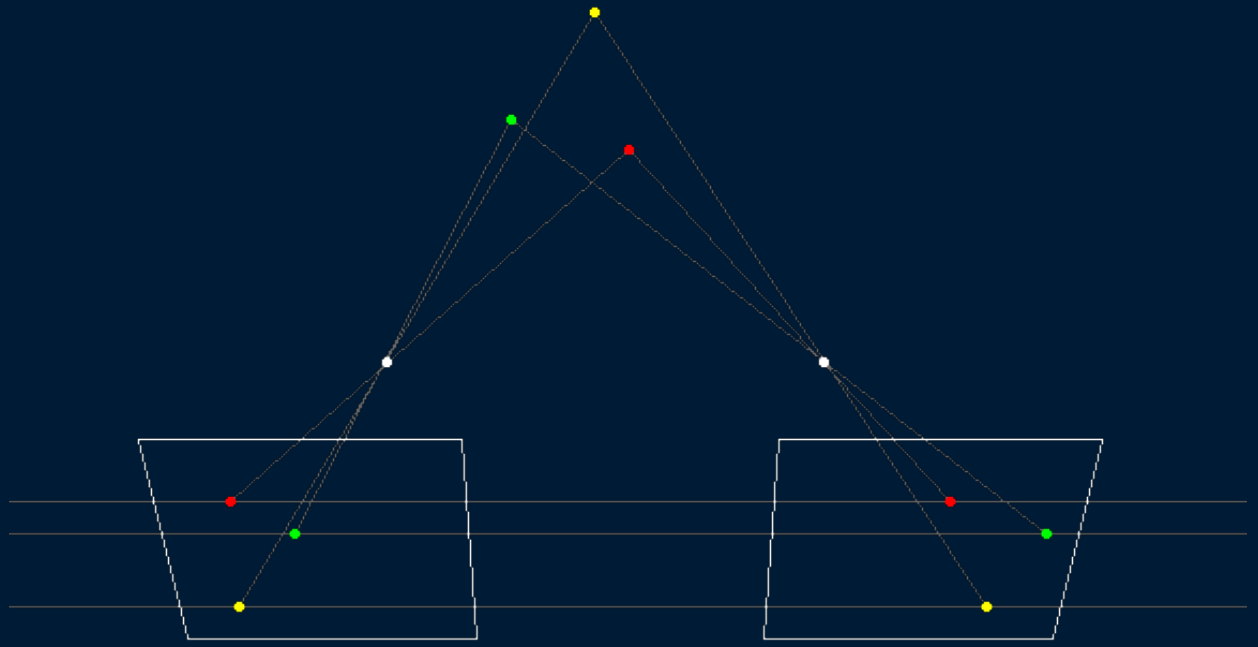
Matching clues

- Epipolar geometry



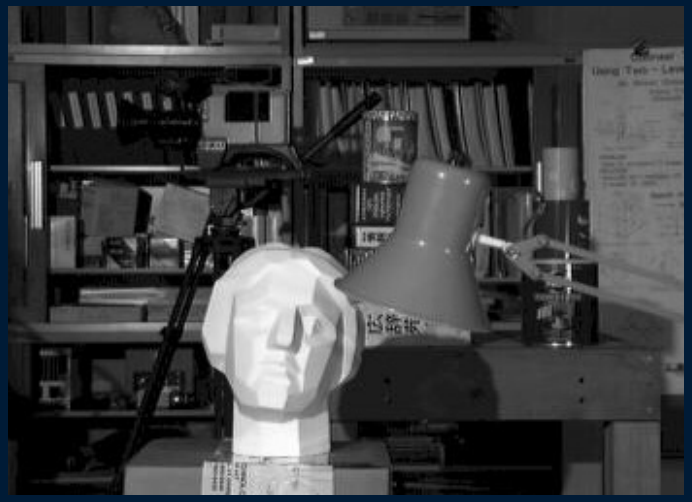
Rectification

- Simplifies the correspondance search
  - Makes all epipolar lines parallel and coincident
  - Corresponds to parallel camera configuration



Goal: disparity map

- Disparity:
  - The horizontal displacement between corresponding points
  - Closely related to scene depth

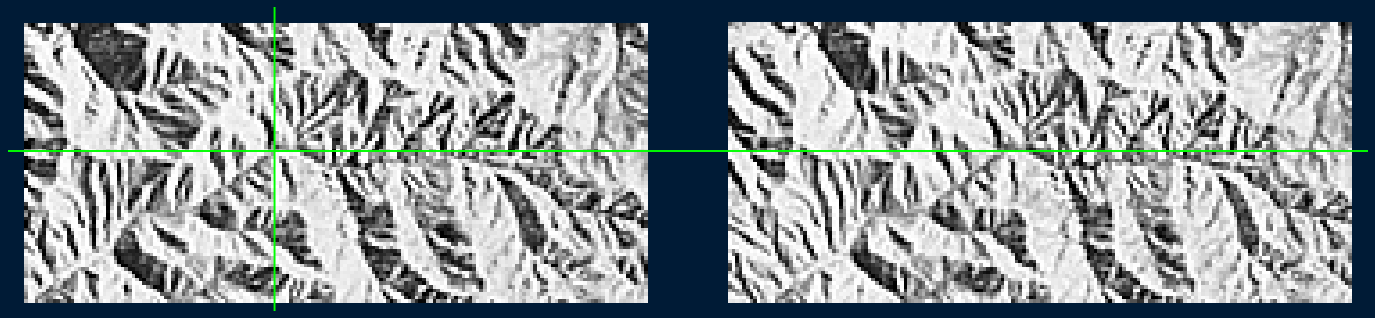


More matching heuristics

- Always valid:
  - (Epipolar line)
  - Uniqueness
  - Minimum/maximum disparity
- Sometimes valid:
  - Ordering
  - Local continuity (smoothness)

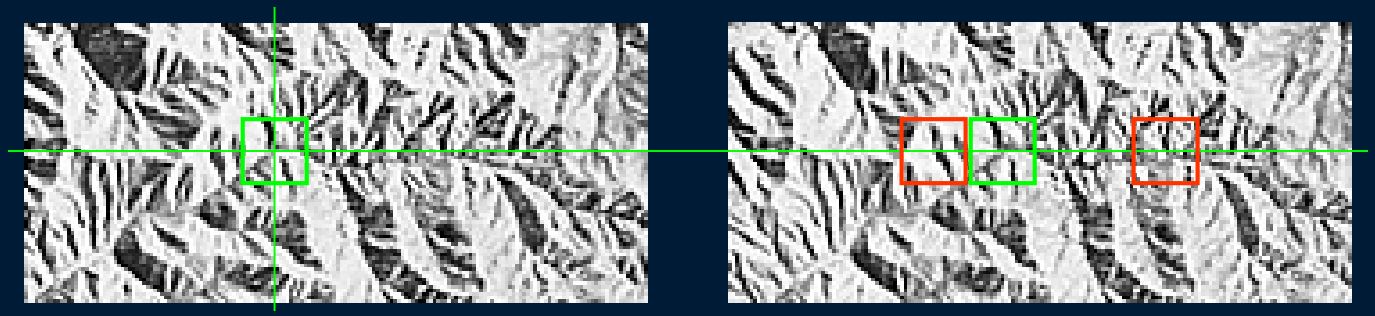
Area-based matching

- Finding pixel-to-pixel correspondences
  - For each pixel in the left image, search for the most similar pixel in the right image



Area-based matching

- Finding pixel-to-pixel correspondences
  - For each pixel in the left image, search for the most similar pixel in the right image
  - Using neighbourhood windows



Area-based matching

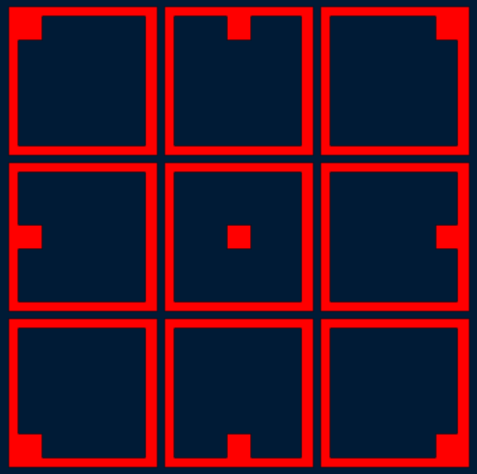
- Similarity measures for two windows
  - SAD (sum of absolute differences)
  - SSD (sum of squared differences)
  - CC (cross-correlation)
  - …

Area-based matching

- Choice of window size
  - Factors to considers:
    - Ambiguity
    - Noise sensitivity
    - Sensitivity towards viewpoint-related distortions
    - Expected object sizes
    - Frequency of depth jumps

Area-based matching

- Variable window position
  - Better matching at depth jumps (disparity edges)



Three or more viewpoints

- More matching information
  - Additional epipolar constraints
  - More confident matches

Summary

- Stereo vision:
  - A method for 3-D analysis of a scene using images from two or more viewpoints
- Two subproblems:
  - Matching
  - Reconstruction
- Most difficult part: Matching
- Area-based matching using windows
  - Low-level matching of image intensity patterns