Div. Ingeniería de Sistemas y Automática

Universidad Miguel Hernández

INTRODUCTION TO 3D VISION

fachhochschule stralsund

Table of contents

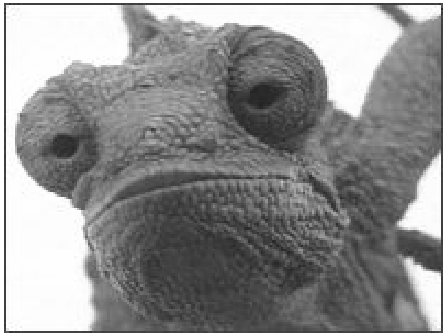
Introduction

- 3D Vision aims at estimating the spatial information of the world by means of different images taken from different viewpoints.
- We will concentrate on stereo vision

Introduction

- Most animals rely on stereo vision to extract 3D information from its environment.



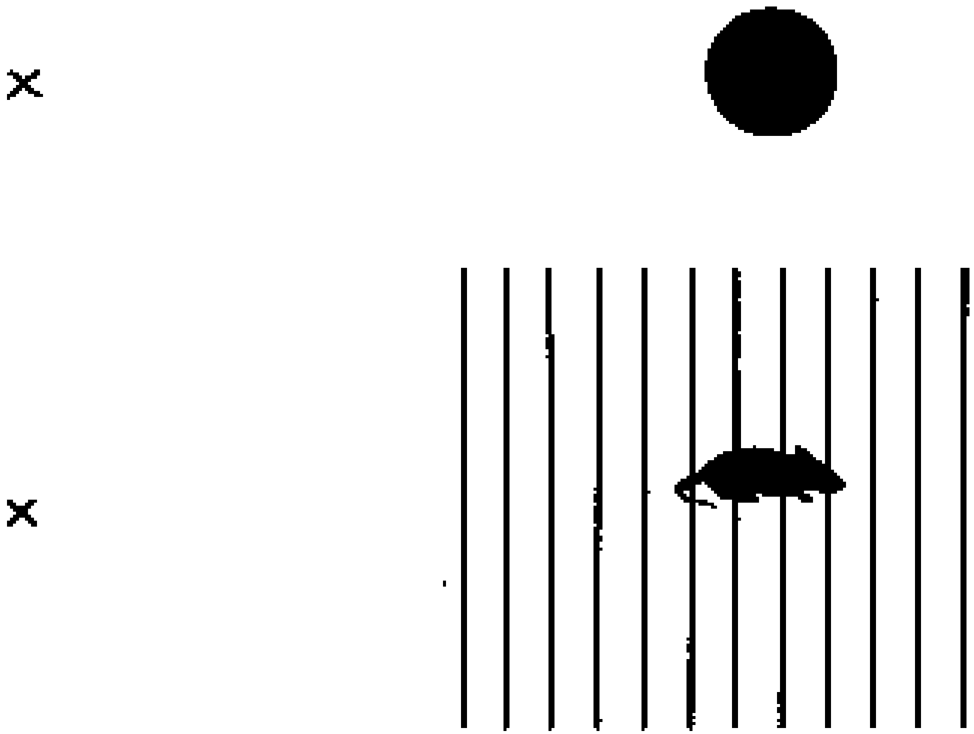


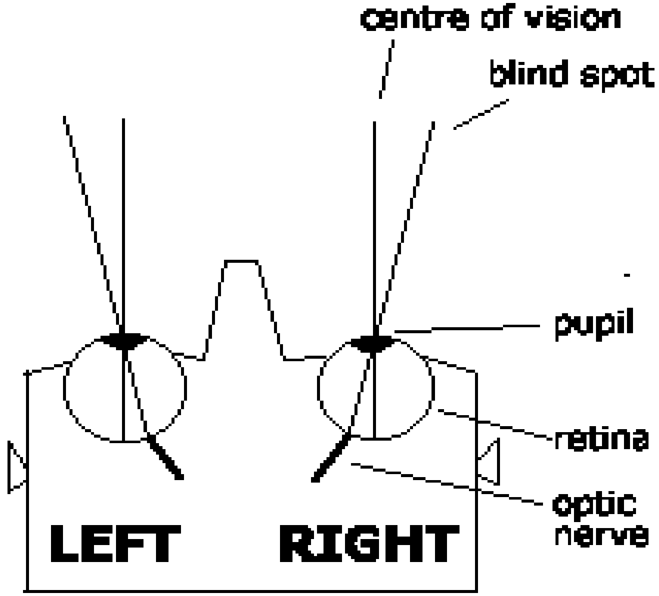

- Experiment:
  - Hold one thumb at arm’s length and close the right and left eye alternatively.
    - With (little) surprise you find that the relative position between thumb and background changes. It is this difference from viewpoint that is used to compute 3D information.
    - But… try to explain why, with both eyes open you only see one thumb.
    - And…. What about the blind spot?... It is filled with information coming from the other eye.

- Experiment 2: Blind spot.
  - Close your left eye and focus on the cross with your right eye; hold the paper at about 25cm apart from your face. Approach or separate the paper slowly until the black point “disappears”.

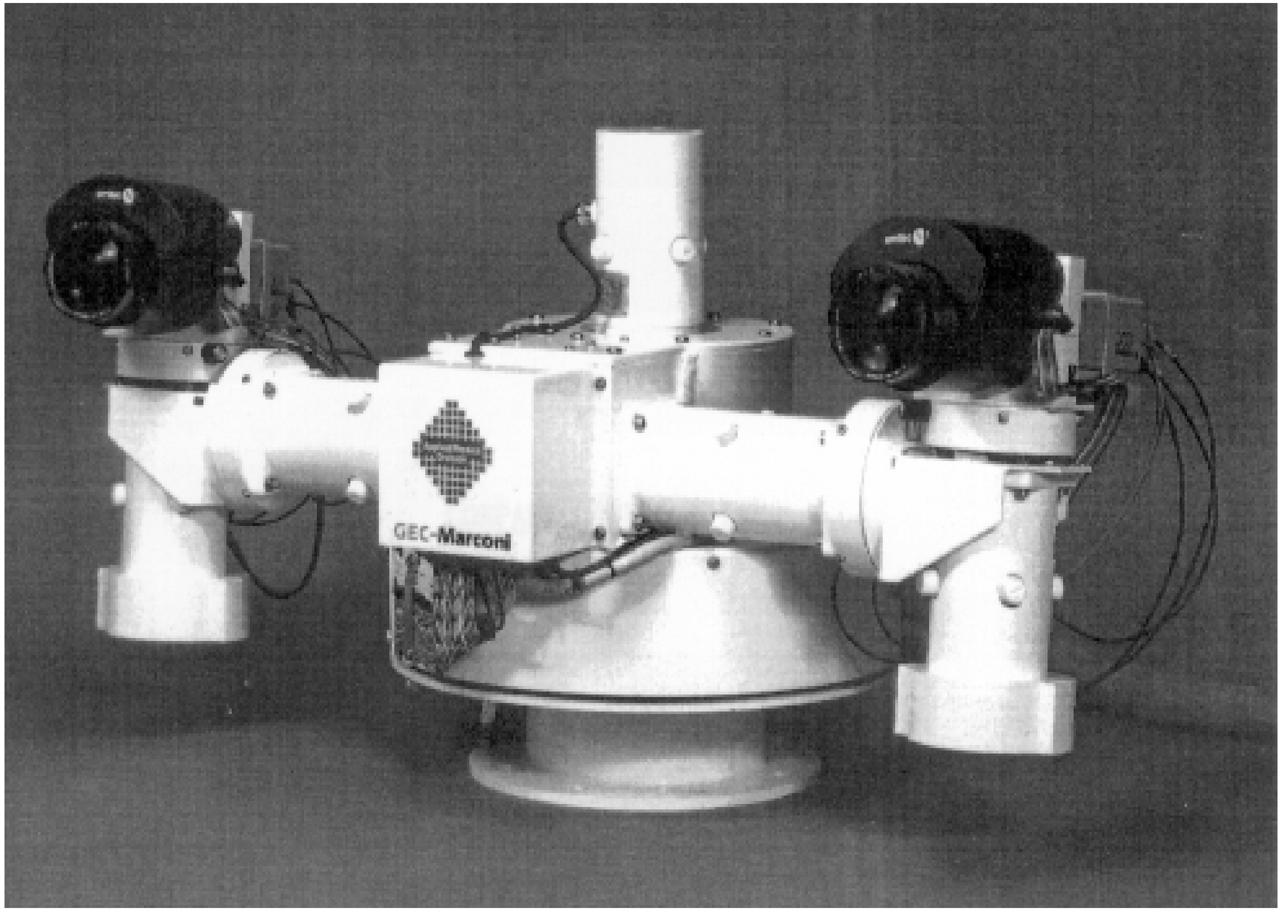
    a) Why does this fact occur?

    b) Repeat the exercise with the image below. What happens with the barrels?





- Two problems in stereo:
  - The problem of correspondence:
    - Which item in the left eye corresponds to which item in the rigth eye?
  - 3D reconstruction:
    - Our brain computes the difference in retinal position between corresponding items (disparity).
    - If the geometry of the stereo system is known, then we can build up a 3D map of the viewed scene.

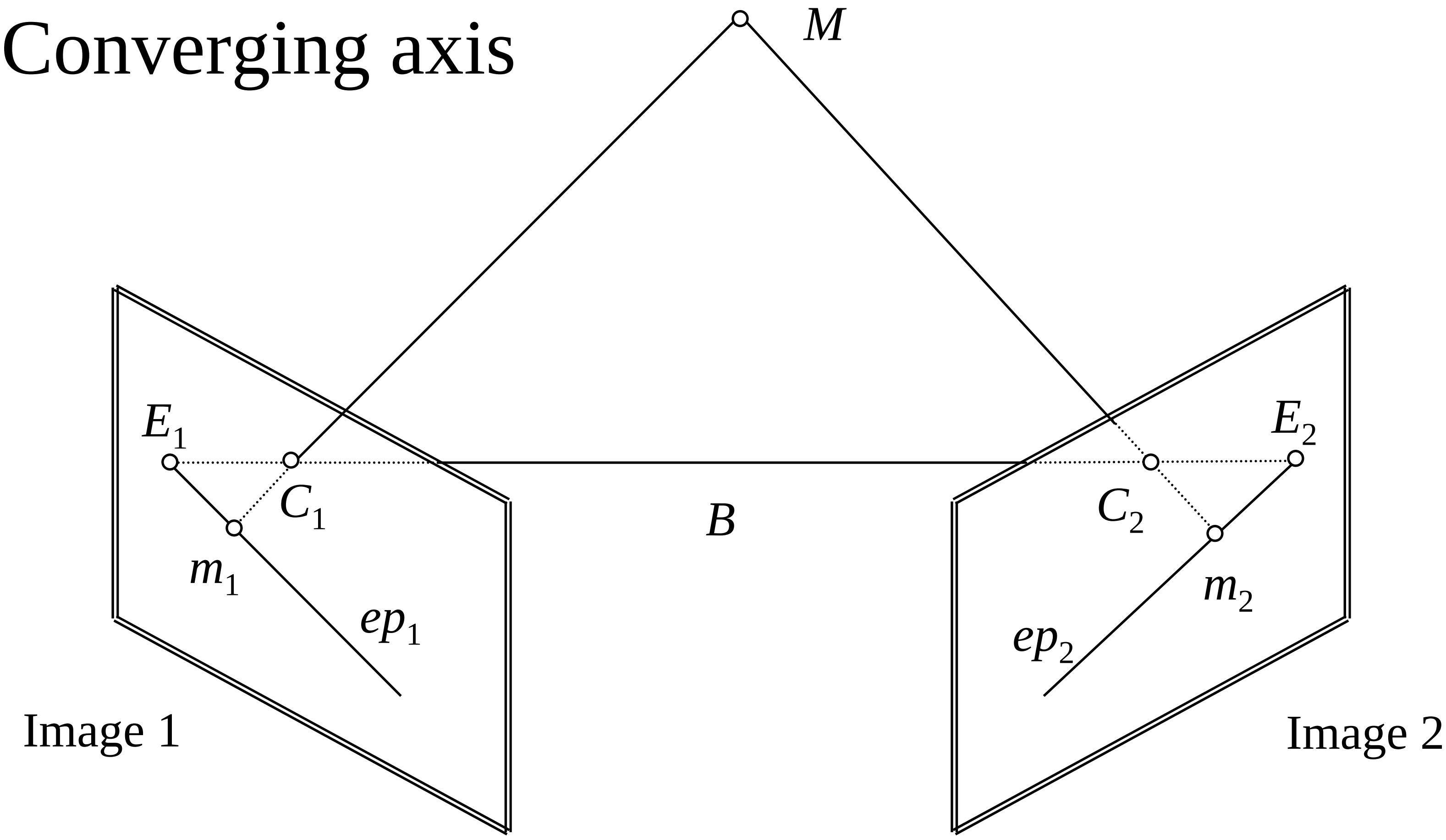


Stereo rig with (3+3+3) DOF

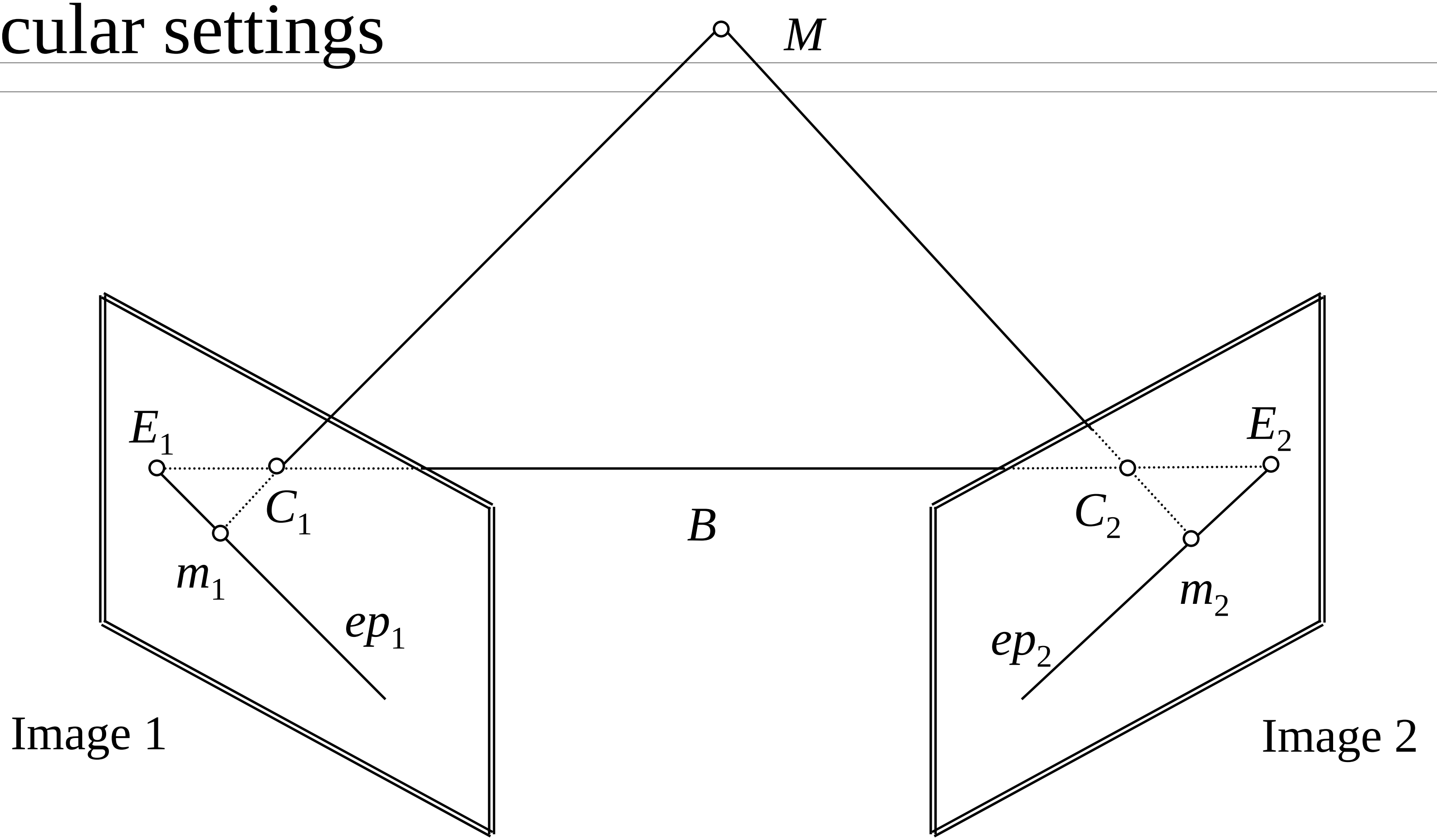
Table of contents

- *Optical axis can have different configurations:*
  - *Parrallel axis*
  - *Converging axis*
  - *Arbitrary (Axis do not cross in space)*

Converging axis



Binocular settings

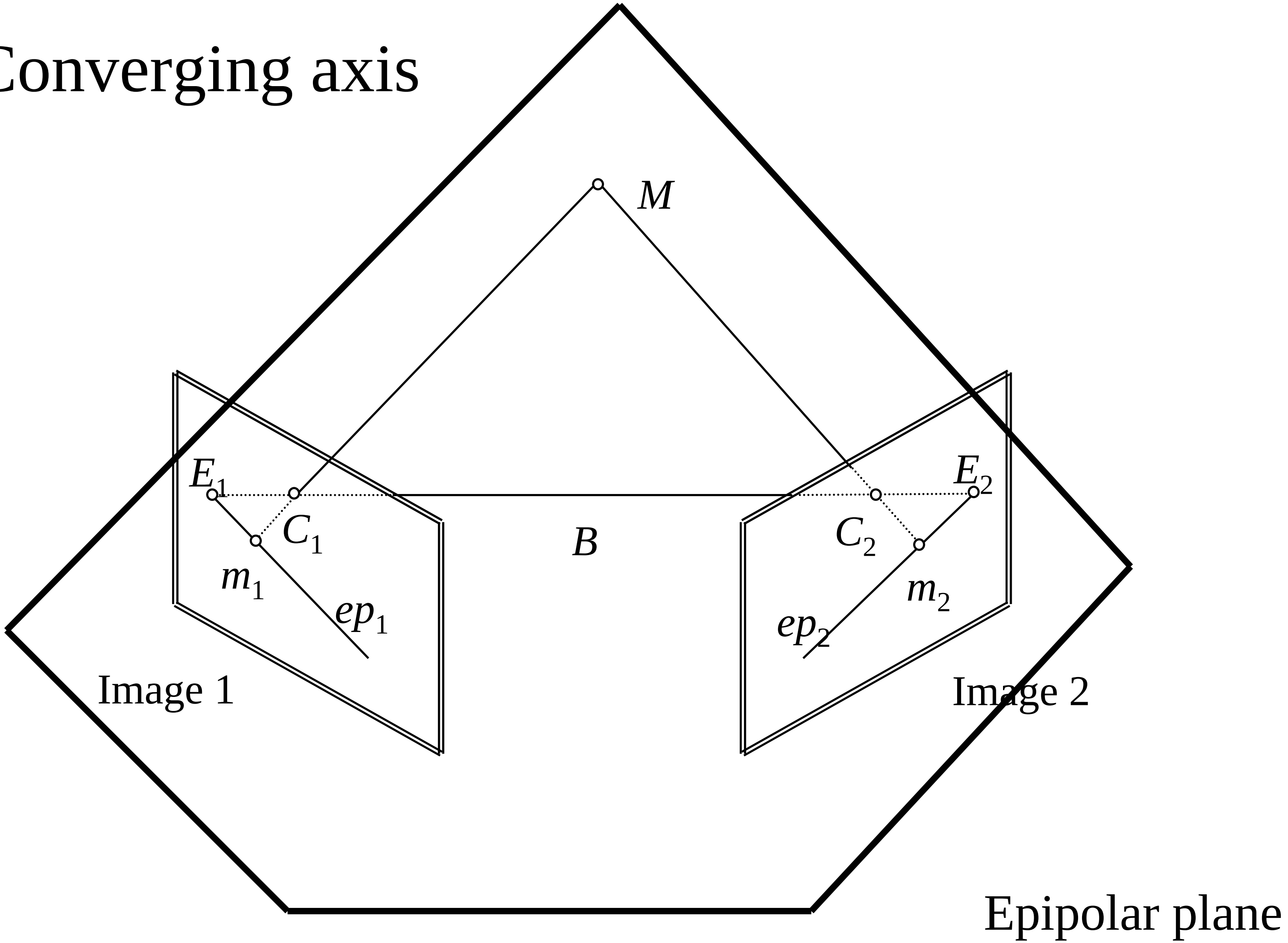


***Baseline**: Line through both optical centers*

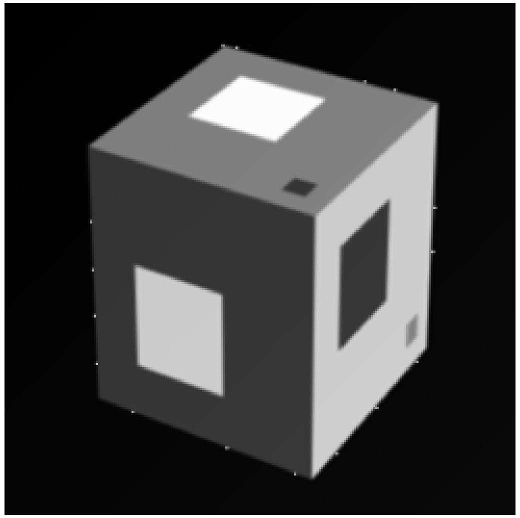
Epipolar plane: Plane that contains the optical centers C_1 and C_2 and the point M.

Epipolar line: Intersection of the epipolar plane and the each image plane.

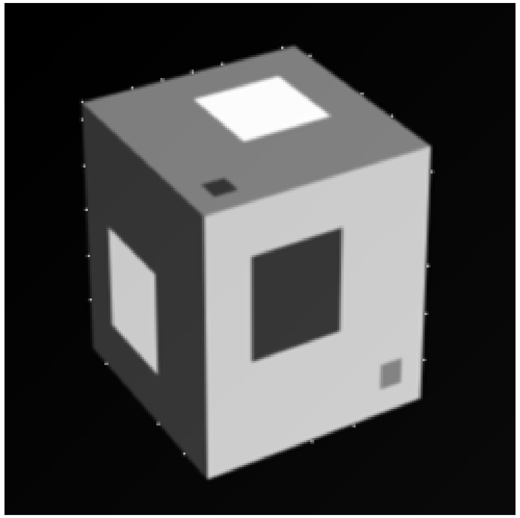
Converging axis

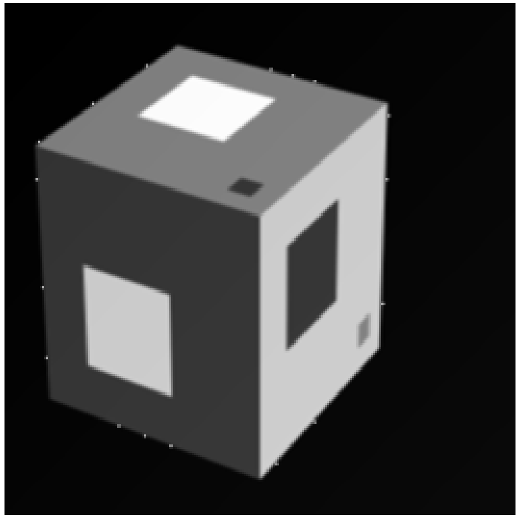


Binocular setting

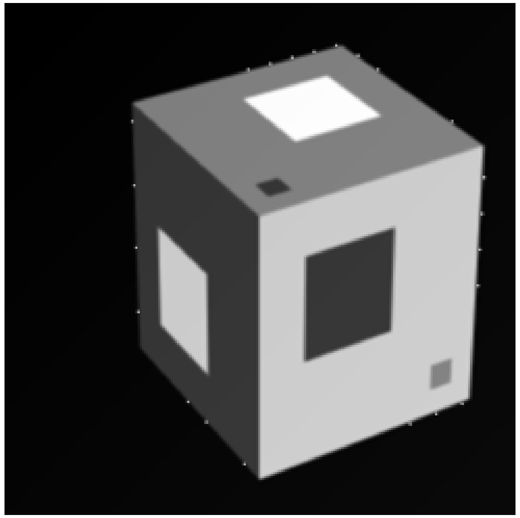


Parallel





Converging



Left camera

Right camera

Different projections

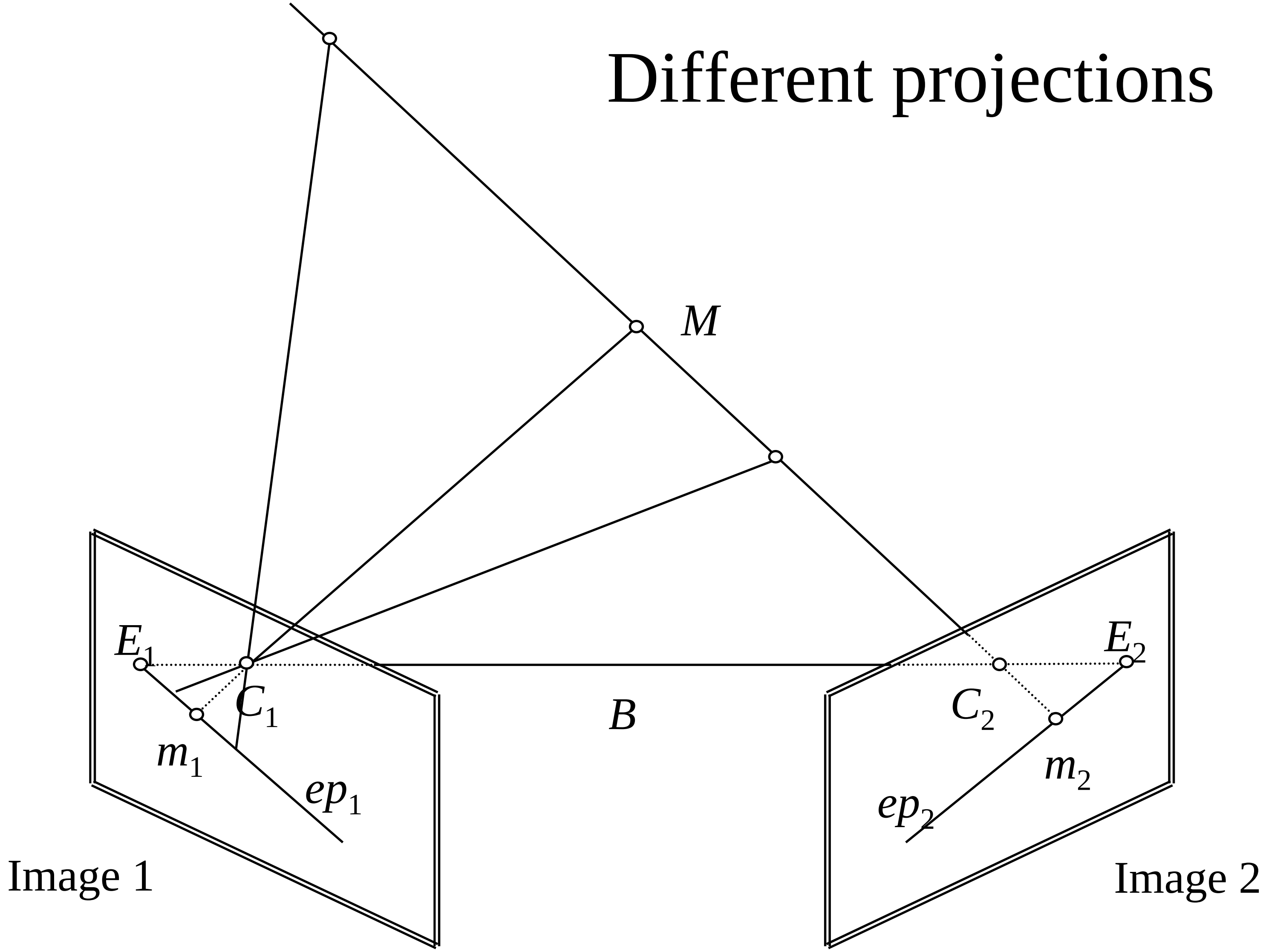
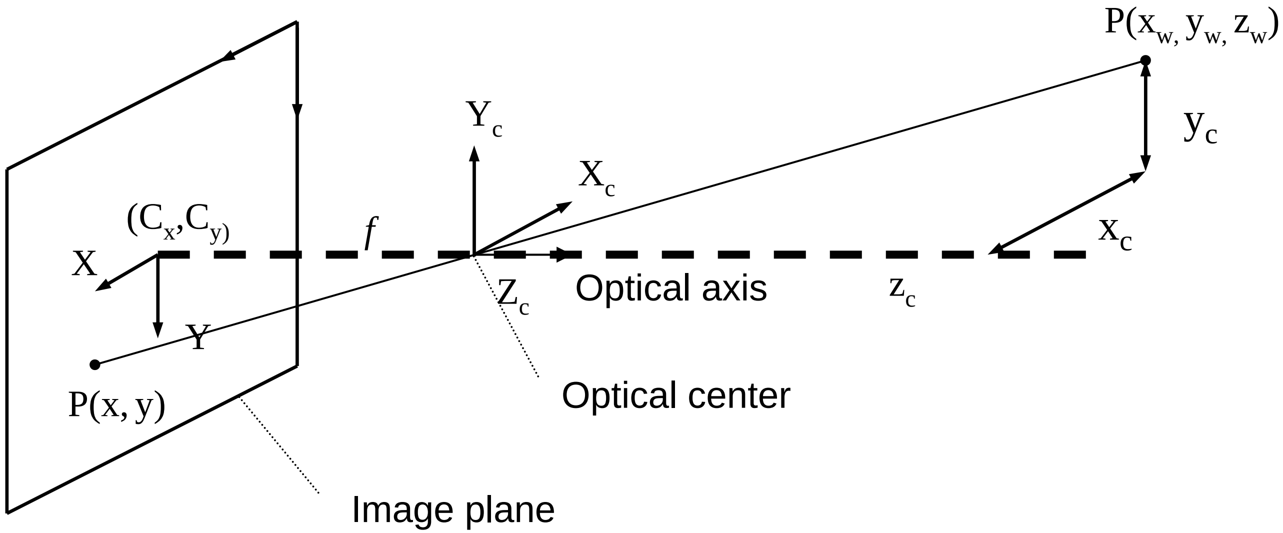


Table of contents

- ✓ Introduction
- ✓ Binocular settings
- Single camera equations
- A simple stereo system
- Correlation based methods
- Feature based techniques

Pin-hole model



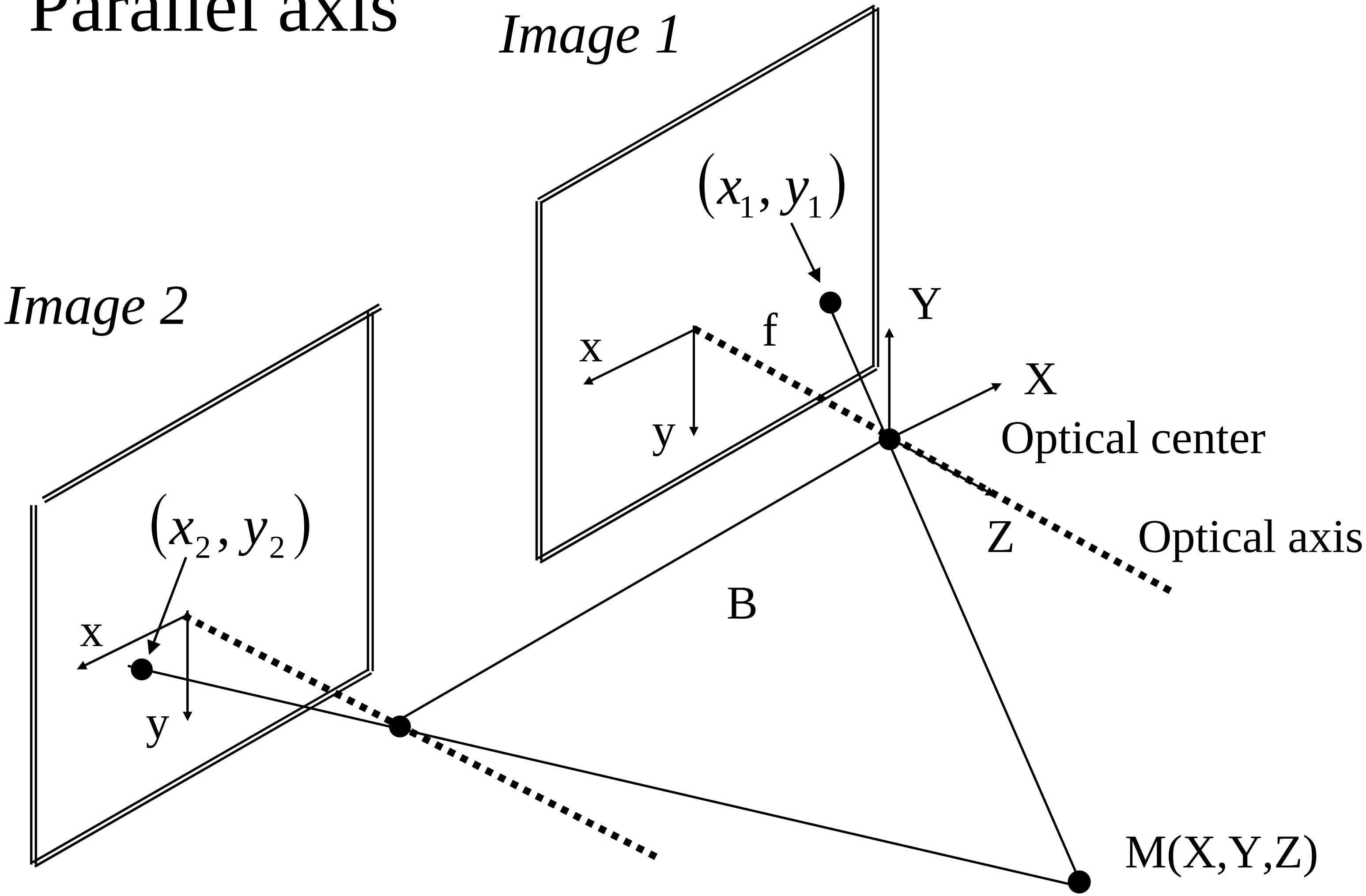
$$\frac{X}{f} = \frac{X_c}{Z_c}$$

$$\frac{Y}{f} = \frac{Y_c}{Z_c}$$

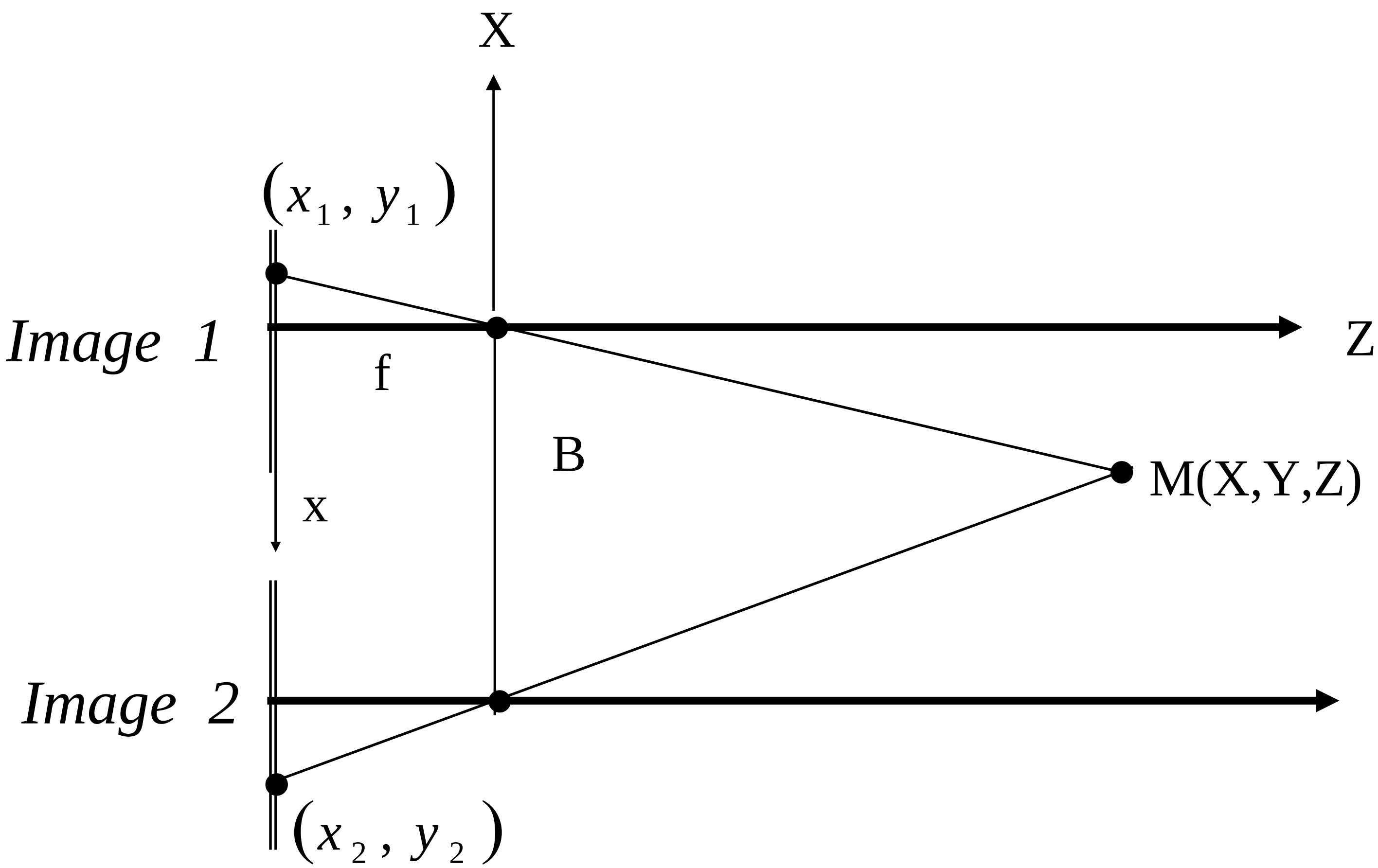
$$\begin{bmatrix} nX \\ nY \\ n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

Table of contents

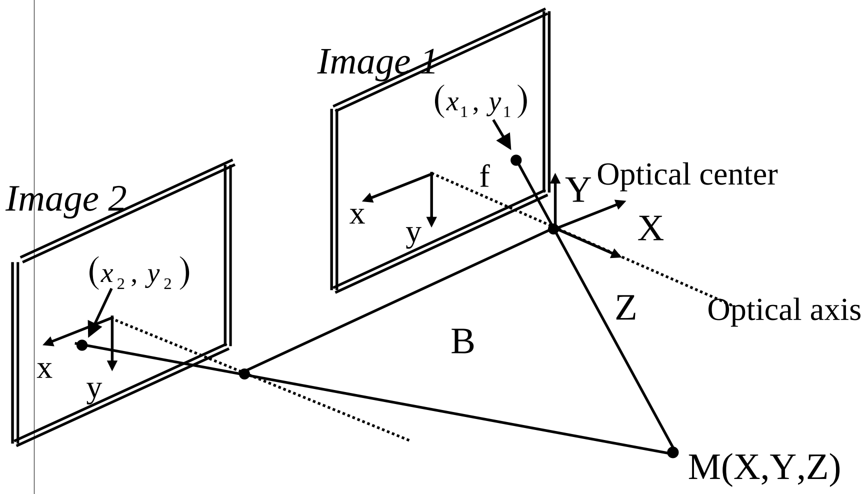
Parallel axis



Parallel axis



3D reconstruction. Parallel axis



For the first image

$$\begin{cases} X = \dfrac{x_1}{f} Z \\ Y = \dfrac{y_1}{f} Z \end{cases}$$

$$X_2 = X_1 - B$$

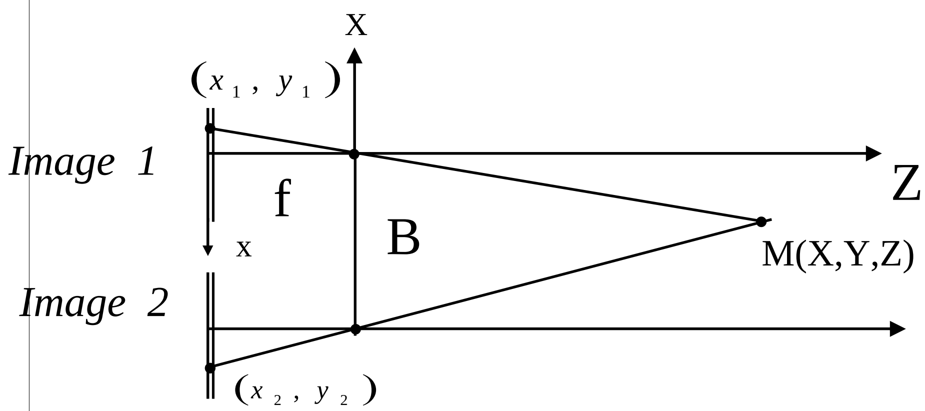
For the second image

$$\begin{cases} X = \dfrac{x_2}{f} Z + B \\ Y = \dfrac{y_2}{f} Z \end{cases}$$

X should be equal,

this lets us obtain Z :

$$Z = \frac{f\ B}{x_1 - x_2}$$



Finally :

$$X = \frac{B\ x_1}{x_1 - x_2}$$

$$Y = \frac{B\ y_1}{x_1 - x_2}$$

3D reconstruction. Parallel axis

- Notice that we are able to obtain the 3D position of point M in space, related to the coordinate system fixed at left camera.
- We can trivially relate this point to a new coordinate system, as long as we know a transformation that relates both systems.

Disparity

Point coordinates in left camera (image 1)

$$\left(x_f^{\,l}, y_f^{\,l}\right) \Rightarrow \left(x_u^{\,l}, y_u^{\,l}\right) \equiv \left(x_1, y_1\right) = m_1$$

Point coordinates in right camera (image 2)

$$\left(x_f^{\,r}, y_f^{\,r}\right) \Rightarrow \left(x_u^{\,r}, y_u^{\,r}\right) \equiv \left(x_2, y_2\right) = m_2$$

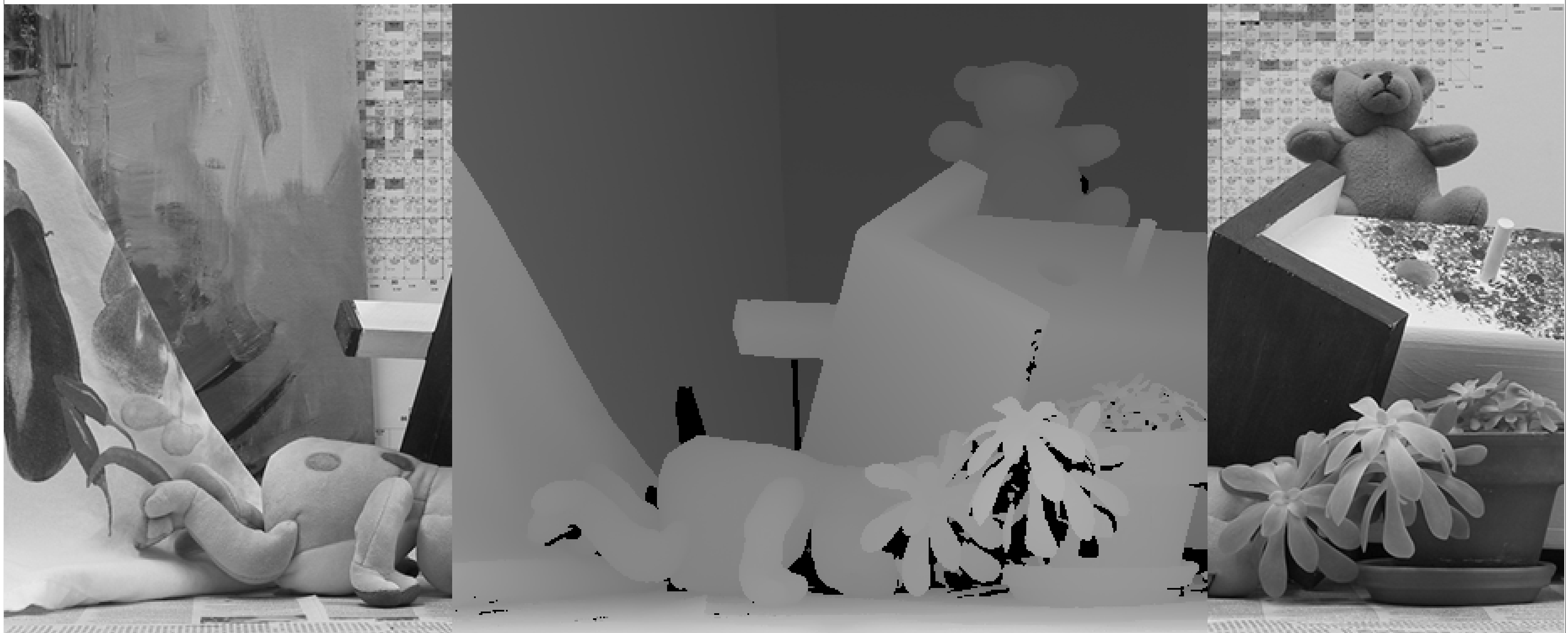
- *Parallel axis:* $d = x_1 - x_2$ **notice that** $y_1 = y_2$
  - *d* is called disparity:

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

- Converging axis $d = \left(d_1, d_2\right) = \left(x_f^l - x_f^r\ , y_f^l - y_f^r\right)$

Disparity

- There is an inverse relation between *Z* and *d*. Fixed B and f: Points with equal disparity possess the same depth *Z*.
- If B is too big, then it is more difficult to find the correspondence of a point accross images.
- If B is too short, then *d* won't be very sensitive to changes in *Z*.



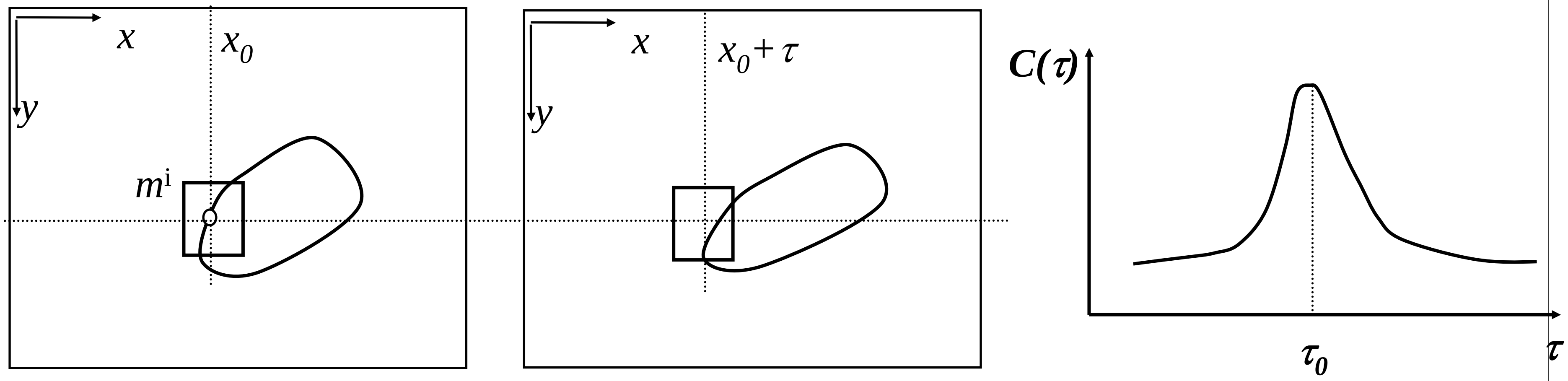
Left eye Right eye

Table of contents

Correspondence

- But… how can we find corresponding points in both images?
- We will make the following assumptions:
  - Most scene points are visible from both viewpoints.
  - Corresponding image regions are similar.
- Correspondence methods can be roughly classified in two groups:
  - Correlation-based methods.
  - Feature-based methods.

- Think… How can we look for a similar point in the other image?
- If we admit that corresponding points look similar in both images we can use a correlation measure to find correspondence.



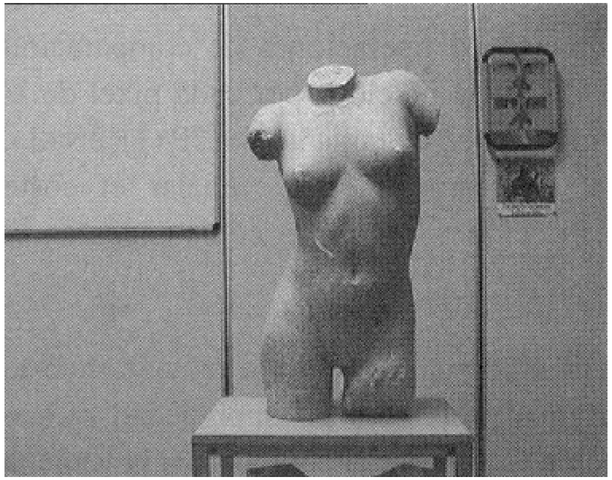
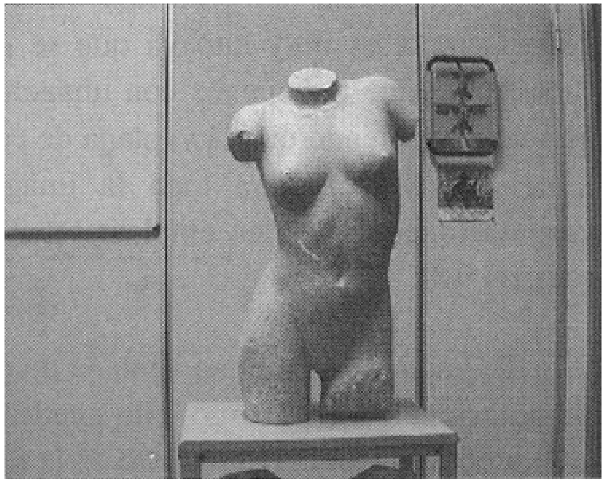
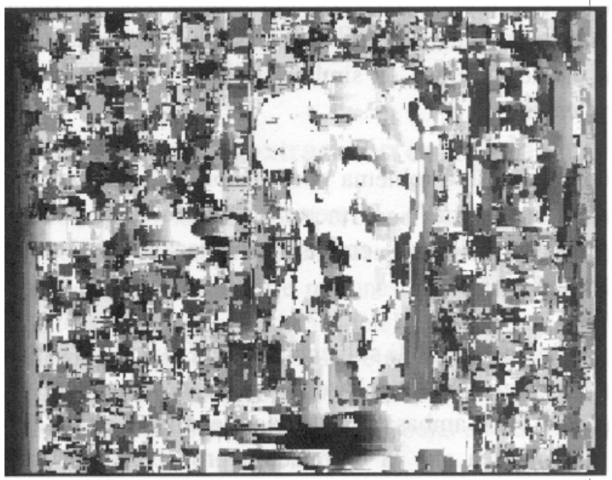
- That is: For each window m^i in the left image we calculate a correlation along the right epipolar line.
- The corresponding point maximizes correlation.

Correlation based methods

- The function to minimize is:

$$SSD(m^l(x,y), m^r(x',y')) = -\frac{1}{N} \sum_{\forall i,j \perp m(x+i,y+j) \in W(m)} \left(I^l(x+i, y+j) - I^r(x'+i, y'+j) \right)^2$$

- We obtain a dense disparity map.

- Problems arising:
  - Illumination differing for both views.
  - Low textured areas can make the method fail.
  - Points can look similar and yet not be correspondant.
  - Occlusions: Points seen in one image are not visible in the other image.

Table of contents

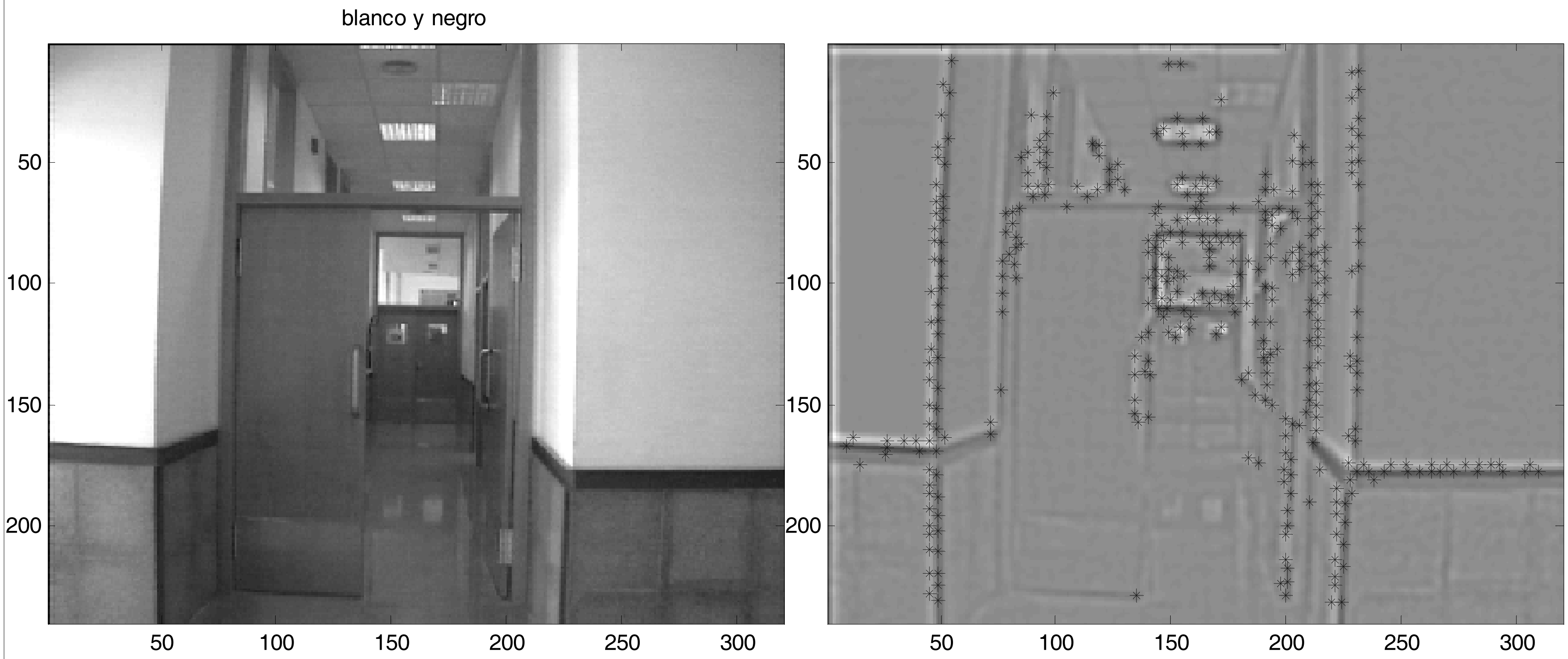
- ✓ Introduction
- ✓ Binocular settings
- ✓ Single camera equations
- ✓ A simple stereo system
- ✓ Correlation based methods
- Feature based techniques

Feature based methods

- Concept:
  - Restrict our search to some very characteristic points.
- First:
  - Extract those characteristic points: Edges, corners… etc, which are of great importance in images.
- Second:
  - Match them accross images.
- Finally:
  - Extract 3D information from correspondences.
- Result: We obtain sparse measures but highly reliable.

Feature based methods

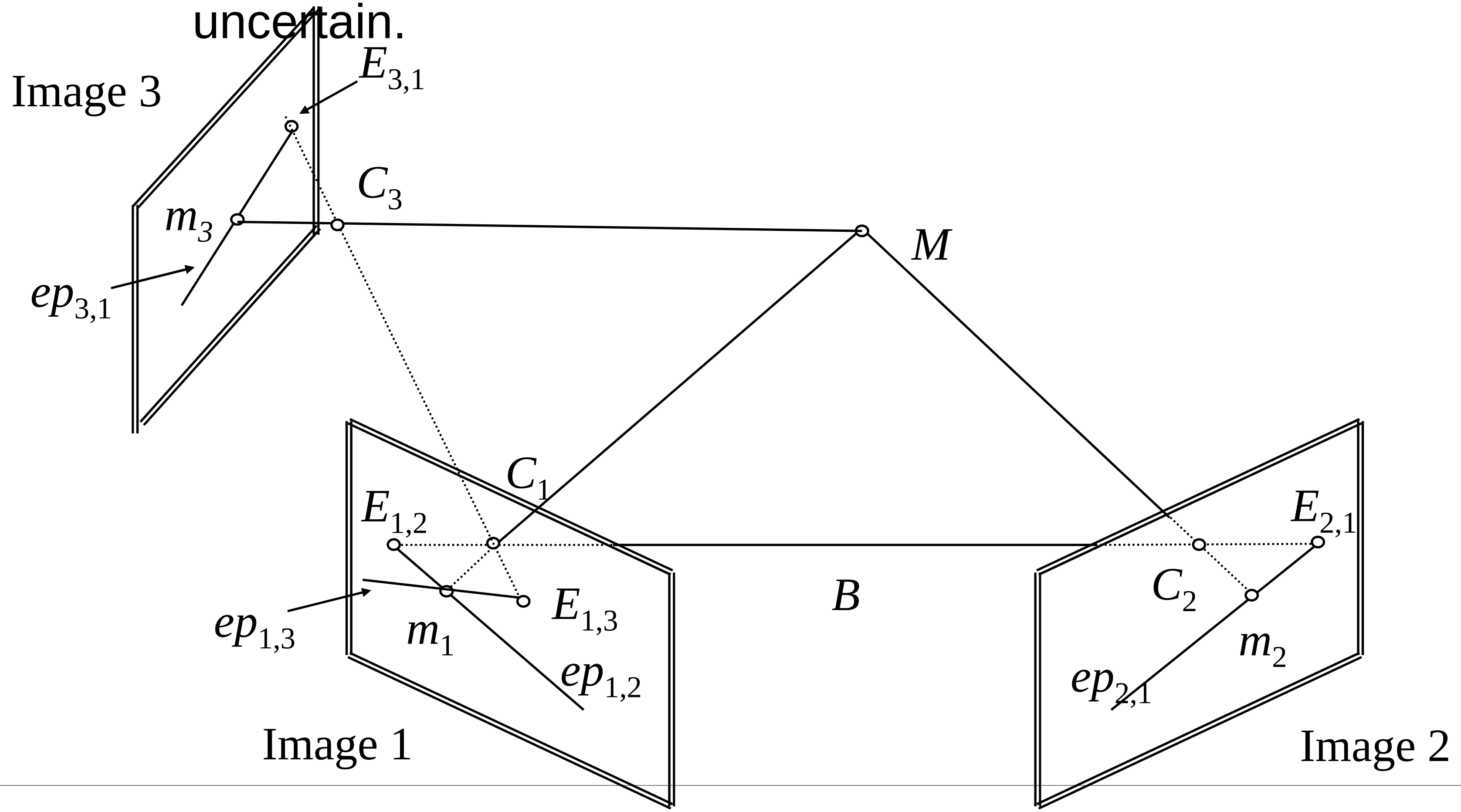
- Example:



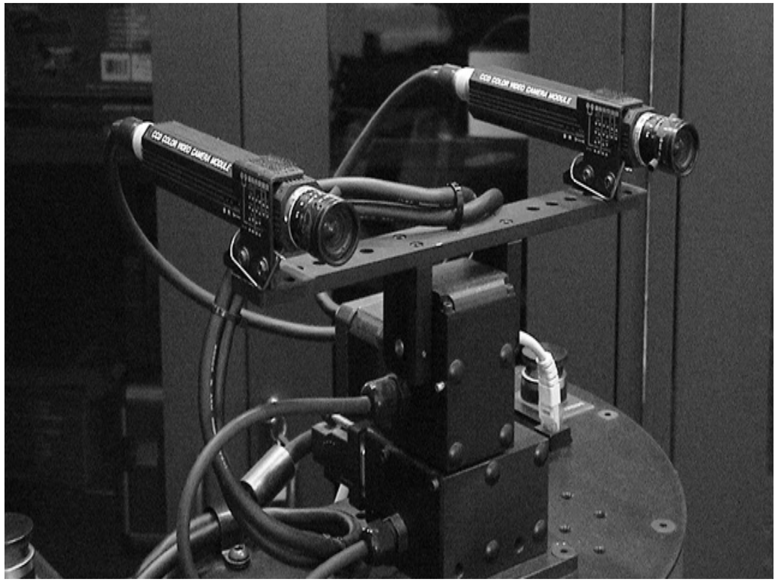
- Example:

Trinocular Vision: Adding more viewpoints

- Trinocular vision:
  - Adds more information with an extra camera.
  - It adds up computational cost.
  - It can be useful if correspondence accross two images is uncertain.



- Applications: Why?
  - Mobile robots: Try to emulate animal stereo-vision. Navigation in indoor environments.



- Monocular vision!