

StereoVision

Paulo Dias





- Introduction - Stereopsis
- Camera models and Camera calibration
- Stereo Vision
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - Stereo Matching



What is stereo?

Where is it coming from?

Where can you see it or use it?



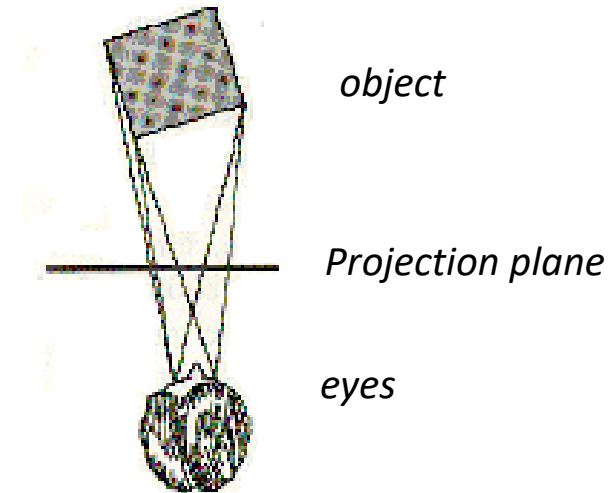
- Introduction - Stereopsis
- Camera models and Camera calibration
- Stereo Vision
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - Stereo Matching



- Many of the **perceptual cues** we use to visualize 3D structure are **available** in **2D projections**
- Such cues include:
 - occlusion (one object partially covering another)
 - perspective (point of view)
 - familiar size (we know the real-world sizes of many objects)
 - atmospheric haze (objects further away look more washed out)
- **Four cues are missing** from single 2D views:
 - stereo parallax - seeing a different image with each eye
 - movement parallax - seeing different images when we move the head
 - Accommodation - the eyes' lenses focus on the object of interest
 - Convergence - both eyes converge on the object of interest

Stereo = "solid" or "**three-dimensional**"
opsis = appearance or **sight**

Also: "binocular vision",
"binocular depth perception"
"stereoscopic depth perception"



- Stereopsis is the impression of **depth** that is perceived when a scene is viewed with **both eyes** by someone with normal binocular vision
- Binocular disparity is due to the **different position** of our two **eyes**



- Depth perception in stereo is based on stereopsis:
 - when the brain registers and fuses two images
 - **Image parallax** means that the two eyes register different images (horizontal shift)
 - The amount of shift depends on the “inter-pupillary distance” (IPD) (varies for each person in the range of 53-73 mm)
 - Works in the **near field** (to a few meters from the eye)

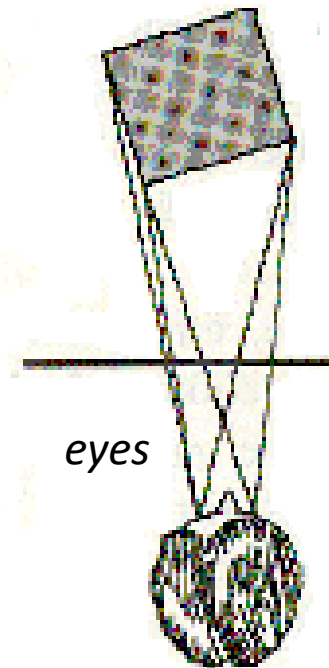


object

(Hearn and Baker, 2002)



Right eye image Left eye image



eyes

*Projection
plane*



- Present two images of the same scene
 - Two images can be presented
 - at the **same time** on two displays
- HMD**
- **time-sequenced** on one display
 - active** glasses
 - **spatially-sequenced** on one display
 - auto-stereoscopic** displays



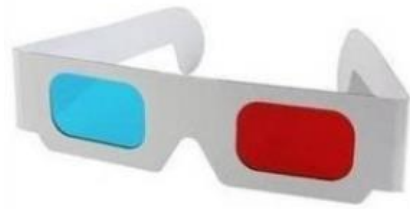
Left eye, right eye images
(Burdea and Coiffet., 2003)



Common ways to produce 3D sensation



- Anaglyphs: two colored images and color coded glasses (red/cyan(green))



- Two images with different light polarization and polarizing glasses
 - Linear and circular



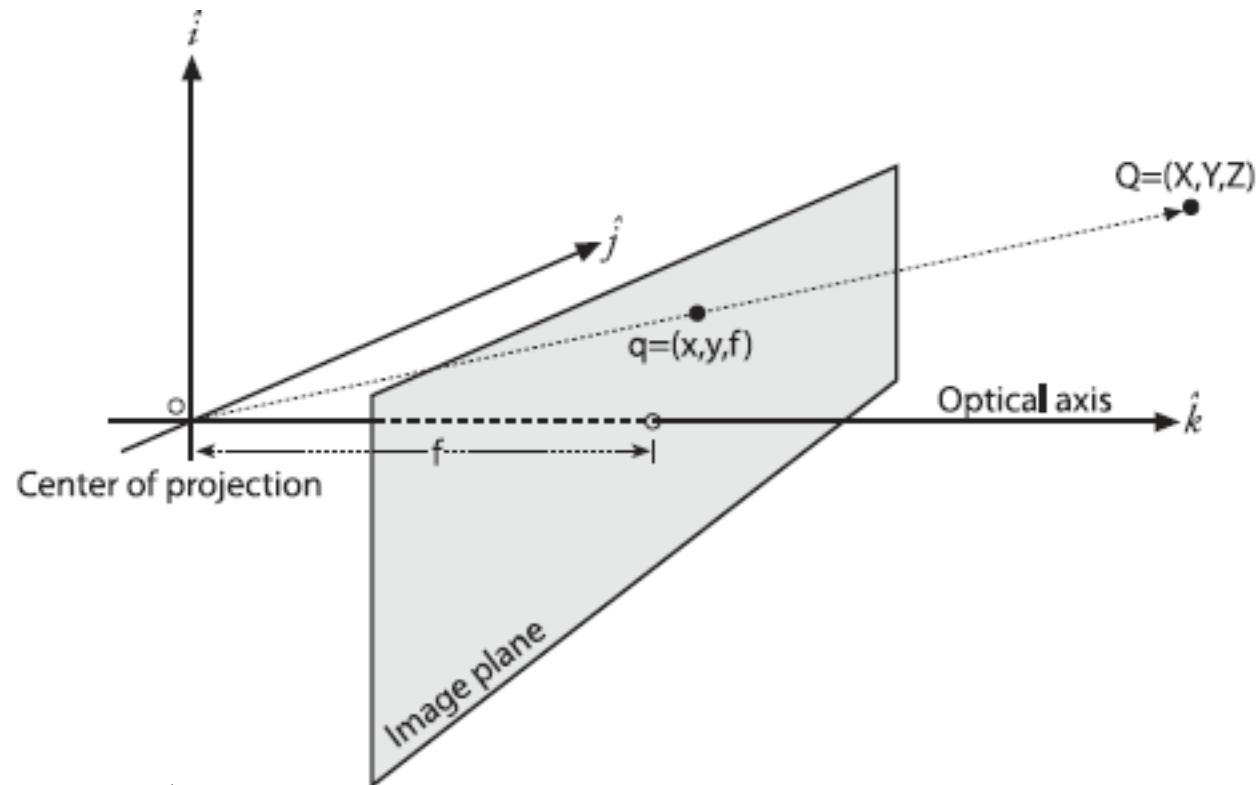
- Double frame-rate displays combined with LCD shutter glasses
- Autostereoscopic displays
 - Parallax barrier and lenticular lens
- Head Mounted Displays (HMDs)
- and “exotic displays”





- Introduction - Stereopsis
- **Camera models and Camera calibration**
- Stereo Vision
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - Stereo Matching

- pinhole camera model



- We get

$$x = f \frac{X}{Z}$$

$$y = f \frac{Y}{Z}$$



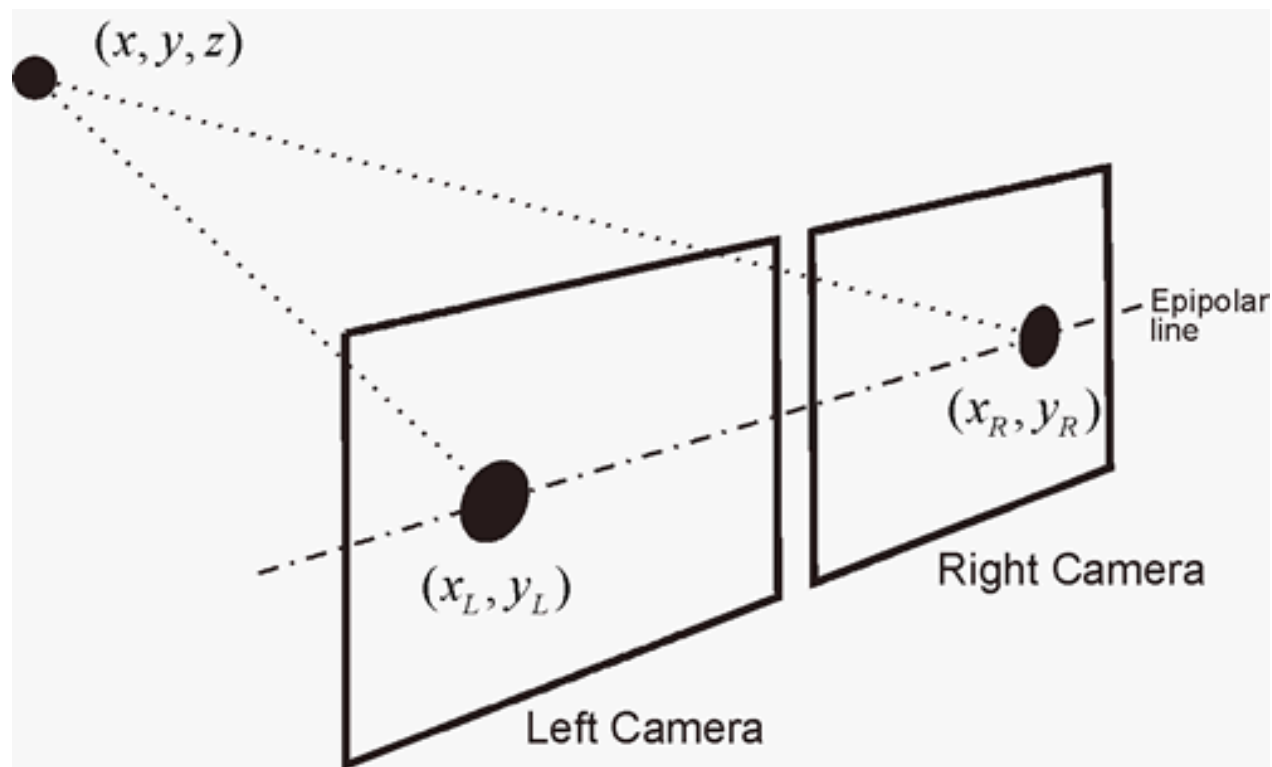
- OpenCV Camera model:
 - 4 intrinsic parameters:
 - Focal distance: f_x, f_y
 - Optical centre: c_x, c_y
 - 5 distortion parameters
 - Lens distortion: k_1, k_2, k_3, p_1, p_2
 - 6 extrinsic parameters:
 - Rotation: r_x, r_y, r_z
 - Translation: t_x, t_y, t_z

Total: **15 parameters**
- Other models: Tsai, Heikkila, Zhang



- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - Stereo Matching

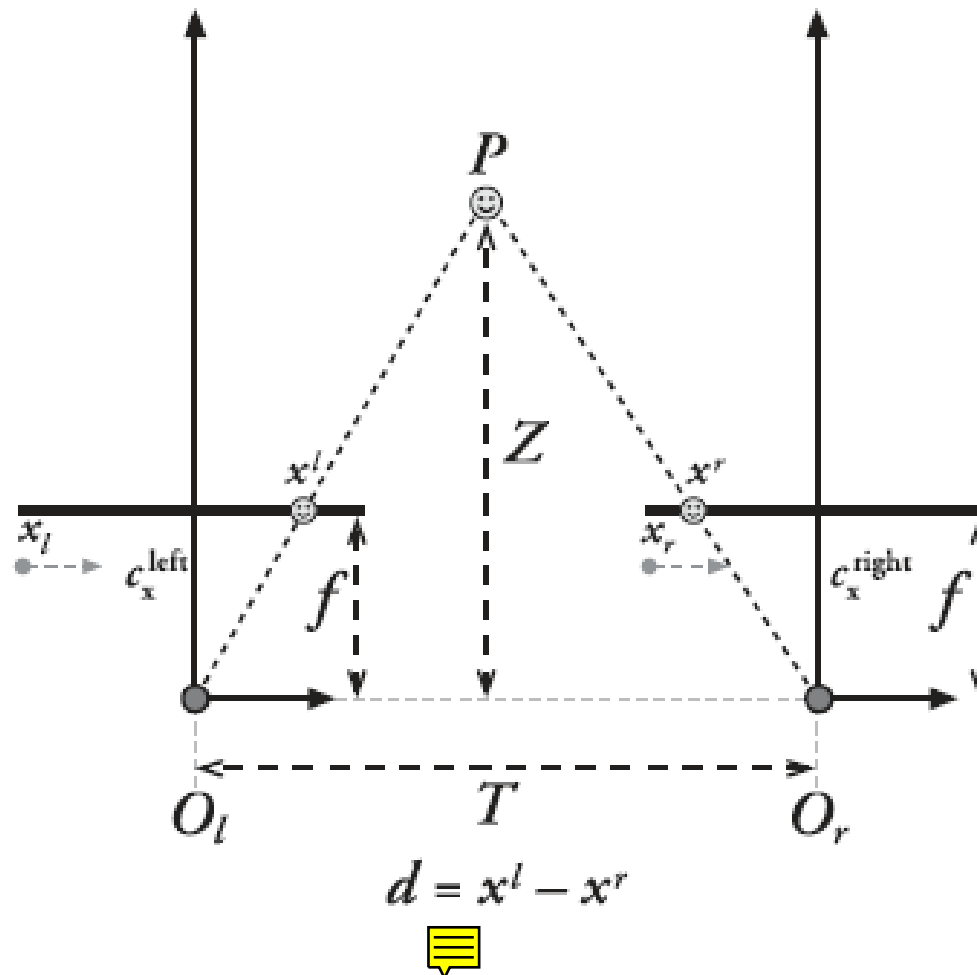
- Capability to define **depth from 2 images**
- Possible by computing **correspondences** between two images



Frontal parallel arrangement

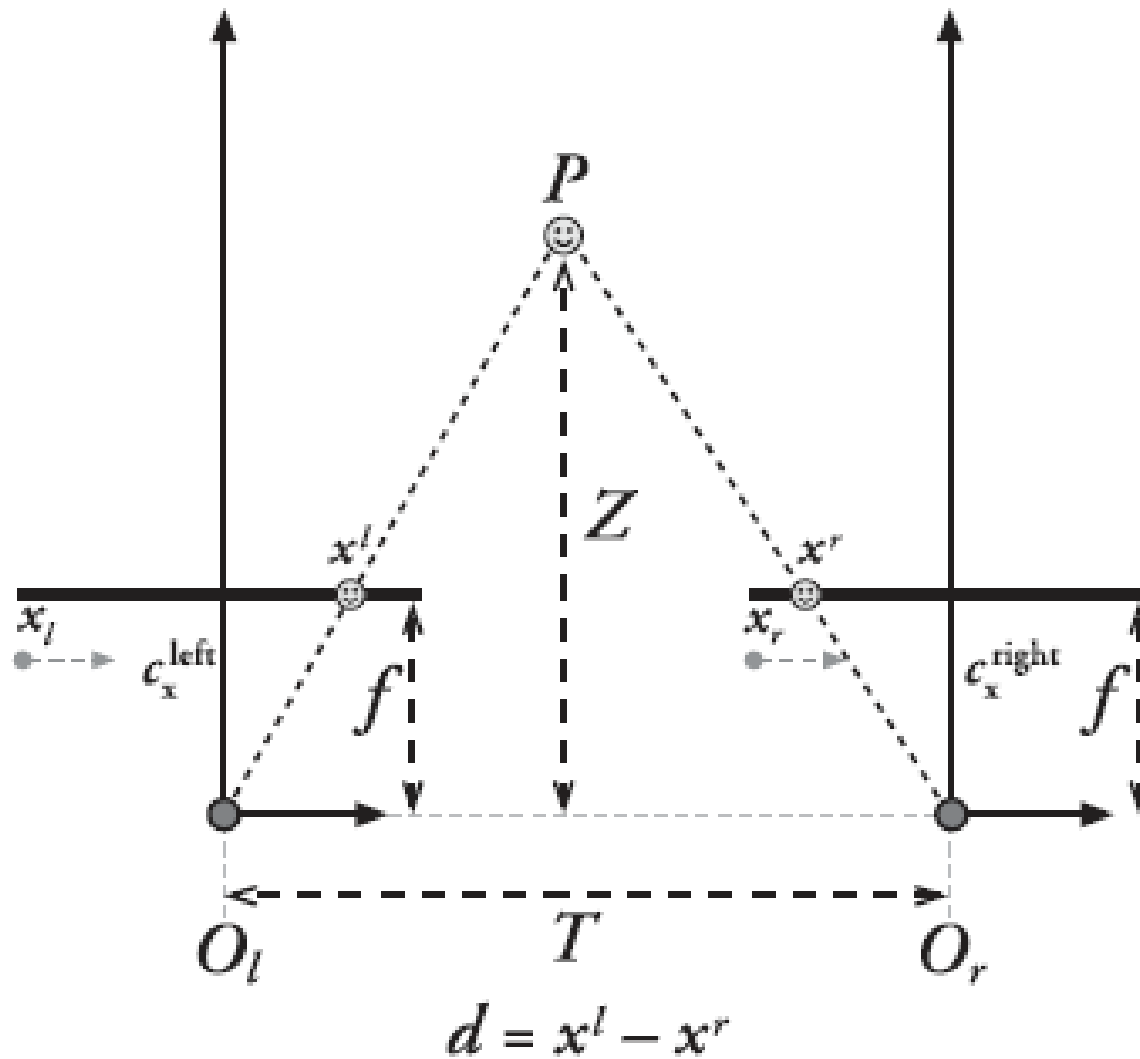


- Two perfectly aligned, coplanar cameras with same focal distance:



d = desparidade - diferença e

Frontal parallel arrangement



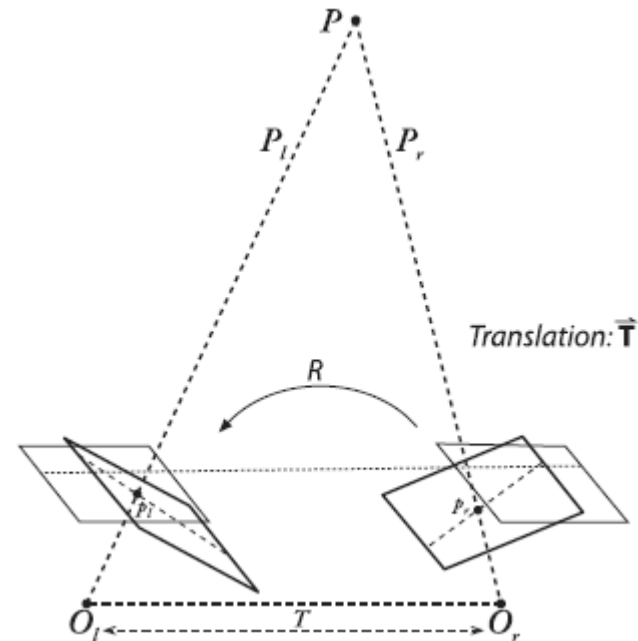
- $Z=?$



- $\frac{T}{Z} = \frac{T-(x^l-x^r)}{(Z-f)}$ then $Z = \frac{T(Z-f)}{T-(x^l-x^r)}$,
- $TZ - Z(x^l - x^r) = TZ - fT$
- So : $Z = \frac{fT}{d}$

=> Stereo system have good depth resolution for **close objects** since depth is inversely proportional to disparity.

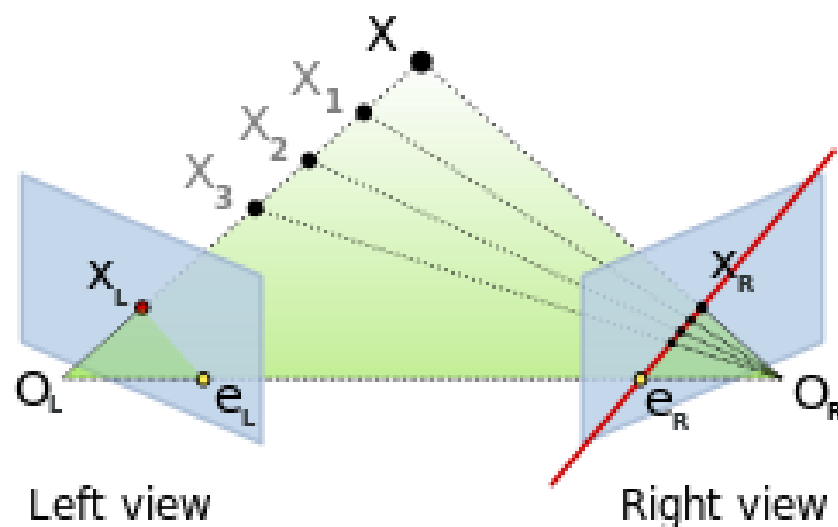
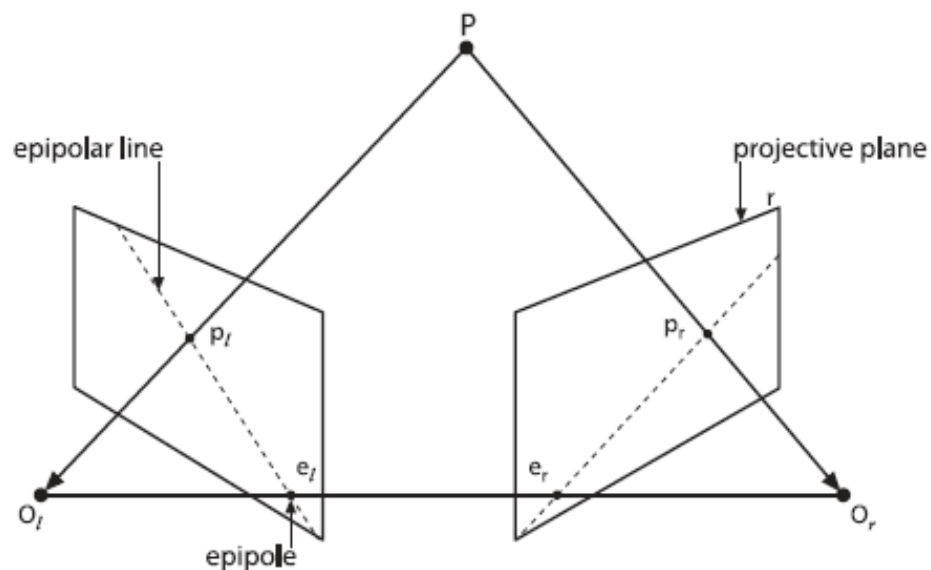
- Easy to relate correspondence to depth in frontal parallel arrangement
- Problem: how to map **real configuration** to frontal **parallel arrangement**.





- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - **Epipolar geometry**
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - Stereo Matching

- An epipole is a **projection of the optical centre** of a camera on the other image plane



- <http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html>



- What is it useful for?:
 - Given a point in an image, its corresponding point in the other image lies on the corresponding epipolar line
 - Order is preserved (given 2 points A e B in a given order in one images, order will be the same in the other image)
- =>Epipolar geometry transform a **2D search** (in image) into a **1D search (along epipolar lines)** saving resources and avoiding errors.



- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - Epipolar geometry
 - **Essential and Fundamental Matrix**
 - Image rectification
 - Template Matching
 - Stereo Matching



- Epipolar Geometry is defined by:
 - Information about relative position between the cameras (rotation and translation) [extrinsic] – **Essential Matrix** (E)
 - Intrinsic parameters of the cameras (focal length, lens distortion, optical centre, etc...) – **Fundamental Matrix** (F)



- Matrix that **maps a 3D point** in one image with its **corresponding 3D point** on the other image considering translation and rotation between cameras:
- $p_l^T E p_r = 0$
- p_l and p_r are in camera 3D coordinate system



- Matrix that **maps a pixel** in one image with its **corresponding pixel** on the other image considering rotation, translation and intrinsic parameters of the cameras:
- $u_l^T F u_r = 0$
- u_l e u_r are in image 2D coordinate system



- Fundamental and Essential matrices represent the **transformation between the stereo pair images**. Fundamental matrix operates in image coordinates (pixels) and Essential matrix operates in physical coordinates.
- Possible to evaluate with 8 point correspondences (eight point algorithm: <http://www.cs.unc.edu/~marc/tutorial/node54.html>)



- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - **Image rectification**
 - Template Matching
 - Stereo Matching

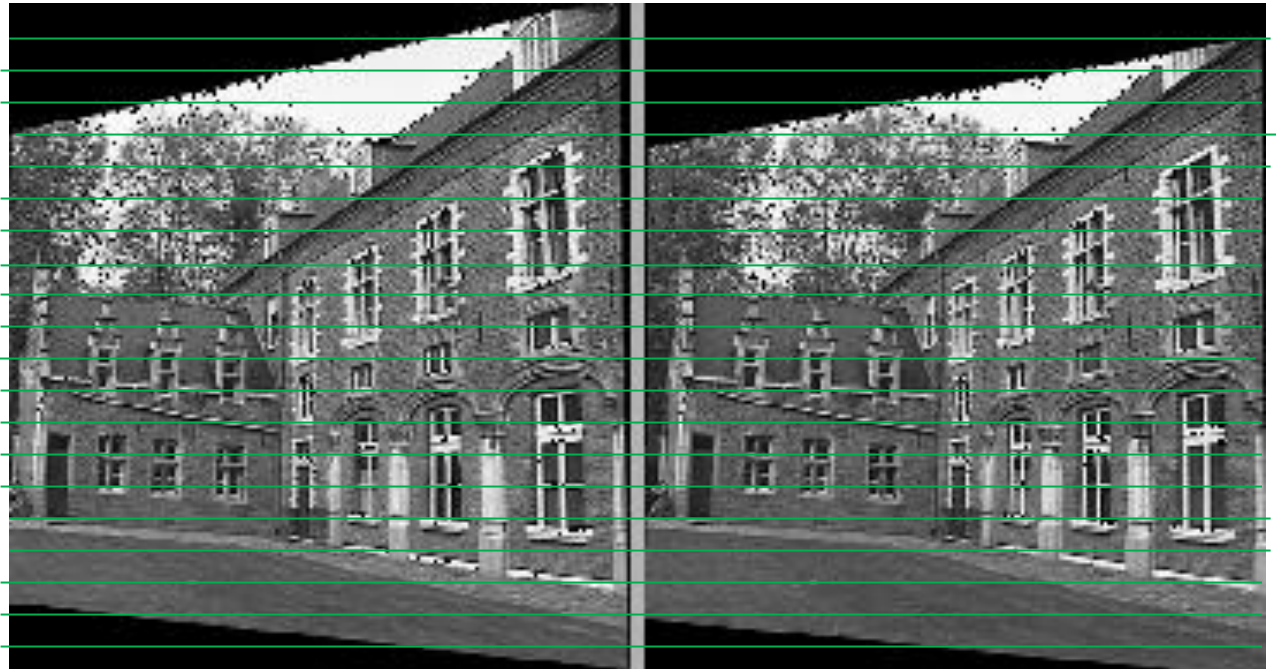


- Given the fundamental matrix, it is possible to **rectify an image by aligning epipolar lines** in rows on the two rectified images getting a **frontal parallel arrangement**.

Original images

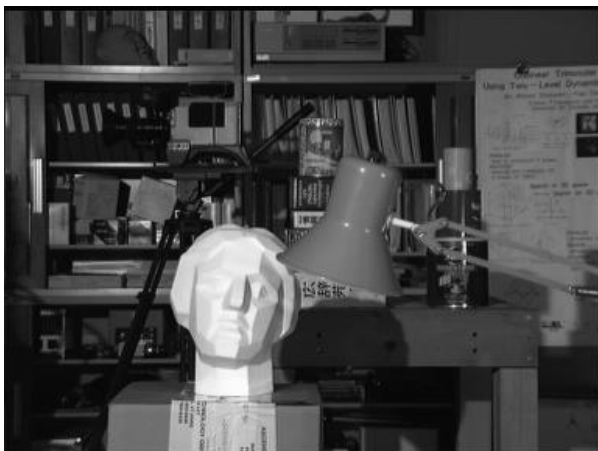


Rectified images



From Visual 3D Modeling from Images (<http://www.cs.unc.edu/~marc/tutorial/>)

- In rectified images, **disparity** is simply **difference** between pixel coordinates x_l and x_r .

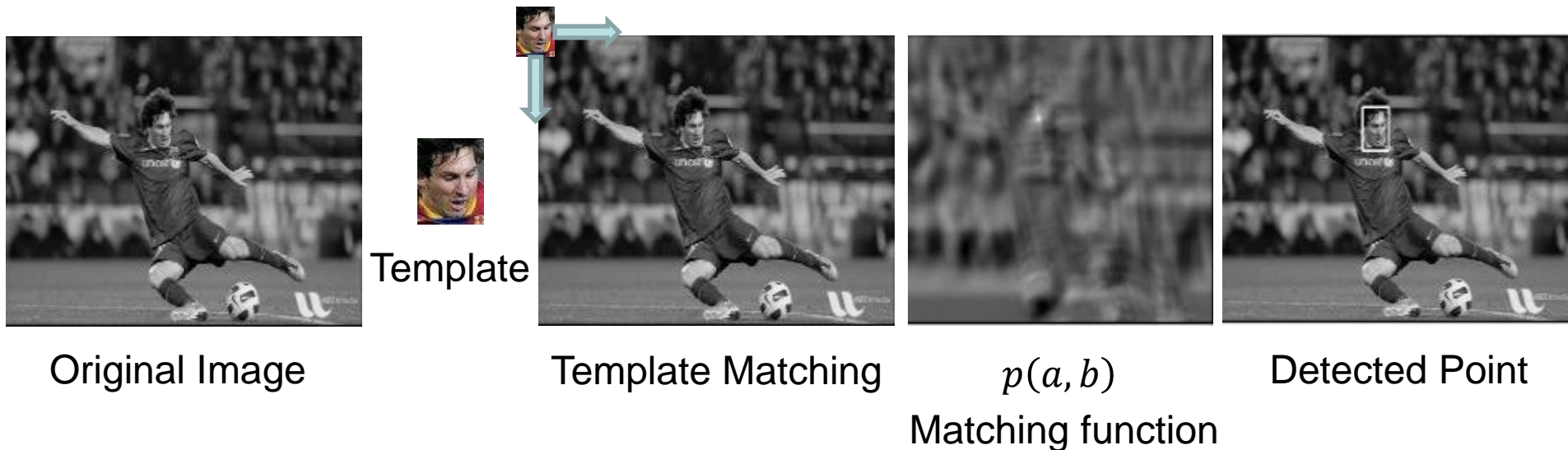


Tsukuba head and lamp stereo dataset



- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - **Template Matching**
 - Stereo Matching

- Template matching is moved to all positions (u, v) in image and computes $p(a, b)$ to evaluate how well template matches image in that position

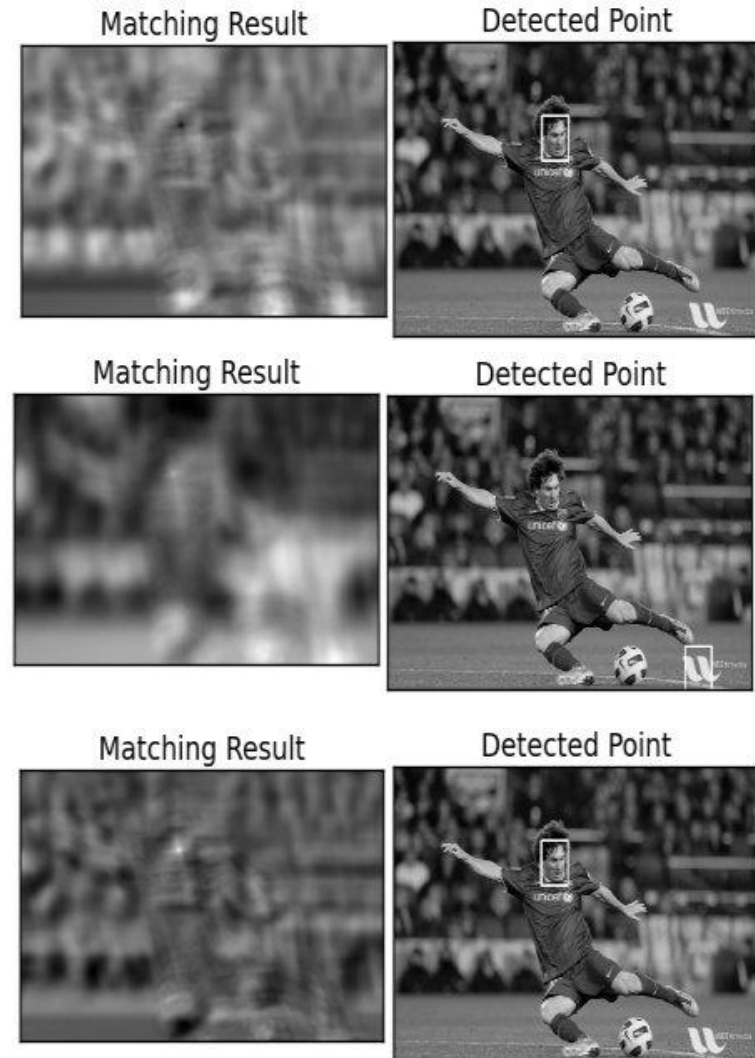


- Object is on image at position where occurs $\max(p(a, b))$ if $\max(p(a, b)) > \text{threshold}$

- Several possible functions to compare template and image $p(a, b)$:

- Square difference matching
- Correlation matching
- Correlation coefficient matching

– ...



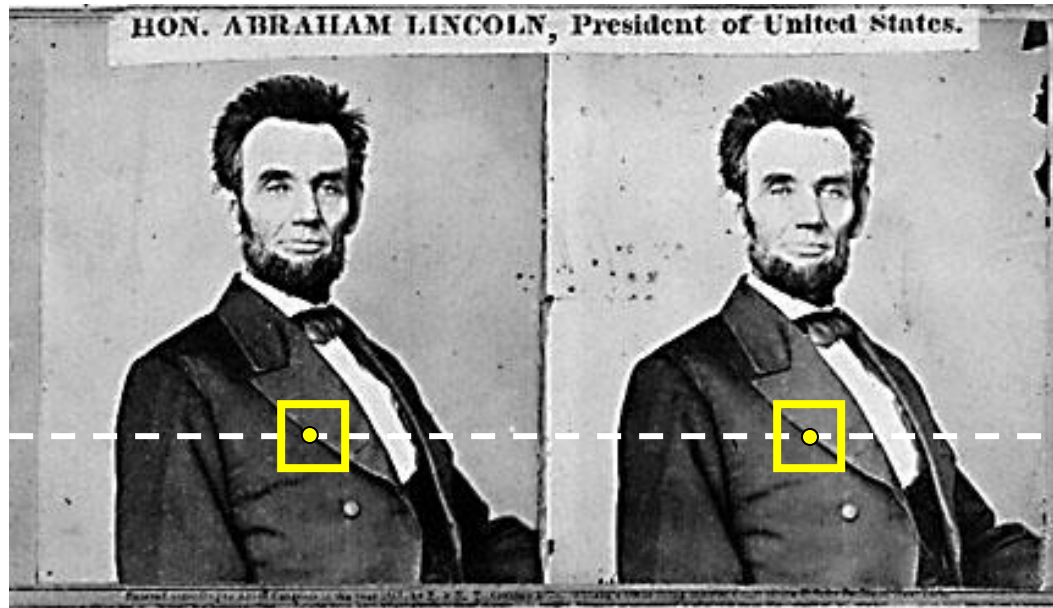


- Introduction - Stereopsis
- Camera models and Camera calibration
- **Stereo Vision**
 - Frontal parallel arrangement
 - Epipolar geometry
 - Essential and Fundamental Matrix
 - Image rectification
 - Template Matching
 - **Stereo Matching**



- Match Pixels in Conjugate Epipolar Lines
 - This is a tough problem
 - Numerous approaches
 - A good survey and evaluation:
<http://vision.middlebury.edu/stereo/>

Basic stereo algorithm

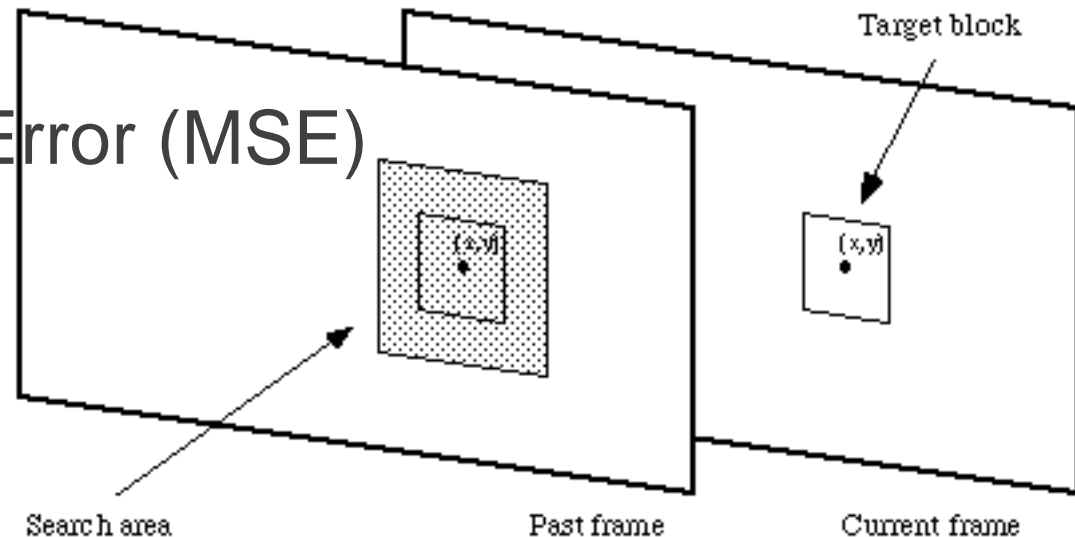


For each epipolar line

For each pixel in the left image

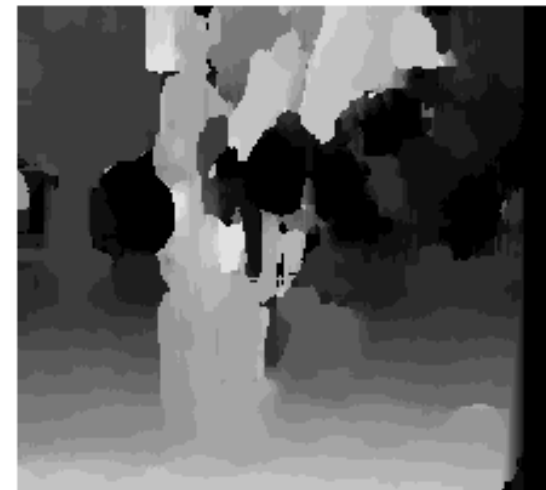
- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

- Block Matching:
 - Divides an image into macroblocks and compare each with a corresponding block and its neighbours in a another image
- Several Metrics
 - Mean difference or Mean Absolute Difference (MAD)
 - Mean Squared Error (MSE)
 - ...





$W = 3$



$W = 20$

size do template

Effect of window size

- Smaller window
- Larger window

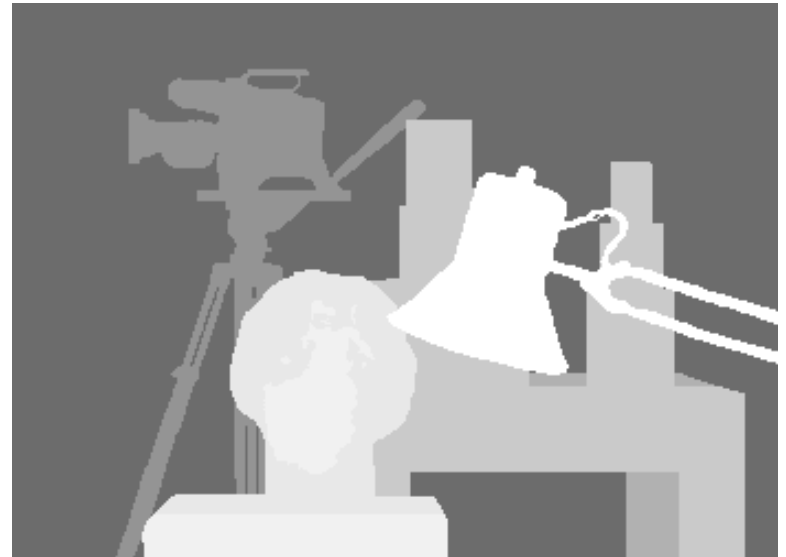
Better results with *adaptive window*

- T. Kanade and M. Okutomi, [A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment](#), Proc. International Conference on Robotics and Automation, 1991.
- D. Scharstein and R. Szeliski. [Stereo matching with nonlinear diffusion](#). International Journal of Computer Vision, 28(2):155-174, July 1998

- Data from University of Tsukuba
- Similar results on other images with ground truth



Scene

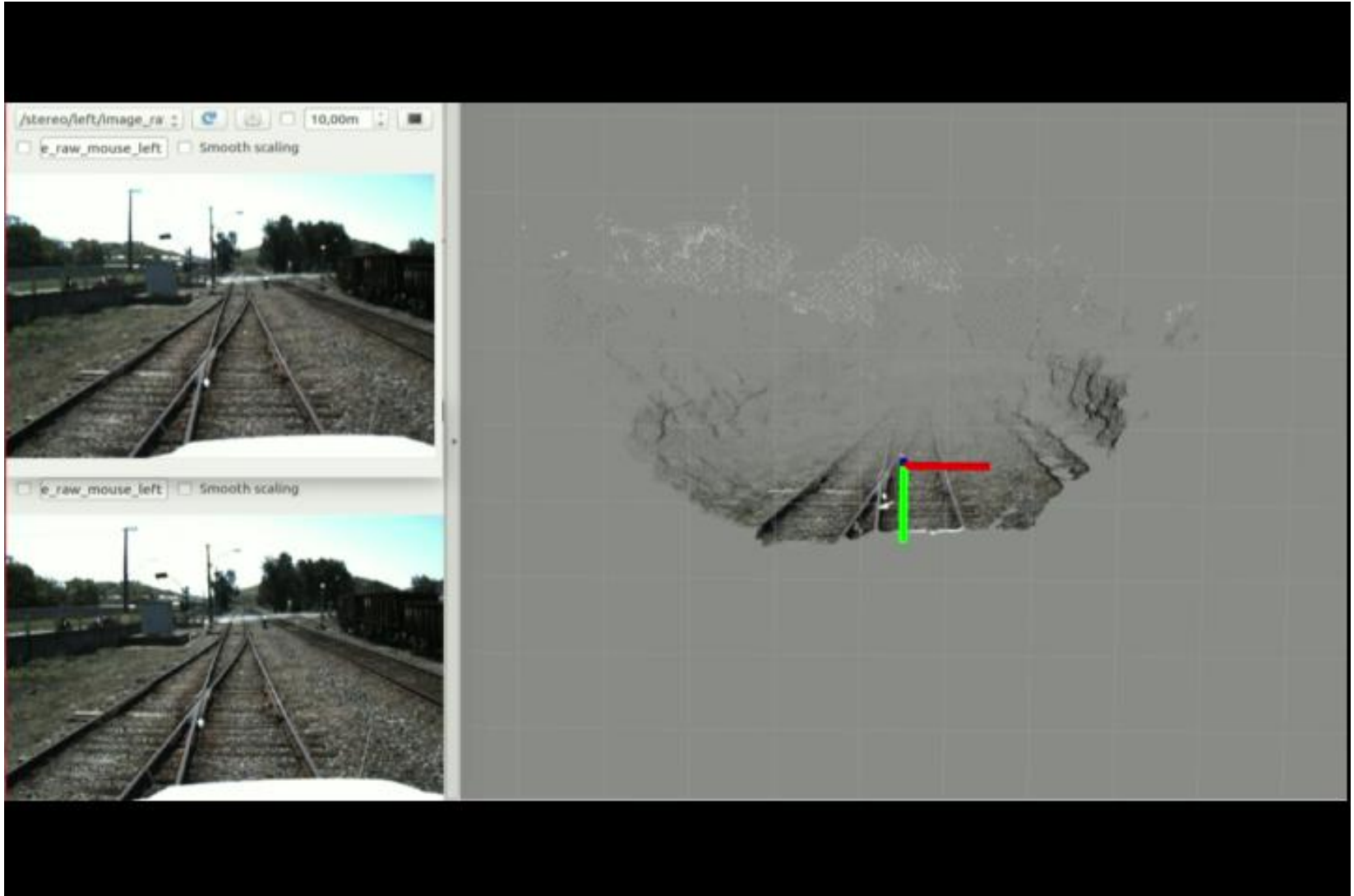


Ground truth

- Sistemas Embarcados de Vistoria (SEV)
 - Stereo vision system with two AVT Mako cameras
 - baseline is around 0.1 meters
 - horizontally aligned.



- Stereo Vision and SLAM





- **Errors** in stereo system comes from
 - Camera calibration errors
 - Poor image resolution
 - Occlusions
 - Violations of brightness constancy (specular reflections)
 - Large motions
 - Low-contrast image regions



- Stereo vision **steps**
 - Calibrate cameras
 - Rectify images
 - Compute disparity
 - Estimate depth



- `cvFindChessboardCorners`: detect chessboard corner in stereo images
- `cvStereoCalibrate`: calibrates stereo rig
- `cvStereoRectify`: computes rotations that make both camera planes the same.
- `cvInitUndistortRectifyMap` and `cvRemap`: use to compute undistortion map and rectified images
- Stereo correspondence (ex: `cvFindStereoCorrespondenceBM`): computes the disparity map.
- `cvReprojectImageTo3D`: disparity map to 3D with calibrated cameras



OPENCV STEREO VISION DEMO



- Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA, 2008.
- Olivier Faugeras Three-dimensional computer vision: a geometric viewpoint. MIT Press Cambridge, MA, USA ©1993
- Szeliski, R. (2010).. *Computer Vision: Algorithms and Applications*, Springer
- Quang-Tuan Luong. "[Learning Epipolar Geometry](#)". Retrieved 2007-03-04.