Stereo and Multiview imaging summary

S. Papadopoulos, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@aiia.csd.auth.gr
www.aiia.csd.auth.gr

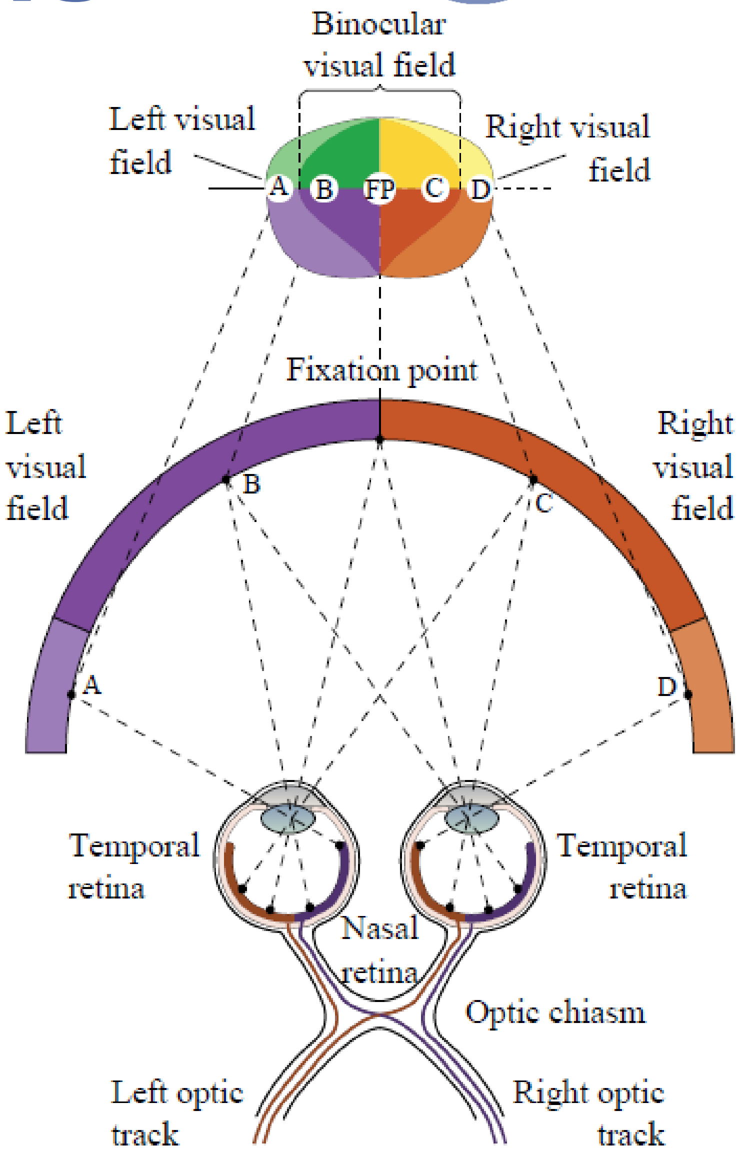
Version 3.2

Artificial Intelligence &
Information Analysis Lab

Introduction to Stereopsis

- The horizontal separation of the eyes leads to a difference, ***stereo parallax***, in image location and appearance of an object between the two eyes, called ***stereo disparity***.

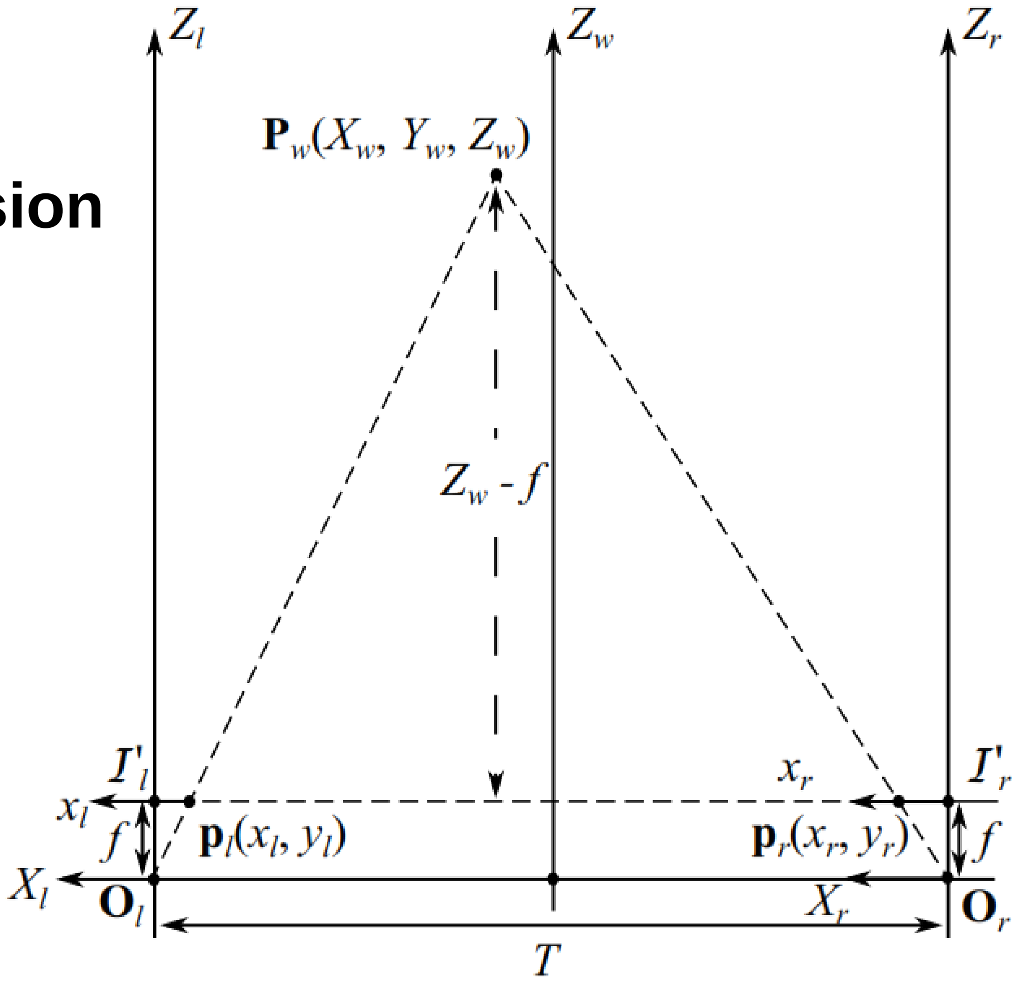
- Stereo parallax is utilized by the brain in order to extract depth information.



Basics of Stereopsis

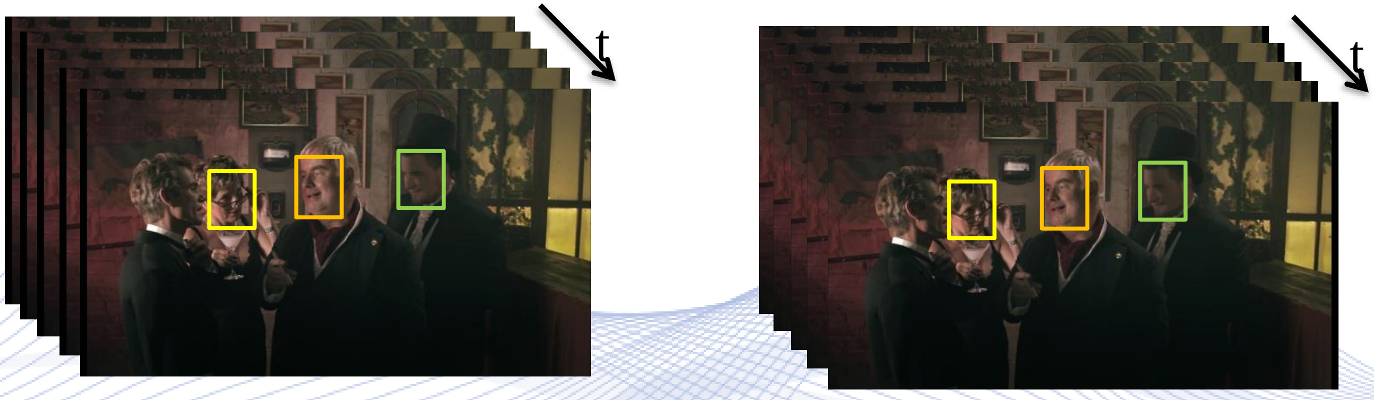
Parallel Stereo vision Geometry

T: baseline
f: focal length



Basics of Stereopsis

- Stereo images and videos:
- Left and right image/video channels

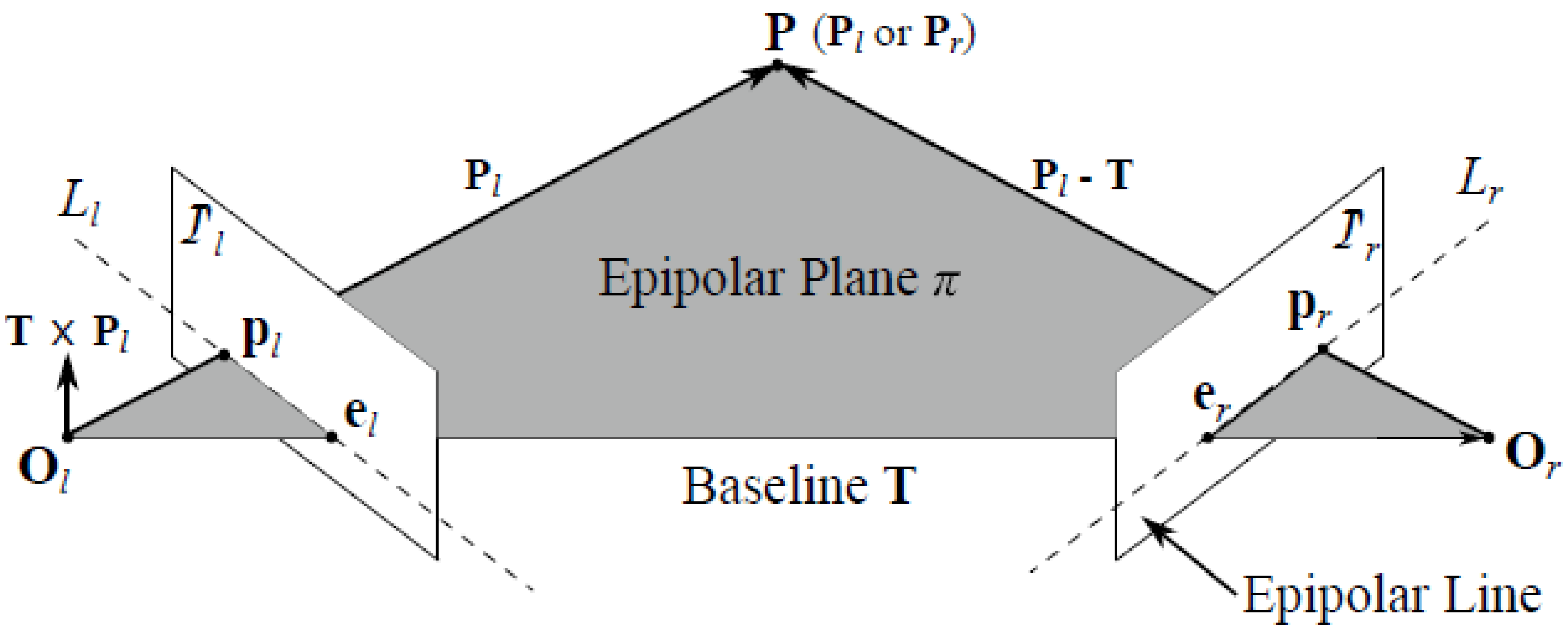


Left and right video channels.

Basics of Stereopsis

A dense disparity map can be estimated from detecting pixel correspondences in left and right images.

Epipolar Geometry



- Epipoles $\mathbf{e}_l$, $\mathbf{e}_r$: intersection points between camera centers projections and image planes.
- Epipolar plane π: 3D plane containing line $\mathbf{T}$ and point $\mathbf{P}$.
- Epipolar lines L_l, L_r: intersection between π and each image plane.

The Essential Matrix $\mathbf{E}$

- The ***Essential Matrix*** $\mathbf{E}$ compactly encodes the epipolar constraint:

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

  where:

$$\mathbf{E} = \mathbf{R}\mathbf{T}_\times = \begin{bmatrix} T_z r_{12} - T_y r_{13} & -T_z r_{11} + T_x r_{13} & T_y r_{11} - T_x r_{12} \\ T_z r_{22} - T_y r_{23} & -T_z r_{21} + T_x r_{23} & T_y r_{21} - T_x r_{22} \\ T_z r_{32} - T_y r_{33} & -T_z r_{31} + T_x r_{33} & T_y r_{31} - T_x r_{32} \end{bmatrix}.$$

- $\mathbf{E}$ is a 3×3 rank-deficient matrix. It is is completely determined by the rotation and translation between the two cameras/views.
  - If the WCS coincides with the coordinate system of the left or right camera, $\mathbf{E}$ encodes extrinsic camera parameters (incl. baseline $\mathbf{T}$).

The Fundamental Matrix $\mathbf{F}$

- The ***Fundamental Matrix*** $\mathbf{F}$ also encodes the epipolar constraint:

$$\mathbf{p}_{dr}^T \mathbf{F} \mathbf{p}_{dl} = 0.$$

- The fundamental matrix $\mathbf{F}$ is related to the essential matrix:

$$\mathbf{F} = (\mathbf{P}_{Ir}^{-1})^T \mathbf{E} \mathbf{P}_{Il}^{-1} = (\mathbf{P}_{Ir}^{-1})^T \mathbf{R} \mathbf{T}_{\times} \mathbf{P}_{Il}^{-1}.$$

- $\mathbf{F}$ is a 3×3 rank-deficient matrix.
- It is defined in *pixel* coordinates, while $\mathbf{E}$ was defined in camera plane or normalized virtual image plane coordinates.

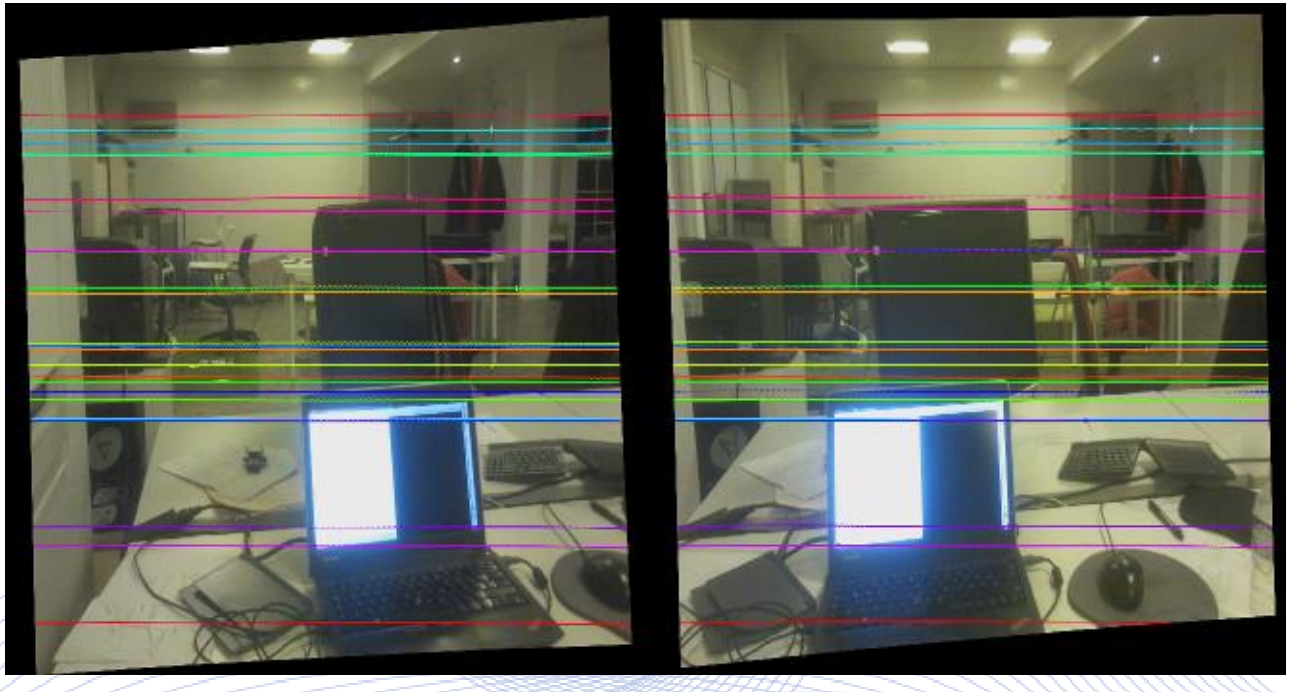
Eight-point Algorithm

- $\mathbf{F}$ can be estimated by employing $K > 7$ left-right pixel correspondences and the fundamental matrix constraint:

$$\mathbf{p}_{dr}^T \mathbf{F} \mathbf{p}_{dl} = 0.$$

- We formulate a homogeneous system $\mathbf{Xu} = 0$, where $\mathbf{X}$ is as $K \times 9$ matrix and $\mathbf{u}$ contains the 9 entries of matrix $\mathbf{F}$.

Rectification

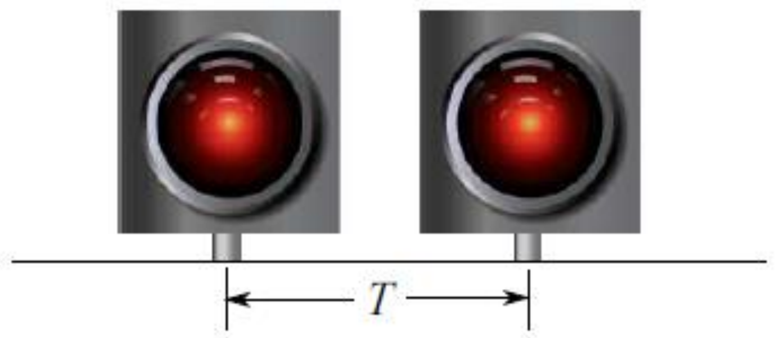


- Rectification simplifies the search for pixel correspondences between views:
  - Search on epipolar lines becomes a search along a horizontal scan line, at the same height as the reference pixel.

Stereo Camera Technologies

- The parallel, *side-by-side stereo rig* design tries to imitate the way eyes are positioned on the human face

- The cameras can:
  - perform horizontal shifts, thus changing their inter-axial (baseline) distance T,
  - converge and diverge,
  - change zoom and focus.



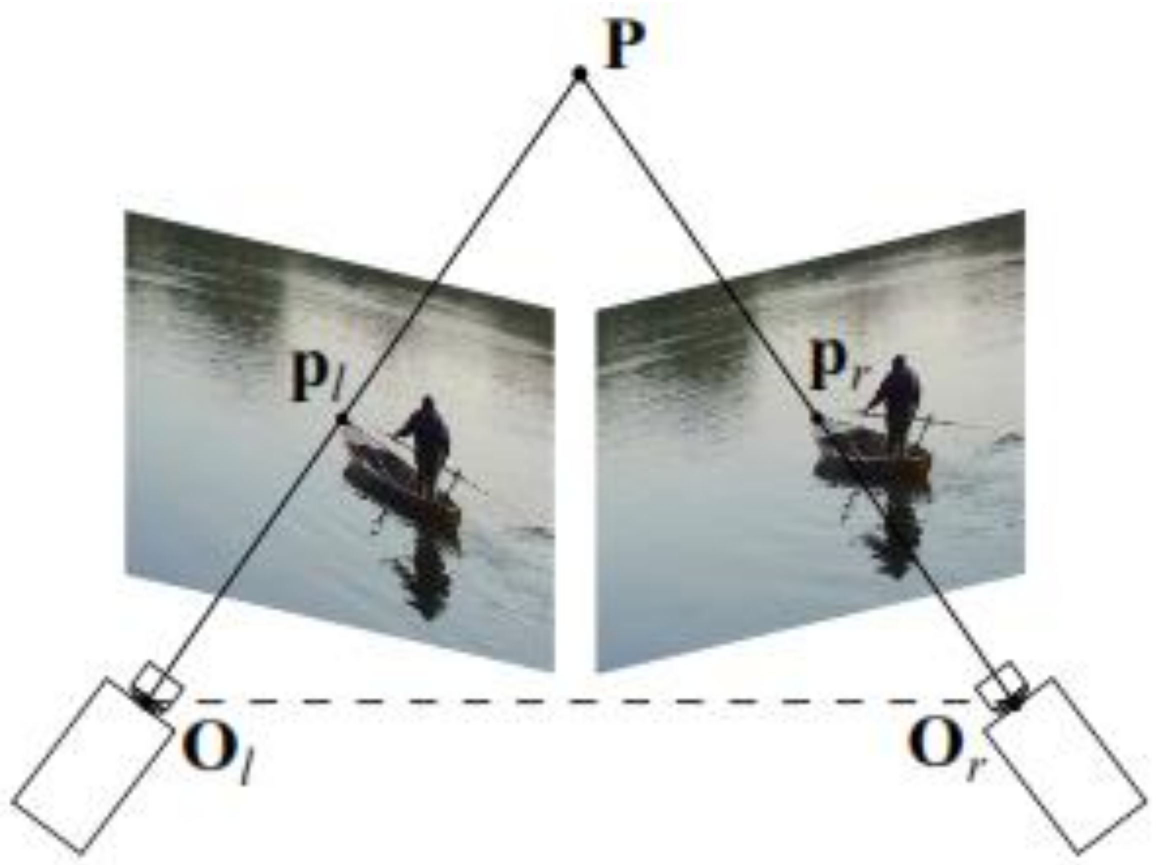
Feature Extraction

- *Feature detectors:*
  - SIFT, AGAST, SURF, Hessian Affine, CeNSuRe, BRISK, ORB, AKAZE, or simply dense sampling.

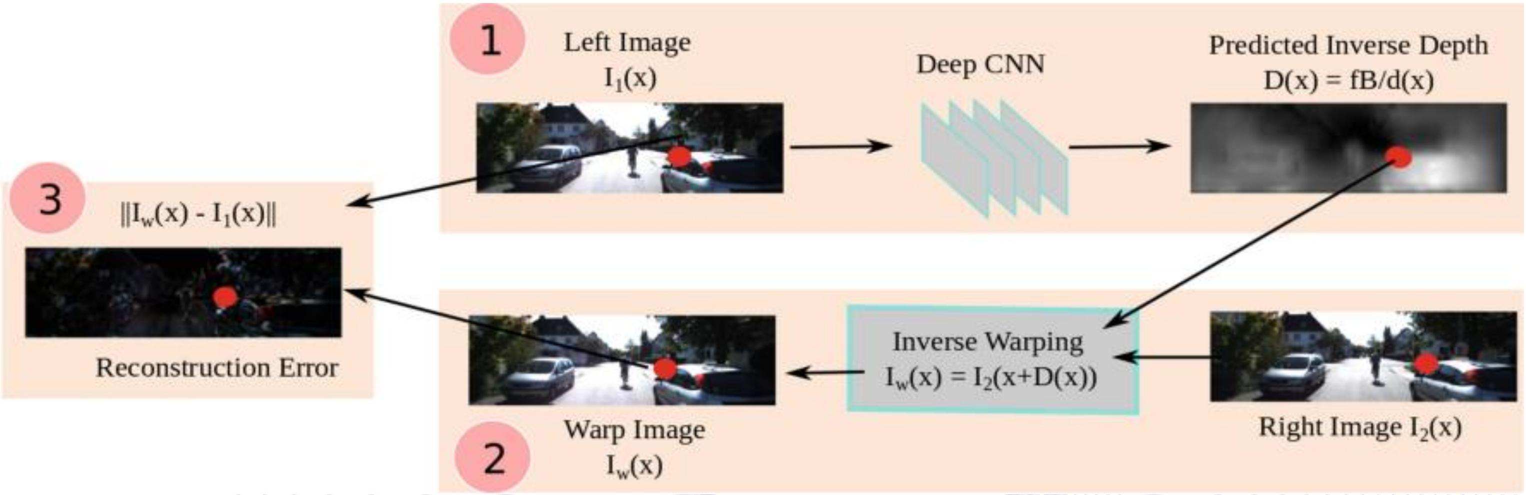
- *Feature descriptors:*
  - SIFT, SURF, DAISY, HOG, LIOP, LUCID, BRIEF, BRISK, FREAK, ORB, AKAZE, LATCH, CENTRIST, BinBoost, LMoD.

- Convolutional neural network features.

Feature Correspondence



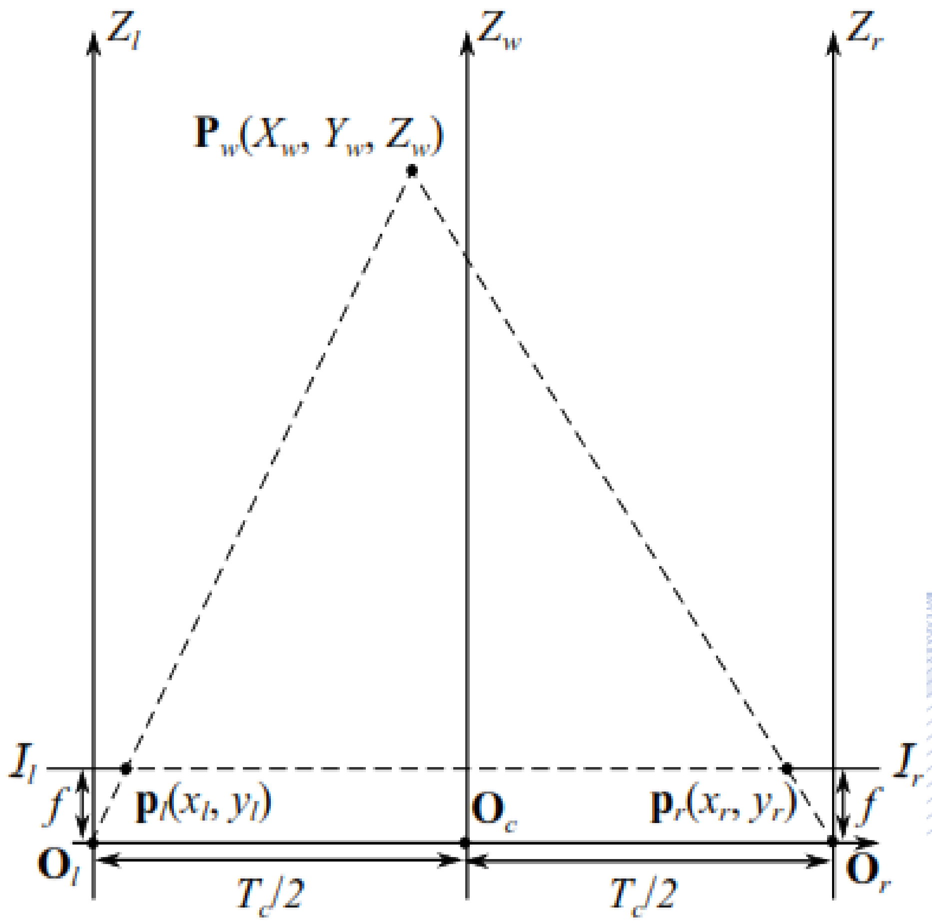
Disparity Estimation with NNs



[GAR2016]

Unsupervised NN depth estimation from stereo image pairs.

Parallel and Converging Camera Setups

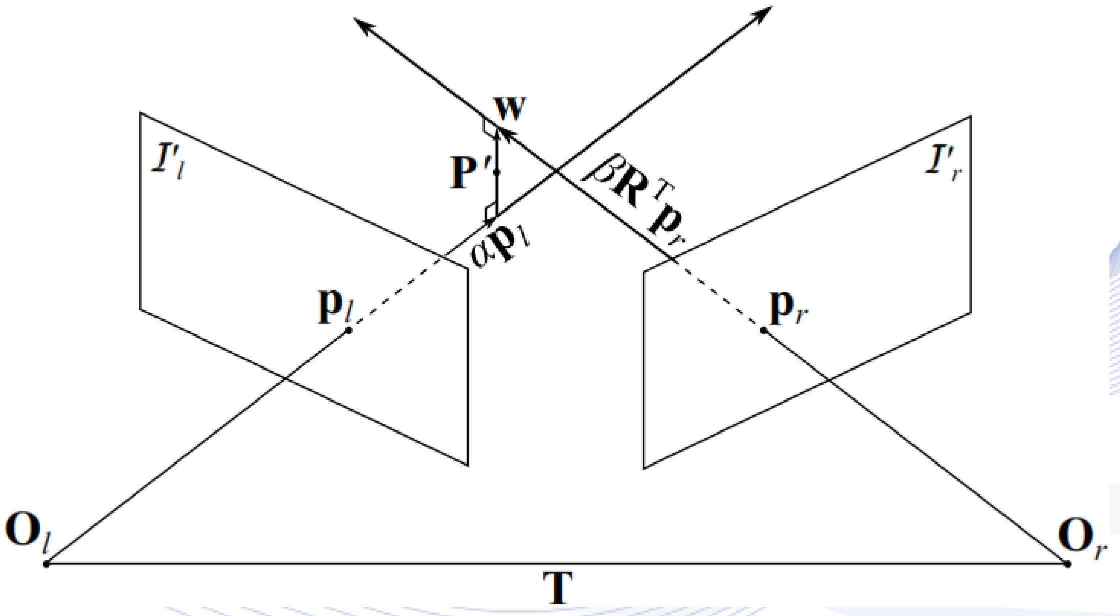


- Transformation from left/right camera coordinates to world coordinates in parallel stereo-rig setup by translation by $T_c/2$.

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} X_l - \frac{T_c}{2} \\ Y_l \\ Z_l \end{bmatrix} = \begin{bmatrix} X_r + \frac{T_c}{2} \\ Y_r \\ Z_r \end{bmatrix}.$$

General 3D reconstruction in a calibrated stereo camera system

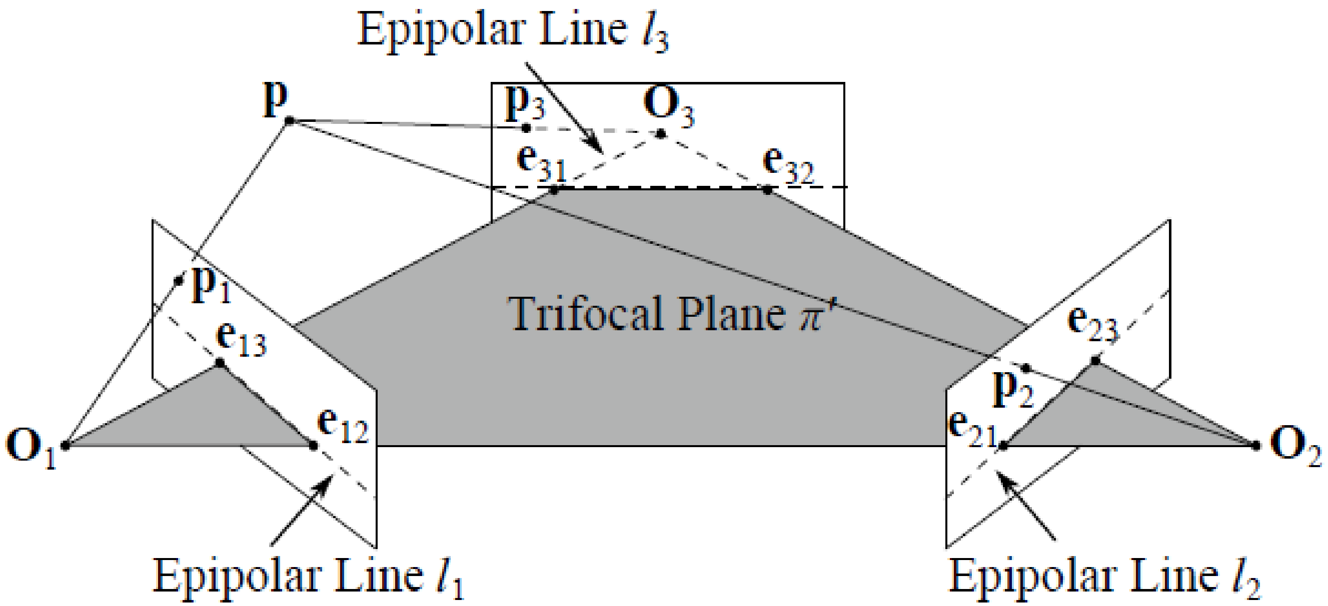
- Due to noise in camera calibration, triangulation refinement may be needed, so that the rays emanating from the optical centers of the cameras and passing through its left and right projections intersect on (or close to) **P**.



Monocular NN Depth Estimation

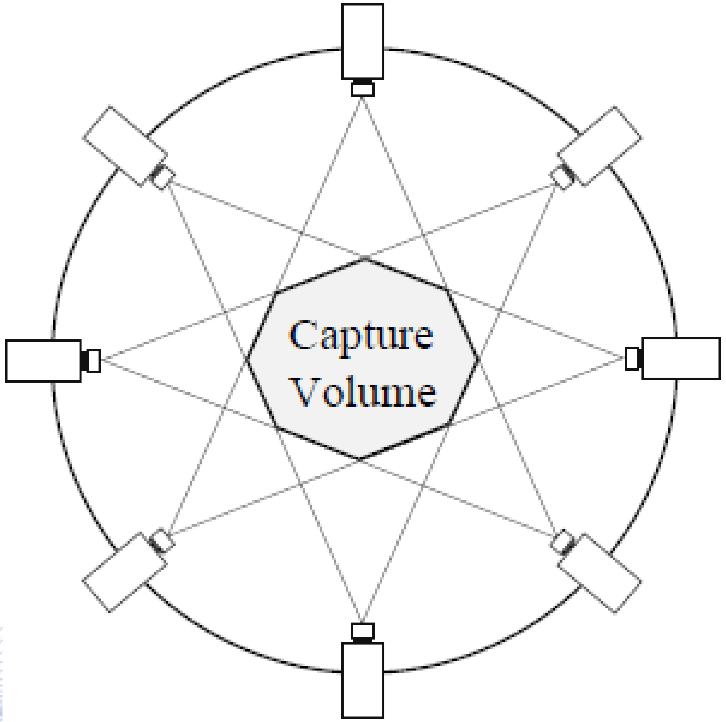
Depth image from monocular video [APOLLO].

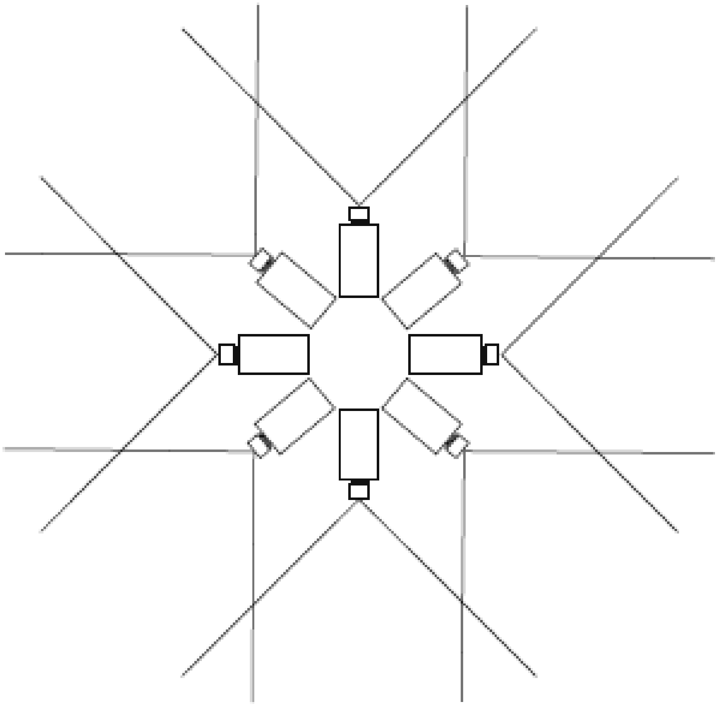
Three-Views and the Point Transfer



- Given two image points $\mathbf{p}_1, \mathbf{p}_2$ on the first and second image plane, respectively, the exact position of the corresponding point $\mathbf{p}_3$ on the third image plane can be completely specified in terms of $\mathbf{p}_1, \mathbf{p}_2$.

Multiple Camera Image Acquisition





Circular camera positioning setups.

3D monument modelling:
Vlatadon Monastery, Thessaloniki

3D monument modelling:
Vlatadon Monastery, Thessaloniki

Artificial Intelligence &
Information Analysis Lab

Bibliography

[PIT2019] I. Pitas, "Computer vision", Createspace/Amazon, in press.
[SZE2011] R.Szelinski, " Computer Vision " , Springer 2011
[HAR2003] Hartley R, Zisserman A. , " Multiple view geometry in computer vision" . Cambridge university press; 2003.
[DAV2017] Davies, E. Roy. "Computer vision: principles, algorithms, applications, learning ". Academic Press, 2017
[TRU1998] Trucco E, Verri A. "Introductory techniques for 3-D computer vision", Prentice Hall, 1998.
[PIT2017] I. Pitas, "Digital video processing and analysis " , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, "Digital Video and Television " , Createspace/Amazon, 2013.
[PIT2000] I. Pitas, Digital Image Processing Algorithms and Applications, J. Wiley, 2000.
[NIK2000] N. Nikolaidis and I. Pitas, 3D Image Processing Algorithms, J. Wiley, 2000.
[ZBO2015] Zbontar, Jure, and Yann LeCun. "Computing the stereo matching cost with a convolutional neural network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

Bibliography

[MAY2016] Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[DOS2015] Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2015.

[GAR2016] Garg, Ravi, et al. "Unsupervised cnn for single view depth estimation: Geometry to the rescue." *European Conference on Computer Vision*. Springer, Cham, 2016.

[ZHO2017] Zhou, Tinghui, et al. "Unsupervised learning of depth and ego-motion from video." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[VIJ2017] Vijayanarasimhan, Sudheendra, et al. "Sfm-net: Learning of structure and motion from video." *arXiv preprint arXiv:1704.07804* (2017).

[YIN2018] Yin, Zhichao, and Jianping Shi. "Geonet: Unsupervised learning of dense depth, optical flow and camera pose." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[BIA2019] Bian, Jiawang, et al. "Unsupervised scale-consistent depth and ego-motion learning from monocular video." *Advances in Neural Information Processing Systems*. 2019.

[GOR2019] Gordon, Ariel, et al. "Depth from videos in the wild: Unsupervised monocular depth learning from unknown cameras." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

[CHE2019] Chen, Yuhua, Cordelia Schmid, and Cristian Sminchisescu. "Self-supervised learning with geometric constraints in monocular video: Connecting flow, depth, and camera." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

[APOLLO] http://apolloscape.auto/

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr