Stereo and Multiview imaging

Contributor: Prof. Ioannis Pitas

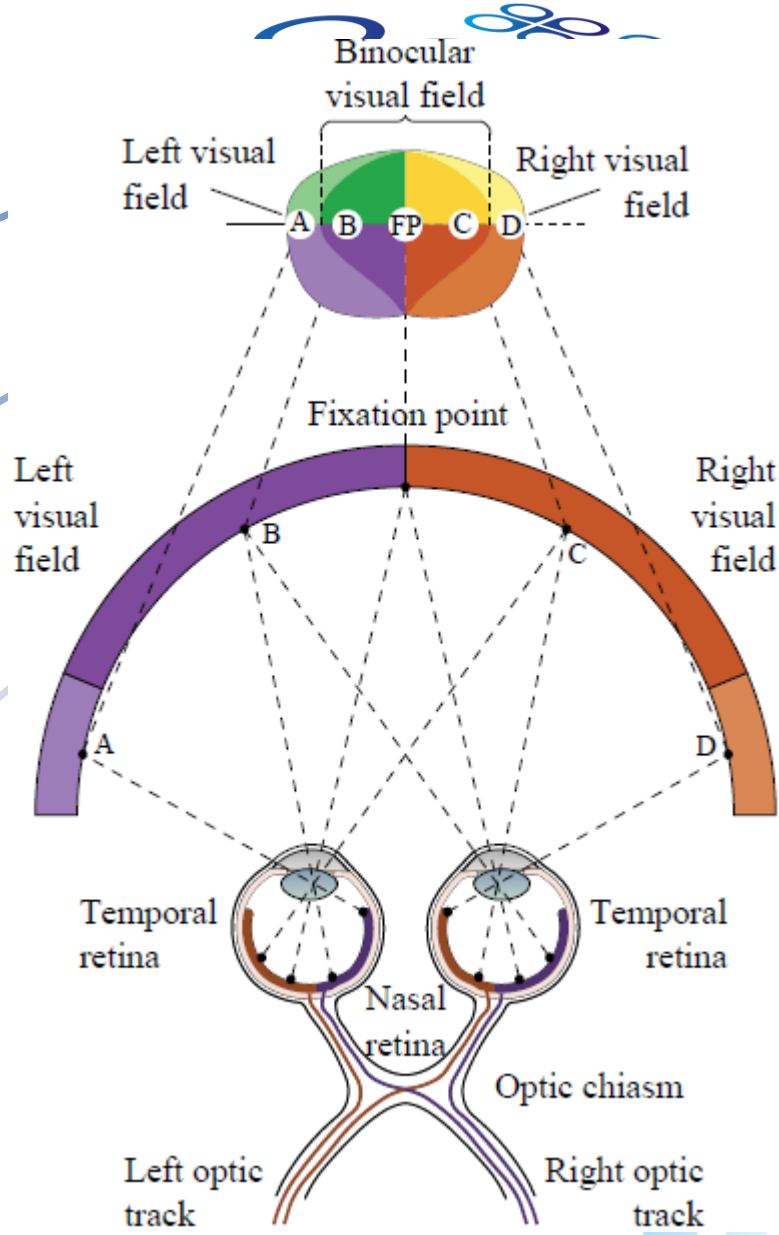
Presenter: Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@aiia.csd.auth.gr
www.multidrone.eu
Presentation version 1.2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- The two monocular views are automatically combined into a single a single subjective view, called *cyclopean view.*

- With stereo vision, the viewer can see where objects are in relation to him/herself, especially when these objects are moving towards or away from him/her in the depth dimension (*binocular visual field*).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- Large binocular disparities, large depth motion and/or frequent changes of the motion direction in depth, induce *visual discomfort*.
- *Binocular rivalry* occurs *w*hen discrepant monocular images are presented to the two eyes and rival for *perceptual dominance*, such that only one monocular image is perceived at a time, while the other is suppressed from awareness.
- Binocular rivalry effects are very disturbing, when the dominant image alternates from one eye to the other.

This project has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- Binocular rivalry can be caused by differences in:
  - size,
  - display scene representation complexity and
  - brightness.
- Due to binocular rivalry, we perceive high stereo image quality, even if only one of the stereo images (typically the right one) is of high quality, while the other one is of low quality,
- This property is extensively used in asymmetric stereo video coding.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- Oculomotor depth cues:
  - *Accommodation*: has to do with the change in ocular focus and can be either a reflex or a consciously controlled action. The ability of accommodation declines with age.
  - *Vergence*: the change in ocular alignment caused by the simultaneous movement of both eyes in opposite direction, in order to maintain single binocular vision.
  - *Myosis*: the eye pupil constriction to less than or equal to two millimeters It is a normal response to an increase in illumination.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- *Eye convergence*: the rotation of the eyes towards each other when we try to look closer at an object (natural movement).

- *Eye divergence*: the opposite phenomenon (rather unnatural movement).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Introduction to Stereopsis

- Accommodation and vergence are combined when we try to focus on an object at a distance, offering oculomotor depth cues.

- *Accommodation and vergence conflict.* When using stereo displays, our eyes focus on the display screen, while object disparity dictates them to converge before/after the screen.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Basics of Stereopsis

- Scene depth can by inferred by simultaneous acquisition of two scene views, from slightly different world positions.
  - Use of disparity maps for depth estimation.

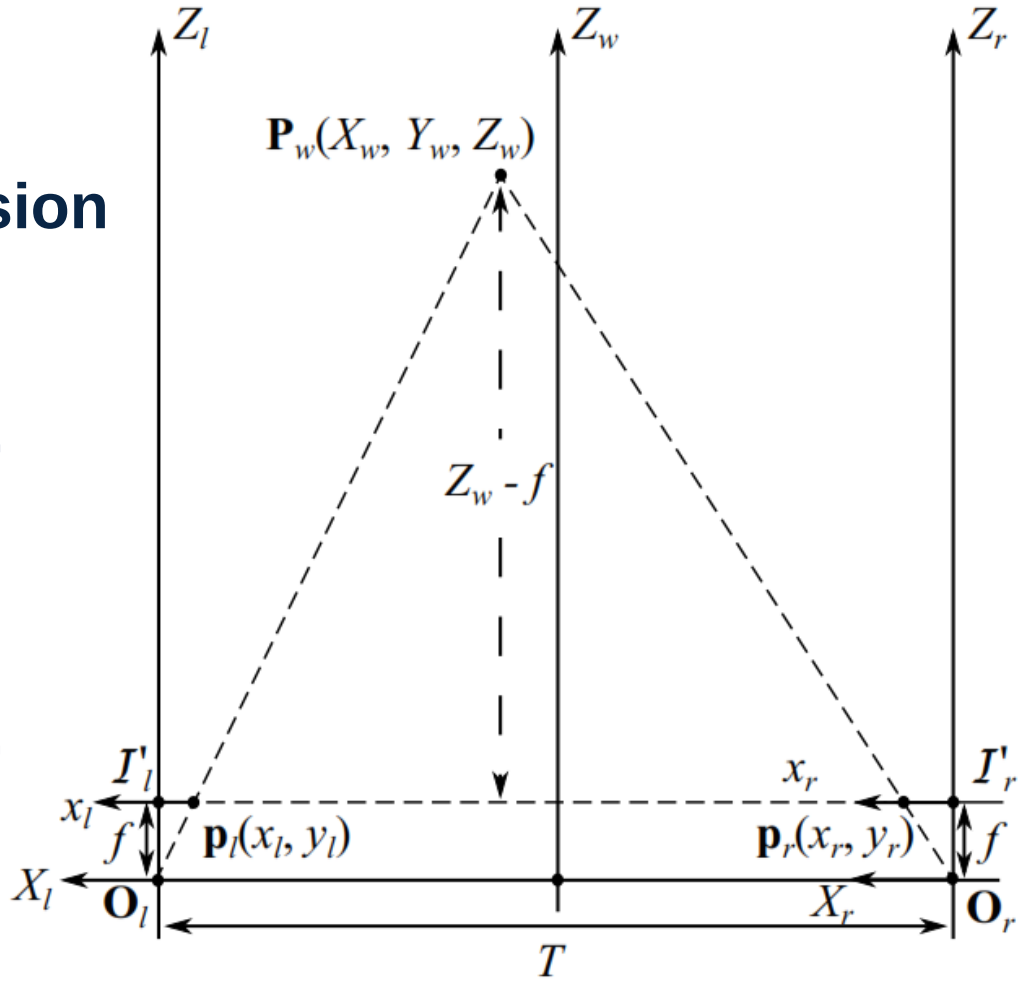
- Stereo camera rig:
  - Parallel (two cameras with parallel optical axes).
  - Converging (two cameras with converging optical axes).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Basics of Stereopsis

Parallel Stereo vision Geometry

T: baseline
f: focal length



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Basics of Stereopsis

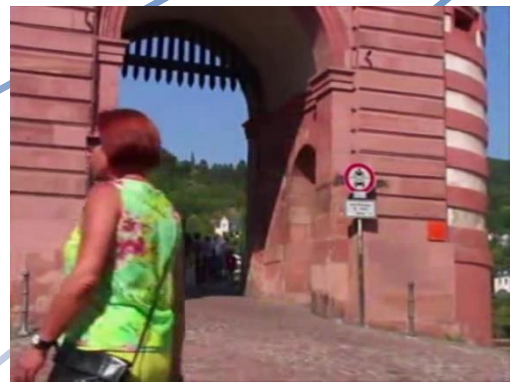
- In a parallel stereo rig, vertical disparity/parallax is zero.
- Horizontal disparity: $d = x_r - xl \leq 0$.
- d is inversely proportional to scene depth Z_w (by triangle similarity):

$$d = -f\frac{T}{Z_w}$$

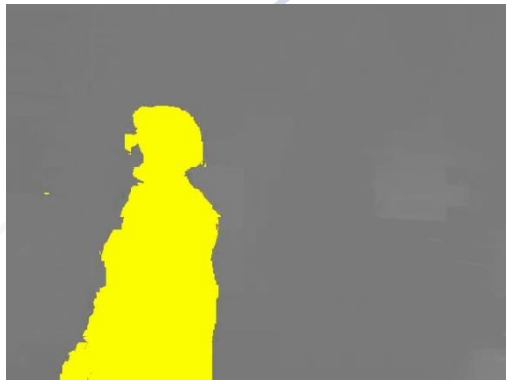
- Thus, $|d|$ decreases as the imaged object distance from the camera increases.
- d is zero for visible scene points at infinity.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Basics of Stereopsis



- A dense disparity map can be estimated from detecting pixel correspondences.

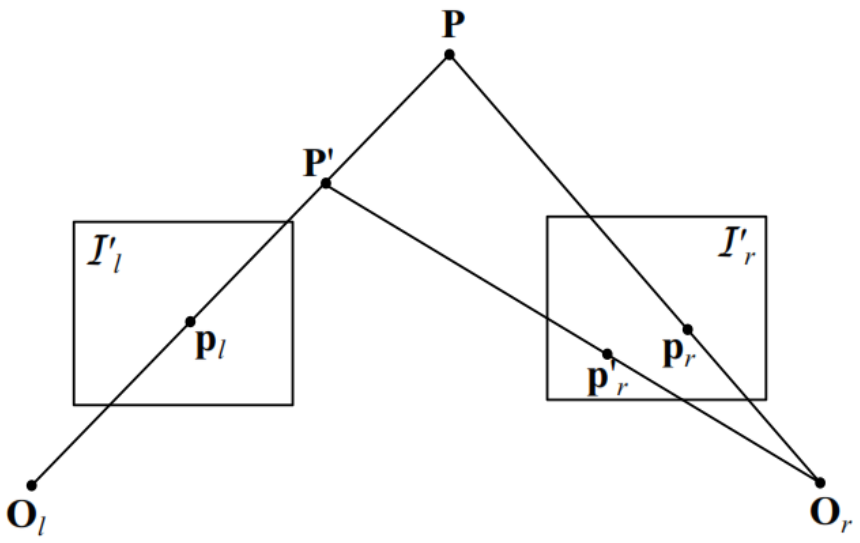


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Epipolar Geometry

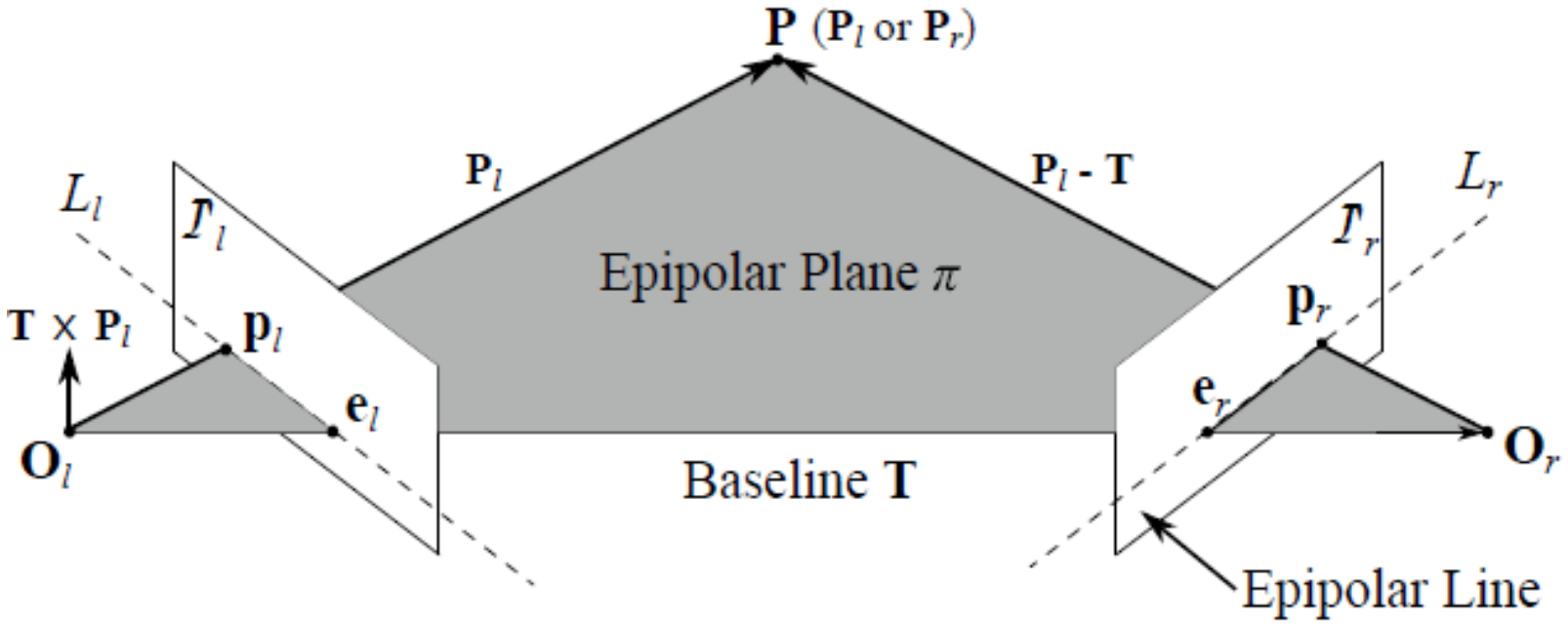
- *Epipolar geometry* is two-view geometry, i.e., the geometry of stereoscopic 3D vision.

- Property: Different 3D scene points projecting to the same left-view 2D point, may project to different right-view 2D points (*parallax effect*).



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Epipolar Geometry

- For a known 3D scene point, its left and right camera-system 3D coordinates $\mathbf{P}_l$, $\mathbf{P}_r$ are related by: $\mathbf{P}_r = \mathbf{R}(\mathbf{P}_l - \mathbf{T})$.
- By applying perspective projection:

$$\mathbf{p}_l = \frac{f_l}{Z_l}\mathbf{P}_l, \qquad \mathbf{p}_r = \frac{f_r}{Z_r}\mathbf{P}_r$$

- *Epipolar Constraint*: The image of line $\mathbf{O}_l\mathbf{P}$ on the left-view is the point $\mathbf{p}_l$, but on the right-view it is the right epipolar line L_r (and vice-versa).
  - Therefore, the match $\mathbf{p}_r$ of $\mathbf{p}_l$ must lie on L_r.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101087(MULTIDRONE)

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 $\mathbf{T}$).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

The Essential Matrix $\mathbf{E}$

- Normalized image plane counterpart:

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

- $\mathbf{E}$ geometrically relates the two views.
- Using $\mathbf{E}$, we can:
  a. map points to their epipolar line and
  b. recover extrinsic camera parameters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

The Fundamental Matrix $\mathbf{F}$

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

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

where the fundamental matrix $\mathbf{F}$ is given by the following relation:

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

- $\mathbf{F}$ 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

The Fundamental Matrix $\mathbf{F}$

- Thus, $\mathbf{E}$ only encodes extrinsic camera parameters, while $\mathbf{F}$ encodes both intrinsic and extrinsic ones.

- If we estimate $\mathbf{F}$ from known pixel correspondences between views, we can obtain $\mathbf{E}$:

$$\mathbf{E} = \mathbf{P}_{Ir}^{T}\mathbf{F}\mathbf{P}_{Il}.$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

The Fundamental Matrix $\mathbf{F}$

- $\mathbf{F}$ can be inferred solely by estimating pixel correspondences (uncalibrated cameras).

- Thus, $\mathbf{F}$ is useful in self-calibration (e.g., for uncalibrated 3D scene reconstruction), i.e., in determining intrinsic camera parameters purely from the visual content.
  - *Kruppa equations* are related to $\mathbf{F}$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Eight-point Algorithm

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

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

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Eight-point Algorithm

- The $i-th$ row of $\mathbf{X}$ has the following form:

$$\mathbf{X}_i = \begin{bmatrix} x_{dli}x_{dri} & x_{dli}y_{dri} & x_{dli} & y_{dli}x_{dri} & y_{dli}y_{dri} & y_{dli} & x_{dri} & y_{dri} & 1 \end{bmatrix}$$

- The system can be solved using SVD decomposition:

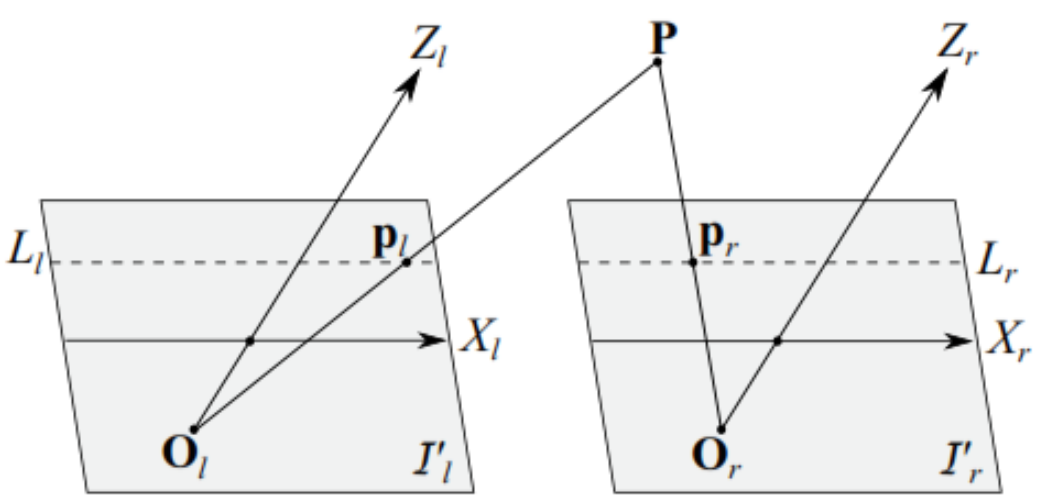
$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values. The solution $\mathbf{u}$ is the column of matrix $\boldsymbol{V}$ corresponding to the zero singular value of $\mathbf{\Sigma}$.
- Further steps alleviate the effect of noise.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

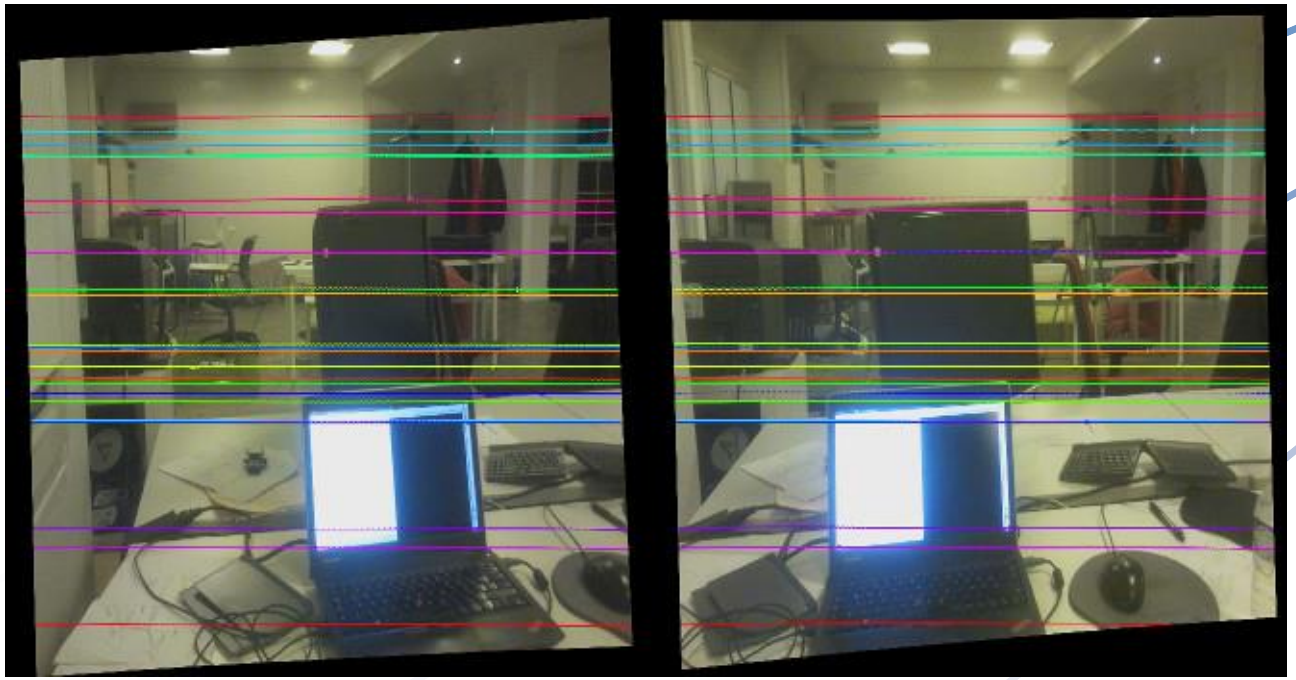
Rectification

- Rectification is the process of rotating each of the two image planes (left-right) around the corresponding optical centers, so that all epipolar lines become horizontal.
- In parallel stereo-rigs with equal focal lengths, the views are already rectified.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Rectification

- Most rectification methods require a number of known point correspondences between views, or the camera parameters.

- They define two virtual cameras with external parameters derived from rotating the actual cameras, so that the virtual image planes become co-planar.
  - Epipolar lines become horizontal:

$$\mathcal{P}_{vl} = \mathbf{P}_{Il}[\mathbf{R}| - \mathbf{R}\mathbf{O}_l], \qquad \mathcal{P}_{vr} = \mathbf{P}_{Ir}[\mathbf{R}| - \mathbf{R}\mathbf{O}_r]$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Rectification

- $\mathbf{R}$ gives the virtual camera orientation. It is an orthogonal change-of-basis matrix.

- It can be determined row-by-row.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Rectification

- The X axis must be parallel to the baseline:

$$\mathbf{R}_1 = \frac{\mathbf{O}_r - \mathbf{O}_l}{|\mathbf{O}_r - \mathbf{O}_l|}$$

- The Y axis must be orthogonal to X and the optical axis $\mathbf{k}$ of one of the initial cameras:

$$\mathbf{R}_2 = \mathbf{k} \times \mathbf{R}_1$$

- The Z axis is orthogonal to both X and Y:

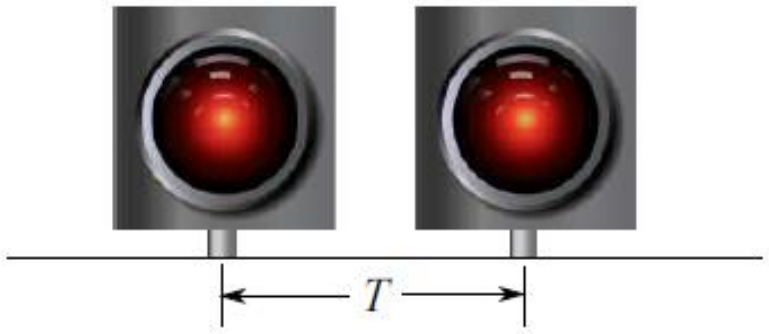
$$\mathbf{R}_3 = \mathbf{R}_1 \times \mathbf{R}_2$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

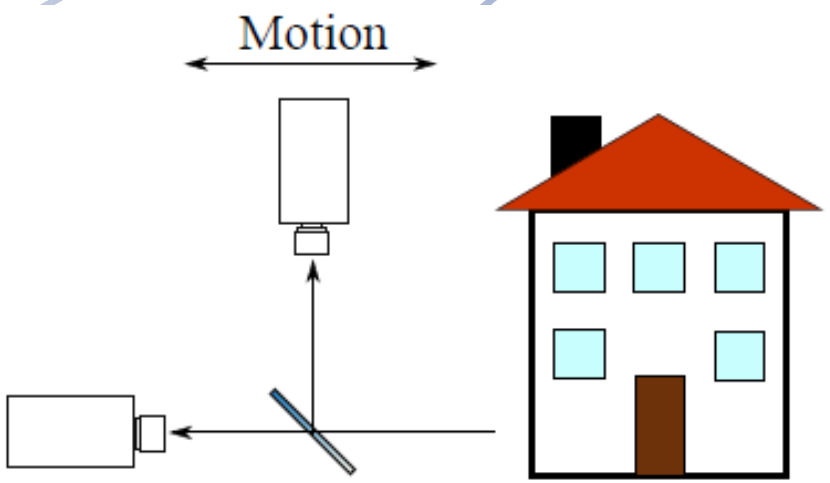
Stereo Camera Technologies

- They perform well as:
  - main cameras in soccer;
  - other large-field sports;
  - back position shots.
- When producing 3D video content, even if the two cameras are of exactly the same model, slight differences in their parameters will lead to discrepancies between the captured images.
- The baseline distance cannot become arbitrarily small.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

- For close-up shots, the *beamsplitter rig* is the best choice:
  - the two cameras are perpendicular to each other,
  - an appropriately positioned half surface mirror splits light between them,
  - the one camera shoots through the mirror,
  - the other captures the reflected light.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

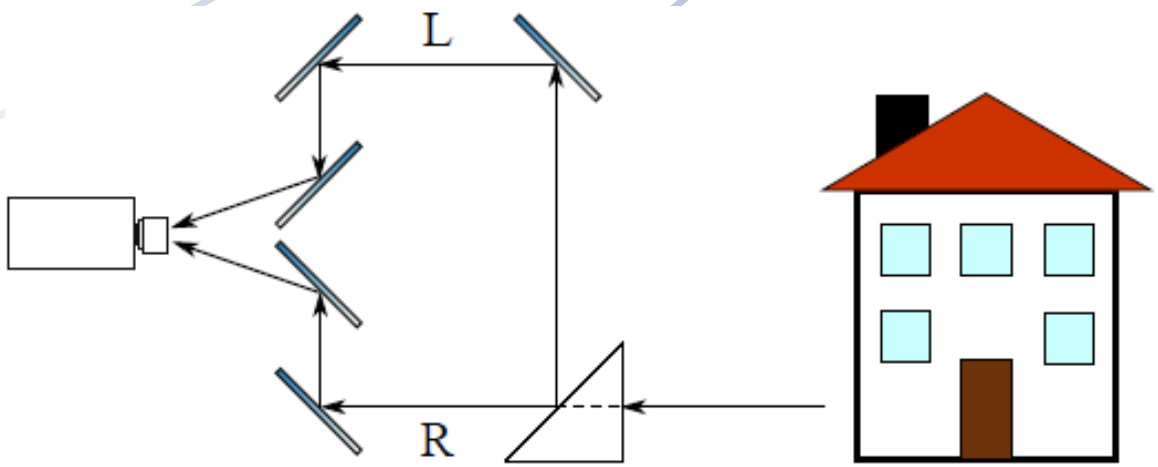
- Main drawbacks arise from the mirror:
  - even a slight mirror movement may cause serious problems;
  - the rig is very fragile;
  - sensitive to rapid movements and dust;
  - the light sent to the vertical camera through the mirror is polarized, due to its reflection;
  - vertical camera recordings need to be flipped before synthesis with the horizontal ones.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

- A different beamsplitter system, also employs mirrors, but consists of only one mono camera with a prism attached in front of the lens, able to split a light beam in two.

- Two, side-by-side images of the scene are produced on the same image frame.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

- Main drawbacks:
  - As the two images pass through two different sides of the lens, they undergo different distortions.
  - Light scattering caused by the mirrors may result in ghosting.
  - Reduced spatial image resolution.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

- *Coupled stereo cameras* can produce:
  - *Two separate video files.*
  - *One side-by-side video file.*
  - *One multi-video coded file.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Stereo Camera Technologies

- Monoscopic cameras with special mounted stereoscopic lenses can also be used for left and right image capturing:
  - mainly used for shots with a fixed interaxial distance,
  - produce flawlessly synchronized and aligned left and right images.
- Main drawback: The image resolution is reduced, as smaller portion of the image sensor is used for each stereo image channel.

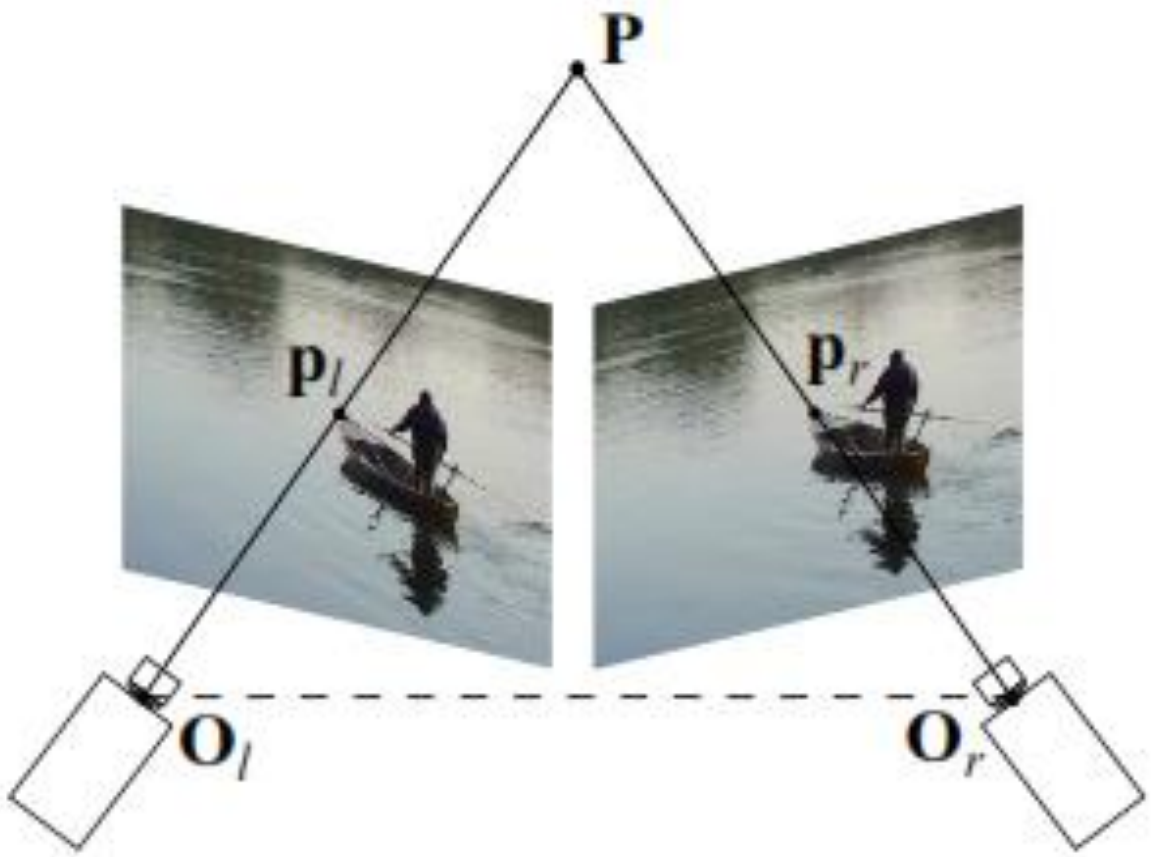
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Correspondence

- The 3D geometry of a scene can be recovered, given multiple 2D scene views.
- The multiple views may come from different view points / cameras:
  - stereoscopic / binocular view, trinocular view, Multiview imaging
- The may come from a moving camera: *Structure-from-Motion* (SfM).
  - In SfM, the scene has to be static.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Correspondence



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

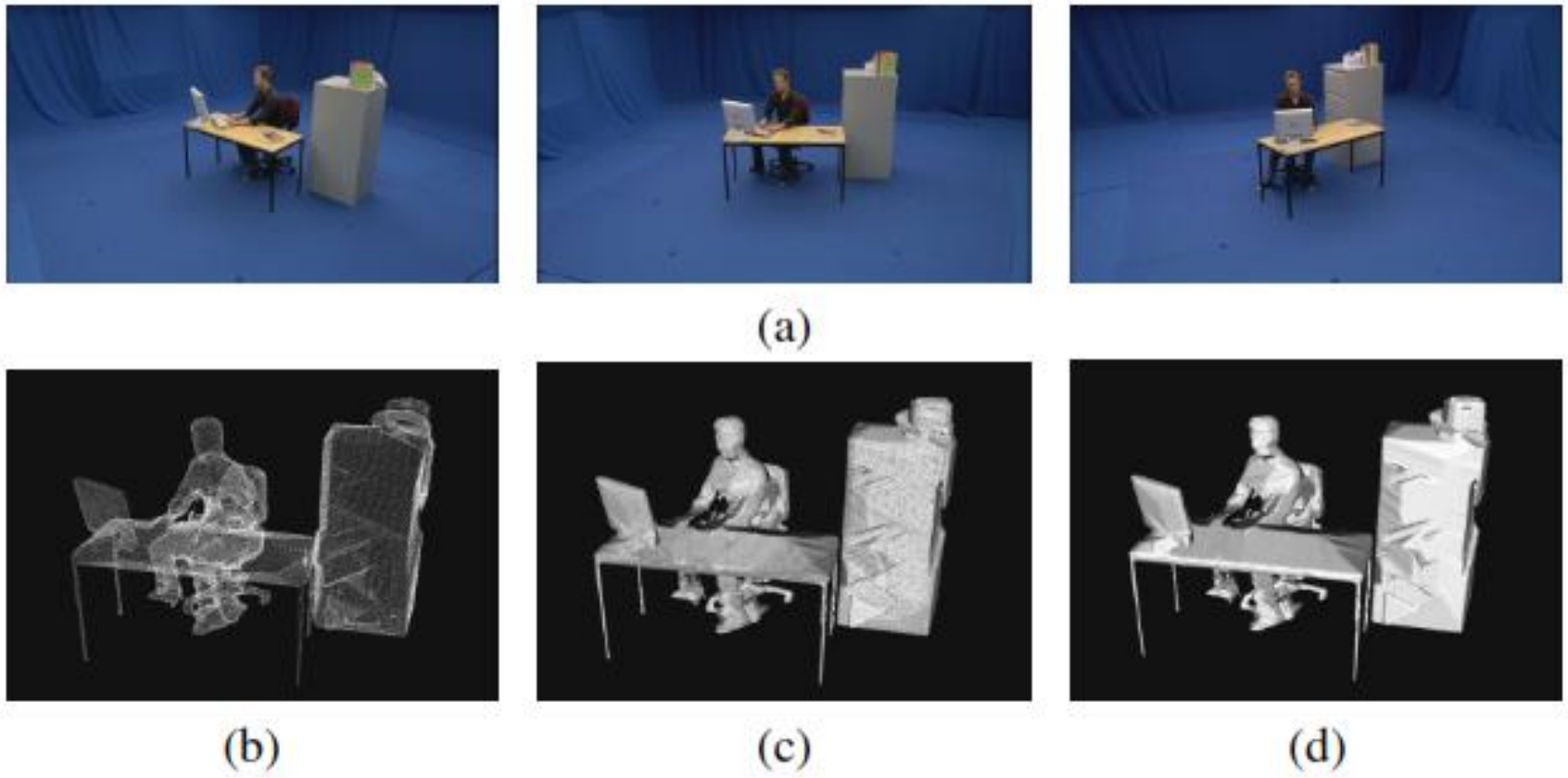
Feature Correspondence

- Point correspondences have to be estimated between all views.

- Then, if all camera parameters are known, the 3D world location of each corresponded point can be found by *triangulation.*

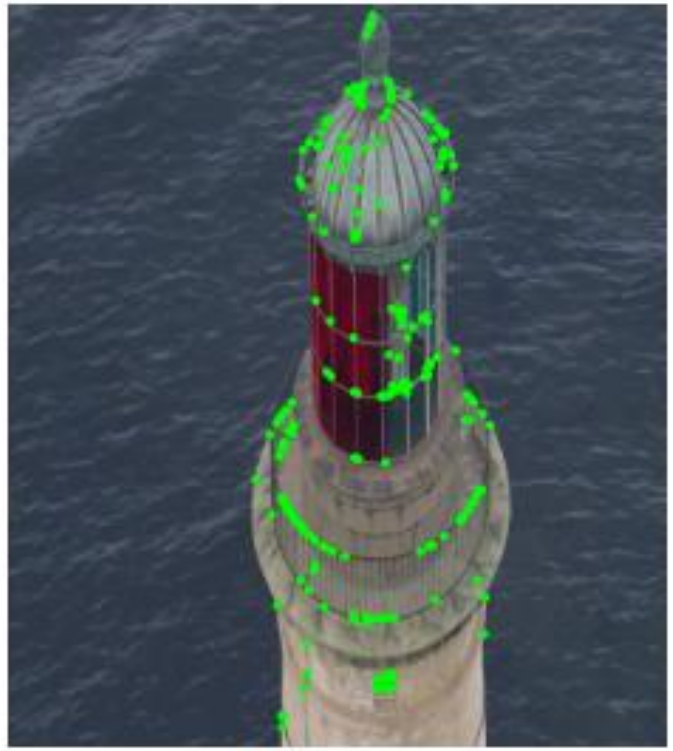
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Correspondence



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Correspondence



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Extraction

- 2D point correspondence
  - Pixel correspondence
    - Feature correspondence.
- Local feature: a small image region having interesting spatial characteristics (e.g., corner).
  - It can be described by a N-dimensional vector.
- Typically (not always), a feature detector/descriptor tries to produce description vectors invariant to several image transformations.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- *Homologous* image features: projections of the same natural 3D point on each camera view, after *feature matching.*

- Commonly used constraints for reduction of the search space for feature correspondences:
  - *Epipolar constraint:* when the projection geometry is known, search for a corresponding feature point can be restricted to the epipolar line on the other image of the stereo pair.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- *Uniqueness constraint:* a point in one view has at most one corresponding match in the other view.

- *Continuity constraint:* adjacent feature points in one view should correspond to adjacent features in the other view.

- *Topological constraint:* the relative position of 3D points remains unaltered in their projections to all views.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Area-based matching algorithms are the oldest matching methods, used mainly for low-level feature matching.

- Matching of two feature points is based on the minimization of some distance measure of the respective local image windows.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Assuming feature points $\mathbf{p}_l = [x_l, y_l]^T$ and $\mathbf{p}_r = [x_r, y_r]^T$ to be matched, the grayscale intensity images $f_l(x, y)$ and $f_r(x, y)$ in a $L = (2N + 1) \times (2M + 1)$ local neighborhood window centered around these points are going to be compared.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Distance-based matching measures which can be used:
  - The *sum of absolute differences* (SAD) or L_1 norm:

$$SAD(\mathbf{p}_l, \mathbf{p}_r) = \sum_{i=-N}^{N} \sum_{j=-M}^{M} |f_l(x_l + i, y_l + j) - f_r(x_r + i, y_r + j)|$$

  - The *sum of squared differences* (SSD) or L_2 norm:

$$SSD(\mathbf{p}_l, \mathbf{p}_r) = \sum_{i=-N}^{N} \sum_{j=-M}^{M} \left(f_l(x_l + i, y_l + j) - f_r(x_r + i, y_r + j)\right)^2$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Correlation-based similarity measures which can be used:
  - The *normalized cross-correlation* (NCC):

$$NCC(\mathbf{p}_l, \mathbf{p}_r) = \frac{\sigma_{lr}^2(p_l, p_r)}{\sqrt{\sigma_l^2(p_l)\sigma_r^2(p_r)}}$$

where:

$$\sigma_{lr}^2(\mathbf{p}_l, \mathbf{p}_r) = \frac{1}{(2N+1)(2M+1)} \sum_{i=-N}^{N} \sum_{j=-M}^{M} \left(f_l(x_l+i, y_l+j) - \bar{f}_l\right) \cdot \left(f_r(x_r+i, y_r+j) - \bar{f}_r\right)$$

$$\sigma_l^2(\mathbf{p}_l) = \frac{1}{(2N+1)(2M+1)} \sum_{i=-N}^{N} \sum_{j=-M}^{M} \left(f_l(x_l+i, y_l+j) - \bar{f}_l\right)^2$$

$$\sigma_r^2(\mathbf{p}_r) = \frac{1}{(2N+1)(2M+1)} \sum_{i=-N}^{N} \sum_{j=-M}^{M} \left(f_r(x_r+i, y_r+j) - \bar{f}_r\right)^2$$

Feature Matching Algorithms

- The somewhat more stable than NCC is the modified *normalized cross-correlation* (MNCC):

$$MNCC(\mathbf{p}_l, \mathbf{p}_r) = \frac{2\sigma_{lr}^2(p_l, p_r)}{\sigma_l^2(p_l) + \sigma_r^2(p_r)}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Other image feature characteristics which can be used for feature matching:
  - Edge attributes (e.g., edge orientation, location, intensity difference between the two sides of the edges).
    - They may suffer from occlusion problems.
  - Corner attributes (e.g., coordinates):
    - Harris detector.
  - Orientation of line segments and coordinates of the end or/and mid points.
    - Detection methods not robust against noise.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Curve segments.
  - Not frequently used because of high computational complexity and matching ambiguities.
- Curve attributes (e.g., turning points).
- Circles.
- Ellipses.
- Polygonal regions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- Most of the feature-based stereo or multiview matching systems use a combination of features and compare the descriptor token vectors containing the attributes of each feature point.
  - Edges, curves, surface and region patches.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- General matching approach based on a similarity metric between a token vector pair – nearest neighbor search:
  - If $\mathbf{x}, \mathbf{y}$ feature descriptor vectors and $\mathbf{w}$ the weight vector of the feature token type, then the similarity is given by:

$$S = \frac{1}{\|\mathbf{w}^T(\mathbf{x} - \mathbf{y})\|}$$

  - $\|\cdot\|$ can be the Euclidean distance metric and $\mathbf{w}$ can be omitted if every feature characteristic is equally important.
- We search for a pair of feature descriptors maximizing the similarity.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Feature Matching Algorithms

- *Naïve nearest neighbor search*: a brute force approach.
  - Calculate the similarity of each feature point on one image with every other feature point on the other image and match the pair with maximum similarity.

- *Best-bin-first search*: a faster but approximate method - modification of kd-tree search.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

3D Reconstruction Techniques in Stereo Vision

- Three general cases of 3D point cloud reconstruction:
  - *Calibrated* cameras: known intrinsic and extrinsic parameters – reconstruction is a simple matter of triangulation.
  - Uncalibrated cameras: some or all camera parameters are unknown – calibration needs to take place.
    - Known intrinsic parameters only: estimation of the extrinsic parameters and the 3D geometry up to an unknown scaling factor can solve the problem.
  - No parameters known: 3D reconstruction only possible up to an unknown projective transformation.

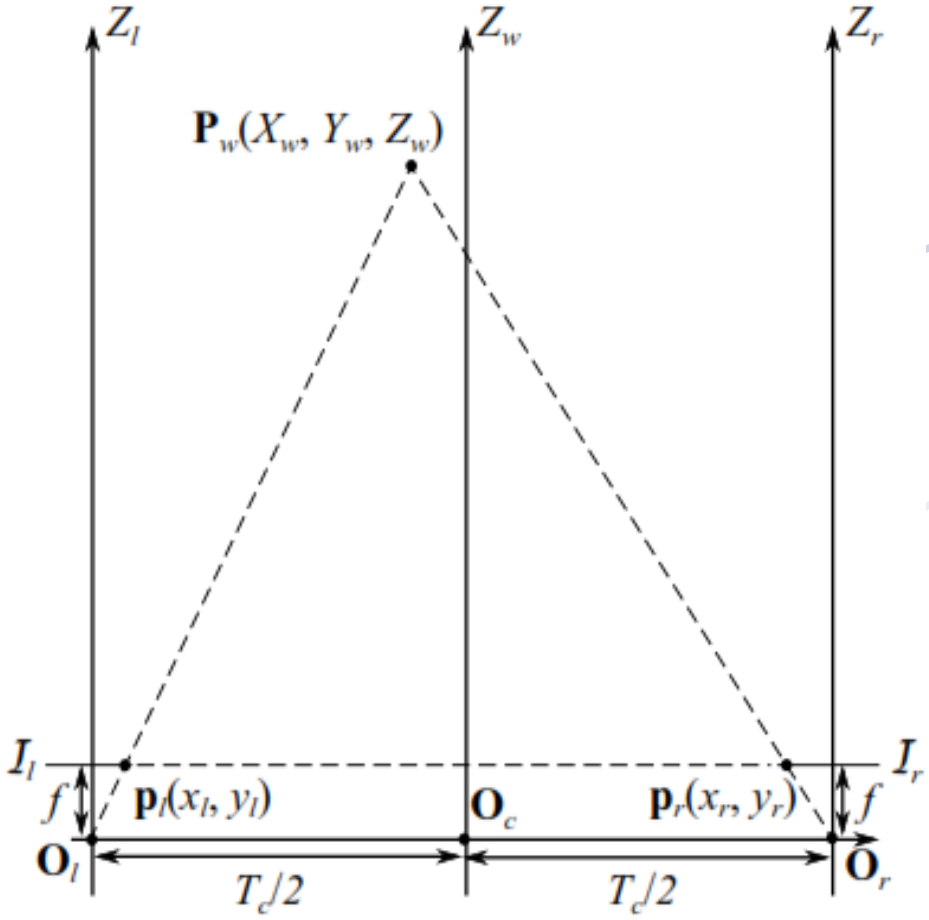
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- $\mathbf{O}_l, \mathbf{O}_r$: the centers of projection of the left/right camera – origins of the coordinate systems (X_l, Y_l, Z_l), (X_r, Y_r, Z_r).
- $\mathcal{I}_l$, $\mathcal{I}_r$: the virtual image planes of the left/right camera.
- T_c: the camera baseline - distance between the two centers of projection.
- f: the camera focal length - distance between the center of projection of a camera and its image plane.
- $\mathbf{O}_c$: the center of the world coordinate system (X_w, Y_w, Z_w) - baseline midpoint.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups



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

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- Triangle similarities can be used to recover 3D world coordinates $\mathbf{P}_w$ from left/right image plane coordinates, assuming all camera parameters and point disparity values $d_c = x_r - x_l$ are known:

$$Z_w = -\frac{fT_c}{d_c}, \qquad X_w = -\frac{T_c(x_l + x_r)}{2d_c}, \qquad Y_w = -\frac{T_c y_l}{d_c} = -\frac{T_c y_r}{d_c}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- Such a camera setup produces non-positive disparity values.
- Thus, during display, all points appear in front of the screen.

- Points at infinity $(Z_l = Z_r = Z_w = \infty)$ produce zero camera disparity and are displayed on the screen.

- The closer the 3D point is to the camera during filming, the larger its camera disparity is (in absolute value).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- In the converging camera setup:
  - the left/right camera optical axes form an angle θ with the coordinate axis Z_w and
  - converge on Z_w at distance $Z_c = \frac{T_c}{2}\frac{1}{\tan\theta} = \frac{T_c}{2}\tan\left(\frac{\pi}{2} - \theta\right)$ from the camera centers.

- Typically, θ is small, so that $\sin\theta \approx 0, \tan\theta \approx 0, \cos\theta \approx 1$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- A point in world space projected on the left and right image planes, can be transformed into the left/right camera systems by translating first by $T_c/2$ and then rotating by an angle $-\theta$ about the Y_w axis:

$$\begin{bmatrix} X_l \\ Y_l \\ Z_l \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} X_w + \frac{T_c}{2} \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} \left(X_w + \frac{T_c}{2}\right)\cos\theta - Z_w \sin\theta \\ Y_w \\ \left(X_w + \frac{T_c}{2}\right)\sin\theta + Z_w \cos\theta \end{bmatrix}$$

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} X_w - \frac{T_c}{2} \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} \left(X_w - \frac{T_c}{2}\right)\cos\theta + Z_w \sin\theta \\ Y_w \\ -\left(X_w - \frac{T_c}{2}\right)\sin\theta + Z_w \cos\theta \end{bmatrix}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Parallel and Converging Camera Setups

- A point in left/right image plane coordinates can be reverted to world space coordinates by:

$$X_w = T_c \frac{x^l + \tan\theta\left(f + \frac{x^l x^r}{f} + x^r \tan\theta\right)}{x^l - x^r + \tan\theta\left(2f + 2\frac{x^l x^r}{f} - x^l \tan\theta + x^r \tan\theta\right)} - \frac{T_c}{2}$$

$$Y_w = T_c \frac{y^l}{f} \frac{\cos\left(\arctan\left(\frac{x^l}{f}\right) + \theta\right)\cos\left(\arctan\left(\frac{x^l}{f}\right)\right)}{\sin\left(\arctan\left(\frac{x^l}{f}\right) + \arctan\left(\frac{x^r}{f}\right) + 2\theta\right)}$$

$$Z_w = T_c \frac{f - \left(d + \frac{x^l x^r}{f}\tan\theta\right)\tan\theta}{x^l - x^r + \tan\theta\left(2f + 2\frac{x^l x^r}{f} - x^l \tan\theta + x^r \tan\theta\right)}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

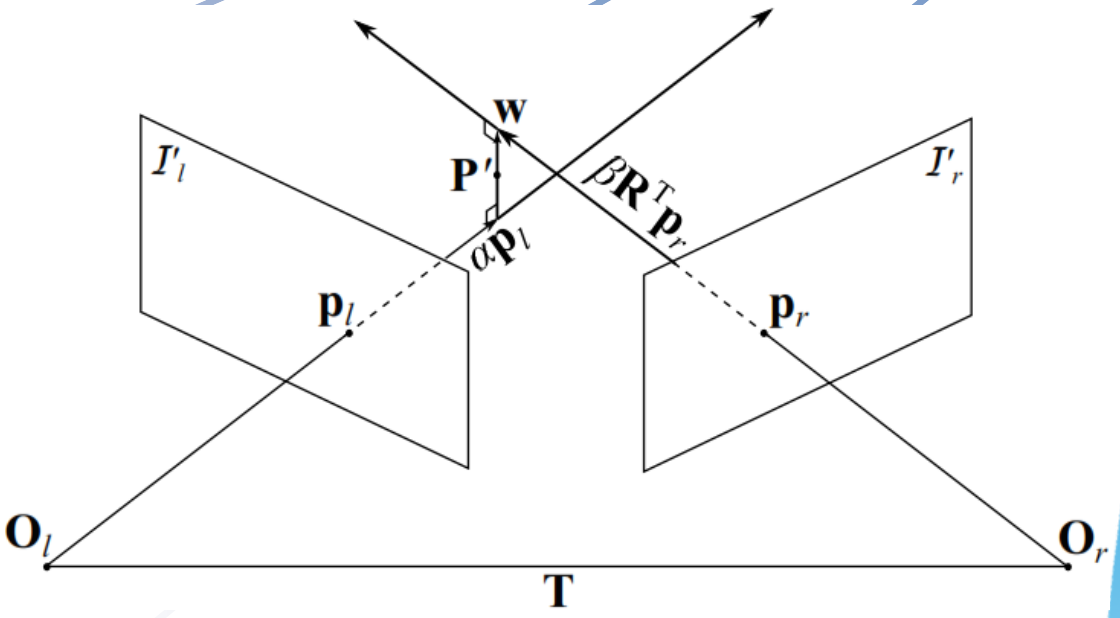
Parallel and Converging Camera Setups

- The converging camera setup produces:
  - Negative camera disparities for object having $Z_w < Z_c$, thus appearing in front of the screen place.
  - Zero camera disparities for object having $Z_w = Z_c$, thus appearing on the screen plane.
  - Positive camera disparities for object having $Z_w > Z_c$, thus appearing behind the screen plane.
- Objects at infinity $(Z_l = Z_r = Z_w = \infty)$ have large positive disparity.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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**.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

3D reconstruction from known intrinsic camera parameters only

- 3D reconstruction is possible even without knowledge of extrinsic camera parameters, up to an unknown scaling factor.
  - The essential matrix $\mathbf{E}$ can be estimated up to an unknown scaling factor using the eight-point algorithm, but has to be normalized.
  - Then, the extrinsic parameters $\mathbf{R}, \mathbf{T}$ can be recovered from $\mathbf{E}$, since $\mathbf{E} = \mathbf{R}\mathbf{T}_{\times}$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

3D reconstruction from known intrinsic camera parameters only

- Normalized $\mathbf{T}$ can easily be found using:

$$\tilde{\mathbf{E}}^T\tilde{\mathbf{E}} = \begin{bmatrix} 1-\tilde{T}_x^2 & -\tilde{T}_x\tilde{T}_y & -\tilde{T}_x\tilde{T}_z \\ -\tilde{T}_y\tilde{T}_x & 1-\tilde{T}_y^2 & -\tilde{T}_y\tilde{T}_z \\ -\tilde{T}_z\tilde{T}_x & -\tilde{T}_z\tilde{T}_y & 1-\tilde{T}_z^2 \end{bmatrix}$$

- Subsequently, vector $\mathbf{w}$ is defined as follows:

$$\mathbf{w}_i = \tilde{\mathbf{E}}_i \times \tilde{\mathbf{T}}$$

- The i-th row of $\mathbf{R}$ can then be recovered in the following manner:

$$\hat{\mathbf{R}}_i = \mathbf{w}_i + \mathbf{w}_j \times \mathbf{w}_k$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Three-Views and the Point Transfer

- *Trifocal geometry*: the geometry we deal with in the case of three views.
- *Trifocal tensor:*
  - It puts all the geometric relations between three views in a nutshell.
    - It can be used to map corresponding points in two views to their correspondences in the third view.
    - It can be applied to straight lines, since the image of a line in one view can be computed from its corresponding images in the two other views.
  - It depends only on the geometric transformation parameters between views and the intrinsic camera parameters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

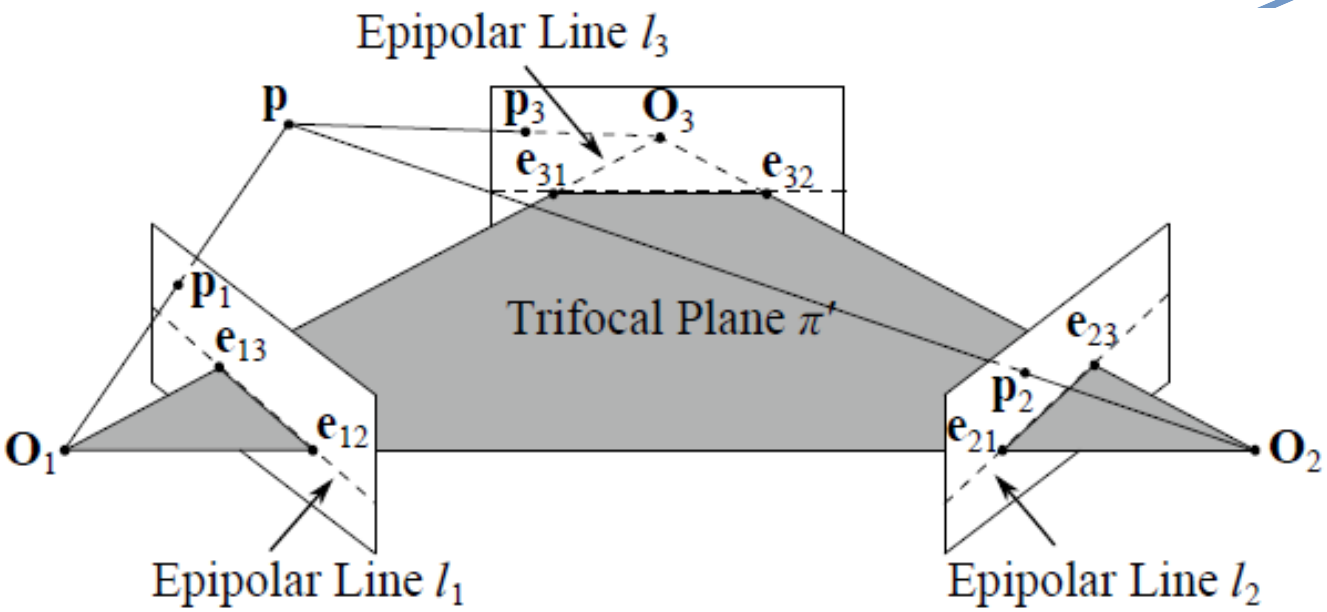
Three-Views and the Point Transfer

- It is uniquely defined by the camera projection matrices.
- It can be computed from image point correspondences, without any prior information about the camera calibration parameters.

- Assuming a thee camera system:
  - The *trifocal plane* can be defined by the three distinct optical centers of the cameras $\mathbf{O}_1, \mathbf{O}_2, \mathbf{O}_3$ (*general viewpoint assumption*).
  - $\mathbf{F}_{12}, \mathbf{F}_{13}, \mathbf{F}_{23}$ the fundamental matrices of the three possible view pairs.
  - $\mathbf{e}_{ij}, \mathrm{i}, \mathrm{j} = 1,2,3$ the trifocal system epipoles.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

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 $\mathbf{p}_1, \mathbf{p}_2$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Three-Views and the Point Transfer

- *Point transfer* or *prediction* of point position in the third image plane by:

$$\mathbf{p}_3 = \mathbf{F}_{13}\mathbf{p}_1 \times \mathbf{F}_{23}\mathbf{p}_2.$$

- Considering that the three camera centers are not collinear, the epipolar line $\mathbf{F}_{23}\mathbf{e}_{31}$ of the epipole $\mathbf{e}_{31}$ on the second image, is the line connecting $\mathbf{e}_{21}$ and $\mathbf{e}_{23}$:
  - The epipolar line $\mathbf{l}_2 = \mathbf{e}_{21} \times \mathbf{e}_{23}$ will be identical to the epipolar line $\mathbf{F}_{23}\mathbf{e}_{31}$.
  - This still applies after permuting the indices $1,2,3$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Three-Views and the Point Transfer

- By noticing that $\mathbf{e}_{21}$ and $\mathbf{e}_{23}$ are corresponding points, since they are both images of the $\mathbf{O}_1$ camera center, and permuting the indices, we get:

$$\mathbf{e}_{31}^T \mathbf{F}_{23} \mathbf{e}_{21} = 0 \qquad \mathbf{e}_{12}^T \mathbf{F}_{31} \mathbf{e}_{32} = 0 \qquad \mathbf{e}_{23}^T \mathbf{F}_{12} \mathbf{e}_{13} = 0$$

- When the optical centers are aligned:
  - there is no trifocal plane;
  - the three camera centers have to lie on the same baseline and
  - the epipoles have to satisfy: $\mathbf{e}_{12} = \mathbf{e}_{13},\ \mathbf{e}_{21} = \mathbf{e}_{23}, \mathbf{e}_{31} = \mathbf{e}_{32}$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

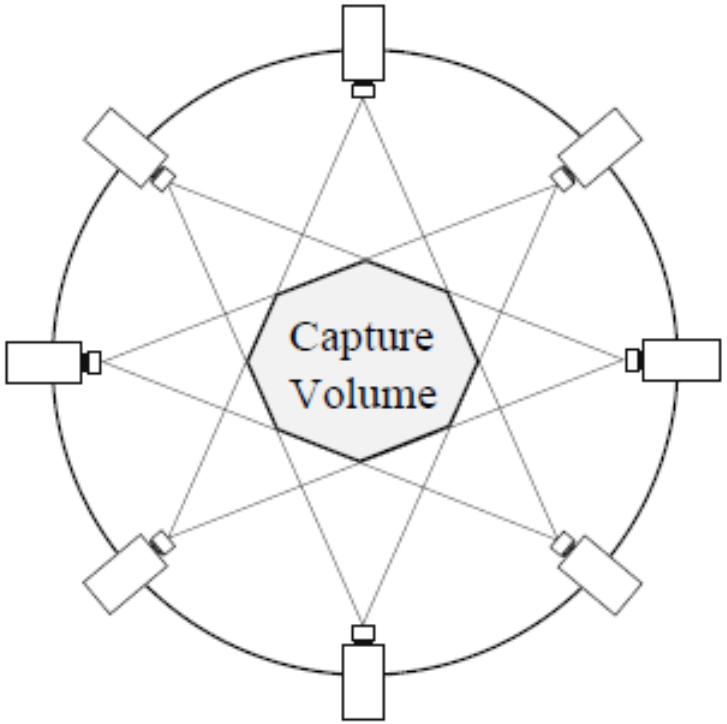
Multiple Camera Image Acquisition

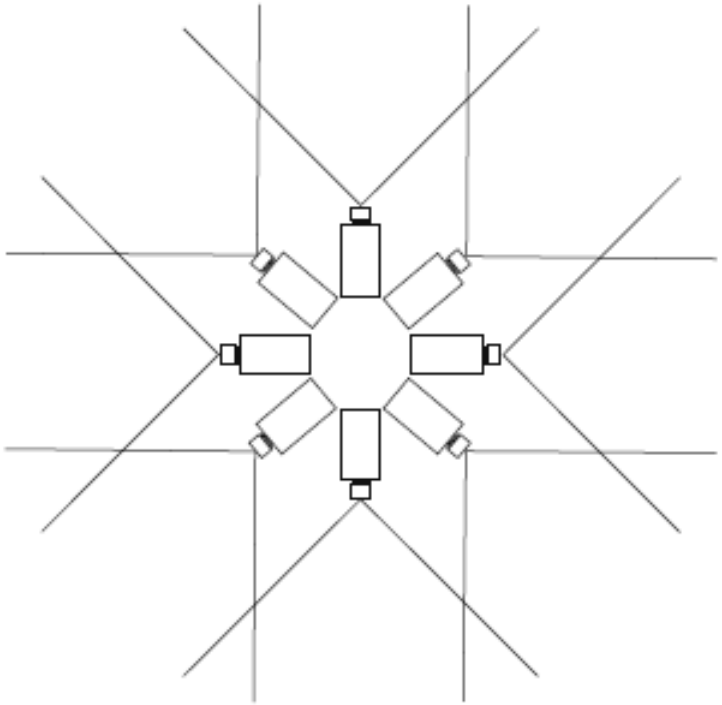
- A multiple *witness camera* setup is usually employed, alongside the primary production cameras, when acquisition of complimentary 3D is desired in professional movie shooting.
- Such systems can also be used for:
  - motion capturing,
  - 2D tracking, when the tracked object is partially occluded.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Multiple Camera Image Acquisition

- Circular camera positioning setups.





This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Multiple Camera Image Acquisition

- A two-way information exchange network is established between the cameras and the processing unit, via a parallel connection.
- For performance reasons, the optimal assignment of the camera units to the interconnection network nodes requires:
  - preservation of the spatial geometry of the units,
  - minimization of the mutual communication/access time between the processing unit and any camera.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Coverage

- Voronoi diagrams can be used for calculating the positioning and coverage of each camera, especially when units are placed over a dome.
- For output image production, each pixel is reconstructed taking into account the contribution of each camera.
- *Gaussian blending* is applied when merging the images acquired by each camera, taking into account:
  - the camera position and
  - the pixel positions on the respective image plane.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Multiple camera calibration with regard to a reference view

- Other cameras are calibrated vs a *reference camera.*
- Most methods based on a moving planar surface with a printed checkerboard pattern:
  - Pros: not requiring a 3D calibration object.
  - Cons: leading to partially calibrated cameras, since the moving calibration grid is not visible from all views.
- *Dominant plane*: extrinsic camera parameters are estimated from moving object trajectories on the basis of a common coordinate system.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Multiple camera calibration with regard to a reference view

- *Planar pattern:* known line parallelism and information extracted from the pattern texture are used to derive constraints and initialize the projection matrices.
- *Bright spot:* a bright spot is moved throughout the 3D scene volume, its projections in each camera are detected and a common matrix accumulating their homogeneous image coordinates is formed. Projective structures are refined to Euclidean and camera projection matrices are recovered.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Self-calibration

- *Self-calibration* or *autocalibration:* a family of camera calibration methods using solely image information.
  - Pros: flexibility, simplicity.
  - Cons: lacking robustness, may fail in degenerate cases.

- Relies on epipolar and projective geometry concepts.

- Multiple views of a 3D scene are needed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Self-calibration

- Recent self-calibration approach: stochastic optimization of a metric that depends on the intrinsic parameters $\mathbf{P}_{Il}, \mathbf{P}_{Ir}$.

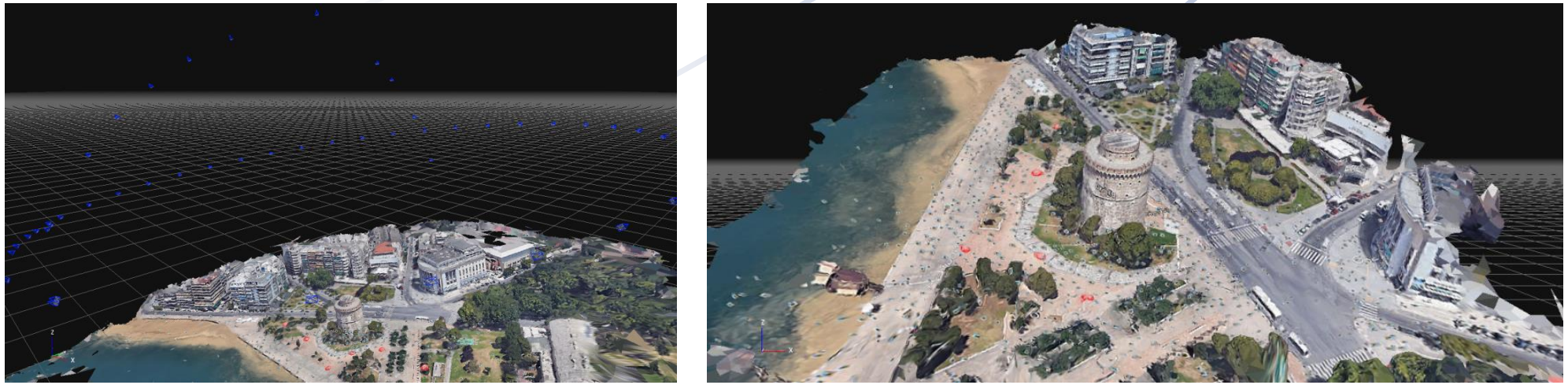
- Optimization algorithms employed to this end:
  - particle swarm optimization,
  - genetic optimization.

- Commonly used metric for minimization:
  - a reformulation of the simplified *Kruppa equation.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

3D Scene Reconstruction from Uncalibrated Multiple Cameras

Images obtained from Google Earth



3D models reconstructed in 3DF Zephyr Free using 50 images from Google Earth

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@aiia.csd.auth.gr
www.multidrone.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)