Computer Vision

Levente Hajder, Dmitry Chetverikov

Eötvös Loránd University, Faculty of Informatics

Basics of Stereo Vision

Outline

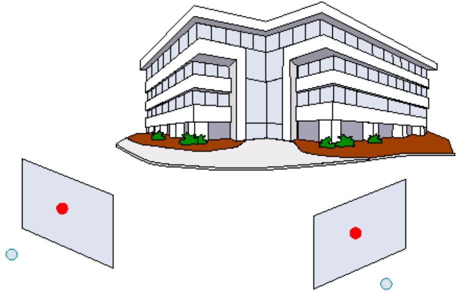
Single, calibrated image 1/2

- Depth cannot be measured
  - at least two cameras required for depth estimation.
- Surface normal can be estimated
  - integration of normals $\longrightarrow$ surface
  - sensitive to depth change
- Surface normal estimation possible in smooth, textureless surfaces
  - **shape from shading**
  - intensity change $\longrightarrow$ surface normal
  - less robust
  - reconstruction ambiguity

Single, calibrated image 2/2

- Texture-change in a smooth, regularly-textured surface
  - **shape from texture**
  - texture change $\longrightarrow$ surface normal
  - less robust
- Illumination change
  - photometric stereo
  - more light sources $\longrightarrow$ surface normal
  - robust, but ambiguity can present
  - high, finer details
  - 3D position is less accurate
- Special scenes
  - e.g. parallel and perpendicular lines
  - $\rightarrow$ buildings, rooms, ...
  - applicability is limited

Stereo vision illustration

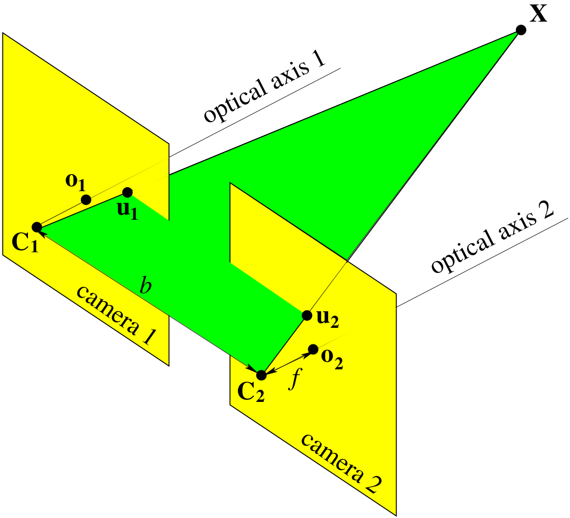


- For reconstructing a 3D scene,
  - at least two, calibrated images required.
  - and point correspondences given in the images.
- The process is called **triangulation.**

Standard stereo

- Same calibrated cameras applied for taking the images
- Optical axes are parallel
- Planes of images are the same, as well as lower and upper border lines
- Baseline between focal points is small
  - *narrow baseline*
- Operating principles
  - correspondences obtained by maching algorithms
  - depth estimation by triangulation
- Following parameters have to know for triangulation:
  - baseline b
  - focal length f
  - disparity d
- Disparity: point location difference between images

Geometry of standard stereo



Wide-baseline stereo

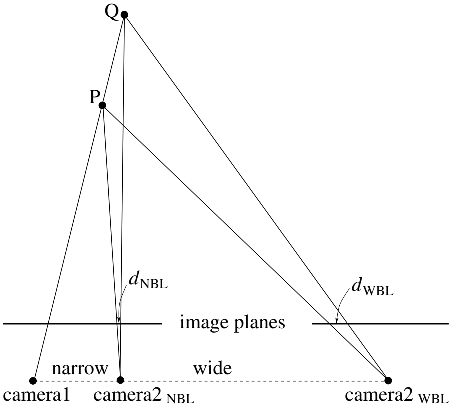
- Calibrated camera(s)
  - two images taken from different viewpoints
- Baseline is larger
  - *wide baseline*
- Benefits over standard stereo
  - larger disparities

  → more accurate depth estimation
- Disadvantages
  - geometric distortion in images are larger
  - more occlusions

  → point maching is more difficult

Example for narrow/wide baseline stereo

- Points P and Q are on the same projective ray
  - $\rightarrow$ First cameras are the same
- $d_{\mathrm{WBL}} \gg d_{\mathrm{NBL}}$
  - $\rightarrow$ more accurate estimation for WBL
- d_{NBL} is very small
  - more correspondences
  - $\rightarrow$ rounding noise
  - $\rightarrow$ depth is layered



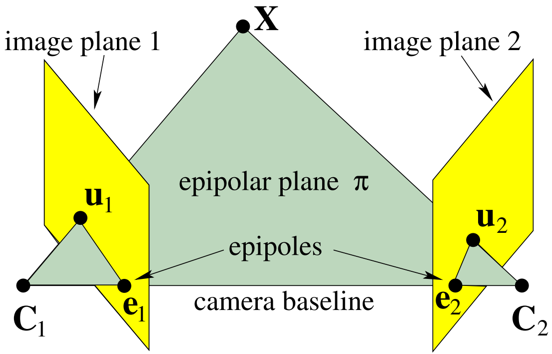
Outline

correspondence-based stereo vision

- Image-based 3D algorithms usually exploit point correspondences in images
  - Pattern matching in images is a challenging task
- Less DoF $\longrightarrow$ faster, more robust solutions
  $\rightarrow$ geometric constraint should be applied
- **Epipolar geometry $\longrightarrow$ epipolar constraint**
  - epipolar lines correspond to each other
  - 2D search $\rightarrow$ 1D-s search
- Stereo geometry
  - uncalibrated cameras $\longrightarrow$ **fundamental matrix**
  - calibrated cameras $\longrightarrow$ **essential matrix**
  - image rectification $\longrightarrow$ 1D matching

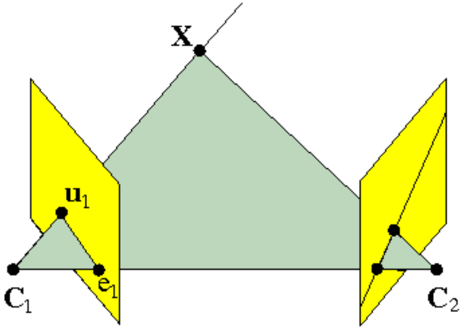
Overview

Geometry of stereo vision



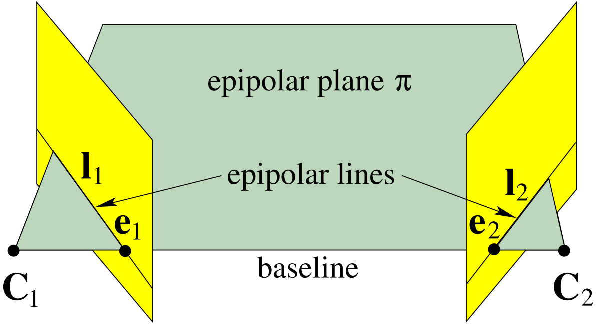
- **Baseline C_1C_2** connects two focal points.
- Baselines intersect image planes at epipoles.
- Two focal points and the spatial point **X** defines **epipolar plane**.

Geometry of stereo vision: a video



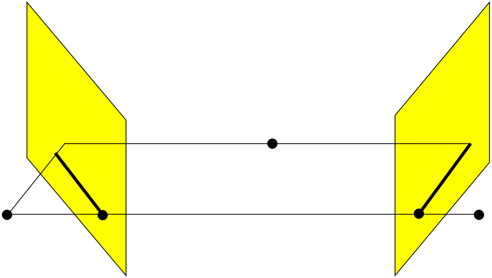
- Point **X** lies on line on ray back-projected using the point in the first image
- Point in the second image, corresponding to $\mathbf{u}_1$, lies on an **epipolar line**
  - → **epipolar constraint**
- Line $\mathbf{u}_1\mathbf{e}_1$ is the related epipolar line in the first image.

Epipolar geometry



- Each plane, containing the baseline, is an epipolar plane
- Epipolar plane π intersects the images at lines $\mathbf{l}_1$ and $\mathbf{l}_2$.
  - $\rightarrow$ Two epipolar lines correspond to each other.

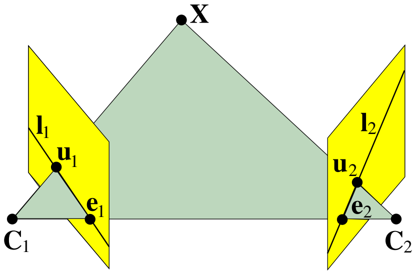
Epipolar geometry: video



- Epipolar plane 'rotates' around the baseline.
- Each epipolar line contains epipole(s).

Overview

Calibrated cameras: essential matrix 1/2



- Calibration matrix **K** is known, rotation **R** and translation **t** between coordinate systems are unknown.
- Lines $\mathbf{C}_1\mathbf{u}_1$, $\mathbf{C}_2\mathbf{u}_2$, $\mathbf{C}_1\mathbf{C}_2$ lay within the same plane:

$$\mathbf{C}_2\mathbf{u}_2 \cdot [\mathbf{C}_1\mathbf{C}_2 \times \mathbf{C}_1\mathbf{u}_1] = 0$$

Calibrated cameras: essential matrix 2/2

- In the second camera system, the following equation holds if homogeneous coordinates are used:

$$\mathbf{u}_2 \cdot [\mathbf{t} \times \mathbf{R}\mathbf{u}_1] = 0$$

- Using the **essential matrix E** (Longuet-Higgins, 1981):

$$\mathbf{u}_2^T \mathbf{E} \mathbf{u}_1 = 0, \quad (1)$$

where essential matrix is defined as

$$\mathbf{E} \doteq [\mathbf{t}]_\times R \quad (2)$$

  - $[\mathbf{a}]_\times$ is the cross-product matrix:

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b} \doteq \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Properties of an essential matrix

- The equation $\mathbf{u}_2^T \mathbf{E} \mathbf{u}_1 = 0$ is valid if the 2D coorinates are normalized by $\mathbf{K}$.
  - Normalized camera matrix: $\mathbf{P} \longrightarrow \mathbf{K}^{-1}\mathbf{P} = [\mathbf{R}| - \mathbf{t}]$
  - $\rightarrow$ Normalized coordinates: $\mathbf{u} \longrightarrow \mathbf{K}^{-1}\mathbf{u}$
- Matrix $\mathbf{E} = [\mathbf{t}]_\times R$ has 5 degree of freedom (DoF).
  - $3(\mathbf{R}) + 3(\mathbf{t}) - 1(\lambda)$
  - λ: (scalar unambigity)
- Rank of essential matrix is 2.
  - $\mathbf{E}$ has two equal, non-zero singular value.
- Matrix $\mathbf{E}$ can be decomposed to translation and rotation by SVD.
  - translation is up to an unknown scale
  - sign of $\mathbf{t}$ is also ambiguous

Uncalibrated case: fundamental matrix

- Longuet-Higgins formula in case of **uncalibrated** cameras

$$\mathbf{u}_2^T F \mathbf{u}_1 = 0, \tag{3}$$

  where the **fundamental matrix** is defined as

$$\mathbf{F} \doteq \mathbf{K}_2^{-T} \mathbf{E} \mathbf{K}_1^{-1} \tag{4}$$

  - $\mathbf{u}_1$ and $\mathbf{u}_2$ are unnormalized coordinates.
- Matrix **F** has 7 DoF.
- Rank of **F** is 2
  - Epipolar lines intersect each other in the same points
  - $\det \mathbf{F} = 0 \longrightarrow$ **F** cannot be inverted, it is non-singular.
- Epipolar lines: $\mathbf{l}_1 = \mathbf{F}^T \mathbf{u}_2$, $\mathbf{l}_2 = \mathbf{F} \mathbf{u}_1$
- Epipoles: $\mathbf{F}\mathbf{e}_1 = \mathbf{0}$, $\mathbf{F}^T \mathbf{e}_2 = \mathbf{0}^T$

Overview

Estimation of fundamental matrix

- We are given N point correspondences: $\{\mathbf{u}_{1i} \leftrightarrow \mathbf{u}_{2i}\}, i = 1, 2, \ldots, N$
  - Degree of freedom for **F** is 7 : $\longrightarrow N \geq 7$ required
  - Usually, $N \geq 8$. (Eight-point method)
  - If correspondences are contaminated $\longrightarrow$ robust estimation needed
  - In case of outliers: $N \gg 7$
- Basic equation: $\mathbf{u}_{2i}^T \mathbf{F} \mathbf{u}_{1i} = 0$
- Goal is to find the singular matrix closest to **F**.

Eight-point method

Input: N point correspondences $\{\mathbf{u}_{1i} \leftrightarrow \mathbf{u}_{2i}\}, N \geq 8$
Output: fundamental matrix $\mathbf{F}$

Algoritmus: *Normalized 8-point method*

1. Data-normalization is separately carried out for the two point set:
   - translation
   - scale
2. Estimating $\hat{\mathbf{F}}'$ for normalized data
   (a) Linear solution by SVD $\longrightarrow \hat{\mathbf{F}}'$
   (b) Then singularity constraint $\det \hat{\mathbf{F}}' = 0$ is forced $\longrightarrow \hat{\mathbf{F}}'$
3. Denormalization
   - $\hat{\mathbf{F}}' \longrightarrow \mathbf{F}$

Data normalization and denormalization

- Goal of **data normalization**: numerical stability
  - **Obligatory step**: non-normalized method is not reliable.
  - Components of coefficient matrix should be in the same order of magnitude.
- Two point-sets are normalized by affine transformations $\mathbf{T}_1$ and $\mathbf{T}_2$.
  - Offset: origin is moved to the center(s) of gravity
  - Scale: average of point distances are scaled to be $\sqrt{2}$.
- **Denormalization:** correction by affine tranformations:

$$\hat{\mathbf{F}} = \mathbf{T}_2^T \hat{\mathbf{F}}' \mathbf{T}_1 \qquad (5)$$

Homogeneous linear system to estimate **F**

- For each point correspondence: $\mathbf{u}_2^T \mathbf{F} \mathbf{u}_1 = 0$, where $\mathbf{u}_k = [u_k, v_k, 1]^T, k = 1, 2$

$\rightarrow$ For element of the fundamental matrix, the following equation is valid:

$$u_2 u_1 f_{11} + u_2 v_1 f_{12} + u_2 f_{13} + v_2 u_1 f_{21} + v_2 v_1 f_{22} + v_2 f_{23} + u_1 f_{31} + v_1 f_{32} + f_{33} = 0$$

- If notation $\mathbf{f} = [f_{11}, f_{12}, \ldots, f_{33}]^T$ is introduced, the equation can be written as a dot product:

$$[u_2 u_1, u_2 v_1, u_2, v_2 u_1, v_2 v_1, v_2, u_1, v_1, 1]\mathbf{f} = 0$$

- For all i: $\{\mathbf{u}_{1i} \leftrightarrow \mathbf{u}_{2i}\}$

$$A\mathbf{f} \doteq \begin{bmatrix} u_{21} u_{11} & u_{21} v_{11} & u_{21} & v_{21} u_{11} & v_{21} v_{11} & v_{21} & u_{11} & v_{11} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{2N} u_{1N} & u_{2N} v_{1N} & u_{2N} & v_{2N} u_{1N} & v_{2N} v_{1N} & v_{2N} & u_{1N} & v_{1N} & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}$$

Sulution as homogeneous linear system of equations

- Estimation is similar to that of **homography**.
- Trivial solution $\mathbf{f} = \mathbf{0}$ has to be excluded.
  - vector $\mathbf{f}$ can be computed up to a scale
  - $\rightarrow$ vector norm is fixed as $\|\mathbf{f}\| = 1$
- If $\operatorname{rank} \mathbf{A} \leq 8$
  - $\operatorname{rank} \mathbf{A} = 8 \longrightarrow$ exact solution: nullvector
  - $\operatorname{rank} \mathbf{A} < 8 \longrightarrow$ solution is linear combination of nullvectors
- For noisy correspondences, $\operatorname{rank} \mathbf{A} = 9$.
  - optimal solution for algebraic error $\|\mathbf{Af}\|$
  - $\|\mathbf{f}\| = 1 \longrightarrow$ minimization of $\|\mathbf{Af}\| / \|\mathbf{f}\|$
  - $\rightarrow$ optimal solution is the eigenvector of $\mathbf{A}^T\mathbf{A}$ corresponding to the smallest eigenvalue
- Solution can also be obtained from SVD of $\mathbf{A}$:
  - $\mathbf{A} = \mathbf{UDV}^T \longrightarrow$ last column (vector) of $\mathbf{V}$.

Singular constraint

- If $\det \mathbf{F} \neq 0$
  - epipolar lines do not intersect each other in epipole.
  - $\rightarrow$ less accurate epipolar geometry $\longrightarrow$ less accurate reconstruction
- Solution of homogeneous linear system does not guarantee singularity: $\det \widehat{\mathbf{F}} \neq 0$.
- Task is to find matrix $\widehat{\mathbf{F}}'$, for which
  - Frobenius norm $\|\widehat{\mathbf{F}} - \widehat{\mathbf{F}}'\|$ is minimal, and
  - $\det \widehat{F}' = 0$
- SVD of $\mathbf{A}$: $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$
  - $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2, \delta_3)$ is the diagonal matrix containing singular values, and $\delta_1 \geq \delta_2 \geq \delta_3$
  - The estimation for closest matrix, fulfilling singularity constraint:

$$\widehat{F}' = \mathbf{U}\,\mathrm{diag}(\delta_1, \delta_2, 0)\mathbf{V}^{\mathrm{T}} \tag{6}$$

Epipoles from fudamental matrix **F**

- The epipoles are the null-vectors of **F** and $\mathbf{F}^T$: $\mathbf{F}\mathbf{e}_1 = \mathbf{0}$, and $\mathbf{F}^T\mathbf{e}_2 = \mathbf{0}$.
- Nullvector can be calculated by e.g. SVD.
- Singularity constraint guarantees that **F** has a null-vector
- Singular Value Decomposition: $\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, and then
    - $\mathbf{e}_1$: last column of **V**.
    - $\mathbf{e}_2$: last column of **U**.

Limits of eight-point method

- Similar to homography/projective matrix estimation
  - Significant difference: singularity constraint introduces

  $\rightarrow$ Similar benefits/weak points to homography/proj. matrix estimation
- Method is not robust
  - RANSAC-like robustification can be applied.
- There are another solution
  - Seven-point method: determinant constraint is forced to linear combination of null-spaces.

Non-linear methods to estimate F

- Algebraic error
  - It yields initial value(s) for numerical optimization.
- Geometric error
  - line-point distance

$$\epsilon = \frac{\mathbf{x'}^T \mathbf{F}\mathbf{x}}{|\mathbf{F}\mathbf{x}|_{1:2}}$$

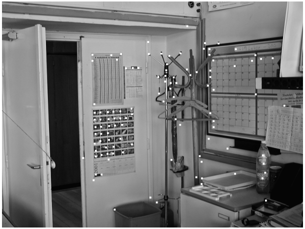
  - Symmetric version

$$\epsilon = \frac{\mathbf{x'}^T \mathbf{F}\mathbf{x}}{|\mathbf{F}\mathbf{x}|_{1:2}} + \frac{\mathbf{x}^T \mathbf{F}^T \mathbf{x'}}{|\mathbf{F}^T \mathbf{x'}|_{1:2}}$$

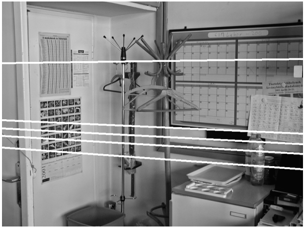
  - where operator $(\mathbf{x})_{1:2}$ denotes the first two coordinates of vector $\mathbf{x}$.
  - Geometric error minimized by numerical techniques.

Estimation of epipolar geometry: 1st example

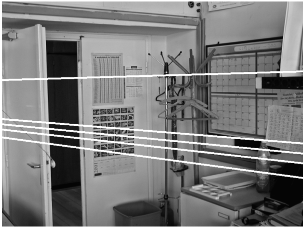
KLT feature points #1



KLT feature points #2

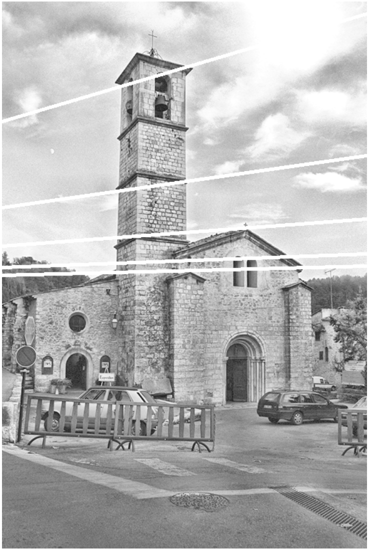


epipolar lines #1



epipolar lines #2

Estimation of epipolar geometry: 2nd example



Outline

Overview

Geometry of standard stereo

$$\frac{u_1}{f} = \frac{h - X}{Z}$$

$$-\frac{u_2}{f} = \frac{h + X}{Z}$$

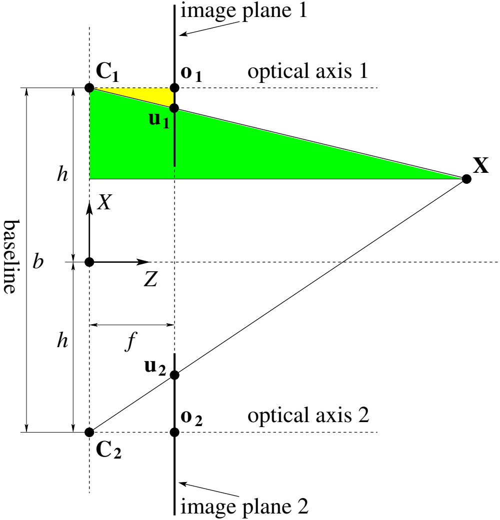
$$v_1 = v_2$$

$$Z = \frac{2hf}{u_1 - u_2} = \frac{bf}{d}$$

$$X = -\frac{b(u_1 + u_2)}{2d}$$

$$Y = \frac{bv_1}{d} = \frac{bv_2}{d}$$

$d \doteq u_1 - u_2$ disparity



Precision of depth estimation

- If $d \to 0$, and $Z \to \infty$
  - Disparity of distant points are small.
- Relation between disparity and precision of depth estimation

$$\frac{|\Delta Z|}{Z} = \frac{|\Delta d|}{|d|}$$

  - larger the disparity, smaller the relative depth error

  $\rightarrow$ precision is increasing
- Influence of base length

$$d = \frac{bf}{Z}$$

  - For larger b, same depth value yields larger disparity

  $\rightarrow$ Precision of depth estimation increasing

  $\rightarrow$ more pixels $\longrightarrow$ precision of diparity increasing

Overview

Goals of rectification

- Input of rectification: non-standard stereo image pair
- Goal of rectification: make stereo matching more accurate
  - After rectification, corresponding pixels are located in the same row

  $\rightarrow$ standard stereo, 1D search
- Rectification based on epipolar geometry
  - Images are transformed based on epipolar geometry

  $\rightarrow$ after transformation, corresponding epipolar lines are placed on the same rows

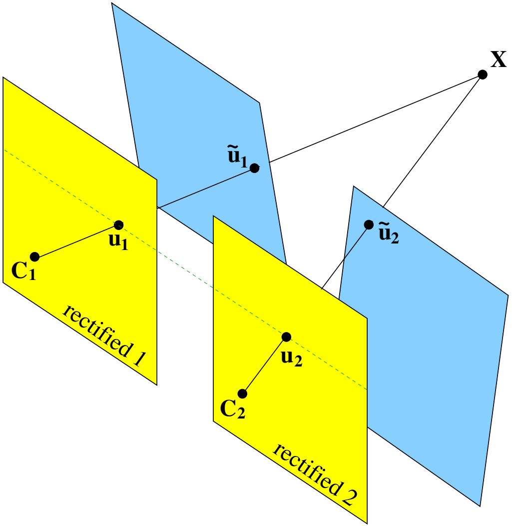
  $\rightarrow$ epipoles are in the infinity
- For rectification, only the fundamental matrix has to be known

  $\rightarrow$ Fundamental matrix represents epipolar geometry

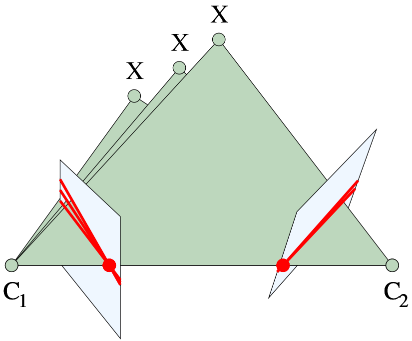
Rectification methods

- Only the general principles are discussed here.
  - Rectification is a complex method.
  - Rectification **is not required**, it has both advantages and disadvantages.
- Rectification can be carried out by homographies.
- It has ambiguity: there are infinite number of rectification transformations for the same image pair.
- The aim is to find a 2D projective transformation that
  - fulfills the requirement for rectification and
  - distorts minimally the images.
- Knowledge of camera intrinsic parameters helps the rectification.

Geometry of rectification

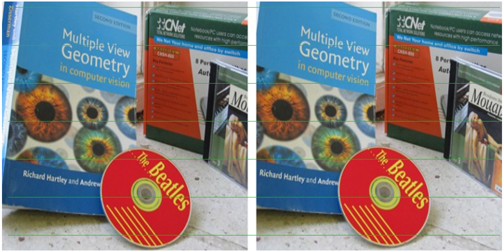


Rectification: a video video

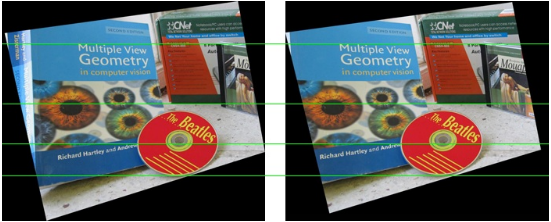


Epipoles transformed to infinity

Rectification: an example



before



after

Benefits of rectifications

- Modify the inage in order to get a standard stereo,
  $\rightarrow$ then algorithms for standard stereo can be applyied.
- The properties of epipolar geometry can be visualized by rectifying the images.
- For practical purposes, the rectification has to be very accurate
  - otherwise there will be a shift between corresponding rows.

  $\rightarrow$ feature matching more challenging, 1D cannot be run.

Weak points of rectification

- Distortion under rectification hardly depends on baseline width.
- For wide-baseline stereo:
  - Rectification significantly destorts the image.
  - $\rightarrow$ Pixel-based method can be applied for feature matching
  - $\rightarrow$ Correspondence-based methods often fail.
- Size and shape of rectified images differ from original ones.
  - $\rightarrow$ Feature matching is more challenging.

$\rightarrow$ Many experts do not agree that rectification is necessary.
  - Epipolar lines can be followed if fundamental matrix is given.
  - Matching can be carried out in original frames.
  - $\rightarrow$ Then noise is not distorted by rectifying transformation.

Outline

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

Types of stereo reconstruction

- **Fully calibrated** reconstruction
  - Known **intrinsic and extrinsic** camera parameters
  - reconstruction by triangulation
  - known baseline $\longrightarrow$ known scale
- **Metric (Euclidean)** reconstruction
  - knonw **intrinsic** camera parameters, $n \geq 8$ point correspondences given
  - Extrinsic camera parameters obtained from **essential matrix**
  - Reconstruction up to a similarity transformation

  $\rightarrow$ up to a scale
- **Projective reconstruction**
  - **unknown** camera parameters, $n \geq 8$ point correspondences are given
  - Composition of projective matrices from a **fundamental matrix**
  - reconstruction can be computed up to a projective transformation

Overview

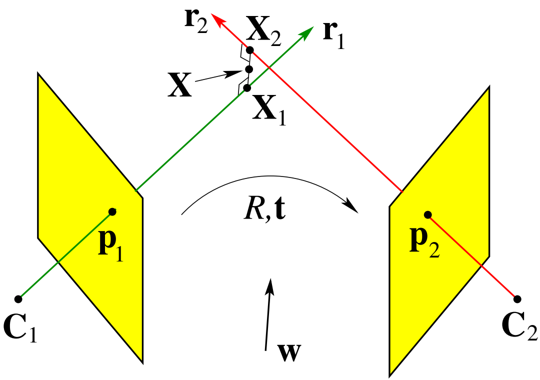
Triangulation

- **Task:**
  - Two calibrated cameras are given, including both intrinsic and extrinsic parameters, and
  - Locations $\mathbf{u}_1, \mathbf{u}_2$ of the projection of spatial point $\mathbf{X}$ are given
  - Goal is to estimate spatial location $\mathbf{X}$.
- Two calibration matrices are known, therefore
  - for a projection matrix: $\mathbf{K}^{-1}\mathbf{P} = [\mathbf{R}| - \mathbf{t}]$ and
  - for calibrated (aka. normalized) coordinates: $\mathbf{p} = \mathbf{K}^{-1}\mathbf{u}$.
- For the sake of simplicity, the first camera gives the world coordinate system
  - **non-homogeneous** coordinates are used

  $\rightarrow$ $\mathbf{p}_2 = \mathbf{R}(\mathbf{p}_1 - \mathbf{t}), \mathbf{p}_1 = \mathbf{t} + \mathbf{R}^T\mathbf{p}_2$
- Image points are bask-projected to 3D space
  - two rays obtained, they usually do not intersect each other due to noise/calibration error

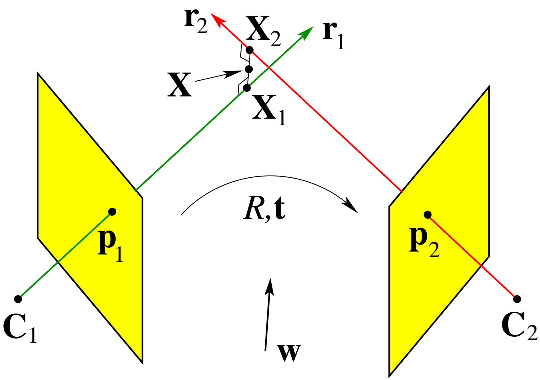
  $\rightarrow$ task is to give an estimate for spatial point $\mathbf{X}$.

Linear triangulation: geometry



- Line $\mathbf{X}_1\mathbf{X}_2$ perpendicular to both $\mathbf{r}_1$ and $\mathbf{r}_2$.
- Estimate $\mathbf{X}$ is the middle point of section $\mathbf{X}_1\mathbf{X}_2$
- Vector $\mathbf{w}$ is parallel to $\mathbf{X}_1\mathbf{X}_2$.

Linear triangulation: notations



- $\alpha \mathbf{p}_1$ is a point on ray $\mathbf{r}_1$ ($\alpha \in \Re$)
- $\mathbf{t} + \beta R^T \mathbf{p}_2$ a point on other ray $\mathbf{r}_2$ ($\beta \in \Re$)
  - $\rightarrow$ coordinate system fixed to the first camera
- Let $\mathbf{X}_1 = \alpha_0 \mathbf{p}_1$, $\mathbf{X}_2 = \mathbf{t} + \mathbf{R}^T(\beta_0 \mathbf{p}_2 - \mathbf{t})$

Linear triangulation: solution

- Task is to determine
  - the middle point of the line section $\mathbf{X}_1\mathbf{X}_2$
  - $\rightarrow$ determination of α_0 and β_0 required
- Remark that
  - Vector $\mathbf{w} = \mathbf{p}_1 \times \mathbf{R}^T(\mathbf{p}_2 - \mathbf{t})$ perpendicular to both $\mathbf{r}_1$ and $\mathbf{r}_2$.
  - Line $\alpha\mathbf{p}_1 + \gamma\mathbf{w}$ parallel to $\mathbf{w}$ and contain the point $\alpha\mathbf{p}_1$ ($\gamma \in \Re$).

$\rightarrow$ α_0, β_0 (as well as γ_0) are given by the solution of the following linear system: :

$$\alpha\mathbf{p}_1 + \mathbf{t} + \beta\mathbf{R}^T(\mathbf{p}_2 - \mathbf{t}) + \gamma[\mathbf{p}_1 \times \mathbf{R}^T(\mathbf{p}_2 - \mathbf{t})] = 0 \quad (7)$$

- Triangulated point is obtained, e.g by $\alpha_0\mathbf{p}_1$
- There is no solution if $\mathbf{r}_1$ and $\mathbf{r}_2$ are parallel

Linear triangulation: an algebraic solution

- Two projected locations of spatial point $\mathbf{X}$ are given:

$$\lambda_1 \mathbf{u}_1 = \mathbf{P}_1 \mathbf{X}$$
$$\lambda_2 \mathbf{u}_2 = \mathbf{P}_2 \mathbf{X}$$

- λ_1 and λ_2 can be eliminated. 2 + 2 equations are obtained:

$$u\mathbf{p}_3^T \mathbf{X} = \mathbf{p}_1^T \mathbf{X}$$
$$v\mathbf{p}_3^T \mathbf{X} = \mathbf{p}_2^T \mathbf{X}$$

- where $\mathbf{p}_i^T$ is the i-th row of projection matrix $\mathbf{P}$.
- Both projections yield 2 equations. Only vector $\mathbf{X}$ is unknown.
- Solution for $\mathbf{X}$ is calculated by solving the homogeneous linear system of equations.
- Important remark: solution is obtained in homogeneous coordinates.

Refinement by minimizing the reprojection error

- Linear algorithm yield points $\mathbf{X}_i$, $i = 1, 2, \ldots, n$ if n point pairs are given
- The solution should be refined
  - minimization of **reprojection error** yields **more accurate** estimate
- For minimizing the reprojection error, the following parameters have to be refined:
  - Spatial points $\mathbf{X}_i$
  - Rotation matrix $\mathbf{R}$ and baseline vector $\mathbf{t}$
  - $\rightarrow$ intrinsic camera parameters are usually fixed as cameras are pre-calibrated
- Initial values for numerical optimization
  - Spatial points $\mathbf{X}_i$ from linear triangulation
  - Initial rotation matrix $\mathbf{R}$ and baseline vector $\mathbf{t}$ by decomposing the essential matrix

Metric reconstruction by decomposing the essential matrix

- Intrinsic camera matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ given, fundamental matrix computed from $n \geq 8$ point correpondences
  - $\mathbf{E}$ can be retrieved from $\mathbf{F}$, $\mathbf{K}_1$ and $\mathbf{K}_2$.
  - from $\mathbf{E}$, extrinsic parameters can be obtained by decomposition
- Unknown baseline $\longrightarrow$ unknown scale
  - baseline normalized to 1
  - $\rightarrow$ Euclidean reconstruction possible up to a similarity transformation
- It is assumed that world coordinate is fixed to the first camera
  - $\rightarrow$ Therefore, $P_1 = [\mathrm{I}|\mathbf{0}]$, where I is the identity matrix
- Position of second camera computed from essential matrix $\mathbf{E}$ by SVD.
  - Four solutions obtained,
  - only one is correct.

Camera pose estimation by SVD

- The Singular Value Decompoisition of $\mathbf{E}$ is $\mathbf{E} = \mathbf{UDV}^T$, where $\mathbf{D} = \mathrm{diag}(\delta, \delta, 0)$
  $\rightarrow$ $\mathbf{E}$ has two equal singuar values
- Four solutions can be obtained as follows:

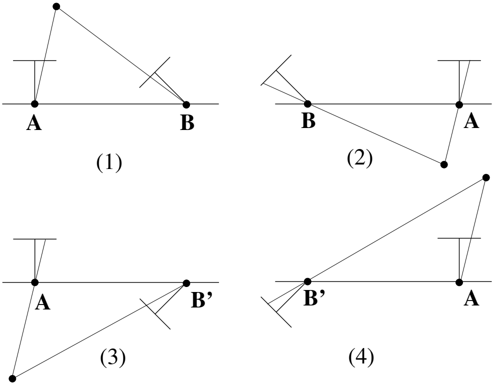
$$\mathbf{R}_1 = \mathbf{UWV}^T \qquad \mathbf{R}_2 = \mathbf{UW}^T\mathbf{V}^T$$
$$[\mathbf{t}_1]_\times = \delta \mathbf{UZU}^T \qquad [\mathbf{t}_2]_\times = -\delta \mathbf{UZU}^T$$

- where

$$\mathbf{W} \doteq \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{Z} \doteq \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Combination of 2-2 candidates for translation and rotation yield 4 solutions.
- Determinants of $\mathbf{R}_1$ and $\mathbf{R}_2$ have to be positive, otherwise matrices should be multiplied by -1.

Visualization of the four solutions



- Left and right: camera locations replaces
- Top and bottom: mirror to base lane
- 3D point is in front of the cameras only in the top-left case.

Overview

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

Projective reconstruction based on fundamental matrix

- Unknown intrinsic parameters, $n \geq 8$ known point correspondences
- Reconstruction can be obtained up to a projective transformation.
  - If $\mathbf{H}$ is a 4×4 projective transformation, then $\mathbf{P}_k\mathbf{X} = (\mathbf{P}_k\mathbf{H})(\mathbf{H}^{-1}\mathbf{X})$, $k = 1, 2$

  $\rightarrow$ if $\mathbf{u}_1 \leftrightarrow \mathbf{u}_2$ are projections of $\mathbf{X}$ by P_k, then $\mathbf{u}_1 \leftrightarrow \mathbf{u}_2$ are those of $\mathbf{H}^{-1}\mathbf{X}$ by $\mathbf{P}_k\mathbf{H}$.

  $\rightarrow$ From fundamental matrix $\mathbf{F}$, matrices $\mathbf{P}_k$ can be computed up to the transformation $\mathbf{H}$
- There is a matrix $\mathbf{H}$ to get the canonical form for $\mathbf{P}_1$ as
  - $\mathbf{P}_1 = [\mathrm{I}|\mathbf{0}]$

Summary of calibrated and uncalibrated 3D vision

	calibrated case	uncalibrated case
epipolar constraint	$\mathbf{u}_2^T K_2^{-T} E K_1^{-1} \mathbf{u}_1 = 0$	$\mathbf{u}_2^T F \mathbf{u}_1 = 0$
fundamental matrix	$E = [\mathbf{t}]_\times R$	$F = K_2^{-T} E K_1^{-1}$
epipoles	$E K_1^{-1} \mathbf{e}_1 = \mathbf{0}$	$F \mathbf{e}_1 = \mathbf{0}$
	$\mathbf{e}_2^T K_2^{-T} E^T = \mathbf{0}^T$	$\mathbf{e}_2 F^T = \mathbf{0}$
epipolar lines	$\mathbf{l}_1 = K_1^{-T} E^T K_2^{-1} \mathbf{u}_2$	$\mathbf{l}_1 = F^T \mathbf{u}_2$
	$\mathbf{l}_2 = K_2^{-T} E K_1^{-1} \mathbf{u}_1$	$\mathbf{l}_2 = F \mathbf{u}_1$
reconstruction	metric: $\mathbf{X}_m$	projective: $\mathbf{X}_p = H \mathbf{X}_m$

Correction of projective reconstruction

- Metric reconstruction is the subset of projective reconstruction
  - How can projective tranformation **H** be computed?
  - What kind of knowledge is required for correction?
- (*Direct*) method
  - 3D locations of five points must be known.

  $\rightarrow$ **H** can be estimated: $\mathbf{X}_m = \mathbf{H}^{-1}\mathbf{X}_p$
- (*Stratified*) method
  - Parallel and perpendicular lines
  - Projective $\longrightarrow$ affine $\longrightarrow$ metric

  $\rightarrow$ For an affine reconstruction, **H** is an affinity

Data for correction of projective reconstruction: a video

- Parallel and
- perpendicular lines

Overview

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

Planar motion

- A vehicle moves on a planar road.
- It can be rotated and translated.
- Coordinate system fixed to the car, axis Z parallel to the road.
- Two frames of the video yields a stereo problem.
- Vehicle is rotated, due to steering, around axis Y by angle β.
- Translation is in plane XZ: its direction represented by angle α.

$$\mathbf{t} = \begin{bmatrix} t_x \\ 0 \\ t_z \end{bmatrix} = \rho \begin{bmatrix} \cos\alpha \\ 0 \\ \sin\alpha \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

Planar motion: essential matrix

- Furthermore

$$\mathbf{t} = \rho \begin{bmatrix} \cos\alpha \\ 0 \\ \sin\alpha \end{bmatrix} \quad \rightarrow \quad [\mathbf{t}]_X = \rho \begin{bmatrix} 0 & -\sin\alpha & 0 \\ \sin\alpha & 0 & -\cos\alpha \\ 0 & \cos\alpha & 0 \end{bmatrix}$$

- Then the essential matrix is as follows:

$$\mathbf{E} = [\mathbf{t}]_X \mathbf{R} \sim \begin{bmatrix} 0 & -\sin\alpha & 0 \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & 0 & \sin\alpha\sin\beta - \cos\alpha\cos\beta \\ 0 & \cos\alpha & 0 \end{bmatrix}$$

Planar motion: essential and fundamental matrices

- After applying trigonometric equalities:

$$\mathbf{E} \sim \begin{bmatrix} 0 & -\sin\alpha & 0 \\ \sin(\alpha+\beta) & 0 & -\cos(\alpha+\beta) \\ 0 & \cos\alpha & 0 \end{bmatrix}$$

- If camera intrinsic matrices are the same for the images, and the common matrix is a so-called semi-calibrated one: $\mathbf{K} = diag(f, f, 1)$, then

$$\mathbf{F} = \mathbf{K}^{-T}\mathbf{E}\mathbf{K}^{-1} \sim \begin{bmatrix} 0 & -\frac{\sin\alpha}{f^2} & 0 \\ \frac{\sin(\alpha+\beta)}{f^2} & 0 & -\frac{\cos(\alpha+\beta)}{f} \\ 0 & \frac{\cos\alpha}{f} & 0 \end{bmatrix}$$

Planar motion: estimation

- Only four out of nine elements in fundamental/essential matrices are nonzero.
  - Essental matrix can be estimated by two point correspondences.
  - Semi-calibrated camera: three correspondences.
- Robustification, e.g. by RANSAC, is fast
- Equation from one correspondence $\mathbf{p}_1 = [u_1, v_1]$, $\mathbf{p}_2 = [u_2, v_2]$ for two angles α and β (calibrated case):

$$\left\langle [v_1, -u_2 v_1, -v_2, v_2 u_1]^T, [\cos\alpha, \sin\alpha, \cos(\alpha+\beta), \sin(\alpha+\beta)]^T \right\rangle = 0$$

- For multiple correspondences, solution can be written as

$$\mathbf{A}_1 \mathbf{v}_1 + \mathbf{A}_2 \mathbf{v}_2 = 0$$

- where $\mathbf{v}_1 = [\cos\alpha, \sin\alpha]^T$ and $\mathbf{v}_2 = [\cos(\alpha+\beta), \sin(\alpha+\beta)]^T$

Planar motion: estimation

- Thus, $\mathbf{v}_1^T\mathbf{v}_1 = \mathbf{v}_2^T\mathbf{v}_2 = 1$.
- Furthermore,

$$\mathbf{A}_1\mathbf{v}_1 + \mathbf{A}_2\mathbf{v}_2 = 0 \quad (8)$$

$$\mathbf{A}_1\mathbf{v}_1 = -\mathbf{A}_2\mathbf{v}_2 \quad (9)$$

$$\mathbf{v}_1 = -\mathbf{A}_1^{\dagger}\mathbf{A}_2\mathbf{v}_2 \quad (10)$$

$$\mathbf{v}_1^T\mathbf{v}_1 = \mathbf{v}_2^T\left(\mathbf{A}_1^{\dagger}\mathbf{A}_2\right)^T\left(\mathbf{A}_1^{\dagger}\mathbf{A}_2\right)\mathbf{v}_2 = 1 \quad (11)$$

$$\mathbf{v}_2^T\mathbf{B}\mathbf{v}_2 = 1 \quad (12)$$

- If $\mathbf{B} = \left(\mathbf{A}_1^{\dagger}\mathbf{A}_2\right)^T\left(\mathbf{A}_1^{\dagger}\mathbf{A}_2\right)$
- Thus, $\mathbf{v}_2$ is given by the intersection of an ellipse and the unit-radius circle as $\mathbf{v}_2\mathbf{B}\mathbf{v}_2 = \mathbf{v}_2^T\mathbf{v}_2 = 1$.

Planar motion: estimation

- Solution is given by Singular Value Decomposition: $\mathbf{B} = \mathbf{U}^T\mathbf{S}\mathbf{U}$.
- Let $\mathbf{r} = [r_x \quad r_y]^T = \mathbf{U}\mathbf{v}_2$.

$$\mathbf{v}_2^T\mathbf{B}\mathbf{v}_2 = 1 \quad (13)$$

$$\mathbf{v}_2^T\mathbf{U}^T\mathbf{S}\mathbf{U}\mathbf{v}_2 = 1 \quad (14)$$

$$\mathbf{r}_2^T\mathbf{S}\mathbf{r}_2 = 1 \quad (15)$$

$$\mathbf{r}_2^T \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \mathbf{r}_2 = 1 \quad (16)$$

- Therefore, $s_1 r_x^2 + s_2 r_y^2 = 1$
- and $r_x^2 + r_y^2 = 1$

→ Linear system for r_x^2 and r_y^2. (Four candidate solutions, similarly to general stereo vision.)

- $\mathbf{v}_2 = \mathbf{U}^T\mathbf{r}$ and $\mathbf{v}_1 = -\mathbf{A}_1^\dagger\mathbf{A}_2\mathbf{v}_2$ gives final solution.

Outline

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

Summary

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

References

- R.Hartley, A.Zisserman: "Multiple View Geometry in Computer Vision", Cambridge University Press
- M.Sonka, V.Hlavac, R.Boyle: "Image Processing, Analysis and Machine Vision", Thomson
- Y. Ma, S. Soatto, J. Kosecka, S. Shankar Sastry: "An Invitation to 3-D Vision", Springer
- D.A. Forsyth, J. Ponce: "Computer Vision: a modern approach", Prentice Hall
- E. Trucco, A. Verri: "Introductory Techniques for 3-D Computer Vision", Prentice Hall
- Kató Zoltán, Czúni László: "Számítógépes látás"
  - `tananyagfejlesztes.mik.uni-pannon.hu/`