3D Computer Vision

Radim Šára   Martin Matoušek

Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

https://cw.fel.cvut.cz/wiki/courses/tdv/start

http://cmp.felk.cvut.cz
mailto:sara@cmp.felk.cvut.cz
phone ext. 7203

rev. December 5, 2023

Open Informatics Master's Course

Module VII

Stereovision

mostly covered by

Šára, R. How To Teach Stereoscopic Vision. Proc. ELMAR 2010 referenced as [SP]

additional references

C. Geyer and K. Daniilidis. Conformal rectification of omnidirectional stereo pairs. In *Proc Computer Vision and Pattern Recognition Workshop*, p. 73, 2003.

J. Gluckman and S. K. Nayar. Rectifying transformations that minimize resampling effects. In *Proc IEEE CS Conf on Computer Vision and Pattern Recognition*, vol. 1:111–117. 2001.

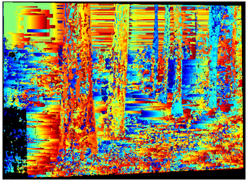
M. Pollefeys, R. Koch, and L. V. Gool. A simple and efficient rectification method for general motion. In *Proc Int Conf on Computer Vision*, vol. 1:496–501, 1999.

Stereovision = Getting Relative Distances Per Pixel given the Epipolar Geometry

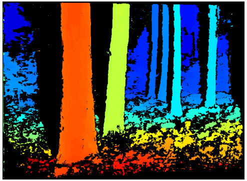
The success of a model-free stereo matching algorithm is unlikely:

WTA Matching:

For every left-image pixel find the most similar right-image pixel along the corresponding epipolar line. [Marroquin 83]



disparity map from WTA



a good disparity map

- monocular vision already gives a rough 3D sketch because we understand the scene
- pixelwise independent matching without any problem understanding is difficult
- matching can benefit from a geometric simplification of the problem: epipolar rectification

▶Linear Epipolar Rectification for Easier Correspondence Search

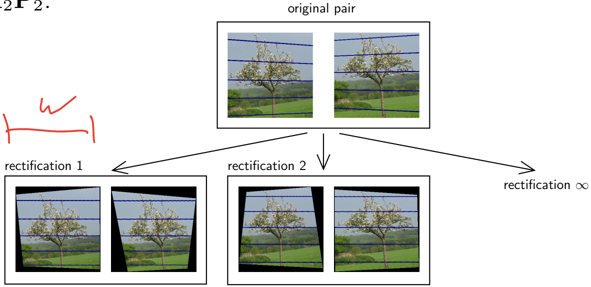
Obs:

- epipoles and epipolars are elements of $\mathbb{P}^2$, they may be mapped by homographies
- if we map epipoles to infinity, epipolars become parallel
- we then rotate them to become horizontal
- we then scale the images to make corresponding epipolars colinear
- this can be achieved by a pair of (non-unique) homographies applied to the images

Problem: Given fundamental matrix $\mathbf{F}$ or camera matrices $\mathbf{P}_1$, $\mathbf{P}_2$, compute a pair of homographies that maps epipolars to horizontal lines with the same row coordinate.

Procedure:

1. find a pair of rectification homographies $\mathbf{H}_1$ and $\mathbf{H}_2$.
2. warp images using $\mathbf{H}_1$ and $\mathbf{H}_2$ and transform the fundamental matrix $\mathbf{F} \mapsto \mathbf{H}_2^{-\top}\mathbf{F}\mathbf{H}_1^{-1}$ or the cameras $\mathbf{P}_1 \mapsto \mathbf{H}_1\mathbf{P}_1$, $\mathbf{P}_2 \mapsto \mathbf{H}_2\mathbf{P}_2$.



$O(n^3)$

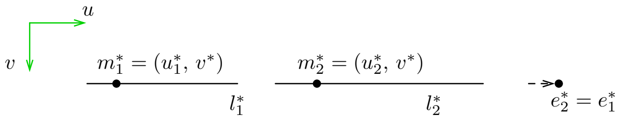
$O(n^2 \log n)$

▸Rectification Homographies

Assumption: Cameras $(\mathbf{P}_1, \mathbf{P}_2)$ are rectified by a homography pair $(\mathbf{H}_1, \mathbf{H}_2)$:

$$\mathbf{P}_i^* \simeq \mathbf{H}_i \mathbf{P}_i = \begin{bmatrix} \mathbf{Q}_i & \mathbf{q}_i \end{bmatrix} = \mathbf{H}_i \mathbf{K}_i \mathbf{R}_i \begin{bmatrix} \mathbf{I} & -\mathbf{C}_i \end{bmatrix}, \quad i = 1, 2$$

rectified entities: $\mathbf{F}^*$, l_1^*, l_2^*, etc:



- the rectified location difference $d = u_1^* - u_2^*$ is called disparity

corresponding epipolar lines must be:

1. parallel to image rows $\Rightarrow$ epipoles become $e_1^* = e_2^* = (1, 0, 0)$
2. equivalent $l_2^* = l_1^*$: $\quad \mathbf{l}_1^* \simeq \mathbf{e}_1^* \times \underline{\mathbf{m}}_1 = [\mathbf{e}_1^*]_\times \underline{\mathbf{m}}_1 \simeq \mathbf{l}_2^* \simeq \mathbf{F}^* \underline{\mathbf{m}}_1 \quad \Rightarrow \quad \mathbf{F}^* = [\mathbf{e}_1^*]_\times$

- therefore the canonical fundamental matrix is

$$\mathbf{F}^* \simeq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

A two-step rectification procedure

1. find some pair of primitive rectification homographies $\hat{\mathbf{H}}_1$, $\hat{\mathbf{H}}_2$
2. upgrade to a pair of optimal rectification homographies while preserving $\mathbf{F}^*$

▶Geometric Interpretation of Linear Rectification

What pair of physical cameras is compatible with $\mathbf{F}^*$?

- we know that $\mathbf{F} = (\mathbf{Q}_1 \mathbf{Q}_2^{-1})^\top [\underline{\mathbf{e}}_1]_\times$ →80
- we choose $\mathbf{Q}_1^* = \mathbf{K}_1^*$, $\mathbf{Q}_2^* = \mathbf{K}_2^* \mathbf{R}^*$; then

$$\mathbf{F}^* \simeq (\mathbf{Q}_1^* \mathbf{Q}_2^{*-1})^\top [\underline{\mathbf{e}}_1^*]_\times \overset{!}{\simeq} (\mathbf{K}_1^* \mathbf{R}^{*\top} \mathbf{K}_2^{*-1})^\top \mathbf{F}^*$$

- we look for $\mathbf{R}^*$, $\mathbf{K}_1^*$, $\mathbf{K}_2^*$ compatible with equations

$$(\mathbf{K}_1^* \mathbf{R}^{*\top} \mathbf{K}_2^{*-1})^\top \mathbf{F}^* = \lambda \mathbf{F}^*, \qquad \mathbf{R}^* \mathbf{R}^{*\top} = \mathbf{I}, \qquad \mathbf{K}_1^*, \mathbf{K}_2^* \text{ upper triangular}$$

- we also want $\mathbf{b}^*$ from $\underline{\mathbf{e}}_1^* \simeq \mathbf{P}_1^* \underline{\mathbf{C}}_2^* = \mathbf{K}_1^* \mathbf{b}^*$ — $\mathbf{b}^*$ in camera-1 frame
- result after equations reduction:

$$\mathbf{R}^* = \mathbf{I}, \quad \mathbf{b}^* = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{K}_1^* = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_2^* = \begin{bmatrix} k_{21} & k_{22} & k_{23} \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{36}$$

- rectified cameras are in canonical relative pose — not rotated, canonical baseline
- rectified calibration matrices can differ in the first row only
- if $\mathbf{K}_1^* = \mathbf{K}_2^*$, the rectified pair is called the standard stereo pair and we have the standard rectification homographies
- standard rectification homographies: points at infinity have zero disparity

$$\mathbf{P}_i^* \underline{\mathbf{X}}_\infty = \mathbf{K} \begin{bmatrix} \mathbf{I} & -\mathbf{C}_i \end{bmatrix} \underline{\mathbf{X}}_\infty = \mathbf{K} \mathbf{X}_\infty \qquad i = 1, 2$$

- this does not mean that the images are not distorted after rectification

▶Primitive Rectification

Goal: Given fundamental matrix $\mathbf{F}$, derive some easy-to-obtain rectification homographies $\mathbf{H}_1$, $\mathbf{H}_2$

1. Let the SVD of $\mathbf{F}$ be $\mathbf{U}\mathbf{D}\mathbf{V}^\top = \mathbf{F}$, where $\mathbf{D} = \mathrm{diag}(1, d^2, 0)$, $\quad 1 \geq d^2 > 0$
2. Write $\mathbf{D}$ as $\mathbf{D} = \mathbf{A}^\top \mathbf{F}^* \mathbf{B}$ for some regular $\mathbf{A}$, $\mathbf{B}$. For instance ($\mathbf{F}^*$ is given →160)

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -d & 0 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & d & 0 \end{bmatrix}$$

3. Then

$$\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^\top = \underbrace{\mathbf{U}\mathbf{A}^\top}_{\hat{\mathbf{H}}_2^\top} \mathbf{F}^* \underbrace{\mathbf{B}\mathbf{V}^\top}_{\hat{\mathbf{H}}_1} = \hat{\mathbf{H}}_2^\top \mathbf{F}^* \hat{\mathbf{H}}_1 \qquad \hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2 \text{ orthogonal}$$

and the primitive rectification homographies are

$$\hat{\mathbf{H}}_2 = \mathbf{A}\mathbf{U}^\top, \qquad \hat{\mathbf{H}}_1 = \mathbf{B}\mathbf{V}^\top$$

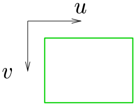
⊛ P1; 1pt: derive some other admissible $\mathbf{A}$, $\mathbf{B}$

- **Hence:** Rectification homographies do exist →160
- there are other primitive rectification homographies, these suggested are just easy to obtain

▸The Set of All Rectification Homographies

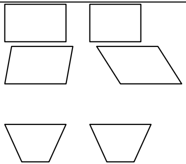
Proposition 1 Homographies $\mathbf{A}_1$ and $\mathbf{A}_2$ are rectification-preserving if the images stay rectified, i.e. if $\mathbf{A}_2^{-\top}\,\mathbf{F}^*\,\mathbf{A}_1^{-1} \simeq \mathbf{F}^*$, which gives

$$\mathbf{A}_1 \cong \begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & s_v & t_v \\ 0 & q & 1 \end{bmatrix}, \qquad \mathbf{A}_2 \cong \begin{bmatrix} r_1 & r_2 & r_3 \\ 0 & s_v & t_v \\ 0 & q & 1 \end{bmatrix}, \tag{37}$$



where $s_v \neq 0,\ t_v,\ l_1 \neq 0,\ l_2,\ l_3,\ r_1 \neq 0,\ r_2,\ r_3,\ q$ are 9 free parameters.

general	transformation	standard
$l_1,\ r_1$	horizontal scales	$l_1 = r_1$
$l_2,\ r_2$	horizontal shears	$l_2 = r_2$
$l_3,\ r_3$	horizontal shifts	$l_3 = r_3$
q	common special projective	
s_v	common vertical scale	
t_v	common vertical shift	
9 DoF		$9 - 3 = 6$ DoF



- q is due to a rotation about the baseline
- s_v changes the focal length

proof: find a rotation $\mathbf{G}$ that brings $\mathbf{K}$ to upper triangular form via RQ decomposition: $\mathbf{A}_1\mathbf{K}_1^* = \hat{\mathbf{K}}_1\mathbf{G}$ and $\mathbf{A}_2\mathbf{K}_2^* = \hat{\mathbf{K}}_2\mathbf{G}$

Corollary for Proposition 1 Let $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$ be (primitive or other) rectification homographies. Then $\mathbf{H}_1 = \mathbf{A}_1\bar{\mathbf{H}}_1, \quad \mathbf{H}_2 = \mathbf{A}_2\bar{\mathbf{H}}_2$ are also rectification homographies, where $\mathbf{A}_1$, $\mathbf{A}_2$ are as in (37).

Proposition 2 Pairs of rectification-preserving homographies $(\mathbf{A}_1, \mathbf{A}_2)$ form a group, with group operation (composition) $\quad (\mathbf{A}_1', \mathbf{A}_2') \circ (\mathbf{A}_1, \mathbf{A}_2) = (\mathbf{A}_1' \mathbf{A}_1, \mathbf{A}_2' \mathbf{A}_2)$.

Proof:

- closure by Proposition 1
- associativity by matrix multiplication
- identity belongs to the set
- inverse element belongs to the set by $\mathbf{A}_2^\top \mathbf{F}^* \mathbf{A}_1 \simeq \mathbf{F}^* \Leftrightarrow \mathbf{F}^* \simeq \mathbf{A}_2^{-\top} \mathbf{F}^* \mathbf{A}_1^{-1}$

▶Primitive Rectification Suffices for Calibrated Cameras

Obs: calibrated cameras: $d = 1 \Rightarrow \hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2$ (→162) are orthonormal ← SVD

E, K1, K2 given

1. determine primitive rectification homographies $(\hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2)$ from the essential matrix
2. choose a suitable common calibration matrix $\mathbf{K}$, e.g. from $\mathbf{K}_1$, $\mathbf{K}_2$:

$$\mathbf{K} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad f = \frac{1}{2}(f^1 + f^2), \quad u_0 = \frac{1}{2}(u_0^1 + u_0^2), \quad \text{etc.}$$

3. the final rectification homographies applied as $\mathbf{P}_i \mapsto \mathbf{H}_i \mathbf{P}_i$ are

$$\mathbf{H}_1 = \mathbf{K}\hat{\mathbf{H}}_1\mathbf{K}_1^{-1}, \quad \mathbf{H}_2 = \mathbf{K}\hat{\mathbf{H}}_2\mathbf{K}_2^{-1}$$

- we got a standard stereo pair (→161) and non-negative disparity:

let $\mathbf{K}_i^{-1}\mathbf{P}_i = \mathbf{R}_i \begin{bmatrix}\mathbf{I} & -\mathbf{C}_i\end{bmatrix}, \quad i = 1, 2$ note we started from $\mathbf{E}$, not $\mathbf{F}$

$$\mathbf{H}_1\mathbf{P}_1 = \mathbf{K}\hat{\mathbf{H}}_1\mathbf{K}_1^{-1}\mathbf{P}_1 = \mathbf{K}\underbrace{\mathbf{B}\mathbf{V}^\top\mathbf{R}_1}_{\mathbf{R}^*}\begin{bmatrix}\mathbf{I} & -\mathbf{C}_1\end{bmatrix} = \mathbf{K}\mathbf{R}^*\begin{bmatrix}\mathbf{I} & -\mathbf{C}_1\end{bmatrix}$$

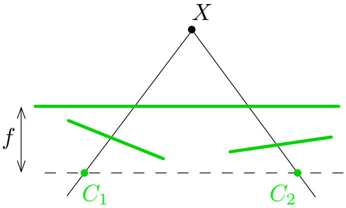
$\mathbf{A}$, $\mathbf{B}$ from →162

$$\mathbf{H}_2\mathbf{P}_2 = \mathbf{K}\hat{\mathbf{H}}_2\mathbf{K}_2^{-1}\mathbf{P}_2 = \mathbf{K}\underbrace{\mathbf{A}\mathbf{U}^\top\mathbf{R}_2}_{\mathbf{R}^*}\begin{bmatrix}\mathbf{I} & -\mathbf{C}_2\end{bmatrix} = \mathbf{K}\mathbf{R}^*\begin{bmatrix}\mathbf{I} & -\mathbf{C}_2\end{bmatrix}$$

Pi X∞

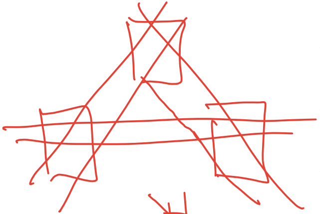
- one can prove that $\mathbf{B}\mathbf{V}^\top\mathbf{R}_1 = \mathbf{A}\mathbf{U}^\top\mathbf{R}_2$ with the help of essential matrix decomposition (15)
- Note that points at infinity project by $\mathbf{K}\mathbf{R}^*$ in both cameras ⇒ they have zero disparity (→168), hence...

►Summary & Remarks: Linear Rectification



. . . It follows: Standard rectification homographies reproject onto a common image plane parallel to the baseline

- rectification is done with a pair of homographies (one per image) →159
  - ⇒ projection centers of rectified cameras are equal to the original ones
  - binocular rectification: a 9-parameter family of rectification homographies
  - trinocular rectification: has 9 or 6 free parameters (depending on additional constrains)
  - in general, linear rectification is not possible for more than three cameras
- rectified cameras are in canonical orientation →161
  - ⇒ rectified image projection planes are coplanar
- equal rectified calibration matrices give standard rectification →161
  - ⇒ rectified image projection planes are equal
- primitive rectification is already standard in calibrated cameras →165



- known $\mathbf{F}$ used alone does not allow standardization of rectification homographies
- for that we need either of these:
  1. projection matrices, or calibrated cameras, or
  2. a few points at infinity calibrating k_{1i}, k_{2i}, $i = 1, 2, 3$ in (36), from $\mathbf{K}_1 \underline{\mathbf{X}}_\infty \simeq \mathbf{K}_2 \underline{\mathbf{X}}_\infty$

Optimal and Non-linear Rectification

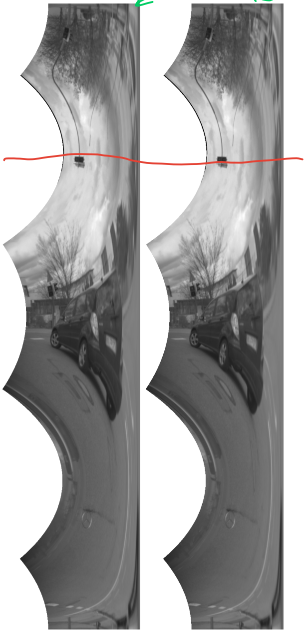
Optimal choice for the free parameters in $\mathbf{H}_{1,2}$

- by minimization of residual image distortion, eg. [Gluckman & Nayar 2001]

$$\mathbf{A}_i^* = \arg\min_{\mathbf{A}_i} \iint_\Omega \left(\det J\big((A_i \circ H_i)(\mathbf{x})\big) - 1\right)^2 d\mathbf{x}\,, \quad i = 1,2$$

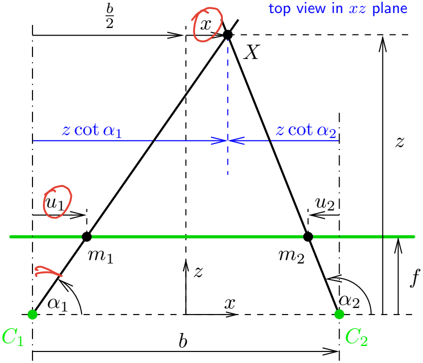
- by minimization of image information loss [Matoušek, ICIG 2004]
- non-linear rectification — suitable for forward motion
  non-parametric: [Pollefeys et al. 1999]
  analytic: [Geyer & Daniilidis 2003]

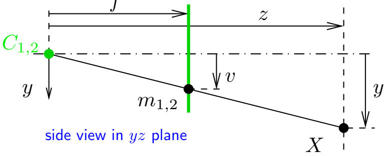
forward egomotion



rectified images, Pollefeys' method

▶Trivializing Epipolar Geometry: Binocular Disparity in a Standard Stereo Pair





- Assumptions: single image line, standard camera pair

$$b = z\cot\alpha_1 - z\cot\alpha_2 \qquad b = \frac{b}{2} + x - z\cot\alpha_2$$

$$u_1 = f\cot\alpha_1 \qquad u_2 = f\cot\alpha_2$$

- eliminate α_1, α_2 and obtain:
  $X = (x, y, z)$ from **disparity** $d = u_1 - u_2$:

$$z = \frac{b\,f}{d}, \quad x = \frac{b}{d}\,\frac{u_1 + u_2}{2}, \quad y = \frac{b\,v}{d}$$

f, d, u, v in pixels, b, x, y, z in meters

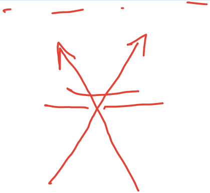
Observations

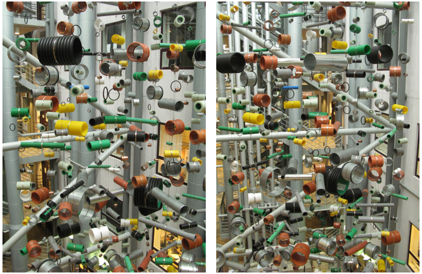
- constant disparity surface is a frontoparallel plane
- distant points have small disparity
- relative error in z is large for small disparity

$$\frac{1}{z}\frac{\mathrm{d}z}{\mathrm{d}d} = -\frac{1}{d}$$

- increasing the baseline or the focal length increases disparity, hence reduces the error

How Difficult Is Stereo?





Centrum för teknikstudier at Malmö Högskola, Sweden

The Vyšehrad Fortress, Prague

- top: easy interpretation from even a single image
- bottom left: we have no help from image interpretation
- bottom right: ambiguous interpretation due to a combination of missing texture and occlusion

Thank You