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 8, 2020

Open Informatics Master's Course

Implementing Simple Linear Constraints

What for?

1. fixing external frame as in $\theta_i = \mathbf{t}_i$, $s_{kl} = 1$ for some i, k, l — 'trivial gauge'
2. representing additional knowledge as in $\theta_i = \theta_j$ — e.g. cameras share calibration matrix $\mathbf{K}$

Introduce reduced parameters $\hat{\theta}$ and replication matrix $\mathbf{T}$:

$$\theta = \mathbf{T}\,\hat{\theta} + \mathbf{t}, \quad \mathbf{T} \in \mathbb{R}^{p,\hat{p}}, \quad \hat{p} \le p$$

then $\mathbf{L}_r$ in LM changes to $\mathbf{L}_r\,\mathbf{T}$ and everything else stays the same →107

$$\mathbf{T} = \begin{array}{c|cccc} & \hat\theta_1 & \hat\theta_2 & \hat\theta_3 & \hat\theta_4 \\ \hline \theta_1 & 1 & & & \\ \theta_2 & & 1 & & \\ \theta_3 & & & & \\ \theta_4 & & & & 1 \\ \theta_5 & & & & 1 \end{array} \qquad \mathbf{t} = \begin{bmatrix} \\ \\ 1 \\ \\ \\ \end{bmatrix}$$

these $\mathbf{T}$, $\mathbf{t}$ represent	
$\theta_1 = \hat{\theta}_1$	no change
$\theta_2 = \hat{\theta}_2$	no change
$\theta_3 = t_3$	constancy
$\theta_4 = \theta_5 = \hat{\theta}_4$	equality

- $\mathbf{T}$ deletes columns of $\mathbf{L}_r$ that correspond to fixed parameters — it reduces the problem size
- consistent initialisation: $\theta^0 = \mathbf{T}\,\hat{\theta}^0 + \mathbf{t}$ — or filter the init by pseudoinverse $\theta^0 \mapsto \mathbf{T}^\dagger \theta^0$
- no need for computing derivatives for θ_j corresponding to all-zero rows of $\mathbf{T}$ — fixed θ

- constraining projective entities →147–149
- more complex constraints tend to make normal equations dense
- implementing constraints is safer than explicit renaming of the parameters, gives a flexibility to experiment
- other methods are much more involved, see [Triggs et al. 1999]
- **BA resource:** http://www.ics.forth.gr/~lourakis/sba/ [Lourakis 2009]

Matrix Exponential: A path to Minimal Parameterizations

- for any square matrix we define

$$\operatorname{expm}(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{1}{k!}\,\mathbf{A}^k \qquad \text{note: } \mathbf{A}^0 = \mathbf{I}$$

- some properties:

$$\operatorname{expm}\mathbf{0} = \mathbf{I}, \quad \operatorname{expm}(-\mathbf{A}) = \left(\operatorname{expm}\mathbf{A}\right)^{-1},$$

$$\operatorname{expm}(a\mathbf{A} + b\mathbf{A}) = \operatorname{expm}(a\,\mathbf{A})\operatorname{expm}(b\,\mathbf{A}), \quad \operatorname{expm}(\mathbf{A}+\mathbf{B}) \neq \operatorname{expm}(\mathbf{A})\operatorname{expm}(\mathbf{B})$$

$\operatorname{expm}(\mathbf{A}^\top) = (\operatorname{expm}\mathbf{A})^\top$ hence if $\mathbf{A}$ is skew symmetric then $\operatorname{expm}\mathbf{A}$ is orthogonal:

$$(\operatorname{expm}(\mathbf{A}))^\top = \operatorname{expm}(\mathbf{A}^\top) = \operatorname{expm}(-\mathbf{A}) = (\operatorname{expm}(\mathbf{A}))^{-1}$$

$$\det\left(\operatorname{expm}\mathbf{A}\right) = e^{\operatorname{tr}\mathbf{A}}$$

$e^0 = 1$ $\qquad (e^A)^\top = (e^A)^{-1}$

Some consequences

- traceless matrices map to unit-determinant matrices ⇒ we can represent homogeneous representatives
- skew-symmetric matrices map to orthogonal matrices ⇒ we can represent rotations
- matrix exponential provides the exponential map from the powerful Lie group theory

Lie Groups Useful in 3D Vision

Matrix

group		matrix	represent
special linear	$\mathrm{SL}(3,\mathbb{R})$	real 3×3, unit determinant $\mathbf{H}$	2D homography
special linear	$\mathrm{SL}(4,\mathbb{R})$	real 4×4, unit determinant	3D homography
orthogonal	$\mathrm{SO}(3)$	real 3×3 orthogonal $\mathbf{R}$	3D rotation
special Euclidean	$\mathrm{SE}(3)$	4×4 $\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$, $\mathbf{R} \in \mathrm{SO}(3)$, $\mathbf{t} \in \mathbb{R}^3$	3D rigid motion
similarity	$\mathrm{Sim}(3)$	4×4 $\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & s^{-1} \end{bmatrix}$, $s \in \mathbb{R} \setminus 0$	rigid motion + scale

- Lie group G = topological group that is also a smooth manifold with nice properties
- Lie algebra $\mathfrak{g}$ = vector space associated with a Lie group (tangent space of the manifold)
- group: this is where we need to work
- algebra: this is how to represent group elements with a minimal number of parameters
- Exponential map = map between algebra and its group $\exp \colon \mathfrak{g} \to G$
- for matrices $\exp = \mathrm{expm}$
- in most of the above groups we have a closed-form formula for the exponential and for its principal inverse
- also Jacobians are readily available

Homography

$$\mathbf{H} = \operatorname{expm} \mathbf{Z}$$

- $\mathrm{SL}(3, \mathbb{R})$ group element

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad \text{s.t.} \quad \det \mathbf{H} = 1$$

- $\mathfrak{sl}(3, \mathbb{R})$ algebra element — 8 parameters

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & -(z_{11} + z_{22}) \end{bmatrix}$$

- note that $\operatorname{tr} \mathbf{Z} = 0$

▶Rotation in 3D

rotation angle [rad]

$$\mathbf{R} = \operatorname{expm}\,[\boldsymbol{\phi}]_\times, \quad \boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3) = \varphi\,\mathbf{e}_\varphi, \quad 0 \le \varphi < \pi, \quad \|\mathbf{e}_\varphi\| = 1$$

axis vector

- $\mathrm{SO}(3)$ group element

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \text{s.t.} \quad \mathbf{R}^{-1} = \mathbf{R}^\top$$

- $\mathfrak{so}(3)$ algebra element — 3 parameters

$$[\boldsymbol{\phi}]_\times = \begin{bmatrix} 0 & -\phi_3 & \phi_2 \\ \phi_3 & 0 & -\phi_1 \\ -\phi_2 & \phi_1 & 0 \end{bmatrix}$$

- exponential map in closed form — Rodrigues' formula

$$\mathbf{R} = \operatorname{expm}\,[\boldsymbol{\phi}]_\times = \sum_{n=0}^{\infty} \frac{[\boldsymbol{\phi}]_\times^n}{n!} = \overset{\circledast\,1}{\cdots} = \mathbf{I} + \frac{\sin\varphi}{\varphi}\,[\boldsymbol{\phi}]_\times + \frac{1-\cos\varphi}{\varphi^2}\,[\boldsymbol{\phi}]_\times^2$$

- (principal) logarithm — log is a periodic function

$$0 \le \varphi < \pi, \quad \cos\varphi = \frac{1}{2}\left(\operatorname{tr}(\mathbf{R}) - 1\right), \quad [\boldsymbol{\phi}]_\times = \frac{\varphi}{2\sin\varphi}(\mathbf{R} - \mathbf{R}^\top),$$

- $\boldsymbol{\phi}$ is rotation axis vector $\mathbf{e}_\varphi$ scaled by rotation angle φ in radians
- finite limits for $\varphi \to 0$ exist: $\sin(\varphi)/\varphi \to 1$, $(1-\cos\varphi)/\varphi^2 \to 1/2$

3D Rigid Motion

$$\mathbf{M} = \operatorname{expm}\,[\boldsymbol{\nu}]_\wedge$$

- $\mathrm{SE}(3)$ group element — 4×4 matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad \text{s.t.} \quad \mathbf{R} \in \mathrm{SO}(3),\ \mathbf{t} \in \mathbb{R}^3$$

- $\mathfrak{se}(3)$ algebra element — 4×4 matrix

$$[\boldsymbol{\nu}]_\wedge = \begin{bmatrix} [\boldsymbol{\phi}]_\times & \boldsymbol{\rho} \\ \mathbf{0} & 0 \end{bmatrix} \quad \text{s.t.} \quad \boldsymbol{\phi} \in \mathbb{R}^3,\ \varphi = \|\boldsymbol{\phi}\| < \pi,\ \boldsymbol{\rho} \in \mathbb{R}^3$$

- exponential map in closed form

$$\mathbf{R} = \operatorname{expm}\,[\boldsymbol{\phi}]_\times, \quad \mathbf{t} = \operatorname{dexpm}([\boldsymbol{\phi}]_\times)\,\boldsymbol{\rho}$$

$$\operatorname{dexpm}([\boldsymbol{\phi}]_\times) = \sum_{n=0}^{\infty} \frac{[\boldsymbol{\phi}]_\times^n}{(n+1)!} = \mathbf{I} + \frac{1-\cos\varphi}{\varphi^2}[\boldsymbol{\phi}]_\times + \frac{\varphi - \sin\varphi}{\varphi^3}[\boldsymbol{\phi}]_\times^2$$

matrix inverse →

$$\operatorname{dexpm}^{-1}([\boldsymbol{\phi}]_\times) = \mathbf{I} - \frac{1}{2}[\boldsymbol{\phi}]_\times + \frac{1}{\varphi^2}\left(1 - \frac{\varphi}{2}\cot\frac{\varphi}{2}\right)[\boldsymbol{\phi}]_\times^2$$

- (principal) logarithm via a similar trick as in $\mathrm{SO}(3)$
- finite limits exist: $(\varphi - \sin\varphi)/\varphi^3 \to 1/6$
- this form is preferred to $\mathrm{SO}(3) \times \mathbb{R}^3$

▶Minimal Representations for Other Entities

- fundamental matrix via $\mathrm{SO}(3) \times \mathrm{SO}(3) \times \mathbb{R}$

$$\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^\top, \quad \mathbf{D} = \mathrm{diag}(1, d^2, 0), \quad \mathbf{U}, \mathbf{V} \in \mathrm{SO}(3), \quad 3 + 1 + 3 = 7 \text{ DOF}$$

- essential matrix via $\mathrm{SO}(3) \times \mathbb{R}^3$

$$\mathbf{E} = [-\mathbf{t}]_\times \mathbf{R}, \quad \mathbf{R} \in \mathrm{SO}(3), \quad \mathbf{t} \in \mathbb{R}^3, \ \|\mathbf{t}\| = 1, \qquad 3 + 2 = 5 \text{ DOF}$$

- camera via $\mathrm{SO}(3) \times \mathbb{R}^3$ or $\mathrm{SE}(3)$

$$\mathbf{P} = \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \end{bmatrix} \mathbf{M}, \qquad 5 + 3 + 3 = 11 \text{ DOF}$$

- $\mathrm{Sim}(3)$ useful for SfM without scale
- closed-form formulae still exist but are a bit messy
- a (bit too brief) intro to Lie groups in 3D vision/robotics and SW:

J. Solà, J. Deray, and D. Atchuthan. A micro Lie theory for state estimation in robotics. arXiv:1812.01537v7 [cs.RO], August 2020.

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.

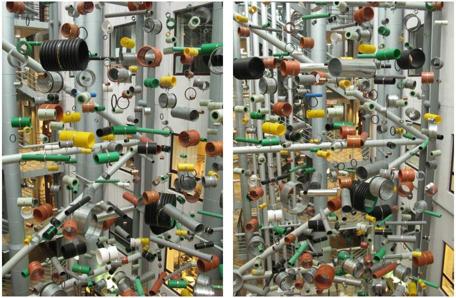
J. Gluckman and S. K. Nayar. Rectifying transformations that minimize resampling effects. In *Proc IEEE CS Conf on Computer Vision and Pattern Recognition*, vol. 1:111–117. 2001.

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.

What Are The Relative Distances?

- monocular vision already gives a rough 3D sketch because we understand the scene

What Are The Relative Distances?

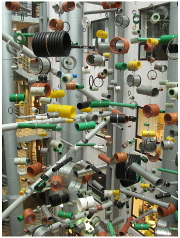


Centrum för teknikstudier at Malmö Högskola, Sweden

The Vyšehrad Fortress, Prague

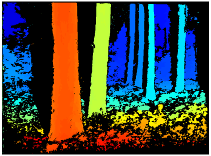
- left: we have no help from image interpretation
- right: ambiguous interpretation due to a combination of missing texture and occlusion

▶How Difficult Is Stereo?

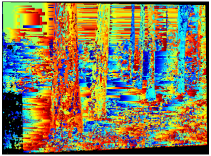


- when we do not recognize the scene and cannot use high-level constraints the problem seems difficult (right, less so in the center)
- most stereo matching algorithms do not require scene understanding prior to matching
- the success of a model-free stereo matching algorithm is unlikely:

left image



a good disparity map



disparity map from WTA

WTA Matching:

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

A Summary of Our Observations and an Outlook

1. simple matching algorithms do not work
2. stereopsis requires image interpretation in sufficiently complex scenes

   or another-modality measurement

we have a tradeoff: model strength $\leftrightarrow$ universality

Outlook:

1. represent the occlusion constraint: correspondences are not independent due to occlusions
   - epipolar rectification
   - disparity
   - uniqueness as an occlusion constraint
2. represent piecewise continuity the weakest of interpretations; piecewise: object boundaries
   - ordering as a weak continuity model
3. use a consistent framework
   - finding the most probable solution (MAP)

▶Linear Epipolar Rectification for Easier Correspondence Search

Obs:

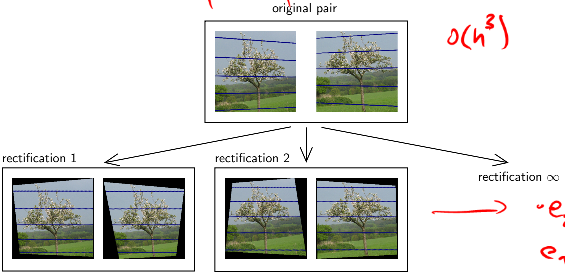
- if we map epipoles to infinity, epipolar lines become parallel
- we then rotate them to become horizontal
- we then scale the images to make correspoding epipolar lines colinear
- this can be achieved by a pair of (non-unique) homographies applied to the images

Problem: Given fundamental matrix $\mathbf{F}$ or camera matrices $\mathbf{P}_1$, $\mathbf{P}_2$, compute a pair of homographies that maps epipolar lines to horizontal 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 $\mathbf{P}_1 \mapsto \mathbf{H}_1\mathbf{P}_1$, $\mathbf{P}_2 \mapsto \mathbf{H}_2\mathbf{P}_2$.



Thank You