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 14, 2021

Open Informatics Master's Course

▶Choleski Decomposition for B. A.

The most expensive computation in B. A. is solving the normal eqs:

$$\text{find } \mathbf{x} \text{ such that } \quad -\sum_{r=1}^{z} \mathbf{L}_r^\top \nu_r(\theta^s) = \Big(\sum_{r=1}^{z} \mathbf{L}_r^\top \mathbf{L}_r + \lambda \operatorname{diag}\big(\mathbf{L}_r^\top \mathbf{L}_r\big)\Big)\mathbf{x}$$

- $\mathbf{A}$ is very large — approx. $3 \cdot 10^4 \times 3 \cdot 10^4$ for a small problem of 10000 points and 5 cameras
- $\mathbf{A}$ is sparse and symmetric, $\mathbf{A}^{-1}$ is dense — direct matrix inversion is prohibitive

> Choleski: symmetric positive definite matrix $\mathbf{A}$ can be decomposed to $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$, where $\mathbf{L}$ is lower triangular. If $\mathbf{A}$ is sparse then $\mathbf{L}$ is sparse, too.

1. decompose $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$ — transforms the problem to $\mathbf{L}\underbrace{\mathbf{L}^\top \mathbf{x}}_{\mathbf{c}} = \mathbf{b}$
2. solve for $\mathbf{x}$ in two passes:

$$\mathbf{L}\,\mathbf{c} = \mathbf{b} \qquad \mathbf{c}_i := \mathbf{L}_{ii}^{-1}\Big(\mathbf{b}_i - \sum_{j<i} \mathbf{L}_{ij}\mathbf{c}_j\Big)$$

forward substitution, $i = 1, \ldots, q$ (params)

$$\mathbf{L}^\top \mathbf{x} = \mathbf{c} \qquad \mathbf{x}_i := \mathbf{L}_{ii}^{-1}\Big(\mathbf{c}_i - \sum_{j>i} \mathbf{L}_{ji}\mathbf{x}_j\Big)$$

back-substitution

- Choleski decomposition is fast (does not touch zero blocks)
  non-zero elements are $9p + 121k + 66pk \approx 3.4 \cdot 10^6$; ca. $250\times$ fewer than all elements
- it can be computed on single elements or on entire blocks
- use profile Choleski for sparse $\mathbf{A}$ and diagonal pivoting for semi-definite $\mathbf{A}$ — see above; [Triggs et al. 1999]
- λ controls the definiteness

Profile Choleski Decomposition is Simple

```
function L = pchol(A)
%
% PCHOL profile Choleski factorization,
%    L = PCHOL(A) returns lower-triangular sparse L such that A = L*L'
%    for sparse square symmetric positive definite matrix A,
%    especially efficient for arrowhead sparse matrices.

% (c) 2010 Radim Sara (sara@cmp.felk.cvut.cz)

 [p,q] = size(A);
 if p ~= q, error 'Matrix A is not square'; end

 L = sparse(q,q);
 F = ones(q,1);
 for i=1:q
  F(i) = find(A(i,:),1); % 1st non-zero on row i; we are building F gradually
  for j = F(i):i-1
   k = max(F(i),F(j));
   a = A(i,j) - L(i,k:(j-1))*L(j,k:(j-1))';
   L(i,j) = a/L(j,j);
  end
  a = A(i,i) - sum(full(L(i,F(i):(i-1))).^2);
  if a < 0, error 'Matrix A is not positive definite'; end
  L(i,i) = sqrt(a);
 end
end
```

$\frac{1}{2}(A+A^T)$

▶Gauge Freedom (kalibrační invariance)

1. The external frame is not fixed: See Projective Reconstruction Theorem →132

$$\underline{\mathbf{m}}_{ij} \simeq \mathbf{P}_j \underline{\mathbf{X}}_i = \mathbf{P}_j \mathbf{H}^{-1} \mathbf{H} \underline{\mathbf{X}}_i = \mathbf{P}'_j \underline{\mathbf{X}}'_i$$

2. Some representations are not minimal, e.g.
   - $\mathbf{P}$ is 12 numbers for 11 parameters
   - we may represent $\mathbf{P}$ in decomposed form $\mathbf{K}$, $\mathbf{R}$, $\mathbf{t}$
   - but $\mathbf{R}$ is 9 numbers representing the 3 parameters of rotation

As a result

- there is no unique solution
- matrix $\sum_r \mathbf{L}_r^\top \mathbf{L}_r$ is singular

Solutions

1. fixing the external frame (e.g. a selected camera frame) explicitly or by constraints
2. fixing the scale (e.g. $s_{12} = 1$)

3a. either imposing constraints on projective entities
   - cameras, e.g. $\mathbf{P}_{3,4} = 1$ — this excludes affine cameras
   - points, e.g. $(\underline{\mathbf{X}}_i)_4 = 1$ or $\|\underline{\mathbf{X}}_i\|^2 = 1$ — the 2nd: can represent points at infinity

3b. or using minimal representations
   - points in their Euclidean representation $\mathbf{X}_i$ — but finite points may be an unrealistic model
   - rotation matrices can be represented by skew-symmetric matrices →149

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 →108

$\mathbf{T} =$

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
θ_1	1			
θ_2		1		
θ_3				
θ_4				1
θ_5				1

$\mathbf{t} =$

1

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 →149–151
- 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

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

$A \in \mathbb{R}^{n \times n}$

$GL(n)$

- some properties:

$$\operatorname{expm}(x) = e^x, \quad x \in \mathbb{R}, \quad \operatorname{expm}\mathbf{0} = \mathbf{I}, \quad \operatorname{expm}(-\mathbf{A}) = \big(\operatorname{expm}\mathbf{A}\big)^{-1},$$

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

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

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

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

Some consequences

- traceless matrices ($\operatorname{tr}\mathbf{A} = 0$) map to unit-determinant matrices $\Rightarrow$ we can represent homogeneous matrices
- skew-symmetric matrices map to orthogonal matrices $\Rightarrow$ we can represent rotations
- matrix exponential provides the exponential map from the powerful Lie group theory

Lie Groups Useful in 3D Vision

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
special 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 a have a closed-form formula for the exponential and for its principal inverse
- Jacobians are also readily available for $\mathrm{SO}(3)$, $\mathrm{SE}(3)$ [Solà 2020]

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

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

- $\mathrm{SO}(3)$ 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

SE(3)

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

- $\mathrm{SE}(3)$ 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

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

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

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

- dexpm: differential of the exponential in $\mathrm{SO}(3)$
- (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 pose 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.

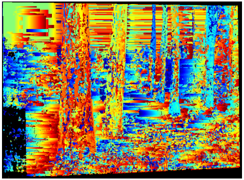
J. Gluckman and S. K. Nayar. Rectifying transformations that minimize resampling effects. In *Proc IEEE CS Conf on Computer Vision and Pattern Recognition*, vol. 1:111–117. 2001.

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.

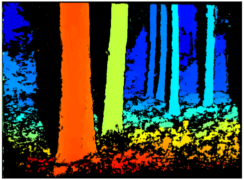
The success of a model-free stereo matching algorithm is unlikely:

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 understanding is difficult
- matching can benefit from a geometric simplification of the problem

▶Linear Epipolar Rectification for Easier Correspondence Search

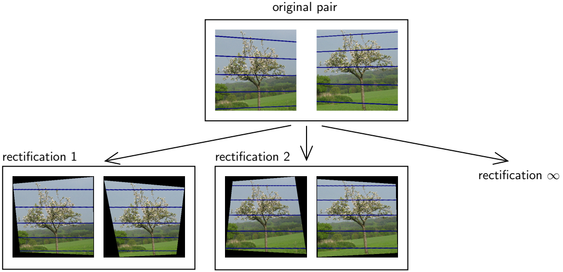
Obs:

- if we map epipoles to infinity, epipolar lines become parallel
- we then rotate them to become horizontal
- we then scale the images to make corresponding 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 $\mathbf{P}_1 \mapsto \mathbf{H}_1\mathbf{P}_1$, $\mathbf{P}_2 \mapsto \mathbf{H}_2\mathbf{P}_2$.

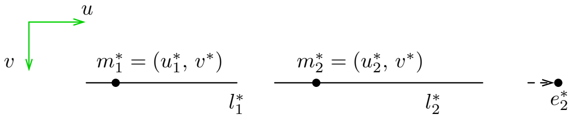


▶Rectification Homographies

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

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

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 $l_2^* = l_1^*$: $\quad \mathbf{l}_1^* \simeq \mathbf{e}_1^* \times \mathbf{m}_1 = [\mathbf{e}_1^*]_\times \mathbf{m}_1 \simeq \mathbf{l}_2^* \simeq \mathbf{F}^* \mathbf{m}_1 \Rightarrow \quad \mathbf{F}^* = [\mathbf{e}_1^*]_\times$

- 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 $\mathbf{F} = (\mathbf{Q}_1 \mathbf{Q}_2^{-1})^\top [\underline{\mathbf{e}}_1]_\times$ →79
- 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

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

- 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 cam. 1 frame
- result:

$$\mathbf{R}^* = \mathbf{I}, \quad \mathbf{b}^* = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{K}_1^* = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_2^* = \begin{bmatrix} k_{21} & k_{22} & k_{23} \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{33}$$

- rectified cameras are in canonical relative pose — not rotated, canonical baseline
- rectified calibration matrices can differ in the first row only
- when $\mathbf{K}_1^* = \mathbf{K}_2^*$ then the rectified pair is called the standard stereo pair and the homographies 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 simple 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 →155)

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

3. Then

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

and the primitive rectification homographies are

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

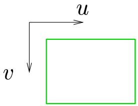
⊛ P1; 1pt: derive some other admissible $\mathbf{A}$, $\mathbf{B}$

- rectification homographies do exist →155
- there are other primitive rectification homographies, these suggested are just simple to obtain

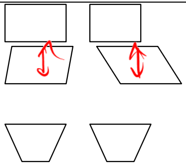
▸The Set of All Rectification Homographies

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

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



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 $\mathbf{A}_2\mathbf{K}_2^* = \hat{\mathbf{K}}_2\mathbf{G}$

The Rectification Group

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.

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

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: $d = 1 \Rightarrow \hat{\mathbf{H}}_1,\ \hat{\mathbf{H}}_2$ (→157) are orthonormal

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.

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

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 (→156) and non-negative disparity:

let $\mathbf{K}_i^{-1}\mathbf{P}_i = \mathbf{R}_i\,[\mathbf{I} \quad -\mathbf{C}_i]\,, \quad i = 1, 2$ — note we started from $\mathbf{E}$, not $\mathbf{F}$

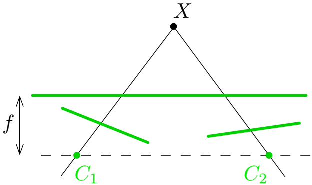
$$\mathbf{H}_1\mathbf{P}_1 = \mathbf{K}\hat{\mathbf{H}}_1\mathbf{K}_1^{-1}\mathbf{P}_1 = \mathbf{K}\,\underbrace{\mathbf{B}\mathbf{V}^\top\mathbf{R}_1}_{\mathbf{R}^*}\,[\mathbf{I} \quad -\mathbf{C}_1] = \mathbf{K}\mathbf{R}^*\,[\mathbf{I} \quad -\mathbf{C}_1]$$

$$\mathbf{H}_2\mathbf{P}_2 = \mathbf{K}\hat{\mathbf{H}}_2\mathbf{K}_2^{-1}\mathbf{P}_2 = \mathbf{K}\,\underbrace{\mathbf{A}\mathbf{U}^\top\mathbf{R}_2}_{\mathbf{R}^*}\,[\mathbf{I} \quad -\mathbf{C}_2] = \mathbf{K}\mathbf{R}^*\,[\mathbf{I} \quad -\mathbf{C}_2]$$

- 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 (13)
- points at infinity project by $\mathbf{K}\mathbf{R}^*$ in both cameras ⇒ they have zero disparity →166

▶Summary & Remarks: Linear Rectification

standard rectification homographies reproject onto a common image plane parallel to the baseline



- rectification is done with a pair of homographies (one per image) →154
  - ⇒ rectified camera centers 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 →156
  - ⇒ rectified image projection planes are coplanar
- equal rectified calibration matrices give standard rectification →156
  - ⇒ rectified image projection planes are equal
- primitive rectification is already standard in calibrated cameras →160

- 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 (33)

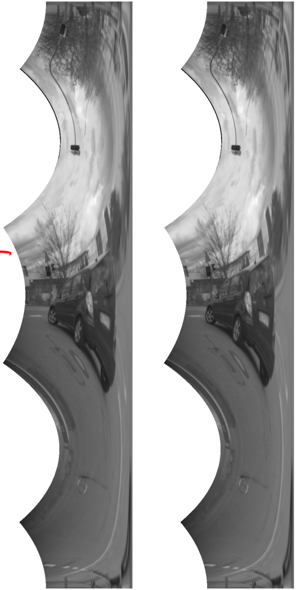
Optimal choice for the free parameters

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

$$\mathbf{A}_1^* = \arg\min_{\mathbf{A}_1} \iint_\Omega \left(\det J(\mathbf{A}_1 \hat{\mathbf{H}}_1 \underline{\mathbf{x}}) - 1\right)^2 d\mathbf{x}$$

- 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

Thank You