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. November 30, 2021

Open Informatics Master's Course

▶Stripping MH Down To Get RANSAC [Fischler & Bolles 1981]

- when we are interested in the best config only. . . and we need fast data exploration. . .

. . . then the sampling procedure simplifies:

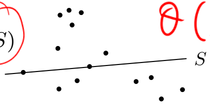
1. ~~given C_t,~~ draw a random sample S from $\cancel{q(S \mid C_t)}$ $q(S)$ independent sampling
   no use of information from C_t
2. ~~compute acceptance probability~~
$$a = \min\left\{1,\ \frac{\pi(S)}{\pi(C_t)} \cdot \boxed{\frac{q(C_t \mid S)}{q(S \mid C_t)}}\right\}$$
   difficult
3. ~~draw a random number u from unit-interval uniform distribution $U_{0,1}$~~
4. ~~if $u \leq a$ then $C_{t+1} := S$ else $C_{t+1} := C_t$~~
5. if $\pi(S) > \pi(C_{\text{best}})$ then remember $C_{\text{best}} := S$

Steps 2–4 make no difference when waiting for the best sample configuration

- . . . but getting a good accuracy configuration might take very long this way
- good overall exploration but slow convergence in the vicinity of a mode where C_t could serve as an attractor
- cannot use the past generated configurations to estimate any parameters
- we will fix these problems by (possibly robust) 'local optimization'

▶RANSAC with Local Optimization and Early Stopping

1. initialize the best configuration as empty $C_{\text{best}} := \emptyset$ and time $t := 0$
2. estimate the number of needed proposals as $N := \binom{n}{s}$ n – No. of primitives, s – minimal config size
3. while $t \le N$:

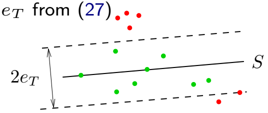


$\theta(S) = n(S)$

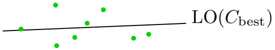
   a) propose a minimal random config S of size s from $q(S)$
   robust
   b) if $\pi(S) > \pi(C_{\text{best}})$ then
      i) update the best config $C_{\text{best}} := S$ $\pi(S)$ marginalized as in (26); $\pi(S)$ includes a prior $\Rightarrow$ MAP

      $q = \binom{n}{s}^{-1}$

      ii) threshold-out inliers using e_T from (27)



      iii) start local optimization from the inliers of C_{best} LM optimization with robustified (→115) Sampson error possibly weighted by posterior $\pi(m_{ij})$ [Chum et al. 2003]



      iv) update C_{best}, update inliers using (27), re-estimate N from inlier counts →124 for derivation

$$N = \frac{\log(1-P)}{\log(1-\varepsilon^s)}, \quad \varepsilon = \frac{|\operatorname{inliers}(C_{\text{best}})|}{m\,n},$$

   c) $t := t + 1$

$\varepsilon_k = (1-\gamma_k)\varepsilon_{k-1} + \gamma_k \varepsilon_k^n$ $\quad \gamma \approx 1/k$

4. output C_{best}

- see MPV course for RANSAC details see also [Fischler & Bolles 1981], [25 years of RANSAC]

Example Matching Results for the 7-point Algorithm with RANSAC

input images | interest points (ca. 3600) | tentative corresp. (416) | matching (340)

- no descriptors used
- notice some wrong matches (they have wrong depth, even negative)

remember: hidden labels →111

- they cannot be rejected without additional constraints or scene knowledge
- without local optimization the minimization is over a discrete set of epipolar geometries proposable from 7-tuples

Beyond RANSAC

By marginalization in (23) we have lost constraints on M (e.g. uniqueness). One can choose a better model when not marginalizing:

$$\pi(M, \mathbf{F}, E, D) = \underbrace{p(E \mid M, \mathbf{F})}_{\text{reprojection error}} \cdot \underbrace{p(D \mid M)}_{\text{similarity}} \cdot \underbrace{p(\mathbf{F})}_{\text{prior}} \cdot \underbrace{P(M)}_{\text{constraints}}$$

this is a global model: decisions on m_{ij} are no longer independent!

In the MH scheme

- one can work with full $p(M, \mathbf{F} \mid E, D)$, then configuration $C = M$ — $\mathbf{F}$ computable from M
  - explicit labeling m_{ij} can be done by, e.g. sampling from

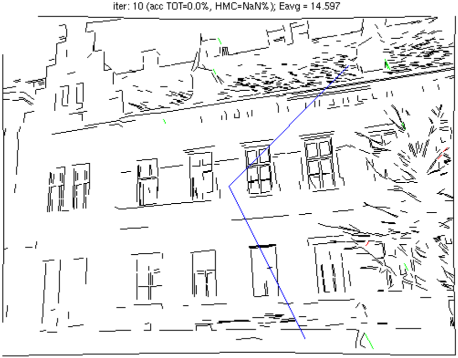
    $$q(m_{ij} \mid \mathbf{F}) \sim \big((1 - P_0)\, p_1(e_{ij} \mid \mathbf{F}),\; P_0\, p_0(e_{ij} \mid \mathbf{F})\big)$$

    when $P(M)$ uniform then always accepted, $a = 1$ — ⊛ derive
  - we can compute the posterior probability of each match $p(m_{ij})$ by histogramming m_{ij} from $\{C_i\}$
  - local optimization can then use explicit inliers and $p(m_{ij})$
- error can be estimated for the elements of $\mathbf{F}$ from $\{C_i\}$ — does not work in RANSAC!
- large error indicates problem degeneracy — this is not directly available in RANSAC
- good conditioning is not a requirement — we work with the entire distribution $p(\mathbf{F})$
- one can find the most probable number of epipolar geometries (homographies or other models) — by reversible jump MCMC and Bayesian model selection

  if there are multiple models explaning data, RANSAC will return one of them randomly

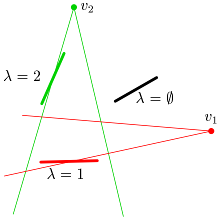
Example: MH Sampling for a More Complex Problem

Task: Find two vanishing points from line segments detected in input image. Principal point is known, square pixel.



video

- primitives = line segments
- latent variables
  1. each line has a vanishing point label $\lambda_i \in \{\emptyset, 1, 2\}$, $\emptyset$ represents an outlier
  2. 'mother line' parameters θ_L (they pass through their vanishing points)
- explicit variables
  1. two unknown vanishing points v_1, v_2
- marginal proposals (v_i fixed, v_j proposed)
- minimal configuration $s = 2$



$$\arg\min_{v_1, v_2, \Lambda, \theta_L} V(v_1, v_2, \Lambda, \theta_L)$$

simplifications

- vanishing points restricted to the set of all pairwise segment intersections
- mother lines fixed by segment centroid, then θ_L uniquely given by λ_i, and the configuration is

$$C = \{v_1, v_2, \Lambda\}$$

Module VI

3D Structure and Camera Motion

6.1 Reconstructing Camera System: From Triples and from Pairs

6.2 Bundle Adjustment

covered by

[1] [H&Z] Secs: 9.5.3, 10.1, 10.2, 10.3, 12.1, 12.2, 12.4, 12.5, 18.1

[2] Triggs, B. et al. Bundle Adjustment—A Modern Synthesis. In *Proc ICCV Workshop on Vision Algorithms.* Springer-Verlag. pp. 298–372, 1999.

additional references

D. Martinec and T. Pajdla. Robust Rotation and Translation Estimation in Multiview Reconstruction. In *Proc CVPR*, 2007

M. I. A. Lourakis and A. A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. ACM Trans Math Software 36(1):1–30, 2009.

▶Reconstructing Camera System by Gluing Camera Triples

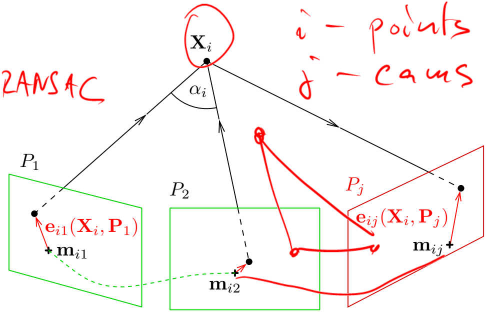
Given: Calibration matrices $\mathbf{K}_j$ and tentative correspondences per camera triples.

Initialization

1. initialize camera cluster $\mathcal{C}$ with P_1, P_2,
2. find essential matrix $\mathbf{E}_{12}$ and matches M_{12} by the 5-point algorithm →88 | RANSAC
3. construct camera pair

$$\mathbf{P}_1 = \mathbf{K}_1 \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}, \ \mathbf{P}_2 = \mathbf{K}_2 \begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}$$

4. triangulate $\{X_i\}$ per match from M_{12} →106
5. initialize point cloud $\mathcal{X}$ with $\{X_i\}$ satisfying chirality constraint $z_i > 0$ and apical angle constraint $|\alpha_i| > \alpha_T$



Attaching camera $P_j \notin \mathcal{C}$

1. select points $\mathcal{X}_j$ from $\mathcal{X}$ that have matches to P_j
2. estimate $\mathbf{P}_j$ using $\mathcal{X}_j$, RANSAC with the 3-pt alg. (P3P), projection errors $\mathbf{e}_{ij}$ in $\mathcal{X}_j$ →66
3. reconstruct 3D points from all tentative matches from P_j to all P_l, $l \neq k$ that are not in $\mathcal{X}$
4. filter them by the chirality and apical angle constraints and add them to $\mathcal{X}$
5. add P_j to $\mathcal{C}$
6. perform bundle adjustment on $\mathcal{X}$ and $\mathcal{C}$

coming next →138

▶The Projective Reconstruction Theorem

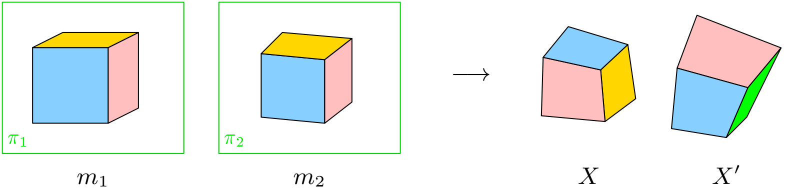
- We can run an analogical procedure when the cameras remain uncalibrated. But:

Observation: Unless $\mathbf{P}_i$ are constrained, then for any number of cameras $i = 1, \ldots, k$

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

- when $\mathbf{P}_i$ and $\underline{\mathbf{X}}$ are both determined from correspondences (including calibrations $\mathbf{K}_i$), they are given up to a common 3D homography $\mathbf{H}$

(translation, rotation, scale, shear, pure perspectivity)



- when cameras are internally calibrated ($\mathbf{K}_i$ known) then $\mathbf{H}$ is restricted to a similarity since it must preserve the calibrations $\mathbf{K}_i$ [H&Z, Secs. 10.2, 10.3], [Longuet-Higgins 1981]

(translation, rotation, scale)

Thank You