Photogrammetry & Robotics Lab

3D Coordinate Systems

(Bsc Geodesy & Geoinformation)

14. Outlier Detection

Wolfgang Förstner

The slides have been created by Wolfgang Förstner.

The problem – step by step (animated)

1st system with 7 points
2nd system translated
2nd system scaled/rotated
two outliers introduced

Which are the outliers?

The problem – step by step

1st system
2nd system
2nd system scaled/rotated
two outliers introduced

Which are the outliers? (5 and 6)

Topics

1. The notion of outliers
2. The complexity of the problem
3. Complete search
4. Random sample consensus

For this video see Förstner/Wrobel (2016), Sect. 4.7

6.1 The notion of outliers

The notion of outliers

Model vs. data errors

e.g. Gauss-Markov model

$$\boldsymbol{l} \quad \text{sample of} \quad \underline{\boldsymbol{l}} \sim \mathcal{N}(\boldsymbol{f}(\boldsymbol{x}), \Sigma_{ll})$$

- No possibility to distinguish
  errors in data and errors in model
- Errors in
  - Function (to many, too few, wrong functions)
  - Covariance matrix (wrong variance, correlations)
  - Distribution (close to or far from normal, one sided, ….)

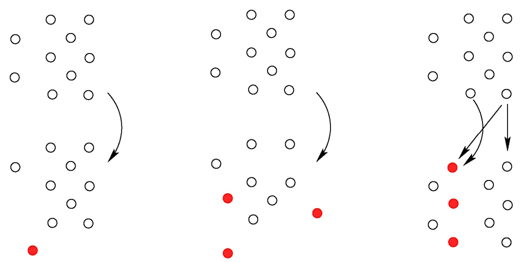
The notion of outliers

- Errors may be
  - Random: meant to be covered by the stochastical model
  - Systematic: wrong # parameters, neglected covariances
    (not necessarily distinguishable)
  - Blunders/outliers: anything else
    mostly wrong identification of correspondences
- Outliers may be
  - Small: negligible, hardly detectable, …
  - Medium: show, but do not severely distort estimate
  - Large: heavily distort estimate

The notion of outliers

Outliers may occur

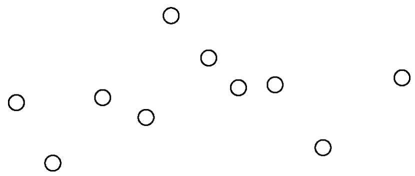
- Seldom/isolated (< 0.1 %) or often (>10 %)
- Random or mimicking good result



6.2 The complexity of the problem

Complexity of problem

Number I of entities/observations/points
Each may be inlier or outlier
Number of alternatives 2^I : exponentially many
e.g. 10 points meant to be on a straight line
→ 1024 alternatives



Two types of approaches

Two alternative views on solution methods

1. **Search** for outliers or inliers: discrete problem
   → use inliers for estimation
2. **Robust estimation**: continuous problem
   "Estimation which is insensitive to deviations from model assumptions"
   → identify inliers

Boundaries between both methods blurry

Two types of approaches

Search

Select minimal subset of observations
Check consistency with model, e.g. using Ω
Problem: How to select? Sequence of trials?

Robust estimation

Perform estimation insensitive to errors in the model
Identify outliers
Problem: How to identify?

The break down point of estimator

The percentage of outlying observations which may
allow estimator to give arbitrary wrong result

Maximum break down point: 50 %
otherwise outliers may mimic good results

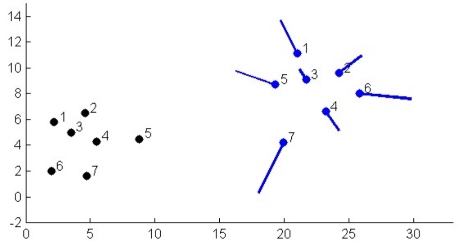
e.g.

- Arithmetic mean of I points: one outlier → bad
  break down point = 0 % → all LSE have BDP = 0
- Median of I points: if outliers are the minority → fine
  break down point = $\lfloor (I-1)/2 \rfloor / I \approx 50\%$

LS estimation for finding outliers?

1. Determine LS estimate of 2D similarity
2. Analyse residuals

→



- All residuals are large
- Largest residual not decisive (in 7, not in 5 or 6)

Large variety of methods depending on …

Preconditions:
availability of approximate values/direct solution

Likelihood of success
number of observations
homogeneity of configuration
error rate

Computational complexity
number of unknown parameters
existence of large outliers

Two algorithms

… selection

- Direct solution: available
- Number of observations: medium $(2U)$ to large $(> 10U)$
- Configuration: may be inhomogeneous
- Error rate may be high (possibly > 50 %)
- Number of parameters is small (≤ 12)
- Large outliers exist

17

Class of algorithms

- Deterministic algorithms
  The algorithm's output is determined by its input
- Stochastic algorithms
  The algorithm's output stochastically varies for same input
  → no solution is guaranteed

→ One deterministic/one stochastic algorithm

18

Complete search

Complete search

- For all possible configurations of inliers/outliers:
  Determine some evaluation function
- Choose the best configuration

Configuration: binary vector

$$\boldsymbol{\delta} = [\delta_i] \quad \text{with} \quad \delta_i = \text{'observation } l_i \text{ is inlier'} \in \{0, 1\}$$

19

Complete search

For 2D line or 2D similarity:
configurations with ≥ 2 or ≥ 3 points

$$K = 2^I - \sum_{k=0}^{1} \binom{I}{k} \quad \text{or} \quad K = 2^I - \sum_{k=0}^{2} \binom{I}{k}$$

For 3D similarity, configurations with ≥ 3 or ≥ 4 points

$$K = 2^I - \sum_{k=0}^{2} \binom{I}{k} \quad \text{or} \quad K = 2^I - \sum_{k=0}^{3} \binom{I}{k}$$

→ K solutions with 2 (3) or 3 (4) to I observed points
Complexity: exponential with the number of points

20

Complete search

Evaluation functions

1. Variance factor for inliers

$$\widehat{\sigma}_0^2(\boldsymbol{\delta}) = \frac{\sum_i \delta_i\,(w_i \widehat{v}_i^2)}{(\sum_i \delta_i) - U} \quad \text{using} \quad \delta_i = |\widehat{v}_i| < c\,\sigma_i \quad \text{with} \quad c \in [2,4]$$

$$\text{if } \sum_i \delta_i = U \text{ then } \widehat{\sigma}_0^2(\boldsymbol{\delta}) := 1$$

→ number of outliers is not counted →

2. Additional constant penalty for each outlier

$$\widehat{\sigma}_0^2(\boldsymbol{\delta}) = \frac{\left(\sum_i \delta_i\,(w_i \widehat{v}_i^2)\right) + c^2 \cdot \#\text{outliers}}{I - U} \qquad (*)$$

6.4 Random sample consensus (RANSAC)

Idea

Assumption: Probability of outliers

$$\varepsilon \leq \tfrac{1}{2}$$

→

Probability of choosing sample of S inliers

$$P = (1-\varepsilon)^S > \left(\tfrac{1}{2}\right)^S$$

e.g. for $S = 3$ (similarity transformation) → $P > 1/8$

Looks like a few samples, say 20, may be sufficient

Idea

Random Sample Consensus (RANSAC)
(Fischler/Bolles 1981)

- Randomly choose large enough number T of trial samples of size S, sufficient to determine parameters, to hit a good one with high probability
- Use the other $I - S$ points for checking consensus

RANSAC Algorithm for spatial similarity

Input: data $\mathcal{D} = \{\boldsymbol{x}_i, \boldsymbol{x}'_i, \sigma_{x'_i}, i = 1, ..., I\}$
outlier rate ε
number T of trials
Iterate: for each trial $t \in \{1, ..., T\}$
choose random set $\mathcal{S}_t$ of S points
determine parameters $\boldsymbol{p}_t(\mathcal{S}_t)$
determine consistency measure $c_t(\boldsymbol{p}_t, \mathcal{D} \backslash \mathcal{S}_t)$
Result: set with best consistency $\widehat{t} = \text{argmin}_t\, c_t$

25

Comments

- Data: variance $\sigma^2_{x'_i}$ for determining consistency
  (methods exist to estimate the variance)
- Number T of trials:
  ← → required probability P of success →
- Determine parameters:
  requires direct solution →
- Consistency measures:
  many alternatives →

26

Line fitting demo

Size of sample $S = 2$

Slow demo:

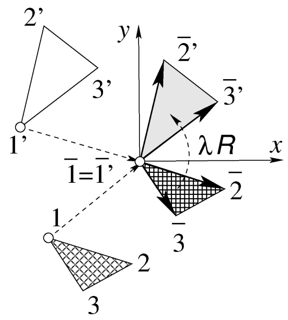
1. Generate in- and outliers
2. Perform inlier detection
   - Complete search. Consensus = variance factor for inliers
   - RANSAC. Consensus = # of inliers

27

Direct solution

Generally: may be hard to find

Here: **Similarity from 3 points**
1. Shift points 1 to origin
2. Determine scaled rotation
   from two direction pairs
   $$\boldsymbol{x}'_i - \boldsymbol{x}'_1 = \lambda R(\boldsymbol{x}_i - \boldsymbol{x}_1)$$
3. Derive similarity
   $$\boldsymbol{x}'_i = \lambda R \boldsymbol{x}_i + \boldsymbol{t} \quad \text{with} \quad \boldsymbol{t} = \boldsymbol{x}'_1 - \lambda R \boldsymbol{x}_1$$



28

Number of trials

Minimal number $T_{\min}$ of trials such that probability of finding at least one good set of S points is at least $P_{\min}$, if error rate is ε

$$\boxed{T_{\min} = \left\lceil \frac{\ln(1 - P_{\min})}{\ln(1 - (1 - \varepsilon)^S)} \right\rceil}$$

Approximation: $T \propto (1 - \varepsilon)^{-S}$

→ **complexity**: $O(k^S)$

Proof using $P = 1 - (1 - (1 - \varepsilon)^S)^T$

29

Proof: P(≥ 1 good S-samples in T trials)

Probability to draw one good point

$$1 - \varepsilon$$

Probability to draw a sample of S good points

$$(1 - \varepsilon)^S$$

Probability to draw bad S-sample - with at least one bad point

$$1 - (1 - \varepsilon)^S$$

Probability to draw only bad S-samples in T trials

$$(1 - (1 - \varepsilon)^S)^T$$

Probability to draw at least one good S-sample in T trials

$$1 - (1 - (1 - \varepsilon)^S)^T \quad \rightarrow \quad T_{\min} = f(P_{\min})$$

30

Number of trials $T_{\min}(S, \varepsilon, P_{\min})$ for $P_{\min} = 99\%$

- Sample size S
- Error rate ε
- Probability of success P

$S\backslash\varepsilon$	10%	20%	30%	40%	**50%**	60%	70%	80%	90%
1	2	3	4	6	7	10	13	21	44
2	3	5	7	11	**17**	27	49	113	459
3	4	7	11	19	**35**	70	169	574	4 603
4	5	9	17	34	72	178	567	2 876	46 050
5	6	12	26	57	146	448	1 893	14 389	460 515
6	7	16	37	97	293	1 123	6 315	71 954	4 605 168
7	8	20	54	163	588	2 809	21 055	359 777	46 051 700
8	9	26	78	272	1 177	7 025	70 188	1 798 893	460 517 017

31

Line fitting demo

Fast demo (→ above)

- Variation of # points
- Variation of precision
- Variation of error rate up to 95 %
- Failure for
  - Too low precision
  - Too low probability/number of trials
  - Too low number of points

32

Comments

✓ Number of trials **independent** on number I of points
→ big advantage for computational efficiency

✓ Sample size $S \leq U$ number of parameters,
(e.g. 3 points vs. 7 parameters)

– Number of necessary trials increases quickly with
- Number of parameters
- Probability approaching 1

→ **Only useful for small problems** ($U \leq 12$)

– **Result** only reachable **with limited probability**

– **Evaluation** of result **difficult**

33

Modifications ..

- Estimate observational noise $\sigma_{x'_i}$ during trials
  → use smallest $\widehat{\sigma}_0^2$ up to current trial
- Control sampling using P of observation being good
  → use classification up to current level
- Stop consensus calculation if no change in decision
  (Wald's (1945) sequential testing)
- Eliminate bad configurations from samples
- If prior knowledge available → exclude solutions

see Raguram, R. et al. (2013).
USAC: A Universal Framework for Random Sample Consensus, IEEE T-PAMI

34

Large number of applications

Estimation from highly perturbed data

- Estimating transformations
  (motions, similarities, affinities, perspective relations, …)
- Estimating geometric entities
  2D/3D lines (2), planes (3),
  circles (3), ellipses (5), spheres (4), cylinders (5),
  algebraic curves/surfaces, …)

Estimating multiple transformations/entities

(finding multiple lines / transformations …)

35

Summary

- Classification of problems for outlier detection
- A deterministic and a stochastic algorithm
- RANSAC as efficient method for finding outliers
  - Complexity exponential in minimum sample size $O(k^S)$
  - Limitations: small U, no guarantee for solution
- There exist algorithms (medium outliers, small ε)
  better than $O(U^3)$ for large U
  (e.g. ML-type estimators minimizing (*) on s. 21)

36

Strategy

1. Find inliers with adequate method
2. Perform ML estimation + evaluation

Thanks for watching

References of video series

Arun, K. S., T. S. Huang, and S. B. Blostein (1987). Least-Squares Fitting of Two 3D Point Sets. IEEE T-PAMI 9 (5), 698-700.

Bishop, C. (2006). Pattern Recognition and Machine Learning. Springer.

Boyd, S. and L. Vandenberghe (2004). Convex optimization. Cambridge University Press.

Cayley, A. (1846). Sur quelques proprétés des déterminants gauches. Journal für die reine und angewandte Mathematik 32, 119-123

Cover, T. and J. A. Thomas (1991). Elements of Information Theory. John Wiley & Sons.

H. S. M. Coxeter (1946): Quaternions and Reflections, The American Mathematical Monthly, 53(3), 136-146

Fallat, S. M. and M. J. Tsatsomeros (2002). On the Cayley Transform of Positivity Classes of Matrices. Electronic Journal of Linear Algebra 9, 190-196.

Fischler, M. A. and R. C. Bolles (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24 (6), 381-395.

Förstner, W. and B. P. Wrobel (2016). Photogrammetric Computer Vision Statistics, Geometry, Orientation and Reconstruction. Springer.

Horn, B. K. B. (1987). Closed-form Solution of Absolute Orientation Using Unit Quaternions. Optical Soc. of America.

Howell, T. D. and J.-C. Lafon (1975). The Complexity of the Quaternion Product. Technical Report TR75-245, Cornell University.

Kanatani, K. (1990). Group Theoretical Methods in Image Understanding. New York: Springer.

Koch, K.-R. (1999). Parameter Estimation and Hypothesis Testing in Linear Models (2nd ed.). Springer.

Li, S. Z. (2000). Markov random field modeling in computer vision. Springer.

McGlone, C. J., E. M. Mikhail, and J. S. Bethel (2004). Manual of Photogrammetry (5th ed.). Maryland, USA: American Society of Photogrammetry and Remote Sensing.

Mikhail, E. M. and F. Ackermann (1976). Observations and Least Squares. University Press of America
Mikhail, E. M., J. S. Bethel, and J. C. McGlone (2001). Introduction to Modern Photogrammetry. Wiley.
Palais, B. and R. Palais (2007). Euler’s fixed point theorem: The axis of a rotation. J. fixed point theory appl. 2, 215-220.
Raguram, R., O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm (2013). USAC: A Universal Framework for Random Sample Consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8), 2022-2038.
Rao, R. C. (1973). Linear Statistical Inference and Its Applications. New York: Wiley.
Rodriguez, O. (1840). Des lois géometriques qui regissent les deplacements d’un système solide independamant des causes qui peuvent les produire. Journal de mathématiques pures et appliquées 1 (5), 380440.
Sansò, F. (1973). An Exact Solution of the Roto-Translation Problem. Photogrammetria 29 (6), 203-206.
Vaseghi, S. V. (2000). Advanced Digital Signal Processing and Noise Reduction. Wiley.
Weber, M. (2003). Rotation between two vectors. Personal communication, Bonn.