RANSAC
Robust Model Estimation
From Data Contaminated By Outliers

Lecturer: Jiří Matas
Authors: Ondřej Chum, Jiří Matas, Ondřej Drbohlav
Czech Technical University in Prague
http://cmp.felk.cvut.cz
What is RANSAC?

- RANSAC = RANdom SAmple Consensus

- **Example**: Finding a line in 2D data.
  - Not all input points are on a line.
  - Finding a line also implicitly divides points to inliers (=those on a line) and outliers (=those not on a line).
Line Fitting: Line Parametrization

- Line parametrization (homogeneous):

  \[ ax + by + c = 0, \quad (a \neq 0 \lor b \neq 0) \]  \hspace{1cm} (1)

  \[ a, b, c \in \mathbb{R} : \text{line parameters} \]  \hspace{1cm} (2)

  \[ (x, y) : \text{point coordinates} \]  \hspace{1cm} (3)

- Line parametrization (radial):

  \[ x \cos \phi + y \sin \phi = r, \]  \hspace{1cm} (4)

  \[ \phi \in [0, \pi[, \quad r \in \mathbb{R} : \text{line parameters} \]  \hspace{1cm} (5)
Line Fitting: Line Parametrization and Residuals

- Line parameters: $\phi \in [0, \pi]$, $r \in \mathbb{R}$

- Point $\mathbf{x} = (x, y)$ on the line:
  
  $$x \cos \phi + y \sin \phi = r$$
  
  $$\Leftrightarrow \mathbf{x} \cdot (\cos \phi, \sin \phi) = r$$

- Point $\mathbf{x} = (x, y)$ not on the line:
  
  $$\mathbf{x} \cdot (\cos \phi, \sin \phi) \neq r$$

- Signed distance $\rho(\mathbf{x})$ from line:
  
  $$\rho(\mathbf{x}) = \mathbf{x} \cdot (\cos \phi, \sin \phi) - r$$

Note: $\mathbf{n} = (\cos \phi, \sin \phi)$ (thus $||\mathbf{n}|| = 1$)
Data points
\[ \mathcal{X} = \{ x_j, j = 1, 2, ..., N_p \} \quad (x_j \in \mathbb{R}^2) \]

Find the line which
“best fits” these points.
Data points
\[ \mathcal{X} = \{\mathbf{x}_j, j = 1, 2, \ldots, N_p \} \quad (\mathbf{x}_j \in \mathbb{R}^2) \]

Find the line which “best fits” the points.

As optimization: Find best line with parameters \( \theta^* \) as

\[ \theta^* = \arg\min_{\theta} \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) \]

For \( f_{LSQ}(\mathbf{x}, \theta) = [\rho(\mathbf{x})]^2 \)

this is easily solvable by Singular Value Decomposition (SVD).
General Case with Outliers, Example 1

Example 1

Least squares fit
General Case with Outliers, Example 2

Example 2

Least squares fit
General Case with Outliers, Robust Cost Function

- $\mathcal{X} = \{x_j\}_{j=1}^{N_p}$ set of data points

Find:

$$\theta^* = \arg \min_{\theta} \sum_{x \in \mathcal{X}} f(x, \theta)$$

$\theta = (r, \phi)$

- No outliers: $f_{LSQ}(x, \theta) = [\rho(x)]^2$

- Use instead:

$$f_{RANSAC}(x, \theta) = \begin{cases} 
0, & \text{if } \rho(x) \leq \text{threshold } \sigma \\
\text{const.}, & \text{otherwise}
\end{cases}$$

- Such cost function is non-convex

- How to find optimal line parameters?
Random Sample Consensus - RANSAC

Select sample of $m$ points at random (here $m=2$)
RANSAC

Select sample of $m$ points at random

Estimate model parameters from the data in the sample
RANSAC

Select sample of $m$ points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point
RANSAC

Select sample of $m$ points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis
RANSAC

Select sample of $m$ points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling
RANSAC

Select sample of \( m \) points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling
Select sample of $m$ points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling
RANSAC [Fischler and Bolles 1981]

**Input:** \( \mathcal{X} = \{x_j\}_{j=1}^N \) data points

\[ e(S) = \theta \] estimates model parameters \( \theta \) given sample \( S \subseteq \mathcal{X} \)

\[ f(x, \theta) = \begin{cases} 
0, & \text{if distance to model } \leq \text{ threshold } \sigma \\
1, & \text{otherwise}
\end{cases} \]

\[ \Rightarrow J(\theta) = \sum_{x \in \mathcal{X}} f(x, \theta) \text{ is } \#\text{outliers} \]

\( \eta \) – required confidence in the solution, \( \sigma \) – outlier threshold

**Output:** \( \theta^* \) parameter of the model minimizing the cost function

1: \( \text{iter} \leftarrow 0, J^* \leftarrow \infty \)

2: \( \text{repeat} \)

3: Select random \( S \subseteq \mathcal{X} \) (sample size \( m = |S| \))

4: Estimate parameters \( \theta = e(S) \)

5: Evaluate \( J(\theta) = \sum_{x \in \mathcal{X}} f(x, \theta) \)

6: If \( J(\theta) < J^* \) then

\[ \theta^* \leftarrow \theta, J^* \leftarrow J(\theta) \]

7: \( \text{iter} \leftarrow \text{iter} + 1 \)

8: \( \text{until} \ P(\text{better solution exists}) = f(|\mathcal{X}|, J^*, \text{iter}) < \eta \)

9: Compute \( \theta^* \) from all inliers \( \mathcal{X}_{in} \): \( \theta^* \leftarrow \text{LocalOptimization}(\mathcal{X}_{in}, \theta^*) \)
RANSAC – how many samples?

- \( N \) Number of points
- \( Q \) Number of inliers, \( Q = N - J^* \)
- \( m \) Size of sample
- \( \epsilon = Q/N \) Inlier ratio

Probability of all-inlier (uncontaminated) sample:

\[
P(\text{inlier sample}) = \frac{\binom{Q}{m}}{\binom{N}{m}} = \frac{Q(Q-1)...(Q-m+1)}{N(N-1)...(N-m+1)} \approx \epsilon^m
\]

Mean time for hitting all-inliers sample is proportional to \( 1/P \).
RANSAC – how many samples?

- How about this formulation:
  - Set the number of samples $k$ such that \textbf{at least one} pair of points from the line has been hit with probability larger than $\eta$
  - Equivalently ... such that \textbf{no} pair of points from the line has been hit with probability lower than $1 - \eta$

- $Q$ \hspace{1cm} Number of inliers, $Q = N - J^*$
- $m$ \hspace{1cm} Size of sample
- $\epsilon = Q/N$ \hspace{1cm} Inlier ratio

Probability of all-inlier (uncontaminated) sample:

$$P(\text{inlier sample}) = \binom{Q}{m} / \binom{N}{m} = \frac{Q(Q-1)...(Q-m+1)}{N(N-1)...(N-m+1)} \approx \epsilon^m$$

We require:

$$P(\text{bad pair } k \text{ times}) = (1-P(\text{inlier sample}))^k < 1 - \eta$$

Finding the solution with confidence $\eta$ therefore requires at least:

$$k \geq \log(1 - \eta) / \log \left(1 - \epsilon^m\right)$$
### Inlier ratio \( \epsilon = Q/N \) [%]

<table>
<thead>
<tr>
<th>Size of the sample ( m )</th>
<th>15%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>132</td>
<td>73</td>
<td>32</td>
<td>17</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5916</td>
<td>1871</td>
<td>368</td>
<td>116</td>
<td>46</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>1.75 \cdot 10^6</td>
<td>2.34 \cdot 10^5</td>
<td>1.37 \cdot 10^4</td>
<td>1827</td>
<td>382</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>1.17 \cdot 10^7</td>
<td>1.17 \cdot 10^6</td>
<td>4.57 \cdot 10^4</td>
<td>4570</td>
<td>765</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>2.31 \cdot 10^{10}</td>
<td>7.31 \cdot 10^8</td>
<td>5.64 \cdot 10^6</td>
<td>1.79 \cdot 10^5</td>
<td>1.23 \cdot 10^4</td>
<td>215</td>
</tr>
<tr>
<td>18</td>
<td>2.08 \cdot 10^{15}</td>
<td>1.14 \cdot 10^{13}</td>
<td>7.73 \cdot 10^9</td>
<td>4.36 \cdot 10^7</td>
<td>7.85 \cdot 10^5</td>
<td>1838</td>
</tr>
<tr>
<td>30</td>
<td>\infty</td>
<td>\infty</td>
<td>1.35 \cdot 10^{16}</td>
<td>2.60 \cdot 10^{12}</td>
<td>3.22 \cdot 10^9</td>
<td>1.33 \cdot 10^5</td>
</tr>
<tr>
<td>40</td>
<td>\infty</td>
<td>\infty</td>
<td>\infty</td>
<td>2.70 \cdot 10^{16}</td>
<td>3.29 \cdot 10^{12}</td>
<td>4.71 \cdot 10^6</td>
</tr>
</tbody>
</table>

computed for \( \eta = 0.95 \)
RANSAC Notes

Pros:
- extremely popular (>17000 citations in Google Scholar)
- used in many applications
- percentage of inliers not needed and not limited
- a probabilistic guarantee for the solution
- mild assumptions: \( \sigma \) known

Cons:
- slow if inlier ratio low
- It was observed experimentally that RANSAC takes several times longer than theoretically expected. This is due to noise – not every all-inlier sample generates a good hypothesis:

\[
P(\text{inlier sample}) \neq P(\text{good model estimate})
\]
RANSAC Issues, Variants

- **Cost function**: MLESAC, Huber loss, ...
- **Outlier threshold** $\sigma$: Least median of Squares, MINPRAN, ...

- **Correctness of the results. Degeneracy.**
  Solution: DegenSAC.

- **Accuracy** (parameters are estimated from minimal samples).
  Solution: Locally Optimized RANSAC

- **Speed**: Running time grows with
  1. number of data points,
  2. number of iterations (polynomial in the inlier ratio)

Addressing the problem:
RANSAC with SPRT (WaldSAC), PROSAC
Locally Optimized RANSAC (LO-RANSAC): Problem Intro

Data: 200 points
LO-RANSAC: Problem Introduction

Data: 200 points
Model, 100 inliers
LO-RANSAC: Problem Introduction

For simplicity, consider only points belonging to the model (100 points)
For simplicity, consider only points belonging to the model (100 points)

RANSAC
Hypothesis generation from 2 points
Will every two points generate the whole inlier set?
This sample: YES. 100 inliers.
For simplicity, consider only points belonging to the model (100 points)

RANSAC

Hypothesis generation from 2 points

Will every two points generate the whole inlier set?

This sample: NO. 45 inliers.
LO-RANSAC: Problem Introduction

For simplicity, consider only points belonging to model (100 points)

The distribution of the number of inliers obtained while randomly sampling points pairs

RANSAC

Hypothesis generation from 2 points

Will every two points generate the whole inlier set?
LO-RANSAC

Input: \( \mathcal{X} = \{x_j\}_{j=1}^N \) data points

\[ e(S) = \theta \]

estimates model parameters \( \theta \) given sample \( S \subseteq \mathcal{X} \)

\[ f(x, \theta) = \begin{cases} 
0, & \text{if distance to model} \leq \text{threshold} \sigma \\
1, & \text{otherwise}
\end{cases} \]

\[ \Rightarrow J(\theta) = \sum_{x \in \mathcal{X}} f(x, \theta) \text{ is \#outliers} \]

\( \eta \) – required confidence in the solution, \( \sigma \) – outlier threshold

Output: \( \theta^* \) parameter of the model minimizing the cost function

1: iter \( \leftarrow 0, J^* \leftarrow \infty \)

2: repeat

3: Select random \( S \subseteq \mathcal{X} \) (sample size \( m = |S| \))

4: Estimate parameters \( \theta = e(S) \)

5: Evaluate \( J(\theta) = \sum_{x \in \mathcal{X}} f(x, \theta) \)

6: If \( J(\theta) < J^* \) then

\[ \theta^* \leftarrow \theta, J^* \leftarrow J(\theta) \]

7: iter \( \leftarrow \) iter + 1

8: until \( P(\text{better solution exists}) = f(|\mathcal{X}|, J^*, \text{iter}) < \eta \)

9: Compute \( \theta^* \) from all inliers \( \mathcal{X}_{in} \): \( \theta^* \leftarrow \text{LocalOptimization}(\mathcal{X}_{in}, \theta^*) \)

12/03/2018

J. Matas, MPV Course
**LO-RANSAC**

**Input:** \( \mathcal{X} = \{\mathbf{x}_j\}_{j=1}^N \) data points

\[
e(S) = \theta \quad \text{estimates model parameters } \theta \text{ given sample } S \subseteq \mathcal{X}
\]

\[
f(\mathbf{x}, \theta) = \begin{cases} 
0, & \text{if distance to model } \leq \text{ threshold } \sigma \\
1, & \text{otherwise} 
\end{cases}
\]

\[
\Rightarrow J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) \text{ is } \#\text{outliers}
\]

\( \eta \) – required confidence in the solution, \( \sigma \) – outlier threshold

**Output:** \( \theta^* \) parameter of the model minimizing the cost function

1: \( \text{iter} \leftarrow 0, J^* \leftarrow \infty \)
2: **repeat**
3: Select random \( S \subseteq \mathcal{X} \) (sample size \( m = |S| \))
4: Estimate parameters \( \theta = e(S) \)
5: Evaluate \( J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta) \)
6: If \( J(\theta) < J^* \) then
   \[
   \theta^* \leftarrow \text{LocalOptimization}(\mathcal{X}_{in}, \theta), J^* \leftarrow J(\theta^*)
   \]
7: \( \text{iter} \leftarrow \text{iter} + 1 \)
8: **until** \( P(\text{better solution exists}) = f(|\mathcal{X}|, J^*, \text{iter}) < \eta \)
9: **gone**
LO-RANSAC: Example

inliers count = 60

12/03/2018

J. Matas, MPV Course
LO-RANSAC: Example

inliers count = 60

Init
Iteration 1
LO-RANSAC: Example

inliers count = 59

Init
Iteration 1
Iteration 2
LO-RANSAC: Example

Init
Iteration 1
Iteration 2
...
Iteration 7

inliers count = 67
LO-RANSAC: Example

inliers count = 97

Init
Iteration 1
Iteration 2
...
Iteration 7
...
Iteration 15
LO-RANSAC: Example

Comparison with model (100 inliers):

$x$ vs. $y$ scatter plot with fitted line, showing 100 inliers and 90 outliers.
Locally Optimized RANSAC

Estimation of (approximate) models with lower complexity (less data points in the sample) followed by LO step estimating the desired model speeds the estimation up significantly.

The estimation of epipolar geometry is up to 10000 times faster when using 3 region-to-region correspondences rather than 7 point-to-point correspondences.

Simultaneous estimation of radial distortion and epipolar geometry with LO is superior to the state-of-the-art in both speed and precision of the model.

Fish-eye images by Braňo Mičušík

Chum, Matas, Obdržálek: Enhancing RANSAC by Generalized Model Optimization, ACCV 2004
LO-RANSAC: Problem Summary

It was observed experimentally that RANSAC takes several times longer than theoretically expected. This is due to the noise – not every all-inlier sample generates a good hypothesis.

By applying local optimization (LO) to the-best-so-far hypotheses:
(i) a near perfect agreement with theoretical performance
(ii) lower sensitivity to noise and poor conditioning.

The LO is shown to be executed so rarely, \( \log(\text{iter}) \) times, that it has minimal impact on the execution time.
RANSAC – Time Complexity

Repeat $k$ times ($k$ is a function of $\eta$, $Q$, $N$)

1. Hypothesis generation
   • Select a sample of $m$ data points
   • Calculate parameters of the model(s)

2. Model verification
   • Find the support (consensus set) by
     • verifying all $N$ data points

$T_{total}$ running time:

$$t = k(t_M + \overline{m}_s N)$$
Randomised RANSAC [Matas, Chum 02]

Repeat \( k/(1-\alpha) \) times

1. Hypothesis generation
2. Model pre-verification \( T_{d, d} \) test
   - Verify \( d \ll N \) data points, reject
   - the model if not all \( d \) data points
   - are consistent with the model
3. Model verification
   - Verify the rest of the data points

\( V \) – average number of data points verified

\( \alpha \) – probability that a good model is rejected by \( T_{d, d} \) test

\[
t = \frac{k}{1 - \alpha} \left( t_M + \bar{m}_s V \right)
\]
Model Verification is Sequential Decision Making

\[ H_g: \ P(x_i = 1|H_g) \geq \varepsilon \]
\[ H_b: \ P(x_i = 1|H_b) = \delta \]
\[ x_i = 1 \quad x_i \text{ is consistent with the model} \]

where

\( H_g \) - hypothesis of a `good` model \((\approx \text{from an uncontaminated sample})\)
\( H_b \) - hypothesis of a `bad` model, \((\approx \text{from a contaminated sample})\)
\( \delta \) - probability of a data point being consistent with an arbitrary model

Optimal (the fastest) test that ensures with probability \( \alpha \) that that \( H_g \) is not incorrectly rejected is the Sequential probability ratio test (SPRT) [Wald47]
SPRT [simplified from Wald 47]

Compute the likelihood ratio

\[ \lambda_i = \prod_{j=1}^{i} \frac{P(x_j|H_b)}{P(x_j|H_g)} \]

if \( \lambda_i > A \) reject the model
if \( i = N \) accept model as 'good'

Two important properties of SPRT:

1. probability of rejecting a \( \text{good} \) model \( \alpha < 1/A \)
2. average number of verifications \( V = C \log(A) \)

\[ C \approx \left( P(0|H_b) \log \frac{P(0|H_b)}{P(0|H_g)} + P(1|H_b) \log \frac{P(1|H_b)}{P(1|H_g)} \right)^{-1} \]
SPRT properties

1. Probability of rejecting a \textit{good} model $\alpha = 1/A$

$$\lambda_i = \prod_{j=1}^{i} \frac{P(x_j|H_b)}{P(x_j|H_g)} = \frac{P(x|H_b)}{P(x|H_g)}, x = (x_1, \ldots, x_i)$$

If $\lambda_i > A$ then $P(x|H_g) < P(x|H_b)/A$, therefore

$$\alpha = \int_{\lambda_i > A} P(x|H_g)dx < \int_{\lambda_i > A} \frac{P(x|H_b)}{A}dx =$$

$$= \frac{1}{A} \int_{\lambda_i > A} P(x|H_b)dx \leq \frac{1}{A} \int P(x|H_b)dx = \frac{1}{A}$$
WaldSAC

Repeat \( k/(1-1/A) \) times

1. Hypothesis generation
2. Model verification, use SPRT

\[
\bar{m}_s \cdot C \log A \\
C \approx ((1 - \delta) \log \frac{1 - \delta}{1 - \varepsilon} + \delta \log \frac{\delta}{\varepsilon})^{-1}
\]

\[
t(A) = \frac{k}{(1 - 1/A)}(t_M + \bar{m}_SC \log A)
\]

In sequential statistical decision problem decision errors are traded off for time. These are two incomparable quantities, hence the constrained optimization.

In WaldSAC, decision errors cost time (more samples) and there is a single minimised quantity, time \( t(A) \), a function of a single parameter \( A \).
Optimal test (optimal A) given $\varepsilon$ and $\delta$

Optimal $A^*$

\[ A^* = \arg \min_A t(A) \]

Optimal $A^*$ found by solving

\[ \frac{\partial t}{\partial A} = 0 \]

\[ A^* = \frac{t_M}{m_s C} + 1 + \log A^* \]

\[ A^* = \lim_{n \to \infty} A_n \]

\[ A_0 = \frac{t_M}{m_s C} + 1, \quad A_{n+1} = \frac{t_M}{m_s C} + 1 + \log A_n \]
Note: the Wald’s test is equivalent to series of $T(d, c)$, where $c = \left( \frac{\log A - d \log \lambda_1}{\log \lambda_0} \right)$
## Exp. 1: Wide-baseline matching

<table>
<thead>
<tr>
<th></th>
<th>samples</th>
<th>models</th>
<th>V</th>
<th>time</th>
<th>spd-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2914</td>
<td>7347</td>
<td>110.0</td>
<td>1099504</td>
<td>1.0</td>
</tr>
<tr>
<td>R-R</td>
<td>7825</td>
<td>19737</td>
<td>3.0</td>
<td>841983</td>
<td>1.3</td>
</tr>
<tr>
<td>Wald</td>
<td>3426</td>
<td>8648</td>
<td>8.2</td>
<td>413227</td>
<td>2.7</td>
</tr>
</tbody>
</table>
Exp. 2 Narrow-baseline stereo

<table>
<thead>
<tr>
<th></th>
<th>samples</th>
<th>models</th>
<th>V</th>
<th>time</th>
<th>spd-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>155</td>
<td>367</td>
<td>600.0</td>
<td>235904</td>
<td>1.0</td>
</tr>
<tr>
<td>R-R</td>
<td>247</td>
<td>587</td>
<td>86.6</td>
<td>75539</td>
<td>3.1</td>
</tr>
<tr>
<td>Wald</td>
<td>162</td>
<td>384</td>
<td>23.1</td>
<td>25032</td>
<td>9.4</td>
</tr>
</tbody>
</table>
Randomised Verification in RANSAC: Conclusions

- The same confidence $\eta$ in the solution reached faster (data dependent, $\approx 10x$)
- No change in the character of the algorithm, it was randomised anyway.
- Optimal strategy derived using Wald`s theory for known $\varepsilon$ and $\delta$.
- Results with $\varepsilon$ and $\delta$ estimated during the course of RANSAC are not significantly different. Performance of SPRT is insensitive to errors in the estimate.
  - $\delta$ can be learnt, an initial estimate can be obtained by geometric consideration
  - Lower bound on $\varepsilon$ is given by the best-so-far support
  - Note that the properties of WaldSAC are quite different from preemptive RANSAC!
• Not all correspondences are created equally
• Some are better than others
• Sample from the best candidates first

Sample from here
Draw $T_l$ samples from $(1 \ldots l)$
Draw $T_{l+1}$ samples from $(1 \ldots l+1)$

Samples from $(1 \ldots l)$ that are not from $(1 \ldots l+1)$ contain $l+1$

Draw $T_{l+1} - T_l$ samples of size $m-1$ and add $l+1$
Degenerate Configurations

The presence of degenerate configuration causes RANSAC to fail in estimating a correct model, instead a model consistent with the degenerate configuration and some outliers is found.

The DEGENSAC algorithm handles scenes with:

- all points in a single plane
- majority of the points in a single plane and the rest off the plane
- no dominant plane present

No a-priori knowledge of the type of the scene is required

**Chum, Werner, Matas**: Epipolar Geometry Estimation unaffected by dominant plane, *CVPR 2005*