RANSAC

Robust model estimation from data contaminated by outliers

Ondřej Chum
Fitting a Line

Least squares fit
• Select sample of m points at random
RANSAC

- Select sample of $m$ points at random

- Calculate model parameters that fit the data in the sample
RANSAC

• Select sample of m points at random

• Calculate model parameters that fit the data in the sample

• Calculate error function for each data point
RANSAC

• Select sample of m points at random

• Calculate model parameters that fit the data in the sample

• Calculate error function for each data point

• Select data that support current hypothesis
RANSAC

- Select sample of m points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- Select data that support current hypothesis
- Repeat sampling
RANSAC

- Select sample of \( m \) points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- Select data that support current hypothesis
- Repeat sampling
RANSAC

- Select sample of m points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- Select data that support current hypothesis
- Repeat sampling
How Many Samples?

On average

\[ N \quad \text{... number of points} \]
\[ I \quad \text{... number of inliers} \]
\[ m \quad \text{... size of the sample} \]

\[ P(\text{good}) = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I - j}{N - j} \]

mean time before the success
\[ E(k) = \frac{1}{P(\text{good})} \]
How Many Samples?

With confidence $p$

How large $k$?

...to hit at least one pair of points on the line $l$ with probability larger than $p$ (0.95)

Equivalently

...the probability of not hitting any pair of points on $l$ is $\leq 1 - p$
How Many Samples?

With confidence \( p \)

\[ N \quad \ldots \quad \text{number of point} \]

\[ I \quad \ldots \quad \text{number of inliers} \]

\[ m \quad \ldots \quad \text{size of the sample} \]

\[
P(\text{good}) = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I - j}{N - j}
\]

\[ P(\text{bad}) = 1 - P(\text{good}) \]

\[ P(\text{bad } k \text{ times}) = \left(1 - P(\text{good})\right)^k \]
How Many Samples?

With confidence $p$

$$P(\text{bad } k \text{ times}) = \left(1 - P(\text{good})\right)^k \leq 1 - p$$

$$k \log \left(1 - P(\text{good})\right) \leq \log(1 - p)$$

$$k \geq \log(1 - p) / \log \left(1 - P(\text{good})\right)$$
## How Many Samples

$$I/N[\%]$$

<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>
RANSAC

\[ k = \frac{\log(1 - p)}{\log \left( 1 - \frac{I}{N} \frac{I-1}{N-1} \right)} \]

- \( k \) … number of samples drawn
- \( N \) … number of data points
- \( I \) … time to compute a single model
- \( p \) … confidence in the solution (.95)
**RANSAC [Fischler, Bolles ’81]**

**In:** $U = \{x_i\}$ set of data points, $|U| = N$

$f(S) : S \rightarrow p$ function $f$ computes model parameters $p$ given a sample $S$ from $U$

$\rho(p, x)$ the cost function for a single data point $x$

**Out:** $p^*$ $p^*$, parameters of the model maximizing the cost function

$k := 0$

Repeat until $P\{\text{better solution exists}\} < \eta$ (a function of $C^*$ and no. of steps $k$)

$k := k + 1$

I. Hypothesis

(1) select randomly set $S_k \subset U$, sample size $|S_k| = m$

(2) compute parameters $p_k = f(S_k)$

II. Verification

(3) compute cost $C_k = \sum_{x \in U} \rho(p_k, x)$

(4) if $C^* < C_k$ then $C^* := C_k, p^* := p_k$

end
Advanced RANSAC

**In:** $U = \{x_i\}$ set of data points, $|U| = N$

$f(S) : S \rightarrow p$ function $f$ computes model parameters $p$ given a sample $S$ from $U$

$\rho(p, x)$ the cost function for a single data point $x$

**Out:** $p^*$ $p^*$, parameters of the model maximizing the cost function

$k := 0$

Repeat until $P\{\text{better solution exists}\} < \eta$ (a function of $C^*$ and no. of steps $k$)

$k := k + 1$

I. Hypothesis

(1) select randomly set $S_k \subset U$, sample size $|S_k| = m$

(2) compute parameters $p_k = f(S_k)$

II. Verification

(3) compute cost $C_k = \sum_{x \in U} \rho(p_k, x)$

(4) if $C^* < C_k$ then $C^* := C_k$, $p^* := p_k$

end

Non-uniform sampling in PROSAC

Many models are bad, no need to verify all data points – RANDOMIZED RANSAC

Improving precision by Local Optimization
RANSAC Makes an Invalid Assumption

Not every all-inlier sample gives a model consistent with all inliers

Lower number of inliers is detected

RANSAC runs longer
Solution: Local Optimisation Step

Repeat k times
1. Hypothesis generation
2. Model verification
   2b. If model best-so-far Execute (Local) Optimisation

Inner RANSAC + Re-weighted least squares:
- Samples are drawn from the set of data points consistent with the best-so-far hypothesis
- New models are verified on all data points
- Samples can contain more than minimal number of data points since consistent points include almost entirely inliers

How often?

\[ \sum_{l=1}^{k} P_l = \sum_{l=1}^{k} \frac{1}{l} \leq \int_{1}^{k} \frac{1}{x} \, dx + 1 = \log k + 1 \]

Conclusion: the LO step 2b is executed rarely, does not influence running time significantly
Validation: Two-view Geometry Estimation

Histograms of the number of inliers returned over 100 executions of RANSAC (top) and LO-RANSAC (bottom)

Result:
(i) variation of the number of inliers significantly reduced
(ii) speed-up up to 3 times (for 7pt EG and 4pt homography est.)
PROSAC – PROgressive SAmple Consensus

• 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 … l)
Draw $T_{l+1}$ samples from (1 … l+1)

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

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

- RANSAC is a standard tool in computer vision
- it is a simple procedure
  - hypothesize and verify loop
- handles large number of outliers
- a number of advanced strategies to
  - increase the stability
  - speed up
- Vanilla RANSAC never used in practice