RANSAC

Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Praha 2, Karlovo nám. 13, Czech Republic

hlavac@fel.cvut.cz, http://cmp.felk.cvut.cz

LECTURE PLAN

Courtesy: O. Chum, J. Matas

1.

2.

3.

RANSAC

[Fischler, Bolles 1981]

- RANSAC = Random Sampling and Consensus.
- One of the most cited papers in computer vision.

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

RANSAC ALGORITHM

$k := 0$

Repeat until P{better solution exists} $< \eta$
(a function of C^* and number of steps k)

$k := k + 1$

I. Hypothesis

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

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

II. Verification

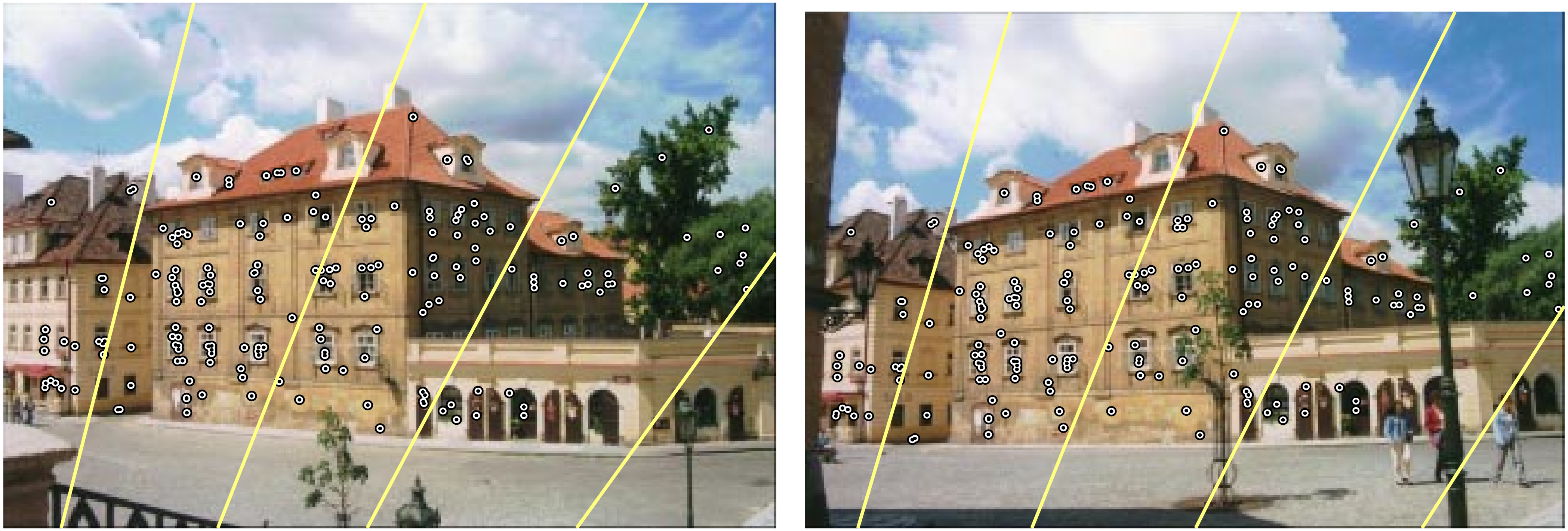
(3) compute cost $C_k = \sum_{x \in U} \rho(p_k, x)$

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

end

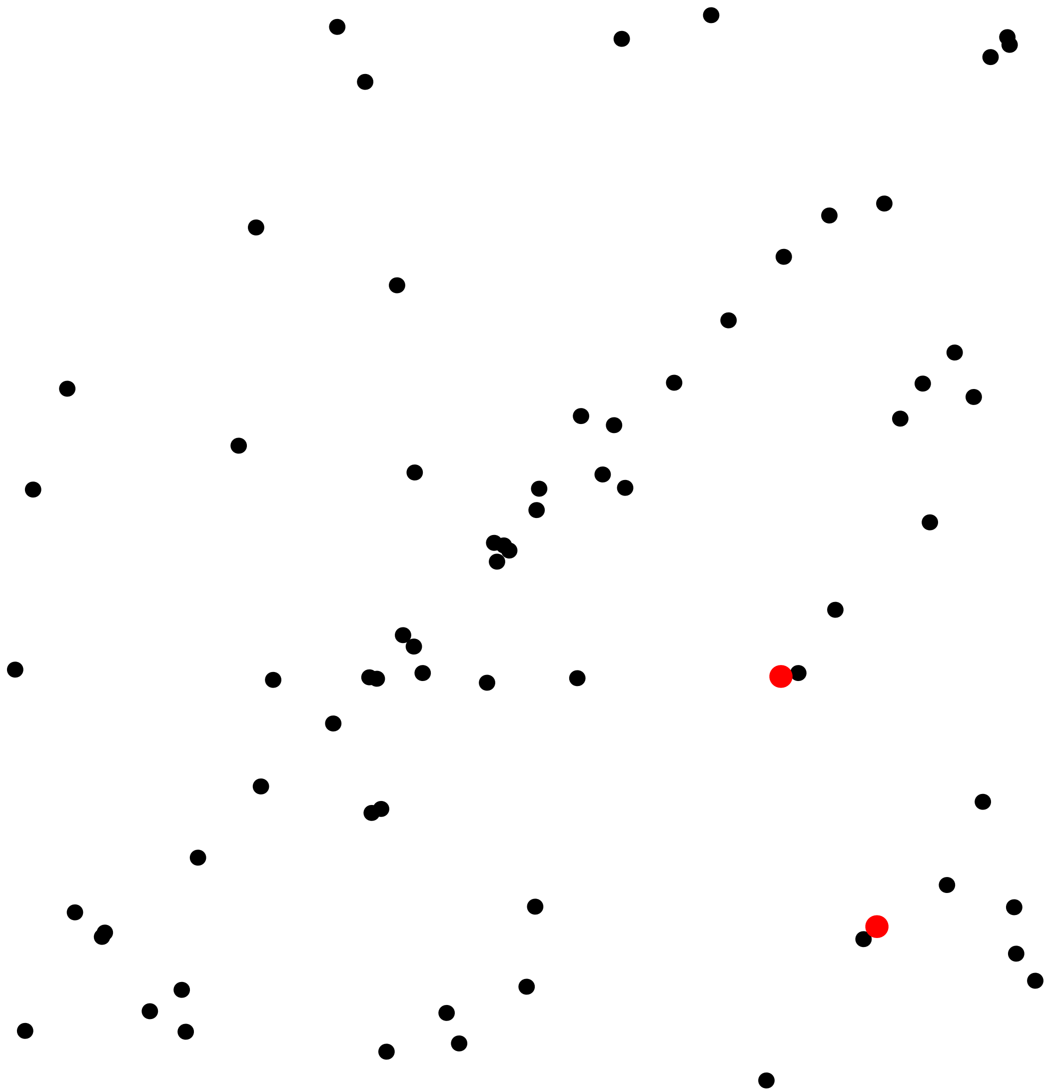
Example I: Epipolar geometry estimation by RANSAC

- Data points U: a set of correspondences, i.e., pairs of 2D points
- Sample size $m = 7$
- Model parameters f: seven-point algorithm - gives 1 to 3 independent solutions
- Cost function ρ: thresholded Sampson's error



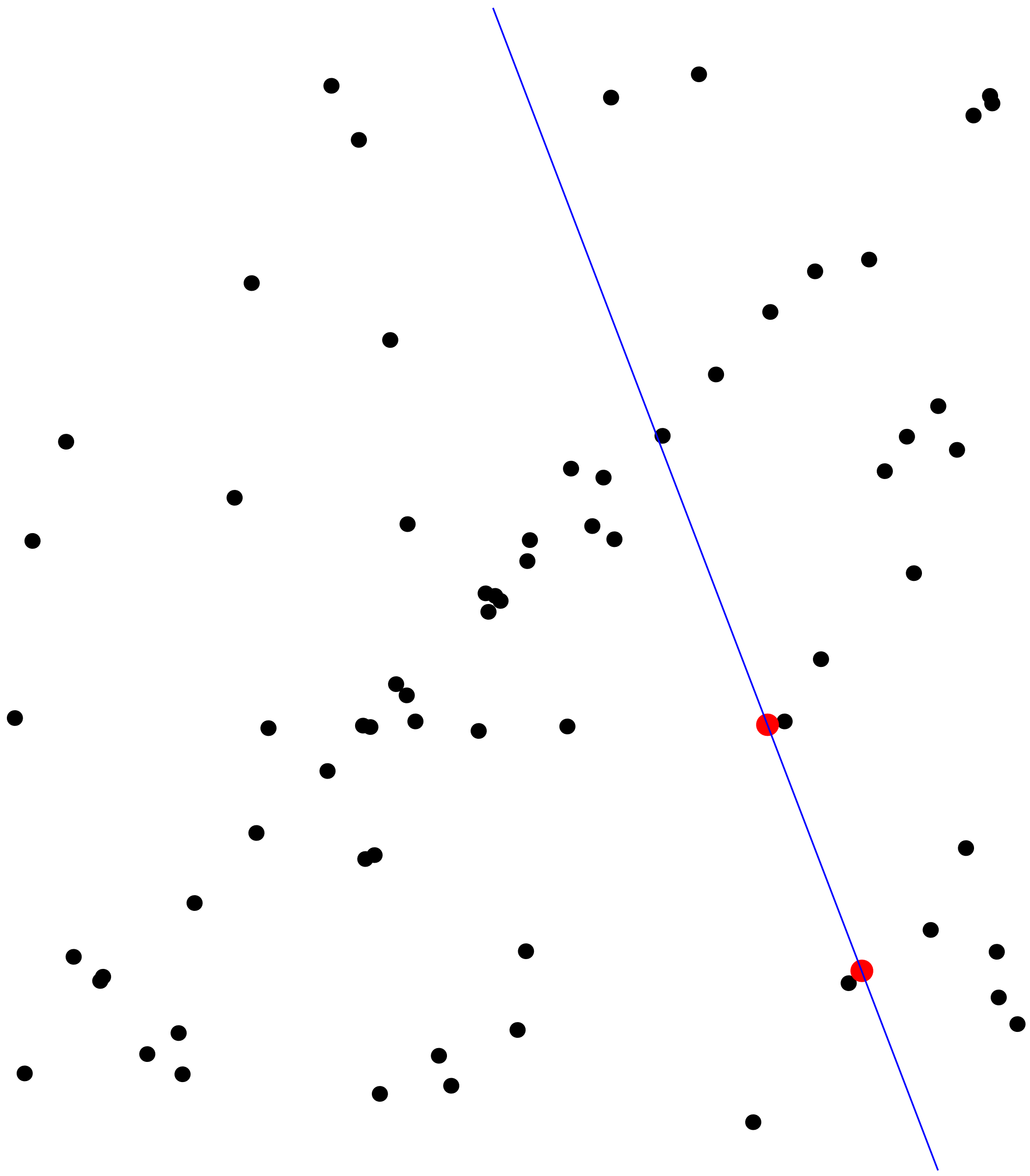
Example II: Line detection by RANSAC.

Example II: Line detection by RANSAC.



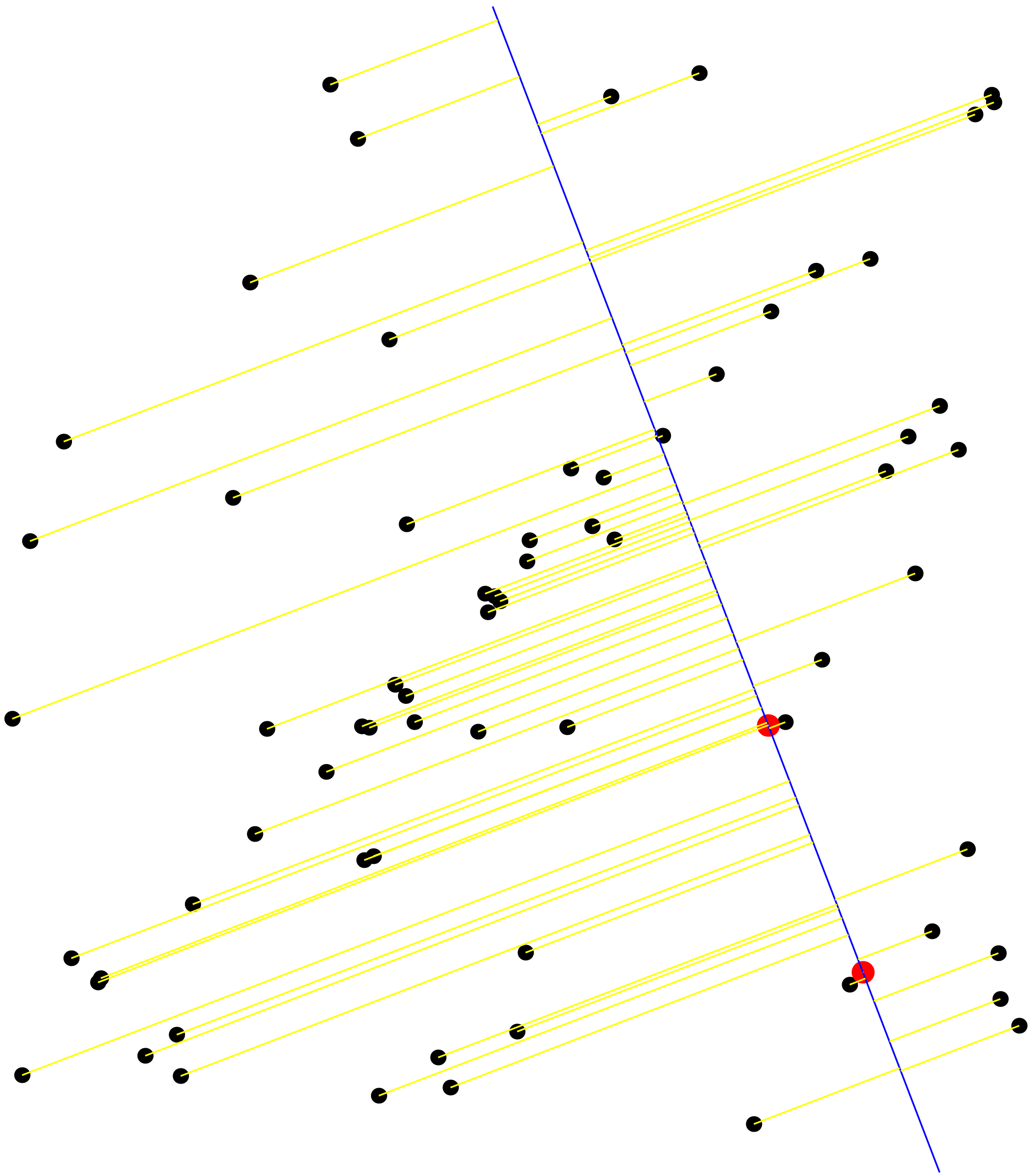
- Select randomly two points.

Example II: Line detection by RANSAC.



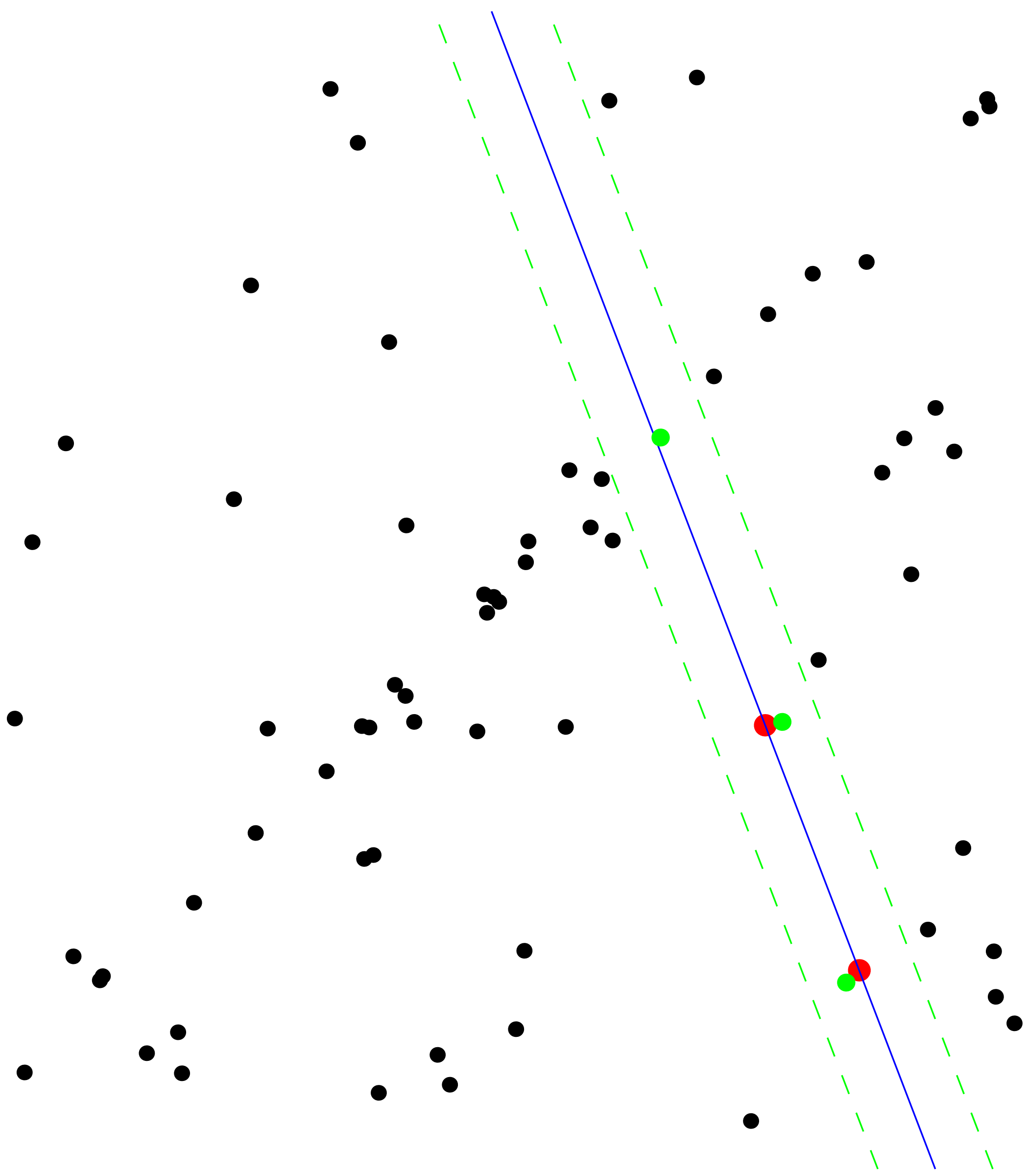
- Select randomly two points.
- The hypothesised model is the line passing through the two points.

Example II: Line detection by RANSAC.



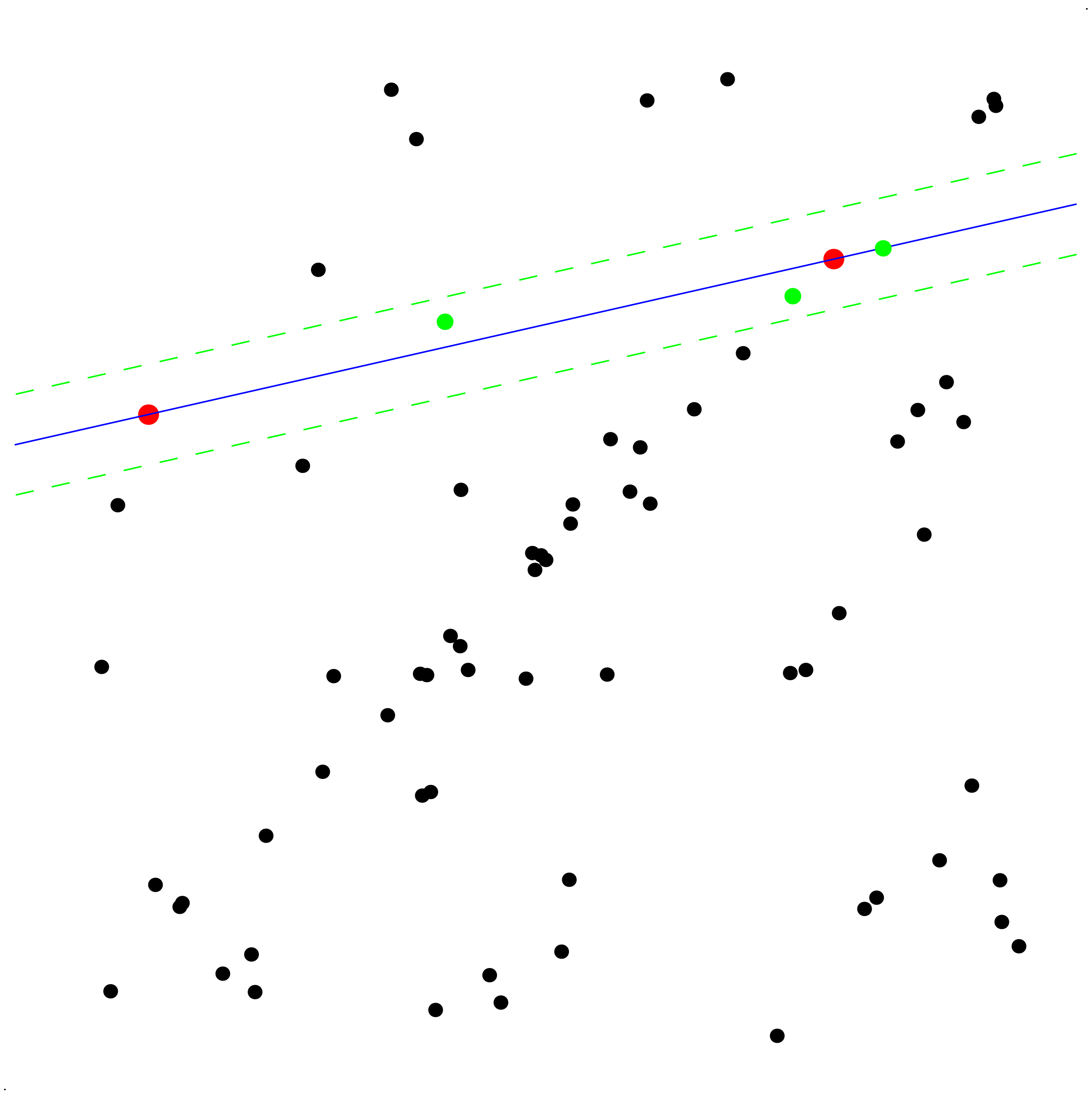
- Select randomly two points.
- The hypothesised model is the line passing through the two points.
- The error function is a distance from the line.

Example II: Line detection by RANSAC.

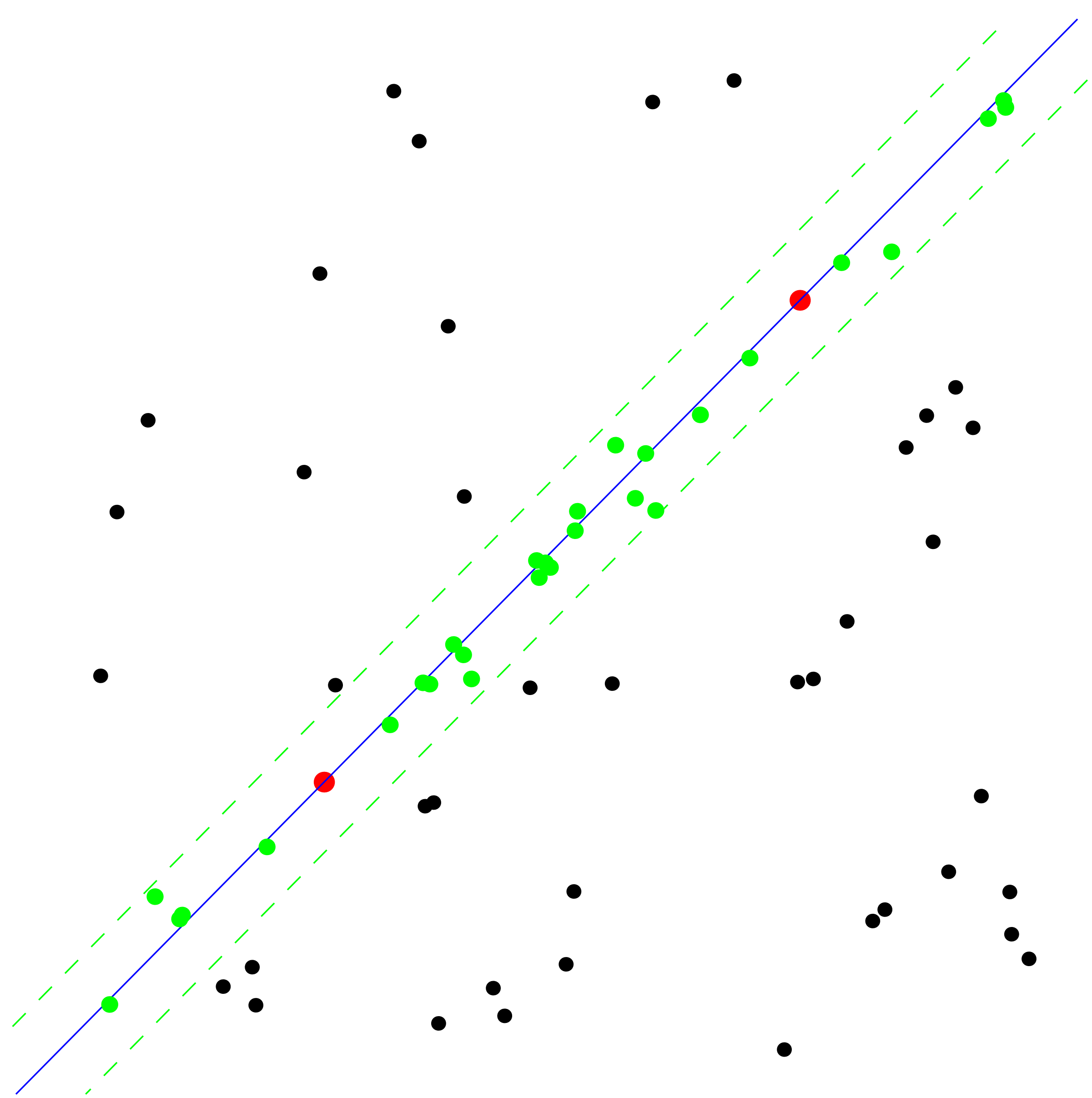


- Select randomly two points.
- The hypothesised model is the line passing through the two points.
- The error function is a distance from the line.
- Points consistent with the model.

Example II: Line detection by RANSAC.



RANSAC Time Complexity



Uncontaminated sample

RANSAC time: $J = k(t_M + N)$

Depends on:

- N - number of data points
- ε - fraction of inliers
- m - size of the sample
- ε^m - probability that uncontaminated sample is selected
- $k = 1/\varepsilon^m$ - the average number of samples before uncontaminated one
- t_M - time to calculate the model

The Number of Data Points

- For each hypothesised model, all the data points are verified.
- The more data points the longer RANSAC takes.
- The majority of samples are contaminated.

Solution:

- Randomize the verification

Randomized RANSAC Algorithm

$k := 0$

Repeat until P{better solution exists} $< \eta$

$k := k + 1$

I. Hypothesis

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

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

II. Preverification

(3) perform test based on $d \ll N$ data points

(4) continue verification only if the test is passed

III. Verification

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

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

end

Preverification Test Analysis

GOOD SAMPLE

P_I $= \varepsilon^m$ probability that all of m sampled data points are inliers, i.e. the probability of taking a good sample

α probability that the good sample passes the preverification test

$\bar{t}_\alpha$ the average number of data points tested in false negative test

$\bar{k}$ $= 1/(\varepsilon^m \alpha)$ the average number of samples before first good sample

CONTAMINATED SAMPLE

P_O $= 1 - \varepsilon^m$ probability that at least one sampled data point is outlier
$P_O \gg P_I$

β the probability that contaminated sample passes the test
$\beta \ll \alpha$

$\bar{t}_\beta$ the average number of data points tested in correct negative test
$\bar{t}_\beta \ll N$

average R-RANSAC time:

$$J = \frac{1}{\varepsilon^m \alpha}\Big(t_M + P_I\ (\alpha\, N + (1-\alpha)\, \bar{t}_\alpha) + P_O\ (\beta\, N + (1-\beta)\, \bar{t}_\beta)\Big)$$

Standard RANSAC:

$$\alpha = 1, \qquad \beta = 1 \qquad \Longrightarrow \qquad J = \frac{1}{\varepsilon^m}(t_M + N)$$

Choosing the Preverification Test : The $T_{d,d}$ Test

$\mathbf{T}_{d,d}$ definition: test passed if all d randomly selected data points from $U \setminus S$ are consistent with the model parameters p

The time spent on the R-RANSAC with $T_{d,d}$

$$J(T_{d,d}) = \frac{1}{\varepsilon^m \, \varepsilon^d} \left(t_M + \varepsilon^m \left(\varepsilon^d N + \frac{1-\varepsilon^d}{1-\varepsilon} \right) + (1-\varepsilon^m) \left(\delta^d N + \frac{1-\delta^d}{1-\delta} \right) \right)$$

where δ is the probability that data point is consistent with a "random" model

$$\alpha = \varepsilon^d$$
$$\beta = \delta^d$$

Optimal d^* satisfying $\frac{\partial J(T_{d,d})}{\partial d} = 0$

$$d^* = \ln \left(\frac{\ln \varepsilon (t_M + 1)}{N \, (\ln \delta - \ln \varepsilon)} \right) / \ln \delta$$

Randomized RANSAC is faster than the standard one, if $J(T_{0,0}) > J(T_{1,1})$

$$N > (t_M + 1) \frac{1-\varepsilon}{\varepsilon - \delta}$$

Epipolar Geometry Estimation: Synthetic Experiment

- 1500 correspondences, 900 outliers, 600 inliers

d	samples	models	tests	inliers	time
0	1866	4569	6821218	600	25.0
1	4717	11536	16311	600	6.0
2	11849	28962	33841	600	15.1

Epipolar Geometry Estimation: Short Baseline Experiment

- 676 correspondences, approx. 60% of inliers
- tentative corr. by Harris operator and cross-correlation
- Leuven castle dataset

d	samples	models	tests	inliers	time
0	480	1146	766875	343	2.6
1	960	2301	83953	342	1.4

Epipolar Geometry Estimation: Wide Baseline Experiment

- 413 correspondences, less than 40% of inliers
- tentative corr. by WBS algorithm [Matas, Chum, Urban, Pajdla '01]
- BOOKSHELF dataset

d	samples	models	tests	inliers	time
0	3094	7582	3078184	161	12.9
1	6366	15583	178217	164	8.7

CONCLUSIONS

- Benefits of RANSAC randomization of hypothesis verification studied.
- A statistical preverification test was proposed.
- The increased performance experimentally verified.

Remaining problems

- The parameter d is fixed at 1, not optimal
- Estimating ε, δ and consequently d during the sampling process