RANSAC
RANdom SAmple Consensus

Tomáš Svoboda, svoboda@cmp.felk.cvut.cz
courtesy of Ondřej Chum, Jiří Matas
Czech Technical University in Prague, Center for Machine Perception
http://cmp.felk.cvut.cz
Last update: November 19, 2007;

Talk Outline
■ importance for computer vision
■ principle
■ line fitting
■ epipolar geometry estimation

Importance for Computer Vision
■ published in 1981 as a model fitting method [2]
■ one of the most cited papers in computer vision
■ widely accepted as a method that works even for difficult computer vision problems
■ recent advancement presented at the “25-years of RANSAC” workshop1. Look at the R. Bowless' presentation.


LSQ does not work for gross errors

2. sketch borrowed from [3]
RANSAC motivations for computer vision

- gross errors (outliers) spoil LSQ estimation
- detection (localization) algorithms in computer vision do have gross error
- in difficult problems the portion of good data may be even less than $1/2$
- standard robust estimation techniques hardly applicable to data with less than $1/2$ good

RANSAC inputs and output

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

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

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

RANSAC algorithm

$k := 0$

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

1. Hypothesis

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

   (2) compute parameters $\theta_k = f(S_k)$

2. Verification

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

   (4) if $C^* < C_k$ then $C^* := C_k$, $\theta^* := \theta_k$

end
• Randomly select two points

● Randomly select two points
● The hypothesised model is the line passing through the two points
Explanation example: line detection

- Randomly select 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

Probability of selecting uncontaminated sample in $K$ trials

- $N$ - number of data points
- $w$ - fraction of inliers
- $s$ - size of the sample

Prob. of selecting a sample with all inliers$^3$: $\approx w^s$
Prob. of not selecting a sample with all inliers: $1 - w^s$
Prob. of not selecting a good sample $K$ times: $(1 - w^s)^K$

The sought probability of selecting uncontaminated sample in $K$ trials at least once: $P = \frac{1}{K} (1 - w^s)^K$

$^3$Approximation valid for $s \ll N$, see the lecture notes
How many samples are needed

How many trials is needed to select an uncontaminated sample with a given probability $P$? We derived $P = 1 - (1 - w^s)^K$. Log the both sides to get

$$K = \frac{\log(1 - P)}{\log(1 - w^s)}$$

Real problem—$w$ unknown

Often, the proportion of inliers in data cannot be estimated in advance.

Adaptive estimation: start with worst case and update the estimate as the computation progress

- set $K = \infty$, $\#\text{samples} = 0$, $P$ very conservative, say $P = 0.99$
- while $K > \#\text{samples}$ repeat
  - choose a random sample, compute the model and count inliers
    - $w = \frac{\#\text{inliers}}{\#\text{data points}}$
    - $K = \frac{\log(1 - P)}{\log(1 - w^s)}$
    - increment $\#\text{samples}$
- terminate

Epipolar geometry estimation by RANSAC

- $U$: a set of correspondences, i.e. pairs of 2D points
data points
- $s = 7$
sample size
- $f$: seven-point algorithm - gives 1 to 3 independent solutionsmodel parameters
- $\rho$: thresholded Sampson’s errorcost function
Besides the main reference [2] the Huber’s book [5] about robust estimation is also widely recognized. The RANSAC algorithm received several essential improvements in recent years [1, 6, 7].

For the seven-point algorithm and Sampson’s error, see [4]