RANSAC (Random Sample Consensus) - Overview

Key Concepts to Master

1. What is RANSAC?

Purpose:

- Robust parameter estimation in presence of outliers
- Fit models (lines, planes, transformations) to noisy data
- Ignore outliers automatically

Core idea:

- Randomly sample minimum points needed for model
- Fit model to sample
- Count how many points agree (inliers)
- Repeat many times
- Keep model with most inliers

2. Algorithm Overview

Input:

- Data points (some are outliers)
- Model to fit
- Distance threshold (inlier criterion)
- Number of iterations

Output:

- Best model parameters
- Set of inliers

Basic Algorithm:

```
best_model = null
best_inliers = 0

For k iterations:
    1. Randomly select minimum sample (e.g., 2 points for line)
    2. Fit model to sample
    3. Count inliers (points within threshold distance)
    4. If inliers > best_inliers:
         best_inliers = inliers
         best_model = model

Return best_model
```

3. Key Components

Minimum sample size:

- Line (2D): 2 points
- Plane (3D): 3 points

- Homography: 4 point pairs
- Fundamental matrix: 8 point pairs

Distance threshold:

- Defines what counts as inlier
- Too small → few inliers
- Too large → outliers accepted
- Typically: 1-3 pixels for image data

Number of iterations:

- More iterations → higher probability of success
- Depends on outlier ratio
- Formula: k = log(1-p) / log(1-w^n)
    - p = desired probability of success (e.g., 0.99)
    - w = estimated inlier ratio (e.g., 0.5)
    - n = minimum sample size

4. Outlier Rejection Strategy

Why traditional least squares fails:

```
Least squares: minimize Σ(error²)
Problem: Outliers have huge squared errors → dominate fit
```

RANSAC approach:

```
Ignore outliers completely
Only fit to inliers
Robust to 50%+ outliers
```

5. Model Fitting Example: Line

Traditional least squares (sensitive to outliers):

```
All points influence fit equally
Outliers pull line away from true solution
```

RANSAC (robust):

```
Iteration 1: Sample points {A, B}
  → Fit line L₁
  → Count inliers = 10

Iteration 2: Sample points {C, outlier}
  → Fit line L₂
  → Count inliers = 2

...

Best line from iteration with most inliers
```

Detailed Algorithm Steps

Step 1: Random Sampling

```
Select n random points (n = minimum for model)
Example for line: randomly pick 2 points
```

Step 2: Model Fitting

```
Fit model to selected points
For line: y = mx + b computed from 2 points
```

Step 3: Consensus Set

```
For each data point:
    distance = perpendicular distance to model
    if distance < threshold:
        point is inlier
```

Step 4: Evaluation

```
Count number of inliers
If count > best_so_far:
    Save this model as best
```

Step 5: Refinement (optional)

```
Refit model using all inliers from best model
Improves accuracy
```

Example Calculation

Problem: Fit line to 10 points, 3 are outliers

Given:

- Inlier ratio w = 0.7 (70% inliers)
- Desired success probability p = 0.99
- Minimum sample n = 2

Calculate iterations needed:

```
k = log(1-p) / log(1-w^n)
k = log(1-0.99) / log(1-0.7²)
k = log(0.01) / log(0.51)
k = -4.605 / -0.673
k ≈ 7 iterations
```

MATLAB Quick Reference

```
% RANSAC for line fitting
function [best_model, inliers] = ransac_line(points, ...
                                  threshold, k)
    best_inliers = [];
    best_model = [];

    for i = 1:k
        % 1. Random sample
        idx = randperm(size(points,1), 2);
        sample = points(idx, :);

        % 2. Fit line
        model = fit_line(sample);

        % 3. Find inliers
        distances = point_to_line_dist(points, model);
        inliers = find(distances < threshold);

        % 4. Evaluate
        if length(inliers) > length(best_inliers)
            best_inliers = inliers;
            best_model = model;
        end
    end

    % 5. Refinement
    best_model = fit_line(points(best_inliers, :));
end

% Built-in MATLAB RANSAC (for various models)
[model, inliers] = fitgeotform2d(matchedPoints1, ...
    matchedPoints2, 'affine', 'MaxDistance', threshold);
```

Study Checklist

- [ ] Understand purpose: robust fitting with outliers
- [ ] Know basic RANSAC algorithm
- [ ] Understand minimum sample sizes
- [ ] Can calculate number of iterations needed
- [ ] Know role of distance threshold
- [ ] Understand why it's robust
- [ ] Can explain difference from least squares

Common Exam Questions

Q1: What does RANSAC do?

- Robust parameter estimation ignoring outliers

Q2: What is the basic algorithm?

1. Random sample minimum points
2. Fit model
3. Count inliers
4. Repeat, keep best

Q3: How many points to fit a line?

- Minimum 2 points

Q4: What determines number of iterations?

- Outlier ratio
- Desired success probability
- k = log(1-p) / log(1-w^n)

Q5: Why is RANSAC robust?

- Outliers don't influence model fit
- Only uses inliers
- Can handle 50%+ outliers

Q6: What is consensus set?

- Set of inliers for a model
- Points within threshold distance

Advantages

- Robust to high outlier ratios (>50%)
- General purpose (works for any model)
- Simple to implement

Disadvantages

- Non-deterministic (random sampling)
- Requires threshold parameter
- Computational cost (many iterations)

Answers to Common Exam Questions

Q1: What does RANSAC do? RANSAC performs robust parameter estimation (e.g., fitting a line or homography) when the data contains outliers. Unlike least squares which lets outliers dominate the fit, RANSAC randomly samples minimal subsets, fits a model to each, and keeps the model that has the most inlier support.

Q2: What is the basic algorithm?

1. Randomly select the minimum number of points needed to fit the model (e.g., 2 for a line)
2. Fit the model to those points
3. Count inliers: all data points within a distance threshold of the model
4. If this inlier count is the best so far, save the model
5. Repeat for k iterations
6. (Optional) Refit the best model using all its inliers for improved accuracy

Q3: How many points to fit a line? Minimum **2 points** define a line. For other models: 3 points for a plane, 4 point pairs for a homography, 8 point pairs for a fundamental matrix.

Q4: What determines the number of iterations? `k = log(1 - p) / log(1 - w^n)` where p = desired success probability (e.g., 0.99), w = inlier ratio, n = minimum sample size. Higher outlier ratio (lower w) or larger sample size (n) requires exponentially more iterations. Example: w=0.5, n=2, p=0.99 gives k = log(0.01)/log(0.75) = 16 iterations.

Q5: Why is RANSAC robust? Outliers never influence the model fit because models are fit only to small random samples. An outlier can only affect a trial if it is selected in the random sample, and such trials will have few inliers and be discarded. RANSAC can handle over 50% outliers.

Q6: What is the consensus set? The set of all data points that are inliers for a given model -- i.e., points whose distance to the model is below the threshold. The model with the largest consensus set is selected as the best fit.

Related Topics

- Lecture 9: RANSAC
- Feature matching (generates data with outliers)
- Hough Transform (alternative robust method)
- Homography estimation