

RANSAC Formula Explanation

The formula for the number of iterations N in RANSAC comes from probability theory and is designed

to ensure a high likelihood of finding at least one subset of points (a sample) that does not include any outliers.

Let me break it down step by step:

What Are We Trying to Achieve?

We want to randomly select subsets of data points in such a way that we have a high probability (p) of selecting at least one subset containing only inliers (no outliers). The formula calculates how many iterations (N) are required to achieve this.

Assumptions:

- Probability of selecting an inlier: Let u be the probability that a single data point is an inlier.

The probability of selecting an outlier is $v = 1 - u$.

- Size of a random sample: To estimate the model, we need a minimum of m data points (e.g., 2 points for a line, 3 points for a plane, etc.).

Probability of picking a good sample:

For a random sample of size m , all m points must be inliers.

The probability of selecting all inliers in one random sample is: u^m

(This is the probability of selecting an inlier raised to the power m , since all m points must be inliers.)

Probability of selecting a bad sample:

A "bad sample" is one that includes at least one outlier.

The probability of selecting a bad sample is: $1 - u^m$

Goal:

We want the probability of selecting only bad samples over N iterations to be very small. In other words:

$$(1 - u^m)^N$$

This represents the probability of selecting bad samples N times in a row.

If we want at least one good sample with high probability (p), we require:

$$1 - (1 - u^m)^N \geq p$$

This ensures that the probability of at least one good sample is p or higher.

Deriving the Formula:

Rearrange the equation:

$$(1 - u^m)^N \leq 1 - p$$

Take the natural logarithm (\ln) of both sides:

$$N * \ln(1 - u^m) \leq \ln(1 - p)$$

Solve for N :

$$N \geq \ln(1 - p) / \ln(1 - u^m)$$

Substitute $u = 1 - v$ (where v is the probability of an outlier):

$$N \geq \ln(1 - p) / \ln(1 - (1 - v)^m)$$

This is the formula for the number of iterations N .

Key Takeaways:

- N increases as:

- The percentage of inliers (u) decreases (because it's harder to find subsets with only inliers).
- The number of points (m) required for the model increases (since larger subsets are less likely to contain only inliers).
- p (the desired probability of success) is typically set to 0.99 for a 99% chance of finding a good sample.