

RANSAC

RANSAC (RANdom SAmple Consensus) 是一种迭代的随机算法，用于在数据中找到最合适的模型参数，特别是在数据中包含大量噪声和异常值（离群点）的情况下。RANSAC 算法的目标是从含有大量异常值的数据中估计出能够描述内点的模型。RANSAC 在计算机视觉和图像处理领域中非常常用，比如图像配准、特征匹配、3D重建等。

RANSAC 算法的步骤

1. 设定模型参数：

- 根据任务设定模型的参数。例如，在图像匹配中，可能是两个图像间的基础矩阵或本质矩阵的参数；在平面拟合中，可能是平面的法向量和偏移量。

2. 随机采样：

- 从数据中随机选择一个最小样本集，该样本集应该足以确定模型的参数。比如，在图像配准中，如果拟合一个基础矩阵，需要至少选择8对匹配点。

3. 模型估计：

- 使用采样得到的最小样本集，估计出模型参数。

4. 计算一致性：

- 根据估计的模型，对所有数据点计算模型的残差或误差，判断哪些点与该模型一致。
- 一致性通常通过设定一个阈值来判断，如果某个数据点与模型的距离小于阈值，则认为该点与模型一致，称为**内点**；否则为**外点**。

5. 记录最优模型：

- 如果当前模型的内点数量大于之前的最优模型，则将当前模型作为最优模型，记录下内点的数量和模型参数。

6. 重复上述步骤：

- 重复以上步骤多个迭代，以找到内点数量最多的模型。迭代次数通常根据数据中异常值的比例和置信度来确定。

7. 输出结果：

- 经过迭代后，输出包含内点数量最多的模型参数，即最符合数据的模型。

公式

假设数据中包含的内点比例为 p ，要以置信度 $1 - \alpha$ 找到一个仅包含内点的样本集，RANSAC 所需的最少迭代次数 N 可通过以下公式计算：

$$N = \frac{\log(\alpha)}{\log(1 - p^s)} \quad (1)$$

其中：

- s 是每次采样中需要的最小样本点数量（例如，拟合基础矩阵时 $s = 8$ ）。
- p 是内点比例的估计值。

参数

- **最小样本集大小 s** ：能够唯一确定模型的最小样本数量，例如，基础矩阵估计需要 8 对点。
- **阈值 ϵ** ：判断数据点是否属于模型的误差阈值，需根据任务和噪声情况选择。
- **最大迭代次数**：用于限制算法的运行时间，若达到最大次数则停止迭代。

优点和缺点

- **优点：**
 - 能够有效处理包含大量异常值的数据。
 - 比较高效，尤其在数据集包含大量异常值时。
 - 可以适用于多种模型拟合问题，例如直线拟合、平面拟合、几何变换等。
- **缺点：**

- 结果不稳定，可能因为随机采样导致不同结果，因此通常通过多次运行取平均来减小影响。
- 对于内点比例较低的数据，需要更多的迭代才能找到合适的模型，计算成本可能较高。
- 阈值的选择较为关键，如果不当会影响内外点的区分效果。

RANSAC 应用示例

1. **图像匹配**：在特征匹配中，可以使用 RANSAC 去除错误匹配的特征点。例如，在估计两个图像间的基础矩阵或本质矩阵时，RANSAC 可以过滤掉错误的匹配点，保留可靠的点对。
2. **几何模型拟合**：如直线拟合、平面拟合、圆拟合等。在二维或三维空间中，通过 RANSAC 可以拟合出包含内点的几何模型，同时排除噪声和异常点。
3. **3D 重建和定位**：在多视图几何中，RANSAC 经常用于估计相机的姿态和位置，重建三维结构。

总结

RANSAC 是一种在数据中存在大量异常值的情况下进行模型拟合的鲁棒算法。它通过随机采样和一致性检测，选择出内点数量最多的模型，最终找到一个适合于内点的模型参数。