RANSAC

노이즈가 심한 측정 데이터로부터 모델 파라미터 결정

Introduction

Fischler 과 Bolles 에 의해서 제안된 RANSAC ("RANdom SAmple Consensus") 알고리즘은 측정 노이즈가 심한 원본 데이터로부터 모델 파라미터를 예측하는 방법이다.

RANSAC 은 전체 원본 데이터 중에서 모델 파라미터를 결정하는데 필요한 최소의 데이터를 랜덤하게 샘플링하면서 반복적으로 해를 계산함으로써 최적의 해를 찾는다. 이 방법은 전통적인 통계적 방법과는 반대의 개념을 가진다. 즉, 대부분의 방법들이 초기의 해를 획득하기 위해서 가능한 많은 데이터를 사용하고 그 결과로부터 유효하지 않은 데이터를 제거한다. 반면에 이 방법은 가능한 적은 양의 초기 데이터를 사용해서 일관된 데이터의 집합(consensus set)을 확장시켜가는 방식을 사용한다.

RANSAC 알고리즘

RANSAC 알고리즘은 주어진 원본 데이터에서 일부를 임의로 선택한 후 최적의 파라미터를 예측하는 과정을 반복하면서 좋은 모델 파라미터를 찾는다.

알고리즘 실행 순서

입력 데이터가 M 개 있고, 모델의 파라미터를 예측하는데 N 개의 데이터가 필요한 경우, 알고리즘은 다음과 같다.

I. Hypothesis:

1. 원본 데이터에서 임의로 N 개의 셈플 데이터를 선택한다.
2. 이 데이터를 정상적인 데이터로 보고 모델 파라미터를 예측한다.

II. Verification:

3. 원본 데이터가 예측된 모델에 잘 맞는지 검사한다.

3-1. 만일 원본 데이터가 유효한 데이터인 경우, 유효한 데이터 집합에 더한다.

4. 만일 예측된 모델이 잘 맞는다면, 유효한 데이터 집합으로부터 새로운 모델을 구한다.

수식으로 표현

Definition:

$U = \langle x_i \mid i = 1, \cdots, N \rangle$: 원본 데이터 집합

$p \leftarrow \mathrm{f}(S)$: 함수 $\mathrm{f}()$ 는 U 로부터 랜덤하게 선택한 데이터 S 로 모델 파라미터 p 를 계산

$\rho(p, x_i)$: 하나의 데이터 x_i로부터 코스트를 계산하는 함수

p^* : 코스트를 최대화 하는 모델 파라미터

Algorithm:

$k := 0$

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

(a function of C^* and no. of steps k)

$k := k + 1$

// I. Hypothesis:

1. select randomly set $S_k \subset U, \quad |S_k| = m$

2. compute parameters $p_k = \mathrm{f}(S_k)$

// 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

Pseudo code

Given:

data - a set of observed data points

model - a model that can be fitted to data points

m - the minimum number of data values required to fit the model

N - the maximum number of iterations allowed in the algorithm

t - a threshold value for determining when a data point fits a model

d - the number of close data values required to assert that a model fits well to data

Return:

bestfit - model parameters which best fit the data (or nil if no good model is found)

```
iterations = 0
bestfit = nil
besterr = something really large
while iterations < N {
    maybeinliers = m randomly selected values from data
    maybemodel = model parameters fitted to maybeinliers
    alsoinliers = empty set
    for every point in data not in maybeinliers {
        if point fits maybemodel with an error smaller than t
            add point to alsoinliers
    }
```

```
    if the number of elements in alsoinliers is > d {
        % this implies that we may have found a good model
        % now test how good it is
        bettermodel = model parameters fitted to all points in maybeinliers and alsoinliers
        thiserr = a measure of how well model fits these points
        if thiserr < besterr {
            bestfit = bettermodel
            besterr = thiserr
        }
    }
    increment iterations
}
return bestfit
```

반복횟수 N 의 결정

반복횟수 N 은 확률 p (일반적으로 0.99 로 설정)를 보장할 수 있도록 충분히 높게 선택되어야 한다. 확률 p 는 최소한 하나의 샘플 집합이 유효한 데이터만을 포함할 확률이다. 데이터의 유효한 확률을 u 라고 하면, 데이터의 유효하지 않을 확률은 $v=1-u$ 이다. 샘플 데이터 수 m 에 대한 반복 횟수 N 은 다음과 같다.

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

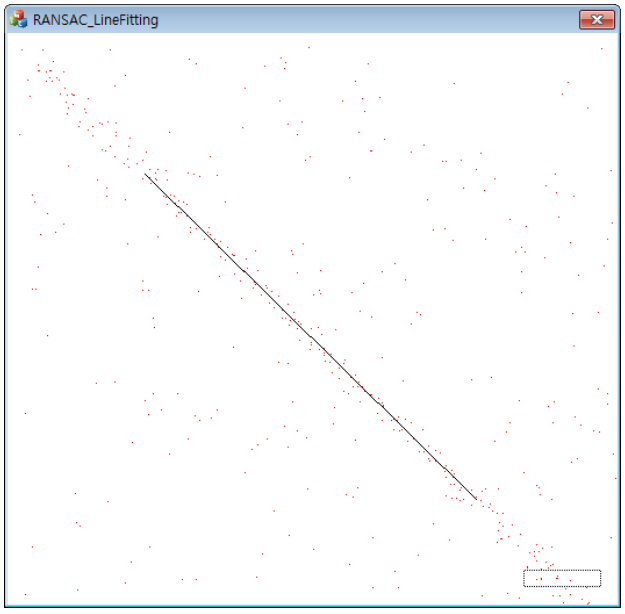
n samples are all inliers

a failure

failure after k trials

$$N=\frac{\log(1-p)}{\log(1-u^m)}$$

Line Fitting Application using RANSAC



주어진 500 개의 입력 데이터(inlier 250 개, outlier 250 개)에서 임의로 3 개의 점을 셈플링 하여 모델 파라미터를 예측한다.

반복횟수는 $N = \frac{\log(1-0.99)}{\log(1-0.5^3)} = 34$ 로 계산된다.

윈도우의 붉은 점들은 입력데이터를 나타내고 검은색 라인은 예측된 모델 파라미터로부터 그린 라인이다.