

## Les algorithmes de RANSAC

RANSAC (RANdom SAmple Consensus) est une méthode fréquemment appliquée pour la segmentation de forme géométrique dans un nuage de points.

RANSAC est un algorithme d'estimation robuste qui a été créé et utilisé par Fischler et Bolles afin de détecter des entités géométriques en 2D mais a été réadapté par la suite à la 3D.

L'algorithme RANSAC est une méthode itérative pour estimer les paramètres d'un certain modèle à partir d'un ensemble de données observées.

L'algorithme RANSAC propose une méthode stochastique, dans laquelle des ensembles minimaux de points sont tirés aléatoirement, et des formes candidates sont générées à partir de ces ensembles minimaux. Cependant, Si le tirage des points au sein du nuage est effectué de manière purement uniforme, la grande majorité des formes candidates construites ne sera pas pertinente, et le nombre de tentatives nécessaires à la reconnaissance d'une forme peut être énorme.

RANSAC atteint son objectif en sélectionnant itérativement un sous-ensemble aléatoire des données d'origine. Ces données sont d'*hypothétiques données pertinentes* et cette hypothèse est ensuite testée comme suit :

1. Un modèle est ajusté aux données pertinentes hypothétiques, c'est-à-dire que tous les paramètres libres du modèle sont estimés à partir de ce sous-ensemble de données.
2. Toutes les autres données sont ensuite testées sur le modèle précédemment estimé. Si un point correspond bien au modèle estimé, alors il est considéré comme une donnée pertinente candidate
3. Le modèle estimé est considéré comme correct si suffisamment de points ont été classés comme données pertinentes candidates.
4. Le modèle est réestimé à partir de ce sous-ensemble des données pertinentes candidates.
5. Finalement, le modèle est évalué par une estimation de l'erreur des données pertinentes par rapport au modèle.

Cette procédure est répétée un nombre fixe de fois, chaque fois produisant soit un modèle qui est rejeté parce que trop peu de points sont classés comme données pertinentes, soit un modèle réajusté et une mesure d'erreur correspondante. Dans ce dernier cas, on conserve le modèle réévalué si son erreur est plus faible que le modèle précédent.

## Principe de l'algorithme

RANSAC se compose essentiellement de trois étapes. Celles-ci seront décrites par rapport à la segmentation d'un ensemble de points P en plusieurs sous-ensembles de points coplanaires.

1. **Proposition** : Dans cette étape, l'algorithme tire aléatoirement un échantillon E (composé de trois points pour définir un plan) à partir de l'ensemble des données P. Un modèle n est alors estimé à partir de cet ensemble

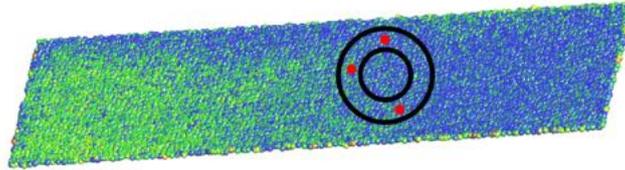


Figure 1: Détails de notre algorithme de RANSAC.

2. **Calcul de la fonction coût** : Une fonction coût est calculée, qui représente le critère que le modèle doit optimiser. En considérant que le critère représente le nombre de points consistants avec un plan, l'expression du modèle optimal peut être écrite comme suit :

$$\mathbf{n}_{opt} = \arg \max_{\mathbf{n}} \mathcal{J}(\mathbb{P}, \mathbf{n})$$

3. **Classement du modèle** : Les modèles sont classés en fonction du nombre de points consistants avec chaque modèle. Le modèle maximisant la fonction coût  $\mathcal{J}(\mathbb{P}, \mathbf{n})$  est considéré comme le meilleur.

$$\mathcal{J}(\mathbb{P}, \mathbf{n}) = |\mathbb{P}_{in}|$$

Où  $|\mathbb{P}_{in}|$  définit le cardinal de l'ensemble de points consistants

Ces trois étapes sont itérées jusqu'à ce que la probabilité de trouver un meilleur modèle devienne faible. En général, cela revient à définir un nombre d'itérations k à partir duquel on dit que le modèle le mieux classé est optimal. Le nombre d'itérations est estimé en considérant q la probabilité de tirer un échantillon non-contaminé par des outliers, et donc  $(1 - q)$  la probabilité de tirer un échantillon contaminé par au moins un outlier.

$$q = \frac{C_{N_I}^3}{C_N^3} \approx \left(\frac{N_I}{N}\right)^3$$

Où –  $N_I$  proportion des inliers.

– N cardinalité de l'ensemble P.

– 3 cardinalités de l'ensemble E.

Si on effectue  $k$  tirages successifs, la probabilité que les échantillons soient tous contaminés par au moins un outliers est  $(1 - q)^k$ .

Ainsi, plus  $k$  est grand plus cette quantité tend vers zéro.

On considère un seuil de probabilité  $\epsilon_{th}$  à partir duquel on estime que l'échantillon minimal n'est pas contaminé :

$$(1 - q)^k \leq \epsilon_{th}$$

On déduit le nombre d'itérations :  $k \approx \text{abs}(\log(\epsilon_{th}) / \log(1 - q))$

### **RANSAC Avantages et inconvénient**

Av :

- ✚ Traiter de manière robuste les valeurs aberrantes
- ✚ Fonctionne bien pour 1 à environ 10 paramètres (selon le nombre de valeurs aberrantes)
- ✚ Facile à comprendre et à implémenter

Inc :

- ✚ Pas de limite supérieure sur le temps de calcul des paramètres. En effet, lorsqu'une limite est utilisée (un nombre maximal d'itérations), la solution obtenue peut ne pas être la solution optimale.
- ✚ Ne peut estimer qu'un seul modèle à un ensemble de données particulier. Comme pour toute approche à modèle unique, lorsque deux (ou plusieurs) modèles coexistent, RANSAC peut ne parvenir à trouver ni l'un ni l'autre.

### **Références :**

<https://hal.archives-ouvertes.fr/tel-01275382/document>

[https://www.youtube.com/watch?v=kDEFUiBV48c&ab\\_channel=DariusBurschka](https://www.youtube.com/watch?v=kDEFUiBV48c&ab_channel=DariusBurschka)

[https://www.youtube.com/watch?v=Cu1f6vpEilg&t=1385s&ab\\_channel=CyrillStachniss](https://www.youtube.com/watch?v=Cu1f6vpEilg&t=1385s&ab_channel=CyrillStachniss)

[https://www.youtube.com/watch?v=yLlSj4yVpss&ab\\_channel=KevinRobbDesigns](https://www.youtube.com/watch?v=yLlSj4yVpss&ab_channel=KevinRobbDesigns)