

# RANSAC

**RANSAC** est une abréviation pour "RANdom SAMple Consensus". Il s'agit d'une méthode itérative pour estimer les paramètres d'un modèle mathématique à partir d'un ensemble de données observées qui contient des valeurs aberrantes (« outliers »). Il s'agit d'un algorithme non-déterministe dans le sens où il produit un résultat correct avec une certaine probabilité seulement, celle-ci augmentant à mesure que le nombre d'itérations est grand. L'algorithme a été publié pour la première fois par Fischler et Bolles en 1981<sup>1</sup>.

L'hypothèse de base est que les données sont constituées de « inliers », à savoir les données dont la distribution peut être expliquée par un ensemble de paramètres d'un modèle, et de « outliers » qui sont des données qui ne correspondent pas au modèle choisi. De plus, les données peuvent être soumises au bruit. Les valeurs aberrantes peuvent venir, par exemple, des valeurs extrêmes du bruit, de mesures erronées ou d'hypothèses fausses quant à l'interprétation des données. RANSAC suppose également que, étant donné un ensemble (généralement petit) d'*inliers*, il existe une procédure qui permet d'estimer les paramètres d'un modèle de telle façon à expliquer de manière optimale ces données.

## Sommaire

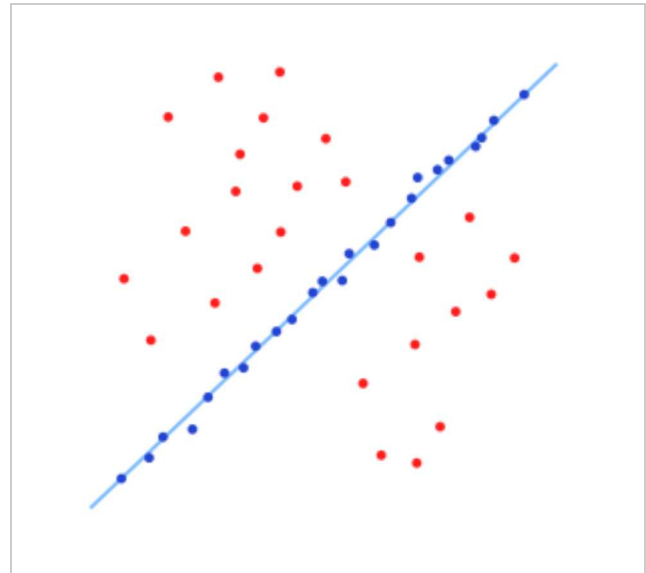
- 1 Exemple
- 2 Présentation
- 3 L'algorithme
- 4 Les paramètres
- 5 Avantages et inconvénients
- 6 Applications
- 7 Références
- 8 Liens externes

## Exemple

Un exemple simple est l'ajustement d'une ligne 2D à une série d'observations. On suppose que cet ensemble contient à la fois des inliers, c'est-à-dire, les points qui peuvent être approximativement ajustés à une ligne, et les outliers (valeurs aberrantes), les points qui sont éloignés de ce modèle de ligne. Un simple traitement par une méthode des moindres carrés donnera une ligne qui est mal ajustée aux inliers. En effet, la droite s'ajustera de manière optimale à tous les points, y compris les valeurs aberrantes (outliers). RANSAC, par contre, peut générer un modèle qui ne tiendra compte que des inliers, à condition que la probabilité de choisir seulement que des inliers dans le choix des données soit suffisamment élevée. Cependant, il n'y a aucune garantie d'obtenir cette situation, et il existe un certain nombre de paramètres de l'algorithme qui doivent être soigneusement choisis pour maintenir ce niveau de probabilité suffisamment élevée.



Un jeu de données avec de nombreuses valeurs aberrantes pour lequel une ligne doit être ajustée.



Ligne ajustée avec la méthode RANSAC, les valeurs aberrantes n'ont aucune influence sur le résultat.

## Présentation

Les données d'entrée de l'algorithme RANSAC sont un ensemble de valeurs des données observées, un modèle paramétré qui peut expliquer ou être ajusté aux observations, et des paramètres d'intervalle de confiance.

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

1. Un modèle est ajusté aux inliers hypothétiques, c'est-à-dire que tous les paramètres libres du modèle sont estimés à partir de cet 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 un inlier candidat.
3. Le modèle estimé est considéré comme correct si suffisamment de points ont été classés comme inliers candidats.
4. Le modèle est ré-estimé à partir de cet ensemble des inliers candidats.
5. Finalement, le modèle est évalué par une estimation de l'erreur des inliers 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 inliers, 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.

## L'algorithme

L'algorithme RANSAC générique, en pseudocode, fonctionne comme suit:

```
entrées :
data - un ensemble d'observations
modele - un modèle qui peut être ajusté à des données
n - le nombre minimum de données nécessaires pour ajuster le modèle
k - le nombre maximal d'itérations de l'algorithme
t - une valeur seuil pour déterminer si une donnée correspond à un modèle
d - le nombre de données proches des valeurs nécessaires pour faire valoir que le modèle correspond
```

```

'sorties :
    meilleur_modèle - les paramètres du modèle qui correspondent le mieux aux données (ou zéro si aucun)
    meilleur_ensemble_points - données à partir desquelles ce modèle a été estimé
    meilleure_erreur - l'erreur de ce modèle par rapport aux données

itérateur := 0
meilleur_modèle := aucun
meilleur_ensemble_points := aucun
meilleure_erreur := infini
'tant que itérateur < k
    points_aléatoires := n valeurs choisies au hasard à partir des données
    modèle_possible := paramètres du modèle correspondant aux points_aléatoires
    ensemble_points := points_aléatoires

    Pour chaque point des données pas dans points_aléatoires
        si le point s'ajuste au modèle_possible avec une erreur inférieure à t
            Ajouter un point à ensemble_points

    si le nombre d'éléments dans ensemble_points est > d
        (ce qui implique que nous avons peut-être trouvé un bon modèle,
         on teste maintenant dans quelle mesure il est correct)
        modèle_possible := paramètres du modèle réajusté à tous les points de ensemble_points
        erreur := une mesure de la manière dont ces points correspondent au modèle_possible
        si erreur < meilleure_erreur
            (nous avons trouvé un modèle qui est mieux que tous les précédents,
             le garder jusqu'à ce qu'un meilleur soit trouvé)
            meilleur_modèle := modèle_possible
            meilleur_ensemble_points := ensemble_points
            meilleure_erreur := erreur

    incrémentation de l'itérateur

retourne meilleur_modèle, meilleur_ensemble_points, meilleure_erreur

```

Variantes possibles de l'algorithme RANSAC :

- Arrêter la boucle principale si un modèle assez bon a été trouvée, c'est-à-dire avec une erreur suffisamment petite. Peut sauver du temps de calcul, au détriment d'un paramètre supplémentaire.
- Calculer *erreur* directement à partir de *modèle\_possible*, sans ré-estimation du modèle à partir de *ensemble\_points*. Peut faire gagner du temps au détriment de la comparaison des erreurs liées à des modèles qui sont estimés à partir d'un petit nombre de points et donc plus sensibles au bruit.

## Les paramètres

Les valeurs des paramètres *t* et *d* doivent être fixées conformément aux exigences spécifiques liées à l'application et à l'ensemble de données. qui peuvent être éventuellement fondées sur l'évaluation expérimentale. Le paramètre *k* (le nombre d'itérations), cependant, peut être déterminé à partir d'un résultat théorique. Soit *p* la probabilité que l'algorithme RANSAC pendant une itération sélectionne uniquement des inliers dans l'ensemble des données d'entrée, lorsqu'il choisit les *n* points à partir desquels les paramètres du modèle seront estimés. Lorsque cela se produit, le modèle qui en résulte est susceptible d'être pertinent, donc *p* donne la probabilité que l'algorithme produise un résultat correct. Soit *w*, la probabilité de choisir un inlier à chaque fois qu'un seul point est sélectionné, c'est-à-dire

$w = \text{nombre de inliers dans les données} / \text{nombre de points dans les données}$

Un cas habituel est que *w* ne soit pas connu à l'avance, mais une valeur approximative peut être estimée. En supposant que les *n* points nécessaires pour l'estimation d'un modèle sont sélectionnées de manière indépendante,  $w^n$  est la probabilité que l'ensemble des *n* points correspond à des inliers et  $1 - w^n$  est la probabilité qu'au moins un des *n* points soit un cas atypique (outlier), un cas qui implique qu'un mauvais modèle sera estimé à partir de cet ensemble de points. Cette probabilité à la puissance de *k* est la probabilité que l'algorithme ne choisissent jamais un ensemble de *n* points qui seraient tous des inliers et cela doit être égale à  $1 - p$ . Par conséquent,

$$1 - p = (1 - w^n)^k$$

qui, en prenant le logarithme des deux côtés, conduit à

$$k = \frac{\log(1 - p)}{\log(1 - w^n)}$$

Il convient de noter que ce résultat suppose que les  $n$  points de données sont sélectionnés de façon indépendante, c'est-à-dire, qu'un point qui a été sélectionné une fois est remis et peut être sélectionné à nouveau dans la même itération. Cela n'est pas souvent une approche pertinente et la valeur calculée pour  $k$  devrait être pris comme une limite supérieure dans le cas où les points sont choisis sans remise. Par exemple, dans le cas de la recherche d'une ligne qui s'ajuste à la série de données illustrée sur la figure ci-dessus, l'algorithme RANSAC choisit généralement deux points à chaque itération et calcule le `modèle_possible` comme la ligne qui relie ces deux points et il est alors important que les deux points soient distincts.

Pour gagner en qualité, l'écart type ou des multiples de celui-ci peuvent être ajouté à  $k$ . L'écart type de  $k$  est définie comme

$$SD(k) = \frac{\sqrt{1 - w^n}}{w^n}$$

## Avantages et inconvénients

Un avantage de RANSAC est sa capacité à faire des statistiques robustes des paramètres du modèle, c'est-à-dire, qu'il peut estimer les paramètres avec un degré élevé de précision, même si une quantité importante de valeurs aberrantes (outliers) est présente dans l'ensemble de données. Un inconvénient de RANSAC est qu'il n'y a pas de limite supérieure sur le temps qu'il faut pour calculer ces paramètres. Quand un temps limite supérieure est utilisé (un nombre maximal d'itérations), la solution obtenue peut ne pas être la solution optimale. Un autre inconvénient de RANSAC est qu'elle suppose de fixer des seuils spécifiques au problème traité.

RANSAC ne peut estimer qu'un seul modèle à un ensemble de données particulier. Comme pour tout approche à modèle unique, lorsque deux (ou plusieurs) modèles coexistent, RANSAC peut ne parvenir à trouver ni l'un ni l'autre.

## Applications

L'algorithme RANSAC est souvent utilisé dans le domaine de la vision par ordinateur, par exemple, pour résoudre simultanément les problèmes de mise en correspondance et d'estimer la matrice fondamentale liée à une paire stéréo de caméras.

## Références

1. Martin A. Fischler and Robert C. Bolles, « Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography », dans *Comm. Of the ACM*, vol. 24, juin 1981, p. 381–395 [lien DOI (<http://dx.doi.org/10.1145%2F358669.358692>) ]
- (en) David A. Forsyth and Jean Ponce, *Computer Vision, a modern approach*, Upper Saddle River NJ, Prentice Hall, 2003, poche (ISBN 978-0-13-085198-7) (LCCN 2002726601 (<http://lccn.loc.gov/2002726601>) )

- **(en)** Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003, 2nd<sup>e</sup> éd.
- P.H.S. Torr, and D.W. Murray, « The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix », dans *International Journal of Computer Vision*, vol. 24, 1997, p. 271–300 [lien DOI (<http://dx.doi.org/10.1023%2FA%3A1007927408552>) ]
- Ondrej Chum, « Two-View Geometry Estimation by Random Sample and Consensus », dans *PhD Thesis*, 2005 [texte intégral (<http://cmp.felk.cvut.cz/~chum/Teze/Chum-PhD.pdf>) ]

## Liens externes

- RANSAC Toolbox for MATLAB (<http://vision.ece.ucsb.edu/~zuliani/Code/Code.html>) . A research (and didactic) oriented toolbox to explore the RANSAC algorithm in MATLAB. It is highly configurable and contains the routines to solve a few relevant estimation problems.
- RANSAC for Dummies (<http://vision.ece.ucsb.edu/~zuliani/Research/RANSAC/docs/RANSAC4Dummies.pdf>) A simple tutorial with many examples that uses the RANSAC Toolbox for MATLAB.
- 25 Years of RANSAC Workshop (<http://cmp.felk.cvut.cz/ransac-cvpr2006/>)

Ce document provient de « <http://fr.wikipedia.org/wiki/RANSAC> ».

Dernière modification de cette page le 26 septembre 2011 à 11:29.

Droit d'auteur : les textes sont disponibles sous licence Creative Commons paternité partage à l’identique ; d’autres conditions peuvent s’appliquer. Voyez les conditions d’utilisation pour plus de détails, ainsi que les crédits graphiques. En cas de réutilisation des textes de cette page, voyez comment citer les auteurs et mentionner la licence.

Wikipedia® est une marque déposée de la Wikimedia Foundation, Inc., organisation de bienfaisance régie par le paragraphe 501(c)(3) du code fiscal des États-Unis.