

SAE - Détection de biomes sur des exoplanètes

11 juin 2024

En juillet 2128, le premier télescope à lentille gravitationnelle solaire (T-LGS) a été mis en route grâce à une coopération entre l'Agence spatiale européenne (ESA) et la National Aeronautics and Space Administration (NASA). Ce télescope s'appuie sur le fait que la gravité de corps massif, comme le soleil, déforme l'espace. Cette déformation est capable de dévier les rayons lumineux d'objets extrêmement éloignés, ce qui permet de voir ces objets comme s'ils étaient beaucoup plus proches. C'est ce qu'on appelle une lentille gravitationnelle comme illustré figure 1.

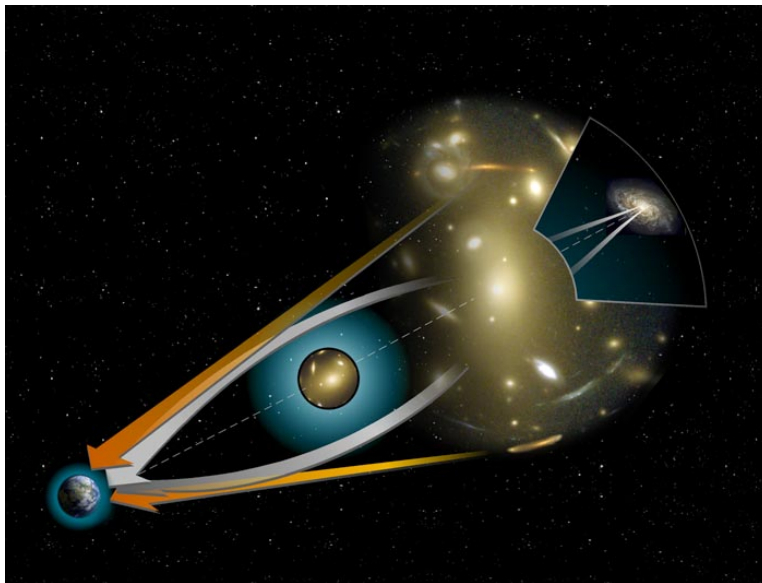


FIGURE 1 – Exemple de lentille gravitationnelle. Les rayons lumineux sont déviés par le corps celeste au centre de l'image, ce qui fait que la galaxie, bien que située derrière ce corps celeste, peut néanmoins être observée de la terre (source Wikipedia – https://fr.wikipedia.org/wiki/Lentille_gravitationnelle).

Cette avancée majeure a permis de photographier pour la première fois des exoplanètes avec une précision inégalée. Le traitement d'un grand nombre de photos d'une planète permet de créer une carte de la surface de cette planète en éliminant les variations de luminosité et la présence de nuages. La figure 2 montre une de ces cartes.

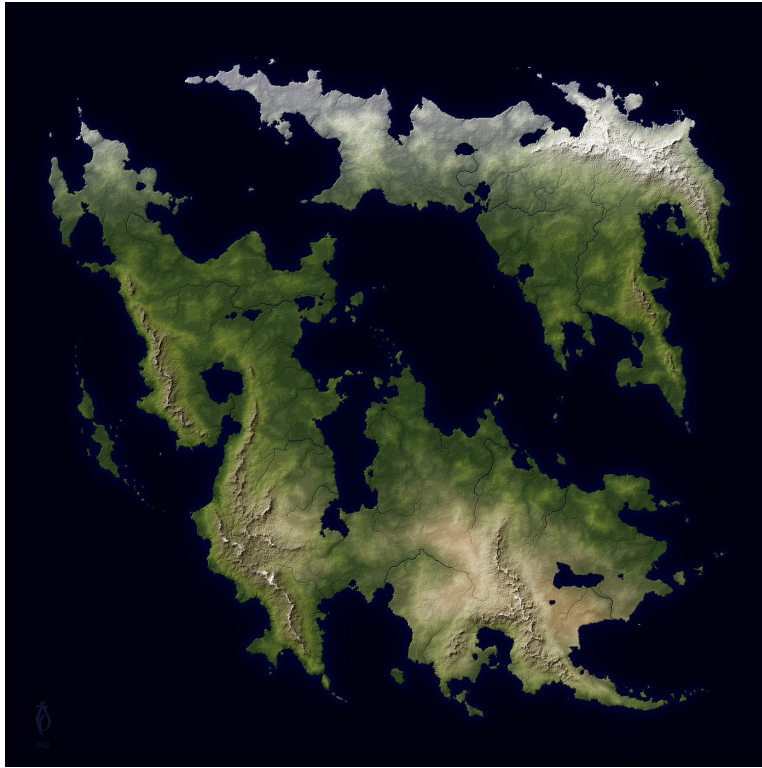


FIGURE 2 – Carte de l'exoplanète *Kepler-442 b* reconstruite à partir de photos du T-LGS.

La présence évidente de trace de vie sur certaines de ces planètes intéresse grandement les exobiologistes. Ils ont besoin d'experts tels que vous pour les aider dans la détection automatique des biomes présents sur ces planètes. Un biome est un ensemble d'écosystèmes aux caractéristiques similaires (par exemple les déserts, les forêts tropicales, les lacs, ...) liées aux conditions environnementales telles que le climat, la topographie et les sols, qui influencent les espèces végétales et animales qui s'y trouvent. L'idée principale est que chaque biome se caractérise par une couleur propre, qui permet de l'identifier sur la carte. Chaque biome est composé d'écosystèmes qui ont les mêmes caractéristiques, mais qui ne sont pas en contact les uns avec les autres.

Par exemple, si on s'intéresse aux zones bleues présentes sur la carte,

- ces zones bleues correspondent à des zones de littoral et de lac, et l'ensemble de ces zones forme le biome « eau peu profonde » ;
- chaque région de ce biome (chaque lac et chaque zone littorale entourant une île) constitue un écosystème à part entière.

L'objectif de la SAÉ sera donc de détecter automatiquement les différents biomes présents sur une exoplanète, puis de définir les différents écosystèmes appartenant à chaque biome.

1 Prétraitement de l'image

Afin de faire ressortir les grands zones de couleurs similaires sur la carte, la première étape sera d'appliquer un filtre de « flou » sur l'image. L'objectif sera de lisser et d'homogénéiser l'image afin de faire disparaître les petits détails qui ne nous intéressent pas.

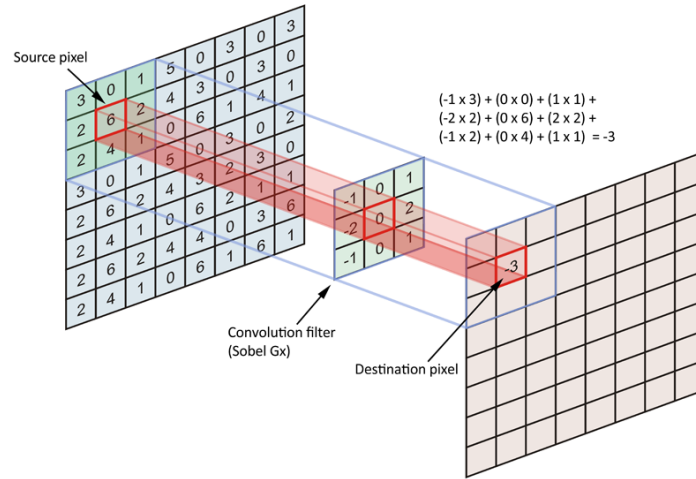


FIGURE 3 – Exemple de filtre appliqué sur un pixel d'un canal.

Pour cela, l'idée est d'appliquer un *filtre de convolution* sur l'image : il s'agit d'une petite matrice de coefficients que l'on applique sur chaque pixel de l'image en considérant ses pixels voisins. La nouvelle valeur du pixel devient la somme des produits des coefficients avec les valeurs initiales du pixel et de ses voisins. Voir figure 3 pour un exemple. Notez bien que le filtre doit être appliqué sur chaque canal de couleur de l'image (R, G et B).

1.1 Flou par moyenne

Le filtre le plus simple est celui dont tout les coefficients sont identiques. Pour que l'image garde ses niveaux de couleur, il est important que la somme des coefficients du filtre soit égal à 1. Par exemple, un filtre 3*3 avec 9 coefficients contiendra la valeur $\frac{1}{9}$ pour tous ses coefficients (voir figure 4). On obtient ce qu'on appelle un « flou par moyenne », parce que la nouvelle valeur de couleur du pixel sera simplement la couleur moyenne de ce pixel et ses voisins.

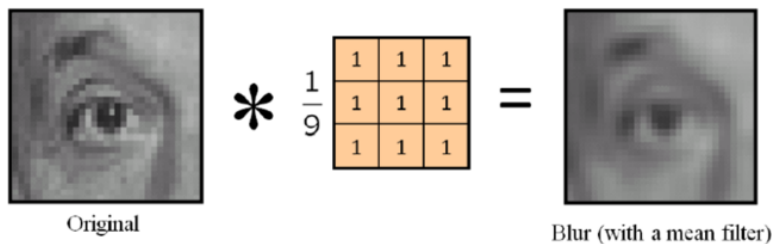


FIGURE 4 – Exemple de filtre de type flou par moyenne.

1.2 Flou gaussien

Une autre façon de faire est d'appliquer des valeurs de coefficients différentes sur le pixel central et sur ses voisins. L'idée est de mettre un plus gros coefficient sur le pixel central et des coefficients de plus en plus petits sur les voisins de plus en plus éloignés. En général, on utilise pour cela un filtre Gaussien qui utilise la formule de Gauss :

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

où x est la distance avec le pixel central sur l'axe horizontal, y est la distance avec le pixel central sur l'axe vertical, et σ est l'écart-type de la distribution gaussienne (un paramètre à choisir).

En pratique, on utilise souvent des filtres pré-calculés avec des valeurs de σ qui dépendent de la taille du filtre choisi, et qui sont normalisées pour que la somme des coefficients soit égale à 1. La figure 5 vous donne des valeurs de coefficients pour des filtres de taille 3*3 à 7*7.

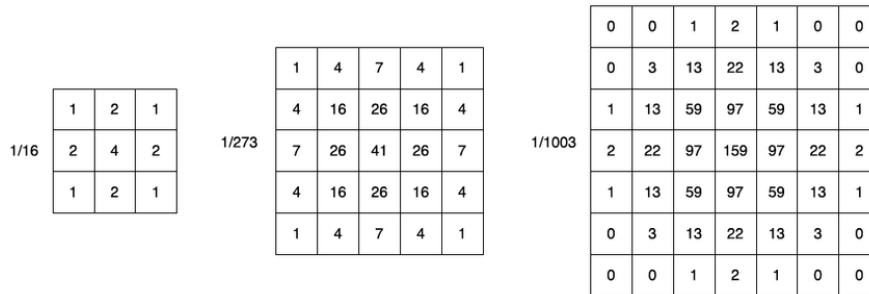


FIGURE 5 – Filtres Gaussiens pré-calculés.



À faire

Testez différents types et différentes tailles de filtres. L'idée est d'homogénéiser les couleurs des zones correspondants aux biomes, tout en conservant les frontières entre les biomes visibles.

2 Détection et visualisation des biomes

2.1 Détection des groupes de pixels de couleur similaire

Les biomes étant définis comme des zones de la carte ayant des couleurs similaires, une bonne approche serait d'utiliser des algorithmes de clustering pour regrouper les pixels en fonction de la similarité de leur couleur.

On supposera que l'on part d'un ensemble de N_{objet} objets à regrouper et que chaque objet dispose de N_{carac} caractéristiques (c'est-à-dire de N_{carac} valeurs représentant les caractéristiques de l'objet).

Les étapes sont les suivantes :

1. Choix de l'algorithme de clustering le plus adapté à cette tâche. Prenez en compte notamment le nombre de pixels à traiter.
2. Choix de la métrique pour calculer la similarité entre couleurs (voir TP de préparation).
3. Choix des paramètres du clustering. Il est conseillé de tester différentes valeurs.
4. Évaluation de la qualité du résultat obtenu. La validation peut être visuelle ou en utilisant des indices de qualité.



À faire

Créez une interface pour les algorithmes de clustering.

Un algorithme de clustering doit être capable de prendre en entrée un tableau de valeurs numériques en deux dimensions (de taille nombre d'objets N_{objet} * nombre de caractéristiques N_{carac}) qui contient la description des objets à classer, la sortie sera un tableau à une dimension (de taille nombre d'objets N_{objet}) qui contient un numéro de cluster pour chaque objet.

Pour visualiser les clusters obtenus, on peut remplacer chaque pixel de la carte par la couleur moyenne de son cluster.

2.2 Étiquetage des biomes

Afin de déterminer un « étiquetage » des biomes, il est possible de préparer une correspondance entre des couleurs et des noms de biomes, puis de calculer la distance minimale entre chaque centre de cluster et chaque biome. La figure 6 vous donne un exemple d'étiquetage possible à adapter selon vos besoins. Vous pouvez utiliser des outils en ligne de conversion de couleur en code RGB, par exemple : https://www.rapidtables.com/web/color/RGB_Color.html.










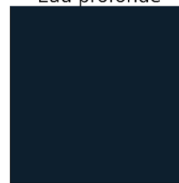
Tundra	Taïga	Forêt tempérée	Forêt tropicale	Savane
				
[71, 70, 61]	[43, 50, 35]	[59, 66, 43]	[46, 64, 34]	[84, 106, 70]
Prairie	Désert	Glacier	Eau peu profonde	Eau profonde
				
[104, 95, 82]	[152, 140, 120]	[200, 200, 200]	[49, 83, 100]	[12, 31, 47]

FIGURE 6 – Exemple d'étiquettes (aussi appelées labels) de biomes avec leurs couleurs RGB.

2.3 Affichage des différents biomes

Pour visualiser les biomes, on commence par préparer un « fond » d'image clair en prenant la carte et en augmentant les canaux RGB d'un certain pourcentage. Les canaux pouvant avoir des valeurs comprises entre 0 et 255, il faut augmenter les valeurs selon un pourcentage de la différence entre la valeur du canal et la valeur maximale de 255, arrondie à l'entier le plus proche. Par exemple, la figure 7 a été obtenue en augmentant les canaux RGB de 75%, ainsi une valeur de 150 devient :

```
int nouveau = Math.round(150 + 75/100 * (255-150)) = 229;
```

Puis, pour chaque cluster, on remplace les pixels du « fond » par les vraies couleurs de l'image afin de visualiser uniquement le biome correspondant. Le titre de l'image peut être le nom de l'étiquette du biome déterminée précédemment. La figure 8 donne un exemple de détection du biome « savanes ».

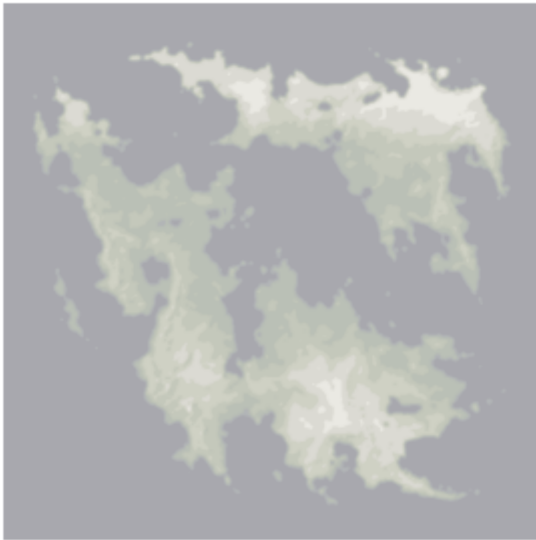


FIGURE 7 – Exemple de fond clair de carte.



FIGURE 8 – Exemple étiqueté de biome détecté sur la carte.

3 Détection et affichage des écosystèmes pour chaque biome

Pour chaque biome, il peut y avoir plusieurs unités d'écosystèmes indépendants séparés spatialement : par exemple différentes montagnes situées à des lieux différents, mais toutes membres du biome « montagnes » détecté par des couleurs similaires sur la carte.

Pour les détecter, on peut appliquer un algorithme de clustering sur la position (abscisses et ordonnées dans l'image) des pixels de chaque biome avec une distance euclidienne.

À faire

À nouveau, respectez bien les étapes de choix et tests :

1. Choix de l'algorithme de clustering le plus adapté à cette tâche.
2. Choix de la métrique pour calculer la similarité entre positions de pixels.
3. Choix et test des paramètres du clustering.
4. Évaluation de la qualité du résultat obtenu.

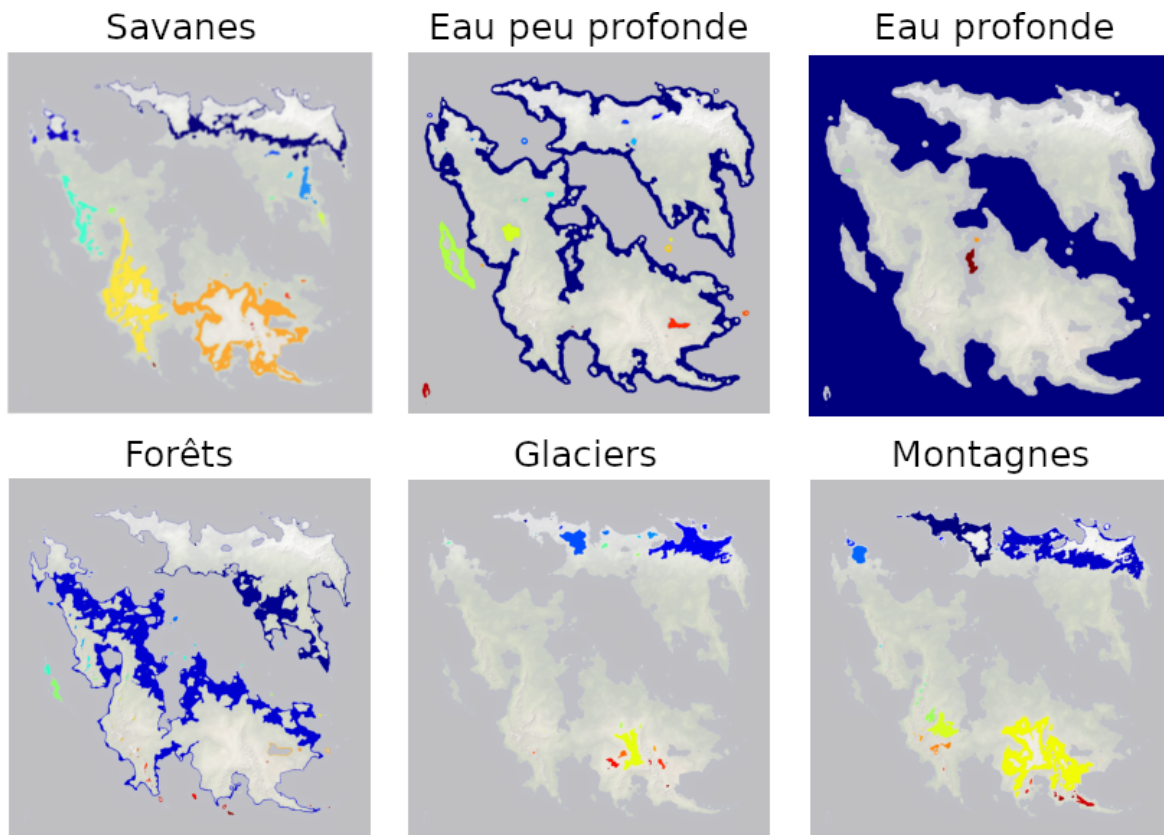


FIGURE 9 – Exemple détection d'écosystèmes.

On remplace alors les pixels de chaque écosystème par une couleur unique, par exemple selon une liste de couleurs prédéfinies.

Notez qu'on peut utiliser la même liste de couleurs pour les différents biomes, à condition de veiller à bien afficher l'étiquette du biome sur chaque figure.

La figure 10 présente un exemple de résultat qu'on peut obtenir pour cette carte. Il ne s'agit pas forcément du meilleur résultat possible pour ce problème.

4 Compte rendu du travail effectué

Afin de présenter les résultats de vos travaux pour évaluation par le comité, il est nécessaire de préparer un certain nombre de documents :

- Une application facile à utiliser, qui peut être appliquée pour une démonstration à partir d'une nouvelle image de carte.
- Un code commenté qui fonctionne.
- Un rapport concis mais complet avec la présentation du travail effectué et une analyse des résultats obtenus.
- Une présentation orale rapide (3-4 slides maximum) présentant vos choix et les résultats qui vous semblent pertinents.

Pour tester votre travail, vous trouverez des exemples de cartes sur Arche. Vous pouvez également générer vous même des cartes à l'aide d'outils en ligne (attention à bien régler les paramètres pour avoir des cartes compatibles avec la méthode mise en oeuvre dans ce projet) :

- <https://topps.diku.dk/torbenm/maps.msp>
- <https://progenesis.com/WorldGen/worldgen.html>
- <https://donjon.bin.sh/scifi/world/>
- <https://www.redblobgames.com/maps/mapgen4/>
- <https://www.redblobgames.com/maps/mapgen2/>

Pour ne pas perdre trop de temps, nous vous incitons à commencer par tester votre approche sur les cartes proposées sur Arche et adaptées à la SAÉ avant de chercher à créer d'autres cartes.

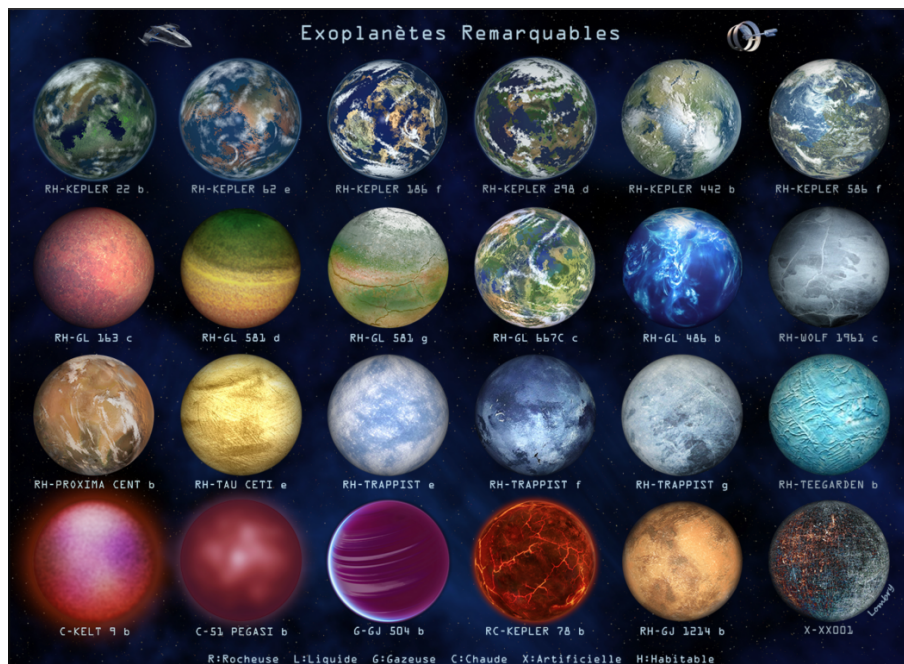


FIGURE 10 – Exoplanètes remarquables (vues d'artiste)