

Jeux de Données d'Observation de la Terre pour la Détection des Changements dans les Forêts

Earth Observation Datasets for Change Detection in Forests

Julius Akinyemi¹, Josiane Mothe², Nathalie Neptune²

¹ MIT Media Lab/UWINCorp Inc., 77 Mass. Ave., E14/E15, Cambridge, MA 02139-4307 USA

akinyemi@media.mit.edu

² Institut de Recherche en Informatique de Toulouse, UMR 5505 CNRS, Université de Toulouse, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

{josiane.mothe,nathalie.neptune}@irit.fr

RÉSUMÉ. La détection automatique des changements qui surviennent dans les forêts (déforestation, reforestation) s'appuie sur des jeux de données divers. Cet article présente une revue des données tant globales que locales qui peuvent servir pour évaluer les tâches de classification de couverture au sol, la détection de changement, la segmentation et l'annotation d'images pour l'analyse des phénomènes de déforestation et de reforestation.

ABSTRACT. The automatic detection of changes in forests (deforestation, reforestation) relies on various data sets. This article reviews data sets both global and local that can be used to evaluate tasks of land cover classification, change detection, segmentation and annotation of images for the analysis of deforestation and reforestation phenomena.

MOTS-CLÉS. Collection d'images, imagerie satellite, détection de la déforestation, masses de données.

KEYWORDS. Image collections, satellite imagery, deforestation detection, big data.

1. Introduction

Des volumes importants de données d'observation de la terre sont depuis plusieurs années disponibles en accès libre et gratuit. En effet, les archives du programme Landsat¹ remontant à 1972 ont été ouvertes en 2008 tandis que le programme Copernicus² a adopté une politique de données libres et gratuites dès ses débuts en 2014 avec le lancement du premier satellite Sentinel-1A.

Nous proposons une revue des principaux jeux de données d'observation de la terre disponibles qui sont pertinents pour détecter les changements de couverture du sol et en particulier la déforestation et la reforestation. Certaines de ces données sont également utilisées dans des publications scientifiques (voir l'étude bibliométrique proposée dans (Neptune *et al.*, 2018)).

Dans les sections suivantes, les problématiques de la classification automatique de la couverture du sol et de la détection de changements à partir des résultats de cette classification seront abordées pour établir le contexte de cette démarche. Les jeux de données annotées et non-annotées de couverture du sols seront ensuite présentés. Finalement, nous terminerons par une discussion sur les données les plus adaptées pour les travaux de détection de la déforestation et de la reforestation.

1. Landsat est un programme des agences américaines, "National Aeronautics and Space Administration"-NASA et "United States Geological Survey"-USGS, pour la collecte d'information spectrale de la surface de la terre depuis les années 70.

2. Copernicus est un programme européen qui vise à l'observation de la terre via un ensemble de satellites chacun dédié à un aspect de l'observation. (voir aussi <http://www.copernicus.eu/>)

2. Le contexte : la couverture du sol et la détection de changements

Les cartes de couverture du sol produites à partir d'images prises par divers satellites d'observation de la terre sont utilisées depuis de nombreuses années pour surveiller l'évolution de l'occupation et l'utilisation du sol. En particulier, les cartes d'occupation du sol où sont présentes une ou plusieurs classes de forêt peuvent être utilisées par les acteurs environnementaux dans leurs processus d'aide à la décision et d'élaboration des politiques.

2.1. La problématique de la classification de la couverture du sol

Les cartes de couverture du sol sont produites soit par interprétation visuelle humaine, soit par un processus de classification automatique, soit par un processus mixte où les résultats de la classification automatique sont vérifiés puis acceptés ou re-classifiés par des humains. Cette approche mixte est notamment celle adoptée par l'agence spatiale brésilienne pour son projet "Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite" (PRODES) qui permet de surveiller la déforestation dans l'Amazonie Légale (INPE, 2011).

2.2. La problématique de la détection de changements

La détection de changements de couverture au sol est le plus souvent réalisée en comparant des images d'une même zone à des temps différents. Deux principales approches de détection de changements existent : la pré-classification et la post-classification. Dans la pré-classification, les images sont d'abord comparées pour trouver les différences entre elles, puis ces différences trouvées sont classifiées. Dans la post-classification, chaque image est d'abord classifiée puis les classifications obtenues sont comparées entre elles pour trouver les changements (Chu *et al.*, 2016).

Il existe de nombreux jeux de données qui portent sur la couverture au sol. Nous avons choisi de n'inclure que ceux qui répondent à au moins un des critères suivants :

- souvent cités dans la littérature,
- produits par d'importants centres de recherche ou institutions,
- pouvant être utilisés gratuitement à des fins de recherche,
- étant librement accessibles ou aisément mis à disposition de tout chercheur qui en fait la demande,
- pertinents pour la détection de la déforestation.

La section suivante présente les jeux de données de couverture du sol et leurs caractéristiques.

3. Les données annotées de la couverture des sols

3.1. Les données globales

3.1.1. Global Forest Change

La carte "Global Forest Change" a été créée par Hansen et ses collègues du département des sciences géographiques de l'Université de Maryland, avec pour objectif de quantifier les changements au niveau des forêts à l'échelle globale, à une résolution spatiale de 30 mètres (Hansen *et al.*, 2013). Pour la première version datant de 2013, des images de Landsat 7 Enhanced Thematic Mapper Plus (ETM+) ³ datant de 2000 à 2012 ont été utilisées. La deuxième version datant de 2017 inclut les données de 2000 à 2016 avec l'ajout d'images provenant de Landsat 8. Les données sont disponibles sous deux formes, des images matricielles ⁴ groupées en tuiles de 10X10 degrés et des images vectorielles ⁵ sur Google Earth Engine ⁶. Les images sont classifiées par pixel et par année.

3. Landsat Enhanced Thematic Mapper Plus est un radiomètre à balayage multi-spectral à huit bandes avec une résolution spatiale de 30 m X 30 m. (voir aussi <https://landsat.gsfc.nasa.gov/the-enhanced-thematic-mapper-plus/>).

4. Les images matricielles aussi appelées images tramées sont des images dans lesquelles les informations sont présentées sous forme de pixels d'information assemblés en grille.

5. Les images vectorielles sont des images composées de points et des chemins entre ces points avec leurs propriétés.

6. Google Earth Engine est une plateforme cloud mise en place par Google pour la visualisation et l'analyse de données spatiales.

3.1.2. Global Land Cover maps - ESA

Les "global land cover maps" sont des cartes de couverture terrestre réalisées par l'initiative sur le changement climatique de l'agence spatiale européenne (ESA) (ESA, 2014). Au total 24 cartes sont proposées sous forme d'images matricielles, pour les années 1992 à 2015, à une résolution spatiale de 300 mètres. Le nombre de classes définies est de 22 dont 8 sont des types de forêt.

3.1.3. Les cartes forêt/non-forêt ALOS PALSAR

Les données globales présentées précédemment ont toutes été réalisées avec des capteurs optiques. Les travaux portant sur la réalisation de cartes globales de couverture au sol à partir d'imagerie radar sont moins courantes. Une contribution notable de production de carte globale permettant de suivre la perte de couverture forestière par année a été faite par (Shimada *et al.*, 2014) en utilisant les données radar ALOS⁷ PALSAR⁸.

<i>Données</i>	<i>Résolution spatiale</i>	<i>Classes</i>	<i>Unité</i>	<i>Années</i>
Global Forest Change	30m	2 à 16	<i>pixel</i>	2000 – 2016
<i>Global LC maps</i>	300m	22	<i>pixel</i>	1992 – 2015
<i>ALOS PALSAR</i>	25m	4	<i>pixel</i>	2007 – 2010

Tableau 3.1. Tableau récapitulatif des données globales

3.2. Les données régionales et locales

3.2.1. Données de déforestation PRODES

Le PRODES (Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite) est un projet pour l'estimation de la déforestation qui fournit des données sur la région amazonienne obtenue à partir de l'interprétation visuelle d'images Landsat et CBERS⁹ préalablement segmentées par régions (Shimabukuro *et al.*, 2000). Les données sont produites chaque année à partir de 1999. Elles sont disponibles au format vecteur et comptent 5 classes dont une classe "forêt" et deux classes de déforestation (INPE, 2011).

3.2.2. Données du système FORMA

Le système FORMA (FORest Monitoring for Action) a été conçu pour fournir des informations pratiquement en temps réel sur la perte de couverture végétale dans les forêts tropicales d'Asie, d'Afrique et d'Amérique du Sud (Wheeler *et al.*, 2014). Les données produites à partir d'images MODIS¹⁰ sont disponibles en ligne. FORMA inclut aussi des informations sur les feux, les précipitations et des alertes déclenchées par une perte de couverture végétale dépassant un seuil de 25% pour 1 pixel. Chaque pixel couvre une surface de 250 x 250 mètres. Contrairement à PRODES, le traitement des images est fait de manière complètement automatique.

3.2.3. Données du bassin amazonien de Kaggle-Planet

Pour la compétition Kaggle¹¹ intitulée "Planet : Understanding the Amazon from Space", Planet Labs Inc¹² a fourni un jeu de données composé d'images satellites du bassin amazonien aux formats GeoTiff et JPG (Kaggle,

7. Advanced Land Observing Satellite (ALOS) est un satellite de l'agence d'exploration aérospatiale japonaise lancé en 2006 et dont la mission a pris fin en 2011. (voir aussi <http://global.jaxa.jp/projects/sat/alos/>)

8. Phased Array L-band Synthetic Aperture Radar (PALSAR) est un capteur radar à synthèse d'ouverture.

9. Les satellites du "China-Brazil Earth Resources Satellite Program" (CBERS) ont été développés dans le cadre d'une coopération entre la Chine et le Brésil (voir également <http://www.cbears.inpe.br/>)

10. Les capteurs MODIS (Moderate Resolution Imaging Spectroradiometer) de la NASA sont des instruments à résolution spatiale moyenne qui produisent des images de toute la surface de la terre tous les 1-2 jours.

11. Kaggle est une plate-forme de compétition en apprentissage automatique et science des données.

12. Planet Labs Inc est une société spécialisée dans les satellites d'observation de la terre basée à San Francisco aux États-Unis.

2017). Les images ont été collectées entre le 1er janvier 2016 et le 1er février 2017. Les images sont présentées sous forme de tuiles annotées avec plusieurs classes représentant les phénomènes d'intérêt dans le bassin amazonien. Les images ont été annotées par interprétation visuelle humaine et 23 classes ont été définies.

3.2.4. Carte de la couverture des sols de l'Afrique

L'agence spatiale européenne propose une carte de couverture du sol pour le continent africain pour l'année 2016, à 20 mètres de résolution spatiale. Cette carte est présentée comme un prototype et a été rendue publique pour obtenir des retours des utilisateurs. Les données utilisées sont issues d'observations du satellite Sentinel-2A de décembre 2015 à décembre 2016 (ESA, 2018). La carte est mise à disposition sous forme d'image matricielle. Dix classes génériques y sont présentes dont une correspondant aux forêts.

<i>Données</i>	<i>Résolution</i>	<i>Satellite</i>	<i>Classes</i>	<i>Années</i>
<i>PRODES (Amazonie)</i>	30m	<i>LandsatCBERS</i>	5	1999–
<i>FORMA (Forêts tropicales humides)</i>	500m	<i>Terra – MODIS</i>	3	2014–
<i>Kaggle – Planet (Amazonie)</i>	3m	<i>PlanetScope/RapidEye</i>	23	2016 – 2017
<i>Carte de l'Afrique</i>	20m	<i>Sentinel – 2</i>	10	2015 – 2016

Tableau 3.2. Tableau récapitulatif des données régionales et locales

4. Les données non-annotées de la couverture des sols

La grande majorité des données d'observation de la terre est disponible sous forme d'images non-annotées. La segmentation de ces images ainsi que leur association à des concepts de manière automatique sont des défis liés à la recherche d'information. Les méthodes de l'état de l'art adressant ces défis dans le contexte de la détection de la déforestation ont été passées en revue par (Huynh *et al.*, 2018).

En 2008, les archives du programme Landsat sont devenues ouvertes et librement accessibles. Le satellite d'observation de la terre Terra de la NASA lancé en 1999, avec son capteur MODIS, réalise des images de la terre entière tous les 1 à 2 jours. Toutes ces images mises à disposition du public sous forme d'images brutes ou de produits d'images pré-traitées sont typiquement non-annotées. Enfin, le programme Européen Copernicus, le plus récent, propose aussi un accès libre aux données des nouveaux satellites Sentinel.

5. Discussion

Malgré la diversité des données présentées dans les sections précédentes, le choix de celles à utiliser pour détecter les changements dans les forêts n'est pas forcément aisé. En effet, les zones d'intérêts ne font pas toujours l'objet d'études préalables aboutissant à des cartes de couverture au sol, d'où l'importance des données non-annotées.

Un autre avantage des données non-annotées est leur mise à disposition rapide. Le processus généralement long d'annotation et de validation des annotations étant absent dans la production de ces données, elles peuvent être rapidement disponibles.

Diverses méthodes de détection de changement de couverture au sol existent pour produire les annotations liées à la déforestation pour les images satellites. Elles comprennent entre autres des approches algébriques, d'autres basées sur des transformations ainsi que de nombreuses méthodes basées sur la classification dont les plus récentes faisant appel aux réseaux de neurones profonds (Huynh *et al.*, 2018).

Il est également possible d'utiliser des données annotées conjointement avec des données non-annotées notamment pour des tâches d'apprentissage semi-supervisé. Plus généralement, ces données peuvent être utilisées dans un système de recherche d'information. En effet, l'association des concepts liés à un phénomène (dans ce cas la déforestation) aux segments d'images concernés rend possible l'appariement image-texte pour des systèmes permettant de rechercher ces segments spécifiques d'images à partir des requêtes textuelles ou d'images.

6. Conclusion

Cet article présente les données d'observation de la terre pouvant servir à la classification automatique de la couverture des sols, en particulier pour analyser les phénomènes de déforestation et de reforestation. Des données annotées et non-annotées tant globales que locales ont été présentées avec une discussion sur leur utilisation notamment pour des tâches de détection de changement et d'annotation d'images.

7. Remerciements

Ces travaux ont été aidés financièrement par la Fondation Schlumberger avec la bourse "Faculty for the Future" dont bénéficie Nathalie Neptune. Ils s'inscrivent également dans le cadre du projet FabSpace 2.0 (voir par exemple (Del Frate *et al.*, 2017)) qui a reçu un financement de l'Union Européenne (programme Recherche et Développement Horizon 2020 - Convention n 693210).

Références

- CHU Y., CAO G. & HAYAT H. (2016). Change Detection of Remote Sensing Image Based on Deep Neural Networks. *Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*, 133(1), 262–267.
- DEL FRATE F., MOTHE J., BARBIER C., BECKER M., OLSZEWSKI R. & SOUDRIS D. (2017). Fabspace 2.0 : The open-innovation network for geodata-driven innovation. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 353–356 : IEEE.
- ESA (2014). Land cover maps | esa cci land cover website. Consulté le : 2018-03-26.
- ESA (2018). Prototype land cover map over africa at 20m released | esa cci land cover website. Consulté le : 2018-03-26.
- HANSEN M. C., POTAPOV P. V., MOORE R., HANCHER M., TURUBANOVA S. A., TYUKAVINA A., THAU D., STEHMAN S. V., GOETZ S. J., LOVELAND T. R., KOMMAREDDY A., EGOROV A., CHINI L., JUSTICE C. O. & TOWNSHEND J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853.
- HUYNH D., MOTHE J. & NEPTUNE N. (2018). Automatic image annotation : the case of deforestation. *Rencontres Jeunes Chercheurs, RJC'18*.
- INPE (2011). PRODES Deforestation Layers Methodology.
- KAGGLE I. (2017). Planet : Understanding the amazon from space | kaggle. Consulté le : 2018-01-25.
- NEPTUNE N., MOTHE J. & AKINYEMI J. (2018). Fouille de publications scientifiques pour une analyse bibliométrique de l'activité de recherche sur la déforestation (regular paper). In *EGC, Atelier Fouille de Textes, Paris*.
- SHIMABUKURO Y. E., DUARTE V., KALIL MELLO E. M. & MOREIRA J. C. (2000). *Presentation of the Methodology for Creating the Digital PRODES*. Rapport interne, São José dos Campos.
- SHIMADA M., ITOH T., MOTOOKA T., WATANABE M., SHIRAISHI T., THAPA R. & LUCAS R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, 13–31.
- WHEELER D., HAMMER D., KRAFT R. & STEELE A. (2014). Satellite-Based Forest Clearing Detection in the Brazilian Amazon : FORMA, DETER, and PRODES In. *World Resources Institute Issue Brief*, p. 1–24.