Extraction automatique de paysages en imagerie satellitaire et enrichissement sémantique

Anne Toulet, Emmanuel Roux, Anne-Elisabeth Laques, Eric Delaître, Laurent Demagistri, Isabelle Mougenot

▶ **To cite this version:**

Anne Toulet, Emmanuel Roux, Anne-Elisabeth Laques, Eric Delaître, Laurent Demagistri, et al.. Extraction automatique de paysages en imagerie satellitaire et enrichissement sémantique. EGC: Extraction et Gestion des Connaissances, Jan 2017, Grenoble, France. hal-01464947

HAL Id: hal-01464947
https://hal.science/hal-01464947

Submitted on 10 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extraction automatique de paysages en imagerie satellitaire et enrichissement sémantique

Anne Toulet*, Emmanuel Roux**, Anne-Élisabeth Laques**
Éric Delaître**, Laurent Demagistri**, Isabelle Mougenot**

*LIRMM Université de Montpellier
161, rue Ada 34095 Montpellier cedex 5
anne.toulet@lirmm.fr
**UMR 228 Espace Dev (UM,UR,UG,UA,IRD), Maison de la Télédétection
500 rue JF Breton 34093 Montpellier Cedex 5
emmanuel.roux@ird.fr, isabelle.mougenot@umontpellier.fr

Résumé. Nous présentons ici une méthode originale pour l'automatisation de la détection de paysages dans une image satellite. Deux enjeux majeurs apparaissent dans ce processus. Le premier réside dans la faculté à prendre en compte l'ensemble des connaissances expertes tout au long du travail d'analyse de l'image. Le second est de réussir à structurer et pérenniser ces connaissances de façon à les rendre interopérables et exploitables dans le cadre du web de données. Nous présentons en quoi la collaboration de plusieurs stratégies alliant les traitements de l'image, le calcul de caractéristiques spécifiques et la programmation logique inductive (PLI), vient alimenter le processus d'automatisation, et comment l'intégration de la connaissance, au travers de la construction d'ontologies dédiées, permet de répondre pleinement à ces enjeux.

1 Introduction

Le contexte qui nous intéresse ici est celui de l'identification de paysages dans une image satellite. L'étude des paysages et leur évolution au cours du temps permet de répondre à de grands enjeux sociaux-économiques et environnementaux (Fahrig, 2003). Pour identifier des paysages, le géographe met en œuvre toutes sortes de techniques : les missions de terrain, les photographies aériennes, les outils cartographiques et l'imagerie satellitaire. Pour l'expert géographe, c'est l'ensemble de ces approches qui permet de caractériser les différents types de paysages et d'en dresser une carte (Enserink, 1999). Or, ce travail est extrêmement coûteux en temps. Le recours à l'imagerie satellitaire est actuellement utilisé comme un moyen d'obtenir plus rapidement et de façon concentrée un grand nombre de ces informations. L'accès aux images a été favorisé par la croissance exponentielle de leur production ainsi que par une plus grande facilité de mise à disposition. Le problème reste que les techniques d'analyse des images satellites sont encore limitées et ne peuvent se faire sans intervention humaine.

L'un de nos objectifs est donc de faciliter le travail du géographe en allant vers une automatisation de l'extraction des paysages dans une image satellite. Atteindre cet objectif nécessite

la mise en œuvre de plusieurs stratégies qui utilisent les connaissances issues du domaine thématique (la géographie du paysage), de l'imagerie satellitaire et de la programmation logique inductive (PLI). Un second objectif est de structurer et pérenniser l'ensemble de ces connaissances au travers de modules ontologiques dédiés, et de les rendre réutilisables et interopérables dans le contexte du web sémantique.

Les images satellites ont fait l'objet de différents travaux mettant à contribution les ontologies (Forestier et al., 2012; Belgiu et Thomas, 2013; Chahdi et al., 2016). Il est à remarquer une grande variété des rôles joués par ces ontologies. En effet, la forte volumétrie des données ainsi que la nature exclusivement numérique des données extraites de l'image obligent à composer et à souvent exploiter les ontologies en complément à d'autres approches, à l'exemple de la classification supervisée ou du clustering. Dans notre étude, c'est la PLI qui est exploitée en lieu et place des techniques de classification "classiques". Le choix de la PLI nous permet de travailler à partir d'une approche combinant apprentissage statistique et programmation logique et ainsi formuler des hypothèses mettant en jeu simultanément plusieurs variables venant décrire nos objets dans l'image. Une autre motivation est de travailler, que cela soit au niveau de la PLI ou des ontologies, avec des formalismes qui s'adossent tous à la logique des prédicats, même si les processus calculatoires sont différents.

L'article s'articule comme suit. En section 2, nous présentons l'approche dans sa globalité. Nous précisons d'abord le contexte thématique, avant d'expliquer l'extraction des paysages par recours à la PLI. Puis nous détaillons comment ces étapes viennent alimenter le processus d'intégration de la connaissance au travers de deux ontologies : l'une dédiée à la géographie et l'autre aux données issues de l'image. En section 3, nous illustrons la démarche à travers un exemple concret. Pour une image satellite donnée, nous expliquons les traitements et calculs préliminaires avant de détailler l'obtention des règles de classification qui permettent de typer les paysages dans l'image. Nous montrons aussi comment nous utilisons ces traitements pour peupler l'ontologie image. En section 4, nous revenons sur les résultats obtenus à la fois dans la prédiction des paysages et dans le rôle joué par les ontologies. Enfin, nous concluons l'article en proposant des perspectives à ce travail.

2 Approche méthodologique

Nous présentons tout d'abord le contexte de la géographie du paysage tel qu'envisagé dans notre étude et les traitements préliminaires appliqués à l'image pour répondre à nos besoins.

2.1 Contexte thématique

Techniquement parlant, l'identification d'un paysage utilise de nombreux critères dont sa composition et sa configuration, mesurées au travers de métriques paysagères appliquées à des composants paysagers (Farina, 2000). À ce niveau d'observation, les Composants Paysagers (CP) s'apparentent à l'occupation du sol ou encore à la « tache » (patch en anglais) en écologie. C'est en particulier les relations qu'entretiennent ces composants les uns par rapport aux autres qui permettent de déterminer l'appartenance du paysage à un Type Paysager (TP). Les métriques paysagères utilisées dans cette étude sont classiques dans ce contexte (Dérioz et al., 2010), à l'exemple de la composition, la diversité ou la fragmentation. Une difficulté réside dans le fait qu'il n'existe pas « naturellement » dans les images satellites d'objets identifiables

à des composants paysagers sur lesquels on pourrait appliquer des métriques paysagères. Nous expliquons ci-dessous les traitements préalables permettant de s'en approcher.

Traitement préalable des images satellites. L'objectif de ces traitements est d'obtenir des objets pouvant s'apparenter à des composants paysagers. La première étape consiste à produire une carte d'occupation du sol qui permet d'obtenir de tels objets. Les labels des classes d'occupation du sol fournissent les labels des objets correspondants dans l'image, à l'exemple de l'« Eau », la « Végétation » ou la « Forêt » (Richards, 2013). Mais ce premier résultat ne permet pas d'atteindre un niveau sémantique suffisamment élevé pour prendre en compte la structure des paysages. En particulier, on ne sait pas appliquer directement les métriques paysagères aux objets obtenus. Deux questions se posent : comment délimiter les zones sur lesquelles appliquer les métriques et quels calculs définir en correspondance aux métriques paysagères ? La première question trouve une réponse dans la mise en place d'une grille sur l'image. Cette grille consiste en un ensemble de mailles régulières carrées, dont la taille est définie en fonction des besoins. Ce procédé permet de définir sur l'image des zones arbitraires bien circonscrites sur lesquelles appliquer les métriques. De nombreux outils sont déjà implantés dans les logiciels de traitement d'images de télédétection, à l'exemple de ENVI [1] ou GRASS [2] qui permettent d'effectuer des calculs de caractéristiques correspondant aux métriques paysagères. Nous reproduisons ainsi la démarche du géographe en appliquant ces calculs aux « composants paysagers » contenus dans chaque maille. Les différentes étapes de notre approche sont récapitulées dans la figure 1 : les connaissances thématiques sont modélisées dans l'ontologie du paysage. Ces connaissances guident en particulier les calculs à effectuer pour reproduire

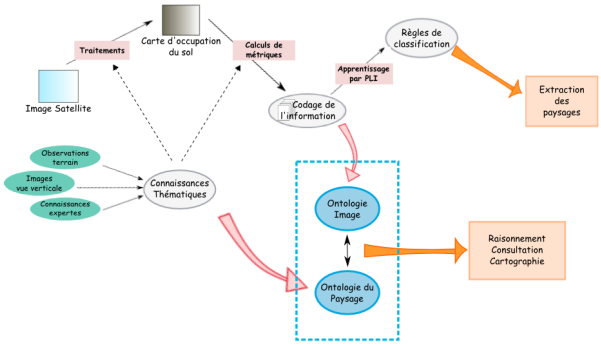


FIGURE 1: Démarche globale.

les métriques paysagères. Les résultats des calculs sont ensuite exploités par la PLI afin de produire des règles de classification qui permettent d'identifier les paysages dans l'image. Ces résultats viennent également peupler l'ontologie image. L'originalité de la méthode est donc de combiner deux approches symboliques : l'une inductive (la PLI) et l'autre déductive (le raisonnement à partir d'ontologies).

1. Environment for Visualizing Images, logiciel commercialisé par la société EXELIS
2. GRASS GIS, logiciel SIG libre développé par le GRASS Development Team

2.2 Extraction des paysages

La programmation logique inductive permet d'obtenir des règles de classification des paysages. Nous expliquons dans cette section la manière dont ces règles sont obtenues.

Obtention de règles par apprentissage automatique. La Programmation Logique Inductive (PLI) a été choisie afin d'induire des règles capables de « prédire » l'appartenance d'une maille à un paysage type, à partir des caractéristiques paysagères de cette maille. La PLI, introduite par Muggleton (Muggleton, 1991), est une méthode d'apprentissage automatique à partir d'exemples (ou méthode d'apprentissage *supervisé*) exploitant le formalisme de la logique du premier ordre. De fait, elle est particulièrement adaptée à la découverte de connaissances relationnelles au sein de données. De plus, l'intérêt d'une telle méthode a déjà été démontré dans le traitement et l'interprétation de données géographiques et de télédétection (Malerba et al., 2003; Vaz et al., 2007).

Programmation Logique Inductive (PLI) : principe général. La PLI (Lavrac et Dzeroski, 1994) admet pour entrées

- une base de connaissances B décrivant un ensemble de connaissances et de contraintes ;
- un ensemble d'exemples E, séparé en deux sous-ensembles E^+ et E^- correspondant respectivement aux exemples positifs et négatifs ;
- un langage de description L.

la PLI génère une « théorie » (*i.e.* un ensemble de règles), H, en utilisant le langage de description L, qui doit i) couvrir les exemples positifs E^+ et ii) ne pas couvrir les exemples négatifs E^-. Parmi les systèmes inductifs existants, le système *Aleph* (Srinivasan, 2007) est retenu ici.

Codage de l'information au sein de la base de connaissances. Chaque métrique paysagère décrite plus haut correspond à un *prédicat* de la base de connaissances, c'est-à-dire une fonction qui retourne une réponse (vrai, ou faux, ou une constante) lorsqu'elle reçoit les informations adéquates en entrée. Par exemple, la métrique *proportion de forêt* correspond au prédicat binaire `proportion_foret/2`, applicable à une maille et retournant la valeur de proportion de forêt pour cette maille. Cependant, la PLI ne permet de raisonner que sur les informations symboliques. Or, les métriques paysagères fournissent des valeurs numériques, qui doivent ainsi être re-codées. Le mode de re-codage choisi ici est celui proposé par (Kavurucu et al., 2011) : pour une variable numérique donnée, chaque valeur observée est re-codée au moyen d'inégalités faisant intervenir des valeurs caractéristiques de la variable considérée. Dans notre application, ces valeurs caractéristiques sont le minimum (min), les 10^{eme}, 20^{eme}, ..., 90^{eme} percentiles et le maximum (max) de la variable considérée. Ainsi, chaque métrique paysagère correspond à un prédicat permettant de tester vingt inégalités. Dans le langage Prolog exploité par le système Aleph, une inégalité s'écrit par exemple :

`proportion_foret_qualitative`$(M, \leq max)$`:- proportion_foret`$(M, X), \quad X \leq max.$

Un tel codage offre un bon compromis entre, d'une part, la perte d'information inhérente au passage de l'information numérique à l'information symbolique, et, d'autre part, la capacité de généralisation au cours du processus d'apprentissage. De plus, elle permet de mettre en évidence des intervalles de valeurs significatifs (Bayoudh et al., 2015).

Base d'exemples et induction des règles. Au regard des objectifs visés (*i.e.* la prédiction de l'appartenance d'une maille quelconque à un paysage type), la PLI nécessite de connaître, pour un nombre significatif de mailles, d'une part, leurs caractéristiques paysagères, et d'autre part, le paysage type auxquels elles appartiennent. Une maille, associée à des valeurs de métriques paysagères et dont l'appartenance à un paysage type est connue, définit ainsi un exemple de l'ensemble d'apprentissage. Notons que la connaissance *a priori* de l'appartenance à un paysage type peut provenir de différentes sources et provient, ici, de l'expertise d'une géographe. Le processus d'apprentissage est lancé autant de fois qu'il y a de paysages types. Pour un paysage type donné, les mailles appartenant à ce paysage constituent les exemples positifs, les autres mailles définissent quant à elles les exemples négatifs. La précision minimale des règles, évaluée au cours du processus d'apprentissage par le rapport $p/(p+n)$, où p et n correspondent aux nombres d'exemples, respectivement positifs et négatifs couverts par la règle, est fixée à 0,7. Une telle valeur est considérée comme un bon compromis entre les exigences antagonistes de précision et de généralisation des règles induites. Enfin, la longueur maximum des prémisses des règles est fixée à cinq littéraux, une telle valeur étant considérée, dans la pratique, comme la limite permettant une lecture correcte des résultats (Michalski, 1983).

2.3 Construction des ontologies

Une ontologie permet la modélisation de connaissances de façon formelle (Gruber, 1993), en s'appuyant pour ce faire sur la définition de concepts d'intérêt et de relations entre ces concepts. Une ontologie facilite notamment le partage des connaissances d'un domaine au travers d'une représentation la plus consensuelle possible. Pour définir nos ontologies, nous utilisons les logiques de description qui permettent de construire et manipuler des bases de connaissances, à partir d'une boîte terminologique (ou *TBox*) qui en modélise les concepts et les relations ; et d'une boîte assertionnelle (ou *ABox*) qui en représente les instances. Les logiques de description s'appuient sur la logique de premier ordre et donnent lieu à des raisonnements par inférence. Ces mécanismes permettent d'enrichir la base de connaissance en explicitant de nouveaux faits déduits des premiers par raisonnement. Nous avons ensuite choisi d'opérationnaliser ces ontologies formelles en OWL2 (Web Ontology Language) (Dean et Schreiber, 2004), langage recommandé par le W3C.

Ontologie du Paysage. L'ontologie du paysage est une ontologie dite de domaine, reflet de l'expertise du géographe, et modélise les concepts qui relèvent de l'organisation d'entités paysagères (type paysager, composant paysager), sous forme d'une arborescence qui exprime la spécialisation de ces concepts (relation de subsomption), ainsi que l'ensemble des propriétés liant ou définissant ces concepts (relation de méronymie, métriques paysagères). Prenons l'exemple volontairement simplifié du type paysager « Écosystème forestier ». Un tel paysage doit contenir nécessairement le composant paysager « Forêt » mais il peut être constitué de composants d'un autre type. Ce paysage est par ailleurs caractérisé par sa diversité faible et peu d'hétérogénéité. Dans cet exemple, nous voyons apparaître les concepts de type paysager et de composant paysager, chacun étant spécialisé par un sous-concept (ici respectivement Écosystème forestier et Forêt). Nous avons par ailleurs besoin des relations *diversité* et *hétérogénéité*, celles-ci étant définies comme des propriétés fonctionnelles au sens mathématique du

terme, à savoir qu'une entité est liée à au plus une autre entité par ces relations. Nous écrirons en logique descriptive :

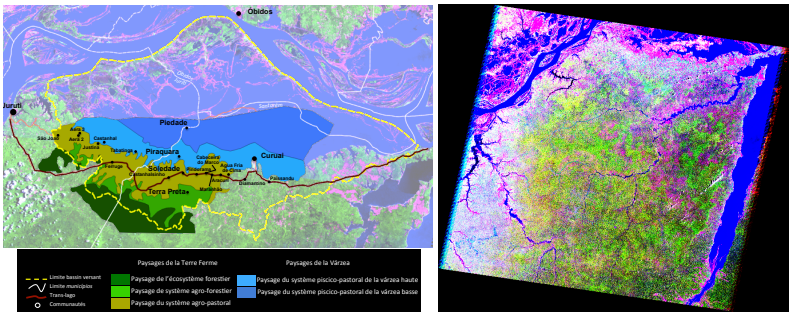
$$\begin{aligned}
&\texttt{ÉcosystèmeForestier} \sqsubseteq \texttt{TypePaysager}\\
&\texttt{Forêt} \sqsubseteq \texttt{ComposantPaysager}\\
&\left.\begin{array}{l}\top \sqsubseteq \leq 1\ a_pour_diversité\\ \top \sqsubseteq \leq 1\ hétérogénéité\end{array}\right\} \text{propriétés fonctionnelles}\\
&\texttt{ÉcosystèmeForestier} \equiv \texttt{TypePaysager} \sqcap \exists estComposéDe.\texttt{Forêt}\\
&\qquad \sqcap \exists a_pour_diversité.\{diversité_faible\} \sqcap \exists hétérogénéité.\{PeuHétérogène\}
\end{aligned}$$

Ontologie image. L'ontologie image vient organiser les diverses connaissances issues de l'image en lien avec l'ontologie du paysage. Elle est construite comme une ontologie applicative, à savoir qu'elle est envisagée comme un conteneur des descripteurs des mailles obtenus à partir des tâches exécutées par le système. Les concepts et les relations de cette ontologie sont fournis par les traitements préalables et le calcul des caractéristiques paysagères. Nous verrons en section 3.3 la manière de peupler cette ontologie.

3 Application

3.1 Données

Nous illustrons la méthode avec une image Landsat 5 acquise en octobre 2009 qui couvre une partie de l'Amazonie située dans l'état du Pará, à la confluence de l'Amazone et du Rio Tapajós. Cinq paysages de type différent (TP1 à TP5) sont décrits et géo-localisés dans la zone d'étude (Fig. 2a), et servent à alimenter la base d'exemples pour la production de règles en PLI. L'objectif est de rechercher ces paysages sur l'ensemble de l'image (Fig. 2b).

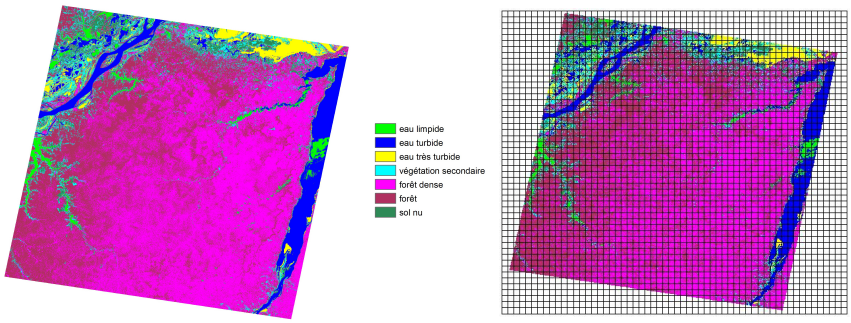


(a) Les 5 paysages labellisés. (b) Image Lansat 2009.

FIGURE 2: Données.

Obtention de la carte d'occupation du sol et maillage. Il existe différentes manières de produire des cartes d'occupation du sol (Richards, 2013). Notre choix s'est porté sur une classification non supervisée (ou *clustering*) par utilisation de l'algorithme des k-means. La carte

d'occupation du sol obtenue est constituée de sept classes labellisées (Fig. 3a) sur laquelle nous apposons une grille de mailles carrées permettant le calcul des caractéristiques paysagères (Fig. 3b).



(a) Carte d'occupation du sol.

(b) Maillage.

FIGURE 3: Cartes de référence.

3.2 Obtention des règles de classification

Nous exploitons les résultats des caractéristiques calculées sur les mailles qui servent à l'apprentissage, et qui sont contenues dans les cinq paysages de référence. Le choix des métriques se fait en cohérence avec la démarche suivie par le géographe, en posant des correspondances entre les métriques paysagères et les caractéristiques calculées dans l'image satellite.

Expert géographe	**Caractéristiques *calculables* dans les mailles**
Composition	Proportion de chaque classe présente
Diversité	Richesse spécifique (nb. de classes présentes)
Hétérogénéité	Indice de Shannon $H = -\sum_{k=1}^{N} p_k \cdot ln(p_k)$
Fragmentation : surface des patches	Surface moyenne des objets
Forme	Périmètre/Aire : moyenne mesurée

TABLE 1: Correspondance entre métriques paysagères et caractéristiques dans l'image.

3.3 Peuplement de l'ontologie image

L'ontologie image modélise les mailles et les propriétés correspondant aux caractéristiques paysagères. Chaque propriété permet d'associer à une maille la valeur obtenue par calcul d'une

caractéristique donnée. Les informations calculées maille par maille viennent ainsi instancier l'ontologie par peuplement de la $ABox$. Prenons l'exemple d'un individu de la classe « Maille » , identifié m59, qui aurait pour proportion de forêt une valeur égale à 0.287, que nous écrirons :

`Maille` $\sqsubseteq Top$
`Maille(m59)`
$\top \sqsubseteq \leq 1\, prop_forêt$ `(propriété fonctionnelle)`
$prop_forêt$`(m59,"0.287")`

De la sorte, toute maille de l'ontologie est décrite par un ensemble d'attributs correspondant aux valeurs des métriques calculées dans les traitements préalables. Chaque maille possède un géoréférencement et une géométrie sous forme d'un polygone prise en charge par emprunt à l'ontologie GeoSPARQL [3] (Battle et Kolas, 2012). Par ailleurs, les règles de classification précisent l'appartenance d'une maille à un type paysager : cette information est ajoutée à la description des mailles. Notons que la structure même de l'ontologie permet d'ajouter à volonté des attributs supplémentaires, à l'exemple de la texture ou de calculs d'indices de végétation.

4 Résultats et discussion

4.1 Résultats

Caractéristiques de la base de règles et performances. Six grilles ont été générées, constituées de mailles carrées allant de 32640m à 1020m de côté (système de mailles emboîtées pour avoir une approche multi-résolution). Les grilles de grande taille (32640m, 16320m, 8160m) se sont révélées inadaptées en rapport à la surface des paysages constituant la base d'apprentissage. Les calculs ont donc été faits sur les grilles de taille 1020, 2040 et 4080. Les résultats les plus pertinents sont obtenus pour les grilles 1020 et 2040, sans véritable différence de l'une à l'autre. Nous présentons ici ceux de la grille 2040. Respectivement 1, 3, 1, 6 et 4 règles pour les Types Paysagers TP1 à TP5 ont été induites pour l'ensemble d'apprentissage complet constitué de 300 mailles de 2040 mètres de côté. Voici un exemple de règle obtenue : $type_paysager(A, tp5) :- pEau_turbide(A, eq0), pForet(A, le0_009)$.
Cette règle signifie qu'une maille A appartient au TP5 si sa proportion d'eau turbide est nulle et si sa proportion de forêt est inférieure à 0.009. Si les règles induites sont intelligibles, y compris par un non expert de la méthode d'apprentissage, les résultats de prédiction sont variables selon le type paysager considéré. Le Tableau 2 présente la sensibilité et la spécificité [4], calculées sur l'ensemble d'apprentissage complet et par validation croisée (valeurs moyenne, minimum et maximum). La précision de la prédiction calculée sur l'ensemble d'apprentissage complet s'élève à 94,7%. Une validation croisée à 10 sous-ensembles stratifiés (*stratified 10-fold cross-validation*) a permis d'estimer à 58,6% la précision de prédiction avec un maximum (respectivement un minimum) sur les 10 sous-ensembles de 71,0% (respectivement 45,7%). Les résultats obtenus pour la sensibilité sont particulièrement bons : celle-ci atteint en effet les 76,3% pour le type paysager 1 et dépasse les 60% pour les types 3 et 5 (respectivement 63,8 et 61,9%). Le type paysager 2 est particulièrement difficile à prédire avec seulement 15% de succès, ce qui s'explique notamment par le nombre restreint d'exemples qui y est associé (15),

3. Standard de l'OGC (Open Geospatial Consortium) pour le géospatial

4. **Sensibilité** : capacité du classifieur à prédire l'appartenance au TP sachant que les objets en question appartiennent à ce TP / **Spécificité** : capacité du classifieur à prédire la non-appartenance au TP sachant que les objets en question n'appartiennent effectivement pas à ce TP

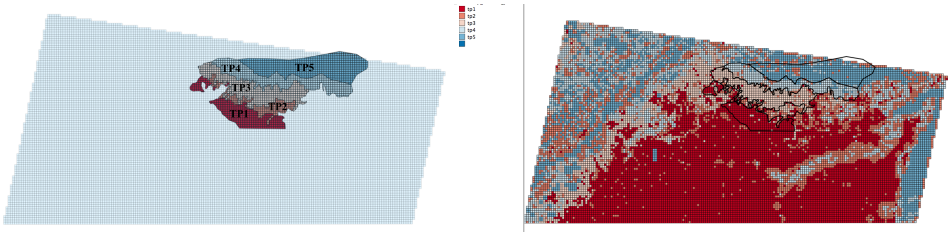
		TP1	TP2	TP3	TP4	TP5
Sur l'ensemble d'apprentissage	Sensibilité	0,98	0,73	0,97	0,93	0,96
	Spécificité	0,98	1,00	0,97	0,99	0,99
Validation croisée moyenne [min ; max]	Sensibilité	0,76 [0,40 ; 1,00]	0,15 [0,00 ; 1,00]	0,64 [0,33 ; 0,83]	0,46 [0,30 ; 0,62]	0,62 [0,25 ; 0,80]
	Spécificité	0,92 [0,82 ; 1,00]	0,92 [0,80 ; 1,00]	0,82 [0,73 ; 0,89]	0,75 [0,60 ; 0.90]	0,83 [0,75 ; 0,94]

TABLE 2: Résultats de prédiction pour la grille de 2040 mètres de résolution.

mais aussi par une identification et une caractérisation moins « franche » de la part de l'expert et de son propre avis.

D'autres tests ont été effectués pour la production des règles par PLI. La relation spatiale d'*adjacence*, deux mailles étant adjacentes si et seulement si elles partagent un même côté, a ainsi été codée dans la base de connaissances du processus d'apprentissage ainsi que le caractère de multi-résolution (relations d'inclusion des mailles d'une grille à l'autre). Cette dernière stratégie d'apprentissage, exploitant l'information de voisinage et le caractère multi-résolution de l'information spatiale, semble particulièrement pertinente : elle correspond à la façon dont l'expert géographe appréhende le territoire, en utilisant différents produits cartographiques, à des emprises géographiques et des résolutions variées. Cependant, cette prise en compte des relations spatiales et de multi-résolution pour la production des règles de classification n'a pas montré d'amélioration véritable des résultats.

La figure 4 présente les paysages qui alimentent la base d'apprentissage pour la PLI (4a) et illustre les résultats de prédiction sur toute la zone d'étude (4b).



(a) Paysages d'apprentissage.

(b) Résultats de prédiction.

FIGURE 4: Paysages de référence et paysages prédits.

Les résultats de prédiction, dont la majeure partie se trouve en dehors de la zone des paysages de référence, débouche sur une caractérisation cohérente du territoire dont la pertinence est confirmée, *a posteriori*, par l'expert.

Modules ontologiques. Différentes passerelles de mise en correspondance des deux modules ontologiques construits ont déjà été explorées. Un premier travail est de mettre en synergie l'expertise du géographe concernant un type paysager tel qu'il est décrit dans l'ontologie du paysage et l'information concernant les mailles typées par la PLI comme appartenant à ce type paysager. Nous définissons à cet effet des requêtes de transformation en langage SPARQL (Ko-

stylev et al., 2015) qui amènent à de nouvelles représentations des connaissances de l'ontologie image, de manière à les rendre conformes au contenu de l'ontologie du paysage. Une réflexion porte ainsi sur ce qui pourrait permettre d'identifier le passage d'un paysage à un autre en se concentrant sur les mailles frontière. La figure 5 fait apparaître le résultat d'une requête (mailles de type TP1 contenant majoritairement des composants paysagers (ou « patchs ») de type forêt et forêt peu dense dans l'ontologie image) que nous traduisons comme étant les mailles de l'écosystème forestier. Ces mailles sont affichées en vert dans la zone d'étude [5]. Dans l'ontologie du paysage, l'écoystème forestier est décrit comme un paysage très majoritairement couvert de forêt et offrant peu de diversité. La tonalité du vert rend compte de la diversité des patchs retrouvés dans les mailles. Le composant forêt est omniprésent quand les mailles sont vert foncé. Les mailles vert clair font apparaître plus de diversité avec en particulier des patchs d'eau ou de végétation intermédiaire. Ces mailles plus riches en composants paysagers sont fréquemment retrouvées sur les limites du paysage écosystème forestier et la diversité est donc un bon indicateur du changement du paysage pour ce qui concerne l'écosystème forestier.

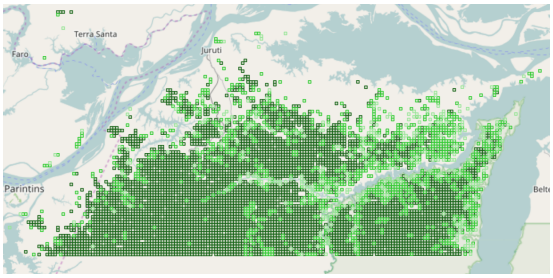


FIGURE 5: Mailles de l'écosystème forestier sur un fond de carte OpenStreetMap.

4.2 Discussion

Nous avons montré tout le potentiel d'une approche permettant, à partir de mécanismes d'induction symbolique, de peupler en partie une ontologie applicative. Les méthodes de fouilles de données habituellement exploitées en traitement d'image, à l'exemple de la classification supervisée, exploitent les attributs en entrée comme des dimensions indépendantes les unes des autres. En ce sens, la PLI qui permet de poser des liens de dépendance entre attributs, est plus à même de tirer parti de l'organisation spatiale de valeurs radiométriques à l'intérieur de mailles vues comme des ensembles de pixels connexes. Nous avons pu ainsi extraire des prédicats mettant en relation plusieurs descripteurs (à l'exemple de la proportion, de la diversité ou de la fragmentation) pour approcher au mieux les types paysagers. Nous avons également montré comment exploiter les relations topologiques avec la PLI : cette piste est encore à creuser. Il nous faut également affiner la description des paysages et améliorer la base d'apprentissage servant à la production des règles de PLI. De la même manière, nous allons faire évoluer nos deux modules ontologiques. En particulier, les relations spatiales entre mailles seront décrites au sein de l'ontologie image et il est attendu également de travailler

5. La visualisation se fait au moyen de la librairie Javascript OpenLayers

sur de nouveaux descripteurs (par exemple, la texture). Enfin, les liens permettant d'exploiter conjointement l'ontologie applicative et l'ontologie des paysages sont également à renforcer. L'objectif est de continuer à construire des règles de transformation en s'appuyant sur SPARQL de façon à pouvoir, à partir d'une régularité sur le graphe de l'ontologie image, en extraire une régularité sur le graphe de l'ontologie des paysages et inversement. La bidirectionnalité nous permettra d'une part, de rapprocher les données de l'image de l'expertise du géographe, d'autre part, de faire évoluer potentiellement les connaissances de l'expert en fonction du contenu de l'image.

5 Conclusion

Nous avons présenté une approche hybride qui se nourrit à la fois de connaissances expertes formalisées au préalable au sein de modules ontologiques, et de nouveaux savoirs apportés par de l'apprentissage symbolique. L'objectif est de coller au plus près de la réalité du géographe expert tout en faisant évoluer les connaissances apprises au moyen de la PLI. Nous avons montré des premiers résultats encourageants dans la capacité à automatiser l'extraction de motifs d'intérêt dans une image satellite. Nous avons comme perspective d'enrichir la représentation ontologique notamment en matière de description de métriques afin d'étoffer les potentialités des mécanismes de raisonnement. La démarche présentée tout au long de cet article, bien que perfectible au niveau de la finesse des résultats, a permis d'atteindre nos objectifs : aller vers une automatisation de l'extraction des paysages dans une image satellite et structurer, pérenniser, enrichir et rendre interopérables les connaissances et données mises en jeu tout au long du processus dans un ancrage résolument orienté web sémantique.

Remerciements. *Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche pour le projet COCLICO portant la référence ANR-12-MONU-0001 au titre du programme MN Modèles Numériques 2012.*

Références

Battle, R. et D. Kolas (2012). Enabling the geospatial semantic web with parliament and geosparql. *Semant. web 3*(4), 355–370.

Bayoudh, M., E. Roux, G. Richard, et R. Nock (2015). Structural knowledge learning from maps for supervised land cover/use classification : Application to the monitoring of land cover/use maps in french guiana. *Computers and Geosciences 76*, 31–40.

Belgiu, M. et J. Thomas (2013). Ontology based interpretation of very high resolution imageries–grounding ontologies on visual interpretation keys. *AGILE 2013—Leuven 2013*(1), 14–17.

Chahdi, H., N. Grozavu, I. Mougenot, Y. Bennani, et L. Berti-Equille (2016). Towards ontology reasoning for topological cluster labeling. In *International Conference on Neural Information Processing*, pp. 156–164. Springer International Publishing.

Dean, M. et G. Schreiber (2004). OWL Web Ontology Language - Reference. W3C recommendation, W3C.

Dérioz, P., P. Béringuier, et A.-E. Laques (2010). Mobiliser le paysage pour observer les territoires : quelles démarches, pour quelle participation des acteurs ? *Développement durable et territoires 1*(2), 11598 – 11607.

Enserink, M. (1999). Biological invaders sweep in. *Science 285*(5435), 1834–1836.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual review of ecology, evolution, and systematics 34*(1), 487–515.

Farina, A. (2000). *Landscape Ecology in Action*. Springer Netherlands.

Forestier, G., A. Puissant, C. Wemmert, et P. Gançarski (2012). Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems 36*(5), 470 – 480.

Gruber, T. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition 5*(2), 199–220.

Kavurucu, Y., P. Senkul, et I. H. Toroslu (2011). A comparative study on ilp-based concept discovery systems. *Expert Systems with Applications 38*(9), 11598 – 11607.

Kostylev, E. V., J. L. Reutter, et M. Ugarte (2015). Construct queries in sparql. In *ICDT*.

Lavrac, N. et S. Dzeroski (1994). *Inductive Logic Programming : Techniques and Applications*. Ellis Horwood.

Malerba, D., F. Esposito, A. Lanza, F. Lisi, et A. Appice (2003). Empowering a gis with inductive learning capabilities : the case of ingens. *Computers, Environment and Urban Systems 27*(3), 265 – 281.

Michalski, R. S. (1983). *Machine learning : An artificial Intelligence Approach*, Chapter a theory and methodology of inductive learning, pp. 110–161. TIOGA Publishing Co.

Muggleton, S. (1991). Inductive logic programming. *New Generation Computing 8*, 295–318.

Richards, J. A. (2013). *Remote Sensing Digital Image Analysis. An Introduction.* Springer-Verlag Berlin Heidelberg.

Srinivasan, A. (2007). The aleph manual.

Vaz, D., M. Ferreira, et R. Lopes (2007). Spatial-yap : a logic-based geographic information system. In *Proceedings of the 23rd international conference on Logic programming*, ICLP'07, Berlin, Heidelberg, pp. 195–208. Springer-Verlag.

Summary

Here we present an innovative method to automatically detect a valuable set of landscape patterns within a satellite image. We are aiming to achieve two major goals. The first deals with capabilities to capture the knowledge shared by experts for image analysis tasks. The second emphasizes ontology supporting roles to retain and reuse knowledge in an extendable and interoperable way. The paper highlights the synergistic use of different strategies including image preprocessing, operations using specific metrics, inductive logic programming (ILP) and knowledge modelling leading to a more global approach. In particular we explore the interplay between an inductive (ILP) and a deductive method (ontology engineering) to fully meet these objectives.