



HAL
open science

Reconstruction automatique de l'occupation du sol à partir d'images aériennes historiques monochromes : une étude comparative

Rémi Ratajczak, Carlos F Crispim-Junior, Élodie Faure, Béatrice Fervers,
Laure Tougne

► To cite this version:

Rémi Ratajczak, Carlos F Crispim-Junior, Élodie Faure, Béatrice Fervers, Laure Tougne. Reconstruction automatique de l'occupation du sol à partir d'images aériennes historiques monochromes : une étude comparative. Conférence Française de Photogrammétrie et de Télédétection (CFPT), Jun 2018, Marne-la-Vallée, France. hal-01827773

HAL Id: hal-01827773

<https://hal.science/hal-01827773v1>

Submitted on 7 May 2019

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.

Reconstruction automatique de l'occupation du sol à partir d'images aériennes historiques monochromes : une étude comparative

R. Ratajczak^{1,2,3} C. Crispim-Junior¹ E. Faure² B. Fervers² L. Tougne¹

¹ Univ Lyon, Lyon 2, LIRIS, F-69676 Lyon, France

² Unité Cancer et Environnement, Centre Léon Bérard, Lyon, France

³ Agence de l'Environnement et de la Maîtrise de l'Energie, Angers, France

{Remi.Ratajczak, Carlos.Crispim-Junior, Laure.Tougne}@liris.cnrs.fr
{Elodie.Faure, Beatrice.Fervers}@lyon.unicancer.fr

Résumé

Cette étude propose trois contributions quant à l'automatisation de la reconstruction de l'occupation du sol à partir d'images aériennes historiques monochromes. Premièrement, un nouveau jeu de données est présenté. Il est constitué d'un total de 4.9 millions d'images acquises entre les années 1970 et 1990, puis annotées par des experts à l'aide de 7 classes à 3 niveaux d'échelle spatiale. Deuxièmement, une étude comparative de l'état de l'art en extraction de caractéristiques et en apprentissage machine a été menée sur 14 filtres, 4 classifieurs et 3 réseaux de neurones profonds à convolution (DCNNs). Enfin, un nouveau filtre de textures et une nouvelle combinaison de filtres texturaux sont présentés. La combinaison proposée a permis d'obtenir des résultats en classification similaires aux DCNNs pour un vecteur de caractéristiques 17 fois plus petit.

Mots Clef

Textures, apprentissage machine, apprentissage profond, occupation du sol, images aériennes historiques.

Abstract

This study proposes three contributions to automate the land cover reconstruction from monochromatic historical aerial images. Firstly, a new dataset is presented. It totalizes 4.9 million historical patches acquired between 1970 and 1990, and annotated by experts with 7 classes over 3 different space scales. Secondly, a comparative study between visual features extractors and machine learning algorithms has been performed. It includes 14 state-of-the-art filters, 4 classifiers, and 3 end-to-end Deep Convolutional Neural Networks (DCNNs). Finally, a novel texture filter and a new combination of texture filters are presented. The proposed combination was able to perform similarly with the DCNNs but, with a 17 times smaller features vector.

Keywords

Textures, machine learning, deep learning, land cover, historical aerial images.

1 Introduction

L'occupation du sol est une donnée qui intéresse de plus en plus la communauté scientifique et les collectivités territoriales pour l'analyse des impacts géo-sociaux des politiques de développement terrestres. Afin d'obtenir des annotations permettant d'observer l'évolution de la terre en continu, les géographes exploitent les données acquises par les programmes satellites tels que LandSat¹, Sentinelle² et Quickbird³. Ces données, mises à la disposition du grand public, représentent des images de télédétection multi-spectrales à hautes résolutions spatiale et temporelle. L'obtention de bases de données de vérité terrain pour ces images est quant à elle facilitée par les efforts menés par les contributeurs aux projets libres tels que OpenStreetMap⁴. La réalisation d'une carte d'occupation du sol détaillée manuellement reste néanmoins une tâche complexe qui mobilise durant plusieurs heures (e.g. 6 à 10h pour 9km²) un géographe expérimenté.

De nombreux algorithmes ont ainsi été développés dans la littérature afin d'aider les praticiens de la télédétection. Issues du domaine de la vision par ordinateur, ces méthodes exploitent pleinement les multiples modalités disponibles ainsi que les éventuelles redondances spatiales présentes au sein d'acquisitions successives. Ces données ont notamment été utilisées pour reconstituer visuellement la surface terrestre masquée par des nuages [12], segmenter sémantiquement des zones urbaines [2] et des parcelles agricoles [13], détecter des bâtiments [3], ou encore générer des modèles d'élévation de terrains [7].

A l'opposé de ces travaux prospectifs réalisés sur des données actuelles, très peu d'études ont à ce jour été menées afin de reconstituer automatiquement des cartes

1. <https://landsat.gsfc.nasa.gov/>

2. <https://sentinel3.cnes.fr/>

3. <http://gslcf.umd.edu/data/quickbird/>

4. <http://www.openstreetmap.fr/>

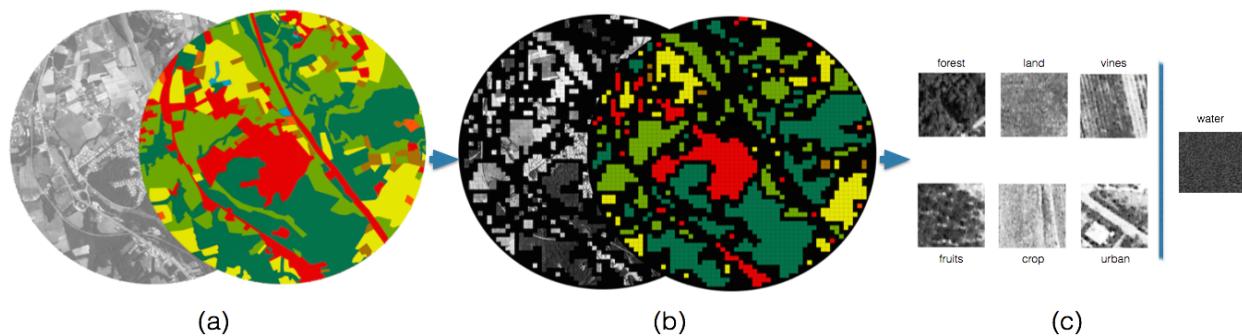


FIGURE 1 – (a) Image aérienne historique et occupation du sol associée, rayon 1.5 km. (b) Après extraction d’imagettes de 100 pixels par 100 pixels. En noir, les imagettes non considérées car correspondant à plusieurs classes. (c) Exemples d’imagettes considérées.

d’occupation du sol en rétrospectif à partir d’images aériennes historiques. Cette information est pourtant essentielle dans de nombreux travaux liés à la santé publique. Les analyses épidémiologiques géocodées à grande échelle, telles que TESTIS [5] et TESTEPERA [4], ont besoin de géolocaliser les champs de culture afin d’inférer la qualité des phytosanitaires utilisés au niveau des lieux de résidences des sujets étudiés. En particulier, l’étude TESTIS [6] a pour but d’analyser l’exposition pendant la vie entière de 1500 sujets adultes, impliquant l’analyse d’images acquises à partir des années 1970.

Actuellement, l’approche la plus couramment employée pour ce type d’analyses consiste en une segmentation manuelle des images aériennes historiques. Ces images sont, pour les moins récentes, disponibles uniquement en niveaux de gris, sans aucun apport colorimétrique, infrarouge ou radar. Cette caractéristique tend à rendre les images historiques plus difficiles à exploiter que les images satellitaires actuelles car moins informatives. Il est aussi important de noter que ces données ont été acquises à une époque où les outils informatiques adéquats pour les analyser n’avaient pas encore été créés. Il n’existe donc qu’un nombre restreint d’annotations disponibles pour ces images. Ces éléments, mis bout-à-bout, représentent un facteur limitant quant à l’ampleur des études épidémiologiques rétrospectives pouvant être menées à l’aide des Systèmes d’Information Géographiques (SIG).

Dans le but de motiver de futurs travaux visant à l’automatisation de la reconstruction de cartes d’occupation sol à partir d’images aériennes historiques, trois contributions sont proposées dans cette étude. Premièrement, un nouveau jeu de données de 4.9 millions d’imagettes annotées a été créé. Celui-ci est constitué de 7 classes d’occupation du sol représentées sur 3 niveaux d’échelle spatiale. Il a été nommé HistAerial. Le nombre d’imagettes présenté (4.9 millions) inclut le jeu de données complet (3.7 millions d’imagettes) ainsi que deux sous-ensembles utilisés pour l’évaluation (1.1 et 0.1 million d’imagettes respective-

ment). Deuxièmement, une comparaison croisée de l’état de l’art en extraction de caractéristiques visuelles et en apprentissage machine a été menée sur HistAerial afin de déterminer les meilleures chaînes de traitements possibles pour l’analyse des images aériennes monochromes. Enfin, deux nouveaux algorithmes d’extraction d’informations texturales sont présentés. Ces deux algorithmes sont capables de générer des vecteurs de caractéristiques de faibles dimensions. Ils ont été intégrés à la comparaison croisée proposée pour un total de 14 filtres, 4 classificateurs et 3 réseaux de neurones profonds à convolution (DCNNs).

2 HistAerial : jeu de données

Le jeu de données qui a été créé puis utilisé dans cette étude a été nommé HistAerial. Il sera publié gratuitement et publiquement⁵.

Images initiales et annotations. HistAerial a été conçu à partir de 81 images aériennes historiques monochromes issues du Géoportail de l’IGN⁶. Ces images ont initialement été acquises en France à l’aide d’un dispositif optique placé sur un avion entre les années 1970 et 1990. La résolution de ces images varie de 0.17 à 1.4 mètres, pour une résolution moyenne estimée à 0.5 mètres, sur un sous échantillon de 25 images. Afin de les analyser, ces images ont été projetées dans le système de coordonnées géographiques RGF93 / Lambert 93 (géoréférencement avec 7 à 8 points clés par image), puis annotées par des géographes de l’unité Cancer et Environnement du Centre Léon Bérard à l’aide de 7 classes d’occupation du sol, à savoir : Verger, Arable, Prairie, Vigne, Urbain, Forêt, Eau. Ces annotations ont été réalisées densément, avec l’ensemble des classes disponibles, pour 56 des 81 images aériennes disponibles. Les autres annotations ont été réalisées de façon éparse afin de combler le manque en données pour certaines classes. Ainsi, 15 images aériennes ont été partiellement annotées avec uniquement la classe Verger, et 10 images ont été partiellement annotées avec uniquement la classe Vigne.

5. <http://eidolon.univ-lyon2.fr/~remi1/HistAerialDataset/>

6. <https://remonterletemps.ign.fr/>

Obtention des imagerie. Des imagerie sans recouvrements et de trois tailles arbitraires (*i.e.* 25 pixels par 25 pixels, 50 pixels par 50 pixels, 100 pixels par 100 pixels) ont été extraites des images aériennes selon une procédure supervisée. Seules les imagerie représentant une seule et unique classe ont été considérées. Pour cela, les images aériennes et leurs annotations ont été appariées et la règle suivante a été établie : si au moins un pixel de l’imagerie n’appartient pas à la même classe que les autres, exclure l’imagerie, sinon la conserver. Ce procédé, partiellement inspiré par les travaux des auteurs de [22], est illustré au travers de la figure 1. La taille maximale des imagerie a été choisie de manière à conserver un nombre suffisant de données pour l’utilisation des algorithmes récents d’apprentissage profond pour chaque échelle d’imagerie proposée. Aucune imagerie plus grande que 100 pixels par 100 pixels n’est donc présente au sein du jeu de données HistAerial. On remarquera que, comme toutes les imagerie ont été extraites de la même base d’images initiales, les imagerie de tailles différentes représentent les mêmes zones géographiques avec un contexte spatial plus ou moins étendu : plus l’imagerie est de grande taille, plus elle intègre d’informations liées à son contexte spatial.

Décomposition du jeu de données. La table 1 représente le jeu de données HistAerial complet. On constate que le nombre d’imagerie diminue exponentiellement avec l’augmentation de leur taille. Cela est dû à la volonté de rendre les imagerie indépendantes les unes des autres via l’absence de recouvrement entre deux imagerie. L’utilisation d’imagerie avec recouvrement pourrait augmenter considérablement la quantité de données disponibles, mais risquerait de générer un biais de représentation dans les données. On remarque aussi que le jeu de données HistAerial complet n’est pas équilibré (*i.e.* certaines classes sont beaucoup plus représentées que d’autres). Afin de mener des études statistiquement viables, deux sous-ensembles du jeu de données HistAerial complet ont été créés par échantillonnage aléatoire. Le premier est dit équilibré en taille (*size-balanced*). Il est représenté par la table 2. Il considère le même nombre d’imagerie par classe et par taille. Ce nombre a été fixé à 6000, ce qui correspond au nombre d’imagerie pour la classe la moins représentée

TABLE 1 – Jeu de données HistAerial complet.

Classes	Nombre d’imagerie par taille (pixels)		
	25 × 25	50 × 50	100 × 100
<i>Verger</i>	319 804	76 866	17 888
<i>Arable</i>	631 015	145 097	30 754
<i>Prairie</i>	348 349	71 334	11 984
<i>Vigne</i>	174 288	40 528	8 889
<i>Urbain</i>	891 500	204 746	43 254
<i>Forêt</i>	443 760	95 945	18 554
<i>Eau</i>	121 294	28 173	6 207
Total	2 930 010	662 689	137 530

TABLE 2 – Jeu de données HistAerial équilibré en taille.

Classes	Nombre d’imagerie par taille (pixels)		
	25 × 25	50 × 50	100 × 100
<i>Verger</i>	6 000	6 000	6 000
<i>Arable</i>	6 000	6 000	6 000
<i>Prairie</i>	6 000	6 000	6 000
<i>Vigne</i>	6 000	6 000	6 000
<i>Urbain</i>	6 000	6 000	6 000
<i>Forêt</i>	6 000	6 000	6 000
<i>Eau</i>	6 000	6 000	6 000
Total	42 000	42 000	42 000

TABLE 3 – Jeu de données HistAerial équilibré en classe.

Classes	Nombre d’imagerie par taille (pixels)		
	25 × 25	50 × 50	100 × 100
<i>Verger</i>	120 000	28 000	6 000
<i>Arable</i>	120 000	28 000	6 000
<i>Prairie</i>	120 000	28 000	6 000
<i>Vigne</i>	120 000	28 000	6 000
<i>Urbain</i>	120 000	28 000	6 000
<i>Forêt</i>	120 000	28 000	6 000
<i>Eau</i>	120 000	28 000	6 000
Total	840 000	196 000	42 000

au sein du jeu de données HistAerial complet (*i.e.* Eau en 100 pixels par 100 pixels). Ce sous-ensemble vise à évaluer les chaînes de traitements pour chaque taille d’imagerie de façon indépendante. Il ne permet cependant pas de comparer l’apport du contexte spatial car la proportion de données échantillonnées pour chaque taille d’imagerie n’est pas identique. Afin de pallier à ce problème, un second sous-ensemble a été créé. Il est dit équilibré en classe (*class-balanced*) et représenté par la table 3. On notera que ces deux sous-ensembles ont été considérés comme des jeux de données indépendants dans cette étude. Les imagerie qu’ils contiennent ont donc été comptées indépendamment de celles de HistAerial complet, portant le nombre total d’imagerie disponible à 4.9 millions sur trois ensembles de données et trois échelles de représentation différentes.

3 Algorithmes

3.1 Filtres : état de l’art

Les algorithmes qui ont été comparés dans cette étude se basent sur l’extraction de caractéristiques de textures. Les textures sont en effet des caractéristiques privilégiées pour l’analyse de données de télédétection. Les images acquises depuis un dispositif aérien ou spatial représentent la surface terrestre d’un point de vue orthogonal aux objets observés. Ces derniers apparaissent alors sous la forme de motifs quasi-réguliers correspondant à la définition des textures inhomogènes en vision par ordinateur.

Textures classiques. Les géographes disposent déjà d'outils leur permettant d'obtenir des informations statistiques considérées comme représentatives de la texture à partir de matrices de co-occurrences (GLCM), aussi connues sous le nom de caractéristiques de Haralick [9]. Depuis ces travaux précurseurs, de nombreuses méthodes ont vu le jour pour l'analyse texturale. Dans cette étude, les filtres basés sur le *Local Binary Pattern* (LBP) [19] ont été privilégiés. Malgré leur faible complexité algorithmique, ces filtres se positionnent actuellement en références sur de nombreux jeux de données représentant des textures [15].

Définition des LBP. Les filtres basés sur le LBP sont des filtres locaux. Ils calculent localement un code binaire représentatif de l'information texturale en utilisant un voisinage circulaire de pixels g_p , centré sur un pixel central g_c . Le voisinage est paramétré par le nombre de pixels voisins à considérer, P , et le rayon du voisinage, R (voir équation (1)). Il est possible de représenter un voisinage (P, R) à l'aide de coordonnées discrètes ou continues [20]. Dans le premier cas, la valeur d'un pixel voisin est obtenue en considérant la valeur discrète la plus proche de la position réelle de l'élément (*i.e.* continue) sur le cercle de rayon R . Dans le second cas, la valeur d'un pixel voisin à la position réelle est obtenue par interpolation bilinéaire. Dans cette étude, seuls des voisinages continus ont été utilisés.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

Représentation. Bien que les LBP reposent sur des relations locales, le résultat d'un tel filtre est généralement représenté à l'aide d'un histogramme de l'image filtrée. Cet histogramme possède un nombre de bins déterminé par la définition du filtre. A titre d'exemple, le LBP original génère un histogramme de taille 2^P , alors que le *Local Ternary Pattern* (LTP) [25] génère ou bien un histogramme de 3^P bins, ou bien un histogramme de 2^{P+1} bins en fonction de la représentation choisie. De plus, certaines méthodes, telles que le *Completed Local Binary Pattern* (CLBP) [8] et le *Extended Local Binary Pattern* (ELBP) [17], concatènent les histogrammes de plusieurs filtres complémentaires (nommés sous-filtres par la suite). Ces dernières méthodes tendent à générer des histogrammes de tailles plus importantes au profit de meilleurs résultats en classification sur les bases de données de textures telles que Outex [18].

Mapping. Afin de réduire la taille des histogrammes générés par les filtres LBP, plusieurs méthodes dites de mise en correspondance (*mapping*) d'histogrammes ont été élaborées. La plus utilisée est celle de l'uniformité et de l'invariance à la rotation (*rotation invariance and uniform, riu²*). Cette opération exploite la symétrie circulaire du voisinage utilisé. L'uniformité correspond à l'utilisation d'un unique bin pour tous les LBP correspondants à un motif similaire, tels qu'une ligne ou un point. L'invariance à la rotation consiste à utiliser un même bin pour tous les motifs iden-

tiques modulo une rotation près. L'application de l'opération *riu²* permet de diminuer la taille d'un histogramme LBP à seulement $P+2$ bins, contre 2^P précédemment, tout en lui inférant la propriété d'invariance aux rotations dans le plan. Son utilisation permet donc d'obtenir de meilleurs résultats sur des jeux de données représentant des textures similaires mais orientées différemment.

Propriétés. Les filtres de type LBP présentent l'avantage d'être, par définition, invariants aux variations globales d'illumination. Cette propriété rend cette catégorie de filtres très adaptée à l'analyse d'images de textures acquises dans des conditions non contrôlées telles que les images aériennes. A l'opposé, ces filtres sont sensibles aux phénomènes de bruits blancs entachant parfois les données. Certaines méthodes ont été définies afin de tenir compte de ce défaut. Elles sont considérées comme étant robustes au bruit (*e.g.* *Robust Local Ternary Pattern* (RLTP) [27], *Median Robust Local Binary Pattern* (MRELBP) [16]).

3.2 Filtres : contributions

Deux nouvelles méthodes d'extraction de textures sont introduites dans cet article. Le développement de ces méthodes a été motivé par la volonté d'obtenir des vecteurs de caractéristiques compréhensibles, performants et de relativement faibles dimensions. En effet, des vecteurs de caractéristiques de faibles dimensions sont synonymes d'apprentissages et de comparaisons plus rapides. Cette propriété facilite l'intégration des algorithmes employés au sein d'architectures matérielles non optimisées pour les calculs telles que les machines utilisées par les praticiens (*e.g.* processeurs non graphiques, contraintes mémoires, *etc.*). Le fait d'avoir un vecteur de caractéristiques de faible dimension permet aussi d'envisager l'utilisation des méthodes proposées au sein d'outils d'apprentissage semi-supervisés intégrant des interactions avec l'utilisateur pour la mise en place d'apprentissages dynamiques (*online*).

Rotated Corner Local Binary Pattern. La première méthode qui est présentée dans cet article est une simplification du *Binary Gradient Contour* (BGC), utilisé par Aguilar et al. [1], lui-même inspiré du LBP présenté en section 3.1. Il a été nommé *Rotated Corner Local Binary Pattern* (R-CRLBP). La simplification proposée réside en une décomposition d'un schéma particulier du BGC à l'aide d'une combinaison orthogonale [28]. Cette combinaison consiste à considérer l'ensemble des pixels sur le voisinage en découplant les motifs orthogonaux. Ici, il s'agit de calculer la différence successive entre les pixels radialement espacés de $\frac{P}{4}$ pixels, P étant un multiple de 8. Afin de considérer l'ensemble des pixels voisins, un total de $\frac{P}{4} - 1$ rotations sont à appliquer au premier motif orthogonal considéré. On notera que, contrairement au LBP classique, le pixel central g_c n'est pas utilisé ici. L'approche par combinaison orthogonale permet de réduire la taille du vecteur de caractéristiques (histogramme) produit par le filtre de 2^P à $\frac{P}{4} \times 2^4$ tout en considérant la même quantité d'information, à savoir l'intégralité des P pixels voisins

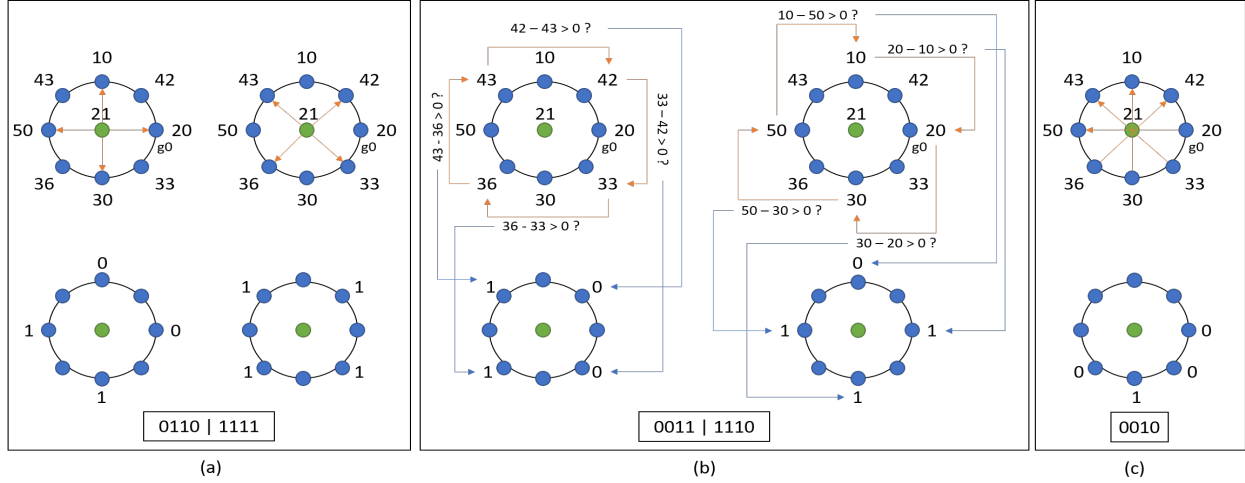


FIGURE 2 – Exemples de filtres LBP avec codes binaires (*i.e.* valeurs retournées) associés : (a) LBP avec combinaison orthogonale, (b) R-CRLBP, (c) CSLBP.

présents sur un cercle de rayon R . Le procédé correspondant au R-CRLBP est représenté par l'équation (2). Sur la figure 2, il est comparé au LBP et au *Center Symmetric Local Binary Pattern* (CSLBP) [10], qui représentent deux autres manières classiques d'utiliser un voisinage circulaire (P, R) pour l'analyse de textures.

$$R - CRLBP_{P,R} = \sum_{p=\alpha, p+=\frac{P}{4}}^{\alpha+\frac{3 \times P}{4}} s(g_p - g_{p-(\frac{P}{4})} 2^i) \quad (2)$$

avec $i = \frac{p-\alpha}{\frac{P}{4}}$ et $\alpha = (0, 1, \dots, \frac{P}{4} - 1)$

Light Combination of Local Binary Pattern. La seconde méthode d'extraction présentée consiste en une combinaison de plusieurs filtres LBP générant des vecteurs de caractéristiques de faibles dimensions. Elle a été nommée *Light Combination of Local Binary Pattern* (LCoLBP). L'hypothèse derrière cette combinaison est qu'un ensemble d'extracteurs générant des vecteurs de caractéristiques de faibles dimensions est capable de capturer plus d'informations discriminantes qu'un unique filtre résultant en un vecteur de dimension supérieure ou égale. En pratique, le LCoLBP concatène les histogrammes issus du R-CRLBP, décrit ci-dessus, du FPLBP [26], du CSLBP [10] et du XCSLBP [24]. Ces trois derniers filtres génèrent chacun un histogramme de $2^{\frac{P}{2}}$ bins pour un voisinage (P, R) donné. Les 4 filtres formant LCoLBP sont, par ailleurs, complémentaires : ils permettent de représenter, pour un voisinage (P, R) donné, les motifs texturaux interne (CSLBP) et robustes au bruit (XCSLBP), ceux présents sur les contours du voisinage (R-CRLBP), et ceux présents à la proximité extérieure du voisinage (FPLBP). La concaténation des histogrammes résultant de ces filtres produit un unique histogramme de $\frac{P}{4} \times 2^4 + 3 \times 2^{\frac{P}{2}}$ bins (voir équation (3)).

$$histo(LCoLBP) = concat(histomap([R - CRLBP, FPLBP, CSLBP, XCSLBP])) \quad (3)$$

Méthodologie d'application. Dans cette étude, 12 filtres de la littérature et 2 nouveaux filtres ont été implémentés en C++ avec OpenCV [6] et comparés avec comme paramètres $P = 8$ et $R = (1, 2, 3)$ (trois échelles de filtres). Ils sont référencés dans la table 4. Les filtres de la littérature ne sont pas détaillés dans cet article par soucis de concision. L'efficacité du *mapping riu*² a été étudiée pour le filtre LBP original. Il a été appliqué de façon systématique pour tous les filtres, ou sous-filtres pour les combinaisons de filtres, générant un histogramme de taille 2^P ou plus.

3.3 Classifieurs

Les algorithmes de classification qui ont été comparés sont très succinctement présentés ci-dessous. Les paramètres de ces méthodes ont été obtenus à l'aide d'une recherche dense sur grille de paramètres (*gridsearch*) via la librairie Scikit-Learn en Python [21].

K plus proches voisins. La méthode des KNN a été utilisée avec $K = 1$ et une distance euclidienne. Cette distance présente l'avantage d'être applicable efficacement sur de nombreuses données, statistiques incluses.

Machine à vecteurs de support. La méthode SVM non linéaire a été utilisée en classification multi-classe avec la Fonction de Base Radiale (RBF) comme noyau.

Fôret aléatoire d'arbres décisionnels. Cette méthode (RF) a été appliquée avec un nombre d'arbres décisionnel arbitrairement fixé à 100.

Perceptron multi-couches. Le MLP a été utilisé avec un nombre de couches cachées sélectionné (automatiquement) entre 1 et 3, et un nombre de neurones par couche proportionnel à la taille T_v du vecteur de caractéristiques v (respectivement $0.75\% \times T_v$, $0.5\% \times T_v$, $0.25\% \times T_v$).

3.4 Réseaux de Neurones à convolution

Les réseaux de neurones profonds à convolution qui ont été utilisés dans cette étude sont issus de la littérature. Ces

TABLE 4 – Meilleurs résultats obtenus sur le jeu de données HistAerial équilibré en taille. Entre parenthèses, le classifieur correspondant.

Filtres texturaux						
Filtre	Paramètres (P, R)	Mapping	Nombre de caractéristiques	Classification - Top-1 (%)		
				25 × 25	50 × 50	100 × 100
LBP [20]	(8, {1,2,3})	riu^2	30	67.5 (RF)	79.0 (RF)	87.4 (KNN)
LBP [20]	(8, {1,2,3})	<i>none</i>	768	66.7 (SVM)	80.5 (KNN)	89.1 (KNN)
VARLBP [20]	(8, {1,2,3})	riu^2	414	69.6 (RF)	80.3 (RF)	84.5 (RF)
CSLBP [10]	(8, {1,2,3})	<i>none</i>	48	60.8 (RF)	68.6 (RF)	80.3 (RF)
XCSLBP [24]	(8, {1,2,3})	<i>none</i>	48	65.9 (RF)	78.3 (RF)	86.0 (RF)
TPLBP [26]	(8, {1,2,3})	riu^2	30	62.3 (RF)	73.6 (RF)	80.1 (RF)
FPLBP [26]	(8, {1,2,3})	<i>none</i>	48	59.9 (MLP)	74.0 (RF)	85.2 (RF)
CLBP [8]	(8, {1,2,3})	riu^2	66	72.1 (RF)	80.9 (RF)	88.1 (RF)
LTP [25]	(8, {1,2,3})	riu^2	60	69.2 (MLP)	80.4 (RF)	88.0 (RF)
RLTP [27]	(8, {1,2,3})	riu^2	60	63.8 (RF)	76.6 (RF)	85.3 (RF)
SCCOCRLTP [27]	(8, {1,2,3})	<i>none</i>	384	54.5 (RF)	76.8 (RF)	85.5 (RF)
ELBP [17]	(8, {1,2,3})	riu^2	66	57.2 (RF)	77.9 (RF)	84.8 (RF)
MRELBP [16]	(8, {1,2,3})	riu^2	66	57.4 (RF)	71.8 (RF)	85.9 (RF)
R-CRLBP	(8, {1,2,3})	<i>none</i>	96	66.9 (MLP)	78.8 (RF)	86.7 (RF)
LCoLBP	(8, {1,2,3})	<i>none</i>	240	72.9 (MLP)	82.9 (RF)	89.3 (RF)

Réseaux de neurones profonds à convolution						
Modèle	Optimisation	Epochs	Nombre de caractéristiques	Classification - Top-1 (%)		
				25 × 25	50 × 50	100 × 100
LeNet [14]	RMSPROP	40	500	60.2	68.3	72.3
AlexNet [11]	SGD	40	4096	73.6	82.0	—
AlexNet [11]	RMSPROP	40	4096	—	—	90.4
VGG-16 [23]	SGD	40	4096	70.3	80.0	—
VGG-16 [23]	RMSPROP	40	4096	—	—	89.1

TABLE 5 – Meilleurs résultats obtenus sur le jeu de données HistAerial équilibré en classe.

Meilleures méthodes appliquées						
Méthode	Paramètres	Mapping	Nombre de caractéristiques	Classification - Top-1 (%)		
				25 × 25	50 × 50	100 × 100
LCoLBP + RF	(8, {1,2,3})	<i>none</i>	240	75.0	84.1	89.3
AlexNet + SGD	40 epochs	—	4096	73.4	85.6	—
AlexNet + RMSPROP	40 epochs	—	4096	—	—	90.4

méthodes ont démontré leur capacité à apprendre automatiquement des représentations de caractéristiques efficaces pour classifier des images. Les architectures LeNet [14], AlexNet [11] et VGG-16 [23] ont été comparées. On notera que LeNet a été utilisé, et ce malgré ses performances généralement inférieures aux réseaux plus récents, car il a initialement été conçu pour des images en niveaux de gris comme celles présentes au sein du jeu de données HistAerial, alors que AlexNet et VGG-16 ont été conçus pour des images couleurs - dans notre cas, nous extrapolons nos images en niveaux de gris sur trois canaux colorimétriques. En pratique, les poids de chaque réseau ont été optimisés sur les jeux de données pour lesquels il ont été créés, puis affinés (*i.e. finetuning*) sur HistAerial. Par ailleurs, il a été supposé que les méthodes de redimensionnement bilinéaires appliquées sur des images carrées ne modifient pas l'information contenue relativement d'une image à

l'autre. Ces dernières ont été utilisées pour générer des images carrées de taille identique à celles attendues en entrée des différents réseaux. Pour chaque architecture, le meilleur taux d'apprentissage (*learning rate*) a été déterminé parmi (0.01, 0.001, 0.0001, 0.00001), l'entraînement a été réalisé sur 40 Epochs (la base de données d'entraînement a été vue 40 fois), et deux méthodes de mise à jour des poids des réseaux (*i.e.* optimisation) ont été comparées (SGD et RMSPROP).

4 Résultats et discussions

Les résultats présentés dans cette section correspondent aux meilleurs taux de réussites obtenus pour chaque filtre et pour chaque taille d'image. Chaque classifieur a été entraîné pour chaque filtre. Les jeux de données équilibrés ont été découpés aléatoirement en ensembles d'entraîne-

ment (63%), de validation (27%) et de test (10%). Les modèles ont été entraînés pour chaque jeu de paramètres sur les données d'entraînement et de validation. Ils ont ensuite été comparés sur les données de test. Les résultats présentés sur la table 4 correspondent aux meilleurs scores obtenus, tous classifieurs confondus, sur l'ensemble de test issu du jeu de données HistAerial équilibré en taille. Les résultats présentés sur la table 5 correspondent aux meilleurs scores obtenus sur le jeu de données HistAerial équilibré en classe avec les meilleures méthodes issues de la table 4.

Evaluation des méthodes. Les résultats obtenus sur le jeu de données HistAerial équilibré en taille (voir table 2) montrent que le filtre LCoLBP proposé obtient les meilleurs résultats parmi les filtres texturaux lorsqu'il est combiné avec une forêt aléatoire d'arbres décisionnels. Les scores obtenus par le LCoLBP sont similaires à ceux du meilleur des réseaux de neurones à convolution qui ont été comparés (ici, AlexNet). Ces deux méthodes obtiennent respectivement des taux de classification de 72.9% et 73.6% sur les imagerie de 25 pixels par 25 pixels, 82.9% et 82.0% sur les imagerie de 50 pixels par 50 pixels et 89.3% et 90.4% sur les imagerie de 100 pixels par 100 pixels. Par ailleurs, le LCoLBP résulte en un vecteur de caractéristiques de 240 bins lorsqu'il est appliqué avec un voisinage de $P=8$ et $R=(1,2,3)$, contre 4096 pour AlexNet. Pour un vecteur de caractéristiques 17 fois plus petit, LCoLBP permet donc d'obtenir des résultats similaires à ceux de AlexNet (légèrement inférieurs, 1.1% avec les imagerie les plus grandes). Dans les deux cas, les classes les plus confondues sont la classe Arable et la classe Prairie, avec, dans le meilleur des cas (100 pixels par 100 pixels), 10% (valeur moyenne) de confusion pour LCoLBP et 12% pour AlexNet. Les classes Verger et Vigne sont quant à elles peu confondues : dans le pire des cas (25 pixels par 25 pixels), 5.5% de confusion Vigne/Verger pour AlexNet et 2% pour LCoLBP.

Importance du contexte spatial. L'importance du contexte spatial a été évaluée sur le jeu de données HistAerial équilibré en classe (voir table 5) à l'aide des méthodes ayant obtenu les meilleurs résultats sur le jeu de données HistAerial équilibré en taille (*i.e.* LCoLBP et AlexNet). Les résultats de cette évaluation sont présentés sur la table 5. On constate que le LCoLBP et AlexNet obtiennent des résultats similaires sur les deux jeux de données. L'apport en données supplémentaires n'a qu'un impact faible sur l'évaluation de ces méthodes (gain de 2.1% (LCoLBP, 25×25) et 3.6% (AlexNet, 50×50) au maximum). Les méthodes comparées obtiennent de meilleurs résultats avec des imagerie de plus grande taille. Le contexte spatial semble donc avoir une importance significative (gain de 15%) sur la classification des images aériennes historiques texturées.

5 Conclusion

Dans cette étude, un nouveau jeu de données a été proposé pour l'analyse d'images aériennes historiques mono-

chromes. Il est composé de 4.9 millions d'imagerie annotées à trois niveaux d'échelle spatiale. Une comparaison croisée de la littérature a été réalisée. Deux nouveaux filtres ont été proposés. Parmi eux, le LCoLBP a permis d'obtenir des résultats similaires aux réseaux de neurones profonds à convolution pour un vecteur de caractéristiques 17 fois plus petit. Enfin, l'importance du contexte spatial dans l'analyse d'images aériennes historiques monochromes a été observée. Dans de futurs travaux, l'utilisation de méthodes hybrides, exploitant des réseaux de neurones à convolution combinés à des méthodes d'extraction de textures classiques, sera étudiée. L'utilisation d'éventuelles corrélations entre les données satellites actuelles et les images aériennes historiques sera analysée.

Annexe

Cette étude a été financée par l'Agence De l'Environnement et de la Maîtrise de l'Energie (ADEME), la Plateforme interface Santé-Environnement en Rhône-Alpes (Envitéra) et le Centre Léon Bérard. Les auteurs remercient Etienne Durant, Géographe, pour les annotations des images aériennes historiques.

Références

- [1] M. A. Aguilar, A. Fernández, F. J. Aguilar, F. Bianconi, and A.G. Lorca. Classification of urban areas from GeoEye-1 imagery through texture features based on Histograms of Equivalent Patterns. *European Journal of Remote Sensing*, 49(1) :93–120, January 2016.
- [2] A. Albert, J. Kaur, and M. C. Gonzalez. Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale. pages 1357–1366. ACM Press, 2017.
- [3] O. Aytekin, A. Erener, I. Ulusoy, and A. Duzgun. Unsupervised building detection in complex urban environments from multispectral satellite imagery. *International Journal of Remote Sensing*, 33(7) :2152–2177, April 2012.
- [4] R. Béranger, J. Blain, C. Baudinet, E. Faure, A. Fléchon, H. Boyle, V. Chasles, B. Charbotel, J. Schuz, and B. Fervers. Testicular germ cell tumours and early exposures to pesticides : The TESTEPERA pilot study. *Bulletin Du Cancer*, 101(3) :225–235, March 2014.
- [5] R. Béranger, O. Pérol, L. Bujan, E. Faure, J. Blain, Ch. Le Cornet, A. Flechon, B. Charbotel, T. Philip, J. Schuz, and B. Fervers. Studying the impact of early life exposures to pesticides on the risk of testicular germ cell tumors during adulthood (TESTIS project) : study protocol. *BMC Cancer*, 14(1), December 2014.
- [6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [7] S. Coveney and K. Roberts. Lightweight UAV digital elevation models and orthoimagery for environmental

applications : data accuracy evaluation and potential for river flood risk modelling. *International Journal of Remote Sensing*, pages 1–22, February 2017.

- [8] Guo, Z., Zhang, L., and D. Zhang. A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *IEEE Transactions on Image Processing*, 19(6) :1657–1663, June 2010.
- [9] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6) :610–621, November 1973.
- [10] M. Heikkila, M. Pietikainen, and C. Schmid. Description of interest regions with center-symmetric local binary patterns. In *5th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'06)*, volume 4338, pages 58–69. Springer-Verlag, 2006.
- [11] Alex K., I. Sutskever, and G. E. Hinton. Image-net classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [12] N. Kussul, M. Lavreniuk, A. Shelestov, and B. Yailymov. Along the season crop classification in Ukraine based on time series of optical and SAR images using ensemble of neural network classifiers. In *The International Geoscience and Remote Sensing Symposium 2016 (IGARSS 2016)*, pages 7145–7148. IEEE, July 2016.
- [13] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5) :778–782, May 2017.
- [14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.
- [15] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen. Local binary features for texture classification : Taxonomy and experimental study. *Pattern Recognition*, 62 :135–160, February 2017.
- [16] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikainen. Median Robust Extended Local Binary Pattern for Texture Classification. *IEEE Transactions on Image Processing*, 25(3) :1368–1381, March 2016.
- [17] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth. Extended local binary patterns for texture classification. *Image and Vision Computing*, 30(2) :86–99, February 2012.
- [18] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen. Outex - new framework for empirical evaluation of texture analysis algorithms. In *Proceedings - International Conference on Pattern Recognition*, volume 1. IEEE Comput. Soc.
- [19] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In *Computer Vision - ECCV 2000*, volume 1842, pages 404–420. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. DOI : 10.1007/3-540-45054-8_27.
- [20] T. Ojala, M. Pietikäinen, and T. Mäenpää. A Generalized Local Binary Pattern Operator for Multi-resolution Gray Scale and Rotation Invariant Texture Classification. In *Advances in Pattern Recognition, ICAPR 2001*, pages 399–408. Springer Berlin Heidelberg, 2001.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [22] A. Porebski, N. Vandenbroucke, L. Macaire, and D. Hamad. A new benchmark image test suite for evaluating colour texture classification schemes. *Multimedia Tools and Applications*, 70(1) :543–556, May 2014.
- [23] Karen S. and Andrew Z. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [24] C. Silva, T. Bouwmans, and C. Frélicot. An extended center-symmetric local binary pattern for background modeling and subtraction in videos. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2015*, 2015.
- [25] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. *IEEE Transactions on Image Processing*, 19(6) :1635–1650, June 2010.
- [26] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Workshop on faces in real-life images : Detection, alignment, and recognition*, 2008.
- [27] X. Wu, J. Sun, G. Fan, and Z. Wang. Improved Local Ternary Patterns for Automatic Target Recognition in Infrared Imagery. *Sensors*, 15(12) :6399–6418, March 2015.
- [28] C. Zhu, C.-E. Bichot, and L. Chen. Image region description using orthogonal combination of local binary patterns enhanced with color information. *Pattern Recognition*, 46(7) :1949–1963, July 2013.