



Vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major du 19 e siècle

Pierre-Alexis Herrault, David Sheeren, Mathieu Fauvel, Martin Paegelow

► To cite this version:

Pierre-Alexis Herrault, David Sheeren, Mathieu Fauvel, Martin Paegelow. Vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major du 19 e siècle. *Revue Internationale de Géomatique*, 2015, 25 (1), pp.35-51. hal-01513411

HAL Id: hal-01513411

<https://univ-tlse2.hal.science/hal-01513411>

Submitted on 25 Apr 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.

Vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major du 19^e siècle

Pierre-Alexis Herrault^{*,**}, David Sheeren^{*}, Mathieu Fauvel^{*},
Martin Paegelow^{**}

^{*}Université de Toulouse, INP-ENSAT, UMR 1201 DYNAFOR
Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, 31326 Castanet Tolosan cedex
pierrealexis.herrault@ensat.fr;david.sheeren@ensat.fr;mathieu.fauvel@ensat.fr

^{**}Université de Toulouse, UTM, UMR 5602 GEODE
5 Allées Antonio Machado, 31058 Toulouse cedex-France
paegelow@univ-tlse2.fr

RÉSUMÉ. Dans cet article, nous proposons une nouvelle méthode pour extraire automatiquement les forêts dans les cartes d'Etat-major du 19^{ème} siècle numérisées. La démarche proposée est constituée de quatre étapes principales : filtrage de l'image, changement d'espace colorimétrique, identification des forêts à l'aide d'un détecteur non paramétrique (SVDD), post-traitement. La méthode est suffisamment robuste pour prendre en compte la diversité des représentations possibles des forêts dans ces cartes anciennes. Les résultats montrent des performances élevées avec une précision globale de détection obtenue de 95%. Cette approche ouvre de nouvelles perspectives pour les différentes études environnementales incluant une dimension historique.

ABSTRACT. In this article, we propose a new method to automatically extract forest features in historical 'Map of France' from 19th century. The proposed method is based on four major steps: filtering of the image, color-space conversion, extraction of the forests based on a non-parametric detection algorithm (SVDD), post-processing. The method is sufficiently robust to take into account the diversity of the representation of the forests in historical map. Results show high performances with a global accuracy of 95%. This approach offers new perspectives in the different environmental studies including a historical dimension.

MOTS-CLÉS : carte ancienne, CIELab, SVDD, morphologie mathématique.

KEYWORDS : historical map, CIELab, SVDD, mathematical morphology.

1. Introduction

Les cartes anciennes contiennent une riche quantité d'informations sur le passé concernant l'occupation du sol, leur structure, la morphologie des villes, les monuments, etc. L'extraction de ces informations présente un intérêt pour de nombreuses études qui portent par exemple sur les transformations des paysages sur le long terme, le développement urbain, l'évolution du trait de côte, et le changement global ([Cousins, 2001](#) ; [Bender *et al.*, 2005](#) ; [Gimmi *et al.*, 2011](#) ; [Smith and Cromley, 2012](#)).

A l'heure actuelle, la vectorisation des éléments figurant sur ces cartes et leur intégration dans une base de données géographiques est généralement menée par digitalisation, avec une forte intervention humaine. Or, pour des raisons évidentes de temps et de coût, cette procédure est peu reproductible sur de vastes étendues ce qui limite la réalisation d'études historiques spatialisées.

Plusieurs travaux se sont intéressés au problème de vectorisation automatique de cartes topographiques, cartes géologiques ou documents plus anciens ([Lichtner, 1985](#) ; [Ansoult *et al.*, 1990](#) ; [Samet et Soffer, 1998](#) ; [Wise, 1999](#), [Leyk *et al.*, 2006](#) ; [Chiang *et al.*, 2012](#) ; [Shaw et Bajcsy., 2011](#) ; [Iwanowski et Kozak 2012](#) ; [Mahmoud, 2012](#)). La difficulté varie suivant le type de carte à traiter (couleur ou noir et blanc), sa qualité (déformation ou non du document numérisé) et le niveau de surcharge de différents éléments planimétriques (superposition des courbes de niveau, toponymes et réseaux divers sur les entités d'occupation du sol). Les méthodes développées sont souvent spécifiques à la carte considérée et peu reproductibles sur une autre source de données. Toutefois, il émerge des différents travaux une démarche commune pour vectoriser automatiquement ces cartes. Elle se compose de trois étapes principales : (1) une phase de « nettoyage », (2) une étape d'identification des éléments d'intérêt et (3) une étape de post-traitement.

Dans cet article, nous proposons une nouvelle méthode de vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major datant du milieu du 19^e siècle ([Dupouey *et al.*, 2007](#)). Il n'existe actuellement aucune méthode de vectorisation automatique dédiée à cette carte qui est pourtant largement exploitée dans différentes études, notamment en écologie historique ([Andrieu *et al.* 2011](#)). Le temps estimé pour la vectorisation manuelle des seules données d'occupation du sol (n'incluant pas les réseaux de routes et chemins) est d'environ 27h, soit environ 4 jours de travail à temps plein (comité de pilotage 2013 – projet de production d'une carte des forêts et autres occupations anciennes des sols de France Métropolitaine). Fondée sur des techniques de traitement d'image, la procédure proposée adopte la démarche générale de vectorisation en incluant un changement d'espace colorimétrique et l'application d'un détecteur supervisé non-paramétrique (Support Vector Data Description).

L'article est structuré de la manière suivante. Une brève description des caractéristiques de la carte d'Etat-Major est donnée en section 2. Nous décrivons en section 3 la méthode proposée avec les principales étapes qui la composent. Nous présentons l'application de cette méthode sur différents extraits de minutes et

discutons des résultats en section 4. Nous concluons en proposant différentes perspectives en section 5.

2. Les minutes de la carte d'Etat-Major

La méthode de vectorisation proposée est définie pour la reconnaissance des formations boisées dans les minutes de la carte d'Etat-Major de la France. Cette carte a été constituée au milieu du 19^e siècle, de 1818 à 1866, à partir de levés d'échelle variable suivant le lieu et des plans cadastraux napoléoniens (Dupouey *et al.*, 2007). La carte générale d'Etat-Major couvre l'ensemble du territoire français au 1:80.000^e. Elle est éditée en noir et blanc. Ses minutes sont en couleur et produites au 1:40.000^e. Leur qualité géométrique est réputée assez fiable mais très variable suivant le lieu. Le système de coordonnées planimétriques de la carte est basé sur la projection dite de Bonne avec comme ellipsoïde de référence celui de Plessis 1817. Il s'agit d'une projection équivalente dont les différents paramètres sont connus.



Figure 1. Exemple original d'une minute d'Etat-Major (feuille n°241NE-SAINT GAUDENS, sud est de Toulouse)

Les minutes de la carte utilisées dans cette étude ont été mises à disposition par l'IGN au format numérique (voir exemple Figure 1). Les images fournies sont codées sur 3 octets (24 bits) correspondant aux 3 couleurs primaires (R, V, B). La résolution spatiale des images est de 4 mètres pour un scannage à 254 dpi. L'IGN fournit les minutes non pas dans le système géodésique initial mais dans le système légal de référence actuel (Lambert-93). Toutefois, il s'agit d'un géoréférencement très sommaire basé sur une transformation affine avec 4 points d'amer. Ce type de transformation n'est pas suffisant pour corriger correctement ces cartes anciennes. Ainsi, des décalages de 60 à 100m peuvent apparaître sur cette carte géoréférencée.

Il est ainsi nécessaire d'améliorer le géoréférencement des minutes pour les rendre superposables aux données cartographiques actuelles. Une procédure de géoréférencement adaptée aux cartes anciennes et fondée sur un modèle de régression à noyau a été proposée et appliquée dans ce sens ([Herrault *et al.*, 2013](#)).

3. Méthode de vectorisation automatique

La méthode de vectorisation automatique des forêts que nous proposons est composée de différentes étapes (Figure 2). Nous détaillons chacune de ces étapes dans les sections suivantes.

3.1. Pré-traitements pour le filtrage d'éléments en surcharge

Avant de procéder à la détection des forêts proprement dite, une étape de nettoyage est nécessaire pour éliminer des éléments planimétriques qui se superposent aux polygones représentant les forêts. Ces éléments concernent principalement certains symboles, la toponymie et les hachures rendant compte du relief. En plaine, ces hachures sont peu nombreuses. En revanche, leur densité est très importante en milieu de montagne ce qui rend le processus de filtrage délicat. La procédure proposée vise à éliminer ces hachures quel que soit le contexte orographique tout en préservant le contour des forêts. Trois étapes successives sont appliquées : (1) dilatation morphologique, (2) filtrage de type médian, (3) filtrage de type passe-bas.

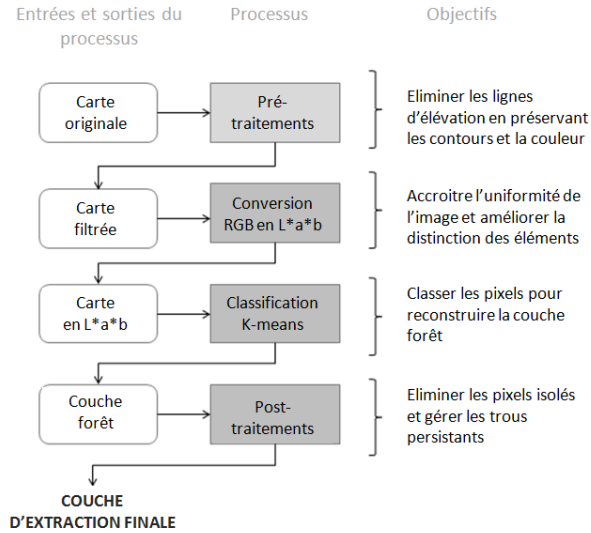


Figure 2. Chaîne de traitement suivie pour l'extraction des forêts sur les minutes d'Etat-Major

La dilatation morphologique (étape 1 des pré-traitements) vise à combler la totalité des trous à l'intérieur des forêts, formés par les symboles, les éléments textuels et les hachures. L'élément structurant choisi est un carré de taille 5x5. Cet opérateur est appliqué à chaque bande de l'image (R, V, B) en niveaux de gris.

Le filtre médian est ensuite mis en œuvre à l'issue de la dilatation pour éliminer les hachures persistantes (étape 2 des pré-traitements). La taille de fenêtre du filtre est la même mais il est appliqué plusieurs fois jusqu'à obtenir une image suffisamment nettoyée. Le nombre d'applications du filtre varie en fonction de la densité des hachures présente sur la carte et donc du relief de la zone. Ce nombre est déterminé de manière empirique. Contrairement à un filtre moyen, ce filtre médian présente l'intérêt de respecter les contours des objets.

Un dernier filtre de type passe-bas et de taille 5x5 est finalement appliqué pour supprimer le bruit résiduel.

3.2. Transformation de l'espace colorimétrique : du RGB vers le CIELab

Les étapes précédentes ont permis de supprimer un certain nombre d'éléments se superposant aux forêts. Néanmoins, le filtrage a également altéré les couleurs de l'image. Les teintes ont été modifiées et ceci, au sein d'une même catégorie d'objet puisqu'en fonction du relief, la surcharge des hachures n'est pas la même. D'une manière plus générale, la couleur est rarement homogène dans les cartes historiques. L'intensité est susceptible de varier plus ou moins fortement suivant le lieu. Le

changement d'espace colorimétrique est nécessaire afin de rendre plus robuste la reconnaissance automatique des objets sur la base de leur couleur.

Le choix d'un espace colorimétrique adapté est une étape importante dans la vectorisation des cartes historiques en couleur. Il conditionne souvent la qualité du résultat final (Ivanowski *et al.*, 2012). L'espace traditionnellement utilisé est l'espace RVB qui présente certaines limites pour la question qui nous occupe : la luminosité n'est pas uniforme en plus du fait qu'il est plus éloigné de la perception humaine (Angulo et Serra, 2003). Les espaces colorimétriques cylindriques dans lesquels la luminosité est isolée sont mieux adaptés à la reconnaissance d'objets dont la luminosité n'est pas uniforme.

Dans notre approche, nous proposons de transformer l'image initiale dans l'espace colorimétrique CIELab. Ce choix s'inspire de travaux précédents montrant que le CIELab est bien approprié aux images bruitées colorées (Angulo et Serra, 2003 ; Ganesan *et al.*, 2010).

Dans l'espace CIELab, la composante de clarté de la couleur est reportée sur l'axe L (Luminosity), variant de 0 à 100, et perpendiculaire à une pile de plan « ab » (Figure 3). Les composantes « a » et « b » sont définies par une gamme de valeurs variant respectivement du rouge au vert en passant par le gris, et du jaune au bleu en passant par le gris. Chacun de ces plans contient ainsi toutes les couleurs possibles pour une valeur de clarté. Dans notre cas, nous pouvons avoir plusieurs valeurs de clarté (i.e. plusieurs valeurs de L) pour une même couleur (ici le vert pour les forêts). La reconnaissance de ces objets est alors fondée sur la recherche des différentes partitions de vert existant pour chaque valeur possible de L . Cette recherche est effectuée à l'aide d'un algorithme de détection non paramétrique décrit ci-dessous.

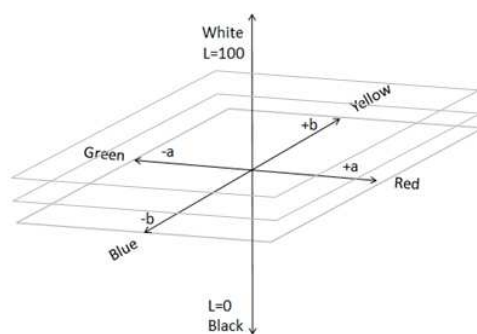


Figure 3. Modèle de représentation cartésien de l'espace CIELab

3.3 Extraction des forêts à partir du détecteur SVDD

La reconnaissance des forêts dans l'espace colorimétrique est réalisée au moyen de l'algorithme de détection supervisée SVDD (*Support Vector Description*) (Sanchez-Hernandez *et al.*, 2007 ; Fauvel *et al.*, 2013). Deux raisons principales expliquent ce choix :

- contrairement à un algorithme de reconnaissance multi-classe, cette méthode supervisée ne nécessite pas un échantillonnage exhaustif de la classe négative (i.e. des exemples correspondant aux autres classes que celle recherchée) ou du moins, pas de manière exhaustive (Tax et Duin, 2004). Cela rend la sélection des exemples d'apprentissage moins contraignante et plus rapide ;
- contrairement à un algorithme de classification supervisée paramétrique (tel que le modèle de mélange gaussien), le détecteur SVDD ne nécessite pas l'estimation des distributions des classes à partir des exemples d'apprentissage. Cela rend l'approche également plus souple car la variabilité des représentations des forêts peut conduire à l'obtention d'une distribution multi-modale de la classe. Par ailleurs, la distribution des classes négatives (autres que forêt) peut être difficile à modéliser en raison d'une grande diversité des objets présents dans la carte. Cela nécessiterait un nombre important d'exemples d'apprentissage négatifs qui rendrait la méthode moins opérationnelle.

Concernant son fonctionnement, l'algorithme SVDD fait partie des méthodes d'apprentissage à noyau (Shawe-Taylor et Cristianini, 2004). Il consiste à trouver l'hypersphère de rayon minimal qui englobe les exemples de la classe d'intérêt en excluant les autres. Lorsque les données ne sont pas distribuées de manière sphérique dans l'espace initial (ici L^*a*b), il est possible de projeter implicitement les données dans un autre espace de représentation H à l'aide d'une fonction noyau k (Shawe-Taylor et Cristianini, 2004) tel que :

$$\begin{aligned} x_i &\rightarrow \phi(x) \\ \mathbb{R}^d &\mapsto H \end{aligned}$$

Pour les séparations non linéaires, le choix d'un noyau adapté est un problème crucial dans l'utilisation de ces méthodes. Un noyau idéal permettrait d'englober les échantillons positifs dans une région de forme sphérique (ou au moins dans une région délimitée) au sein de l'espace des données d'apprentissage. Les noyaux doivent donc être à support compact. C'est la raison pour laquelle nous avons fait le choix d'utiliser la fonction noyau RBF :

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

et non les noyaux polynomiaux par exemple.

De manière plus formelle, appelons $S = (x_i, y_i)$ le jeu d'exemples d'apprentissage avec $x_i \in \mathbb{R}^d$ (où \mathbb{R}^d est l'espace vectoriel de dimension d) et $y_i = 1$ si x_i est positif et $y_i = -1$ si c'est un exemple négatif. Les paramètres $\hat{\beta}$ de l'hypersphère optimale sont déterminés en résolvant le problème d'optimisation suivant (pour les détails, voir Tax et Duin 2004) :

$$\hat{\beta} = \max_{\beta} [\sum_{i=1}^n \beta_i y_i k(x_i, x_i) - \sum_{i,j=1}^n \beta_i \beta_j y_i y_j k(x_i, x_j)] \quad (2)$$

avec $i, j \in \{1, \dots, n\}$ et sous les contraintes $C \geq \beta_i \geq 0$ et $\sum_{i=1}^n \beta_i y_i = 1$. (3)

Où C est le paramètre de régularisation compris entre 0 et 1 qui pénalise les erreurs durant l'optimisation. Ce paramètre permet de régler la proportion d'erreurs que l'on souhaite autoriser durant l'apprentissage. Quand $C = 1$, aucune erreur n'est admise. Des valeurs de C inférieures permettent des erreurs pour une part des échantillons négatifs et positifs ce qui augmente la capacité d'induction.

La valeur au carré \mathcal{R}^2 du rayon de l'hypersphère est donnée par (1) à l'optimal. La distance euclidienne au carré d_c au centre c de l'hypersphère pour un exemple z est donnée par :

$$d_c(z) = k(z, z) + \sum_{i,j=1}^n \hat{\beta}_i \hat{\beta}_j y_i y_j k(x_i, x_j) - 2 \sum_{i=1}^n \hat{\beta}_i y_i k(z, x_i) \quad (4)$$

La détection est effectuée en évaluant si la distance d_c pour un exemple z est plus grande ou plus petite que le rayon de l'hypersphère \mathcal{R}^2 . Dans le premier cas, il s'agit d'une classe négative (l'exemple a dépassé la frontière de l'hypersphère). Dans le second, il s'agit d'une classe positive.

L'hyper paramètre σ est déterminé comme la moyenne des distances entre les échantillons positifs.

$$\sigma^2 = \frac{2}{(n_+-1)(n_+-1)} \sum_{i=1}^{n_+} \sum_{j=i+1}^{n_+} \|x_i - x_j\| \quad (5)$$

Avec n_+ le nombre d'échantillons de la classe positive.

3.4 Post-traitement

A l'issue de la détection des forêts effectuée à l'étape précédente, un post-traitement est réalisé sur le masque extrait pour corriger certains artéfacts. Un opérateur morphologique d'ouverture est d'abord appliqué pour supprimer les pixels isolés. Certains trous résiduels sont ensuite comblés par l'application d'une règle

contextuelle : les groupes de pixels de petite taille « non forêt » (taille = 10*10) totalement bordés par des forêts sont reclassés en forêt. L'image originale ne contenant que les forêts est finalement reconstruite.

4. Résultats

La méthode proposée a été testée sur trois extraits différents des minutes d'Etat-major. Ils se différencient par leur contexte orographique, leur surcharge d'éléments planimétriques et la couleur qui est utilisée pour représenter les entités « forêt » (différents niveaux de vert). Le premier extrait est issu d'une carte dans la Région de Saint-Gaudens au sud-est de Toulouse (feuille n°241NE). Le second est situé à proximité de Mende (feuille n°196SE). Le troisième extrait se situe près de Tarbes (feuille n°240NE) (Figure 4).



Figure 4. Trois extraits sélectionnés pour l'application de la méthode proposée.

L'étape de filtrage opérée en première partie du protocole de traitement a permis d'éliminer les éléments planimétriques se superposant aux polygones représentant les forêts. Premièrement, l'opérateur de dilatation a réduit efficacement le nombre de pixels dans les basses fréquences. Ces pixels qui correspondent aux lignes d'élévation et à la toponymie ont été supprimés tout en préservant les contours d'objet.

Cette étape est apparue toutefois insuffisante pour procéder directement à la classification des pixels d'intérêt. Un nombre trop important d'éléments sombres est encore visible au sein des entités à extraire à la suite de ce processus. L'application successive d'un filtre médian a alors permis d'éliminer ces artefacts persistants en transférant les pixels sombres vers les hautes fréquences. Enfin, l'ultime opération de cette étape de filtrage a pour but de lisser le bruit de fond granuleux et améliorer l'uniformité de l'image. L'application du filtre passe-bas a bien répondu à ce problème en conservant correctement les détails de l'image. L'image obtenue par cette procédure permet de procéder à l'extraction des pixels d'intérêt (Figure 5).



Figure 5. Résultat du filtrage pour l'extrait 2.

La méthode de classification SVDD offre la possibilité d'introduire des éléments négatifs et de régulariser la proportion d'éléments d'apprentissage à maintenir en dehors de l'hypersphère. Comme indiqué précédemment, fixer le paramètre C à 1 revient à n'accepter aucune erreur dans le processus d'apprentissage. L'objectif étant d'aboutir à un compromis intéressant pour les trois extraits testés, nous décidons de fixer la valeur de C à 0,5.

La figure 6 illustre l'influence du paramètre de régularisation du détecteur SVDD sur la procédure d'apprentissage pour l'extrait 1. Pour cela, nous avons effectué trois phases d'apprentissage à l'aide du même jeu d'échantillon (négatifs et positifs) basées sur trois valeurs différentes du paramètre C ($C=0,1$; $C=0,5$; $C=1$). À noter que cette phase d'apprentissage et la procédure de classification qui en résulte sont à présent effectuées dans l'espace CIELab. Comme attendu, une faible valeur de C (ex : 0,1) conduit à accepter une plus grande part des échantillons négatifs dans l'hypersphère. Détecter les forêts sur l'image entière avec cette valeur de C conduirait à classer trop de pixels « non-forêt » dans la classe « forêt ». Lorsque C est élevé (ex : 1), le détecteur ajuste parfaitement le nuage des échantillons positifs. Il y aura alors de très forts risques de sur-ajustements quand il s'agira de détecter la classe d'intérêt sur l'image entière. En pratique, ce compromis est appris automatiquement par validation croisée (Fauvel *et al.*, 2013). Une valeur de C à 0,5 respecte ces deux critères. L'ensemble du nuage d'échantillons positifs est contenu dans l'hypersphère mais une part raisonnable de pixels négatifs est autorisé dans l'hypersphère afin d'éviter les problèmes de sur-ajustement (*overfitting*).

La chaîne de traitement globale a été appliquée aux trois extraits présentés précédemment. La validation des résultats est effectuée par comparaison avec les couches de forêts obtenues par vectorisation manuelle sur l'ensemble des extraits. Le tableau 1 fournit les résultats d'extraction calculés à partir de matrices de confusion.

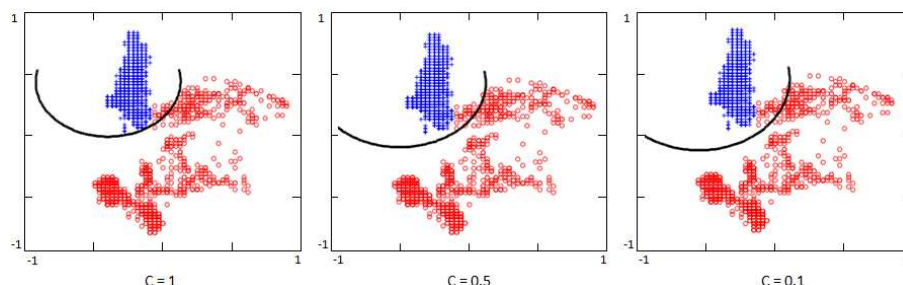


Figure 6. Impact du paramètre de régularisation C sur la procédure d'apprentissage pour l'extrait 1. Les pixels d'apprentissage de la classe positive sont représentés en bleu et les pixels de la classe négative sont représentés en rouge. La ligne noire représente la limite de décision trouvée par le détecteur SVDD.

D'une manière générale, on observe une très bonne performance de la méthode avec une précision globale variant approximativement de 95 à 97% et un indice Kappa supérieur à 0,90 en moyenne. Si on se focalise sur la qualité de l'extraction des forêts proprement dite, on observe également de bons taux d'accord (>92%) avec des erreurs d'excédents et de déficits variant approximativement de 1,5 à 10 %. Pour les extraits 1 et 3, les taux de déficits sont très proches (~7%) et dépassent les taux d'excédents (tendance à la sous-détection). L'effet inverse est observé pour l'extrait 2 (tendance à la sur-détection).

Tableau 1. Résultats de l'extraction opérée sur les trois extraits sélectionnés

	Extrait 1	Extrait 2	Extrait 3
Précision globale (toutes classes) %	96,7	95,35	97,03
Indice Kappa	0,91	0,89	0,93
Taux d'accord (classe Forêt) %	92,3	95,71	92,3
Excédent (classe Forêt) (%)	5,45	10,32	1,53
Déficit (classe Forêt) (%)	7,67	4,29	7,66

La méthode d'extraction automatique a été développée sous la plate-forme MatLab et le logiciel ENVI 4.8. Sous ENVI, une interface dédiée a été conçue pour faciliter l'usage de cette méthode par les thématiciens (Figure 7). Chaque étape de la méthode fait appel à des fonctions qui peuvent être utilisées de manière indépendante. Cela permet à l'utilisateur, s'il le souhaite, d'adapter la démarche selon sa propre source ou ses propres objectifs d'extraction.

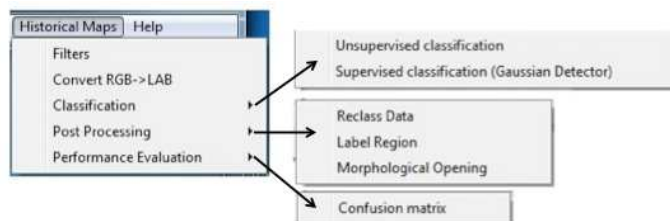


Figure 7. Aperçu de l'interface « Historical Maps » disponible sous ENVI/IDL. Cette extension regroupe la totalité des outils nécessaires à la mise en place de la démarche proposée. L'outil est disponible à cette adresse : <http://fauvel.mathieu.free.fr/software.html>

5. Discussion

Comme indiqué précédemment, les résultats montrent globalement une très bonne performance de la méthode de vectorisation automatique. Les erreurs de déficits et d'excédents proviennent en majorité de défauts d'extraction au niveau du contour des objets. Ces défauts sont principalement liés à l'étape de filtrage réalisée en amont ou en aval de la détection des forêts. Ils ne portent donc surtout que sur des parties d'objets (les lisières). Toutefois, on observe également des erreurs qui couvrent une superficie plus importante et qui concernent la totalité des objets (Figure 8). Pour ces erreurs, l'objet peut ne pas avoir été détecté en raison d'une surcharge trop importante de hachures conduisant à la perte de la couleur et donc de l'objet, après filtrage (cas du déficit de grande superficie sur l'extrait 1 ou 2, Figure 8). Les erreurs « grossières » d'excédents apparaissent quant à elles lorsque la forêt est adjacente à une prairie et que la surcharge de hachures est également importante. Etant donné que les prairies sont de couleur bleu tirant sur le vert, l'effet du lissage peut rendre les couleurs des prairies plus proches de celles des forêts ce qui favorise la sur-détection (cf. extrait 2, Figure 8).

D'une manière plus générale, les erreurs de déficits et d'excédents indiquent que la méthode proposée a tendance tantôt à sur-détecter les forêts, tantôt à les sous-détecter. En plus des raisons précédemment évoquées, deux autres explications peuvent être avancées :

- le compromis déterminé dans la mise en œuvre du détecteur SVDD conduit à accepter une part d'échantillons négatifs dans l'hypersphère afin d'éviter les problèmes de sur-ajustement des échantillons positifs. Les paramètres retenus dans la phase d'apprentissage et appliqués ensuite pour extraire les pixels d'intérêt dans l'image entière jouent un rôle déterminant. Certains pixels appartenant à d'autres classes ont été classifiés comme appartenant à la couche « forêt » ;

- malgré une phase de post-traitement visant en partie à éliminer les pixels isolés, les paramètres de l'élément structurant utilisé pour l'ouverture morphologique (3*3) n'ont pu corriger l'ensemble des imperfections observées dans les masques résultants.

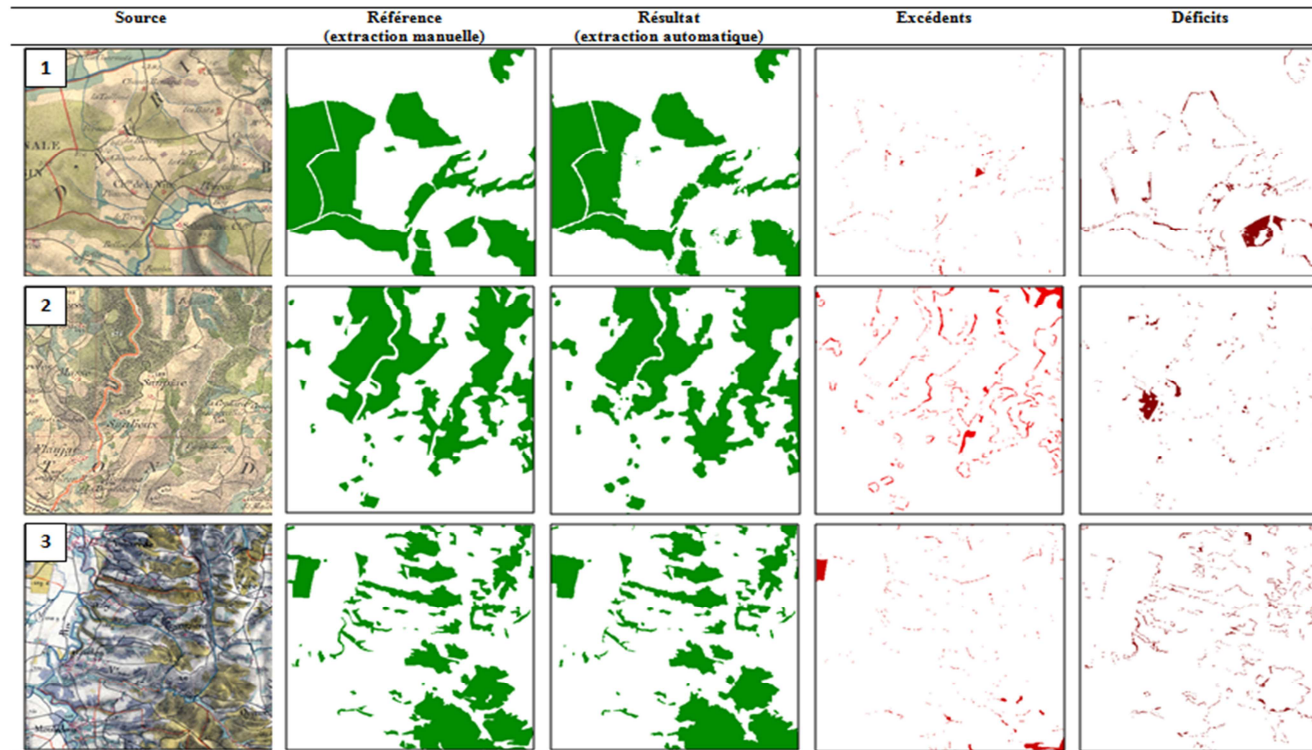


Figure 8. Comparaison de l'extraction automatique et de la référence avec une spatialisation des erreurs d'excédents et de déficits

La présence d'erreurs et leur impact potentiel dépendent de l'application à laquelle sont destinées ces données. Dans notre cas, les résultats d'extraction produits par l'outil développé ont vocation à être utilisés pour des travaux en écologie historique (modélisation des relations espèces/habitats en tenant compte de l'état du paysage passé). Cela requiert de minimiser les erreurs de confusion et d'exhaustivité mais également les erreurs géométriques. En effet, les modèles écologiques envisagés s'appuient sur des variables morphologiques (surface des forêts, longueur de lisières) et des variables contextuelles (connectivité, densité de l'espace boisé environnant). La qualité de représentation des forêts extraites est donc tout aussi importante que la qualité sémantique et l'exhaustivité.

Afin de déterminer l'adéquation de la méthode au cadre d'application écologique, nous avons comparé les valeurs de quatre indicateurs géométriques mesurés à l'échelle des forêts et testé s'il existait une différence significative entre les objets vectorisés automatiquement et ceux vectorisés manuellement. Les boîtes à moustaches (Figure 9) et les résultats du test statistique de Wilcoxon (tableau 2) montrent qu'il n'existe pas de différences significatives entre les variables géométriques comparées. Seule la distance au plus proche voisin pour l'extrait 2 diffère significativement. Cela suggère que la méthode proposée, même entachée d'erreurs, ne devrait pas biaiser de manière importante la construction de modèles écologiques.

Tableau 2. Résultats des tests de comparaisons de Wilcoxon effectués entre les variables calculées à partir de la démarche automatique proposée et la digitalisation manuelle. Les résultats significativement différents au seuil alpha de 5% ($p\text{-value} \leq 0,05$) sont représentés en gras.

	Extrait 1	Extrait 2	Extrait 3
	p-value	p-value	p-value
Surface	0,6467	0,2424	0,2735
Périmètre	0,1417	0,4039	0,7057
Périmètre/surface	0,6181	0,123	0,2903
Distance au plus proche voisin	0,5571	0,0482	0,3442

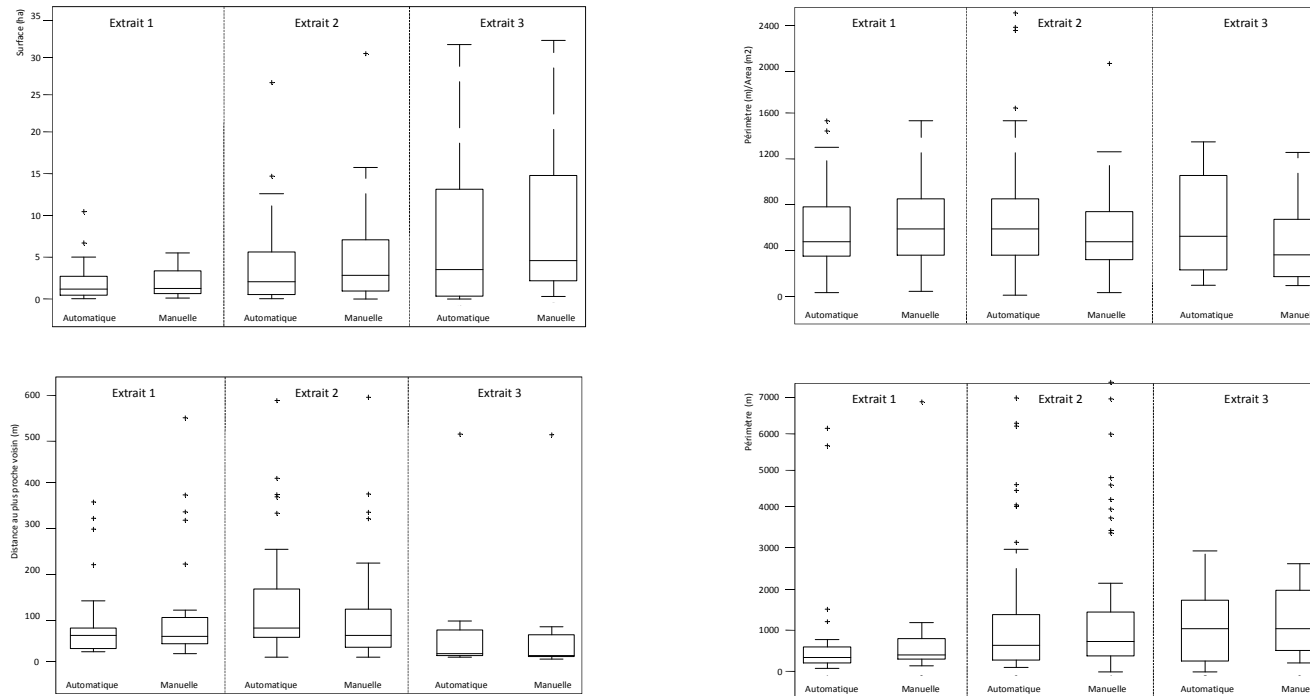


Figure 9. Comparaisons de la distribution des indicateurs morphologiques calculés à partir de la vectorisation automatique et de la vectorisation manuelle de référence.

6. Conclusion

Dans cet article, une méthode d'extraction automatique des forêts sur les minutes de la carte d'Etat-Major du 19^e siècle est développée. La procédure montre des résultats performants avec une précision globale autour de 95 %. La première phase de filtrage a permis d'éliminer efficacement les objets se superposant aux forêts par une succession de filtres morphologiques et de convolution. En ce qui concerne la modification de l'espace colorimétrique de la carte, l'espace CIELab paraît bien approprié au traitement des cartes historiques. En accroissant l'uniformité de l'image, il offre la possibilité d'effectuer une classification des pixels d'intérêt de manière efficace. Quant au détecteur SVDD, il fournit des résultats d'extraction très précis. La possibilité d'introduire des échantillons négatifs sans pour autant échantillonner de manière exhaustive la classe négative rend plus facile l'identification de la classe d'intérêt. Dans cette étude le noyau gaussien est utilisé et les résultats expérimentaux montrent qu'il remplit correctement l'objectif d'extraction.

Les résultats pourraient être améliorés en considérant trois points majeurs. Premièrement, une recherche approfondie du paramètre de régularisation C du détecteur SVDD permettrait d'établir une limite de décision de manière plus fine. Deuxièmement, il serait intéressant d'adapter la procédure de pré-traitement en fonction de la densité des lignes de relief observée. Le recours à un modèle numérique de terrain (MNT) permettrait par exemple de stratifier le processus de filtrage en fonction de pentes plus ou moins fortes. Finalement, les post-traitements devraient faire l'objet d'une attention particulière afin d'accroître la précision des résultats issus de la procédure de classification.

Les perspectives de recherche portent sur l'application de la démarche proposée à d'autres types de cartes présentant des caractéristiques similaires (surcharge d'éléments planimétriques, représentation en couleur des catégories d'occupation des sols). Le fait de tester la méthode sur d'autres cartes scannées permettrait d'accroître la généralité de la démarche et d'alimenter les recherches dans ce domaine.

Remerciements

Les auteurs tiennent à remercier Perrine Loussert et Kevin Gouerec, étudiants du Master SIGMA (Science de l'Information Géoréférencée pour la Maîtrise de l'environnement et l'Aménagement des territoires, INP-ENSAT/UTM), pour l'implémentation de l'outil « Historical Maps » dans la plateforme ENVI/IDL.

Bibliographie

- Andrieu E., Ladet S., Heintz W., Deconchat M. (2011). History and spatial complexity of deforestation and logging in small private forests. *Landscape and Urban Planning*, vol. 103, n°2, p. 109-117.
- Angulo J., Serra J. (2003). Mathematical morphology in color spaces applied to the analysis of cartographic images. *Actes du colloques GEOPRO 2003*, Mexico City, Mexique.
- Ansault M., Soille P., Loodts J. (1990). Mathematical morphology: a tool for automated GIS data acquisition from scanned thematic maps. *Photogrammetric Engineering and Remote Sensing*, vol. 56, n° 9, p. 1263-1271.
- Bender O., Boehmer H.J., Jens D., Schumacher K.P. (2005). Using GIS to analyse 200 years of cultural landscape change in Southern Germany. *Landscape and Urban Planning*, vol. 70, p. 111-125.
- Chiang Y.-Y., Leyk S., Knoblock C. A. (2012). Efficient and Robust Graphics Recognition from Historical Maps. *Proceedings of the Ninth International Workshop on Graphics Recognition GREC 2011*, Séoul, Corée du Sud.
- Cousins S. (2001) Analysis of land-cover transitions based on 17th and 18th century cadastral maps and aerial photographs. *Landscape Ecology*, vol. 16, p. 41-54.
- Dupouey J.L., Bachacou J., Cosserat R., Aberdam S., Vallauri D., Chappart G., Corvisier de Villèle, M.A. (2007). Vers la réalisation d'une carte géoréférencée des forêts anciennes de France. *Le Monde des Cartes*, vol. 191, p. 85-98.
- Fauvel M., Arbelot B., Benediktsson J.A., Sheeren D., Chanussot J. (2013). Detection of Hedges in a Rural Landscape Using a Local Orientation Feature: From Linear Opening to Path Opening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, n°1, p.15-26.
- Ganesan P., Rajini V., Rajkumar I R. (2010). Segmentation and Edge Detection of Color Images Using CIELAB Color Space and Edge Detectors, *Actes du colloque INTERACT 2010*, Chennai, Inde.
- Gimmi U., Lachat T., Burgi M. (2011). Reconstructing the collapse of wetland networks in the Swiss lowlands 1850–2000. *Landscape ecology*, vol. 26, n°8, p. 1071-1083.
- Herrault P.-A., Sheeren D., Fauvel M., Monteil C., Paegelow M. (2013). A comparative study of geometric transformation models for the historical 'Map of France' registration. *Geographia Technica*, vol. 8, n°1, p. 34-46.
- Iwanowski M., Kozak J. (2012). Automatic detection of forest regions on scanned old maps, <http://pe.org.pl/articles/2012/4a/62.pdf>
- Leyk S., Boesch R., Weibel R. (2006). Saliency and semantic processing: Extracting forest cover from historical topographic maps. *Pattern recognition*, vol. 39, n° 5, p. 953-968.
- Lichtner W. (1985). Pattern recognition procedures for automatic digitizing of cadastral Maps. *Actes du colloque AutoCarto 7 (Digital Representations of Spatial Knowledge)*, Washington D.C, USA.

- Mahmoud M.I. (2012). *Information extraction from paper maps using oriented object analysis*. Thèse en gestion de l'information géographique, Faculty of Geo Information Science and Earth Observation of the University of Twente.
- Samet H., Soffer A. (1998). Magellan: Map acquisition of geographic labels by legend analysis. *International journal on document analysis and recognition*, vol. 1, n° 2, p. 89-101.
- Sanchez-Hernandez C., Boyd D. S., Foody G. M. (2007). One class classification for mapping a specific land-cover class: SVDD classification of fenland. *IEEE Transactions Geosciences Remote Sensing*, vol. 45, n° 4, p. 1061-1073.
- Shaw T., Bajcsy P. (2011). Automation of Digital Historical Map Analyses. *Proceedings of iS&T/SPIE Electronic Imaging 2011*, San Francisco, USA.
- Shawe-Taylor J., Cristianini N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge University Press.
- Smith M.J., Cromley R.G. (2012). Measuring Historical Coastal Change using GIS and the Change Polygon Approach. *Transactions in GIS*, vol. 16, n°1, p. 3-15.
- Tax D. M. J., R. Duin P. W. (2004). Support vector data description. *Machine Learning*, vol. 54, n° 1, p. 45-66.
- Wise S. (1999). Extracting raster GIS data from scanned thematic maps. *Transactions in GIS*, vol. 3, n° 3, p. 221-237.