



HAL
open science

Vectorisation semi-automatique de planches scannées du cadastre ancien

Charlotte Odie

► **To cite this version:**

Charlotte Odie. Vectorisation semi-automatique de planches scannées du cadastre ancien. Sciences de l'ingénieur [physics]. 2017. dumas-01835690

HAL Id: dumas-01835690

<https://dumas.ccsd.cnrs.fr/dumas-01835690>

Submitted on 11 Jul 2018

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.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
ECOLE SUPERIEURE DES GEOMETRES ET TOPOGRAPHES

MEMOIRE

présenté en vue d'obtenir

le DIPLOME D'INGENIEUR CNAM

SPECIALITÉ : Géomètre et Topographe

par

Charlotte ODIE

Vectorisation semi-automatique de planches scannées du cadastre ancien

Soutenu le 06 Juillet 2017

JURY

PRESIDENT: M. CHARLET Christophe

MEMBRES:

M. FOLLIN Jean-Michel	Maître de stage
Mme SERVIÈRES Myriam	Maître de stage
Mme SIMONETTO Élisabeth	Maître de stage
M. TOURRE Vincent	Maître de stage
Mme THOMMERET Nathalie	Professeur référent

Remerciements

Tout d'abord, Je tiens à remercier mes maîtres de stage, Mme Elisabeth SIMONETTO et M. Jean-Michel FOLLIN pour leur aide, leur compétence ainsi que leur temps. Je les remercie tout spécialement pour le suivi et l'encadrement qu'ils m'ont apportés. Je remercie également Mme Myriam SERVIÈRES et M. Vincent TOURRE mes responsables au sein de l'IRSTV.

Ensuite, je remercie mon professeur référent, Mme Nathalie THOMMERET pour ses remarques avisées. Je la remercie également pour ses remarques lors de la phase de rédaction et de correction de ce rapport.

Je remercie également M. Gilles BERTEAU pour son expertise au sujet du cadastre et tout particulièrement pour avoir organisé une visite au sein du SDNC. Je remercie les agents du SDNC pour leur temps et pour leurs renseignements.

Je remercie aussi M. Frédéric DURAND d'avoir pris le temps de m'expliquer certains aspects liés à l'utilisation du serveur de calculs.

Je remercie mes camarades de la T0 pour leurs conseils et leur convivialité pendant ces cinq mois de stage.

Glossaire

Une **classification** est une opération de traitement d'image qui fournit une carte thématique, où chaque pixel est assigné à une classe thématique définie par l'opérateur. Une méthode **supervisée** s'appuie sur des échantillons de chaque classe, généralement saisis manuellement.

Un **point d'appui** est un point qui possède à la fois des coordonnées dans le système de référence et qui est repérable sur la donnée à géoréférencer. Dans notre cas, il servira au calcul des paramètres de la transformation géométrique à appliquer aux données à géoréférencer.

Un **point de contrôle** est un point d'appui, qui n'est pas utilisé dans le calcul de la transformation géométrique mais pour estimer la qualité de celle-ci.

Un **point de liaison** est un point repérable sur les données à relier entre elles. Dans notre cas, le point de liaison est connu en coordonnées dans le système terrain grâce à la donnée de référence. C'est donc aussi un point d'appui.

La **squelettisation** consiste à retirer au fur et à mesure les points du contour de la forme, tout en préservant ses caractéristiques topologiques.

La **transformée de Hough** est une technique de reconnaissance de formes inventée en 1962 par Paul Hough, utilisée dans le traitement d'images numériques.

La **vectorisation** est le « procédé de conversion de symboles et de lignes d'une carte raster vers un objet vecteur : un point, une ligne ou un polygone [...] Une représentation vecteur d'une carte est plus abstraite et possède plus de fonctionnalités que son homologue raster, mais perd une partie des informations contextuelles » (Gregory et Ell, 2007).

Table des matières

REMERCIEMENTS	2
GLOSSAIRE	3
TABLE DES MATIERES	4
INTRODUCTION	6
1 ÉTAT DE L'ART	8
1.1 L'ETABLISSEMENT DES PLANS CADASTRAUX DEPUIS 1813	8
1.1.1 LES PLANS DU CADASTRE NAPOLEONNIEN	8
1.1.2 LA RENOVATION ET LA CONSERVATION DU PLAN CADASTRAL	9
1.2 LES PROBLEMATIQUES DE LA VECTORISATION DES DONNEES CARTOGRAPHIQUES	12
1.2.1 NUMERISATION ET VECTORISATION DU CADASTRE EN FRANCE	12
1.2.2 ETAT DE L'ART SUR LA VECTORISATION DE CARTES	13
1.3 LE PROJET DU LABORATOIRE GEOMATIQUE ET FONCIER DU CNAM	15
1.3.1 LA CHAINE METHODOLOGIQUE ELABOREE PAR L'EQUIPE L2G DU LABORATOIRE GEF	15
1.3.2 LES ETAPES TRAITEES DE LA CHAINE METHODOLOGIQUE DU GEF	16
1.3.3 LES REALISATIONS DE CE TFE – LES ETAPES TRAITEES DE LA CHAINE METHODOLOGIQUE	17
2 CHAINE METHODOLOGIQUE	18
2.1 METHODOLOGIES MISES EN PLACE EN FONCTION DE L'ANCIENNETE DES PLANCHES	18
2.2 ETAPES COMMUNES AUX DEUX METHODES : LE PRETRAITEMENT	21
2.2.1 CONVERSION EN NIVEAU DE GRIS	21
2.2.2 APPLICATION D'UN MASQUE ET DETECTION DES POINTS DE CONTOUR PAR SEUILLAGE	21
2.2.3 SUPPRESSION D'OBJETS ISOLES	22
2.2.4 SQUELETTISATION	22
2.2.5 OPERATEURS MORPHOLOGIQUES	23
2.3 PREMIERE METHODE : EXTRACTION DES SEGMENTS PAR TRANSFORMEE DE HOUGH PROBABILISTE.	24
2.3.1 EXTRACTION DE SEGMENTS PAR TRANSFORMEE DE HOUGH PROBABILISTE :	25
2.3.2 DETECTION DES SEGMENTS ET CREATION DE SOUS-SEGMENTS	25
2.3.3 CONNEXION ENTRE SOUS-SEGMENTS VALIDES	26
2.3.4 CONVERSION AU FORMAT VECTORIEL (SHP) :	26
2.4 SECONDE METHODE : VECTORISATION PAR « SUIVI DE CHEMIN »	26
2.5 ETAPES FINALES COMMUNES AUX DEUX METHODES	28
2.5.1 NETTOYAGE TOPOLOGIQUE DES VECTEURS	28
2.5.2 TRANSFORMATION DES LIGNES EN OBJETS SURFACIQUES	28
3 APPLICATION DE LA METHODOLOGIE SUR LES DONNEES ET SES RESULTATS	30
3.1 PRESENTATION DES DONNEES – ANALYSE DES DIFFICULTES	30
3.1.1 CAS DES PLANCHES LES PLUS ANCIENNES (1813)	31
3.1.2 CAS DES PLANCHES INTERMEDIAIRES (1850)	32
3.1.3 CAS DES PLANCHES LES PLUS RECENTES (1972)	33
3.1.4 DIFFERENCE DE QUALITE RADIOMETRIQUE DES IMAGES	34
3.2 APPLICATION DE LA METHODOLOGIE SUR LES DONNEES	35
3.2.1 LE PRETRAITEMENT	35

3.2.2	PREMIERE METHODE : EXTRACTION DES SEGMENTS PAR TRANSFORMEE DE HOUGH PROBABILISTE.	40
3.2.3	SECONDE METHODE : VECTORISATION PAR « SUIVI DE CHEMIN »	45
3.2.4	ETAPES FINALES COMMUNES AUX DEUX METHODES	46
3.3	VALIDATION ET RESULTATS	50
3.3.1	DANS LE CAS DES LINAIRES	50
3.3.2	DANS LE CAS DES POLYGONES	51
3.3.3	CORRECTION MANUELLE DES VECTEURS	52
CONCLUSION		53
PERSPECTIVES		54
BIBLIOGRAPHIE		55
TABLES DES FIGURES		58
LISTE DES TABLEAUX		59

Introduction

Les cartes anciennes sont une part considérable de l'héritage cartographique et sont une source d'informations importante pour différents types de recherche. En effet, elles peuvent montrer l'évolution d'un territoire dans le temps et permettent plus particulièrement d'analyser et d'étudier les changements environnementaux ainsi que le développement de la démographie. Nous retrouvons cette problématique dans le cas des plans cadastraux. En effet, dans de nombreux départements, les services des archives diffusent leurs fonds documentaires numérisés en ligne, incluant entre autres les plans cadastraux anciens. Nous avons alors à disposition des plans cadastraux en format raster datant du début du 19ème siècle jusqu'à nos jours, cela recouvre une période de 200 ans pendant laquelle des plans ont été mis à jour à des instants « t ». Ainsi, ces plans constituent une source d'information précieuse pour le territoire et retracent l'évolution des modes de vie et ce à différentes échelles, de la commune à la section cadastrale.

Le format raster n'étant pas adapté aux analyses que nous souhaitons mener, il est indispensable de les obtenir au format vecteur. En effet, d'une part, la donnée vectorielle permet aux historiens de détecter et analyser plus facilement les changements spatiaux et temporels d'un territoire. D'autre part, elle permet, de rendre plus facile la combinaison et l'analyse entre les données historiques et modernes dans le but de visualiser et étudier plus facilement leurs différences. Une vectorisation manuelle serait trop coûteuse et chronophage, il faudrait donc envisager une méthodologie permettant de convertir automatiquement des données rasters en données vecteurs.

Pour ce faire, nous nous placerons dans le contexte de la reconnaissance puis de la vectorisation d'objets surfaciques « parcelles » à partir de planches cadastrales scannées anciennes. La vectorisation semi-automatique sera mise en œuvre selon deux approches : la première méthode basée sur l'extraction des segments par la transformée de Hough probabiliste et la seconde basée sur la vectorisation par « suivi de chemin » à l'aide d'utilitaires du logiciel GRASS. Néanmoins, ces méthodes montrent leurs limites, en effet, selon l'ancienneté des plans, ces derniers contiennent des informations de natures très variées et sont de qualité inégale. Aussi, la vectorisation est plus difficile pour les planches anciennes au vu de leurs dégradations causées par le temps. Les données issues de la vectorisation dépendent donc du type et de la qualité de la donnée en entrée.

Ainsi, avant de commencer la vectorisation, un prétraitement doit être appliqué aux images, notamment dans le cas des images les plus anciennes. Ce prétraitement a pour objectifs de faire ressortir les contours et de nettoyer l'image du bruit afin de faciliter la vectorisation.

L'objectif de ce travail de fin d'études (TFE) est de mettre en place un outil sous la forme d'une chaîne méthodologique, reproductible et semi-automatique de vectorisation de planches cadastrales scannées anciennes, à l'aide de scripts Python et de logiciels Open Source. Les données produites pourront être ensuite comparées à des références vectorisées manuellement. Cette chaîne s'inscrit dans un projet plus global dont le résultat final sera une base de données (BDD) vectorielles multi-dates et géoréférencée. Cette BDD, permettra d'analyser des phénomènes anthropiques (remembrement, opération d'aménagement) ou naturels (évolution d'un cours d'eau) ayant affectés un territoire. Dans un premier temps, ce travail s'appliquera à un échantillon de deux communes, Vaas et Aubigné-Racan, situées au sud du département de la Sarthe (72).

Notre problématique est la suivante : quelle est la meilleure méthode pour vectoriser semi-automatiquement des plans scannés du cadastre ancien.

Dans une première partie, nous nous intéresserons à l'établissement des plans cadastraux de 1813 jusqu'à nos jours, cette étude historique permet d'expliquer les résultats obtenus. Ensuite, nous exposerons dans un état de l'art la problématique de la vectorisation de données cartographiques. Enfin, nous présenterons le contexte général du projet du laboratoire Géomatique et Foncier (GeF) du CNAM en partenariat avec l'IRSTV (institut de recherche en science et techniques de la ville) « Le cadastre comme outil d'analyse de l'évolution des territoires », dans lequel s'insère ce TFE.

Ensuite, dans une deuxième partie, les étapes de la méthodologie de vectorisation semi-automatique réalisée dans ce TFE seront détaillées. Les étapes de prétraitement ainsi que deux approches mises en œuvre seront développées, les méthodes et les principes de fonctionnement des outils utilisés et créés seront expliqués.

Enfin, dans une troisième et dernière partie, les données cartographiques cadastrales mobilisées dans le projet seront présentées. Une étude a été effectuée pour les différentes périodes de conception ou mise à jour du plan. Ceci permettra d'identifier les facteurs susceptibles d'affecter la qualité de la vectorisation. Enfin, l'application et les résultats des deux méthodes seront présentés et analysés, et ce, pour chaque date.

1 État de l'art

L'objectif de cette partie est, dans un premier temps, de comprendre les méthodes d'établissement des plans cadastraux depuis 1813. Dans un deuxième temps, de faire un état de l'art des travaux de vectorisation et de mosaïquage de données cartographiques scannées anciennes. Et dans un dernier temps, de présenter la chaîne méthodologique globale conçue par l'équipe de Géodésie et Géomatique (L2G) du laboratoire Géomatique et Foncier (GeF) et d'en distinguer les différentes étapes.

1.1 L'établissement des plans cadastraux depuis 1813

1.1.1 Les plans du cadastre napoléonien

En 1791, l'Assemblée Constituante décide de supprimer les anciens impôts pour les remplacer par « une contribution foncière unique établie sur toutes les propriétés foncières à raison de leur revenu net » (Rodallec 2015). Cet impôt permettra de couvrir les dépenses de l'Etat. Ainsi, « la confection d'un cadastre général permettant de déterminer sur tout le territoire la contenance et le revenu de ces propriétés s'avérait nécessaire » (Rodallec, 2015). En aout 1791, il est permis aux communes de « lever à leurs frais » un cadastre afin de répartir l'impôt le plus justement possible. Afin de normaliser les méthodes d'établissement du plan cadastral, des règles à suivre sont énumérées par le décret du 16-23 septembre 1791.

« La loi du 15 septembre 1807, suivie du règlement impérial du 27 janvier 1808 décida la mise en œuvre d'un cadastre parcellaire général »¹, entièrement financé par l'Etat. Les travaux débutent en 1808 et ne s'achèvent qu'à la fin des années 1840. Ce cadastre est connu sous le nom de Cadastre Napoléonien ou d'Ancien Cadastre et constitue la base du Cadastre français contemporain.

Parmi les trois méthodes de réalisation du cadastre parcellaire qui ont été proposées, la solution retenue est celle de Hautier, géomètre en chef et ancien professeur de cours de géométrie pratique du cadastre. Sa solution consiste « à faire la triangulation calculée de la commune, à lever le périmètre de cette commune, à lever en outre les chemins et ruisseaux, à calculer les contenances des masses ou polygones formés par ces divisions et subdivisions et à mesurer ensuite les parcelles de propriétés contenues dans ces grandes masses. Ce mode permettrait de tirer parti des divers degrés de talents des arpenteurs en ce qu'il chargerait les plus simples arpenteurs de lever sous la direction des premiers le détail des propriétés » (Clergeot, 2009).

Une commission réunie sous l'impulsion de Martin Gaudin, ministre des finances en 1808 et présidée par le mathématicien Jean-Baptiste Delambre met au point un code cadastral publié en 1811, il s'agit du Recueil Méthodique des lois, décret, règlements, instructions et décisions sur le cadastre de la France. On y apprend par exemple au Titre IV – « Arpentage : les procédures de confection du plan », qu'un plan parcellaire est divisé en feuilles et que chaque feuille doit, autant qu'il est possible, contenir une section². Une section est d'après le paragraphe 20 de la documentation officielle du cadastre, CAD – Descriptif, usage et diffusion de la documentation cadastrale « [...] une partie du territoire communal

¹ Paragraphe 1.2 Généralités sur le plan cadastral du descriptif du contenu de la BD PARCELLAIRE® V1.2, IGN, juin 2012.

² Article 213.

déterminée dans le but de faciliter l'établissement, la consultation et la tenue des documents cadastraux. Son périmètre est constitué dans la mesure du possible par les limites présentant un caractère suffisant de stabilité (ex : route, chemins, cours d'eau...) ». Néanmoins, si une section ne peut pas tenir sur une feuille, elle peut être divisée en deux feuilles³ et au contraire, deux sections peuvent être sur la même feuille⁴. Une section est identifiée par une lettre simple : A, B, C, etc. Une fois le plan parcellaire terminé, l'ensemble de la commune est représentée sur une feuille de papier grand aigle (75 × 106 cm) : il s'agit du tableau d'assemblage. Son échelle peut s'étendre du 1/5000^{ème} au 1/20000^{ème} selon l'étendue de la commune. Ces documents graphiques s'accompagnent d'une documentation littéraire telle qu'un état de section (tableau répertoriant les différentes propriétés représentées sur une section) et une matrice cadastrale (registre récapitulatif renseignant pour chaque propriétaire la liste de ses biens et leurs valeurs)⁵.

En 1827, le règlement du 25 Mars introduit des améliorations qui vont permettre une meilleure qualité géométrique du plan cadastral. Pour commencer, les travaux de triangulation sont réalisés par un spécialiste, le « triangulateur » alors qu'auparavant, ils étaient réalisés par un géomètre de première classe moins qualifié. En 1829, le triangulateur doit, pour chaque commune à cadastrer, créer un étalon qui permettra aux géomètres de vérifier leur chaîne avant de commencer leur lever (Clergeot, 2009). En 1846, l'arrivée du théodolite permet d'effectuer des mesures de meilleure qualité et plus rapidement.

1.1.2 La rénovation et la conservation du plan cadastral

Suite à la pression des communes et l'augmentation des contentieux liée à l'absence de régularisation du plan, il est nécessaire de tenir les plans à jour et d'organiser la conservation du cadastre, alors que le Cadastre Napoléonien n'est pas encore achevé. Ainsi, une loi du 7 août 1850 autorise les communes cadastrées depuis au moins 30 ans, à leurs frais, à renouveler leur cadastre (Rodallec, 2015). Il s'agit d'un lever parcellaire nouveau, ces plans ont une codification différente de ceux antérieurs à 1840, la numérotation des parcelles a été refondue et de nouvelles catégories d'éléments bâtis sont représentées. C'est le cas des bâtiments légers qui sont représentés par un rectangle avec les deux diagonales, comme on peut le voir sur la Figure 1-1.

Devant l'urgence du renouvellement du cadastre dans certaines communes, intervint la loi du 17 mars 1898. Cette loi permettait aux communes cadastrées depuis au moins 30 ans d'obtenir le renouvellement et la révision de leur cadastre, ce à condition d'en assurer la conservation (Rodallec, 2015). La révision ou mise à jour consiste à actualiser le plan napoléonien à l'aide de mesures s'appuyant sur des éléments parcellaires anciens inchangés. Cette méthode était appliquée à condition que la densité de limites anciennes conservées et que la valeur métrique de l'ancien plan soient suffisantes (Clergeot, 2009). Cette loi permet également la création d'un service technique du cadastre chargé de ces opérations, dont le service ne dura que 9 ans (Clergeot, 2009). Néanmoins, le directeur de ce service, M Charles Lallemand, conduit à une amélioration de la qualité géométrique des

³ Article 214.

⁴ Article 216.

⁵ Définitions issues du site « Histoire du cadastre français ». ([Http://cadastre.pagesperso-orange.fr/napo2.htm](http://cadastre.pagesperso-orange.fr/napo2.htm)) consulté le 26/05/2017

plans. En effet, il est à l'origine d'une instruction portant sur « des méthodes, des procédés de calculs et des instructions de détails » (Clergeot, 2009).



Figure 1-1 : : Extraits de plans cadastraux napoléoniens de 1813 (à gauche) et 1850 (à droite) sur la commune d'Aubigné-Racan, lieu-dit de la Motte de Cœur (source : Archives Départementales de la Sarthe).

C'est en 1930 qu'une réforme complète voit le jour : il s'agit de la loi du 16 avril 1930 qui pose le principe d'une rénovation générale de l'ancien cadastre. Elle doit s'accompagner de la conservation annuelle des plans rénovés (Rodallec, 2015). En 1941, la loi du 17 décembre est à l'origine de la création d'un service unique du cadastre. Elle rend les modalités d'exécution de la réfection identique dans toute la France, à partir de cette date, la réfection remplacera l'opération de renouvellement.

Dans le cadre de la Réforme de la publicité foncière de 1955, le décret du 30 avril 1955 définit le cadre légal de la conservation et de la rénovation du cadastre. Dans le but qu'il assure l'identification et la détermination physique correcte des immeubles. Actualisation possible

Les modes de rénovation du cadastre sont :

- La rénovation par voie de « **mise à jour** » ou « **révision** », chaque fois qu'il était possible de procéder à une simple actualisation du plan Napoléonien. Cette méthode était appliquée à condition que la densité de limites anciennes conservées et que la valeur métrique de l'ancien plan soient suffisantes. Les sections révisées sont identifiées par lettre simple, A, B, C, etc.
- La rénovation par voie de « **réfection** ». La réfection consiste à effectuer un lever parcellaire nouveau appuyé sur une triangulation cadastrale, accompagné d'une délimitation des propriétés publiques et privées sous l'égide d'une commission communale de délimitation créée à cet effet. La réfection a lieu lorsqu'une actualisation du plan Napoléonien n'est pas possible. Les sections sont identifiées par deux lettres, AA, AB, AC, etc.

La révision a généralement été appliquée aux zones non bâties et la réfection aux communes urbaines (Clergeot P, 2009). Cette rénovation fut réalisée entre 1931 et 1974.

Avec l'extension des zones urbaines, certains plans ont besoin d'une autre « rénovation », il s'agit du remaniement, il est créé par la loi du 18 juillet 1974. L'article 6 de cette même loi le défini : « le remaniement recouvre toutes les opérations ayant pour objet

d'assurer une nouvelle rénovation du cadastre, lorsque le plan cadastral antérieurement rénové présente des insuffisances qui ne permettent plus d'en assurer la conservation annuelle de manière satisfaisante »⁶.

La dernière opération qui marque la modification des plans cadastraux par voie de réfection est l'Aménagement Foncier Agricole et Forestier (AFAF), ancien remembrement. Le principe de l'AFAF « est de mettre fin à la dispersion des parcelles d'un exploitant en regroupant ses terres, si possible aux abords du siège de son exploitation, afin d'améliorer les conditions de culture » (Rodallec, 2015). L'attribution des nouvelles parcelles doit respecter le principe d'équivalence « c'est-à-dire fournir à chaque propriétaire une surface équivalente en valeur de productivité réelle à celle des terrains apportés [...] » (Macario M, 2008). De plus, cette ré-attribution de parcelles implique la confection de nouveaux plans cadastraux et une mise à jour des matrices cadastrales. Cette procédure est dirigée par le conseil départemental et les plans sont établis par un géomètre privé, (Rodallec, 2015). Le plan est ensuite « labellisé » par le service du cadastre. Les sections cadastrales remembrées sont identifiées par deux lettres selon l'ordre suivant, ZA, ZB, ZC, ..., YA, YB, etc. Le champ d'application de ces opérations, initialement limité à la réorganisation des terres agricoles, a été élargi, par le décret du 10 avril 1963, à la construction d'autoroutes et de grands ouvrages publics (ordonnance du 3 octobre 1958 confirmée par la loi du 8 août 1962 et complétée par les lois du 30 décembre 1967 et du 4 juillet 1980) (Macario, 2008).

La Figure 1-2 récapitule les différentes opérations cadastrales de 1807 à nos jours.

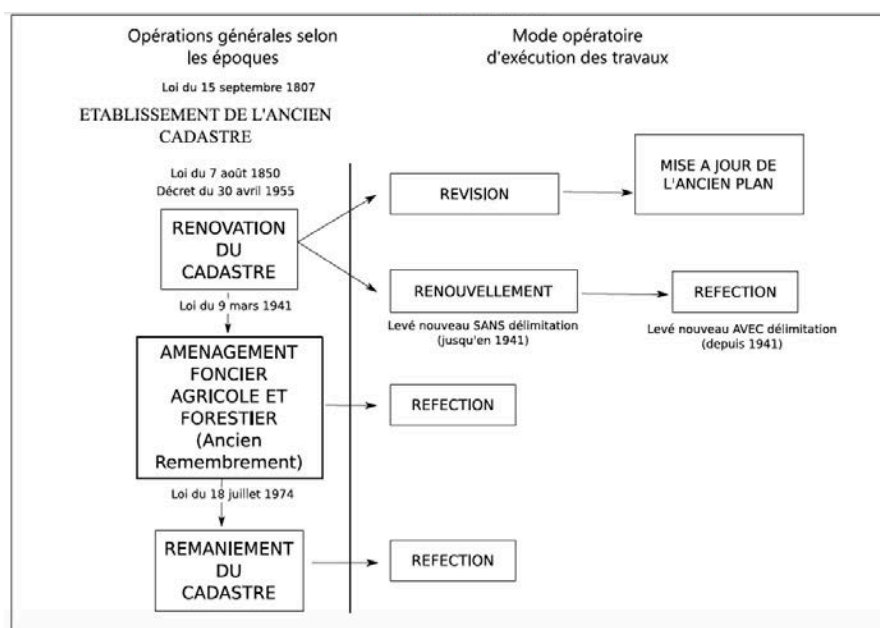


Figure 1-2 : Terminologie concernant les opérations cadastrales de rénovation des plans (source : Rodallec D, 2016).

⁶ Article 6 de la loi n°74-645 du 18 juillet 1974 sur la mise à jour périodique de valeurs locatives servant de base aux impositions directes locales, page 07539 du JO du 19/04/1974, consulté le 26/05/17

1.2 Les problématiques de la vectorisation des données cartographiques

1.2.1 Numérisation et vectorisation du cadastre en France

Nous nous sommes tout d'abord intéressés au cas du cadastre. En effet, les plans cadastraux actuels appelés Plans Cadastre Informatisés (PCI) sont dématérialisés. Nous nous sommes alors questionnés sur la façon dont cette donnée était obtenue.

Sous l'élan des collectivités locales, la vectorisation des plans cadastraux a débuté dans les années 1990. En 1993, le protocole d'accord national sur la numérisation du plan cadastral formule le cadre légal de l'informatisation du plan cadastral. La numérisation est le scan d'une planche alors que la vectorisation d'une image numérique permet d'obtenir des objets géométriques et des primitives géométriques (segments de droite, arcs de cercle, polygones, etc).

La vectorisation se faisait manuellement à l'aide de table à digitaliser. Cette démarche avait pour but de :

- Moderniser et de rationaliser les méthodes d'établissement, de diffusion et de gestion de la documentation ;
- Mieux satisfaire les besoins divers des usagers.

Malgré des attentes fortes de la part des collectivités locales, cette procédure s'est avérée très lente. Ainsi, l'utilisation du scanner a été privilégiée et de 2000 à 2003, 460 000 feuilles cadastrales ont été scannées en format A0. Et à la fin de l'année 2004, l'ensemble des plans cadastraux sont dématérialisés. D'après Rodallec (2015), la dématérialisation du plan cadastral comprend deux volets :

- La vectorisation des plans neufs produits à partir des opérations de remaniements et de remembrement, celle des plans obtenus par numérisation dans le cadre de la politique conventionnelle⁷ ou les opérations de vectorisation complémentaire pour les portions de territoire communal⁸ ;
- Le scannage des plans cadastraux qui ne rentrent pas dans les critères précédents ;

Il existe donc deux formats distincts pour gérer le plan cadastral informatisé :

- PCI Vecteur pour les plans numérisés au format vectoriel ;
- PCI Image pour ceux scannés dans un format « image ».

Actuellement, 92% des plans cadastraux sont au format PCI Vecteur. L'objectif est que la totalité des plans le soit en 2020.

Pour passer du plan papier au format PCI Vecteur, plusieurs opérations doivent s'enchaîner. Cela nécessite des allers-retours entre le Centre des Impôts Fonciers (CDIF), le prestataire (labélisé par convention) chargé de la vectorisation et le Service de la Documentation Nationale du Cadastre (SDNC). Il y a plusieurs CDIF par département, notamment 3 en Sarthe, pour un total de 311 en France et ce sont les seuls à avoir le droit de modifier les données⁹. Le CDIF veille à la conformité des fichiers informatiques avec les plans papiers

⁷ La politique conventionnelle est une convention bilatérale entre, d'une part, la collectivité locale désireuse d'un cadastre numérique, et la DGFIP, seule détentrice légale du plan cadastral concerné.

⁸ Il s'agit de portions importantes du territoire communal qui ne donnent pas lieu à constitution de parcelles, ces terrains sont exemptés de la taxe foncière des propriétés non bâties (Rodallec, 2015)

⁹ Entretien du 18 Mai 2017 dans les locaux du SDNC avec Gilles Berteau, administrateur des finances publiques adjoint.

et est aussi le lien entre tous les acteurs, alors que le SDNC quant à lui contrôle la structuration des fichiers. En effet, il vérifie que la structure du fichier utilisée par le prestataire est conforme à la norme EDIGEO. En effet, comme indiqué dans les paragraphes 50 et 60 de la documentation officielle du cadastre, CAD – Informatisation du plan cadastral : « Deux standards d'échange sont agréés par la DGFIP pour réaliser les travaux de vectorisation du plan cadastral : le standard d'échanges fondé sur la norme EDIGÉO ; le standard d'échanges fondé sur le format DXF-PCI. Le format EDIGÉO étant plus riche, il lui sera donné la préférence, pour toute nouvelle convention, [...] ».

A l'heure actuelle, le SDNC a de moins en moins de contrôles à effectuer puisque le volume de plans à vectoriser diminue. Ainsi, les prestataires sont moins sollicités et la vectorisation est effectuée par le SDNC.

1.2.2 Etat de l'art sur la vectorisation de cartes

Il s'agit d'un état de l'art non exhaustif sur la vectorisation de cartes anciennes.

Lorsque dans les années 1990, le protocole d'accord national sur la numérisation du plan cadastral est signé, l'équipe du laboratoire méthodes d'analyses pour le traitement d'images et la stéréorestoration (MATIS) située dans les locaux de l'Institut national de l'information géographique et forestière (IGN) a effectué quelques travaux en lien avec la vectorisation semi-automatique des plans numérisés. Il y avait notamment un projet « cartes scannées » pour lequel « une méthode d'extraction et d'identification de chaînes de caractères dans le contexte cartographique » a été développée par Pierrot-Deseilligny M *et al.* (1995) notamment. Le but de cette étude était de « convertir automatiquement le fonds cartographique existant, sous forme papier, en une base de données d'objets géographiques directement utilisable dans un Système d'Information Géographiques (SIG) ». L'IGN était le producteur des cartes analysées. De façon générale, une carte se différencie des autres documents graphiques par sa densité importante d'informations et sa légende très riches et par une normalisation des règles de rédaction. Néanmoins, la première spécificité induit des difficultés qui peuvent être réduites avec une utilisation des règles de rédaction. L'article précise également que les caractères ayant une forme simple, telles que le « I », « l », « j », etc. peuvent être facilement confondu avec des traits caractéristiques tels que les routes. La méthode utilisée dans la chaîne est la reconnaissance de formes. Une autre méthode pour détecter et extraire les chaînes de caractères d'une carte est basée sur le fait que les traits constituant les caractères sont en général des segments courts contrairement à ceux des routes (Cao, Tan 2002).

Par la suite, Viglino et Pierrot-Deseilligny (2003) ont travaillé sur un dispositif d'interprétation des plans cadastraux avec pour objectif la reconstitution complète des parcelles et des bâtiments afin de les incorporer dans un SIG. Ce travail a été réalisé sur les images numériques issues du protocole d'accord national sur la numérisation du plan. Le dispositif est basé sur trois étapes, la première est la vectorisation et la classification d'objets à « faible niveau » à savoir les hachures, les contours fermés et les lignes. Dans un deuxième temps, ces objets sont interprétés, dans un ordre particulier, comme des éléments caractéristiques du plan. Tels que les bâtiments (qui sont représentés par les contours fermés hachurés), les hangars (qui sont eux représentés par des contours fermés avec deux diagonales), les flèches qui relient les parcelles à leurs numéros. La dernière étape consiste à créer la structure SIG des parcelles.

Baily (2007) propose une méthodologie complète pour la vectorisation des cartes topographiques historiques de Grande-Bretagne. Il commence d'abord par géoréférencer les images numériques correspondant aux cartes topographiques, et ensuite, pour l'extraction des données vecteurs, la méthode consiste en une approche semi-automatique. Il s'agit d'une classification supervisée suivie d'un nettoyage effectué manuellement puis d'une conversion en format vecteur. Les résultats de cette méthode ont été comparés à une donnée de référence vectorisée manuellement. On constate donc que cette méthode semi-automatique offre de bons résultats, avec un pourcentage de correspondance entre les régions allant de 69 à 97%.

D'autre part, Iosifescu *et al.* (2016) proposent une méthodologie complète pour la vectorisation de cartes topographiques Suisses du 19^{ème} et 20^{ème} siècle. Ici aussi, la première étape consiste à géoréférencer les images numériques correspondant aux cartes. Mais la particularité est qu'ensuite un prétraitement est appliqué à l'image. Cette étape comprend un traitement des couleurs qui permettra de mettre en évidence des caractéristiques spécifiques de l'image. Ensuite, l'image est convertie en image binaire, c'est à dire, une image pour laquelle chaque pixel ne peut avoir pour valeur que 0 ou 1, en effet, pour la vectorisation, il est important d'avoir des images binaires contenant les objets à extraire, ici, il s'agit des bâtiments, de la rivière et les contours. Le prétraitement se termine pas un nettoyage de l'image binaire grâce à des opérateurs morphologiques. L'étape suivante est la vectorisation automatique, elle est faite grâce à des outils du SIG, il s'agit d'une conversion des points de contours en données vectorielles. Les lignes sont vectorisées grâce à l'utilitaire `gdal_contour` de la bibliothèque GDAL¹⁰ alors que pour les polygones, l'utilitaire `gdal_polygonize` est utilisé. Et la dernière étape consiste à nettoyer automatiquement les vecteurs obtenus, cela peut être fait avec un logiciel de SIG, notamment le logiciel libre QGIS. Néanmoins, les caractéristiques de ces cartes anciennes diffèrent de celles du cadastre. Effectivement, il s'agit de cartes topographiques couleurs et elles contiennent plus d'informations que les planches cadastrales anciennes.

Aussi, les travaux concernant l'extraction de routes se rapprochent plus des particularités des données cadastrales. Chiang et Knoblock (2010) et Chiang et Knoblock (2013) présentent une technique complète pour extraire puis pour vectoriser le réseau routier d'images scannées. Les trois grandes étapes de cette approche sont l'extraction des routes, la détection de leurs intersections et la vectorisation. La première difficulté est l'extraction des pixels « route ». En effet, différentes couleurs représentent différentes couches (les routes, les lignes de contours, les écritures). En identifiant les couleurs représentant les routes, il est possible d'extraire les pixels « route » de la carte grâce à une classification. Ensuite, les intersections des routes sont détectées, elles serviront de point de liaison lors du géoréférencement. Puis, des opérateurs morphologiques sont appliqués à l'image, notamment une squelettisation. Puis, la génération des vecteurs se fait à partir des axes des routes, ils sont obtenus en appliquant la transformée de Hough à l'image numérique.

¹⁰ Geospatial Data Abstraction Library, bibliothèque logicielle spécialisée dans les données vectorielles et rasters

1.3 Le projet du laboratoire Géomatique et Foncier du CNAM

Ce TFE s'inscrit dans le projet « Le cadastre comme outil d'analyse de l'évolution des territoires » du laboratoire Géomatique et Foncier (GeF) du CNAM. Il est en partenariat avec l'IRSTV (Institut de recherche en sciences et techniques de la Ville), une fédération de recherche dont fait partie le GeF et dont l'un des centres d'intérêt est la constitution de bases de données géographiques historiques en lien avec la ville. Ce travail doit aboutir à la création d'une base de données vectorielles multi-dates et géoréférencées. De plus, le volume des données à traiter (taille des images) et la consommation de ressources de certains des algorithmes utilisés ont nécessité l'utilisation du serveur de calcul du GeF.

Des travaux réalisés à l'ESGT en 2015 et 2016 ont permis de mettre en œuvre l'étape du géoréférencement des données vectorielles anciennes en recherchant la transformation la plus appropriée (Labrousse *et al.* (2015) et Fahrasmane (2016)). Cependant ce travail doit être complété afin d'aboutir à une chaîne de traitement complète, reproductible, la plus automatisée possible et utilisant des logiciels open source. Celle-ci s'étend depuis les planches scannées jusqu'à la base de données vectorielles multi-date, support d'études sur l'évolution du parcellaire cadastral de 1813 à aujourd'hui. Ce projet a pour finalité l'analyse de l'évolution du territoire à une échelle fine (par exemple l'impact d'un remembrement ou la variation de l'emprise d'un cours d'eau) observée par le prisme des changements du parcellaire cadastral.

1.3.1 La chaîne méthodologique élaborée par l'équipe L2G du laboratoire GeF

Pour obtenir la base de données vectorielles multi-dates à partir des données cadastrales anciennes, l'équipe de Géodésie et Géomatique (L2G) du laboratoire GeF (Géomatique et Foncier) a mis en place une chaîne de traitement semi-automatique. Le critère semi-automatique permet un gain de temps important dans le traitement des étapes.

La Figure 1-3 illustre le principe d'enchaînement de la méthodologie globale.

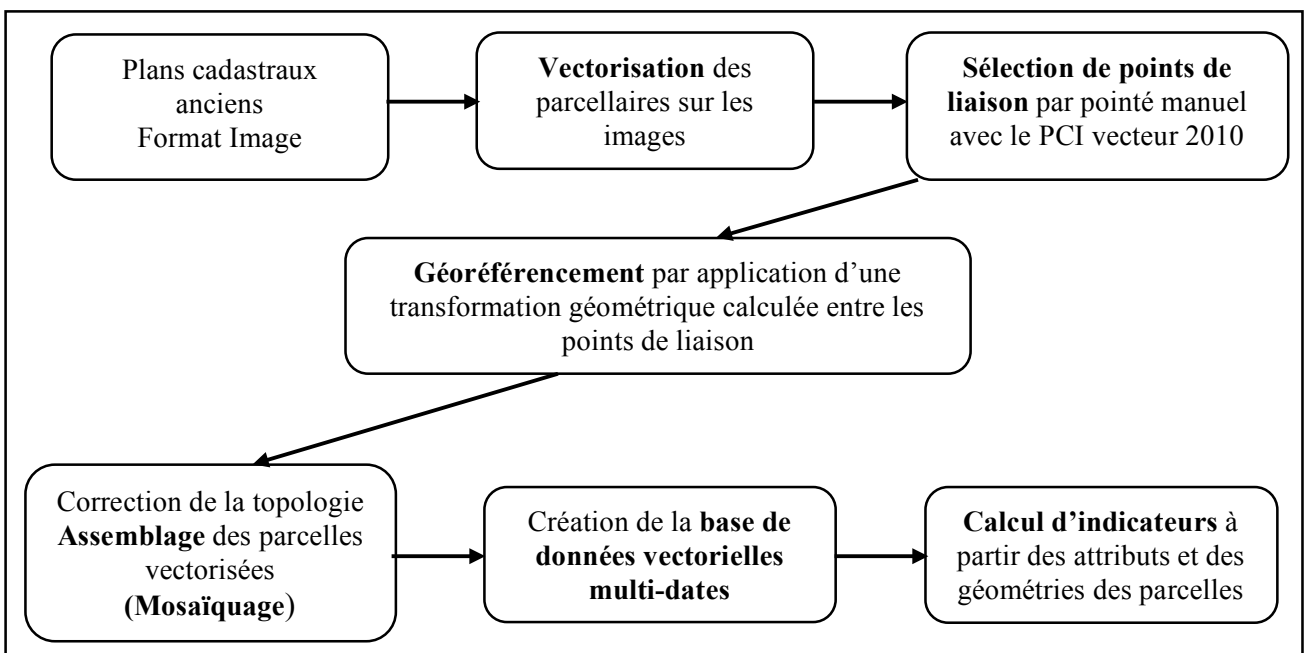


Figure 1-3 : Principe d'enchaînement des étapes de la chaîne méthodologique du laboratoire du GeF, source : d'après Fahrasmane (2016).

Les données en entrée sont des plans cadastraux scannés anciens, ces plans sont conservés aux archives départementales.

La première étape est la vectorisation semi-automatique des images correspondant aux planches cadastrales anciennes, qui sera largement abordée dans ce TFE. Cette étape est préliminaire au géoréférencement, en effet, les parcelles représentées dans les images n'ont subi aucune déformation liée au géoréférencement.

La deuxième étape est le géoréférencement, traité par Fahrasmane (2016) « il faut transformer les coordonnées des points de la données vectorielle ancienne selon le modèle géométrique estimé entre les points de liaison [...] l'objectif est de pouvoir superposer des parcellaires vectorisés avec une qualité géométrique et visuelle optimale, au parcellaire cadastral du PCI-vecteur de 2010. La transformation géométrique optimale est choisie parmi sept transformations ». Néanmoins, suite à cette étape, des défauts de chevauchement ou des trous peuvent apparaître, ainsi, il faut s'assurer de la cohérence de l'assemblage. Le mosaïquage par traitement topologique permet de les solutionner.

La troisième étape est le mosaïquage (assemblage) des feuilles cadastrales vectorisées et géoréférencées. En effet, une commune est composée de plusieurs feuilles cadastrales représentant une section, qu'il est nécessaire de mosaïquer afin d'obtenir une continuité dans le plan. Le mosaïquage des vecteurs pourra s'appuyer sur un traitement topologique des données. Cette étape permet d'obtenir un plan vectorisé continu qui se rapproche de la donnée PCI-vecteur actuelle.

La cinquième étape est la création de la base de données multi-dates qui permet de lier les données entre elles et constitue le support des analyses menées sur l'évolution du parcellaire cadastral.

1.3.2 Les étapes traitées de la chaîne méthodologique du GeF

Des travaux réalisés à l'ESGT en 2015 et 2016 ont permis de traiter certaines étapes de la chaîne méthodologique.

En 2015, dans le cadre des projet pré-professionnel (PPP), des élèves de l'ESGT, Labrousse et al (2015) avaient pour objectif d'étudier l'évolution du parcellaire cadastral dans deux communes rurales du sud de la Sarthe traversée par le Loir : Vaas et Aubigné-Racan. L'étude s'est basée sur les données cadastrales à proximité de cette rivière. Les données à disposition étaient des plans cadastraux scannés de trois époques (1813, 1850, 1974) et le PCI vecteur datant de 2012. Les étudiants devaient créer une base de données vectorielles multi-dates et géoréférencées.

Au cours de ce projet, les étudiants ont réalisé, dans l'ordre, le géoréférencement des feuilles cadastrales numérisées, leur vectorisation et une évaluation de la qualité des méthodes de géoréférencement, et la constitution d'une base de données multi-dates par appariement (sélection des points de liaison des différentes données) des parcelles ayant évolué au cours du temps.

Pour le géoréférencement, ils ont utilisé le logiciel Open Source QGIS et plus particulièrement le plugin « Géoréférencer GDAL » de la bibliothèque GDAL. La vectorisation a elle aussi été effectuée grâce aux outils disponibles sur QGIS, notamment le plugin « AutoTrace ». Ensuite, l'équipe a mis en place des calculs d'indicateurs, le premier

permettant d'étudier l'évolution des limites des parcelles. Le second portant sur l'évolution des parcelles elles-mêmes en tant qu'entité, en étudiant les deux types de mutation qu'elles peuvent subir : la fusion ou la division. Un dernier cas a été étudié, celui où la parcelle ne subit aucune mutation. Les calculs ont été réalisés à partir de requête en langage SQL dans l'environnement PostgreSQL/PostGIS.

Finalement, concernant le géoréférencement et la vectorisation, plusieurs étapes ont été effectuées manuellement, ce qui s'avère chronophage. De plus, la méthodologie élaborée pour le géoréférencement est pertinente mais jugée peu rigoureuse (Fahrasmane, 2016).

En 2016, le TFE de Maïté Fahrasmane avait pour objectif de sélectionner la meilleure fonction mathématique permettant de réaliser un géoréférencement de parcelles cadastrales anciens vectorisés. Ce travail s'est appliqué à la même zone d'étude que le PPP de 2015. Il a permis d'élaborer une méthode rigoureuse de géoréférencement et d'évaluation des transformations géométriques, qu'elles soient conventionnelles ou innovantes. Cela a été réalisé grâce à l'utilisation de programmes informatiques en langage Python qui ont permis d'automatiser les différents calculs et d'évaluer les transformations géométriques. Un total de sept transformations a été testé :

- Fonctions globales : fonctions polynomiales de degrés 1, 2 et 3, la transformée d'Helmert, fonction spline à plaques minces TPS (Thin Plate Spline),
- Fonctions locales : la triangulation de Delaunay,
- Une méthode innovante : le noyau gaussien

Ensuite, l'évaluation de la qualité des transformations a été faite sur les points de contrôle. La méthode de la validation croisée a été choisie, il s'agit d'une méthode d'estimation de fiabilité qui tient compte de tous les points d'appui et permet de les considérer successivement comme des points de contrôle. Cette redondance permet de calculer la qualité moyenne de la transformation sur tout l'échantillon de points. La redondance des calculs de modèles rend cette méthode plus robuste que celle utilisée dans les travaux du PPP de 2015 où l'erreur moyenne quadratique est calculée une seule fois entre les points d'appui géoréférencés (Fahrasmane, 2016). La réalisation des calculs et des évaluations des transformations ont été faite à l'aide de programme en langage Python.

1.3.3 Les réalisations de ce TFE – les étapes traitées de la chaîne méthodologique

Ce TFE, quant à lui, traite l'étape de la vectorisation.

La partie 2 traite de la mise en place d'une méthode de vectorisation semi-automatique des planches cadastrales scannées anciennes et plus particulièrement de la reconnaissance et de la vectorisation d'objets surfaciques « parcelles ». Ainsi, une procédure semi-automatique utilisant des logiciels open source a été mise en place. Une approche est basée sur l'extraction de segment par la transformée de Hough probabiliste alors que la seconde est basée sur le « suivi de chemin ». Des étapes seront communes aux deux méthodes. La partie 3 traite de l'application de ces méthodes sur trois planches scannées de trois époques différentes (1813, 1850 et 1972) qui présentent des caractéristiques différentes comme vu dans la partie 3.1. Les deux approches ont été testées puis validées à partir de planches références.

2 Chaîne méthodologique

L'objectif de cette partie est la reconnaissance puis la vectorisation d'objets surfaciques « parcelles » à partir de planches cadastrales scannées anciennes. Pour ce faire, nous proposons une procédure semi-automatique et se limitant à l'utilisation de logiciels open source. Le choix de privilégier les logiciels open source est motivé par l'envie de garder la main sur les codes, d'éviter les « boîtes noires » et il est lié à la maturité des logiciels et langages de programmation open source. De plus, nous souhaitons limiter le nombre d'outils utilisés pour éviter une méthodologie trop complexe. Enfin, certaines étapes de la procédure font appel à des algorithmes ou outils qui doivent être robustes aux changements de paramètres afin de s'adapter facilement aux caractéristiques des différentes données.

2.1 Méthodologies mises en place en fonction de l'ancienneté des planches

Afin de réaliser la vectorisation semi-automatique des images numériques obtenues par scannage des planches, deux approches vont être envisagées. Une première centrée sur la transformée de Hough probabiliste et une seconde centrée sur la routine d'extraction de lignes du logiciel GRASS. Les deux méthodes seront mises en œuvre dans ce TFE, ce qui conduira à une comparaison des résultats obtenus dans une partie suivante.

Compte tenu de la différence de qualité radiométrique importante des images selon leur année de réalisation, la chaîne de traitement appliquée pour les planches anciennes (figure 2-1) sera légèrement différente de celle appliquée pour les autres planches (figure 2-2), notamment la phase de prétraitement et celle d'extraction des lignes.

Pour les deux approches envisagées, une partie des étapes de la chaîne sera commune.

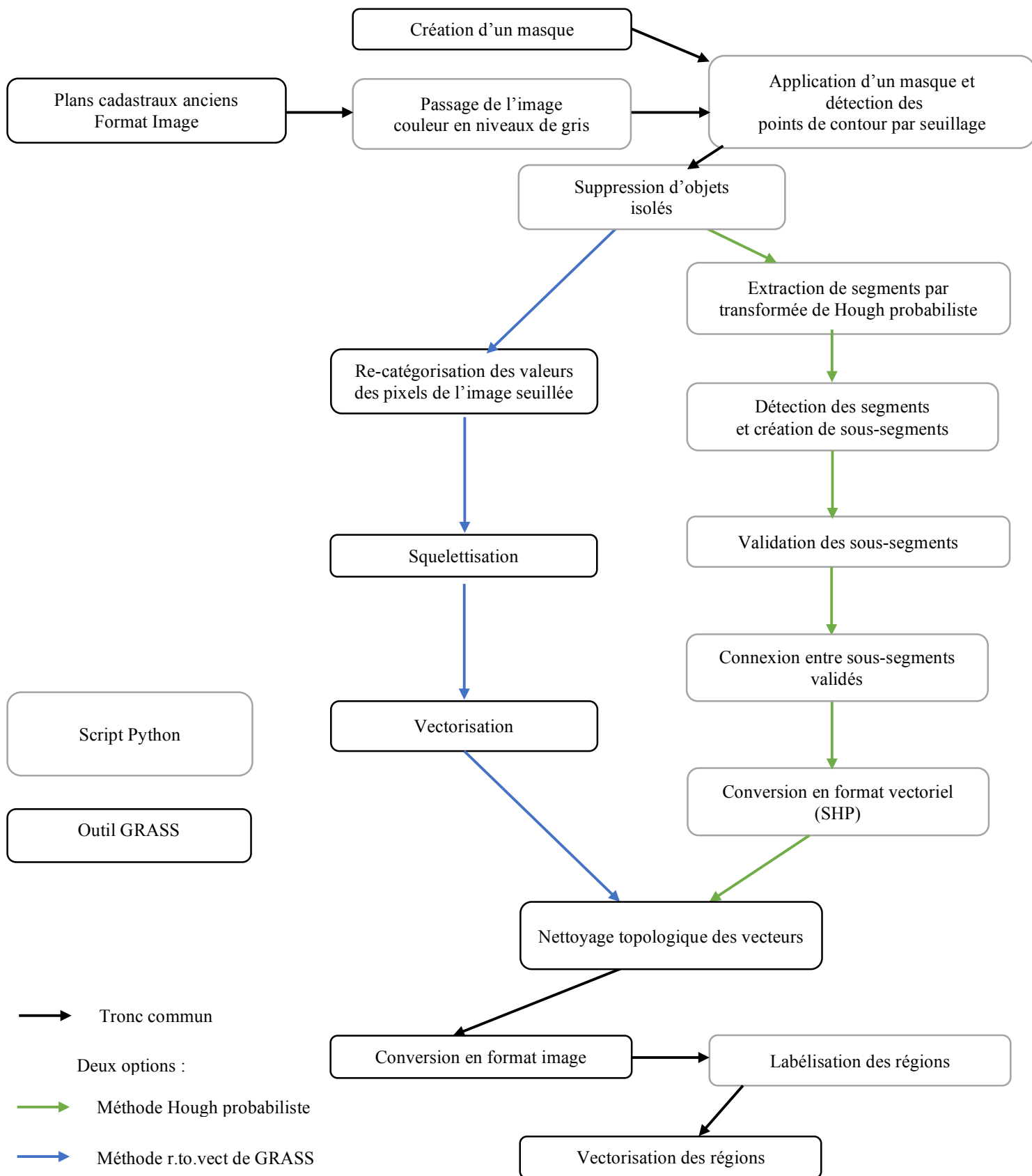


Figure 2-1 : Méthodologie pour les planches anciennes (1813 et 1850).

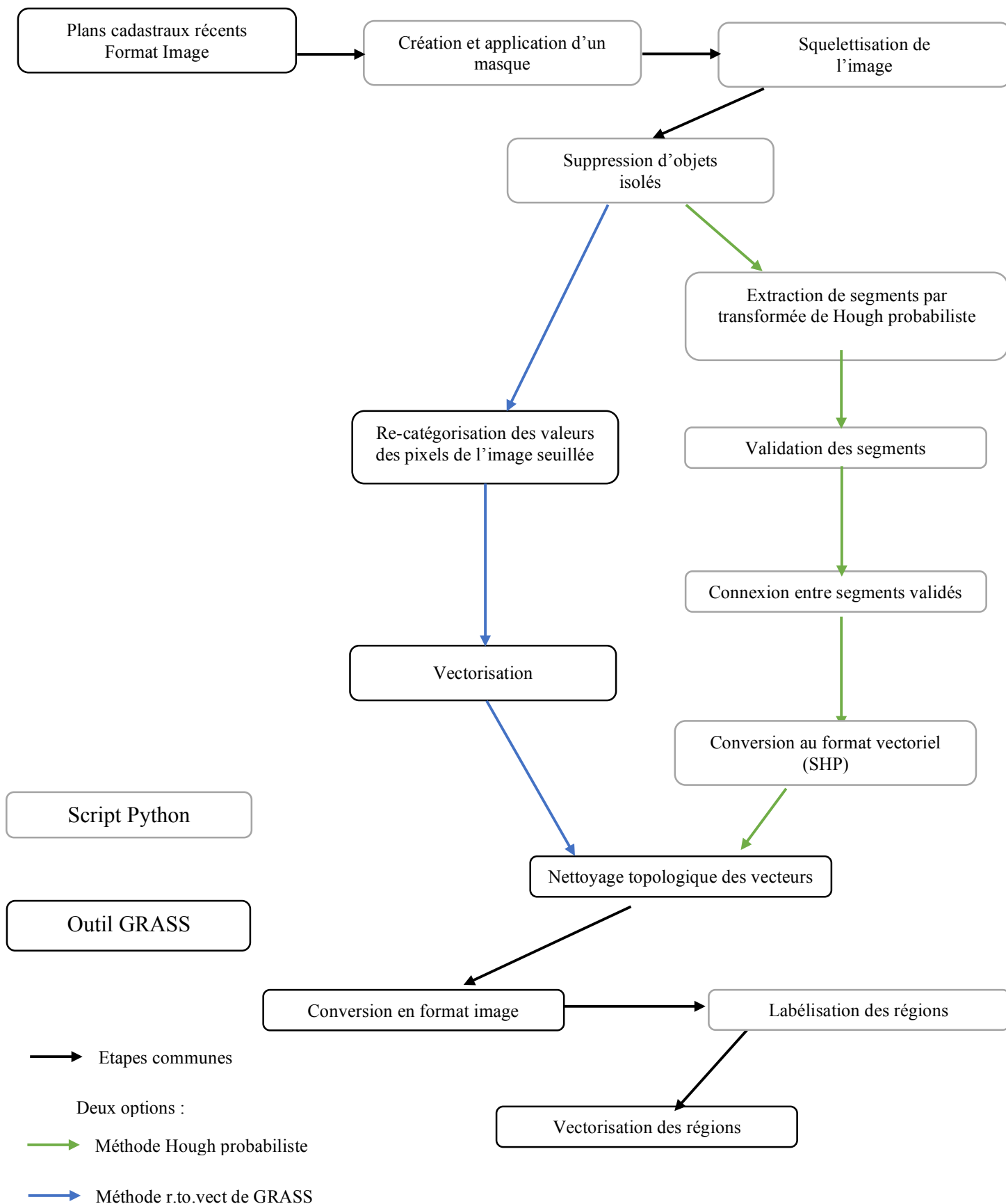


Figure 2-2 : Méthodologie pour les planches plus récentes (1972).

2.2 Etapes communes aux deux méthodes : le prétraitement

Il s'agit des traitements préliminaires à exécuter sur l'image, pour les deux traitements et pour les deux époques (1813 et 1850 - 1972).

2.2.1 Conversion en niveau de gris

Dans la méthodologie employée, la détection des points de contour par seuillage sera réalisée sur une image d'entrée en niveau de gris. Ainsi, si l'image initiale est en couleur, il faudra réaliser une conversion en noir et blanc. La conversion d'une image couleur en niveaux de gris consiste à remplacer, pour chaque pixel, les trois valeurs représentant les niveaux de rouge, vert et de bleu, en une seule valeur représentant la luminance. Ainsi, la première étape consiste donc à convertir l'image couleur en niveaux de gris.

La formule utilisée par la fonction *rgb2grey* (du module *color* de la bibliothèque de traitement des images scikit-image disponible sous Python) pour cette conversion est définie par la Commission Internationale de l'Eclairage (CIE) dans sa recommandation 709 :

$$\text{Niveau de gris} = 0,2125 \text{ Rouge} + 0,7154 \text{ Vert} + 0,0721 \text{ Bleu}$$

Cette formule rend compte de la manière dont l'œil humain perçoit les trois composantes rouge, vert et bleu, de la lumière. La somme des trois coefficients vaut un. Il y a une forte inégalité de la perception des apparences colorées.

2.2.2 Application d'un masque et détection des points de contour par seuillage

L'étape suivante consiste à appliquer un masque et à réaliser une détection de points de contour par seuillage. Ce masque a été appliqué à l'image convertie en niveau de gris et contourne les limites cadastrales extérieures. Cela permet de découper l'image, et à la fois de gagner du temps dans le traitement de la chaîne et de limiter les problèmes à la zone d'étude.

Ce masque a été réalisé manuellement à partir du logiciel ENVI, en effet, la mise en œuvre de cette étape depuis Python ou GRASS a posé des difficultés.

L'application de ce masque se fait lors du seuillage de l'image, qui aboutit à une image binaire, l'image ne comportera plus que deux valeurs, 0 et 1.

Une image peut être seuillée par exemple par seuillage simple, par seuillage par Hystérésis ou encore par seuillage adaptatif.

Le principe du seuillage simple est que si T est un seuil, et que la valeur du pixel au point (i, j) est supérieure à T , alors le point est dans la classe 1, sinon il est dans la classe 2.

Le principe du seuillage par Hystérésis est le suivant :

- 1- Choix de deux seuils : seuil haut et seuil bas.
- 2- Si la valeur du pixel est supérieure au seuil haut, c'est un point de contour sûr qu'on code par 1.
- 3- Si la valeur du pixel est inférieure au seuil bas, alors ce n'est pas un point de contour qu'on code par 0.
- 4- Si la valeur du pixel est entre les deux, il est considéré comme point de contour s'il est connecté à un pixel de valeur supérieure au seuil haut.

La méthode de la moyenne glissante (méthode de seuillage adaptatif) consiste à déterminer localement un seuil en chaque pixel à partir de la moyenne locale des niveaux de gris. La moyenne locale est calculée pour un voisinage de taille $n \times n$.

Le seuil en chaque pixel vaut : $T_{i,j} = c \times M_{i,j}$

Avec $M_{i,j}$ la moyenne locale au pixel (i,j) et c une constante à déterminer. Si la valeur du pixel au point (i,j) est supérieure à $T_{i,j}$, il prend la valeur 1, sinon, il prend la valeur 0.

L'inconvénient de cette méthode est que la détermination de la constante c se fait empiriquement.

De plus, un seuillage par Hystérésis locale a été imaginé, il s'agit d'une association entre le seuillage par moyennes glissantes et le seuillage par Hystérésis. Ceci permet d'éliminer des faux contours ou une partie du bruit. L'objectif est d'obtenir une meilleure continuité dans les lignes de contours.

La mise en œuvre consiste à déterminer des seuils bas et hauts localisés. Le seuil haut en chaque pixel vaut : $Seuil\ haut_{i,j} = h \times M_{i,j}$ et le seuil bas en chaque pixel vaut $Seuil\ bas_{i,j} = b \times M_{i,j}$ avec $M_{i,j}$ la moyenne locale au pixel (i,j) , h et b des constantes à déterminer. Cela permettra de garder localement les contours les plus marqués sur l'image mais en essayant d'assurer leur continuité.

Comme dans le cas de la moyenne glissante, la détermination des paramètres se fait cependant empiriquement.

2.2.3 Suppression d'objets isolés

Ensuite, l'étape suivante de la chaîne est la suppression d'objets isolés, notamment le bruit.

La méthode consiste, à partir d'une image binaire, à supprimer les « objets » constitués de points de contour connectés, dont la taille (nombre de pixels) est inférieure à une valeur spécifiée. La connectivité est de deux types, 8-connecte (8 voisins), 4-connecte (4 voisins).

Une connectivité de 4 signifie que pour que deux pixels soient connectés, ils doivent partager une face, alors que pour une connectivité de 8, un coin suffit.

Ceci est effectué grâce à la fonction (*remove_small_object*) du module *morphology* de la bibliothèque scikit-image. Pour cette fonction, il faut préciser la donnée en entrée :

- Une image qui doit être de type booléen ou entier.

Puis les paramètres :

- La taille maximale des régions de pixels connectés à supprimer,
- Le type de connectivité (4 ou 8),
Un critère pour décider de modifier directement l'image en entrée ou pour créer une nouvelle image.

2.2.4 Squelettisation

La deuxième étape consiste à squelettiser l'image, la squelettisation consiste à retirer au fur et à mesure les points du contour de la forme, tout en préservant ses caractéristiques topologiques.

En effet, cette étape facilite la suite du traitement dans le cas de la méthode par transformée de Hough et est obligatoire dans le cas de la méthode r.to.vect de GRASS. Ce traitement permet d'obtenir un résultat satisfaisant, la continuité des lignes est maintenue.

L'outil utilisé est *skeletonize* du module *morphology* de la bibliothèque scikit-image. L'algorithme est décrit par Zhang et Suen (1984). Pour cette fonction, l'image en entrée doit être binaire et les points de contour doivent être en blancs. La méthode utilisée dans l'algorithme consiste à considérer l'image comme une matrice où chaque pixel a pour valeur 0 ou 255. Puis à appliquer un processus itératif à chaque pixel de valeur 255 de cette matrice en tenant compte des valeurs des 8 pixels voisins, en effet, on considère un voisinage de 3x3 pixels. Le but de cette méthode est de supprimer les points de contour qui n'appartiennent pas au squelette et d'obtenir un squelette qui a une largeur de 1 pixel.

Pour ce faire, il faut diviser chaque itération en deux « sous-itérations ». Pour la première sous-itérations, le pixel considéré sera supprimé s'il répond à 4 conditions, comme par exemple :

- Le nombre de pixels voisins égal à 255 doit être compris entre 2 et 6,
- Le produit de 3 pixels voisins doit être égal à 0,

Cette première sous-itération permet de supprimer les pixels situés au sud-est et le pixel situé dans le coin au nord-ouest du voisinage d'un pixel du squelette, ils n'appartiennent donc pas au squelette final.

Pour la seconde sous-itération, il y a deux conditions qui concernent également la valeur du produit de 3 pixels voisins doit être égal à 0, mais il s'agit de pixels différents. Ces deux conditions permettent de supprimer les pixels situés au nord-ouest et le pixel situé dans le coin au sud-est du voisinage d'un pixel du squelette. De plus, grâce à ces conditions, les points de départ et d'arrivée du squelette sont préservés ainsi que la continuité du squelette. Les itérations continuent jusqu'à ce qu'il n'y ait plus de point à supprimer.

Une fois que le pré-traitement est effectué et que les points de contour sur les images sont déterminés, l'objectif suivant est l'extraction des polygones en format vectoriel Shapefile.

Pour cela, deux méthodes vont être envisagées, une première basée sur la transformée de Hough probabiliste, pour cela on utilise des fonctions issues des bibliothèques en langage Python. La seconde quant à elle centrée sur la routine d'extraction de lignes du logiciel de SIG, GRASS.

2.2.5 Opérateurs morphologiques

Des opérations morphologiques ont été appliquées à l'image seuillée de la planche de 1813 afin d'améliorer la détection des points de contour. Chiang YY *et al.* (2005) ainsi que Chiang YY, Knoblock C.A (2010) ont réalisé des traitements tels qu'une dilatation suivie d'une érosion (ce qui revient à effectuer une fermeture). Les traitements morphologiques utilisent un élément, l'élément structurant (ES), voir figure 2-3 qui est appliqué à tous les pixels de l'image un à un. L'opération est itérée jusqu'à stabilisation du résultat.

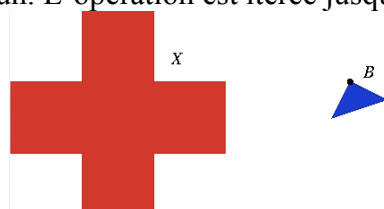


Figure 2-3 : A gauche, c'est une figure X, à savoir un ensemble de pixels et à droite, c'est l'élément structurant B, où la boule noire indique la position du point de référence.

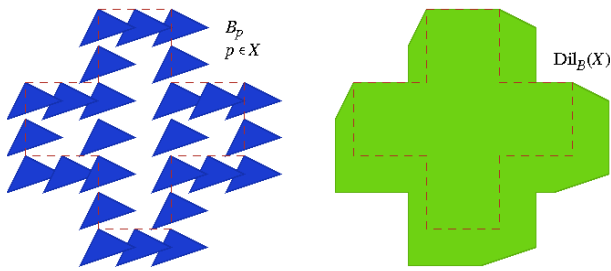


Figure 2-4 : A gauche, les fenêtres B_p , obtenue en positionnant B sur p, pour un certain nombre de pixels p appartenant à X. A droite, la dilatation de X par B.

L'effet de la dilatation (figure 2-4) est d'abord d'élargir la figure, la hauteur et largeur de la figure dilatée seront les sommes respectivement des hauteurs et largeurs de la figure originelle et de l'élément structurant.

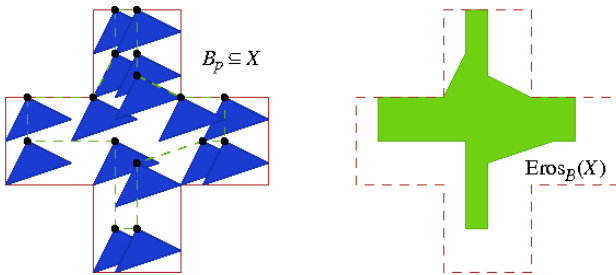


Figure 2-5 : À gauche, plusieurs positions (boules noires) de pixels p dont les fenêtres B_p sont incluses dans X. À droite, l'érosion de X par B.

L'effet de l'érosion (figure 2-5) est de rétrécir la figure, la hauteur et largeur de la figure érodée seront les différences respectivement des hauteurs et largeurs de la figure originelle et de l'élément structurant. On considère les mêmes figure X et élément structurant B que plus haut.

Une opération de fermeture a été appliquée à l'image binaire, mais le résultat obtenu n'est pas très pertinent.

2.3 Première méthode : Extraction des segments par transformée de Hough probabiliste.

L'image en entrée est une image seuillée, et le premier traitement à être effectué est l'étape de l'extraction des segments par la méthode de la transformée de Hough probabiliste, comme dans l'article suivant (Chiang, Knoblock, 2009).

En annexe 3, est expliquée la méthode de la transformée de Hough (TH).

La TH permet donc de déterminer des droites, cependant, l'objectif de cette étape est de déterminer des segments, ainsi, la méthode utilisée sera la TH probabiliste (qui est une méthode localisée). En effet, la TH probabiliste prend en compte la longueur de pixels connectés formant un linéaire.

De plus, contrairement à la TH standard qui est appliquée sur tous les pixels de l'image, la TH probabiliste est appliquée sur une proportion des pixels de l'image. Ainsi, la complexité de calcul est réduite¹³.

La méthode est expliquée en figure 2-6 : deux paramètres sont à prendre en compte, une longueur d'espacement (de trou) qui représente l'écart maximal entre deux pixels pour toujours former un segment et une longueur de segment minimale.

¹³ Tiré de la présentation de HASNAOU Y et al. « Détection des droites par la transformée de Hough » (<https://fr.slideshare.net/khaledfayala/dtection-des-droites-par-la-transforme-de-hough>) consultée de 2/05/17



Figure 2-6 - Méthodologie d'extraction d'un segment (extrait de Bernard, 2015).

2.3.1 Extraction de segments par transformée de Hough probabiliste :

L'extraction des segments se fera à l'aide de la méthode de la TH probabiliste. L'outil utilisé dans la chaîne de traitement est la fonction *probabilistic_hough_line* de la bibliothèque *scikit-image*. L'explication de l'algorithme est donnée dans (Galamhos, *et al*, 1999).

Pour cette fonction, il faut préciser la donnée en entrée :

- Une image seuillée, avec les contours en blanc,

Puis les paramètres :

- Un seuil, il faut que la valeur dans la nappe d'accumulation soit supérieure à ce seuil pour que l'algorithme passe à l'étape suivante : la recherche du segment le plus long,
- Une longueur de segment minimale,
- Une longueur « de trou » qui représente l'écart maximal entre deux pixels pour toujours former un segment,
- Une liste correspondant aux angles theta, ces angles définissent les segments en coordonnées polaires.

2.3.2 Détection des segments et création de sous-segments

L'étape suivante est la détection d'intersections entre les segments obtenus par la TH probabiliste et la décomposition des segments qui s'intersectent en sous-segments.

En effet, lorsque deux segments *s1* et *s2* se croisent, ils forment une intersection. Quatre sous-segments ont donc été créés :

- Point de départ de *s1* – intersection
- Point d'arrivée de *s1* – intersection
- Point de départ de *s2* – intersection
- Point d'arrivée de *s2* – intersection

Comme nous le verrons dans la partie 3.1.1 des traits de construction sont encore visibles sur ce type de planche et ceux-ci sont détectés par la TH probabiliste. L'intérêt de cette étape est tout d'abord de nettoyer le résultat obtenu à l'étape précédente. Ensuite, le second but est d'individualiser les sous-segments pour reconstruire les contours des parcelles en établissant la règle topologique suivante : chaque limite de parcelle doit être connectée à chacune de ses extrémités à avec une autre limite.

Et par la suite, chaque sous-segment passe un test de validation afin d'éliminer les cas de fausses alarmes. Cette étape permet de nettoyer le résultat. Pour se faire, les outils *intersects* et *intersection* de la bibliothèque Shapely sont utilisés, ainsi, des sous-segments ont été créés et placés dans une liste. Les segments qui n'intersectent aucun autre ont également été gardés et placés dans la même liste, qu'on appellera liste des sous-segments.

Ensuite, dans une troisième étape, validation des sous-segments, il convient de supprimer les sous-segments non valides, c'est à dire, ceux qui ne comptent pas suffisamment de points de contour.

Le but est de compter, pour chaque sous-segment, le nombre de point de contours : il faut que la valeur du pixel soit de 255 dans l'image seuillée.

Pour cela, la fonction *line_aa* du module *draw* de la bibliothèque scikit-image a été utilisée. Cette fonction permet d'obtenir, pour une image donnée, la valeur des pixels appartenant à un segment donné.

Les segments sont validés si :

$$\frac{\text{Nombre de points de contour}}{\text{Longueur du sous-segment}} > \text{seuil}$$

2.3.3 Connexion entre sous-segments validés

L'étape suivante de cette méthode consiste à connecter, lorsque c'est justifié, les extrémités des sous-segments valides. En effet, afin d'obtenir une continuité des limites, il est nécessaire, pour certains des sous-segments, de les connecter. Pour ce faire, on considère les extrémités d'autres sous-segments présentes dans un certain voisinage, il y a donc un paramètre de taille à prendre en compte afin de gagner du temps lors du traitement. Ensuite, la jonction n'est effectuée que si entre les deux extrémités, il y a un certain nombre de points de contours, la fonction *line_aa* a été utilisée, comme c'était le cas à l'étape précédente.

La connexion est validée si, entre les deux extrémités, le rapport entre le nombre de pixels ayant une valeur de 255 et la longueur qui sépare les deux extrémités est supérieur à un seuil.

$$\frac{\text{Nombre de points de contour entre les deux extrémités}}{\text{Longueur qui sépare les deux extrémités}} > \text{seuil}$$

2.3.4 Conversion au format vectoriel (SHP) :

La dernière étape est l'enregistrement des sous-segments valides en format Shapefile. Il se fait grâce à la fonction *Writing Shapefiles* de la bibliothèque *pyshp*.

2.4 Seconde méthode : Vectorisation par « suivi de chemin »

Cette méthode quant à elle fera appel, via les lignes de commande de l'outil GRASS, à des utilitaires de vectorisation.

Avant de commencer le traitement des données, il y a des détails techniques de mise en œuvre à prendre en compte :

Pour les données rasters, les commandes utilisées commenceront par « r. », alors que pour les données vecteurs, les commandes utilisées commenceront par « v. ».

Ensuite, les outils GRASS ne peuvent être appliqués qu'aux données appartenant à un projet GRASS. La première chose à faire est donc d'importer la donnée à traiter dans le projet GRASS.

Pour cela, il faut utiliser la commande : *r.in.gdal*

De plus, afin de limiter la zone géographique dans laquelle GRASS doit travailler, il faut définir une région. Cette notion de région permet aussi de ne travailler que sur une portion de la carte qui nous intéresse afin d'accélérer le traitement. De plus, il faut s'assurer que la région soit bien définie avant d'effectuer certains traitements. En effet, de nombreuses commandes ne fonctionnent que sur la région décrite comme notamment les commandes d'export et la plupart des modules liés aux données rasters. La fonction permettant de définir la région est le module *g.region*.

Une fois ces détails techniques pris en compte, le traitement peut commencer. La première étape consiste à re-catégoriser l'image, en effet, les outils utilisés dans la suite nécessitent certaines conditions : ils ne s'appliquent que sur les pixels qui ont une valeur non nulle et il faut que les contours aient une valeur de pixel à 0. Ainsi, il faut créer un fichier *regles.txt* contenant les modifications à appliquer et utiliser la commande *r.reclass*.

L'étape suivante consiste à squelettiser l'image, le principe est d'amincir les cellules qui désignent des lignes. Pour correspondre à des linéaires, ces objets doivent présenter une largeur unique de un pixel. En effet, pour pouvoir utiliser la fonction de vectorisation, il est nécessaire que l'image soit squelettisée. La fonction utilisée pour effectuer la squelettisation est la fonction *r.thin* qui utilise l'algorithme décrit par Jang et Chin R.T (1990). L'objectif est d'obtenir un squelette de un pixel d'épaisseur qui respecte les caractères topologiques de l'objet en entrée, dans le même cas que la fonction utilisée précédemment. Comme pour l'algorithme de squelettisation précédent, nous retrouvons le voisinage de 3x3 ainsi que le principe des sous-itérations, néanmoins, il ne s'agit d'une méthode différente. Pour la première sous-itérations, quatre modèles sont appliqués pour supprimer les pixels situés dans les coins nord-est, nord-ouest, sud-est et sud-ouest du voisinage d'un pixel appartenant au squelette. Pour la seconde sous-itération, quatre modèles sont également appliqués, ils permettent de supprimer les pixels situés au sud, nord, est et ouest du voisinage d'un pixel appartenant au squelette. Et il y a des itérations jusqu'à ce qu'il n'y ait plus de pixel à supprimer. Il s'agit d'une autre implémentation du même principe de squelettisation.

Et la dernière étape permet de vectoriser l'image, en effet, le but est d'obtenir des polylignes au format vectoriel. Cela se fait grâce à la fonction *r.to.vect*. Il faut préciser si on veut convertir le raster en points, lignes ou en polygone. Dans le cas présent, il s'agit de lignes. L'algorithme¹⁴ utilisé est basé sur le suivi de chemin : en effet, pour un pixel de contour donné, il va rechercher les pixels voisins connexes en considérant son voisinage et en ne revenant jamais en arrière.

¹⁴L'algorithme utilisé pour la fonction *r.to.vect* est disponible sur https://trac.osgeo.org/grass/browser/grass/branches/releasebranch_7_2/raster/r.to.vect/lines.c, consulté le 16 juin 2017.

2.5 Etapes finales communes aux deux méthodes

2.5.1 Nettoyage topologique des vecteurs

Une fois les vecteurs obtenus, il est nécessaire de les nettoyer, cela peut se faire à l'aide de différents outils de GRASS. Il y a notamment la fonction *v.clean* qui est constituée d'un ensemble d'outils permettant de nettoyer la topologie de données vectorielles, il faut préciser l'outil à utiliser dans les paramètres de la fonction. Il est également possible d'utiliser l'outil *v.generalize* avec l'algorithme de Douglas-Peucker qui permet de lisser des données vectorielles ainsi que l'outil *v.build.polylines* qui construit des polygones à partir de lignes.

2.5.2 Transformation des lignes en objets surfaciques

Les données actuelles sont les segments en format vecteur SHP nettoyés, cette étape permettra donc de transformer ces segments en polygones. Néanmoins, il est nécessaire de passer par le format raster pour obtenir des polygones, en effet, il n'existe à notre connaissance pas d'outil ou de fonction implémentée pour passer de polygones en format vectoriel aux polygones en format vectoriel. Alors qu'à partir de données rasters, cela se fait grâce à une labélisation.

2.5.2.a Conversion en format image

Cette étape consiste à rasteriser les données vectorielles, cela se fait grâce à l'outil *v.to.r*. Il faut ensuite l'exporter du projet GRASS.

2.5.2.b Labellisation des régions

Une région ou objet est un ensemble de pixels connectés de même valeur. Il y a deux types de connectivités, 4-c et 8-c (voir section 2.2.3). Le but de cette étape est que chaque ensemble de pixels connectés et de même valeur soit transformé en région et que chacune d'entre elle soit labellisée par un numéro différent. Il s'agit de la labellisation.

Une fois les images rasters obtenues, on peut procéder à la labellisation, elle se fait grâce à l'outil *label* du module *measure* de la bibliothèque scikit-image de Python. L'algorithme est basé sur l'article de Fiorio et Gustedt (1996).

Il faut préciser la donnée en entrée :

- Une image

Les paramètres en entrée sont :

- Le type de connectivité (4 ou 8)
- La valeur des pixels qui seront considérés comme arrière-plan, par défaut, les pixels de valeur 0 sont considérés comme l'arrière-plan.

Pour chaque ensemble de pixels connectés et de même valeur, la fonction crée une région et lui confère un label.

Une fois la labélisation effectuée, afin d'éliminer les pixels de valeur 0 entre chaque limite de parcelle, nous avons mis en place un filtre majoritaire. Pour chaque pixel de valeur 0, et dans une fenêtre 5x5, la valeur de ce pixel est remplacée par la valeur des pixels majoritaires, mais différent de 0.

2.5.2.c Vectorisation des régions

Ensuite, l'image labellisée est importée dans GRASS à l'aide de la fonction `r.in.gdal` puis vectorisée avec la fonction `r.to.vect` (option « `type=area` ») . L'algorithme transforme des régions de même label en polygone.

3 Application de la méthodologie sur les données et ses résultats

3.1 Présentation des données – analyse des difficultés

Les données mobilisées dans ce projet sont des planches cadastrales scannées, datant pour les plus anciennes de 1813 et pour les plus récentes de 1970. Ces planches sont établies pour les communes de Vaas et Aubigné-Racan, toutes les deux situées dans le département de la Sarthe. Les données les plus anciennes, datées de 200 ans, ont subi des dégradations liées à leurs conditions de conservation. Il y a donc un nombre important de difficultés rencontrées, tels que des éléments ou du bruit pouvant perturber la détection des linéaires des limites de parcelles. On peut alors distinguer les causes suivantes : âge, actions humaines, quantité des informations représentées. Il faut également tenir compte des caractéristiques des fichiers issus de la numérisation. De plus, depuis 1813, les méthodes pour la réalisation de plans cadastraux se sont nettement améliorées (règlement du 25 Mars 1827). En 1930, la rénovation du cadastre est décidée, elle se déroule entre 1930 et 1975 (loi du 16 Avril 1930). Il n'y a pas eu de plan cadastral établi entre 1850 et 1974 pour Vaas et Aubigné-Racan¹⁵.

De plus, nous constatons une différence importante entre les planches de 1813 et de 1850, en effet, les planches de 1850 semblent bien mieux préservées que celles de 1813. Cependant, nous n'avons pas d'informations concernant les conditions de conservation des planches à cette date. Nous pouvons seulement émettre des hypothèses¹⁶ telles que l'évolution de la qualité du papier, ou la construction de bâtiments (mairie, école, ...) offrant de meilleures conditions de conservation. De plus, ces plans étaient réalisés en plusieurs exemplaires et chacun était probablement conservé aux archives départementales, au service du cadastre et à la mairie. Il est donc certain que ceux conservés à la mairie ont été plus manipulés que ceux qui étaient dans des locaux administratifs non-ouverts au public. Enfin, la loi du 10 mai 1838 fait de la conservation des archives une dépense obligatoire pour les départements.¹⁷

Ainsi, le traitement sera effectué sur trois planches test, chacune établie à une date différente (1813, 1850 et 1972) et ceci pour chacune des communes. Le travail de vectorisation a été réalisé sur trois planches tests de la commune d'Aubigné-Racan, une première planche n°008 datant de 1813, une deuxième, n°90 de 1850 et une dernière, n°155 de 1972, elles sont visibles en annexe 1. Il s'agit de planches de qualité hétérogène dans lesquelles nous retrouvons certaines caractéristiques communes mais aussi des problématiques différentes.

¹⁵ Entretien du 17 Mars 2017 avec Gilles Berteau, administrateur des finances publiques adjoint.

¹⁶ Entretien du 30 Mai 2017 avec Pierre Clergeot, ancien professeur de cadastre à l'ESGT et spécialiste du cadastre ancien.

¹⁷ Cette loi est citée dans le Répertoire méthodique et alphabétique de législation de doctrine et de jurisprudence en matière de droit civil, commercial, criminel, administratif, de droit des gens et de droit public, Volume 5.

3.1.1 Cas des planches les plus anciennes (1813)

Les planches de 1813 sont celles ayant subi le plus de dégradations.

- Les dégradations liées au vieillissement

A l'époque où elles ont été scannées (années 2000), les planches étaient âgées de près de 200 ans. Il y a donc des dégradations importantes liées à cela.

Par exemple, les planches sont jaunies, sur les côtés, il y a des marques liées à l'humidité, et sur les bords le papier est dégradé.

- Les dégradations liées aux actions humaines

L'action humaine a également dégradé les planches. En effet, sur certaines il y a des taches d'encre, des marques de plis, comme le montre la Figure 3-1 les plis sont généralement horizontaux et verticaux et peuvent être confondus avec des limites. Et, sont encore visibles sur les planches, ce qu'on imagine être des traits de constructions, c'est à dire, les traits d'appui qui permettaient de calculer les surfaces des parcelles¹⁸. En effet, le calcul de la surface d'une parcelle singulière sur plan papier se fait en décomposant la parcelle en éléments géométriques simples (triangles, trapèzes) puis en calculant et additionnant les surfaces de tous ces éléments.



Figure 3-1 : Extrait d'une planche datant de 1813.

- Les difficultés en lien avec les informations représentées

Les informations graphiques spécifiées sur les planches représentent des difficultés pour la vectorisation. Il y a pour commencer des informations textuelles telles que le cartouche, les numéros de parcelle, les numéros des sections, le numéro de la feuille, les noms de rues, les noms de lieux-dits. On note aussi la présence en bordure de liseris de couleur rouge ou noire qui représentent des limites de section. Il y a également des informations graphiques spécifiques aux cartes, telles que l'échelle, la flèche nord. Concernant les limites de parcelles, leur continuité n'est pas parfaitement respectée, à certains endroits, les traits ont été effacés.

¹⁸ Entretien du 17 Mars 2017 avec Gilles Berteau, administrateur des finances publiques adjoint.

- Les caractéristiques des fichiers issus du scannage des planches

Le scannage des planches a permis la création d'une base d'images numériques. La numérisation a été effectuée par le service des Archives Départementales ou le SDNC. Il est donc intéressant de connaître les paramètres utilisés lors de cette étape Fahrasmane, (2016). Ces informations sont récapitulées dans le Tableau 3-1.

L'ensemble des images numériques de cette période sont codées sur trois canaux (rouge, vert, bleu) et chaque canal est codé entre 0 et 255. L'image test qui sera utilisée est de dimensions 8081 x 5160 : elle correspond à la planche 008.

Tableau 3-1 : Propriétés des images scannées des plans de 1813 d'après Fahrasmane, 2016

Année	1813	
Commune	Aubigné-Racan	Vaas
Échelle	1/2500	1/1250
		1/2500
Effectifs de planches	3	1
		8
Pas du scanner	200 dpi	200 dpi
Taille d'un pixel dans l'image (µ)	127	127
Taille d'un pixel sur le terrain (m)	0,3175	0,16
		0,32

3.1.2 Cas des planches intermédiaires (1850)

Il s'agit des planches de 1850, on retrouve certains aspects déjà présents pour les planches plus anciennes et certains disparaissent, comme nous allons le détailler dans les paragraphes suivants.

- Les dégradations liées au vieillissement

On retrouve la même problématique que précédemment, en effet, les planches ont été conservées 150 ans ce qui explique qu'elles soient jaunies. Néanmoins, il n'y a pas de trace liée à l'humidité ce qui implique que les planches sont dans un meilleur état général que celles de 1813.

- Les difficultés en lien avec les informations représentées

Bien entendu, on retrouve les informations graphiques et textuelles spécifiques aux planches cadastrales : la flèche nord, l'échelle, le cartouche, les numéros de parcelle, les numéros des sections, le numéro de la feuille, le nom des rues, le nom des lieux-dits. On note aussi la présence en bordure de liseris de couleur jaune qui représentent des limites de section. S'ajoute à cela, la présence de signes d'appartenance et d'un carroyage horizontal et vertical qui peut être confondu avec les limites cadastrales comme nous pouvons le constater dans la Figure 3-2.

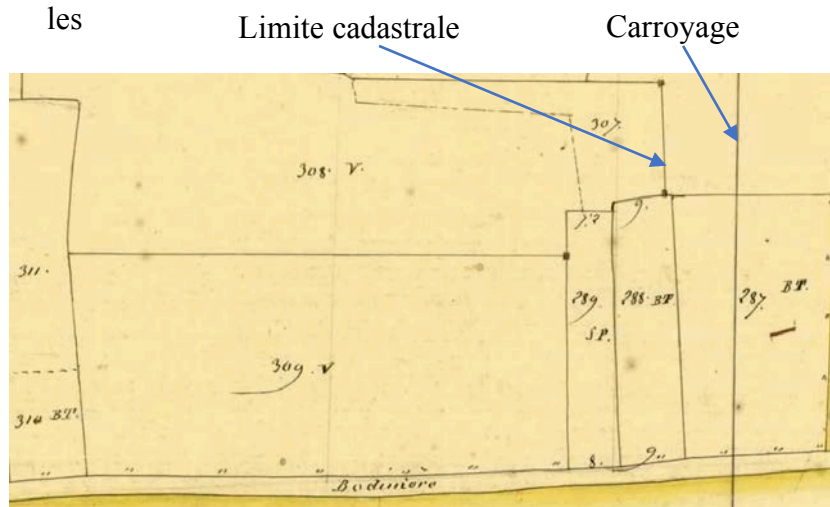


Figure 3-2 : Extrait d'une planche datant de 1850.

En effet, les traits de carroyage sont généralement horizontaux et verticaux et de même couleur et épaisseur que les limites de parcelles. De plus, certaines limites apparaissent en pointillées, il s'agit de biens non délimités (BND) pour lesquels il y a une incertitude sur l'emplacement de la limite¹⁹. Pour terminer, des lettres apparaissent à droite des numéros de parcelles, il s'agit du type d'occupation des sols.

Les limites de parcelles sont bien distinguées, elles sont continues (elles ne sont pas interrompues, il n'y a « pas de trou »).

- Les caractéristiques des fichiers issus du scannage des planches

Tableau 3-2 : Propriétés des images scannées des plans de 1850 d'après Fahrasmane, 2016.

Le pas du scannage est le même que précédemment, la seule différence est l'échelle des planches qui est ici plus grande que pour les planches de 1813.

La résolution spatiale est donc meilleure. Ces informations sont récapitulées dans le Tableau 3-2. L'ensemble des images numériques de cette période sont codées sur trois canaux (rouge, vert, bleu) et chaque canal est codé entre 0 et 255. L'image test qui sera utilisée correspond à planche 090 de dimensions 8036x5424.

Année	1850	
	Aubigné-Racan	Vaas
Commune	Aubigné-Racan	Vaas
Échelle	1/2000	1/500
		1/1000
		1/2000
Effectifs de planches	6	4
		7
		12
Pas du scanner	200 dpi	200 dpi
Taille d'un pixel dans l'image (μ)	127	127
Taille d'un pixel sur le terrain (m)	0,254	0,06
		0,13
		0,25

3.1.3 Cas des planches les plus récentes (1972)

Les planches traitées les plus récentes datent de 1972 il s'agit de planches rénovées, aussi la qualité du plan est nettement supérieure aux planches anciennes (voir section 1.1). De plus, les planches ont été conservées durant une période nettement plus courte et les méthodes de conservation se sont améliorées, ainsi les dégradations liées au vieillissement sont absentes. Ces planches ont beaucoup moins de défauts.

¹⁹ Entretien du 18 Mai 2017 avec Gilles Berteau, administrateur des finances publiques adjoint

- Les difficultés en lien avec les informations représentées

Pour ces planches, les seules difficultés sont la présence des différentes informations textuelles telles que la flèche nord, l'échelle, le cartouche, les numéros de parcelle, les numéros des sections, le numéro de la feuille, le nom des rues, le nom des lieux-dits. Il y a également des informations graphiques telles que la présence de bâtiments représentés par des rectangles hachurés, les croix de carroyage, les signes d'appartenances. On note aussi la présence de linéaires (lignes non continues) comme nous le voyons sur Figure 3-3 matérialisant des informations autres que les limites de parcelles : limites administratives (limites de communes, limites d'un lieu-dit) et ligne de chemin de fer. Mais au-delà de ces difficultés, les planches sont très nettes, le fond est homogène, le tracé est franc, les lignes sont parfaitement continues (il n'y a pas de « trous »), l'épaisseur de trait est relativement constante.



Figure 3-3 : Extrait d'une planche datant de 1972.

- Les caractéristiques des fichiers issus du scannage des planches

Tableau 3-3 : Propriétés des images scannées des plans de 1972-1974 d'après Fahrasmane, 2016

Année	1972-1974	
	Aubigné-Racan	Vaas
Commune	Aubigné-Racan	Vaas
Échelle	1/2000	1/1000
		1/2000
Effectifs de planches	6	10
		10
Pas du scanner	400 dpi	400 dpi
Taille d'un pixel dans l'image (μ)	63,5	63,5
Taille d'un pixel sur le terrain (m)	0,127	0,0635
		0,127

On retrouve sensiblement les mêmes échelles que pour les planches précédentes. La seule différence est le pas du scannage qui est supérieure ce qui signifie que les images numériques récentes ont une meilleure résolution spatiale, comme nous le constatons dans le tableau 3-3. L'ensemble des images numériques de cette période sont à valeurs binaires (0 pour le noir et 255 pour le blanc). L'image test qui sera utilisée correspond à la planche 155 de dimensions 16630x11533.

Au vu des différences de qualité entre les planches anciennes et plus récentes, certains traitements devront être réalisés pour les planches anciennes qui ne seront pas nécessaires pour les planches de 1972.

3.1.4 Différence de qualité radiométrique des images

Afin de démontrer quantitativement la différence de qualité radiométrique des images, la variation de luminance pour chacune des trois images a été mesurée. Il s'agit de déterminer si chaque pixel représente de manière correcte une caractéristique du plan.

Pour cela, une dizaine de zones de différentes tailles ne contenant pas de contours a été sélectionnée sur chaque image convertie en niveaux de gris (voir section 2.2.1). Pour chaque

zone, l'écart-type²⁰ a été calculé et le bruit moyen (variation de luminance) correspond à la moyenne des écart-types de chaque zone. Ce travail a été réalisé manuellement à partir du logiciel de traitement d'images de télédétection ENVI.

Pour l'image correspondant à la planche 008 (1813), le bruit moyen est de 8,20 et pour celle correspondant à la planche 090 (1850), il est de 4,7. Pour l'ensemble des images numériques correspondant aux planches datant de 1972, la valeur du bruit moyen est à 0 puisque qu'elles sont à valeur binaire. Pour les planches de 1813 et 1850, cela signifie qu'il y a une variation de la luminance entre des pixels voisins et cette variation est plus importante dans le cas des planches plus anciennes.

Finalement, la vectorisation semi-automatique va être réalisée sur trois images :

- L'image correspondant à la planche 008 (1813) de la commune d'Aubigné-Racan,
- L'image correspondant à la planche 090 (1850) de la commune d'Aubigné-Racan,
- L'image correspondant à la planche 155 (1972) de la commune d'Aubigné-Racan.

Ces images se trouvent en annexe 1.

De plus, avant d'évaluer quantitativement les résultats, ces mêmes images vont être vectorisées manuellement, il s'agira de références.

3.2 Application de la méthodologie sur les données

3.2.1 Le prétraitement

3.2.1.a Dans le cas des données anciennes (1813 et 1850)

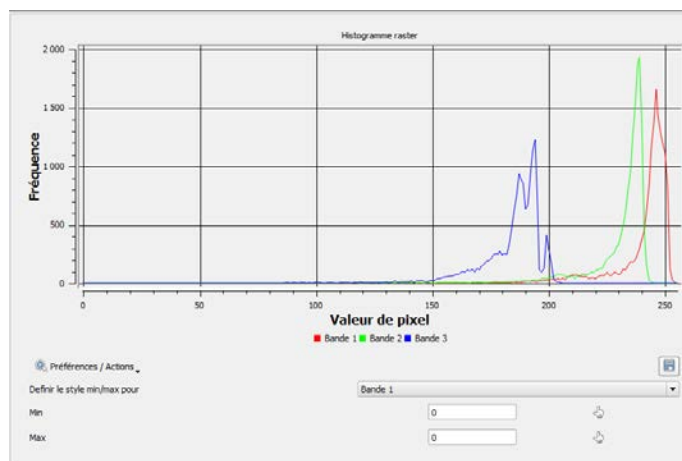
- **Conversion en niveau de gris**

Les données en entrée sont des images couleurs. Le résultat de la conversion est présenté en annexe 2.

Néanmoins, on constate que l'image est soumise à un bruit important qu'on a estimé dans la section 3.1.4.

La Figure 3-4 représente l'histogramme de l'image en entrée. Les niveaux plus importants dans les canaux rouges et vert que dans le canal bleu expliquent l'apparence jaunie de la planche scannée.

Ainsi, lors de la transformation de l'image couleur en niveaux de gris, nous modifierons la formule initiale et le coefficient appliqué au canal bleu est



²⁰ L'écart-type est la mesure de dispersion de données. Il est défini comme la moyenne quadratique des écarts par rapport à la moyenne.

augmenté. Après plusieurs tests pour différentes valeurs de coefficients, dans une zone de l'image censée être homogène, le résultat offrant une variation de la valeur de pixels voisins la moins importante est obtenu avec la formule suivante :

$$\text{Niveau de gris} = 0,1 \text{ Rouge} + 0,1 \text{ Vert} + 0,8 \text{ Bleu}$$

De plus, les niveaux de gris sont inversés pour faire apparaître les points de contours en niveaux de gris clairs, alors qu'ils étaient foncés initialement. Le résultat de cette conversion est illustré en annexe 2.

Nous pouvons voir l'amélioration apportée par la formule dont les coefficients ont été réajustés. En effet, dans une zone censée être homogène de l'image, la variation de la valeur de pixels voisins est moins importante que dans le cas du résultat obtenu par défaut, ce qui facilitera l'étape du seuillage. Cette différence se voit plus particulièrement sur les côtés de l'image et lors de l'étape suivante.

- **Application d'un masque et détection des points de contour par seuillage.**

L'application du masque se fait au moment du seuillage, le résultat de la méthode du seuillage simple est illustré en annexe 2. On constate que de nombreux points de contour disparaissent. Ainsi, nous avons appliqué un seuillage par Hystérésis, le résultat de la méthode du seuillage par Hystérésis sur l'image de la planche 008 est illustré en annexe 2. De nombreux points de contour ne sont pas détectés même si l'image n'est pas énormément bruitée. Dans ce cas, cette méthode ne procure donc pas de résultat satisfaisant. En effet, l'image montre des variations importantes d'éclairement et des seuils uniques sur toute l'image ne sont pas adaptés.

Ainsi, la méthode de la moyenne glissante a été envisagée, elle consiste à déterminer localement un seuil en chaque pixel à partir de la moyenne locale des niveaux de gris.

Cette méthode a été appliquée à deux images converties en niveaux de gris : une première par la formule de la CIE, et une seconde par la formule adaptée. Les résultats sont en annexe 2. Le résultat semble meilleur dans le cas de l'image qui a été convertie en niveaux de gris par la formule adaptée. En effet, il est beaucoup moins bruité. Néanmoins, par endroits, certains contours ne sont pas continus. Nous avons alors imaginé un seuillage par hystérésis local qui est une association entre le seuillage par moyennes glissantes et le seuillage par Hystérésis. Le résultat est illustré en annexe 2.

Ensuite, afin de valider cette dernière méthode, les résultats ont été comparés avec ceux obtenus par moyennes glissantes.

- Validation de la méthode de seuillage par Hystérésis local :

Pour commencer, une comparaison visuelle des résultats a été effectuée, elle est illustrée en Figure 3-5. Une meilleure continuité des contours est obtenue dans le cas de la méthode par Hystérésis locale, cependant, il y a plus de bruit, c'est-à-dire des points de contours isolés et en dehors des limites de parcelles recherchées mais ceci pourra être corrigé dans un second temps.

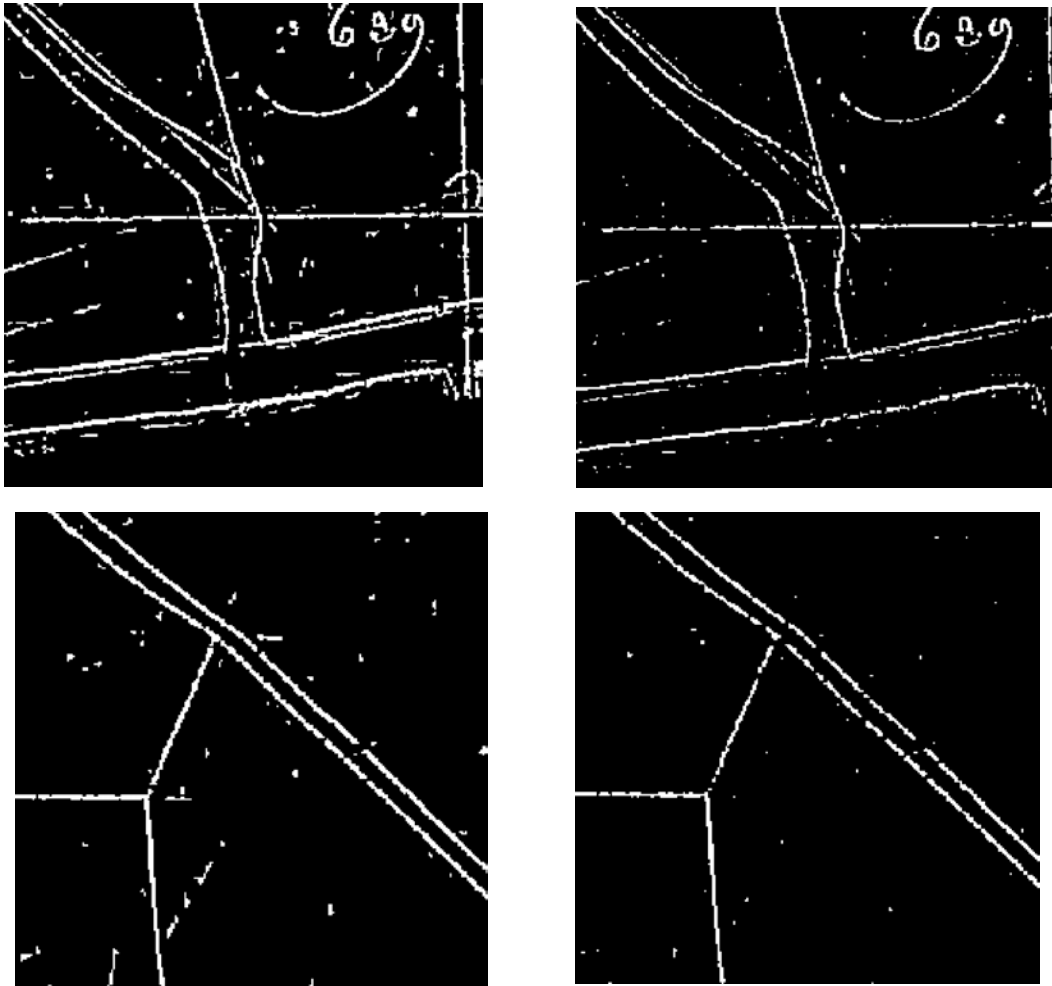


Figure 3-5 : Images seuillées par la méthode Hystérésis locale (à gauche) et par la méthode de la moyenne glissante (à droite)

De plus, une évaluation quantitative des qualités de détection a aussi été effectuée, en effet, l'image correspondant à la planche 008 a été vectorisée manuellement, puis rastérisée, ce qui permet de comparer les deux méthodes (seuillage par Hystérésis locale et moyenne glissante) avec une référence établie manuellement.

Deux probabilités ont été calculées, avec « référence », l'image correspondant à la planche 008 vectorisée manuellement et « résultat », le résultat du seuillage.

$$\text{Probabilité de détection} = \frac{\text{Nombre de points de contour dans référence et dans résultat}}{\text{Nombre de points de contour dans référence}}$$

$$\text{Probabilité de fausses alarmes} = \frac{\text{Nombre de points de contour dans résultat et pas dans référence}}{\text{Nombre de points de contour dans résultat}}$$

Tableau 3-4 : Comparaison de la méthode de la moyenne glissante et Hystérésis locale

Moyenne glissante	n=5	Hystérésis locale	n=5
	c = 1,12		h : 1,20 b : 1,10
Probabilité détection	92%	Probabilité détection	94%
Probabilité fausses alarmes	89%	Probabilité fausses alarmes	87%

Dans le Tableau 3-4 la variable n représente le voisinage, il s'agit d'une fenêtre carrée centrée sur le pixel considéré.

La méthode par Hystérésis locale permet de détecter davantage de points de contour avec une probabilité de fausses alarmes plus faible. Notons que la forte valeur obtenue pour la probabilité de fausses alarmes est liée au fait que les écritures sont elles aussi détectées. Ainsi, la méthode par Hystérésis locale est

retenue pour la suite du traitement.

Ces deux méthodes sont dépendantes de choix réalisés de façon empirique (taille du voisinage, constantes), ce qui pose problème pour leur utilisation dans un contexte semi-automatique.

Aussi, les deux méthodes vont être comparées afin de tester leur robustesse. Pour chaque méthode, les valeurs du voisinage, de ou des constantes vont être modifiées afin de tester leur sensibilité, les résultats sont dans les Tableau 3-5 et Tableau 3-6

Tableau 3-5 : Probabilités de détection et de fausses alarmes des deux méthodes avec variation de la valeur de ou des constantes de 0,1

Moyenne glissante	c = 1,10			Hystérésis locale	h = 1,20 b = 1,0		
	n=3	n=5	n=9		n=3	n=5	n=9
Probabilité détection	90%	93%	95%	Probabilité détection	96%	97%	98%
Probabilité fausses alarmes	87%	90%	92%	Probabilité fausses alarmes	99%	99%	99%

Moyenne glissante	c = 1,20			Hystérésis locale	h = 1,30 b = 1,10		
	n=3	n=5	n=9		n=3	n=5	n=9
Probabilité détection	77%	88%	92%	Probabilité détection	91%	94%	96%
Probabilité fausses alarmes	81%	86%	90%	Probabilité fausses alarmes	81%	87%	89%

Pour commencer, dans le Tableau 3-5, les valeurs de la ou des constantes sont proches de celles déterminées de façon empirique. Les résultats montrent que les probabilités de détection sont du même ordre de grandeur.

Ensuite, d'après les résultats de ce tableau, nous constatons que, dans le cas des deux méthodes, plus la taille du voisinage (n) est grande, plus la probabilité de détection augmente, mais c'est aussi le cas de la probabilité de fausses alarmes.

De plus, nous constatons que, lorsque pour les deux méthodes, les valeurs de ou des constantes varient de 0,1, la probabilité de détection obtenue par la méthode de la moyenne

glissante a une amplitude de variation plus importante. En effet, pour les deux constantes et toutes les tailles de voisinages confondus, la probabilité de détection s'étend de 77% à 95%. Par comparaison, pour la méthode d'Hystérésis locale, la probabilité de détection s'étend de 91% à 98%. La méthode de la moyenne glissante est plus sensible au choix de la constante « c ».

Tableau 3-6 : Probabilités de détection et de fausses alarmes des deux méthodes avec variation importante des valeurs de ou des constantes.

Moyenne glissante	c = 1,0			Hystérésis locale	h = 1,50 b = 0,50		
	n=3	n=5	n=9		n=3	n=5	n=9
Probabilité détection	96%	97%	97%	Probabilité détection	100%	100%	100%
Probabilité fausses alarmes	98%	98%	98%	Probabilité fausses alarmes	99%	99%	99%

Moyenne glissante	c = 0,7			Hystérésis locale	h = 1,0 b = 0,4		
	n=3	n=5	n=9		n=3	n=5	n=9
Probabilité détection	99%	100%	100%	Probabilité détection	100%	100%	100%
Probabilité fausses alarmes	99%	99%	99%	Probabilité fausses alarmes	99%	99%	99%

Dans le Tableau 3-6, les valeurs des constantes choisies sont assez éloignées des valeurs « idéales », de plus, l'amplitude entre les seuils hauts et bas est assez importante. Les résultats, pour les deux méthodes, s'approchent tous de 100%, donc tous les points de contour sont détectés, mais il y a aussi énormément de fausses alarmes.

Finalement, nous considérons que la méthode par moyenne glissante est moins robuste dans le sens où elle est plus sensible aux variations de paramètres.

- **Suppression d'objets isolés**

Dans le cas de l'image étudiée, les points de contour détectés ne sont pas parfaitement continus, il y a des « trous ». Ainsi, la taille des objets à supprimer ne doit pas être trop importante au risque de supprimer des ensembles de pixels représentant des contours corrects.

3.2.1.b Dans le cas des données récentes (1972)

Les données en entrée sont des images binaires de bonne qualité. L'image utilisée dans cet exemple est l'image n°155. Il n'est donc pas nécessaire de convertir l'image et réaliser un seuillage.

- **Application d'un masque**

Pour cette image, la première étape est l'application du masque. Cette étape a la même utilité que précédemment.

- **Squelettisation**

La deuxième étape consiste à squelettiser l'image. La continuité et l'absence de trous dans les points de contour permet d'obtenir un résultat satisfaisant, la continuité des lignes est maintenue.

- **Suppression d'objets isolés**

Ensuite, cette étape permet la suppression des numéros de parcelles et nom des rues, le résultat est visible en Figure 3-6. Néanmoins, certaines écritures sont connexes aux lignes et n'ont donc pas été supprimées. Ainsi, il n'y a pas de perte d'information puisqu'il n'y a pas de bout de contour isolé et que ceux-ci sont continus. La même fonction que précédemment a été utilisée (*remove_small_object*).

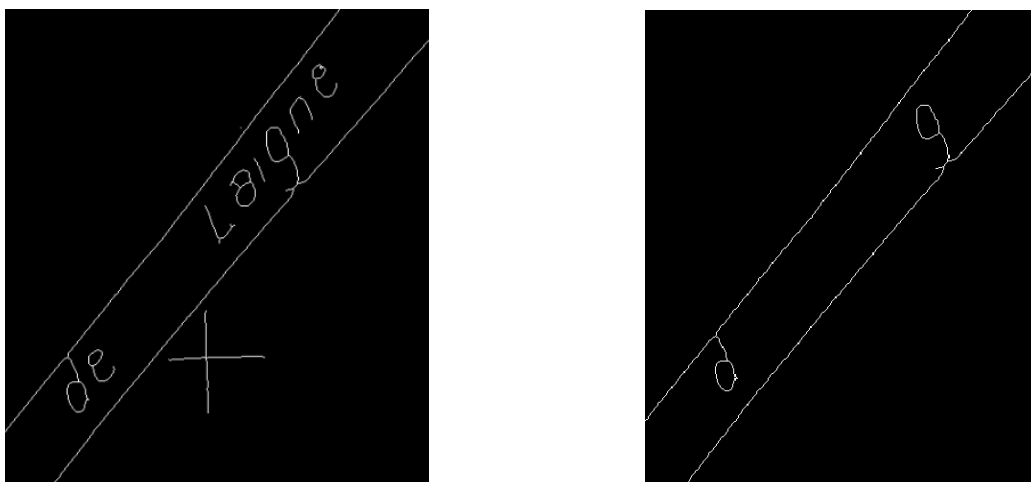


Figure 3-6 : Image squelettisée (à droite) et image après suppression d'objets isolés (à gauche).

Une fois que le pré-traitement est effectué et que les contours sur les images sont déterminés, l'objectif suivant est l'extraction des polygones en format vectoriel Shapefile.

3.2.2 Première méthode : Extraction des segments par transformée de Hough probabiliste.

3.2.2.a Dans le cas des données anciennes (1813 et 1850)

- **Extraction des segments**

La fonction utilisée nécessite quatre paramètres, un seuil, une longueur de segment minimale, une longueur « de trou » et une liste correspondant aux angles theta.

Tout d'abord, concernant la liste d'angles theta, il s'agit d'une liste d'angles variant entre 0 et 2π de pas $0,5^\circ$. Pour le seuil, une valeur de 15 a été décidée : en effet, une valeur trop petite entraînerait plus de cas de fausses alarmes. Ces paramètres sont ceux qui influencent

le moins les résultats, leur détermination est donc plus facile. Aussi, dans les exemples à suivre, ce sont ces valeurs qui auront été choisies.

Ensuite, le but est de trouver le bon compromis sur les autres paramètres pour détecter un maximum d'éléments tout en évitant les cas de fausses alarmes, dues aux pliures et aux traits de constructions. Ainsi, une longueur de segment minimale importante permettra de détecter des segments plus grands, d'éviter certains cas de fausses alarmes dues aux écritures et aux traits de constructions, donc le résultat est moins bruité. Néanmoins, certains segments qui devraient être détectés ne le sont pas car ils ne sont pas suffisamment prononcés. Nous pouvons le constater sur la Figure 3-7 qui permet de comparer un extrait d'une image ayant subi la TH probabiliste avec des paramètres identiques sauf pour la longueur de segment minimale.

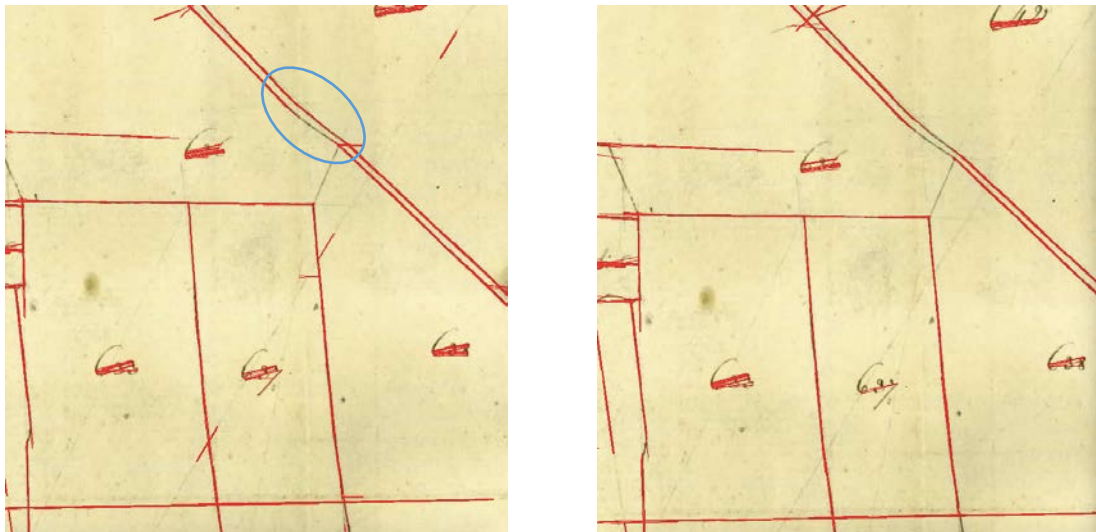


Figure 3-7 : Extraits de la planche 008 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 30 et de 15 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 50 et de 15.

On constate que l'image de droite possède moins de cas de fausses alarmes, cependant, un segment n'est pas détecté alors qu'il l'est dans l'image de gauche.

De plus, on retrouve ces problématiques pour le paramètre de la longueur de trou. Un paramètre de valeur trop grande ne permettra pas d'obtenir un résultat satisfaisant, car trop de cas de fausses alarmes. Et avec un paramètre de valeur trop petite, certains segments ne seraient pas détectés alors qu'ils le devraient. La Figure 3-8 permet de comparer un extrait d'une image après la TH probabiliste avec des paramètres identiques sauf pour la longueur de trou.

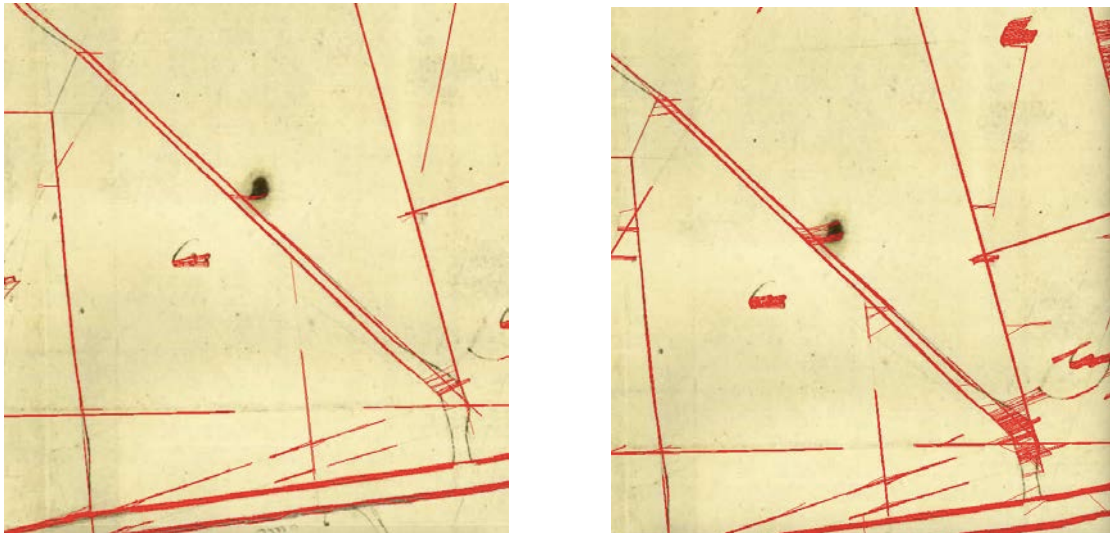


Figure 3-8 : Extraits de la planche 008 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 30 et de 15 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 30 et de 30.

Dans le cas de l'image de droite, le résultat n'est pas exploitable, en effet, il y a énormément de cas de fausses détections qui ne peuvent pas tous être corrigés.

Les valeurs des paramètres fournissant le meilleur compromis pour l'image de la planche 008 sont :

- 15 pour le seuil,
- 30 pour la longueur de ligne minimale,
- 15 pour la longueur d'intervalle
- $0,5^\circ$ pour l'intervalle d'angles theta.

Une des limites de cette méthode est que, ici aussi, la détermination de ces paramètres se fait de manière empirique.

• Détection des segments et créations de sous-segments

Cette étape permet d'éliminer un nombre important de fausses alarmes, on peut le constater dans la Figure 3-9. Le résultat est ainsi moins bruité.

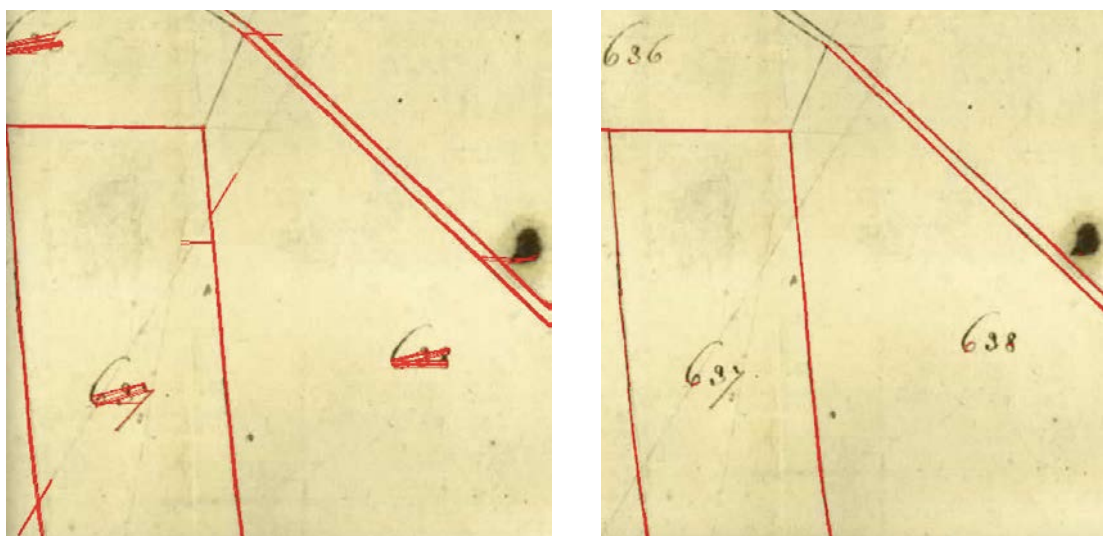


Figure 3-9 : Extrait de la planche 008 avant et après validation des sous segments avec un seuil de 80%.

Néanmoins le risque est que des sous-segments valides qui correspondent aux limites d'une parcelle disparaissent parce que les contours ne sont pas suffisamment marqués dans l'image initiale.

- **Connexion entre les sous-segments validés**

Cette étape permet de connecter, lorsque c'est justifié, les extrémités de deux sous-segments valides afin d'obtenir une continuité des limites. La valeur du seuil est élevée afin d'éviter des créer des cas de fausses détections.

Le résultat est visible à la Figure 3-10. On constate que des extrémités ont été connectées, en effet, il s'agit de limite de parcelle.

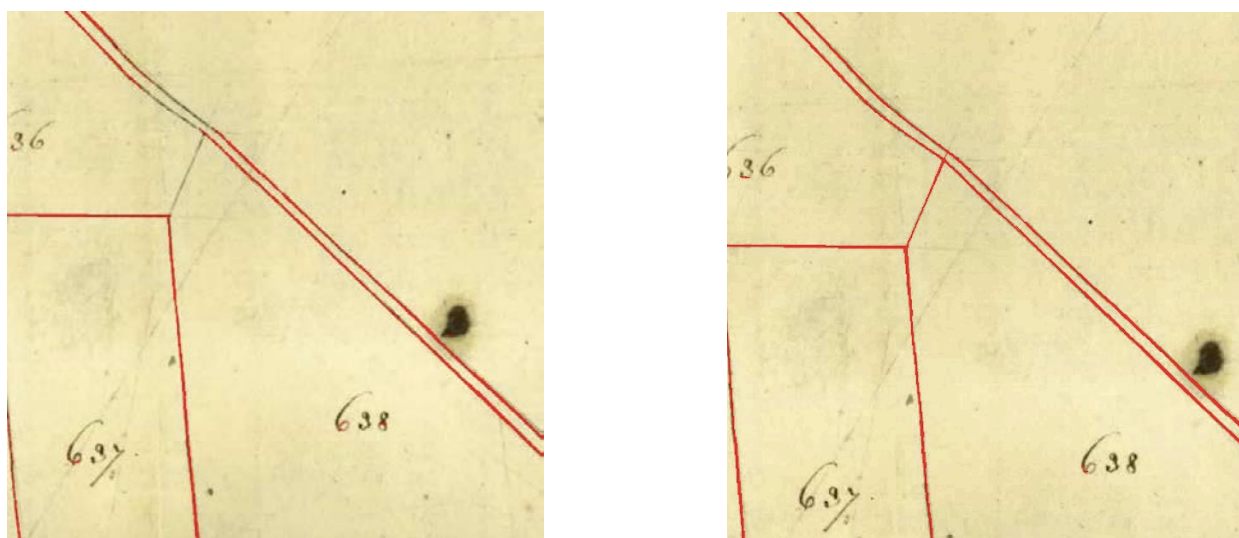


Figure 3-10 : Extraits de la planche 008 avant et après connexion des sous-segments. A gauche, il s'agit de l'extrait de l'image après validation des sous-segments (étape 3) et à droite, l'extrait de l'image après jonction des sous-segments avec un seuil de 90%.

3.2.2.b Dans le cas des données récentes (1972)

- **Extraction de segments**

Dans le cas de cette image, il faut également trouver le bon compromis pour détecter le maximum de lignes en évitant les fausses détections liées aux écritures.

Tout d'abord, concernant la valeur de theta, un pas de $0,5^\circ$ a été décidé. Pour le seuil, une valeur de 5 a également été décidée. En effet, la planche est très nette, le fond est homogène, le tracé est franc, il n'y a pas de trace de construction ni de pliure, il y a donc beaucoup moins de risque de fausses détections.

Concernant les deux autres paramètres, si la longueur des lignes est trop grande, certaines courbes ne sont pas détectées, et si elle est trop petite, les écritures sont davantage détectées. Cela est visible sur la Figure 3-11.

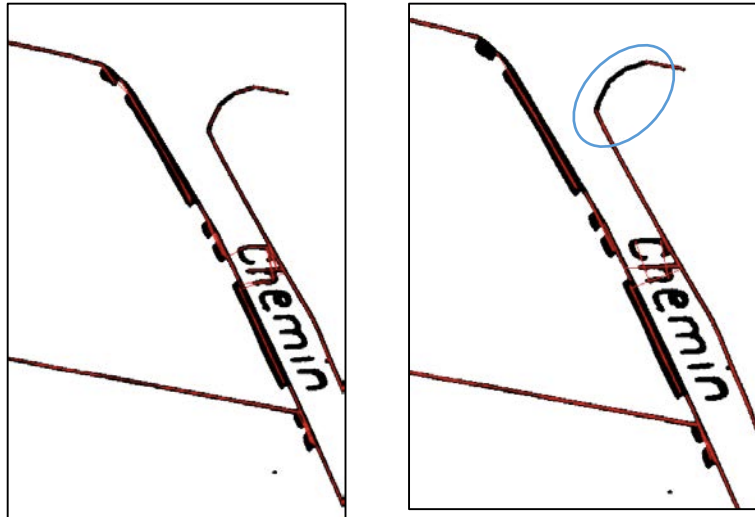


Figure 3-11 : Extraits de la planche 155 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 20 et de 20 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 40 et de 20.

Et pour le paramètre de la longueur de l'intervalle, s'il est trop grand, là aussi, les écritures sont détectées, et s'il est trop petit, les courbes ne sont pas détectées. Les valeurs des paramètres fournissant le meilleur compromis pour l'image de la planche 155 sont :

- 5 pour le seuil,
- 20 pour la longueur de ligne minimale,
- 20 pour la longueur d'intervalle
- 0,5° pour l'intervalle d'angles theta.

- **Validation des segments**

La deuxième étape est la validation de ces droites, elle permettra d'éliminer les cas de fausses détections, notamment les écritures.

En effet, il n'y a pas de recherche d'intersection et de sous-segments nécessaire au vu de l'état de la planche. La validation se fait comme pour le cas des planches anciennes, il faut un nombre de point de contours suffisant. Le résultat est visible en Figure 3-12.

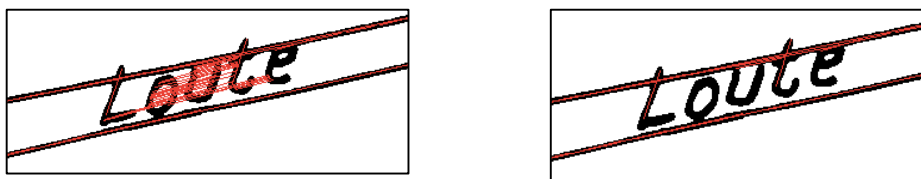


Figure 3-12 : Extraits de la planche 155 avant et après validation des sous segments A gauche, il s'agit de l'extrait de l'image après la transformée de Hough (étape 1) et à droite, l'extrait de l'image après suppression des segments non valides avec un seuil de 80% (étape 2).

Pratiquement toutes les droites ont été supprimées sauf celles qui sont à l'intérieur de la typographie.

- **Connexion entre sous-segments validés**

La troisième étape permet de connecter les extrémités des segments à condition que le seuil soit respecté. Cela permet d'assurer la continuité dans les courbes et dans les cas où l'épaisseur de la ligne change. Le résultat est visible en Figure 3-13.

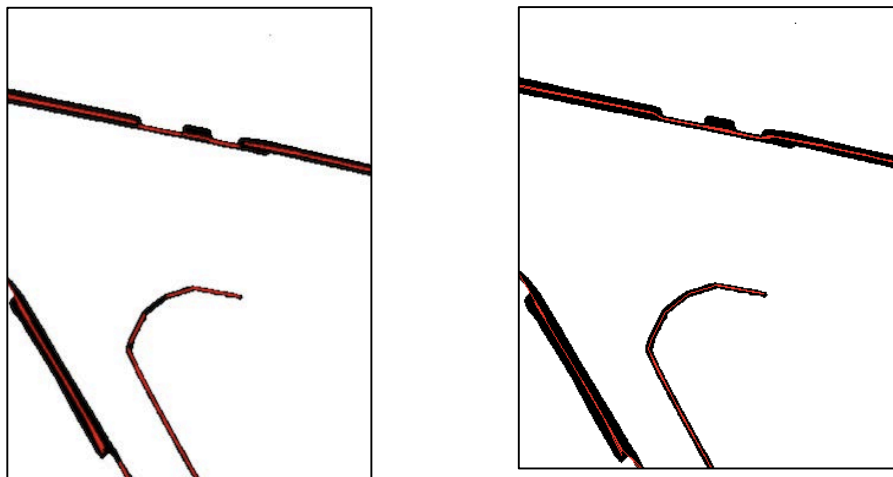


Figure 3-13 : Extraits de la planche 155 avant et après connexion des sous segments A gauche, il s'agit de l'extrait de l'image après validation des sous segments (étape 2) et à droite, l'extrait de l'image après jonction des sous segment avec un seuil de 90% (étape 3).

3.2.2.c Conversion au format vectoriel

La dernière étape est l'enregistrement des sous-segments valides en format Shapefile. Il se fait grâce à la fonction *Writing Shapefiles* de la bibliothèque *pyshp*. Cette étape est commune aux deux époques.

3.2.3 Seconde méthode : Vectorisation par « suivi de chemin »

3.2.3.a Dans le cas des données anciennes (1813 et 1850)

L'image en entrée est l'image binaire issue du pré-traitement. Pour l'importer dans le projet GRASS, il faut utiliser la commande :

```
r.in.gdal input="../../../008_binaire.png" output=008_import
```

Ensuite, afin de délimiter la région à la taille du raster en entrée, la commande *g.region* est utilisée.

Afin de pouvoir appliquer les outils suivants, il faut re-catégoriser les valeurs de l'image.

```
g.region rast= 008_import
```

Dans le cas de l'image en entrée, il s'agit d'une image binaire donc la valeur des pixels est de 0 ou 255 (dans le cas des contours). Le fichier en format txt précise donc que les pixels de valeur 255 doivent prendre la valeur 0, et les pixels de valeur 0 doivent prendre une valeur nulle :

```
255 = 0
```

```
* = NULL
```

```
r.reclass input=008_import output=008_reclass rules="../../../regles.txt"
```

L'étape suivante consiste à squelettiser l'image, l'outil *r.thin* sera utilisé. Le nombre d'itérations maximum par défaut est de 50 et c'est largement suffisant.

```
r.thin input=008_reclass output=008_thin iterations=50
```

Une fois ces étapes effectuées, la vectorisation de l'image peut avoir lieu. Le but est d'obtenir des polygones : ainsi dans la fonction *r.to.vect*, le type à choisir est « line ».

```
r.to.vect input=008_thin output=008_vect type=line
```

3.2.3.b Dans le cas des données récentes (1972)

Il faut également importer l'image, effectuer une re-catégorisation et la vectoriser. La squelettisation a été faite précédemment dans le script python, il n'est donc pas nécessaire de la refaire. Globalement les étapes sont les mêmes pour les planches anciennes et plus récentes.

Contrairement à la méthode précédente, il n'y a pas de paramètre à préciser ce qui rend ce traitement plus rapide et plus robuste.

3.2.4 Etapes finales communes aux deux méthodes

3.2.4.a Nettoyage topologique des vecteurs

A l'aide d'outils de nettoyage topologique des vecteurs, il est possible d'améliorer le résultat obtenu.

Lorsque les segments obtenus sont en « escalier », il est possible de les lisser avec l'outil *v.generalize*. Les contours en escalier sont généralement obtenus après la vectorisation d'une image raster par l'outil *r.to.vect*. Le résultat est illustré en Figure 3-14.

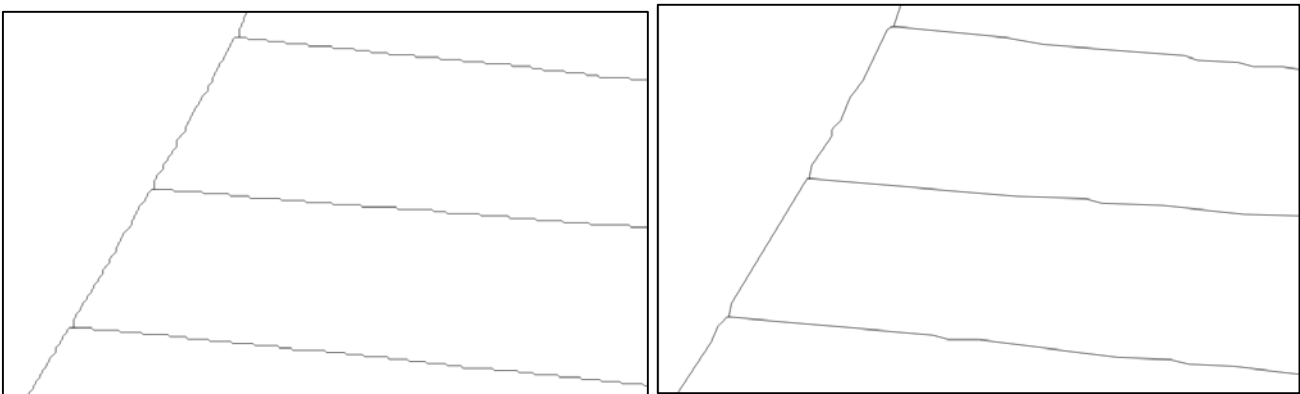


Figure 3-14 : A gauche, les segments avant lissage et à droite, les segments après lissage.

Lorsque qu'il reste des segments détectés sur les écritures qui doivent être supprimés, l'outil *rmdangle* de la fonction *v.clean* permet de supprimer les nœuds pendants de longueur inférieure à un seuil. Les nœuds pendants sont le résultat des lignes vectorielles qui doivent se connecter les unes aux autres mais qui ne se touchent pas ou alors dans le cas où une ligne se termine au-delà de la ligne avec laquelle elle doit être connectée. D'un point de vue du nombre d'entités l'application de cet outil permet de passer de 2827 entités à 2138. Le résultat est illustré en Figure 3-15.

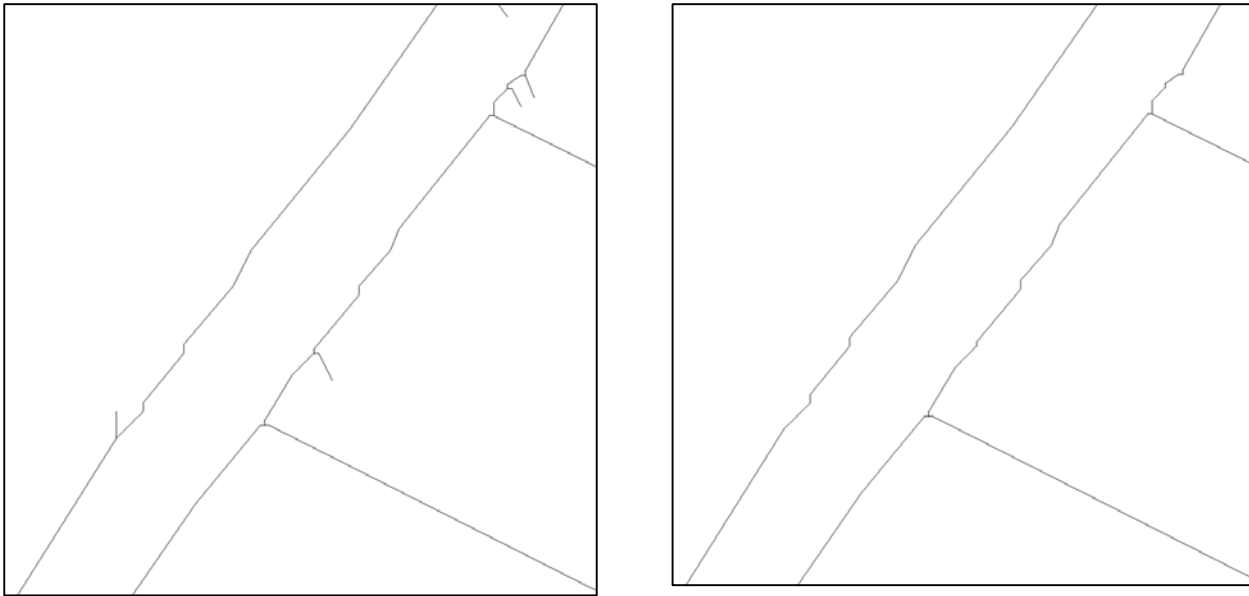


Figure 3-15 : A gauche, les segments avant suppression des nœuds pendants, à droite les segments après suppression.

Lorsque qu'il y a un nombre important de segments détectés qui ne sont pas continus pour un même contour, l'outil *prune* de *v.clean* peut être utilisé. Il simplifie les lignes en supprimant les sommets en fonction d'un seuil. Cet outil préserve la topologie de la zone. Le résultat est illustré en Figure 3-16.

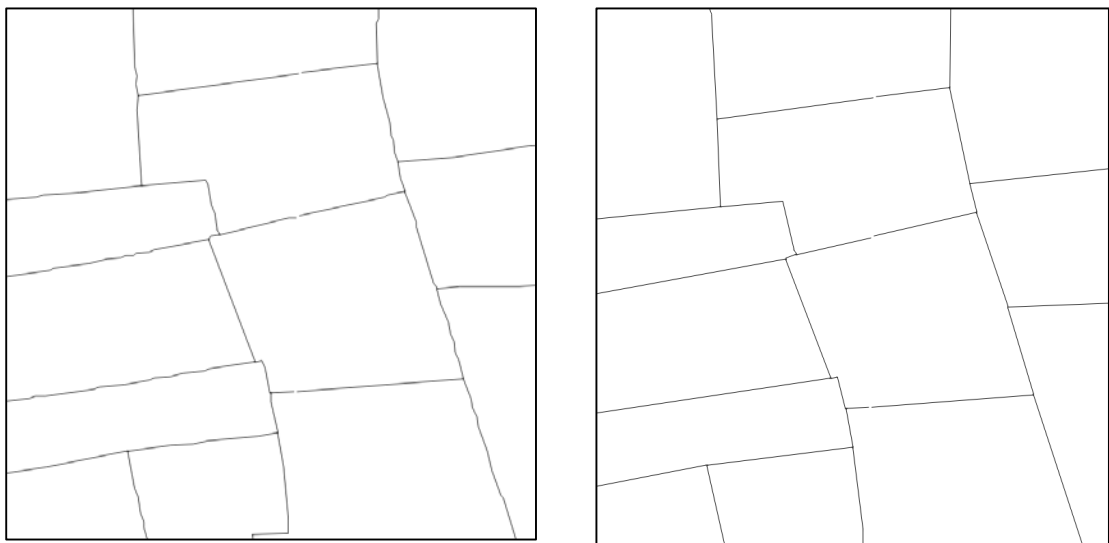


Figure 3-16 : A gauche, les segments avant amélioration de la continuité, à droites les segments après amélioration de la continuité.

Lorsque deux sommets ne sont pas accrochés l'un à l'autre, il est possible d'utiliser l'outil *snap* de *v.clean*. Il accroche un sommet à un autre si la distance est inférieure à un seuil, mais si le seuil est trop important, cela peut endommager la topologie de la zone. Cet outil permet de passer de 2138 entités à 741, ce qui prouve son intérêt. Le résultat est illustré en Figure 3-17.

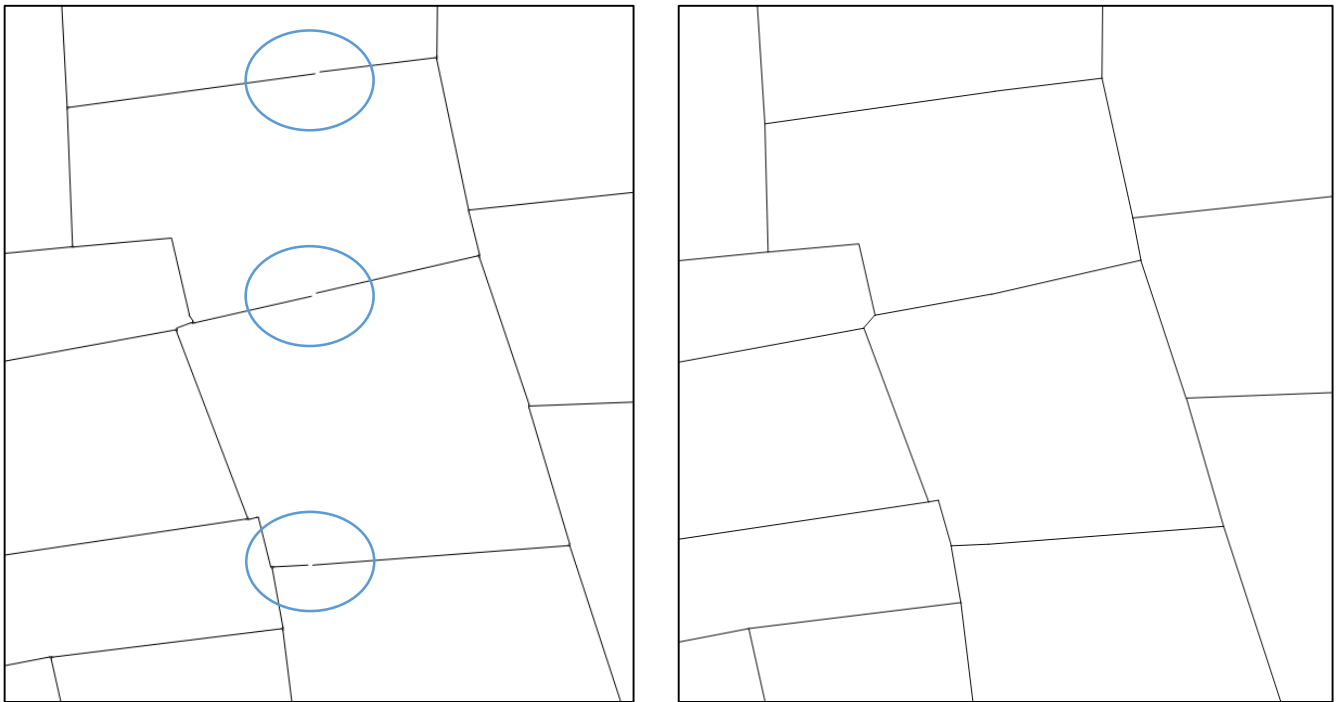


Figure 3-17 : A gauche, les segments avant connexion des extrémités, à droites les segments après connexion des extrémités.

Afin de supprimer les petites surfaces de taille inférieure à un seuil donné et les remplacer par une seule limite, l'outil *rmaea* de la fonction *v.clean* est utilisé. En effet, la limite commune la plus longue avec une zone contiguë plus grande est enlevée ainsi la petite surface est fusionnée avec la zone contiguë.

Selon la méthode (transformée de Hough probabiliste ou extraction de lignes du logiciel GRASS), l'étape de nettoyage est différente. En effet, davantage d'étapes sont nécessaires dans le cas de la seconde méthode.

De plus, pour la méthode de la TH probabiliste, la première étape est d'utiliser l'outil *snap* de la fonction *v.clean* afin de rattacher entre eux tous les segments. Alors que pour la méthode d'extraction de ligne du logiciel GRASS, il faut réaliser un lissage puis un *rmdangle*.

3.2.4.b Transformation des polygones en objets surfaciques

Les différentes étapes sont illustrées en Figure 3-18. A la fin de cette étape, nous obtiendrons des polygones ou polygones fermés représentant les parcelles en format vectoriel SHP.

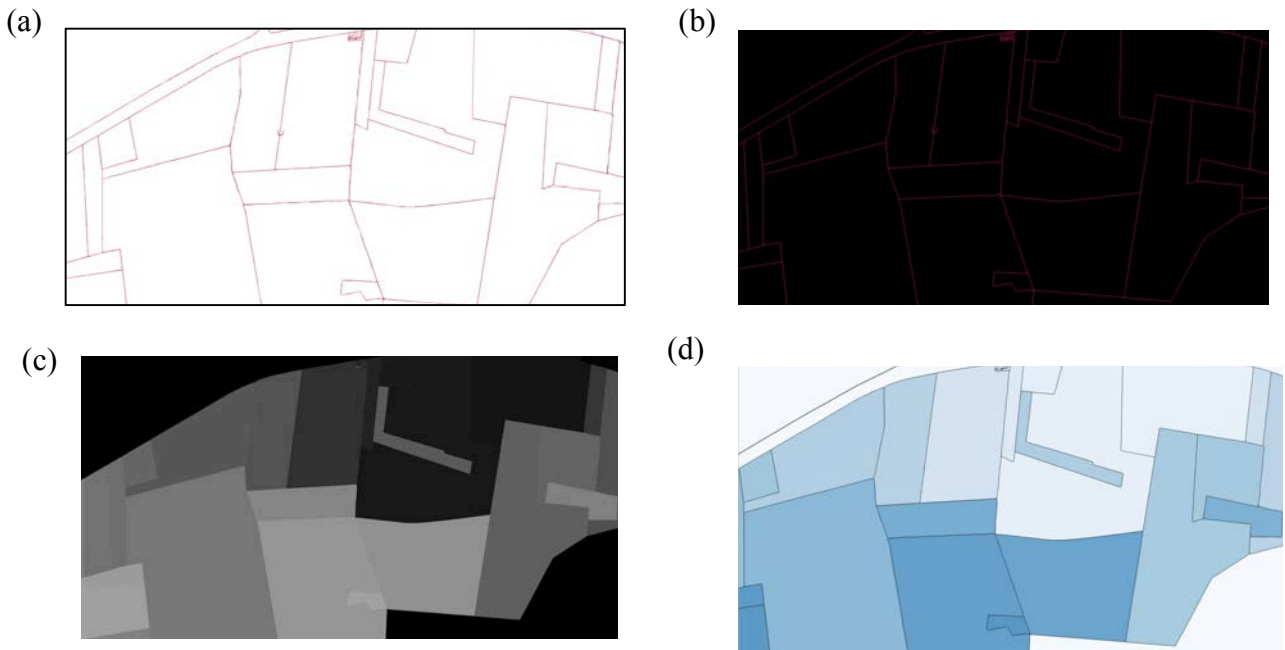


Figure 3-18: Les différentes étapes pour obtenir des polygones fermés en format vectoriel à partir de lignes en format vectoriel. (a) : les vecteurs issus du nettoyage topologique, (b) : l'image de ces vecteurs après rasterisation, (c) : l'image après labellisation et (d) : couche vectorielles après vectorisation des régions.

- **Conversion en format image**

La rasterisation des données vectorielles se fait grâce à l'outil *v.to.r*

```
v.to.r input=008_vect_clean output=008_rast
```

Ensuite, il faut exporter le raster du projet GRASS.

```
r.out.gdal input=008_rast output= ".../008_rast.tif"
```

- **Labellisation des régions**

Cette étape permet d'obtenir des régions à partir des ensembles de polygones correspondant à des contours fermés.

- **Vectorisation des régions**

Cette étape permet de vectoriser le raster obtenu à l'étape précédente.

```
r.to.vect input=008_label output=008_vect_area type=area
```

Il faut choisir le type *area* puisqu'il s'agit de polygone.

La vectorisation peut créer des contours en escalier qui peuvent être lissés à l'aide de l'outil *v.generalize*.

3.3 Validation et résultats

3.3.1 Dans le cas des linaires

A partir des données linéaires rastérisées obtenues dans la partie précédente, nous pouvons valider des premiers résultats.

En effet, ces trois images peuvent être comparées avec les images correspondant aux planches 008, 090 et 155 qui ont été vectorisées manuellement, puis rastérisées.

A partir d'un script Python, deux probabilités ont été calculées, avec « référence », l'image correspondant à la planche vectorisée manuellement et « img_rast », l'image rastérisée obtenue avec le nettoyage topologique des vecteurs pour les deux méthodes. Une zone tampon a été appliquée à la référence afin d'obtenir des résultats qui ne sont pas biaisés, elle a été effectuée grâce à une dilatation (voir section 3.1.1) avec un ES de 3x3, cela a été fait avec la fonction *dilation* du module *morphology* de la bibliothèque scikit-image. Les résultats sont dans le Tableau 3-7.

$$\text{Probabilité de détection} = \frac{\text{Nombre de points de contour dans référence et dans img_rast}}{\text{Nombre de points de contour dans référence}}$$

$$\text{Probabilité de fausses alarmes} = \frac{\text{Nombre de points de contour dans img_rast et pas dans référence}}{\text{Nombre de points de contour dans img_rast}}$$

Tableau 3-7 : Comparaison pour les trois planches et pour les deux méthodes des probabilités de détection et de fausses alarmes.

Année de la planche	Probabilité	1813 (008)	1850 (090)	1972 (155)
Méthode TH probabiliste	Détection	60%	93%	88%
	Fausses alarmes	50%	6%	13%
Vectorisation par « suivi de chemin »	Détection	55%	82%	90%
	Fausses alarmes	59%	16%	10%

Concernant la probabilité de fausses alarmes, on constate que pour les deux méthodes et pour les époques que 1850 et 1972, la valeur est très correcte. Cela est dû au nettoyage de la topologie qui est effectué précédemment et à la qualité des planches. En effet, nous avons remarqué précédemment que pour ces deux planches, les limites de parcelles étaient bien distinguées, le tracé était relativement net et il n'y avait pas de trous dans les contours, particulièrement pour la planche de 1972.

Néanmoins, pour la planche datant de 1813, les résultats sont justes supérieurs à la moitié, cela est dû à l'état de la planche d'origine. En effet, les points de contours ne sont pas tous détectés.

Dans le cas de la probabilité de détection, les valeurs sont également correctes pour les planches de 1850 et 1972. On remarque que pour la planche de 1972, les valeurs obtenues pour les deux méthodes sont proches. En effet, l'image obtenue par le scannage de la planche de 1972 était binaire, ce qui a facilité l'application des deux méthodes. Alors que pour la planche de 1850, il y a une différence légère selon la méthode utilisée. Lors de l'étape du seuillage, certains points de contour n'ont pas été détectés, cela est en partie dû au carroyage.

Ensuite, ces trous n'ont pas pu être comblés avec la méthode de vectorisation par « suivi de chemin » alors que la méthode de la TH probabiliste a permis de rapprocher les extrémités, mais sans combler parfaitement les trous. Ceci peut expliquer les différents résultats.

Comme précédemment, les valeurs concernant la planche la plus ancienne sont mauvaises. Cela est expliqué par le nombre important de cas de fausses détections, notamment les écritures. Ces écritures n'ont pas pu être toutes supprimées lors de l'étape de suppression d'objets isolée lors du prétraitement. En effet, il y avait un risque de supprimer des points de contour corrects isolés puisque les points de contour ne sont pas tous connectés dans le cas de cette planche.

D'autres résultats peuvent être obtenus à partir des polygones en format vectoriel SHP, en effet, ces trois couches SHP peuvent être comparées avec les images correspondant aux planches 008, 090 et 155 qui ont été vectorisées manuellement.

3.3.2 Dans le cas des polygones

A partir des polygones représentant les objets « parcelles » obtenus en format vectoriel à la dernière étape, il est possible de valider certains résultats. Les résultats de cette partie seront liés aux résultats précédents, en effet, les polygones sont créés à partir du linéaire. Ce n'est alors pas surprenant d'obtenir des résultats très corrects dans le cas des planches de 1850 et de 1972.

- Cas des planches de 1850 et de 1972 :

Pour les deux méthodes, la surface de chaque polygone obtenu a été comparée avec la surface du polygone référence correspondant. Pour cela, à l'aide des utilitaires du logiciel QGIS, nous avons compté le nombre de polygones pour lesquels la surface est à 15% près égale à celle de la référence correspondante, il s'agira de la probabilité de détection. Il y a également le cas où la méthode utilisée ne détecte qu'un seul polygone alors qu'en réalité, il y en a deux, cela est lié au fait qu'il y a un trou dans les limites détectées, il s'agira de la probabilité de fausses alarmes. Les résultats sont dans le Tableau 3-8.

Tableau 3-8 : Comparaison pour deux planches et pour les deux méthodes des probabilités de détection et de fausses alarmes.

Année de la planche	Probabilité	1850 (090) 106 parcelles	1972 (155) 54 parcelles
Méthode TH probabiliste	Détection	92%	96%
	Fausses alarmes	8%	4%
Vectorisation par « suivi de chemin »	Détection	93%	100%
	Fausses alarmes	7%	0%

Les résultats sont légèrement moins bons dans le cas de la planche de 1850 : en effet, certaines limites non délimitées (BND, voir section 3.1.2) sont représentées en pointillés et elles n'ont pas pu toutes être détectées comme des lignes continues où chaque extrémité est connectée à une autre limite. Cette difficulté a été présente pour les deux méthodes, ce qui explique que les résultats sont très proches.

Dans le cas de la planche de 1972, les résultats sont très corrects et pour les deux méthodes. L'absence de trous dans les contours explique ces résultats presque parfaits.

- Cas de la planche de 1813 :

Pour cette méthode la probabilité de détection a été calculée de la même façon que précédemment. Les nombreux trous dans les limites détectées font qu'il y a très peu de polygones corrects, cela dans le cas des deux méthodes. Les résultats sont dans le Tableau 3-9.

Tableau 3-9 : Comparaison pour la planche de 1813 et pour les deux méthodes des probabilités de détection.

Année de la planche	Probabilité	1813 (008) 66 parcelles
Méthode TH probabiliste	Détection	10%
Vectorisation par « suivi de chemin »	Détection	11%

Le nombre peu élevé de polygones corrects ainsi que les probabilités de fausses alarmes élevées dans le cas des linaires (Tableau 3-7) expliquent les nombreux cas de fausses alarmes dans cas des polygones. En effet, comme c'est le cas des deux autres planches, la méthode utilisée ne détecte qu'un seul polygone alors qu'en réalité il y en a plusieurs. De plus, des polygones sont formés avec des polylignes qui ne sont pas des limites de parcelles, il s'agit de cas de fausse détection linéaire.

Les résultats tels quels ne sont alors pas exploitables, il est nécessaire de réaliser un nettoyage manuel.

3.3.3 Correction manuelle des vecteurs

Cette correction peut être réalisée sur toutes les planches mais elle est indispensable dans le cas de la planche de 1813. Ce nettoyage peut être réalisé sur les polylignes avant l'étape de transformation en polygones ou directement sur les polygones. Dans le cas présent, la correction a été réalisée sur les polylignes. Il est donc intéressant de comparer le temps nécessaire pour une vectorisation entièrement manuelle au temps nécessaire pour une vectorisation semi-automatique (Tableau 3-10)

Tableau 3-10 : Pour les trois planches, durée nécessaire pour une vectorisation semi-automatique et une vectorisation manuelle

Etapas	Planche de 1813			Planche de 1850			Planche de 1972		
	Méthode TH probabiliste	Vectorisation par « suivi de chemin »	Vectorisation manuelle	Méthode TH probabiliste	Vectorisation par « suivi de chemin »	Vectorisation manuelle	Méthode TH probabiliste	Vectorisation par « suivi de chemin »	Vectorisation manuelle
Temps total	34'	24'	1h15'	17'	12'	50'	13'	10'	30'

Nous pouvons constater que la durée nécessaire aux différents traitements est liée à l'anciennement des planches. Une vectorisation semi-automatique avec une correction manuelle prendra moins de temps qu'une vectorisation manuelle, ce qui démontre l'intérêt d'une procédure semi-automatique.

Conclusion

De nombreuses publications traitent de la vectorisation de cartes anciennes, et ce, en France ou à l'étranger, il y a également eu des travaux sur la vectorisation de réseaux routiers. Cependant, l'objectif est le même, il s'agit de développer une méthodologie complète de vectorisation qui s'adapte aux données en entrée. Cette méthodologie permettra d'obtenir des données vectorielles afin d'analyser les changements spatiaux et temporels d'un territoire.

Dans le cadre de ce TFE, nous avons mis en place une chaîne méthodologique pour la vectorisation de planches scannées du cadastre ancien. Les différentes étapes se feront à l'aide de programmes et logiciels informatiques afin d'automatiser la chaîne. Nous avons privilégié l'utilisation de logiciels Open Source en raison de leur maturité et pour pouvoir maîtriser les outils créés.

Les données à disposition sont des planches cadastrales scannées, datant pour les plus anciennes de 1813 et pour les plus récentes de 1970. Il y a donc des différences importantes entre les données, qui ont été prises en compte dans la création de la chaîne méthodologique. En effet, deux méthodologies ont été mises en place en fonction de l'ancienneté de la planche, de plus, pour chaque méthodologie, deux méthodes ont été testées puis comparées.

Les résultats obtenus sont cohérents puisque la planche la plus ancienne offre les résultats les moins satisfaisants, et ce, pour les deux méthodes testées. Alors que les résultats concernant les planches de 1850 et 1972 sont très corrects, et ce, pour les deux méthodes également. Les résultats obtenus sont sensiblement les mêmes dans le cas des deux méthodes. La méthode de vectorisation par « suivi de chemin » ne nécessite aucun paramètre, elle est donc jugée autonome et est plus facile d'utilisation. Alors que la méthode par extraction de segments permet des applications plus ciblées grâce aux choix des différents paramètres.

Perspectives

La chaîne méthodologique mise en place pour les données anciennes peut être améliorée, en effet, dans le cas de la planche de 1813, les résultats ne sont pas satisfaisants. Une autre méthode de détection de segments peut être envisagée, il s'agit de l'outil LSD (Line Segment Detector). Cette méthode ne nécessite aucun paramètre et contrôle elle-même son nombre de fausses détections. (Grompone von Gioi *et al*, 2012)

Concernant les planches de 1850, les limites approximatives des biens non délimités (BND) sont représentées en pointillés sur les planches, ainsi, les limites n'ont pas pu être détectées. Il faudrait alors envisager de réaliser deux transformées de Hough probabilistes, une première avec les paramètres définis précédemment afin de détecter les contours. Une seconde avec des paramètres plus larges afin de détecter les limites pointillées, la difficulté est ensuite de déterminer les segments « nouveaux » à conserver.

Ce TFE traite de la vectorisation, selon la chaîne méthodologique globale du projet du GeF, les étapes suivantes sont le géoréférencement et le mosaïquage. L'étape du géoréférencement a été traitée dans le cadre d'un TFE en 2016, il reste alors l'étape du mosaïquage.

Cette étape doit permettre d'assurer une continuité et une cohérence dans le plan entre les couches vectorielles géoréférencées. Le mosaïquage des vecteurs peut s'appuyer sur un traitement topologique des données. Dans un premier temps, il faut corriger les erreurs topologiques apparaissant après le géoréférencement au niveau des contacts entre chaque planche (trous, chevauchements) pour s'assurer de la continuité de la donnée. Ensuite, il faut éliminer les chevauchements pour s'assurer de la continuité de la donnée. Ces erreurs topologiques peuvent être détectées et traitées à l'aide de logiciels de SIG. L'outil vérificateur de topologie de QGIS permet de détecter les erreurs. Ensuite, les outils de GRASS peuvent les corriger, notamment les outils *v.clean* et *v.build* utilisés dans le cadre de ce TFE.

Enfin, dans la perspective d'une application de la chaîne dans sa totalité, il est nécessaire de modifier les dernières étapes de la chaîne de vectorisation. En effet, dans la chaîne de traitement actuelle, lors de l'étape de la transformation des polylignes en polygones, seuls les objets « parcelles » sont labélisés puis vectorisés. Les linéaires qui représentent les bords extérieurs des routes ne sont pas conservés, néanmoins ils sont nécessaires pour le recalage des données.

Bibliographie

Article de revue scientifique et ouvrage :

- Baily B, (2007). The extraction of digital vector data from historic land use maps of Great Britain using image processing techniques. *e_Perimetron*, vol 2 (n°4), pp209-223.
- Berteau G, Clergeot P, (2009). Du cadastre napoléonien au cadastre en ligne sur Internet. *Revue XYZ n°119*, pp.49-59.
- Cao R, Tan C.L, (2002). Text/Graphics Separation in Maps. In: Blostein D., Kwon YB. (eds) Graphics Recognition Algorithms and Applications. GREC 2001. Lecture Notes in Computer Science, vol 2390. Springer, Berlin, Heidelberg, pp167-177.
- Chiang YY, Knoblock C.A, Chen CC, (2005) Automatic Extraction of Road Intersections from Raster Maps. In: 13th ACM International Symposium on Advances in Geographic Information Systems (ACMGIS'05), Bremen, Germany, November 2005. USA, New York: ACM, pp 267-276.
- Chiang YY., Knoblock C.A. A Method for Automatically Extracting Road Layers from Raster Maps. In: 10th International Conference on Document Analysis and Recognition (ICDAR), Barcelona, 26–29 July 2009, pp. 838–842.
- Chiang YY, Knoblock C.A, (2010) Extracting Road Vector Data from Raster Maps. In: Ogier JM., Liu W., Lladós J. (eds) Graphics Recognition, Achievements, Challenges, and Evolution. GREC 2009. Lecture Notes in Computer Science, vol 6020. Springer, Berlin, Heidelberg, pp93-105.
- Chiang YY, Knoblock C.A, (2013) A general approach for extracting road vector data from raster maps. *IJDAR*, n°16, No.1, pp 55-81.
- Fahrasmane M, (2016). Géoréférencement de cadastres anciens vectorisés au moyen de logiciels libres : Application sur les communes de Vaas et Aubigne-Racan (Sarthe). Mémoire de soutenance de DIPLÔME D'INGÉNIEUR CNAM Spécialité : Géomètre Topographe, ESGT, 2016, pp.
- Fiorio C, Gustedt J, (1996). Two linear time Union-Find strategies for image processing. *Theoretical Computer Science* n°154, pp. 165-181.
- Follin J.-M., Fahrasmane M., Simonetto E., (2016). An open-source based toolchain for the georeferencing of old cadastral maps. Actes de OGRS2016 Open Source Geospatial Research & Education Symposium, Pérouse (Italie), 12-14 octobre 2016.
- Galamhos C, Matas J, Kittler J, (1999). Progressive probabilistic Hough transform for line detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- Gregory I.N. et Ell P.S, (2007). Historical GIS: Structuring, mapping and analysis of urban social structures. Cambridge studies in Historical Geography, Cambridge: Cambridge University Press.

Grompone von Gioi R, Jakubowicz J, Morel JM, Randall G, (2012). LSD: a Line Segment Detector. *Image Processing On line (IPOL)*, pp 35-55.

Herrault P-A, (2015). Extractions de fragments forestiers et caractérisation de leurs évolutions spatio-temporelles pour évaluer l'effet de l'histoire sur la biodiversité : une approche multi-sources. Mémoire de thèse, école doctorale ED TESC : Géographie et aménagement, Université Toulouse le Mirail – Toulouse II, 2015, pp. 139-166.

Iosifescu I, Tsorlini A, Hurni L, (2016). Towards a comprehensive methodology for automatic vectorization of raster historical maps. *E-Perimtron*, [en ligne], n°11. Disponible sur : <http://www.e-perimtron.org/Vol_11_2/Iosifescu_et_al.pdf>, consulté le 03/03/2017.

Jang B.K, Chin R.T, (1990). Analysis of thinning algorithms using mathematical morphology, *IEEE transactions on pattern analysis and machine intelligence*, vol. 12 (n°6), pp. 541-551.

Labrousse E., Lambert P., Thouzeau V., Follin J.-M., Fournier M., Bonnefond M., (2015). Évolution du parcellaire cadastral à l'échelle communale : propositions méthodologiques et retour d'expérience à partir des cas des communes d'Aubigné-Racan et de Vaas (Sarthe) de 1813 à nos jours. Laboratoire GeF, document interne.

Macario M, (2008). Aménagement foncier rural. Site de la Direction technique Territoires et ville/ Cerema, rubrique Aménagement Construction Urbanisme Opérationnel, consulté en ligne le 29/05/2016. ([Http://www.outil2amenagement.certu.developpement-durable.gouv.fr/amenagement-foncier-rurala541.html](http://www.outil2amenagement.certu.developpement-durable.gouv.fr/amenagement-foncier-rurala541.html)).

Pierrot-Deseilligny M, H. Le_Men, G. Stamon, (1995). Lecture automatique des écritures sur cartes scannées. *Traitement du Signal*, vol. 12 (n° 6), pp.639-652.

Trias-Sanz R, Pierrot-Deseilligny M, Louchet J, Stamon G, (2007). Methods for Fine Registration of Cadastre Graphs to Images, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 29, pp. 1990-2000.

Rodallec D, (2015). Présentation générale du cadastre. Fascicule de cours réédité en 2015 par le service de reprographie de l'ESGT.

Viglino J-M, Pierrot-Deseilligny M, (2003) A vector approach for automatic interpretation of the French cadastral map. In: 7th International Conference on Document Analysis and Recognition (ICDAR), Edinburgh, 3-6 August 2003.

Zhang T.Y, Suen C.Y, (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM*, vol 27, pp 236-239.

Publication institutionnelle et document juridique

Article 6 de la loi n°74-645 du 18 juillet 1974 sur la mise à jour périodique de valeurs locatives servant de base aux impositions directes locales, page 07539 du JO du 19/04/1974, consulté le 26/05/17

CAD – Descriptif, usage et diffusion de la documentation cadastrale – Documentation cadastrale (BOI-CAD DIFF-10-20121224) consulté sur le site officiel du Bulletin Officiel des Finances Publiques-Impôts (bofip.impots.gouv.fr) le 26/05/2017.

CAD – Informatisation du plan cadastral - Les conventions de vectorisation - Phase administrative (phase projet) (BOI-CAD-INFO-10-10-20140203) consulté sur le site officiel du Bulletin Officiel des Finances Publiques-Impôts (bofip.impots.gouv.fr) le 24/05/2017

Site internet

ISIMA. Python et traitement d'images, [en ligne]. Disponible sur : <http://www.isima.fr/~vbarra/IMG/pdf/traitement_Images.pdf>. (Consulté le 10/03/2017)

GéoPeople. Protocole de vectorisation manuelle sous Qgis, [en ligne]. Disponible sur <http://geopeople.ign.fr/doc/rapports/L2.0-1_Protocole_de_vectorisation_manuelle_sous_Qgis.pdf>. (Consulté le 20/03/2017)

ESRI France. Projet Alpage : du plan Vasserot au SIG participatif, [en ligne]. Disponible sur <http://www.esrifrance.fr/sig2008/univ_larochelle.htm>. (Consulté le 13/02/2017).

Descriptif du contenu de la BD PARCELLAIRE® V1.2, IGN, juin 2012, pp 3-4, [en ligne] Disponible sur <http://professionnels.ign.fr/sites/default/files/DC_BDPARCELLAIRE_1_2.pdf>. (Consulté le 07/06/2017)

Site de la bibliothèque de traitement des images « scikit-image », [en ligne]. Disponible sur <<http://scikit-image.org>>.

Bernard S, (2015). Extraction de segments pour la reconnaissance de symboles : Une approche robuste par Transformée de Hough, disponible sur <<http://slideplayer.fr/slide/3511777/>>. (Consulté le 02/06/2017).

RONSE C, (2013). Opérations morphologiques de base : dilatation, érosion, ouverture et fermeture binaires, disponible sur <<https://dpt-info.u-strasbg.fr/~cronse/TIDOC/MM/deof.html>>. (Consulté le 15/06/2017).

Tables des figures

Figure 1-1 : Extraits de plans cadastraux napoléoniens de 1813 (à gauche) et 1850 (à droite) sur la commune d'Aubigné-Racan, lieu-dit de la Motte de Cœur (source : Archives Départementales de la Sarthe).....	10
Figure 1-2 : Terminologie concernant les opérations cadastrales de rénovation des plans (source : Rodallec D, 2016).	11
Figure 1-3 : Principe d'enchaînement des étapes de la chaîne méthodologique du laboratoire du GeF, source : Fahrasmane, 2016.	15
Figure 2-1 : Méthodologie pour les planches anciennes (1813 et 1850).	19
Figure 2-2 : Méthodologie pour les planches plus récentes (1972).	20
Figure 2-3 : A gauche, c'est une figure X, à savoir un ensemble de pixels et à droite, c'est l'élément structurant B, où la boule noire indique la position du point de référence	23
Figure 2-4 : A gauche, les fenêtres Bp, obtenue en positionnant B sur p, pour un certain nombre de pixels p appartenant à X. A droite, la dilatation de X par B.	24
Figure 2-5 : À gauche, plusieurs positions (boules noires) de pixels p dont les fenêtres Bp sont incluses dans X. À droite, l'érosion de X par B.	24
Figure 2-6 : Méthodologie d'extraction d'un segment (extrait de Bernard, 2015). longueur d'intervalle. Alors qu'à droite, elles sont de 30 et de 30, de plus les contours détectés apparaissent en rouge.	25
Figure 3-1 : Extrait d'une planche datant de 1813.	31
Figure 3-2 : Extrait d'une planche datant de 1850.	32
Figure 3-3 : Extrait d'une planche datant de 1972.	34
Figure 3-4 : Histogramme d'une planche datant de 1813.	35
Figure 3-5 : Images seuillées par la méthode Hystérésis locale (à gauche) et par la méthode de la moyenne glissante (à droite)	37
Figure 3-6 : Image squelettisée (à droite) et image après suppression d'objets isolés (à gauche). ..	40
Figure 3-7 : Extraits de la planche 008 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 30 et de 15 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 50 et de 15.....	41
Figure 3-8 : Extraits de la planche 008 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 30 et de 15 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 30 et de 30.....	42
Figure 3-9 : Extrait de la planche 008 avant et après validation des sous segments avec un seuil de 80%.	42
Figure 3-10 : Extraits de la planche 008 avant et après connexion des sous-segments A gauche, il s'agit de l'extrait de l'image après validation des sous-segments (étape 3) et à droite, l'extrait de l'image après jonction des sous-segment avec un seuil de 90%.	43
Figure 3-11 : Extraits de la planche 155 avec les segments détectés par TH probabiliste (en rouge). A gauche, les valeurs des paramètres sont de 20 et de 20 respectivement pour la longueur de ligne et la longueur d'intervalle. Alors qu'à droite, elles sont de 40 et de 20.....	44
Figure 3-12 : Extraits de la planche 155 avant et après validation des sous segments A gauche, il s'agit de l'extrait de l'image après la transformée de Hough (étape 1) et à droite, l'extrait de l'image après suppression des segments non valides avec un seuil de 80% (étapes 2).	44
Figure 3-13 : Extraits de la planche 155 avant et après connexion des sous segments A gauche, il s'agit de l'extrait de l'image après validation des sous segments (étape 2) et à droite, l'extrait de l'image après jonction des sous segment avec un seuil de 90% (étape 3).	45
Figure 3-14 : A gauche, les segments avant lissage et à droite, les segments après lissage.	46
Figure 3-15 : A gauche, les segments avant suppression des nœuds pendant, à droites les segments après suppression.	47
Figure 3-16 : A gauche, les segments avant amélioration de la continuité, à droites les segments après amélioration de la continuité.	47
Figure 3-17 : A gauche, les segments avant connexion des extrémités, à droites les segments après connexion des extrémités.	48

Figure 3-18: Les différentes étapes pour obtenir des polygones fermés en format vectoriel à partir de lignes en format vectoriel. (a) : les vecteurs issus du nettoyage topologique, (b) : l'image de ces vecteurs après rasterisation, (c) : l'image après labellisation et (d) : couche vectorielles après vectorisation des régions. 49

Liste des tableaux

Tableau 3-1 : Propriétés des images scannées des plans de 1813 d'après Fahrasmane, 2016	32
Tableau 3-2 : Propriétés des images scannées des plans de 1850 d'après Fahrasmane, 2016.	33
Tableau 3-3 : Propriétés des images scannées des plans de 1972-1974 d'après Fahrasmane, 2016...34	
Tableau 3-4 : Comparaison de la méthode de la moyenne glissante et Hystérésis locale	38
Tableau 3-5 : Probabilités de détection et de fausses alarmes des deux méthodes avec variation de la valeur de ou des constantes de 0,1	38
Tableau 3-6 : Probabilités de détection et de fausses alarmes des deux méthodes avec variation importante des valeurs de ou des constantes.	39
Tableau 3-7 : Comparaison pour les trois planches et pour les deux méthodes des probabilités de détection et de fausses alarmes.	50
Tableau 3-8 : Comparaison pour deux planches et pour les deux méthodes des probabilités de détection et de fausses alarmes.	51
Tableau 3-9 : Comparaison pour la planche de 1813 et pour les deux méthodes des probabilités de détection.	52
Tableau 3-10 : Pour les trois planches, durée nécessaire pour une vectorisation semi-automatique et une vectorisation manuelle.....	52

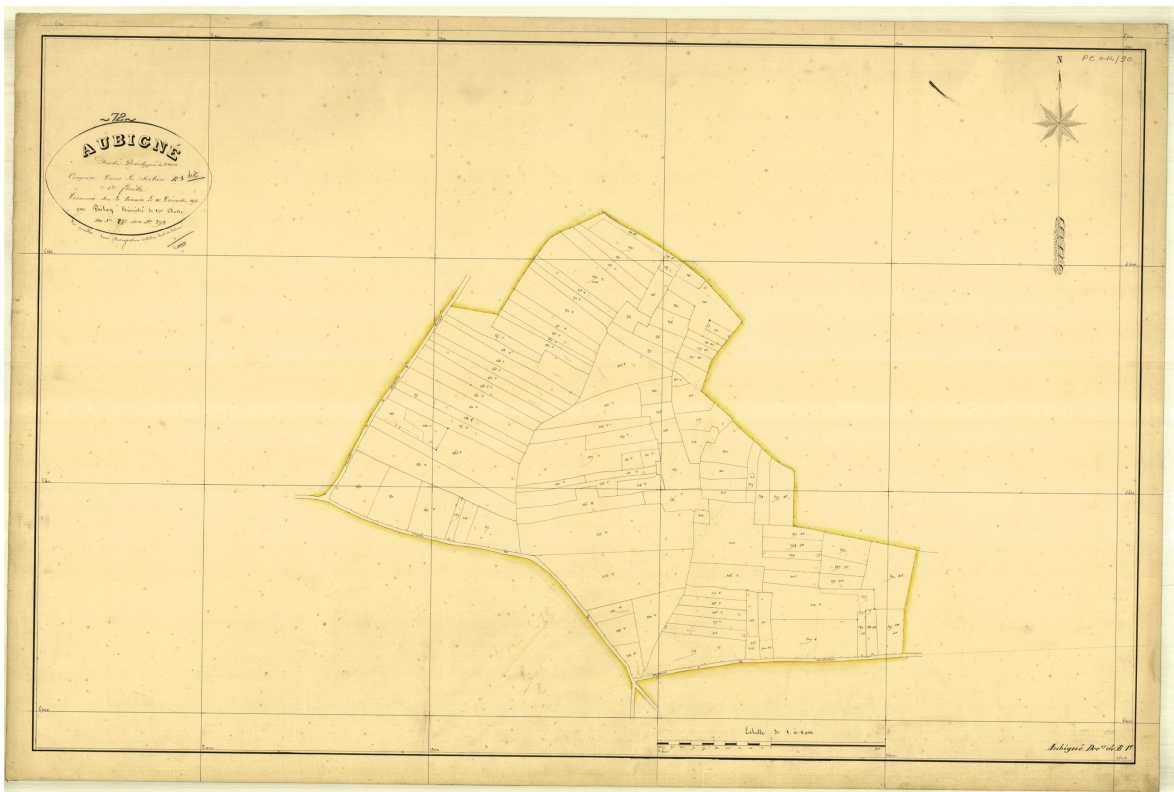
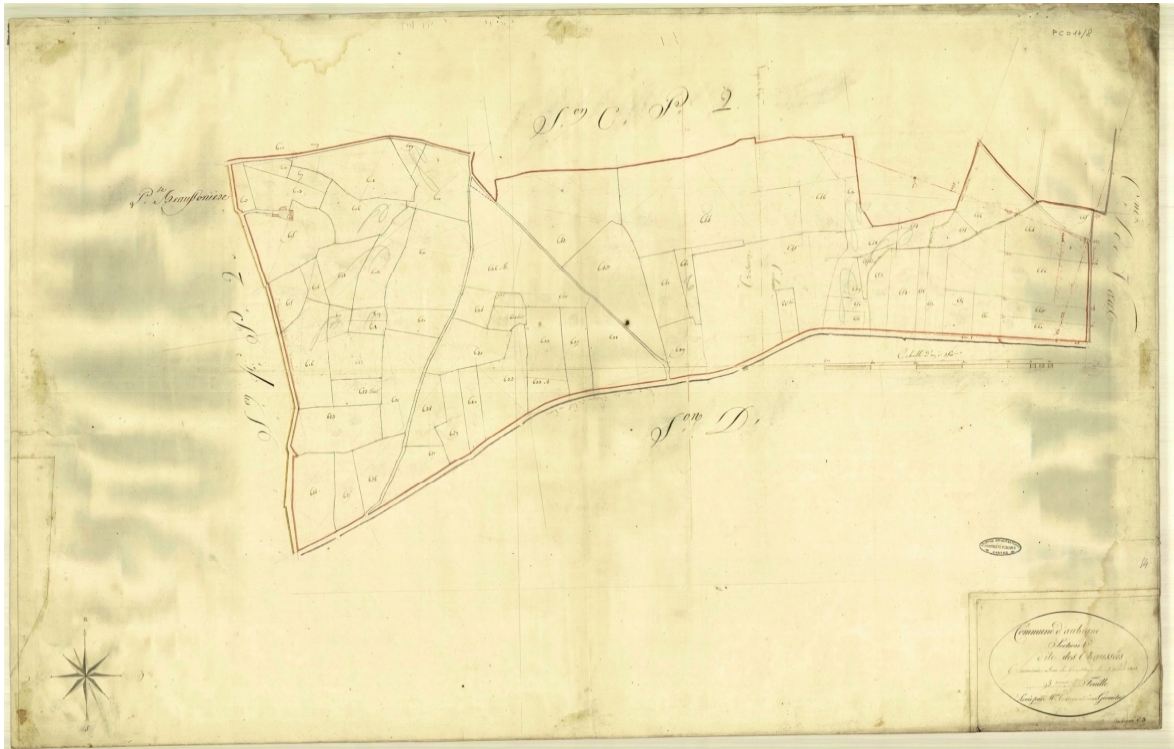
Tables des annexes

Annexe 1 : Images utilisées en entrée dans l'application des différentes méthodes

Annexe 2 : Résultats obtenus après application des différentes méthodes

Annexe 3 : La Transformée de Hough

Annexe 1 : Images utilisées en entrée dans l'application des différentes méthodes





Annexe 2 : Résultats obtenus après application des différentes méthodes

Image convertie en niveaux de gris avec la formule de la luminance par défaut



Image convertie en niveaux de gris avec la formule de la luminance adaptée



Image seuillée par méthode du seuillage simple avec conversion en niveaux de gris par la formule de la luminance adaptée.



Image seuillée par méthode hystérésis avec conversion en niveaux de gris par la formule de la luminance adaptée.



Image seuillée par méthode de la moyenne glissante avec conversion en niveaux de gris par la formule de la luminance par défaut



Image seuillée par méthode de la moyenne glissante avec conversion en niveaux de gris par la formule de la luminance adaptée



Image seuillée par méthode Hystérésis locale avec conversion en niveaux de gris par la formule de la luminance adaptée.



Annexe 3 : Transformée de Hough

La Transformée de Hough (TH) est introduite en 1962 qui permet de détecter des objets paramétrés dans une image (droite, cercle, parabole).

Ici, l'objectif est de détecter des morceaux de droites, sachant qu'une droite est décrite par l'équation : $y = ax + b$

Avec $a = \frac{y_2 - y_1}{x_2 - x_1}$ la pente de la droite, et $b = y_1 - ax_1$ l'ordonnée du point d'intersection de cette droite sur l'axe y .

Le principe de la méthode est de passer de l'espace image 2D à l'espace de Hough 2D (2D parce qu'une droite est caractérisée par deux paramètres, a et b).

Ainsi, une droite dans l'espace image (x, y) sera transformée en un point dans l'espace de Hough (a, b) . Figure 1.

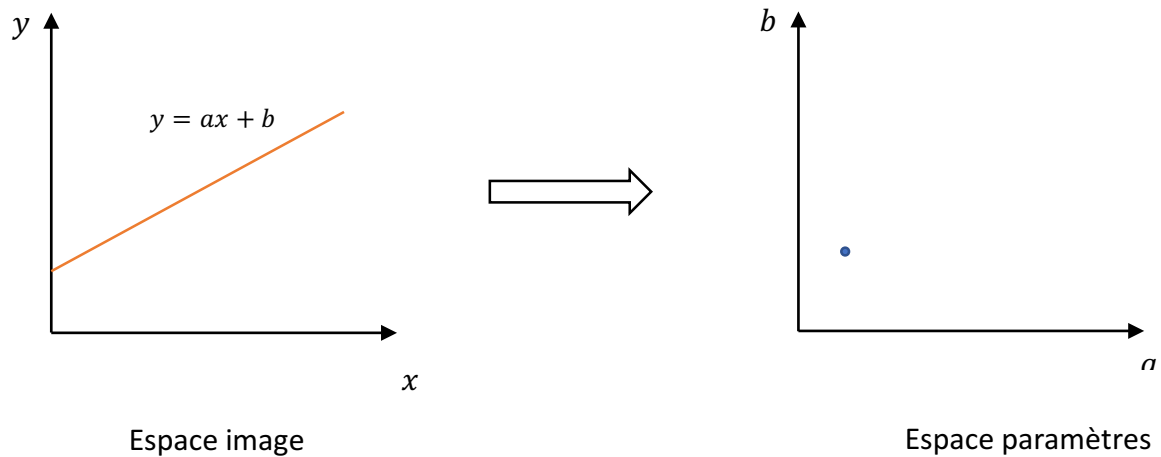


Figure 1 : Principe de détection d'une droite.

Et par analogie, un point dans l'espace image sera transformé en une droite d'équation $b = -ax + y$ dans l'espace paramètres.

Ensuite, pour des points situés sur une même droite, ils sont représentés par des droites qui se coupent en (a, b) dans l'espace de Hough. (Figure 2).

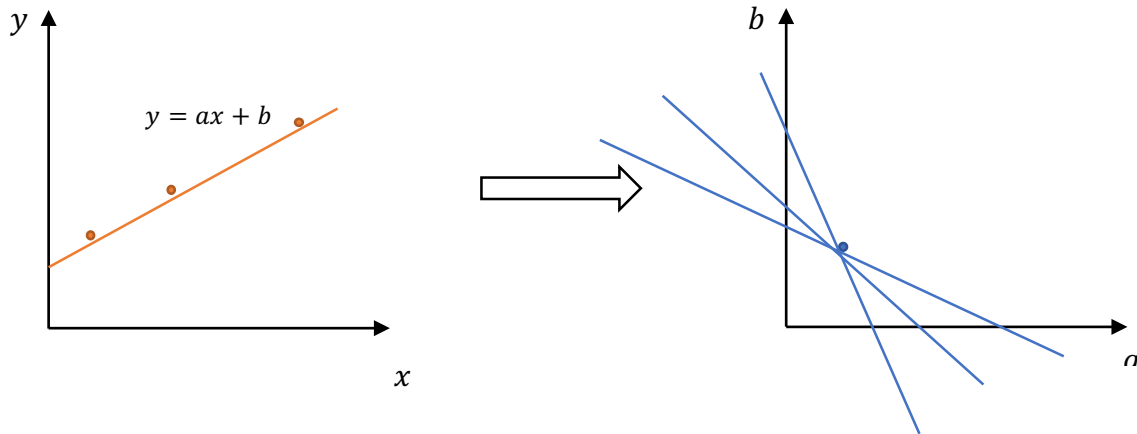


Figure 2 : Principe de détection de plusieurs points sur une droite.

Néanmoins, la représentation $y = ax + b$ pose quelques problèmes. En effet, l'espace des paramètres doit être borné, ainsi, une droite verticale ne peut pas être représentée ($a = \infty$). C'est pourquoi, une représentation polaire (r, θ) est envisagée.

$$x = r \times \cos \theta$$

$$y = r \times \sin \theta$$

$$r = x \times \cos \theta + y \times \sin \theta$$

Avec r la distance entre l'origine et la ligne et θ l'angle entre l'axe x et le vecteur r .

L'espace paramétré (r, θ) est donc borné par :

$r \in [-d, d]$ avec d la taille de la diagonale de l'image

$\theta \in [-\pi/2, \pi/2]$

La mise en œuvre consiste à comptabiliser le nombre de pixels contribuant à une droite (de l'espace image) dans l'espace des paramètres. Cela constituera une nappe d'accumulation, où un pic indique la probable présence d'une droite.

Plus précisément, la méthode consiste à⁴:

1. Appliquer une détection de points de contour
2. Initialiser un accumulateur
3. Pour chaque point de contour :
 - a. Déterminer sa droite image dans l'espace des paramètres
 - b. Incrémenter l'accumulateur sur les points de cette droite
4. Rechercher des maxima \rightarrow paramètres de la droite

⁴ Tiré de la présentation de GRAF B et ESPIC C « Transformée de Hough » (<http://elynxsdk.free.fr/ext-docs/Demosaicing/more/news1/article-graf-esplic.pdf>) consultée le 2/05/17

Résumé

Les premiers plans cadastraux remontent aux années 1810. Ils ont fait et font encore l'objet de mises à jour, ce qui fait du cadastre une source d'informations pour étudier l'évolution d'un territoire. Les plans cadastraux anciens peuvent être scannés et rendus disponibles en format raster, néanmoins ce format n'est pas suffisant pour réaliser une analyse complète des changements spatiaux d'un territoire au cours du temps. Un format vectoriel est beaucoup plus adapté. L'opération qui permet de passer des données rasters aux données vectorielles est la vectorisation. De nombreuses publications traitent de la vectorisation automatique et semi-automatique et les méthodes diffèrent selon les données de base. Une attention toute particulière est donc apportée à l'étape du prétraitement, préalable à la vectorisation. Dans ce travail, nous proposons une chaîne méthodologique pour vectoriser de façon semi-automatique des plans cadastraux scannés anciens à l'aide de logiciels Open Source. Une première méthode est basée sur l'extraction des segments alors que la seconde est basée sur la vectorisation par « suivi de chemin ». Les résultats obtenus doivent subir un nettoyage manuel, plus ou moins conséquent selon l'ancienneté de la planche traitée.

Mots clé : Cadastre Ancien, Vectorisation, Données Raster, Prétraitement, Traitement Numérique Des Images, Automatisation, Transformée De Hough Probabiliste, SIG, GRASS, Logiciel Open Source.

Abstract

The first cadastral maps date back to the 1810s, they have been and are still being updated, making the cadastre a source of information to study the evolution of a territory. Historical cadastral maps can be scanned and made available in raster format, but this format is not sufficient to carry out a complete analysis of the spatial changes of a territory over time. A vector format would be much more suitable. The operation that allows to pass from raster to vector data is the vectorization. Numerous publications deal with automatic and semi-automatic vectorization and the methods differ according to the input data. Particular attention is therefore paid to the pre-processing step prior to the vectorization. In this work, we propose a methodological toolchain for semi-automatic vectorization of old scanned cadastral maps using free and open source software. A first method is based on the extraction of the segments whereas the second is based on the vectorization by tracking. The obtained results must undergo a manual cleaning, more or less substantial according to the age of the cadastral map.

Keywords: Historical Cadastral Map, Vectorization, Raster Data, Preprocessing, Digital Image Processing, Automation, Probabilistic Hough Transform, GIS, GRASS, Free and Open Source Software.