
Vers une semi-automatisation du processus d'intégration de plan cadastral ancien dans une base de données multi-dates

Jean-Michel Follin¹, Elisabeth Simonetto¹

1. Laboratoire GeF, Ecole Supérieure des Géomètres et Topographes, le Cnam
1 boulevard Pythagore, 72000 Le Mans, France
prenom.nom@lecnam.net

RESUME. Les recherches en géomatique s'intéressant au patrimoine géohistorique connaissent un essor depuis plusieurs années comme en attestent les nombreux projets et les conférences et groupes de travail dédiés. Elles se nourrissent de la disponibilité d'un nombre croissant de documents anciens géographiques. Le cadastre napoléonien, diffusé par les archives départementales, fournit en particulier la description la plus détaillée du territoire français dans sa globalité durant le 19^{ème} siècle. Grâce à cette donnée l'évolution du parcellaire cadastral sur une période de deux siècles peut être étudiée en lien avec les transformations affectant le territoire. Dans cette perspective les travaux présentés ici portent sur une chaîne de traitement semi-automatique permettant de vectoriser les planches scannées de cadastre ancien, de les géoréférencer, de les assembler puis de les intégrer dans une base de données multi-époques. Dans cet atelier, nous abordons les solutions adoptées plus particulièrement pour les étapes de vectorisation et de mosaïquage avec une application sur les données de trois époques (1813, 1850 et 1972-74) d'une commune rurale du sud de la Sarthe pour laquelle les premiers résultats obtenus seront présentés.

ABSTRACT. Studies of geo-historical heritage have recently picked up momentum, as evidenced by the many projects, conferences and working groups dedicated to it. They thrive on an increased availability of ancient geographical documents. For instance, in France, the Napoleonic cadastre, archived in the "departmental" libraries, provides the most accurate representation of France's whole territory during the 19th century. It is thus now possible to study two centuries of cadastral parcels evolution driven by territorial transformations. This article details a semi-automatic processing chain tailored to vectorize the old scanned cadaster sheets, geo-reference them, assemble them, and finally feed them into a historical database. We focus especially here on the two steps of vectorization and tiling, using a three date dataset (1813, 1850, 1972-74) covering one rural municipality located in the south of the Sarthe "departement". First results will be presented and discussed.

MOTS-CLÉS : Cadastre ancien, vectorisation, mosaïquage de données vectorielles

KEYWORDS: Old cadastre, vectorization, tiling of vector data

1. Introduction

Actuellement, de plus en plus de documents anciens, en particulier géographiques (cartes, plans, clichés aériens), sont dématérialisés et mis à disposition sous format numérique. Ils constituent une source d'information précieuse sur le territoire, notamment pour les chercheurs en sciences humaines. Dans de nombreux départements français, les services des archives diffusent ainsi leurs fonds documentaires numérisés en ligne dont notamment les plans cadastraux anciens scannés. Or, selon la période à laquelle ils ont été dressés, ces derniers contiennent des informations très variées, et qui sont plus ou moins détaillées ou précises (Clergeot P. et Bertheau G., 2008).

Afin d'étudier l'évolution du territoire à travers les changements opérant au cours du temps dans la forme de son parcellaire cadastral, nous proposons une chaîne méthodologique reproductible, semi-automatique et basée sur des outils libres comprenant une première étape de préparation des données (vectorisation, géoréférencement et mosaïquage de feuilles cadastrales anciennes) et une deuxième étape d'intégration des informations dans une base de données spatio-temporelle modélisant l'évolution des parcelles au cours du temps qui sera utilisée pour mener des analyses (quantification et qualification des changements, mise en relation avec d'autres facteurs, identification de motifs récurrents, ...)

Nous allons broser un état de l'art succinct sur les mises en œuvre de bases de données cartographiques multi-dates avant de faire une présentation globale de notre démarche et des difficultés rencontrées.

2. État de l'art

Les études menées sur des données cartographiques anciennes se sont considérablement développées ces dernières années en France et dans le monde. Elles démontrent l'intérêt récent pour les bases de données géographiques historiques et leur utilisation à des fins analytiques ou prospectives. En France, nous pouvons notamment citer les projets de recherche GéoPeuple (Costes *et al.*, 2012), GeoHistoricalData (Cura *et al.*, 2018), ALPAGE (Noizet et Grosso, 2012), MODE RESPYR (Herrault *et al.*, 2013) et ModelSpace / Architerre (Le Couédic *et al.*, 2012). Le développement de chaînes de traitements semi-automatiques des plans anciens est un sujet très actif comme en témoignent par exemple les travaux de Iosifescu *et al.* (2016) et Arteaga (2013) qui emploient plusieurs outils (tels que GDAL et ImageMagick), nécessitent des paramétrages manuels lors du prétraitement et peuvent faire appel à une démarche collaborative dans l'étape de validation ("Building Inspector" de NYPL).

Ces différents projets ne s'intéressent pas nécessairement au même type de document et ne partagent pas notre objectif de semi-automatisation du processus complet depuis l'image jusqu'à la base de données historiques.

3. De la vectorisation au mosaïquage : propositions méthodologiques

3.1. La chaîne de traitement semi-automatique

La première étape est la vectorisation semi-automatique des images correspondant aux planches cadastrales anciennes. Deux méthodes ont ici été évaluées : l'une basée sur la transformée de Hough dite probabiliste, proposée dans la bibliothèque scikit-image en langage Python, et l'autre sur une méthode dite du « suivi de chemin » du logiciel GRASS. La deuxième étape est le géoréférencement dont l'objectif est de pouvoir rendre superposables, avec une qualité optimale, le parcellaire vectorisé ancien au parcellaire numérique récent. Elle est précédée par une procédure manuelle de sélection des points de liaison qui servent au calcul d'un modèle géométrique de transformation des coordonnées. En reprenant la méthodologie de (Herrault *et al.*, 2013) la fonction « ridge par noyau gaussien » s'est avérée être la meilleure parmi celles évaluées à l'aide de l'erreur moyenne quadratique pour les plans les plus anciens (Follin *et al.*, 2016). La troisième étape est le mosaïquage par traitement topologique des feuilles cadastrales vectorisées et géoréférencées qui a pour finalité de contrôler la cohérence topologique de l'assemblage. En effet en sortie du traitement de géoréférencement des défauts topologiques (chevauchement, trou) entre les parcelles voisines des différentes feuilles cadastrales sont constatés. La quatrième étape est la création de la base de données multi-dates qui permet de modéliser les changements d'état (division, fusion) au cours du temps et constitue le support des analyses menées sur l'évolution du parcellaire cadastral.

3.2. Application

Nous avons appliqué notre méthode sur des feuilles du cadastre de la commune d'Aubigné-Racan (Sarthe) couvrant trois époques : 1813, 1850 et 1972-74. Les données de 1813, et dans une moindre mesure celles de 1850, sont difficiles à traiter en raison du jaunissement du papier lié au temps et de nombreuses traces laissées par l'homme qui compliquent la vectorisation automatique. Les planches numérisées de 1972-74 se présentent sous forme d'images binaires en noir et blanc qu'il est plus aisé de traiter (figure 1). Les caractéristiques sont résumées dans la table 1.



FIGURE 1. Extraits de feuilles cadastrales d'Aubigné-Racan (1813, 1951 et 1972) couvrant une zone d'environ 300 mètres sur 400 mètres

TABLE 1. *Propriétés des feuilles scannées d'Aubigné-Racan*

Année	1813	1850	1972-1974
Échelle	1/2500	1/2000	1/2000
Effectifs de planches	22	45	44
Pas du scanner	200 dpi	200 dpi	400 dpi
Taille d'un pixel dans l'image (µm)	127	127	63,5
Taille d'un pixel sur le terrain (cm)	31,8	25,4	12,7

Dans cet atelier, nous présenterons en détail les performances et les limites des méthodes proposées pour les étapes de vectorisation et de mosaïquage.

Concernant la vectorisation, les deux méthodes évaluées visent à extraire des segments de contour des parcelles qui seront converties en polygones dans un deuxième temps. Elles se basent sur des étapes identiques pour le prétraitement (conversion en niveaux de gris de valeur 0 ou 255, application d'un masque dessiné manuellement et détection de points de contour avec un seuillage par hystérésis à seuils adaptatifs localement suivi d'une suppression des objets isolés puis d'une squelettisation) et le post-traitement (simplification et nettoyage topologique des lignes puis transformation de celles-ci en objets surfaciques après un passage en mode raster pour procéder à un étiquetage des composantes connexes). Le seuillage par hystérésis à seuils adaptatifs localement qui hybride le seuillage simple par moyennes glissantes et celui par hystérésis permet d'obtenir un prétraitement robuste en limitant sa sensibilité aux variations de paramètres.

L'évaluation des deux méthodes - basée sur la transformée de Hough probabiliste (THP) et dite de « suivi de chemin » (SdC) - pour chaque époque a porté sur les résultats obtenus avant les étapes de transformation des polygones en objets surfaciques (table 2). Elle a consisté à comparer les linéaires obtenus à ce stade avec des données de référence, saisies manuellement, par superposition.

TABLE 2. *Comparaison pour trois planches et pour les deux méthodes des probabilités de détection et de fausses alarmes obtenues pour les données linéaires*

Méthode	Probabilité	Planche de 1813	Planche de 1850	Planche de 1972
THP	Détection	60%	93%	88%
	Fausses alarmes	50%	6%	13%
SdC	Détection	55%	82%	90%
	Fausses alarmes	59%	16%	10%

Les erreurs en sortie de la vectorisation, qui restent importantes pour les plans les plus anciens, nécessitant des interventions manuelles. Néanmoins le temps global de traitement (processus automatique et correction manuelle) est inférieur au temps nécessaire à une vectorisation entièrement manuelle du plan (table 3).

TABLE 3. *Durée nécessaire pour une vectorisation semi-automatique (THP, SdC) et une vectorisation manuelle (M) calculée pour chacune des planches*

	Planche de 1813			Planche de 1850			Planche de 1972		
Méthode	THP	SdC	M	THP	SdC	M	THP	SdC	M
Temps total	34'	24'	1h15'	17'	12'	50'	13'	10'	30'

Les deux méthodes évaluées offrent donc des résultats de qualités sensiblement équivalentes mais la deuxième présente l'avantage d'être un peu plus rapide, de ne nécessiter aucun paramètre et est donc jugée plus autonome et simple d'utilisation.

L'étape du mosaïquage vise à corriger les erreurs observées entre des parcelles voisines appartenant à des planches différentes après les étapes de vectorisation et de géoréférencement. La feuille cadastrale décrit des espaces cadastrés (les parcelles sur lesquelles sont situés des bâtiments) et non cadastrés (la voirie, les cours d'eau), ces derniers permettant de délimiter les îlots (définis comme des regroupements de parcelles contiguës). Notre méthode de mosaïquage permet de réaliser deux corrections : suppression des chevauchements et comblement des trous entre les parcelles des différentes planches qui devraient être contiguës. Elle permet également de construire des polygones modélisant les espaces non cadastrés (ENC) qui doivent être distingués des trous qui correspondent à des défauts topologiques. Cette étape s'appuie sur des géotraitement tels que l'union, la fusion de polygones et le calcul d'enveloppe convexe par triangulation de Delaunay. Elle requiert un traitement des triangles superflus, coupant les concavités du contour.

Si elle permet de traiter en grande partie les problèmes rencontrés à l'issue de l'étape précédente, nous avons constaté quelques erreurs en sortie de ce processus (duplications de géométries ou création d'artefacts) qui sont liées à l'étape de vectorisation. Par ailleurs elle ne traite pas tous les problèmes topologiques que l'on pourrait rencontrer (par exemple des parcelles en contact alors qu'elles sont normalement séparées par un ENC).

4. Conclusion et perspectives

Nous avons conçu une chaîne méthodologique semi-automatique basée sur des outils open-source pour la vectorisation, le géoréférencement et le mosaïquage du cadastre ancien. Elle permet d'obtenir les entités pour construire une base de données modélisant les parcelles ainsi que les espaces non cadastrés. Notre méthode a été appliquée sur des planches scannées d'une commune rurale du sud de la Sarthe sur lesquelles de bons résultats ont été obtenus avec un temps d'intervention manuelle limité. Cependant il reste des points à améliorer qui peuvent concerner l'automatisation du processus ou la qualité des résultats. Nous pouvons citer notamment, pour la vectorisation semi-automatique, l'application d'une transformée de Hough non probabiliste. Plus globalement une modélisation par graphe d'adjacence granulaire qui permet de considérer différents niveaux détails pourrait

améliorer l'ensemble du processus : contrôler et corriger les résultats de la vectorisation ou guider les corrections à apporter lors du mosaïquage.

Remerciements

Nous tenons à remercier les anciens étudiants de l'esgt stagiaires du laboratoire GeF, Maïté Fahrasmane, Charlotte Odie et Jean-Marc Beveraggi qui ont fait avancer ce projet, les collègues Marie Fournier et Mathieu Bonnefond ainsi que Gilles Berteau de la DGFIP.

Bibliographie

Arteaga, M. G. (2013). Historical map polygon and feature extractor. In Proceedings of the 1st ACM SIGSPATIAL Int. Workshop on MapInteraction, pp. 66-71.

Clergeot P. et Berteau G. (2008). Du cadastre napoléonien au cadastre en ligne sur Internet : 1^{ère} partie. Revue XYZ n° 119, pp 49-59.

Costes B., Grosso E. et Plumejeaud C. (2012). Géoréférencement et appariement de données issues des cartes de Cassini – Intégration dans un référentiel topographique actuel. Actes de SAGEO 2012, Liège, Belgique.

Cura, R., Dumenieu, B., Abadie, N., Costes, B., Perret, J., and Gribaudi, M. (2018). Historical collaborative geocoding. ISPRS Int. J. of Geo-Information, 7(7), pp. 262-290.

Follin J.-M., Fahrasmane M., Simonetto E., (2016). An open-source based toolchain for the georeferencing of old cadastral maps. Actes OGRS2016, Pérouse, Italie.

Herrault P-A., Sheeren D., Fauvel M., Monteil C., Paegelow M., (2013). A comparative study of geometric transformation models for the historical 'Map of France' registration. Geographia Technica n° 1, pp. 34 - 46.

Iosifescu I, Tsorlini A, Hurni L, (2016). Towards a comprehensive methodology for automatic vectorization of raster historical maps. e-Perimtron vol. 11, n°2, pp. 57-76.

Le Couédic M., Leturcq S., Rodier X., Hautefeuille F., Fieux E. et Jouve B. (2012). Du cadastre ancien au graphe. Les dynamiques spatiales fiscales et modernes. Revue ArchéoSciences 36. Presses universitaires de Rennes. pp 71-84.

Noizet H. et Grosso E. (2012). Mesurer la ville : Paris de l'actuel au Moyen âge. Les apports du système d'information géographique d'ALPAGE. Revue du comité français de cartographie, n°211, pp. 85-100