

# Romain RAVEAUX

## 1 ÉTAT CIVIL, GÉNÉRALITÉS

---

Nom : RAVEAUX

Prénoms : Romain

Situation : Célibataire

Date de naissance : 02/08/1981

Lieu de Naissance : Lisieux (France)

Nationalité : Française

- Coordonnées :

*Adresse personnelle* :(à privilégier)

Appt A010

27, rue enrico fermi

17000 La Rochelle

portable: 06 66 85 40 73 (à privilégier)

page web : <http://romain.raveaux.free.fr>

*Adresse professionnelle* :

L3i, Bat Pascal, Université La Rochelle

Avenue M. Crépeau

17042 La Rochelle cedex 01

tél : 05 46 45 82 62

email : [romain.raveaux01@univ-lr.fr](mailto:romain.raveaux01@univ-lr.fr)

- Situation actuelle :

**A.T.E.R. à l'IUT d'informatique de La Rochelle (Institut Universitaire de Technologie)**

- Enseignement (détail page 4) :

Architecture des matériels, systèmes d'exploitation, bases de données (TD, TP), réseaux (cours, TD, TP), statistiques (cours, TD), vision par ordinateur (TD, TP), encadrement de projets. Pour un volume total de 236H équivalent TD.

- Domaine de recherche (détail page 10) :

code 22 : Graphes, combinatoire, complexité

code 92 : Vision par ordinateur, reconnaissance de formes

code 91 : Traitement et analyse d'images, de son, de signaux

(codes fournis par la nomenclature thématique section 27 du CNU)

- Responsabilités administratives et charges collectives (détail page 18) :

Membre du comité d'organisation du workshop d'audience internationale sur la reconnaissance d'informations graphiques (GREC), 2009.

Participation à la rédaction des rapports d'activité du projet ANR ALPAGE.

Présentation de projets (ALPAGE et NAVIDOMASS) lors de l'évaluation du L3i par le CNRS, 2007.

## 2 CURRICULUM VITAE

---

### 2.1 Cursus

- 2010 : **Attaché Temporaire d’Enseignement et de Recherche (service complet) à l’IUT d’informatique de La Rochelle.**
- 2006-2010 : **Doctorat de l’Université de La Rochelle.**  
Mention : Très honorable avec félicitations.  
Intitulé : *Fouille et classification de graphes :  
Application à l’analyse d’images cadastrales couleurs.*  
Jury :  
M. Josep LLADÓS    Université autonome de Barcelone      *Rapporteur*  
M. Chew Lim TAN    Université de Singapour                      *Rapporteur*  
M<sup>me</sup> Nicole VINCENT                                      Université de Paris 5                          *Rapporteur*  
M. Jean-Marc OGIER                                      Université de La Rochelle                      *Directeur*  
M. Jean-Christophe BURIE                              Université de La Rochelle                      *Encadrant*  
M. Umapada Pal    Institut de la statistique d’Inde              *Examineur*  
M. Salvatore-Antoine TABBONE                      Université de Nancy 2                          *Examineur*  
M. Rolf INGOLD    Université de Fribourg                          *Président*
- 2004-2006 : **Master Génie Informatique, Université de Rouen.**  
Option : Extraction et Indexation de l’information.  
Mémoire : Reconnaissance de symboles à partir de schémas électriques.  
Sous la direction de Sébastien Adam et Pierre Héroux.
- 2004-2006 : **Master Génie Électrique et Informatique Industrielle, Université de Rouen.**  
Option : Réseaux et Télécoms.
- 2003-2004 : **Licence Génie Électrique et Informatique Industrielle, Université de Rouen.**  
Mention : Bien.
- 2002-2003 : **Niveau licence Informatique, Institute of Technology Carlow (Ireland).**  
Option : structures de données.
- 2000-2002 : **DUT Génie Électrique et Informatique Industrielle, IUT de Cherbourg.**  
Option : Réseaux locaux.
- 2000 : **Baccalauréat série Science des Techniques Industrielles ( STI ).**  
Lycée Paul Cornu, Lisieux.  
Mention : Bien.

## 2.2 Formations complémentaires

- Stage “*Outils professionnels et méthodes de travail*”, animé par Anne Marie Pauquet (Training & Coaching, Enseignante associée à Université La Rochelle).
- Stage “*Rédaction de textes scientifiques en anglais*”, animé par Francis Healy (Université de La Rochelle).
- Cours “*Etre soi, de la formation à l’emploi*”, animés par Mme Deloffre-Vye Frédérique (Consultante RH, Croissens Consulting).

## 2.3 Stages et mobilité

- |                  |   |
|------------------|---|
| Mai 2008         | <b>Séjour de trois semaines au laboratoire LITIS, Université de Rouen.</b><br>Invité par Sébastien Adam et Éric Trupin.   |
| Mai 2007         | <b>Séjour de deux semaines au Computer Vision Center (CVC), Université de Barcelone (Espagne).</b><br>Invité par Josep Lladós.  |
| Mai-Octobre 2006 | <b>Stage de fin d’études ALGOTECH informatique.</b><br>Sujet : Implémentation d’algorithmes de reconnaissance des formes au sein d’un logiciel de rétro-conversion d’images de schémas électriques (ScanBuilder). |
| Juin-Août 2005   | <b>Stage de Maîtrise au LITIS, Rouen.</b><br>Sujet : Mesure de similarité entre graphes.  |

### 3 ACTIVITÉS D'ENSEIGNEMENT

---

#### 3.1 Enseignements

Depuis quatre ans que j'assure des enseignements à l'IUT dans les départements d'informatique et réseaux et télécommunications ainsi qu'au pôle science de l'université de La Rochelle, j'ai été amené à intervenir dans diverses spécialités, comme le réseau, la vision par ordinateur ou les statistiques. J'ai aussi encadré divers projets et stages. **Je suis actuellement ATER à l'IUT de La Rochelle et j'interviens au Département Informatique.**

##### Année 2006-2007 :

Intitulé	Public	Nature et durée	Statut
Vision par ordinateur	année 2, Master Info	4H TD	Vacataire,
Vision par ordinateur	année 2, Master Info	12H TP	Vacataire

##### Année 2007-2008 :

Intitulé	Public	Nature et durée	Statut
Image et Ontologie	année 2, Master Info	6H CM	Vacataire
Image et Ontologie	année 2, Master Info	1H30 TD	Vacataire
Image et Ontologie	année 2, Master Info	3H TP	Vacataire
Vision par ordinateur	année 1, Master Info	6H TD	Vacataire
Vision par ordinateur	année 1, Master Info	12H TP	Vacataire

##### Année 2008-2009 :

Intitulé	Public	Nature et durée	Statut
Statistiques	année 1, Master Géomatique	6H CM	Vacataire
Statistiques	année 1, Master Géomatique	27H TD	Vacataire
Réseaux (IPv6)	année 1, DUT Réseaux et Télécoms	4H30 CM	Vacataire
Réseaux (IP)	année 1, DUT Réseaux et Télécoms	16H TD	Vacataire
Réseaux (IP)	année 1, DUT Réseaux et Télécoms	27H TP	Vacataire
Architecture système	année 1, DUT Informatique	6H TD	ATER
Architecture système	année 1, DUT Informatique	12H TP	ATER
Système d'Exploitation	année 1, DUT Informatique	31H TP	ATER
Système d'Exploitation	année 2, DUT Informatique	21H TP	ATER

### Année 2009-2010 :

Intitulé	Public	Nature et durée	Statut
Réseaux	année 1, DUT Informatique	10H TD	ATER
Réseaux	année 1, DUT Informatique	24H TP	ATER
Réseaux	année 2, DUT Informatique	32H TP	ATER

### Total des heures 2006-2010 :

Nature	Durée	Heure équivalent TD
CM	16H30	24H45
TD	70H30	70H30
TP	211H	140H35
<b>Total</b>		<b>236H</b>

## 3.2 Contenu des enseignements

### 3.2.1 Réseaux

Ce cours est une introduction aux couches basses des réseaux ainsi qu'à la couche application : les architectures de réseaux, le modèle OSI et TCP/IP. Il couvre les aspects suivants :

- Couche application : technologies d'Internet, services HTTP, FTP, DNS, Samba et Telnet.
- Structure matérielle et logicielle d'un réseau d'ordinateurs : topologie, commutation et modèle en couche.
- Couche physique : bande passante, débit, interface physique, transmission numérique et canaux numériques.
- Contrôle de liaison : contrôle d'erreurs, contrôle de flux, multiplexage, concentration et contrôle d'accès au médium
- Couche réseau : IPv6, datagramme, circuit virtuel, routage statique et dynamique, adressage, ponts et passerelles. Mise en place d'un réseau sous UML (User-Mode Linux) : configurations de passerelles, de switches et de routeurs.

Le choix de commencer ce module par la couche application permet de donner une vision concrète des services réseaux aux étudiants avant de comprendre les mécanismes des couches bases qui assurent ces services. Cette démarche est consolidée par l'ajout de notions sur la Supervision des réseaux (Nagios) fournissant ainsi une vision d'ensemble du domaine.

- Les éléments dont j'ai réalisé la rédaction sont disponibles à cette adresse <http://romain.raveaux.free.fr/> → Section "Teaching". Cela concerne les concepts suivants: IPv6, filtrage et translation d'adresses, supervision Nagios.

### 3.2.2 Traitement de données

**Statistiques** Ce cours propose d'aborder l'analyse statistique de grands tableaux de données. Dans la pratique, on dispose souvent de nombreuses variables mesurées sur les différents individus de la population étudiée. Les méthodes d'analyses des données dites "multidimensionnelles" vont permettre de prendre en compte une information importante, en ayant pour objectif d'en apporter une représentation simplifiée et organisée. L'analyse multidimensionnelle permet, en plaçant le regard au niveau des variables, de repérer les combinaisons qui sont les plus structurantes de l'espace étudié. (quel groupe de variables permet de bien décrire les différences entre les unités observées ?) En plaçant le regard au niveau des individus, elle repère les unités élémentaires qui se ressemblent, et permet par là même d'aborder la notion de typologie. Je me suis efforcé de bien énoncer les résultats théoriques (sans les démontrer sur le plan mathématique) et de mettre ces résultats en application sur des cas concrets, puisés dans le domaine de la géographie.

- Élaboration d'un cours magistral sur l'Analyse en Composantes Principales (ACP) et sur l'Analyse Factorielle des Correspondances (AFC).
- Création de deux TD sur l'ACP et l'AFC pour un volume de total de 27H.
- Ce matériel pédagogique est disponible à l'adresse <http://romain.raveaux.free.fr/> → Section "Teaching".

**Vision par ordinateur** Ce module met en lumière les briques fondatrices du traitement d'image et de la reconnaissance des formes par la conception par les étudiants d'un mini-OCR. Les trois phases classiques d'un OCR sont abordées :

- Le pré-traitement : Filtrage pour la suppression du bruit.
- Extraction de caractéristiques pour des images de caractères.
- Classification prédictive fondée sur la théorie bayésienne.

**Image et Ontologie** Ce cours aborde la représentation des connaissances sous le point de vue des ontologies.

- Présentation des ontologies et langages associés (RDF, OWL).
- Relations sémantiques.
- Relations de composition et d'héritage.

Ensuite, j'aborde la question de l'Ontologie des images dans un but de reconnaissance des formes. En Intelligence Artificielle les objets logiques reliés par des propriétés, axiomes et règles. Les ontologies sont exploitées par des moteurs d'inférence pour induire une classification des objets. Enfin, les étudiants mettent en application ces notions dans un cadre pratique (xxH TP) dans le but de catégoriser des régions couleurs par le biais d'outils comme OWL API, Protégé et Pellet.

- Ce TP est disponible à l'adresse <http://romain.raveaux.free.fr/> → Section "Teaching".

### 3.2.3 Systèmes informatiques

**Architecture des matériels** Étude de l'organisation et du fonctionnement des composants d'un ordinateur :

- Codage des informations (Entier, réel, chaîne de caractères)
- Organisation matérielle et logique (adressage mémoire, pointeur)
- Dialogue entre les composants (Bus, protocole)
- Fonctionnement logique interne des composants (Assembleur x86)

**Systèmes d'exploitation** L'objectif de ce module est l'acquisition des connaissances de base pour le système UNIX et la compréhension des mécanismes utilisés dans les scripts shell : utilisation avancée d'emacs, commandes unix, scripts shell (variables d'environnement, bash, processus, fonctions).

**Bases de données** Ce cours s'intéresse aux fondements des systèmes de gestion de base de données (SGBD) par l'introduction des notions suivantes:

- Algèbre relationnelle et la définition des opérateurs du langage algébrique
- Requête en langage algébrique
- Initiation à l'optimisation des requêtes par arbre syntaxique
- Calcul relationnel et langage prédictif à variables n-uplets (LPN)
- SQL : implantation des langages formels

### 3.3 Encadrements de projets et stages

J'ai eu le plaisir de co-encadrer avec Jean-Marc Ogier (Professeur à l'université de La Rochelle) et Sébastien Adam (Maître de conférences à l'université de Rouen) un ensemble de projets et stages. J'ai proposé à plusieurs reprises des projets de traitement d'images et d'analyse de données liés à mon domaine de recherche tout en gardant à l'esprit que ces projets, pluridisciplinaires, sont destinés à faciliter pour l'étudiant l'acquisition d'un " savoir-faire " et d'un " savoir-être " dans une optique professionnelle.

<b>Novembre 2009 :</b>	<i>Niveau :</i>	Stage de fin étude en Master 2 Informatique à l'Université de La Rochelle.
	<i>Encadrement :</i>	Romain Raveaux, Jean-Marc Ogier.
	<i>Sujet :</i>	Fouille de graphes et Classification de graphes : Application au "Symbol Spotting".
	<i>Étudiant :</i>	Nguyen Quoc Toan.
	<i>Durée :</i>	6 mois.

- 2007-2008** : *Niveau* : Travaux d'Étude et de Recherche (TER) en Master 2 Informatique à l'Université de Rouen.  
*Encadrement* : Romain Raveaux, Sébastien Adam.  
*Sujet* : Segmentation par approche contours.  
*Étudiant* : Julien Dubois.  
*Durée* : Tout au long de l'année.
- 2007-2008** : *Niveau* : Travaux d'Étude et de Recherche (TER) en Master 2 Informatique à l'Université de Rouen.  
*Encadrement* : Romain Raveaux, Sébastien Adam.  
*Sujet* : Segmentation d'image couleur, du pré-traitement à l'évaluation  
*Étudiant* : Nicolas Martin.  
*Durée* : Tout au long de l'année.
- 2007-2008** : *Niveau* : Stage de fin d'année Master 2 Informatique à l'Université de La Rochelle.  
*Encadrement* : Romain Raveaux, Jean-Marc Ogier et Josep Lladós.  
*Sujet* : Segmentation couleur de plans cadastraux de Paris - 19<sup>ème</sup> siècle.  
*Étudiant* : Alexandre Hamon.  
*Durée* : Trois mois.
- 2007-2008** : *Niveau* : Projet tutoré en Master 1 Informatique à l'Université de Rouen.  
*Encadrement* : Romain Raveaux, Sébastien Adam.  
*Sujet* : Recherche d'espaces couleurs hybrides.  
*Étudiant* : Mickaël Fainet, Nicolas Martin, Julien Dubois.  
*Durée* : Tout au long de l'année.

### 3.4 Projet d'enseignement

Les évolutions récentes dans le monde de l'informatique (taille et nature des logiciels, etc.) créent une demande croissante de la part de l'industrie aux concepteurs informaticiens. Mon parcours m'a donné beaucoup de recul par rapport à la conception d'un système d'information au sens large du terme, du traitement des données à l'architecture réseaux en passant par les bases de données. Mes objectifs s'orientent d'une part, vers l'enseignement des outils d'analyse de données en parallèle à l'étude des réseaux informatiques et d'autre part, l'élargissement à l'enseignement des systèmes à base d'ontologie (RDF, OWL) pour les systèmes d'information dans le cursus universitaire.



J'ai pris beaucoup de plaisir à effectuer tous les enseignements que j'ai donnés (Architecture réseau, système, traitement des données,...) et je me sens prêt à investir du temps dans d'autres domaines. D'autre part, je souhaite pouvoir faire profiter les étudiants de mes connaissances acquises lors de ma formation polyvalente (pratique et théorique), notamment dans ma spécialisation qui concerne les systèmes ontologiques et l'analyse de données structurées. Mes activités me rendent particulièrement compétent sur les domaines suivants :

- Architectures et protocoles des réseaux (modèle OSI, TCP/IP, UMTS),
- Analyse de données et modélisation de connaissance,
- Programmation et conception (C, C++, Java, UML),
- Systèmes d'exploitation (Unix, Linux),
- Bases de données.

Enfin, je reste bien sûr flexible et cela ne me déplairait pas de devoir préparer et enseigner un cours pour lequel je n'ai pas encore d'expérience.

## 4 ACTIVITÉS DE RECHERCHE

---

### 4.1 Thème de recherche

#### 4.1.1 Contexte

Le cadre de ma thèse s'inscrit dans le projet ALPAGE. AnaLyse diachronique de l'espace urbain PARisien : approche GEomatique : ALPAGE est un programme de recherche, initié en 2006, grâce au soutien de l'Agence nationale de la recherche. Fondé sur l'association de 4 laboratoires, et la collaboration de nombreux autres partenaires, il est coordonné par Hélène Noizet, membre du LAMOP.

Ce projet vise à mettre en place des outils de travail mutualisés STIC-SHS permettant de développer les recherches concernant l'espace urbain parisien, à savoir des logiciels de reconnaissance des formes adaptés aux plans cadastraux anciens et un SIG (Système d'Information Géographique) comprenant des couches cadastrales et historiques. C'est une nouvelle approche du milieu urbain, intégrant réellement la dimension spatiale, qui pourra être mise en oeuvre grâce aux apports de disciplines récentes (imagerie industrielle, géomatique et archéogéographie). Fondés sur la volonté de développer l'interdisciplinarité au sein des SHS (histoire, géographie, archéologie, histoire de l'art, architecture, urbanisme..) et de mettre en place des synergies STIC/SHS, les objectifs scientifiques sont multiples :

- Construire des outils de reconnaissance des formes innovants et adaptés aux plans cadastraux anciens
- Utiliser des modèles explicatifs pour expliquer la répartition géographique des objets
- Analyser la morphologie du parcellaire à l'échelle de la ville

Le patrimoine culturel et scientifique de l'Europe est un bien public unique qui représente la mémoire collective et vivante de nos différentes sociétés et qui forme une base solide pour le développement des industries s'appuyant sur le contenu numérique dans une société de la connaissance durable. De nombreuses démarches de sauvegarde et de valorisation de données patrimoniales sont engagées, comme en attestent les impulsions prises au niveau européen par les représentants nationaux des grands organismes de gestion du patrimoine, se traduisant par des processus massifs de numérisation d'ouvrages anciens. Ces collections d'ouvrages numérisés constituent d'ores et déjà des entrepôts gigantesques de données, représentés sous forme d'images scannées, pour lesquelles les techniques traditionnelles des bases de données sont inopérantes. L'exploitation et la valorisation à venir de ces collections d'images n'ont toujours pas trouvées de réponses satisfaisantes, du fait même de leurs caractères faiblement structurés. D'autre part, la diversité des usages liée aux différents besoins, nous amène aux des questions dites de « re-documentarisation » personnalisée de la part des experts des Sciences Humaines et Sociales.

L'interprétation de documents graphiques est un enjeu crucial. Le document technique, que représente le cadastre par îlots de Philibert Vasserot, joue un rôle tout à fait stratégique car il constitue une version graphique du patrimoine. La conservation et la mise à disposition de cette information entre dans la problématique de la Gestion Électronique de Document (G.E.D.), dont les systèmes ont pour but de permettre une circulation efficace de l'information et de capitaliser la connaissance. Si l'analyse d'un document se réduisait au passage du document papier à une image " pixellaire ", le problème serait trivial. En réalité les problèmes scientifiques sous-jacents sont très complexes car l'objectif est bien plus ambitieux, passer du document papier à son interprétation sémantique. La

question peut se résumer sous cette forme : Comment construire à partir d'amas de pixels, à partir d'entrepôts de données inertes, une information porteuse de sens ? Ce domaine couvre plusieurs disciplines scientifiques : le traitement du signal et de l'image, la reconnaissance de forme, l'intelligence artificielle, l'histoire, les interfaces homme-machine ainsi que la modélisation cognitive des processus d'interprétation. Dans le cadre de la vectorisation des plans Vasserot, les outils d'analyses informatiques exploitent les informations colorimétriques des plans numérisés afin d'extraire des vecteurs facilitant la reconstitution du plan parcellaire de Paris.

#### 4.1.2 Résumé de thèse

- Sujet : Fouille et classification de graphes : Application à l'analyse d'images cadastrales couleurs.
- Laboratoire d'accueil : L3i, équipe IMeDoc
- Directeur de thèse : Pr Jean-Marc Ogier
- Soutenue publiquement le 25 novembre 2010
- Obtenue avec la mention *très honorable avec félicitations*
- Composition du jury :

M. Josep LLADÓS	Université autonome de Barcelone	<i>Rapporteur</i>
M. Chew Lim TAN	Université de Singapour	<i>Rapporteur</i>
M <sup>me</sup> Nicole VINCENT	Université de Paris 5	<i>Rapporteur</i>
M. Jean-Marc OGIER	Université de La Rochelle	<i>Directeur</i>
M. Jean-Christophe BURIE	Université de La Rochelle	<i>Encadrant</i>
M. Umapada Pal	Institut de la statistique d'Inde	<i>Examineur</i>
M. Salvatore-Antoine TABBONE	Université de Nancy 2	<i>Examineur</i>
M. Rolf INGOLD	Université de Fribourg	<i>Président</i>

- **Mots-clès :**  
Classification de graphes, représentation à base de graphes, traitement d'images couleurs, reconnaissance des formes, interprétation d'image de document, modélisation de connaissance.
- **Résumé :**  
Les travaux présentés dans ce mémoire de thèse abordent sous différents angles très intéressants, un sujet vaste et ambitieux : l'interprétation de plans cadastraux couleurs. Dans ce contexte, notre approche se trouve à la confluence de différentes thématiques de recherche telles que le traitement du signal et des images, la reconnaissance de formes, l'intelligence artificielle et l'ingénierie des connaissances. En effet, si ces domaines scientifiques diffèrent dans leurs fondements, ils sont complémentaires et leurs apports respectifs sont indispensables pour la conception d'un système d'interprétation. Le centre du travail est le traitement automatique de documents cadastraux du 19e siècle. La problématique est traitée dans le cadre d'un projet réunissant des historiens, des géomaticiens et des informaticiens. D'une part nous avons considéré le problème sous un angle systémique, s'intéressant à toutes les étapes de la chaîne de traitements mais aussi avec un souci évident de développer des méthodologies applicables dans

d'autres contextes. Les documents cadastraux ont été l'objet de nombreuses études mais nous avons su faire preuve d'une originalité certaine, mettant l'accent sur l'interprétation des documents et basant notre étude sur des modèles à base de graphes. Des propositions de traitements appropriés et de méthodologies ont été formulées. Le souci de combler le gap sémantique entre l'image et l'interprétation a reçu dans le cas des plans cadastraux étudiés une réponse.

## 4.2 Contributions

**1) Interprétation de plans cadastraux.** Des détecteurs ont été spécialement conçus pour récupérer et identifier différents composants métiers tels que des caractères de texte, le cadre entourant chaque plan, les quartiers, les rues, et les parcelles. Ces outils spécifiques sont exécutés séquentiellement dans l'objectif d'obtenir les contours de ces objets métiers. Ensuite, ces éléments sont insérés dans une représentation à base de graphe. Cette représentation structurelle est mise en concordance avec la modélisation de connaissance (méta-modèle) définie par les experts. Cette comparaison est effectuée grâce à un algorithme de mise en correspondance de graphes. La "Métamodélisation" fait référence à la construction d'un ensemble de «concepts» (classes, relations, etc) dans un domaine particulier. Cette méthodologie a été publiés dans [5,6].

**2) Évaluation de la vectorisation des plans de cadastre.** Cette étude concerne la conception d'un banc d'essai pour évaluer les systèmes de conversion Image vers Vecteur. Plus précisément, ce protocole est conçu pour évaluer la performance des systèmes de détection et d'approximation de polygones. Ce protocole d'évaluation de performance donne une comparaison scientifique à un niveau polygone de la cohérence d'un document vectorisé. Cet outil utilise des méthodes concrètes d'évaluation de performance qui peuvent être appliquées à des systèmes complets de polygonisation. D'ailleurs, un système dédié à la vectorisation de plans cadastraux a été évalué selon ce point de vue et les résultats en termes de qualité de détection et de précision de l'approximation polygonale sont présentés dans [2]. Ensuite, le comportement de notre série d'indices a été analysé en faisant augmenter le niveau de dégradation de deux jeux de tests contenant des images de documents. Par cette mise à l'épreuve d'un algorithme de polygonisation traditionnel, nous démontrons que notre protocole peut révéler les forces et les faiblesses d'un système de conversion image vers polygones. Enfin, nous espérons que ce protocole d'évaluation permettra d'évaluer sous un autre angle, plus proche de la sémantique et des objets manipulés par l'homme, les outils de retro-conversion de documents de la communauté de la reconnaissance de graphiques.

Notre contribution est double, un algorithme de mise en correspondance d'objets pour localiser spatialement les erreurs dans les documents vecteurs, puis une distance entre graphes qui rend compte de la précision de l'approximation polygonale. L'évaluation de performance intègre de nombreux aspects et facteurs fondés sur des unités uniformes tout en restant générique et sans seuil. Ces résultats ont été publiés dans [2].
--

**3) Mise en correspondance de graphes et distance entre graphes fondées sur l'assignement de sous-graphes.** Au cours de la dernière décennie, l'utilisation d'objets représentés à base de graphes a considérablement augmenté. En effet, la représentation d'objets par le biais de graphes a de nombreux avantages sur la représentation traditionnelle par des vecteurs de caractéristiques. Par conséquent, les méthodes pour comparer les graphes sont devenus de premier intérêt.

Dans ma thèse, une méthode de mise en correspondance de graphes et une distance entre graphes relationnels attribués sont définies. Les deux approches sont fondées sur la décomposition en sous-graphes. Dans ce contexte, les sous-graphes peuvent être considérés comme des caractéristiques structurelles extraites à partir d'un graphe donné. La nature structurelle de nos caractéristiques de graphe permet de représenter l'information locale d'un nœud racine. Étant donné deux graphes  $g_1$ ,  $g_2$ , la mise en correspondance peut être exprimé comme l'adéquation minimale entre les sous-graphes de  $g_1$  et les sous-graphes de  $g_2$  et ceci en respectant une fonction de coût. Par des expérimentations sur 4 jeux de données différents, la distance induite par mise en correspondance de graphes a été appliquée pour mesurer l'exactitude de l'appariement de graphes. Enfin, nous démontrons un gain important en rapidité par rapport aux méthodes conventionnelles tout en gardant une précision pertinente.

Nous avons proposé un **cadre formel** pour la mise en correspondance de graphes, applicable à de larges types de graphes. Les points clés de notre méthode (SGMD) sont:

- Une complexité algorithmique polynomiale ( $O(n^3)$ ).
- Nous avons prouvé les relations suivantes entre notre distance et la distance théorique (la distance d'édition de graphes ED) :
  - SGMD est une borne inférieure pour la distance d'édition.
  - Une relation de rangs existe entre SGMD et ED.

Ces résultats ont été publiés dans [3].

### 4.3 Mémoire de Master : Reconnaissance de symboles à partir de schémas électriques

Mon mémoire de Master a été effectué au sein de l'entreprise ALGOTECH informatique dans le département R&D en partenariat avec le laboratoire Litis de Rouen sous la direction de Sébastien Adam (Litis) et Élise Gabarra (ALGOTECH). nous avons proposé deux approches différentes de reconnaissance de symboles.

(1) L'approche statistique décrite dans mon mémoire est associée à une combinaison parallèle de classifieurs. Cette méthode permet de confronter des points de vue différents d'un même problème afin d'obtenir une classification finale plus précise. Cette méthode fonctionne de manière contrastée, de bons résultats sont obtenus à partir de la base GREC alors que les taux de reconnaissances sont décevants sur la base ELEC. Une partie de l'explication vient d'une « mauvaise » qualité des symboles de la base ELEC[8 000 pixels par image], en effet ces derniers possèdent une densité de pixels bien inférieure à celle de la base GREC[250 000 pixels par image]. Notre approche statistique est sensible au contexte d'acquisition de l'image

(2) L'approche structurelle permet de représenter des symboles graphiques matérialisés par des graphes structurels comportant des attributs nominaux. Lors de la construction des graphes, l'utilisation de leur longueur relative comme attributs sur les sommets et des angles relatifs entre les vecteurs comme attribut sur les arcs rend nos graphes invariants à la rotation et au changement d'échelle. Une mesure de similarité rapide et significative est décrite pour comparer de tels graphes étiquetés et ainsi trouver le meilleur appariement possible entre deux graphes. Pour tenir compte de la présence de bruit et de distorsions, notre approche est basée sur un appariement de graphes tolérants aux erreurs qui utilisent des valeurs nominales comme attributs. Les résultats obtenus sur les corpus de tests sont encourageants mais des efforts restent à fournir sur l'extraction des primitives tel que les arcs de cercle et les vecteurs. Ces derniers manquent de précision et engendrent des erreurs de représentation de l'information.

## 5 PUBLICATIONS ET PROJETS

---

### 5.1 Publications

#### References

##### Revue d'audience internationale avec comité de rédaction

- [1] Romain Raveaux, Sébastien Adam, Pierre Héroux and Eric Trupin. Learning Graph Prototypes for Shape Recognition. *Computer Vision and Image Understanding (CVIU)*, (Révision : 2nd round).
- [2] Romain Raveaux, Jean-Marc Ogier and Jean-Christophe Burie. A Local Evaluation of Vectorized Documents by means of Polygon Assignments and Matching. *Journal: International Journal on Document Analysis and Recognition (IJ DAR)*, (In press), 2010.
- [3] Romain Raveaux, Jean-Christophe Burie and Jean-Marc Ogier. A graph matching method and a graph matching distance based on subgraph assignments. *Pattern Recognition Letters*, 31(5):394–406, 2010.

##### Actes de conférence d'audience internationale avec comité de sélection

- [4] Romain Raveaux and Guillaume Hillairet. Model Driven Image Segmentation Using a Genetic Algorithm for Structured Data. In Manuel Grana Romay, Emilio Corchado, and Teresa Garcia-Sebastian, editors, *Hybrid Artificial Intelligence Systems, 5th International Conference, HAIS 2010*, pages 311–318. Springer, Lecture Notes in Computer Science, 2010.
- [5] Romain Raveaux, Jean-Christophe Burie, and Jean-Marc Ogier. A colour text/graphics separation based on a graph representation. In *19th International Conference on Pattern Recognition (ICPR)*, pages 1–4, Washington, DC, USA, 2008. IEEE Computer Society.
- [6] Romain Raveaux, Jean-Christophe Burie, and Jean-Marc Ogier. Object Extraction from Colour Cadastral Maps. In *DAS '08: Proceedings of the 2008 The Eighth IAPR International Workshop on Document Analysis Systems*, pages 506—514, Washington, DC, USA, 2008. IEEE Computer Society.
- [7] Romain Raveaux, Jean-Christophe Burie, and Jean-Marc Ogier. A Colour Space Selection Scheme dedicated to Information Retrieval Tasks. In *Pattern Recognition in Information Systems, Proceedings of the 8th International Workshop on Pattern Recognition in Information Systems, PRIS 2008, In conjunction with ICEIS 2008, Barcelona, Spain, June 2008*, pages 123–134. INSTICC PRESS, 2008.
- [8] Romain Raveaux, Jean-Christophe Burie, and Jean-Marc Ogier. *A Segmentation Scheme Based on a Multi-graph Representation: Application to Colour Cadastral Maps*, volume 5046 of *Lecture Notes in Computer Science*, pages 202–212. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

- [9] Romain Raveaux, J.-C. Burie, and J.-M. Ogier. A Colour Document Interpretation: Application to Ancient Cadastral Maps. In *ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition*, pages 1128–1132, Washington, DC, USA, 2007. IEEE Computer Society.
- [10] Romain Raveaux, Barbu Eugen, Hervé Locteau, Sébastien Adam, Pierre Héroux, and Eric Trupin. *A Graph Classification Approach Using a Multi-objective Genetic Algorithm Application to Symbol Recognition*, volume 4538 of *Lecture Notes in Computer Science*, pages 361–370. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [11] Eugen Barbu, Romain Raveaux, Herve Locteau, Sebastien Adam, Pierre Heroux, and Eric Trupin. Graph Classification Using Genetic Algorithm and Graph Probing Application to Symbol Recognition. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 296—299, Washington, DC, USA, 2006. IEEE Computer Society.
- [12] Herve Locteau, Romain Raveaux, Sebastien Adam, Yves Lecourtier, Pierre Heroux, and Eric Trupin. *Polygonal Approximation of Digital Curves Using a Multi-objective Genetic Algorithm*, volume 3926 of *Lecture Notes in Computer Science*, pages 300–311. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [13] Hervé Locteau, Romain Raveaux, Sébastien Adam, Yves Lecourtier, Pierre Héroux, and Éric Trupin. Approximation of Digital Curves using a Multi-Objective Genetic Algorithm. In *18th International Conference on Pattern Recognition (ICPR)*, pages 716–719, Washington, DC, USA, 2006. IEEE Computer Society.

### **Exposés sans publication**

- [14] Romain Raveaux. Interprétation de documents dirigée par les modèles. *Université de Rouen*, Mai 2008.
- [16] Romain Raveaux. A Colour Space Selection Scheme dedicated to Information Retrieval Tasks. *Computer Vision Center (CVC)*, Barcelona, Juin 2008.

## **5.2 Responsabilités scientifiques**

Relecteur pour la revue suivante depuis 2009 :

- ELCVIA : Electronic Letters on Computer Vision and Image Analysis

Relecteur pour les conférence suivantes depuis 2007:

- ICPR : International Conference on Pattern Recognition
- ICDAR : International Conference on Document Analysis and Recognition
- HAIS : International Conference on Hybrid Artificial Intelligence Systems

Membre du comité *scientifique* pour la conférence suivante depuis 2009 :

- HAIS : International Conference on Hybrid Artificial Intelligence Systems



## 5.3 Participation à des projets nationaux et internationaux

### 1) Projet ALPAGE

**Dimension** ANR  
**Date** 2006-2009  
**Thème** AnaLyse diachronique de l'espace urbain PARisien : approche GEomatique  
**Animation** MdC Hélène Noizet (LAMOP)  
**Partenaires** LAMOP, L3i, ArcScan, Lienss.

Dans ce projet, j'ai mené à bien l'interprétation d'images cadastrale couleur.

### 2) Projet NAVIDOMASS

**Dimension** ANR  
**Date** 2007-2009  
**Thème** Navigation dans des masses de données  
**Animation** Pr Jean-Marc Ogier (L3i)  
**Partenaires** Loria, Litis, L3i, IRISA, LI, CESR, CRIP5

Mon implication dans ce projet se résume à la réalisation d'un exposé [14] et j'ai assisté à plusieurs réunions de projet à Paris.

## 6 ADMINISTRATION, CHARGES COLLECTIVES ET IMPLICATIONS DIVERSES

---

- Membre du comité d'organisation du workshop d'audience internationale sur la reconnaissance d'informations graphiques (GREC : International Workshop on Graphics RECOgnition), 2009. J'ai notamment travaillé sur le portail web et la gestion des inscriptions <http://grec2009.univ-lr.fr>.
- Participation à la rédaction des 3 rapports d'activité du projet ANR ALPAGE. Ces rapports sont disponibles à l'adresse suivante : <http://lamop.univ-paris1.fr/alpage/index.php/documentation>.
- Participation à la rédaction à 10 rapports d'activité du projet ANR ALPAGE. Ces rapports sont disponibles à l'adresse suivante : <http://lamop.univ-paris1.fr/alpage/index.php/documentation>.
- Présentation de projets (ALPAGE et NAVIDOMASS) lors de l'évaluation du L3i par le CNRS, 2007.
- Membre du jury de soutenance de 2<sup>ème</sup> année Master Info, Université de Rouen, 2008-2009.
- Membre du jury de soutenance de 1<sup>ère</sup> année Master Info, Université de Rouen, 2007-2008.
- Membre du jury de soutenance de 2<sup>ème</sup> année Master Info, Université de La Rochelle, 2007-2008.
- Membre du jury de soutenance de 2<sup>ème</sup> année Master Info, Université de La Rochelle, 2008-2009.
- Membre du groupe de travail IMeDoc .
- Participation à l'École d'Hiver sur l'Image Numérique Couleur (EHINC'2007), Poitiers, France, 9-11 Janvier 2007.
- Participation à la journée TPIVC, Gdr ISIS, Thème B : Image et Vision, Action : Traitements perceptuels pour les images et vidéos couleur ) l'ENST Paris, en décembre 2006