Propositions pour la représentation et l'analyse de documents numériques

Jean-Yves Ramel

▶ **To cite this version:**

Jean-Yves Ramel. Propositions pour la représentation et l'analyse de documents numériques. Traitement du texte et du document. Université Francois Rabelais de Tours, 2006. tel-01075984

HAL Id: tel-01075984
https://hal.science/tel-01075984

Submitted on 20 Oct 2014

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.

Université François Rabelais de Tours
Ecole Doctorale : Santé, Sciences et Technologies
Année Universitaire 2005-2006

HABILITATION A DIRIGER DES RECHERCHES

UNIVERSITE FRANCOIS RABELAIS DE TOURS

Discipline : **Informatique**

Présentée et soutenue publiquement

par :

Jean-Yves RAMEL

le 8 novembre 2006

Propositions pour la représentation et l'analyse de documents numériques

---

Jury

Cardot Hubert	Professeur	Université de Tours
Garbay Catherine	Directeur de recherche	CNRS Grenoble
Jolion Jean-Michel	Professeur	INSA de Lyon
Ogier Jean-Marc	Professeur	Université de La Rochelle
Tabbone Antoine	Maitre de conférences	Université de Nancy 2
Tombre Karl	Professeur	INPL de Nancy
Vincent Nicole	Professeur	Université Paris 5

T A B L E D E S M A T I E R E S

PARTIE 1 : CV et Activités scientifiques

Fiche synthétique 3
Présentation générale 5
Formation 7
Encadrements 8
Contrats et projets de recherche 10
Activités administratives et collectives 12
Activités d'enseignement 14
Liste de mes publications 17

PARTIE 2 : Propositions pour la représentation et l'analyse de documents numériques

Introduction 27
1. Préambule 27
2. Qu'est ce qu'un document ? 28
2.1. Définitions 28
2.2. Représentations des documents numériques 30
3. Comment structurer les traitements ? 31
3.1. Source d'inspiration 31
3.2. Traitements incrémentaux, contextuels et interactifs 32

Chapitre 1 : Représentation de documents numériques (images et tracés en ligne) 35
1.1. Caractéristiques des images de documents graphiques, textuels et des tracés en ligne 35
1.1.1. Documents graphiques 35
1.1.2. Documents textuels 38
1.1.3. Tracés en ligne 39
1.2. Extraction de primitives, caractérisation et segmentation 41
1.2.1. Primitives structurelles et segmentation 42
1.2.2. Signatures statistiques et caractérisation des contenus 53
1.2.3. Structuration des primitives et représentation 55
1.3. Problèmes ouverts 57
1.4. Contributions 59
1.4.1. Une représentation pour les images de document et les tracés en ligne 59
1.4.2. Exemples d'exploitation de la représentation obtenue 67

Chapitre 2 : Analyse et reconnaissance des contenus 73
2.1. Stratégies d'analyse et architectures des systèmes 73
2.1.1. Analyse guidée par les données 75
2.1.2. Analyse guidée par le modèle 75
2.1.3. Intelligence artificielle distribuée, vision cognitive et approche psycho-visuelle 77
2.2. Méthodes de reconnaissance des contenus 80
2.2.1. Méthodes statistiques 80
2.2.2. Comparaison directe 81
2.2.3. Méthodes structurelles 82

2.3. Problèmes ouverts 84
2.4. Contributions 86
2.4.1. Analyses incrémentales et interactives 86
2.4.2. Reconnaissance structurelle 91
2.4.3. Fusion d'informations pour la reconnaissance 93

Chapitre 3 : Documents numériques et interaction 97
3.1. Documents numériques adaptatifs 98
3.1.1. Définitions 98
3.1.2. Architectures des systèmes adaptatifs 99
3.1.3. Hypermédias adaptatifs 101
3.2. Problèmes ouverts 103
3.3. Contributions 104
3.3.1. Structuration des documents numériques adaptatifs 105
3.3.2. Structuration des environnements hypermédias adaptatifs 107
3.3.3. Présentations adaptatives 109

Conclusion générale 115

Références 119

ANNEXE : Sélection d'articles 133

J.Y. Ramel, S. Leriche, ML Demonet, S. Busson. User-driven Page Layout Analysis of historical printed Books. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition Accepté, à paraitre 2006.

Qureshi RJ., Ramel JY, Cardot H., Graphic Symbols Recognition Using Flexible Matching of Attributed Graphs. The Sixth IASTED International Conference VIIP 2006. August 28-30, 2006 Palma de Mallorca, Spain. p.383-388.

Hocquet S., Ramel JY, Cardot H., User specific parameters in one-class problems. International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong. p449-452.

Journet N., Eglin V., Ramel J.Y., Mullot R., Dedicated texture based tools for characterisation of old books. 2nd IEEE International Conference on Document Image Analysis for Libraries. Lyon, France, April 27-28, 2006. p60-70.

Wirotius M., Ramel JY., Vincent N., Selection of Points for On-Line Signature Comparison. IWFHR. Tokyo (Japon). p503-508. Octobre 2004.

Benadi S., Ramel JY, Prévot P.,AHXEL: Adaptive Hyperdocument using XML for E-Learning. 5th IEEE Int. Conf. on Information Technology Based Higher Education and Training: ITHET'04, Istambul. Turquie. 31 Mai - 2 Juin, 2004. CD-ROM

J.Y. Ramel, N. Vincent, H. Emptoz. A structural representation for understanding line drawing images. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition, Volume 3 n°2 - Décembre 2000. p58-66.

Remerciements

Avant toute chose, je tiens à remercier les membres de mon jury, qui ont su trouver du temps, dans des agendas que je sais très remplis, pour lire ce mémoire, l'évaluer et participer à ma soutenance.

Je remercie tout d'abord Catherine Garbay, directeur de recherches au CNRS pour le temps qu'elle a consacré à la lecture de ce mémoire. Ses recherches concernant la conception de systèmes capables de développer des schémas d'interactions complexes avec leur environnement m'ont toujours passionné et ont fortement influencé mes travaux. Ses compétences pluridisciplinaires et ses nombreuses activités sur le plan national m'impressionnent ; j'apprécie donc énormément qu'elle ait accepté d'être rapporteur de mon travail.

Merci également à Jean-Michel Jolion, Professeur à l'INSA de Lyon, pour l'intérêt qu'il a toujours porté à mon travail, pour les conseils et les aides qu'il m'a prodigués, déjà lorsque j'étais en thèse, puis à de nombreuses autres occasions par la suite. Sa puissance de travail, sa rigueur scientifique et ses qualités humaines ont été pour moi des modèles.

Je remercie Karl Tombre, Professeur à l'école des Mines de Nancy, pour le plaisir qu'il m'a fait en acceptant d'être rapporteur de ce mémoire. La qualité de ces recherches et de celles de son équipe, son investissement scientifique et son charisme contribuent de manière très forte à la représentation de la France au niveau international. La pertinence de son jugement scientifique et sa disponibilité m'apparaissent exemplaires. Je lui suis donc extrêmement reconnaissant d'avoir trouvé du temps pour juger mon travail.

Je remercie Jean-Marc Ogier, Professeur à l'université de La Rochelle, d'avoir accepté d'être examinateur de mon travail et de présider le jury. Depuis presque toujours, nos travaux ont de nombreux points communs (surtout terminologiques) mais aussi de nombreuses différences qui les rendent complémentaires. Je suis heureux d'avoir entamé des collaborations fortes avec son laboratoire. Travailler avec Jean-Marc est un réel plaisir ; sa convivialité, son dynamisme et ses compétences dans des domaines diversifiés provoquent toujours un foisonnement d'idées et de projets nouveaux que l'on retrouve nulle part ailleurs !

Je remercie Antoine Tabbone, Maitre de conférences HDR à l'Université de Nancy 2, d'avoir accepté, dans des délias plus que court, de lire mon mémoire et de participer à mon jury. Son jugement sur mon travail a fait naitre des discussions fructueuses qui continueront sans nul doute à alimenter nos futures collaborations.

Merci à Nicole VINCENT, Professeur à l'université PARIS V, pour l'aide et le soutien qu'elle m'a toujours accordés, d'abord durant ma thèse sous sa direction à Lyon, puis à Tours en tant que collègue. Je la remercie de m'avoir initié à la recherche et, par la suite, de m'avoir permis de poursuivre les recherches qui me tenaient à cœur dans le domaine de l'analyse d'images de documents. C'est un grand plaisir pour moi qu'elle fasse partie de ce jury.

Merci à Hubert CARDOT, Professeur à l'université de TOURS et responsable de l'équipe RFAI, pour la confiance qu'il a su m'accorder depuis son arrivée à Tours, pour l'ambiance de travail convivial qu'il entretient au sein de l'équipe et pour les conseils qu'il m'a donnés afin d'améliorer la version finale de ce document.

Mes plus vifs remerciements vont également aux doctorants avec qui j'ai travaillé dans le passé : Sofiane et Matthieu, aux doctorants avec qui je travaille encore actuellement : Nicholas, Sylvain et Rashid et aux étudiants du Département Informatique et stagiaires de Master dont j'ai encadré les travaux. Sans eux, toutes ces recherches n'auraient pu avoir lieu : Ils ont tous largement contribué aux résultats présentés dans cette habilitation.

Merci à l'ensemble des membres du Laboratoire et du Département Informatique que j'ai côtoyés durant ces dernières années ; travailler avec eux est un réel plaisir. Un grand merci à tous ceux avec qui j'ai des échanges privilégiés, que ce soit sur le plan scientifique, culturel, sportif ou tout simplement humain. Je remercie spécialement Mohand Slimane pour sa disponibilité, ses conseils et ses encouragements notamment lors de la rédaction de ce manuscrit.

Enfin, j'adresserai mon dernier remerciement, mais non le moindre, à ma famille (Sandrine, Thomas, Océane et tous les autres) pour son indispensable soutien et pour tout le reste ...

P a r t i e 1

CV & Principales Activités scientifiques

FICHE SYNTHETIQUE

Nom Prénom :	**RAMEL Jean-Yves**
Date et lieu de naissance :	**18 juillet 1970, à Rumilly (74, France)**
Nationalité :	**Française**
Situation de famille :	**Marié - 2 enfants**
Fonction :	**Maître de Conférences - section 27 (depuis septembre 1998)**
Etablissement :	**Département Informatique – Laboratoire Informatique** **Ecole Polytechnique de l'Université de Tours**
Recherche :	**Traitement d'images, Reconnaissance de Formes, Analyse de Documents** **Interaction Homme-Machine, Télé-enseignement.**
Publications :	**2 revues internationales, 2 revues nationales, 3 sélections en LNCS** **20 conférences internationales, 12 conférences nationales,** **18 communications et rapports**
Encadrement, co-encadrement :	**3 thèses en cours, 2 thèses soutenues, 4 DEA-M2R** **PEDR depuis septembre 2004** **4 jurys de thèse**
Activités contractuelles :	**Projet Technovision EPEIRES, ACI ministérielle Madonne, Projet RNTL Kénobi, Projet INSA-V (Campus Numérique National), Conception coopérante (Région Rhône-Alpes), SESAME (CNET, CCETT, FT)**

Animation & Rayonnement : **Membre de la commission de spécialistes en section 27 (Université de Tours)**
Membre de la commission de spécialistes en section 27 (Université Paris V)
Responsable adjoint de l'équipe RFAI (20 chercheurs) du Laboratoire d'Informatique de Tours
Membre du conseil du Laboratoire d'Informatique de Tours (2002-2005)
Membre du Conseil Administration de l'AFRIF (depuis 2004)
Participations à des comités de relecture, programme et d'organisation de colloques
Participations à plusieurs groupes de travail du GDR I3 et ISIS

Enseignements : **Informatique, Réseaux, Analyse de données, Images, Interfaces Homme - Machine, Encadrement de projets collectifs et de fin d'études.**
Niveau : **école d'ingénieurs, DEA d'informatique, DEUG**
Volume horaire : **environ 200 heures en équivalent TD par an**

Adresse personnelle :
27, rue du Stade
37250 VEIGNE
(02-47-34-08-82

Adresse professionnelle :
Laboratoire Informatique
Polytech'Tours
64, avenue Jean Portalis
37200 TOURS
(02-47-36-14-26

PRESENTATION GENERALE

Cette partie décrit l'originalité de mon parcours en soulignant la double compétence et les nombreux contacts que j'ai pu acquérir grâce à mon passage dans plusieurs laboratoires de recherche.

- **Laboratoire Reconnaissance de Formes et Vision (1993-1998)**

Thèmes : **Reconnaissance de formes, Interprétation automatique de documents manuscrits**

De 1993 à 1998 (DEA, thèse, ATER), j'ai mené mes activités de recherche au sein du Laboratoire **Reconnaissance de Formes et Vision (RFV)** de l'INSA de Lyon. Le thème principal de mes travaux était alors la reconnaissance des formes et le traitement d'images appliqués à l'interprétation automatique de dessins manuscrits. Ces travaux ont abouti à la soutenance de mon doctorat en novembre 1996 et à la réalisation d'un prototype logiciel de lecture automatique de formules chimiques manuscrites valorisé auprès d'une PME grenobloise.

- **Laboratoire Interaction Collaborative, Télétravail, Téléformation (1998-2002)**

Thèmes : **Interaction Homme-Machine et architectures informatiques pour le Télé-enseignement**

Mon recrutement comme maître de conférences, en septembre 1998, au sein du Laboratoire **Interaction Collaborative, Télétravail, Téléformation (ICTT – INSA de Lyon)** s'est traduite par un élargissement de mes thématiques de recherche vers le domaine de l'interaction Homme-Machine. J'ai ainsi pu acquérir de nouvelles compétences tout en essayant d'utiliser les résultats de mes travaux antérieurs dans un nouveau contexte. Mon intégration au sein d'un laboratoire **pluridisciplinaire** m'a également offert l'opportunité de donner une dimension nouvelle à mon expérience de travail en équipe, au travers de collaborations avec des chercheurs en **Sciences Humaines et Sociales** impliqués dans les mêmes projets que moi.

Mes réflexions se sont d'abord portées sur les architectures informatiques des plateformes de télé-enseignement. Ce travail a servi de base à la mise en place d'un portail de télé-formation (Owasis) aujourd'hui utilisé à l'INSA de Lyon. Dans un deuxième temps, la poursuite de cette étude a été confiée à Mr Sofiane Benadi qui, dans le cadre de sa thèse (que j'ai co-encadré à 50% avec Patrick Prévot professeur à l'INSA de Lyon), propose un système produisant des interfaces pédagogiques adaptatives.

En parallèle à ces réflexions et dans la continuité des travaux que j'avais réalisés à RFV, j'ai entrepris l'adaptation du système développé durant ma thèse pour créer un système de reconnaissance de documents en-ligne (réalisés avec un stylo électronique). Ces travaux ont été réalisés **en collaboration avec Nicole Vincent** alors Professeur à l'université de Tours.

- **Laboratoire Informatique de Tours (depuis 2002)**

Thèmes : **Analyse de l'écriture manuscrite et analyse de documents (anciens et graphiques)**

En 2002, j'ai souhaité me recentrer sur des recherches dans les domaines de la reconnaissance des formes et de l'interaction Homme-Document électronique en joignant mon expérience à celle de l'équipe de RFAI du **Laboratoire Informatique de Tours**. Cette **mutation** m'a donné la possibilité d'approfondir certaines études me tenant très à cœur.
Ma collaboration, durant l'année 2001, avec Nicole Vincent sur l'écrit en-ligne m'a permis de co-encadrer une thèse en bourse CIFRE dès mon arrivée à Tours en septembre 2002. Cette thèse avait comme objectif l'authentification de signatures manuscrites réalisées sur PDA. Elle a été soutenue par Matthieu Wirotius le 10 décembre 2005. Dans la continuité de cette thèse, mes travaux dans le domaine de la biométrie se poursuivent actuellement dans le cadre de la thèse de Sylvain Hocquet que je co-encadre avec Hubert Cardot (soutenance prévue en novembre 2006).

Depuis mon arrivée à Tours, j'ai également pu reprendre mes activités de recherche dans le domaine de l'analyse d'images de documents. Différentes collaborations (avec le Centre d'Etude Supérieur de la Renaissance de Tours, avec d'autres laboratoires de recherche durant l'**ACI ministérielle Madonne**) m'ont permis d'encadrer la thèse de Nicholas Journet en collaboration avec Remy Mullot, Professeur à La Rochelle et Véronique Eglin, maître de conférences à l'INSA de Lyon (soutenance prévue en novembre 2006). Concernant la reconnaissance de documents graphiques, je co-encadre actuellement la thèse de Rashid-Jalal Qureshi qui travaille sur les méthodes structurelles de reconnaissance de formes et dont les propositions seront évaluées durant le projet **Technovision EPEIRES** (soutenance prévue en novembre 2007).

FORMATION

- 1993-1996 : **Doctorat d'Informatique** obtenu le 26 novembre 1996 à l'INSA de Lyon avec mention Très Honorable. Thèse préparée au sein du laboratoire Reconnaissance de Formes et Vision.

  Sujet : Lecture automatique de dessins, méthodes d'analyse et de reconnaissance, application aux plans cinématiques
  Directeur de thèse : Nicole Vincent, Maitre de Conférences HDR à l'INSA de Lyon

Rapporteurs :	*Jean-Pierre COCQUEREZ, Professeur* *Université de Cergy-Pontoise*	*Guy LORETTE, Professeur* *IRISA, Université de Rennes 1*
Jury :	*J.M. BRUN, Professeur* *Université Claude Bernard (Lyon I)*	*K. TOMBRE, Professeur* *INRIA Lorraine*
	H. EMPTOZ, Professeur *I.N.S.A. (Lyon)*	*N. VINCENT, Professeur* *Université Paris V*

- 1993 : **D.E.A. d'Ingénierie Informatique** de l'INSA de Lyon avec mention Assez Bien.

  Sujet de stage : Etude et mise en oeuvre d'un système de lecture automatique de partitions musicales. Responsable : Nicole Vincent, Professeur à l'université Paris V.

- 1992 : **Maîtrise de Sciences et Techniques "Informatique et Télécommunications"** avec mention Assez Bien.

- 1990 : **DUT Génie Electrique et Informatique Industrielle** option automatismes et systèmes.

- 1988 : **Baccalauréat C** avec mention Assez Bien.

ENCADREMENT DOCTORAL

Les thèses et les DEA que j'ai encadrés traduisent ma volonté de collaboration avec d'autres chercheurs et une certaine mobilité (co-encadrements de thèse avec des professeurs de plusieurs universités).

THESES SOUTENUES

¤ Co-encadrement à 50% avec Mr P. Prévot (Professeur à l'INSA de Lyon) de la thèse de Mr Sofiane Benadi « Structuration des services et des données dans les systèmes d'enseignement à distance ». Financement : Bourse CIES algérienne. Thèse de doctorat en Informatique soutenue à l'INSA de Lyon le 29 septembre 2004. (Mention Très Honorable).

¤ Co-encadrement à 50% avec Mme N. Vincent (Professeur à Paris V) de la thèse de Mr Matthieu Wirotius. « Authentification forte par analyse biométrique de la signature manuscrite en ligne». Financement : Bourse CIFRE. Thèse de doctorat en Informatique soutenue à l'Université de Tours le 10 novembre 2005. (Mention Très Honorable).

THESES EN COURS

¤ Co-encadrement à 50% avec Mr R. Mullot (Professeur à l'Université de La Rochelle) de la thèse de Mr Nicholas Journet depuis octobre 2003 : « Analyse de documents : classification des contenus et indexation». Financement : Bourse Région Poitou-Charente. Soutenance prévue en novembre 2006.

¤ Co-encadrement à 50% avec Mr H. Cardot (Professeur Université de Tours) de la thèse de Mr Sylvain Hocquet depuis octobre 2003 : «Authentification d'un utilisateur par mesures biométriques acquises à partir d'interfaces standards ». Financement : Bourse CIFRE. Soutenance prévue en novembre 2006.

¤ Co-encadrement à 50% avec Mr H. Cardot (Professeur à l'Université de Tours) de la thèse de Mr Rashid-Jalal Qureshi depuis octobre 2004 : « Reconnaissance de symboles complexes dans les images de documents ». Financement : Bourse SFERE Pakistan-France. Soutenance prévue en novembre 2007.

DEA / MASTER

¤ Encadrement du Master 2 recherche de Mr Aurélien Hublier en 2004-2005. « Représentation et Localisation de symboles dans les images de documents techniques ».

¤ Encadrement du DEA de Mlle Nouha Bouteldja en 2003-2004. « Système multi-agents pour l'interprétation de dessins manuscrits ».

¤ Encadrement du DEA de Mr Stéphane Leriche en 2003-2004. « AGORA : un système d'analyse graphique adapté aux ouvrages de la Renaissance ». Ce DEA a donné lieu à une publication dans la revue Traitement du Signal et des Images [Ramel05a].

¤ Encadrement du DEA de Mlle Anne-Sophie Darnand en 2002-2003. « Reconnaissance automatique de gestes graphiques en ligne ».

¤ Encadrement du DEA de Mr Gérome Billois en 2001-2002. « Scalabilité, mobilité et qualité de services dans les grands réseaux de transmission multicast : application au télé-enseignement » [Billois02].

¤ Co-encadrement du DEA de Mr Sofiane Benadi en 1999-2000. «Technologies XML : concepts et applications. Utilisabilité dans le projet Web'Activ-Stratélia».

JURYS DE THESES

¤ Participation (invité) au **jury de thèse** de Mr Mathieu Delalandre . Thèse de doctorat en informatique intitulée « Analyse de documents graphiques : une approche par reconstruction d'objets » et soutenue le 13 décembre 2005 à l'Université de Rouen.

¤ Participation au **jury de thèse** de Mr Matthieu Wirotius. Thèse de doctorat en informatique intitulée « Authentification forte par analyse biométrique de la signature manuscrite en ligne» et soutenue le 10 novembre 2005 à l'Université de Tours.

¤ Participation au **jury de thèse** de Mr Sofiane Benadi. Thèse de doctorat en informatique intitulée « Structuration des services et des données dans les systèmes d'enseignement à distance » et soutenue le 29 septembre 2004 à l'INSA de Lyon.

¤ Participation au **jury de thèse** de Mr Pascal Mercy. Thèse de doctorat en informatique intitulée « Une méthode de vectorisation des plans mécaniques » et soutenue le 7 décembre 2000 à l'Université du Havre.

CONTRAT ET PROJETS

¤ **Projet Techno-vision EPEIRES (2004-2006) :** Ce projet est soutenu par le ministère de la Recherche et de la Défense. L'objectif de la proposition ÉPEIRES est de mettre en place un environnement d'évaluation des performances de certains des traitements requis pour réaliser la reconnaissance et la localisation de symboles dans les images de documents intégrant des composantes graphiques. Ce projet regroupe de nombreux partenaires comme le Laboratoire L3i (université de La Rochelle), le Laboratoire PSI (université de Rouen), l'Équipe QGAR du LORIA, l'Équipe DAG du Computer Vision Center, (Barcelone), le Laboratoire ONE (France Télécom R&D), et la société Algo'Tech Informatique. J'ai la charge de ce projet pour le Laboratoire Informatique de Tours : gestion du budget (12000 Euros TTC), du personnel (stagiaires, Masters, thèses en lien avec ce projet) et animation des recherches scientifiques faites à Tours.

¤ **ACI Madonne (2003-2006) :** Ce projet s'inscrit dans une démarche de sauvegarde et de valorisation de données patrimoniales dont la communauté internationale a pris conscience de l'intérêt récemment. Pour cette ACI, les 6 équipes participantes, expertes et complémentaires dans le domaine de l'analyse d'images de documents, proposent d'apporter des solutions scientifiques et technologiques innovantes à cette problématique liée aux masses de données constituées par des collections d'ouvrages numérisés. La génération de ces entrepôts de données, présentés sous forme de collections de documents hétérogènes faiblement structurés soulève le problème de la recherche d'information et de la navigation au sein de ces corpus. Il s'agit de déterminer des indices s'adaptant aux différentes représentations de l'information que l'on peut rencontrer dans ces documents patrimoniaux comme des zones textuelles, imprimées ou manuscrites, des images, des illustrations graphiques. Ces signatures apporteront des connaissances spécifiques qui aideront la navigation et la recherche d'informations. J'ai la charge de ce projet pour le Laboratoire Informatique de Tours : gestion du budget (12 000 Euros), du personnel (stagiaires, DEA, Postdoc potentiel) et animation des recherches scientifiques faites à Tours.

¤ **Projet RNTL Kenobi (2002-2003) :** Ce projet réalisé en collaboration avec le laboratoire LTCI de l'ENST Paris et la société Sharing Technologie avait pour but la conception d'un système d'extraction de la structure logique de documents électroniques au format PGF afin de pouvoir éditer et partager leur contenu facilement. Dans ce projet, mon rôle a été de participer aux travaux de recherche par l'intermédiaire d'encadrements de stagiaires durant la dernière année du projet et de participer à la rédaction du rapport final rendu au ministère.

¤ **Projet BVH-CESR (2003-2004) :** Ce projet vise à concevoir un système de numérisation et d'indexation dédié aux ouvrages datant de la Renaissance afin que le CESR (Centre d'étude Supérieur de

la Renaissance, UMR CNRS) puisse alimenter plus facilement le fonds documentaire de sa Bibliothèque Virtuelle Humaniste (BVH). Il a débouché sur une collaboration avec le Laboratoire L3i de La Rochelle (co-encadrement avec le Professeur Remy Mullot de la thèse de Nicholas Journet) et le Centre d'Etude Supérieur de la Renaissance de Tours qui fournit le fonds (bases d'images) et l'expertise nécessaire à la réalisation de ces travaux.

¤ **Projet "INSA-V" (2001-2002) :** Ce projet était financé par le Ministère de l'Education Nationale et le Ministère de la Recherche dans le cadre de l'action **Campus numériques français.** Dans ce projet, ma contribution a concerné la rédaction du dossier de demande de financement, l'encadrement de DEA et de stagiaires travaillant sur la conception d'architectures informatiques pour le Télé-enseignement et la structuration des documents pédagogiques (interfaces adaptatives) et a globalement servi de cadre fédérateur aux recherches que j'ai menées au sein du laboratoire ICTT.

¤ **Projet SESAME** : Afin de diffuser et d'exploiter des documents audiovisuels numériques, le **CNET-CCETT France Télécom** (Centre National des Etudes en Télécommunication, Centre Commun d'Etudes en Télédiffusion et Télécommunications) a lancé en 1996 un appel d'offre dans le cadre d'une CTI (Concertation Thématique Informelle) sur le thème *Indexation et recherche par le contenu pour les services multimédias*. Dans ce cadre, le projet SESAME (Système d'Exploration de Séquences Audiovisuelles et Multimédia enrichi par l'Expérience) avait comme objectif d'étudier les possibilités offertes par l'exploitation de documents audiovisuels numériques sous les angles suivants *:* traitement d'image pour l'indexation *(RFV) ;* distribution et accès parallèles à des données audiovisuelles *(LIP) ;* bases de données audiovisuelles *(LISI) ;* aide à l'utilisateur fondée sur l'expérience des sessions d'exploitation d'un système de recherche d'information audiovisuelle *(LISA-LISI)*. J'ai participé à ce projet de septembre 1997 à septembre 1998 et mes travaux ont concerné la mise en place d'outils de traitement et d'analyse de séquences vidéo codées en MPEG2.

¤ **Projet Conception Coopérante :** Mon travail de thèse s'est inscrit dans le cadre du projet «Conception Coopérante» soutenu par la **Région Rhône-Alpes** et mené en partenariat avec différents laboratoires de la région (INPG, ECL, Université Lyon I, INSA de Lyon) de 1994 à 1996. Le but recherché dans ce projet était de faciliter le travail coopératif des concepteurs en mécanique durant la phase de création grâce à la mise en place de dispositifs d'interaction homme-machine intelligents.

ACTIVITES ADMINISTRATIVES ET COLLECTIVES

¤ Pour la recherche

- Co-responsable de l'équipe RFAI depuis septembre 2005 et animation des recherches sur l'Ecrit et le Document
- Membre élu de la commission de spécialistes 27e section de l'Université de Tours (depuis 2003)
- Membre élu de la commission de spécialistes 27e section de l'Université Paris V (depuis 2005)
- Membre du conseil du Laboratoire d'Informatique de l'Université de Tours de 2002 à 2005
- Membre élu du Conseil d'Administation de l'AFRIF (Association Française en Reconnaissance et Interprétation des Formes) depuis juin 2004.
- Membre du GDR I3 (Information, Interaction, Intelligence) – Administration et mise à jour du site Web du Groupe interaction, coopération et communication (SA5.2) en 2003-2004
- Membre du GRCE (Groupe de Recherche en Communication Ecrite)
- Membre de l'AFIHM (Association Francophone d'Interaction Homme-Machine) et de l'ATIEF (Association des Technologies de l'Information pour l'Education et la Formation) de 2000 à 2002
- Participation aux comités de relecture de conférences, de revues et d'ouvrages de la série Lecture Notes in Computer Science de Springer Verlag (IEEE/PAMI, IJDAR, ELCVIA, CIFED, GREC)
- Participation aux comités d'organisation de conférences et colloques (VISUAL'00, TICE'02, LFA'03, RFIA'06)

¤ Pour l'enseignement

- Responsable adjoint du Département Informatique de Polytech'Tours depuis juillet 2006
- Responsable pédagogique de la 3e année du Département Informatique de Polytech'Tours de septembre 2003 à juillet 2006 : Gestion des emplois du temps et des intervenants, suivi pédagogique, relation avec les élèves, …
- Responsable des projets de fin d'études pour le département Informatique de Polytech'Tours depuis septembre 2003 : Centralisation et distribution des sujets, gestion des conventions avec les industriels, suivi et organisation des soutenances, archivage des rapports …
- Responsable de l'option « Indexation de Documents Mutlimédias » du Master 2 Recherche d'Informatique de l'Université de Tours. Gestion de 25 heures de cours avec des intervenants extérieurs

- Elu au conseil de département Génie Productique de l'INSA de Lyon de 2000 à 2002
- Responsable de l'Intranet Pédagogique du département Génie Productique (INSA de Lyon) de 1998 à 2002. L'objectif était d'encourager le corps enseignant à utiliser les TICE pour améliorer ou compléter leurs enseignements traditionnels.

¤ Valorisation

Les recherches menées dans le cadre de l'analyse de l'écriture manuscrite vont constituer une brique fondamentale dans la mise en place d'une salle de visio-enseignement innovante au sein de l'université de Tours. Polytech'Tours s'est allié, depuis cette année, avec l'IAE et s'est attaché la collaboration de l'ESCEM, afin de monter une action originale en matière de création d'entreprises : des étudiants des deux formations se sont associés, sur des dossiers mis en route dans le cadre des Projets de Fin d'Etudes, afin de tenter de créer leurs entreprises innovantes. Ce projet a débouché sur le dépôt d'un dossier de candidature au concours « Entreprises émergentes 2004 » de l'ANVAR [Migeon04].

Suite aux recherches menées en collaboration avec le CESR de Tours, un logiciel d'aide à l'indexation d'ouvrages anciens est développé au sein de l'Université de Tours. Il met en oeuvre des techniques innovantes d'analyse d'images et de reconnaissance des formes : nous aimerions donc le protéger grâce au dépôt **d'une demande de protection du logiciel auprès de l'APP** (en cours via les services valorisation du CNRS et de l'Université Francois Rabelais de Tours) [Ramel05a].

¤ Technique

- Conception et Administration de divers serveurs Web : GT7 du PRC Communication Homme-Machine durant l'année 1996, département Génie Productique de 2000-2002, laboratoire ICTT 2000-2002, Groupe interaction, coopération et communication (SA5.2) de 2003-2004, équipe RFAI du Laboratoire Informatique de Tours
- Administrateur système (stations HP sous système UNIX) pour le laboratoire Reconnaissance de Formes et Vision (RFV) de 1995 à 1997

ACTIVITES D'ENSEIGNEMENT

J'assure actuellement mes enseignements (maître de conférences - 27e section) au sein du **Département Informatique de l'école Polytechnique de l'Université de Tours** après avoir enseigné 4 ans à l'INSA de Lyon, au sein du département Génie Productique (de 1998 à 2002). Auparavant, j'ai enseigné successivement avec le statut de vacataire (1994-1996) à l'**Université Lyon I**, d'ATER à l'INSA de Lyon (1996-1998) et de maître de conférences (depuis 1998) à l'INSA de Lyon.

Outre les responsabilités décrites précédemment, mes enseignements se situent dans le domaine de l'**Analyse de Données,** des **Réseaux** (Transmission de l'information et principes de base des réseaux) et du **Génie Logiciel** (Interface Homme Machine) et concernent des élèves en 2e cycle. Je suis aussi responsable d'un module du Master recherche en Informatique de Tours qui traite de l'indexation des documents multimédias. Je participe également chaque année à l'encadrement de plusieurs projets de fin d'études d'élèves ingénieurs.

Le volume horaire annuel de mes interventions est d'environ 200 heures équivalent TD.

¤ Mes enseignements à Polytech'Tours

- Enseignements en Analyse de données (32 heures de cours + 3x16 heures de TD)
- Cours de Transmission de l'Information (24 heures)
- Enseignements en Réseaux (24 heures de cours+ 3x6 heures de TD + 3x4 heures de TP)
- Enseignements en Interface Homme-Machine (8 heures de cours + 3x8 heures de TP)
- Enseignements en Indexation de Documents Multimédias au sein du Master Recherche Informatique de Tours. Il s'agit de 6 heures de cours d'initiation à la recherche présentant les principaux concepts du domaine.
- Encadrement de plusieurs mini-projets dans différentes matières (Analyse de données, C, java, …)
- Encadrement de stages (2 étudiants/an) et de Projets de Fin d'Etudes (2 étudiants/an).

Le Projet de Fin d'Etudes (P.F.E.) constitue le travail terminal que doit fournir l'élève-ingénieur en vue de l'obtention de son diplôme. Il est mené dans un domaine de l'Informatique sur la base d'un sujet de type recherche et/ou de type industriel. Le but poursuivi est double : éprouver, en situation de projet, la capacité du futur ingénieur à faire preuve d'initiative, en appréciant son degré de maturité face à un problème précis, de haut niveau, qu'il doit résoudre ; faire produire à l'étudiant un travail personnel aboutissant en général à une réalisation personnelle. Le PFE est une mission similaire à celle d'un ingénieur. Il s'agit en général d'adopter un comportement de chef de projets :

phases d'analyse et de conception, organisation du système, mise en oeuvre. Le PFE, correspondant à deux jours de travail par semaine tout au long de l'année, fait l'objet d'un mémoire et d'une soutenance effectuée devant un jury en fin d'année.

¤ Mes enseignements à l'INSA de Lyon

Les matières que j'ai enseignées à l'INSA de Lyon comme ATER puis Maître de Conférences concernaient principalement : l'**Informatique** (algorithmique, architecture, vision, …) et les **Réseaux** (réseaux locaux, interconnexion, Internet, …). De plus, j'assurais, chaque année, l'encadrement technique de deux **projets collectifs** (clés de voûte de la 4e année du département Génie Productique).
En outre, j'ai dispensé un cours sur les **Interfaces Homme-Machine** au sein du **DEA «**Informatique et Systèmes Communicants pour l'Entreprise» de l'école doctorale Informatique, Information et Société de Lyon durant l'année 2002.

D'une durée de 6 mois, le projet collectif correspond à un investissement de près de 2000 heures par projet conduit par un groupe de 7 ou 8 étudiants (maître d'ouvrage). Fil conducteur de la 4e année du département Génie Productique, les objectifs du projet sont d'apprendre aux élèves à « manager » collectivement un projet réel en partenariat avec une entreprise (maître d'œuvre), d'acquérir des compétences sur l'usage des T.I.C. et de formaliser leur connaissance en productique. Voici, à titre d'exemple les sujets des deux projets que j'ai suivis en 2002 :

w Le Petit Monde : Certains examens, comme celui de l'I.R.M., sont très contraignants et engendrent souvent beaucoup de stress notamment chez les enfants. Ceci peut obliger à avoir recours à l'anesthésie générale alors que l'examen n'a qu'un but diagnostique. Dans le cadre de l'humanisation et l'amélioration de la Qualité de Vie dans l'hôpital, l'association Le Petit Monde veut contribuer à limiter de telles pratiques en mettant en place des actions de remplacement moins agressives. En collaboration avec le service d'imagerie pédiatrique de l'hôpital Debrousse, l'association propose de supprimer l'anesthésie générale et de la remplacer par une préparation physique de l'enfant grâce à un jouet simulateur qui intègre des jeux multimédia. La conception et la mise en place de ce simulateur constitue le projet confié à un groupe d'étudiants [Ramel02a].

w **INSAonLine** : Dans le cadre de sa politique de développement forte de la formation en ligne (projet de Campus Numérique INSA-V), l'INSA de Lyon a choisi d'exploiter les compétences des départements Génie Productique et Télécom au sein d'un projet nommé INSA-Online. Le travail demandé aux étudiants est la mise en place d'une architecture permettant de diffuser des cours (vidéo, powerpoint…) à des apprenants à distance en temps réel (streaming) et en différé (vidéo à la demande). Outre la partie logicielle, deux salles d'enregistrement et une régie centrale de diffusion (portail unique) doivent être déployées dans les locaux de l'INSA [Billois02].

¤ Tableau récapitulatif :

Matière	Nature	Niveau	Volume horaire Lieu	Année
Analyse de Données	Cours et TD	2e cycle Bac+4	32 heures + 3x16heures DI – Polytech	2002-2006
Normes et principes des réseaux	Cours, TD, TP	2e cycle Bac+3	24 heures + 3x10 heures DI – Polytech	2002-2006
Transmission de l'information	Cours	2e cycle Bac+3	24 heures DI – Polytech	2002-2006
Interface Homme-Machine	Cours et TP	2e cycle Bac+3	4heures + 3x4heures DI – Polytech	2003-2005
Indexation de Documents Multimédias	Cours	M2R Bac+5	6 heures Université de Tours	2002-2006
Réseaux Locaux **(Normes & Protocoles)**	Cours et TD	2e cycle Bac+4	2x28 heures GPr – INSA	1998-2002
Réseaux Locaux Industriels	Projet	2e cycle Bac+4	3x32 heures GPr – INSA	1996-2000
Projets collectifs	Projet	2e cycle Bac+4	2x50 heures GPr – INSA	1998-2002
Interfaces Homme-Machine	Cours	3e cycle DEA	10 heures DEA de Lyon	2001-2002
Internet - Intranet	Cours et TD	2e cycle BAC + 3	2x12 heures GPr – INSA	1998-2002
Internet - Intranet	TP	2e cycle BAC + 3	3x12 heures GPr – INSA	1998-2000
Architecture des ordinateurs - Programmation Système	TP	2e cycle Bac+3	3x12 heures GPr – INSA	1996-1998
Algorithmique et Langage	Cours, TD et TP	DEUG A 1ere année	5x48 heures Université Lyon I	1994-1997
Vision Industrielle	TP	2e cycle Bac+5	3x8 heures GPr – INSA	1996-1998
Mathématiques	Cours et TD	1er cycle Bac+1	2x25 heures EURINSA	1995-1996
Géométrie Fractale	TP	1er cycle Bac+1	2x4 heures EURINSA	1995-1996

MES PUBLICATIONS

Les noms soulignés correspondent aux personnes (doctorants, stagiaires) pour lesquelles j'ai participé activement à l'encadrement. De même, dans la suite du mémoire, les références soulignées correspondent aux publications auxquelles j'ai participé.

w Revues à audience Internationale

2006:

[Ramel06c] Ramel JY, Wirotius M., Vincent N., A Improvement of alignment approaches for on-line handwritten signature authentication. IEEE Transactions on Systems, Man and Cybernetics –Part B. Special Issue on Recent Advances in Biometrics Systems. **Soumis.**

[Hocquet06a] Hocquet S., Ramel JY, Cardot H., Keystroke Dynamics: the dawn of a Biometric Method. IEEE Transactions on Systems, Man and Cybernetics –Part B. Special Issue on Recent Advances in Biometrics Systems. **Soumis.**

[Ramel06b] Ramel JY, Let's add interaction in document images analysis systems. Electronic Letters on Computer Vision and Image Analysis (ELCVIA). **Soumis.**

[Ramel06a] Ramel JY, Leriche S., Demonet ML, Busson S., User-driven Page Layout Analysis of historical printed Books. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition Accepté, à paraitre 2006.

2000:

[Ramel00a] Ramel JY, Vincent N., Emptoz H., A structural representation for understanding line drawing images. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition, Volume 3 n°2 - Décembre 2000. p58-66.

w Revues à audience nationale

2005:

[Ramel05a] Ramel JY, Leriche S., Segmentation et analyse interactives de documents anciens imprimés. Revue Traitement du Signal (Presse universitaire de Grenoble). Vol 22 numero 3 – novembre 2005. p209-222

1998 :

[Ramel98a] Ramel JY, Vincent N., Emptoz H., Interprétation de documents techniques par "cycles perceptifs" à partir d'une perception globale du document. Revue Traitement du Signal (Presse universitaire de Grenoble). Vol. 15 n°2 - 1998. p1-20.

w Livre LNCS (publication post-workshop)

2004:

[Ramel04a] Ramel JY, Vincent N., Strategy for line drawing understanding. Graphics Recognition. Recent advances and perspectives. LNCS 3088. Springer Verlag: Editors : Josep Lladós and Young-Bin Kwon. 2004. p1-12

2000:

[Ramel00a] Ramel JY, Boissier G., Emptoz H., Graphics Recognition - Recent Advances. Lecture Notes in Computer Science. Volume 1941. Springer Verlag. Editeurs : A. K. Chhabra

and D. Dori. p227-236. 2000. Article sélectionné pour publication post-workshop : A structural representation adapted to handwritten symbol recognition.

1998:

[Ramel00a] Ramel JY, Vincent N., Emptoz H., Graphics Recognition - Algorithms and Systems. Lecture Notes in Computer Science. Volume 1389. Springer Verlag. Editeurs : K. Tombre and A. Chhabra. p48-55. 1998. Article sélectionné pour publication post-workshop : A coarse vectorization as an initial representation for the understanding of line drawing images.

w Conférences internationales avec comité de lecture et publication des actes

2006:

[Qureshi06a] Qureshi RJ., Ramel JY, Cardot H., Graphic Symbols Recognition Using Flexible Matching of Attributed Graphs. The Sixth IASTED International Conference VIIP 2006. August 28-30, 2006 Palma de Mallorca, Spain. p383-388.

[Hocquet06b] Hocquet S., Ramel JY, Cardot H., Combination of classifiers with user specific parameters in keystroke dynamics. The Sixth International Workshop on Pattern Recognition in Information Systems (PRIS-2006) May 23-24, 2006, Cyprus.

[Hocquet06c] Hocquet S., Ramel JY, Cardot H., User specific parameters in one-class problems. International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong. p349-352.

[Ramel06d] Ramel JY, Busson S., Demonet ML., AGORA: the Interactive Document Image Analysis Tool of the BVH Project. 2nd IEEE International Conference on Document Image Analysis for Libraries. Lyon, France, April 27-28, 2006. p145-155.

[Journet06] Journet N., Eglin V., Ramel J.Y., Mullot R., Dedicated texture based tools for characterisation of old books. 2nd IEEE International Conference on Document Image Analysis for Libraries. Lyon, France, April 27-28, 2006. p60-70.

2005:

[Hocquet05] Hocquet S., Ramel JY, Cardot H., Fusion of Methods for Keystroke Dynamic Authentication. IEEE Workshop on Automatic Identification Advanced Technologies. 17-18 October 2005, Buffalo, New York, USA. p224-229.

[Wirotius05a] Wirotius M., Ramel JY, Vincent N., Distance and Matching for Authentication by On-Line Signatures. IEEE Workshop on Automatic Identification Advanced Technologies. 17-18 October 2005, Buffalo, New York, USA. p230-235.

[Qureshi05] Qureshi RJ., Ramel JY, Ali U., Cardot H., Graph Based Recognition of Isolated Graphic Symbols. ICET'05. International Conference on Emerging Technologies. Islamabad, Pakistan, 17-18 September 2005.

[Journet05a] Journet N., Mullot R., Ramel JY, Eglin V., Ancient Printed Documents indexation: a new approach. ICAPR'05. Third International Conference on Advances in Pattern Recognition, Pattern Recognition and Data Mining: Bath, United Kingdom. Lectures Notes in Computer Science - LNCS 3686. Springer-Verlag, . p513-522. ISSN 0302-9743. August 2005.

[Journet05b] Journet N., Eglin V., Ramel JY, Mullot R., Text/Graphic labelling of Ancient Printed Documents. ICDAR'05. International Conference on Document Analysis and Recognition. Séoul. Corée. August 2005. p1010-1014.

[Wirotius05b] Wirotius M., Ramel JY., Vincent N., Contribution of global temporal information for authentication by on line handritten signaures. International Graphonomics Society (IGS), Salerno (Italie), p. 266-270, June 2005.

[Wirotius05c] Wirotius M., Ramel JY., Vincent N., Comparison of point selection for characterizing on-line signature. SPIE Defence and Security Symposium. Floride (USA). Mars 2005. CDROM.

2004:

[Wirotius04a] Wirotius M., Ramel JY., Vincent N., Selection of Points for On-Line Signature Comparison. IWFHR. Tokyo (Japon). p503-508. Octobre 2004.

[Wirotius04b] Wirotius M., Ramel JY., Vincent N., Improving DTW for online handwritten signature verification. ICIAR'04. International Conference on Image Analysis and Recognition. Porto. octobre 2004. .LNCS 3212. Part II. p786-793.

[Wirotius04c] Wirotius M., Ramel JY., Vincent N., New features for authentication by on-line handwritten signatures. ICBA'04. International Conference on Biometric Applications. Honk Hong. 14-17 juillet 2004. .LNCS 3072. p577-583

[Benadi04] Benadi S., Ramel JY, Prévot P.,AHXEL: Adaptive Hyperdocument using XML for E-Learning. 5th IEEE Int. Conf. on Information Technology Based Higher Education and Training: ITHET'04, Istambul. Turquie. 31 Mai - 2 Juin, 2004. CD-ROM

2003:

[Ramel03a] Ramel JY., Vincent N., Strategies for line drawing understanding. GREC'03. Fifth International Workshop on Graphics Recognition. Barcelona - Spain. 30-31 juillet 2003. p13-20.

[Ramel03b] Ramel JY., Crucianu M., Vincent N., Faure C., Detection, Extraction and Representation of Tables. ICDAR'03. International Conference on Document Analysis and Recognition. Edingburg - Scotland, August 2003. p374-378.

2002:

[Benadi02b] Benadi S., Ramel JY, Beuchot G., Using XML for the structuration of E-learning documents. 13th EAEEIE Conference on Innovation in Education

2001:

[Benadi01] Benadi S., Ramel JY, Beuchot G., Structuration XML de Documents Pédagogiques dans un système de télé-enseignement. Conférence internationale Human Computer Interaction (IHM-HCI'2001). Lille - France. 10-14 septembre 2001. Vol II. p123-126.

1999:

[Ramel99a] J.Y. Ramel, G. Boissier, H. Emptoz.. A structural representation adapted to handwritten symbol recognition. Third IAPR International Workshop on Graphics Recognition. Jaipur (India). 26-27 Septembre 1999. p259-266.

[Ramel99b] Ramel JY, Boissier G., Emptoz H., Automatic reading of handwritten chemical formulas from a structural representation of the image. International Conference on Document Analysis and Recognition, 20-22 Septembre 1999, Bangalore (India). Vol. 1. p83-86.

1998:

[Ramel98b] Ramel JY, Vincent N., Emptoz H., A new approach in contextual extraction of graphical entities in drawings. International Conference on Communication,

Computer and Power 7-10 décembre 1998 à Muscat (Sultanat d'Oman). p124-129.

1997:

[Ramel97a] Ramel JY, Vincent N., Brun JM., Bézier curves as a tool to describe kinetic drawings. In : Proceedings of the 4th International Conference on Document Analysis and Recognition (ICDAR'97). 18-21 août 1997 à Ulm (Allemagne). p780-783.

[Ramel97b] Ramel JY, Vincent N., Emptoz H., A coarse vectorization as an initial representation for the understanding of line drawing images. In : Proceedings of the 2nd International Workshop on Graphics Recognition (iapr/tc10). août 1997 à Nancy (France). p30-37.

1996:

[Ramel96a] Ramel JY, Vincent N., Emptoz H., Combining Global and Local Vision for Technical Document Understanding. In : Proceedings of the 13th International Conference on Pattern Recognition (ICPR'96). 25-30 août 1996 à Vienne (Autriche). p773-777.

1995:

[Ramel95] Ramel JY, Vincent N., Emptoz H., A hierarchical method of extraction and description of shapes in technical drawings. In : Proceedings of the International Conference on Image Processing and its Applications (IPA'95). 4-8 juillet 1995 à Edinburg (Ecosse). p379-383.

w **Conférences nationales avec comité de lecture et publication des actes**

2006:

[Qureshi06b] Qureshi RJ., Ramel JY, Cardot H., De l'appariement de graphes symboliques à l'appariement de graphes numériques : Application à la reconnaissance de symboles Conférence Internationale Francophone sur l'Ecrite et le Document. Septembre CIFED 2006. Fribourg (Suisse). A paraître septembre 2006.

[Hocquet06d] Hocquet S., Ramel JY, Cardot H., Fusion de Méthodes pour l'analyse de la dynamique de frappe. Reconnaissance de Formes et Intelligence Artificielle 2006. Tours. CDROM

2005:

[Journet05c] Journet N., Ramel JY, Eglin V., Mullot R., Caractérisation de la mise en page des documents imprimés de la Renaissance par une analyse des orientations. Colloque GRETSI, Septembre 2005, Louvain (Belgique), p122-129..

2004:

[Migeon04] Migeon C., Ramel JY, Renon S., Visio-enseignement » : une solution possible pour un enseignement coopératif et distribué. Colloque TICE Méditerranée 2004, Nice, Novembre 2004. CDROM.

2003:

[Bénadi03] Bénadi S., Ramel JY, Prévot P., Modélisation et Mise en Oeuvre d'Hyperdocuments Pédagogiques Adaptatifs. IHM'2003. Conférence Francophone sur l'Interaction Homme-Machine (ACM), Caen, France p32-39.

2002:

[Ramel02a] Ramel JY, Prévot P., Environnement multimédia pour augmenter la convivialité des examens d'IRM. IHM'2002. Conférence Francophone sur l'Interaction Homme-

Machine (ACM). Poitiers - France. novembre 2002. p259-262.

[Ramel02b] Ramel JY, Milliot S., Prévot P., Une proposition de structuration pour les systèmes et activités de Télé-enseignement. Colloque TICE'2002. Lyon - France. novembre 2002. p213-221.

[Billois02] Billois G., Bernard J., Ramel JY, Nait F., Milliot S., IOL : Une plate forme multimédia pour la diffusion d'enseignements synchrones sur Internet. Colloque TICE'2002. Lyon - France. novembre 2002. p355-356.

[Benadi02a] Benadi S., Ramel JY, Prévot P., Hyperdocuments pédagogiques adaptatifs : Modélisation et mise en oeuvre. CIDE'5. Colloque International sur le Document Electronique. Hammamet - Tunisie. octobre 2002. p139-152.

2000:

[Tarpin00] Tarpin-Bernard F., Ramel JY, Prévot P., Animation à distance de jeux d'entreprise multimédias coopéraztifs : Quelques recommandations. Ergo-IHM'2000. Biarritz - France. octobre 2000. p271-278.

[Ramel00b] Ramel JY, Prévot P., Environnement hypermédias pédagogiques : Quelques recommandations. Colloque international TICE'2000 : Technologie de l'Information et de la Communication dans les Enseignements d'ingénieurs et dans l'Industrie. Troyes - France. octobre 2000. p53-61.

1998:

[Ramel98d] Ramel JY, Vincent N., Emptoz H., Extraction contextuelle d'entités graphiques dans les dessins : du plus simple au plus complexe... In : actes du 1er Colloque International Francophone sur l'Ecrit et le Document (CIFED'98). 11-13 mai 1998 à Québec (Canada). p 453-462.

[Ramel98e] Ramel JY, Vincent N., Emptoz H., Une représentation utilisant les courbes de Bézier pour l'analyse de dessins manuscrits. In : Actes du 11e congrès Reconnaissance de Formes et Intelligence Artificielle (RFIA'98). 20-22 janvier 1998 à Clermont-Ferrand. Vol. 3. p123-132.

1996:

[Ramel96b] Ramel JY, Vincent N., Emptoz H., Coopération entre vision globale et analyse locale pour une meilleure interprétation des documents techniques. In : Actes du 4e Colloque National sur l'Ecrit et le Document (CNED'96). 3-5 juillet 1996 à Nantes. p117-124.

1994:

[Ramel94] Ramel JY, Vincent N., Emptoz H., La reconnaissance des partitions musicales. Actes du 3e Colloque National sur l'Ecrit et le Document (CNED'94). 6-8 juillet 1994 à Rouen. p 325-334.

w **DEA et thèse**

2005:

[Wirotius05] Matthieu Wirotius. Authentification forte par analyse biométrique de la signature manuscrite en ligne. Thèse de doctorat en Informatique soutenue à l'Université de Tours le 10 novembre 2005.

[Hublier05] Aurélien Hublier. Représentation et Localisation de symboles dans les images de documents techniques. Rapport de Master 2 Recherche. Université de Tours. 2005.

2004:

[Benadi04] Sofiane Benadi. Structuration des services et des données dans les systèmes d'enseignement à distance. Thèse de doctorat en Informatique soutenue à l'INSA de Lyon le 29 septembre 2004.

[Boutelja04] Nouha Bouteldja. Système multi-agents pour l'interprétation de dessins manuscrits. Rapport de DEA. Université de Tours. 2004.

[Leriche04] Stéphane Leriche. AGORA : un système d'analyse graphique adapté aux ouvrages de la Renaissance. Rapport de DEA. Université de Tours. 2004.

2003:

[Darnand03] Anne-Sophie Darnand. Reconnaissance automatique de gestes graphiques en ligne. Rapport de DEA. Université de Tours. 2003.

2002:

[Billois02] Gérome Billois. Scalabilité, mobilité et qualité de services dans les grands réseaux de transmission multicast : application au télé-enseignement. Rapport de DEA. INSA de Lyon. 2002.

2000:

[Benadi00] Sofiane Benadi. Technologies XML : concepts et applications. Utilisabilité dans le projet Web'Activ-Stratélia Rapport de DEA. INSA de Lyon. 2000.

1996:

[Ramel96] JY Ramel. Lecture automatique de dessins, méthodes d'analyse et de reconnaissance, application aux plans cinématiques. Thèse de doctorat (INSA de Lyon). 1996. 156 pages.

1993:

[Ramel93] JY Ramel. Lecture automatique de partitions musicales. Rapport de DEA. INSA Lyon. 1993. 61 pages.

w Conférences sans comité de lecture, rapports de recherche et Séminaires

2004 :

- *Bibliothèques Virtuelles Humanistes : Analyse de structure et transcription automatique* M-L Demonet (CESR-Tours), J-Y RAMEL (LI-Polytech'Tours), M-E Boutoue (IRHT-CNRS), F Lebourgeois (LIRIS/INSA-Lyon), S. Busson (CESR-Tours) Premiere rencontre Numérisation et Patrimoine. 21-24 juin 2004 à La Rochelle. 2004.

1998 :

- *Une nouvelle stratégie pour l'extraction d'entités graphiques dans les dessins.* J.Y. Ramel. Exposé dans le cadre des séminaires du laboratoire LI/E3I. Université Francois Rabelais de Tours. 2 avril 1998.
- *De l'image au sens : cas des dessins techniques.* J.Y. Ramel. Table ronde organisée durant le 11e congrès Reconnaissance de Formes et Intelligence Artificielle. 20-22 janvier 1998 à Clermont-Ferrand. 1998. Vol. 2. p467.

1997 :

- *Extraction, Gestion et Utilisation du Contexte en Reconnaissance de Documents Techniques.* J.Y. Ramel. Communication sous forme d'affiche dans le cadre des journées thématiques du GRCE : Journée "Interprétation de documents techniques et cartographiques". IGN (Paris). 2 avril 1997.

1996 :

- *Utilisation d'un mécanisme de "cycles perceptifs" pour l'interprétation de documents techniques.* J.Y. Ramel. Exposé dans le cadre des séminaires du laboratoire ERIC. Université Lumière (Lyon II). 7 octobre 1996.
- *Interface Homme-Machine intelligente : Reconnaissance automatique des intentions conceptuelles des concepteurs.* J.Y. Ramel, N. Vincent. Elément pour le rapport de fin de contrat «Conception Coopérante». Projet réalisé en commun par l'INSA de Lyon, l'UCBL, l'INPG et soutenu par la Région Rhônes-Alpes. février 1996. 15 pages.
- *Une nouvelle architecture pour l'interprétation des document techniques.* J.Y. Ramel. Présentation orale durant le 2e Forum "Images du centre". Ecole d'Ingénieurs en Informatique pour l'Industrie (E3i). 5 décembre 1996 à Tours.
- *Interprétation des document techniques.* J.Y. Ramel. Présentation sous forme d'affiche durant la Journée sur l'Ecrit et le Document organisée par le PRC Communication Homme-Machine (GT7). 3 juillet 1996 à Nantes.
- *Lire et comprendre les documents.* J.Y. Ramel. Présentation sous forme d'affiche au 2e colloque des doctorants de l'INSA de Lyon. 8 avril 1996 à Villeurbanne.
- *Coopération entre différents niveaux de vision : Application à la lecture de plans.* J.Y. Ramel. Présentation orale dans le cadre de l'action Inter-PRC 10.2 «Interaction système environnements pour l'interprétation des Signaux et des Images» (PRC ISIS & CHM). 22 mars 1996 à Paris.
- *Analyse de documents techniques : De l'image à l'interprétation.* J.Y. Ramel. Présentation sous forme d'affiche aux Carrefours de la Fondation Rhônes-Alpes Futur. 16 février 1996 à Charbonnière les Bains.

1995 :

- *La saisie automatique de schémas cimématiques mécaniques.* J.Y. Ramel. Présentation orale à la Journée des projets Régionaux : Modélisation pour la conception de produits. 10 mars 1995 à Voiron.

- *Influence des méthodes et traitements de bas niveau sur l'interprétation.* J.Y. Ramel. Exposé dans le cadre des séminaires internes à l'équipe. 7 septembre 1995.
- *Une méthode d'extraction et de description hiérarchique des formes adaptée à l'analyse de documents techniques.* J.Y. Ramel. Rapport de recherche interne n° 95.06. Equipe RFV. 1995. 15 pages.

1994 :

- *Analyse automatique de documents techniques.* J.Y. Ramel. Présentation sous forme d'affiche au 1er colloque des doctorants de l'INSA de Lyon. 21 avril 1994 à Villeurbanne. p195-196.

P a r t i e 2

Introduction

1. Préambule

Mon cursus scientifique se caractérise par trois périodes correspondant chacune à une activité de recherche bien identifiée : l'analyse de documents graphiques, les interfaces homme-document multimédia et l'indexation de documents peu structurés. Ce rebouclage thématique vers l'analyse de documents après être passé par des applications plus directement accessibles aux utilisateurs, en l'occurrence les environnements pour l'édition et la consultation de documents multimédias, me permet aujourd'hui d'avoir une vision de ce qu'est un document numérique à partir de deux points de vue bien différents. Il me semble intéressant d'essayer de confronter ces deux visions pour proposer de nouvelles approches permettant de mieux appréhender les documents numériques. C'est ce que je tente de faire dans ce mémoire en retraçant mon parcours scientifique et en développant les problématiques auxquelles j'ai été confronté depuis mon arrivée dans le monde de la recherche :

- la spécification de représentations abstraites permettant une description du contenu des images de documents de façon à pouvoir les analyser tout en restant suffisamment générique,
- la mise en place d'algorithmes de construction automatique de ces représentations abstraites à partir des données directement fournies par les dispositifs d'acquisition en hors ligne (images) et en ligne (suite de points horodatés),
- la proposition d'architectures et de méthodes d'analyse et de reconnaissance utilisables sur ces représentations et permettant une interprétation du contenu des documents,
- la proposition d'une structuration des données et des services pour les applications de gestion de documents multimédias de façon à mieux formater et présenter les informations demandées par les utilisateurs (hyperdocuments et interfaces adaptatifs pour le télé-enseignement).

Une autre caractéristique importante de mes travaux est le choix d'intégrer et d'évaluer sur le terrain les résultats fondamentaux obtenus notamment par le biais de prototypes. Ainsi, la plupart de mes recherches se sont situées dans le cadre de projets précis et se proposent d'apporter une réponse viable aux besoins exprimés par les utilisateurs.

Après une introduction précisant comment il est possible de définir la notion de document numérique et quels sont les traitements pouvant lui être appliqués, ce mémoire est organisé en trois chapitres principaux correspondant aux axes de recherche précédemment cités :

- les modes de construction des représentations utilisables pour les images de documents et les tracés en ligne,
- les techniques d'analyse et de reconnaissance des contenus,
- les interactions Homme-Documents numériques.

2. Qu'est ce qu'un document ?

Lors de la conception d'un système d'information manipulant des documents, il me semble crucial de souligner l'importance d'effectuer une bonne modélisation des données afin de pouvoir mieux les traiter par la suite. Une bonne modélisation signifie pour moi que les représentations choisies soient bien adaptées et donc suffisamment génériques pour pouvoir répondre aux différents besoins exprimés. Ces représentations doivent, de plus, être structurées et organisées de façon à constituer des ressources facilement accessibles et modifiables par les processus de traitement. Pour construire de telles représentations, il est indispensable d'étudier minutieusement les caractéristiques des données initiales (contenu des documents et environnements associés) et de bien identifier les objectifs à atteindre (besoins des utilisateurs finaux). Ces derniers jouent d'ailleurs un rôle primordial dans le cycle de vie d'un document et donc dans les chaînes de traitement qui doivent être mises en place.

J'essaie dans cette partie de synthétiser les caractéristiques fondamentales concernant les documents numériques et la structuration des informations qu'ils contiennent.

2.1. Définitions

À l'ère des hypermédias, des DVD, des téléphones mobiles, la notion de document devient difficile à définir car on ne peut plus se référer à des repères précis. En effet, différentes caractéristiques, parfois contradictoires, peuvent être considérées pour établir une définition du document. En voici quelques unes souvent rencontrées [Roisin99, Tran02, Blasselle97, LeMaitre00], chacune ne couvrant que partiellement la notion de document.

<u>Un document est un regroupement d'éléments de contenu sur un support</u>

Un document peut être défini par son support physique. L'exemple classique est celui du livre présenté sur support papier sous la forme d'une suite de pages et identifié par un numéro unique fourni par l'ISBN. On se rapproche également de cette notion traditionnelle de document dans le cas où le support

électronique représente l'information sous forme d'un codage en pages numériques (Postscript, pdf) ou d'images de pages scannées.

Ce qui compte ici est la composition graphique et spatiale du document appelée aussi sa structure physique ou sa forme. Cette présentation spatiale des documents s'appuie d'une part sur une expression des propriétés de style, et d'autre part sur un processus de formatage dont le résultat est un document formaté : les éléments le constituant, que nous appellerons dans la suite « éléments de contenu (EdC) », sont alors disposés physiquement en fonction du support de sortie. Le terme élément de contenu doit ici être pris dans le sens d'une entité élémentaire contenue comme les éléments textuels ou graphiques, les images, voire les sons ou les vidéos. Outre la difficulté à définir le niveau adéquat de granularité de ces objets élémentaires, cette définition n'est pas directement applicable à l'information disponible sur le web du fait des hyperliens puisque, dans ce cas, il n'existe alors plus de limites physiques tant pour le support que pour les EdC.

Un document est un ensemble organisé d'EdC : Structure logique et sémantique

Avec cette définition, il devient nécessaire d'identifier des éléments intermédiaires qui définissent des regroupements logiques d'objets (des paragraphes formant une section par exemple). On parle alors de structure logique du document (ou plus simplement de sa structure). Cette approche peut être élargie pour couvrir les besoins de l'hypertexte pour lesquels une organisation de liens doit être ajoutée à l'organisation hiérarchique des informations. Par exemple, on peut utiliser les liens comme support de relations sémantiques entre composants ou pour fragmenter les informations documentaires de grande taille (exemple : une explication éclatée en un ensemble de pages HTML).

La structure d'un document a pour fonction d'organiser le contenu de façon à en simplifier la lecture. Elle apporte une information implicite qui indique au lecteur la fonction de chaque EdC. La structure traduit aussi l'intention de l'auteur, elle donne un éclairage particulier en mettant intentionnellement en évidence certaines parties. Elle participe donc fortement à la définition du contenu informatif du document. Tout document a pour but de véhiculer un message, son contenu sémantique, qui ne peut être facilement et complètement compris que si la structure logique a été correctement appréhendée ou mise en place.

La représentation de la structure logique s'appuie la plupart du temps sur des opérateurs de composition (comme l'agrégation, la liste et le choix) [Roisin99, Bonhomme98, Akpotsui97] et sur 3 types d'entités :

- Les éléments de base non décomposables qui constituent le contenu. Ici, la notion d'EdC, utilisée également dans la définition précédente, n'est plus caractérisée par ses propriétés physiques mais plutôt par sa sémantique ou tout au moins par le contexte environnant (son voisinage),
- Les éléments composites obtenus par composition d'éléments de base ou d'éléments composites,
- Les attributs qui peuvent être associés aux éléments pour leur adjoindre des informations sémantiques ou autres.

Ces principes permettent d'organiser les documents selon une structure hiérarchique d'éléments typés, décrite linéairement comme un emboîtement de marques (ou balises) encapsulant les éléments de base. Un tel document peut alors être représenté sous forme XML et être transformé avec l'aide des technologies gravitant autour de ce standard (XSLT, XML Schéma, …) [Michard98].

Un document est un enchaînement temporel d'EdC

Cette définition permet de caractériser les présentations multimédias dans lesquelles des objets médias (son, texte, image, vidéo) sont organisés temporellement. L'introduction de la dimension temporelle dans les documents conduit nécessairement à de nouveaux besoins d'expression tels que la durée des objets, leur placement temporel et leur synchronisation. Ces informations temporelles doivent s'intégrer dans l'ensemble des informations attachées aux documents. Par exemple, pour spécifier le déplacement d'un stylet sur un support d'acquisition ou d'un objet sur l'écran, il est nécessaire d'exprimer une information spatiale pour la trajectoire, comme les positions initiales et finales correspondant au déplacement, ainsi qu'une information temporelle comme la durée du déplacement et sa synchronisation avec d'autres objets ou actions.

Les acteurs humains ou logiciels peuvent aussi agir sur les documents ou tout au moins sur leur représentation à un instant donné. Un document est alors non seulement défini par un contenu mais aussi par les traitements qu'il est possible de lui appliquer. Cette définition englobe les problématiques de compositions dynamiques des documents hypertextuels. Par exemple, l'activation d'un lien hypertextuel peut être vue comme une action d'un utilisateur entraînant une réaction de la part du document. On parle alors de documents numériques adaptatifs.

2.2. Représentations des documents numériques

Ces multiples facettes constituent une difficulté pour la mise en place de modèles généraux qui intègrent les différents besoins de traitements des documents. C'est donc en premier lieu un travail de modélisation qu'il faut effectuer. Dans les domaines des IHM (Interfaces Homme-Machine) et des documents multimédias, on est ainsi passé de représentations sous forme simple (texte ascii) à des modèles riches à base de multiples structures (arbres, technologies XML, …). Il paraît intéressant d'essayer de généraliser ce genre d'approches à l'analyse de documents numériques en vue du traitement ou de la reconnaissance automatique de leur contenu.

Depuis de nombreuses années, les documents électroniques ont fait l'objet d'études qui ont conduit à l'identification de caractéristiques attachées aux documents et classées selon différentes dimensions. Pour chacune de ces dimensions, le travail de modélisation consiste à identifier d'une part les entités de base, et d'autre part, leur mode de composition qui doit permettre de couvrir l'ensemble des besoins liés aux traitements qui devront être réalisés sur ces structures. On parle même actuellement de multi-

représentations des documents [Chatti06]. Comme nous le verrons dans la suite, mes travaux se sont focalisés sur les représentations sous forme de graphes ou d'arbres au travers desquels je pense qu'il est possible de construire une description des contenus des documents pouvant couvrir le domaine de l'édition et de la consultation, mais pouvant aussi être étendue au domaine du traitement d'images de documents. Les éléments constituant le document (dont la granularité et le type restent à définir) sont alors représentés par les nœuds d'un graphe tandis que les arcs du graphe décrivent les relations (topologiques, temporelles, imbrication, …) existant entre ces nœuds Des attributs peuvent ensuite être ajoutés aux arcs et aux nœuds afin de les caractériser aussi finement que nécessaire physiquement (forme) et sémantiquement (fond) de manière statique ou dynamique au cours des traitements. Ce choix est complètement en adéquation avec l'usage de plus en plus courant d'arbres XML pour la production et le traitement des documents multimédias (par transformations successives). Dans tous les cas, il est nécessaire de mener une réflexion sur ce que seront les éléments (nœuds) entrant dans la composition du graphe et sur ce que caractériseront les arcs liant ces nœuds.

3. Comment structurer les traitements ?

3.1. Sources d'inspiration

Nous assistons aujourd'hui à de nombreux échanges entre les sciences informatiques et les sciences humaines : pour les uns, l'Homme constitue une référence par rapport à la machine, qu'il serait souhaitable d'égaler, voire de dépasser, dans la mesure où la machine doit réaliser des tâches humaines. Pour les autres, il s'agit, à l'aide d'ordinateurs et de modèles, de valider certaines théories sur les fonctionnements cérébraux.

La question qui se pose alors est : faut-il se référer aux mécanismes humains lorsqu'on désire mettre en place une chaîne de traitements dédiée aux documents ? La réponse est multiple : d'une part, les problématiques à résoudre sont largement basées sur des problématiques humaines (vision, reconnaissance de formes, lecture,...) ; d'autre part, si nous parvenons aujourd'hui à résoudre correctement certains problèmes précis en vision artificielle, les réponses apportées sont rarement généralisables et nous manquons alors de solutions globalement satisfaisantes. Par opposition, les mécanismes humains remplissent ce critère de généralisation et deviennent alors une référence indéniable.

Des travaux antérieurs indiquent qu'il est possible d'opposer la notion de perception visuelle qui correspond à des mécanismes de type réflexe, à la notion de vision (lecture) qui peut supposer une élaboration intelligente sous forme de concepts de ce qui est perçu [Duperthuy97]. Le processus de perception peut alors être vu comme l'élaboration d'une représentation initiale du contenu d'un document de la même manière que David Marr {Marr82] définit son esquisse initiale sur les images naturelles.

Ainsi replongé dans le modèle de Marr, l'esquisse initiale correspond à une première étape d'extraction d'indices visuels bi-dimensionnels à partir d'une image. Cela suppose que la représentation obtenue de la scène soit générale (elle ne doit pas dépendre, en principe, de l'objectif poursuivi) et complète (elle doit représenter tous les éléments importants de la scène). L'objectif visé n'est pas de comprendre ce qui est vu, mais simplement de mettre en forme les informations pertinentes acquises.

Mais, la distinction entre perception visuelle et vision n'est pas aussi tranchée qu'elle n'y paraît puisque, d'un point de vue neuro-anatomique, il n'est pas possible de séparer clairement les différents modules fonctionnels du cerveau en des modules de type réflexe et des modules intelligents. Ainsi, le modèle de Marr a été de nombreuses fois critiqué principalement à cause de l'aspect trop séquentiel des traitements, de la non prise en compte des connaissances a priori ou du contexte (objectif et point de vue). On l'oppose donc souvent à la notion de vision active [Hudelot05] qui considère le processus de vision comme une activité exploratoire. On s'intéresse alors autant à la manière dont sont traitées les informations qu'à leur type. Les premières aires visuelles du cerveau semblent jouer un rôle d'extraction et de répartition des informations en provenance de la rétine et à destination d'aires plus en aval, gérant des informations de plus en plus complexes [Duperthuy97]. De plus, le modèle se complique encore en constatant que toutes ces aires ne fonctionnent pas en cascade (du bas niveau vers le haut niveau) mais qu'il existe aussi des connexions en retour des aires d'ordre supérieur vers les premières aires. Ainsi, le traitement de l'information visuelle n'est pas réalisé de manière séquentielle mais en parallèle et en continu par un ensemble de modules spécialisés [Bullier95]. Ce constat est à l'origine de l'hypothèse selon laquelle, une fois l'image éclatée suivant plusieurs caractéristiques, il n'existe pas de lieu centralisant à nouveau ces informations pour dire que tel objet de telle couleur se déplace dans telle direction mais que l'acquisition des images serait continue afin d'activer et de synchroniser les aires visuelles jusqu'à l'obtention d'un tout cohérent. On peut alors considérer le système de vision comme un système qui relève d'un ensemble de processus permettant de faire évoluer une représentation.

3.2. Traitements incrémentaux, contextuels et interactifs

Il a été indiqué précédemment qu'une représentation structurelle des contenus des documents paraissait satisfaisante et générique puisqu'adaptable à la plupart des types de documents numériques (hypermédias, images de documents graphiques, composites et tracés en ligne). Nous avons également noté, lors de nos références au système visuel humain, qu'une spécialisation des traitements d'analyse peut être efficace lorsque ceux-ci ne restent pas complètement séquentiels et indépendants les uns des autres. De ce point de vue, le traitement d'un document comme l'interprétation d'une scène correspond à des transformations successives appliquées aux informations acquises et mises en commun via une ou plusieurs représentations.

Le problème de l'interprétation (traitement) réside alors dans l'identification d'un ensemble de traitements spécialisés mais interdépendants permettant d'obtenir des applications plus génériques ou

tout au moins plus facilement adaptables en séparant bien les données initiales, les connaissances a priori, et les algorithmes de transformation / reconnaissance.
La (ou les) représentation(s) (structurelles ou autres) devra (devront) permettre de gérer et d'utiliser facilement les informations dont on dispose sur un document numérique à un instant donné. Ces représentations devront utiliser un nombre limité de primitives (élémentaires au début puis de plus en plus évoluées) pour fournir une description précise de toutes les parties composant le document. Mais, de plus, cette structure de description doit aussi pouvoir évoluer en même temps que l'interprétation de manière à correspondre constamment à l'état courant d'analyse.
L'autre point principal ayant retenu mon attention est la capacité de l'esprit à se raccrocher aux informations déjà connues (données fiables) pour mieux interpréter ce qui paraît moins évident à un instant donné. Pour reproduire cette faculté d'analyse incrémentale, il est possible de faire en sorte que les processus d'analyse s'efforcent d'extraire les nouvelles connaissances en prenant un minimum de risque (analyse incrémentale). Ces dernières pourront alors être utilisées pour résoudre les problèmes les plus complexes à partir de données non ambiguës (analyse contextuelle). L'utilisation conjointe d'une représentation évolutive pour le partage de l'information, et de plusieurs processus réalisant chacun une tâche précise doit simplifier la reconstruction d'un tout cohérent correspondant aux contenus des documents.

Pour ce qui est de la production et de la consultation de documents hypermédias, en plus des EdC, on parle de services logiciels ou composants (widgets) mais l'idée de transformations successives et de granularité des contenus des documents reste globalement identique. Nous avons déjà dit qu'un document numérique pouvait être vu de différentes manières : notion de « niveaux d'analyse » ou de « niveaux de structuration » (voir section « Qu'est ce qu'un document ? »). Au niveau le plus bas, les documents numériques peuvent être vus comme un assemblage d'objets élémentaires ou primitives sur un support physique (niveau support). Ensuite, il est possible de s'intéresser à l'enchaînement et aux liens existant entre les EdC (niveau structure). Enfin, on peut s'intéresser au sens et au contenu véhiculés par le document (niveau sémantique). Dans le cadre de mes recherches en IHM, j'ai montré qu'il est possible de tirer parti de cette structuration en 3 niveaux pour simplifier ou améliorer la production et la consultation des documents hypermédias. Nous verrons une nouvelle fois que la principale difficulté réside dans le choix de la granularité des objets considérés comme éléments de base constituant un document. Cette décision correspond au choix des primitives à utiliser pour bien représenter le contenu d'un document et permettre les transformations et descriptions adéquates.

Les chapitres suivants montrent quelles réponses peuvent être apportées à différents verrous scientifiques tant dans le cadre de l'analyse d'images de documents que dans le cadre de la production de documents hypermédias adaptatifs lorsqu'on essaie d'appliquer les principes exposés ci-dessus. Ainsi, dans les chapitres 1 puis 2, j'essaie de démontrer quels sont les intérêts de cette manière de procéder en la

comparant aux méthodes traditionnellement utilisées pour segmenter puis reconnaître le contenu d'une image de document ou le sens de tracés en ligne. Le chapitre 3 explique comment ces propositions contribuent à simplifier la production et la consultation des documents numériques adaptatifs. Pour chaque cas, je tenterai également de préciser les nombreuses voies de recherche que j'aimerais explorer dans le futur.

Chapitre 1 :
Représentation de documents
(images et tracés en ligne)

Une grande partie de mes travaux de recherche concerne l'analyse d'images de documents graphiques, textuels et tracés en ligne. Dans la première partie de ce chapitre, les techniques classiquement utilisées pour caractériser ou segmenter une image de document (traitement d'images) sont décrites en détail. Ensuite, nous montrons l'intérêt de ne plus envisager cette étape comme une étape classique de segmentation ou de caractérisation d'un type précis de contenu mais comme une phase de construction d'une représentation initiale suffisamment générique (esquisse initiale). Pour terminer le chapitre, des exemples d'utilisation de la représentation préconisée seront présentés.

1.1. Caractéristiques des images de documents et des tracés en ligne

Un document a normalement pour but de transmettre un ensemble d'informations à des destinataires suivant une symbolique prédéfinie. Ainsi, dans le domaine de l'analyse de documents, le problème essentiel est d'extraire de l'image l'information graphique et alphanumérique afin de reconstituer la sémantique associée aux différents EdC et à leurs relations. Pour cela, il paraît intéressant d'étudier les caractéristiques principales des documents que l'on désire analyser. Notons dès à présent que ce mémoire ne traite que des images en niveaux de gris (voire binaires) et que la problématique de la couleur n'est pas abordée.

1.1.1. Documents graphiques

Les documents graphiques forment une catégorie particulière de documents : ils sont constitués de lignes, de régions pleines, de régions hachurées, de texte, ... Ils contiennent toujours une quantité importante d'informations, peuvent être d'une grande complexité et leur domaine d'application est très étendu. Chaque type possède une structure qui lui est propre, choisie de manière à ce que les informations qu'il véhicule soient aisément compréhensibles et réutilisables par les destinataires (figure 1.1).

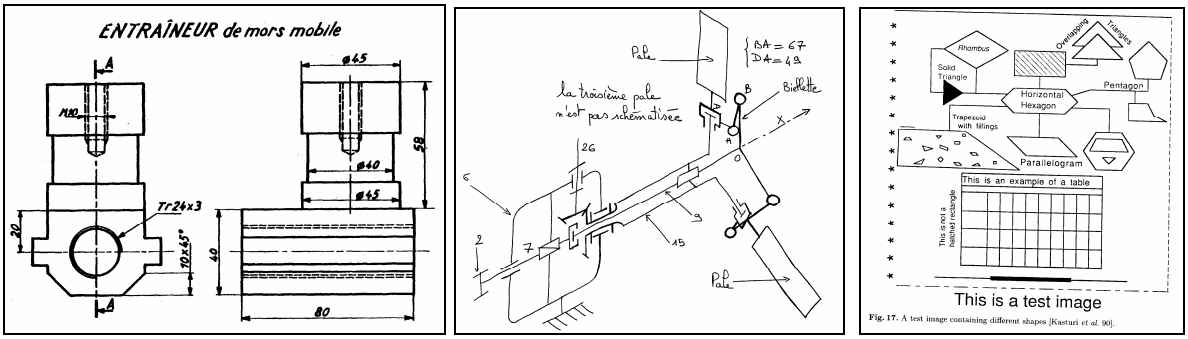


Figure 1.1 : Exemple de documents graphiques

Il peut sembler hasardeux de tenter de définir une structure commune à tous ces documents généralement peu structurés. Une étude des différents types de dessins permet cependant de mettre en évidence un certain nombre de points communs.

Il est la plupart du temps possible de mettre en évidence une organisation en couches de tous ces documents [Tombre92, Ah-Soon01, Cordella00]. Cependant, ce découpage peut se faire suivant différents critères qui aboutissent à des approches complètement différentes lors de l'analyse du contenu du document.

Si l'on se positionne d'un point de vue logique ou sémantique, on peut rechercher des couches qui dépendent fortement du domaine d'application auquel on s'intéresse. On trouvera par exemple :

- une couche correspondant aux symboles (comme les transistors ou les diodes dans les schémas électroniques)
- une couche constituée des liens mettant en relation ces symboles
- une couche « lignes de portée » dans les partitions musicales
- une couche « légende » qui comprend uniquement les traits et les zones textuelles précisant la cotation des objets.

La position des EdC les uns par rapport aux autres est une information cruciale. Par exemple, la position du texte par rapport aux graphiques indique à quelle partie du dessin se rapporte une légende. Il est donc important de ne pas perdre ce type d'information et d'être à même de l'utiliser facilement durant la phase de classification. Le vocabulaire utilisé est souvent, pour chaque type de dessins, relativement réduit et les caractères composant le texte sont plus ou moins aisément séparables par des critères topologiques selon la nature du dessin. Si on conçoit une chaîne de traitement à partir de ce découpage, on aboutit à un système dédié uniquement à un type précis de documents et qui fonctionne plus ou moins bien selon :

- les ambiguïtés pouvant exister au sein des conventions ou des règles qui sont définies dans différentes normes régissant la manière de dessiner dans le domaine d'application concerné
- la complexité de traduction, de mise en place et d'utilisation des règles définissant la structure pour guider l'analyse
- la rigueur avec laquelle les dessins sont exécutés (respect des règles)

- la qualité des images et des algorithmes effectuant l'analyse

Il est à noter qu'en plus des règles lexicales et syntaxiques, il serait souvent possible d'utiliser des informations sémantiques concernant, par exemple, la cohérence du « système » reconstruit pour aider ou plutôt pour valider l'analyse du contenu du document. Ces informations sont potentiellement utilisables durant les différentes phases du traitement mais, jusqu'à présent, peu de systèmes arrivent à les exploiter autrement que pour valider a posteriori des hypothèses d'interprétation proposées.

Il est également possible d'utiliser des critères purement physiques ou géométriques pour décomposer le document en couches. Ces dernières traduisent la structure physique du document qui se compose alors par exemple :

- d'une couche constituée de traits fins rectilignes
- d'une couche constituée de traits fins non rectilignes comme les courbes, les arcs et les cercles. Une partie de ces tracés peut représenter physiquement un mécanisme ou un objet ou avoir une signification purement symbolique. Une autre partie de ces traits peut faire référence à la cotation des objets représentés.
- d'une couche « formes pleines » correspondant également à des parties de symboles ou à des parties de la cotation (têtes de flèches, disques, ...)
- d'une couche « texte » qui se présente généralement sous la forme de groupes de composantes connexes de petites tailles mais qui peut se confondre avec la couche « traits fins ».

Contrairement au découpage précédent, celui-ci s'applique quel que soit le type de documents (ou l'application visée) et les algorithmes effectuant un tel découpage permettent de rester générique. Malheureusement, arrêter l'analyse à ce niveau (extraction de primitives de description) est insuffisant ; d'autres traitements doivent venir compléter l'interprétation et il devient alors difficile de rester générique.

Selon ces deux points de vue, le problème de l'interprétation automatique de document semble résider dans la phase de séparation puis de reconnaissance de chacune de ces couches et dans le choix de l'ordre d'extraction le plus approprié. Ce qui n'est pas sans difficulté car, comme nous venons de le voir, toutes les couches sont constituées d'objets géométriques eux-mêmes formés à partir de primitives d'un même type (arcs et segments). Néanmoins, les remarques précédentes me paraissent justifier qu'un système d'analyse de documents graphiques manipule des entités de type structurel formées à partir de formes pleines, de segments de droite et d'arcs en relation ou connectés les uns avec les autres. Le traitement de symboles graphiques complexes fait aussi apparaître un besoin de modéliser et reconnaître les symboles à différents niveaux si l'on veut rester générique le plus longtemps possible. Pour des symboles complexes, l'approche classique (plutôt statistique) qui consiste à apprendre un modèle par symbole n'est plus adaptée. Il est nécessaire de voir les symboles complexes comme des assemblages de symboles

plus simples. Ceux-ci peuvent être vus comme des compositions de symboles graphiques élémentaires d'où l'intérêt d'utiliser une représentation évolutive ou plusieurs représentations successives du contenu de l'image. Ces nouvelles représentations et ces nouveaux modèles, hiérarchiques ou incrémentaux, qui permettent d'une part d'exprimer et d'autre part d'apprendre ce type d'information restent encore aujourd'hui à développer.

Rappelons enfin que l'analyse porte autant sur la position relative des objets, sur les liens qui existent entre eux que sur leur reconnaissance.

1.1.2. Documents textuels

Les documents majoritairement textuels sur lesquels se sont focalisées mes recherches sont les documents anciens remontant aux débuts de l'imprimerie. Les polices utilisées, la présentation des pages et l'utilisation de l'espace étaient alors très proches de celles des ouvrages manuscrits. Les langues utilisées, le latin ou le français, amènent un facteur supplémentaire de diversité pour ces ouvrages. Les typographies médiévales ont une variabilité de forme importante. Des exemples d'images d'ouvrages anciens sont présentés figure 1.2.

Pour la mise en page, les contraintes techniques de l'époque imposent des présentations particulières. Les irrégularités des mises en page sont beaucoup plus nombreuses que dans les ouvrages actuels à cause d'imprécisions ou de libertés prises par l'imprimeur. Néanmoins, la plupart du temps, un corps de texte occupe la majorité de la page avec des notes en marge de chaque côté du texte. La page peut aussi contenir des illustrations graphiques de différentes tailles, des ornements et des lettrines. Dans le texte, on retrouve des structures connues comme les titres et les sous-titres, les paragraphes, les numéros de page, et d'autres plus particulières comme les réclames. Une autre particularité provient des faibles séparations entre les blocs de texte (notes en marge et corps du texte par exemple). Enfin, sur certaines pages, les règles de mise en page d'aujourd'hui ne sont pas mises en œuvre : par exemple, une illustration peut déborder sur les marges (figure 1.2). Dans les ouvrages de la Renaissance, les illustrations ont généralement été imprimées à l'aide de plaques de bois ou de métal, gravées avec le motif à reproduire et encrées. Ce type d'illustrations est très différent des photographies que l'on peut trouver aujourd'hui majoritairement dans les documents composites. D'autres particularités sont dues aux procédés de numérisation employés. Elles concernent les défauts d'éclairage dans la reliure, la courbure des lignes de texte, l'inclinaison des pages, l'élimination imparfaite des taches.

Figure 1.2 : Exemples de pages d'ouvrages

Ces quelques remarques permettent de dresser une liste de caractéristiques dont il est indispensable de tenir compte lors de la conception d'algorithmes visant à construire une représentation du contenu de tels documents. Comme nous l'avons fait pour les documents graphiques, voici donc les connaissances a priori qui peuvent être utilisées lors de la mise en place d'algorithmes d'analyse « bas niveau » :

- Chaque page est constituée d'un ensemble de zones textuelles et/ou graphiques juxtaposées (structure physique),
- Présence de notes en marges imprimées ou manuscrites,
- Faible espacement entre les lignes provoquant des contacts entre les caractères,
- Espacements entre caractères, mots et blocs de texte non constant,
- Images dégradées même après restauration (apparition des caractères du verso),
- Mise en page plus ou moins complexe qui peut présenter plusieurs colonnes avec des tailles de corps et d'interlignes irrégulières (structure logique),
- Pas de charte éditoriale ou de structure logique toujours identifiable.

La typographie et la technologie de l'imprimerie ont fait énormément de progrès et les ouvrages actuels répondent à des normes bien différentes en matière de présentation. Les logiciels conçus pour reconnaître les documents actuels s'avèrent, en conséquence, bien souvent médiocres pour le traitement des ouvrages anciens. A l'inverse, il semble qu'un système capable d'appréhender des documents anciens pourra facilement interpréter les documents contemporains. La seule chose à vérifier sera le comportement des algorithmes vis à vis des photographies (éléments de contenu particuliers n'apparaissant jamais dans les documents anciens).

1.1.3. Tracés en ligne

Aujourd'hui, de nouveaux dispositifs d'interaction Homme-Machine (figure 1.3) apparaissent sur le marché (PDA, téléphone portable, tableau interactif, papier électronique, ...) à des prix attrayants. Ceux-

ci ont tendance à remettre en question l'utilisation de la souris et du clavier (problème d'ergonomie mais aussi d'encombrement nuisant à la mobilité) en proposant de les remplacer par des périphériques plus habituels et plus transportables comme les stylos électroniques.
Pour l'instant, dans la plupart des cas, les nouveaux dispositifs d'interaction proposés se contentent de se substituer à la souris car les applications n'ont pas encore été adaptées pour tirer parti des avantages liés à la reconnaissance automatique de gestes manuscrits.

Figure 1.3 : Nouveaux dispositifs : papier électronique, tableau interactif [Schilit98]

Pourtant, probablement parce qu'il reste de nombreux progrès à faire dans le domaine de l'analyse de documents et de l'écriture pour obtenir une interaction Homme-Machine satisfaisante, la demande de fonctionnalités de reconnaissance automatique revient très souvent chez les acquéreurs qui ne comprennent pas pourquoi on ne répond pas à leur besoin [Long99, Long00] :

- pouvoir exprimer rapidement des idées sous forme visuelle,
- retrouver l'intégration naturelle entre la gestuelle graphique et la pensée : tant qu'on commencera par apprendre à écrire et à dessiner avec un stylo sur du papier à l'école, ce dernier restera un support privilégié.

Il semble que les propositions actuelles pèchent par leur difficulté d'utilisation (apprentissage long, gestes en petit nombre, non intuitifs, difficiles à mémoriser). D'autres applications pourraient même se démocratiser si on réussissait à améliorer les taux de reconnaissance de documents manuscrits en ligne, notamment autour de l'interaction Homme-Documents :

- édition, correction, annotation de documents électroniques (figure 1.4) [Andre99],
- utilisation de tableaux interactifs et intelligents,
- reconstruction de version idéale à partir du brouillon [Moran95, Julia96].

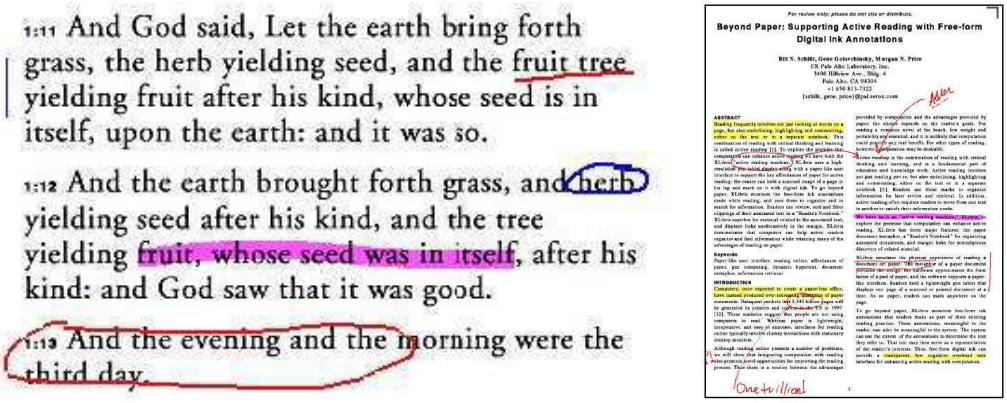


Figure 1.4 : Correction, annotation de documents électroniques [Schilit98]

En ligne, les données (provenant la plupart du temps d'une tablette graphique ou d'un stylo électronique) correspondent à une liste de points horodatés (donc ordonnés). A chacun de ces points peuvent être associées d'autres informations comme la pression exercée par le stylo ou l'inclinaison du stylo vis à vis de la tablette. Il est également possible d'obtenir des informations sur la vitesse et l'accélération du stylo pendant le geste. Un tracé en ligne ne peut pas correspondre à une forme pleine et n'a pas d'épaisseur. Néanmoins, le paramètre «pression» peut parfois être assimilable à une épaisseur. Le tracé peut être plus ou moins hésitant et peut avoir été réalisé avec plus ou moins d'application. De même sa taille peut être très variable (échelle) ce qui complique la reconnaissance.

Ainsi les données initiales, en ligne, sont plus structurées qu'en hors ligne et les principales différences entre les données initiales en ligne (suite de points) et en hors ligne (image) résident dans la présence d'informations temporelles supprimant les problèmes de reconstruction de la chronologie des tracés auxquels on est confronté lors de l'analyse d'images statiques.

1.2. Extraction de primitives, caractérisation et segmentation

Au plus bas niveau de la chaîne usuelle de traitement d'images de documents et de tracés en ligne se trouvent les algorithmes opérant directement sur les données acquises (pixels de l'image ou listes de points) pour en extraire des informations plus structurées, telles que le suivi de contours, la vectorisation, la séparation texte/graphique, l'analyse de texture, ... On découpe souvent les principales thématiques de l'analyse d'image de document en la reconnaissance de l'écriture manuscrite, l'interprétation de documents structurés et l'interprétation de documents graphiques… En pratique, il me paraît inopportun de les dissocier. En effet, les documents peuvent combiner des parties manuscrites et graphiques, avec des éléments propres aux documents structurés [Kasturi92, Nagy00]. Divers outils spécifiques, dédiés à cette phase d'extraction de primitives de description, ont été développés de façon à traiter chaque type de formes de la manière la plus adéquate possible. Si on considère autant les travaux en analyse de

documents textuels que les travaux en analyse de documents graphiques, les méthodes utilisées peuvent, à mon avis, être découpées en deux catégories :

- les méthodes cherchant à construire des représentations structurelles à partir des contours des formes,
- les méthodes cherchant à caractériser les régions ou les formes à l'aide de signatures statistiques.

Dans la suite, nous faisons un bref rappel des principales méthodes pour chacune de ces approches.

1.2.1. Primitives structurelles et segmentation

L'objectif principal des méthodes classées dans cette première catégorie est de transformer les données brutes fournies par les dispositifs d'acquisition en des entités plus facilement manipulables durant la suite des traitements. Cette section explique donc comment et pourquoi les techniques de filtrage, les techniques d'extraction de composantes connexes, l'analyse de projections et la vectorisation sont majoritairement utilisées sur les images de documents.

Filtrage

Le problème de la segmentation physique d'un texte en blocs, puis des blocs en lignes, puis des lignes en mots, puis des mots en signes peut aussi être formulé en termes d'analyse morphologique, de filtrage ou de lissage. Ce type de méthodes a été beaucoup mis en œuvre et testé sur les ouvrages imprimés pour la fusion des caractères en mots puis en lignes et pour l'élimination du bruit [Lebourgeois99]. Il est également possible d'effectuer des opérations locales appelées nettoyage et bouchage qui, par étude du voisinage d'un point à l'aide de masques définissant les configurations de pixels à transformer, améliorent la qualité de l'image [Chen95, OGorman95]. La figure 1.5 illustre les résultats pouvant être obtenus à l'aide de ce type de méthodes.

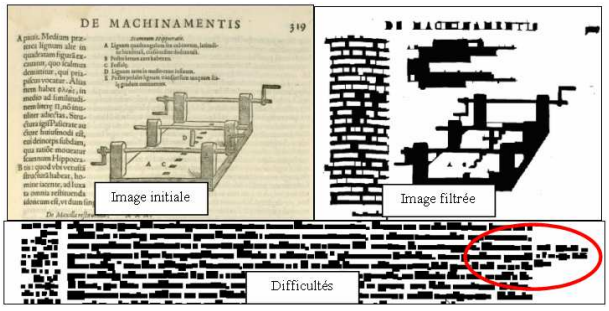


Figure 1.5 : Filtrage morphologique sur une image binarisée

Le classique algorithme RLSA (Run Length Smearing Algorithm) [Wong82] très souvent utilisé fait partie de ces méthodes. Cet algorithme cherche à noircir les plages blanches de longueur inférieure à un seuil S, comprises entre deux pixels noirs. On obtient alors des blocs noirs continus à l'emplacement de chaque ligne. Wong, Casey et Wahl appliquent ce lissage horizontalement et verticalement sur l'image,

puis ils font un ET logique entre les deux matrices obtenues pour obtenir les EdC d'un document. Des variantes ont été introduites [Tsujimoto92] pour éviter de fixer des seuils arbitraires sur le paramètre S. Il est également possible de travailler directement sur les images en niveaux de gris par filtrage différentiel. L'idée est d'utiliser des filtres permettant d'agglomérer les variations d'intensité périodiquement produites par les contours des caractères puis de rechercher des alignements horizontaux pour les lignes de texte. Cette technique, présentée par Wang et Srihari [Wang89], est comparable à une analyse de texture appropriée à la détection de caractères. Pour l'analyse de dessins, certains chercheurs proposent de privilégier l'extraction de zones d'intérêt au moyen de traitements de type morphologique (ouverture/fermeture, érosion/dilatation) afin d'éliminer certaines couches d'information et, au contraire, d'en conserver d'autres (séparation formes fines / formes pleines). Kasturi utilise par exemple ce principe pour localiser et extraire les régions pleines d'un document [Kasturi90]. Une succession d'érosions permet d'éliminer les segments, les arcs et les symboles de l'image. Reste ensuite à redonner aux objets encore présents dans l'image leur forme initiale par dilatations successives. La dilatation n'étant pas exactement l'opération inverse de l'érosion, des traitements supplémentaires sont encore nécessaires pour obtenir une restitution exacte des symboles pleins originels.

Composantes connexes et occlusions

De nombreux systèmes d'analyse de documents effectuent un étiquetage des composantes connexes [Baird92, Anderrsen03]. Sur la figure 1.6, on peut constater que la taille, la proximité et la position relative des composantes connexes peuvent être utilisées pour extraire la structure physique d'une page. Les rectangles circonscrits aux composantes connexes se chevauchent fréquemment dans les graphiques et rarement dans le texte. Ainsi, il est possible de faire l'hypothèse que les zones graphiques correspondent aux composantes connexes de cet ensemble de rectangles circonscrits dont les dimensions (largeur ou hauteur) dépassent un seuil choisi souvent empiriquement.

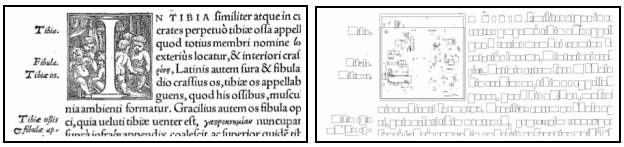


Figure 1.6 : Composantes connexes et segmentation

A priori, un bloc de texte est un ensemble de petites composantes connexes « proches ». Deux caractères sont dits voisins si la distance les séparant est inférieure à un espace maximal [Jaekyu95]. OGorman a proposé d'utiliser uniquement les distances entre composantes pour localiser les zones de texte dans une image [OGorman93]. Partant de cette même idée, de nombreux auteurs [Andersen03, Mitchell01] proposent d'utiliser un plus ou moins grand nombre de caractéristiques (de 5 à 20), extraites sur les composantes connexes, pour effectuer un apprentissage puis une classification de ces dernières. Le nombre et le type des classes varient énormément d'un auteur à l'autre en fonction de

l'objectif visé (séparation caractère/image/bruit ou identification des différentes polices de caractères utilisées).

La figure 1.7 illustre certains problèmes rencontrés avec ce type de méthodes. Il est parfois très compliqué, voire impossible, de fixer des seuils relatifs à la taille des séparateurs de zones. Par exemple, l'irrégularité de ces derniers amène à des situations où l'espace inter-mots est inférieur ou égal à celui qui sépare le corps de texte et les notes de marge (point 1 de la figure 1.7). Le bruit génère aussi des problèmes au moment de fixer des paramètres liés à la taille des composantes. En effet, de petites taches (point 2 figure 1.7) sont souvent identifiées comme étant des caractères. Une autre caractéristique des documents est que certains dessins sont composés d'une multitude de traits qui ne sont pas tous liés les uns aux autres. Cette caractéristique est à l'origine d'un grand nombre de composantes incluses les unes dans les autres et au final rend compliquée leur analyse (point 3 figure 1.7).

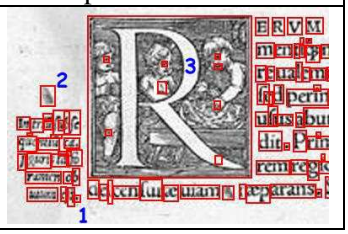
Image d'origine	Extraction des composantes connexes
	

Figure 1.7 : Diversité des composantes connexes dans un document ancien [Journet05a]

Kiochi Kise [Kise98] propose d'utiliser un pavage de Voronoï en plus des composantes connexes pour extraire les blocs de texte d'un document. Le pavage permet de construire un graphe de voisinage simplifiant la localisation des lignes de texte par des regroupements des formes proches. Cette méthode permet d'obtenir une segmentation réellement non rectangulaire de la structure physique indépendamment de l'inclinaison des lignes de texte.

Dans l'analyse de documents graphiques, les composantes connexes détectées procurent également d'importantes informations pour la suite du traitement, plus particulièrement pour l'extraction des composants « texte » ou l'extraction de certains symboles électriques. Une Transformée de Hough peut permettre de regrouper en mots ou en phrases les composantes connexes alignées correspondant (parfois!) aux caractères [Gloger92]. L'hypothèse utilisée est encore une fois que les caractères correspondent, lorsqu'ils ne touchent pas une autre partie du dessin, aux petites composantes connexes. La plupart des méthodes publiées constituent des variations autour de ce principe. La méthode la plus citée dans la littérature est celle de Fletcher et Kasturi [Fletscher88] récemment améliorée dans [Tombre02].

Des traitements complexes et variés peuvent alors être réalisés pour, d'une part supprimer les composantes ne correspondant pas à des caractères (pointillés, ...), et d'autre part localiser les caractères

manquants. Desseilligny [Desseilligny95] étudie très précisément tous les voisinages des composantes connexes de petites tailles pour les regrouper en mots. Pour cela, il utilise des critères de taille, de voisinage, d'orientation et effectue même une reconnaissance de la police. Cette solution n'étant pas satisfaisante, de nombreuses autres techniques ont été développées pour extraire les caractères attachés aux graphiques en minimisant à la fois l'altération du caractère et celle du graphique [Boatto92, Gao95, Dori96, Cao01, Wang01].

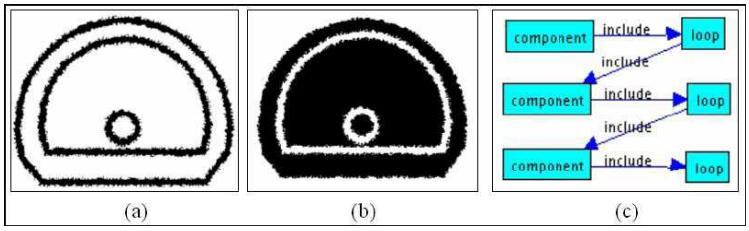


Figure 1.8 : Composantes et occlusions [Delalandre05]

Pour reconnaître des symboles graphiques, dans sa thèse Matthieu Delalandre [Delalandre05], propose d'utiliser les occlusions en plus des composantes connexes. Il construit alors des graphes qu'il nomme « hybrides » traduisant à la fois les relations d'inclusion et de voisinage entre les composantes et les occlusions (loop) figure 1.8.

Analyse des espaces et projections

Comme il est difficile de reconstruire les EdC d'un document à l'aide de critères locaux, des approches partant du point de vue inverse, tentent de localiser les séparations (espaces) entre les blocs de manière plus globale. En effet, deux EdC sont nécessairement séparés par un espace blanc de surface importante, que ce soit dans le sens horizontal ou vertical.

La localisation des séparations blanches peut se faire par analyse de la forme de l'histogramme du nombre de pixels noirs sur les lignes et les colonnes de l'image. On peut considérer une différence importante entre deux valeurs successives dans l'histogramme comme une délimitation entre deux EdC. Le problème est l'estimation des différences significatives dans l'histogramme. En général, on utilise des connaissances a priori sur le document en cours d'analyse (nombre de colonnes, marges, …) pour localiser plus facilement les séparations. Il est également possible, comme [Antonaco98, Breuel03, Baird90], de rechercher et d'analyser les plages blanches de l'image, d'étudier la disposition, la taille ou encore le contenu des composantes extraites. Parmi les méthodes proposées pour traiter les mises en page complexes des documents composites (journaux, pages de magazines, …), la méthode classique « Split and Merge » peut être adaptée comme proposée par Hadjar [Hadjar01] dans le cadre du « segmentation contest » de ICDAR'01. Le but est de localiser les zones blanches homogènes dans une image. Cette méthode peut être comparée à d'autres techniques à base de pavages plus ou moins évolués [Nagy84, Akindele93]. Initialement décrite par Pavlidis et Zhou [Pavlidis91], la méthode de

segmentation Recursive XY-Cut est basée sur la détection des grandes zones d'espaces (white streams) entre les blocs imprimés. L'angle d'inclinaison de la page est alors estimé à partir de l'analyse de ces espaces et les zones délimitées par ces espaces sont considérées comme des blocs de la structure physique. Malheureusement cette méthode ne fonctionne que sur des documents organisés en blocs rectangulaires. Plusieurs variantes permettent aujourd'hui de traiter des documents quelconques à partir de l'analyse des espaces et rendent cette méthode performante sur des documents variés. De plus, l'analyse du voisinage (réalisée avec l'aide de graphes ou d'arbres) de chaque région blanche découverte par découpage permet, ensuite, de localiser et de caractériser les blocs contenus dans la page à un niveau plus ou moins fin.

Dans [Akindele93], Akindele propose une méthode pour segmenter la structure physique en blocs non rectangulaires. Il extrait tous les espaces rectangulaires horizontaux et verticaux ayant une taille significative, puis les lignes centrales qui traversent le milieu de ces espaces rectangulaires horizontaux et verticaux. A partir de l'analyse des points d'intersection entre ces lignes, une heuristique permet de délimiter des blocs non rectangulaires. Cependant, cette approche peut délimiter des zones ne contenant ni texte, ni image. On élimine ces fausses détections en ne conservant que les zones de l'image qui possèdent une densité significative de pixels noirs. Dans tous les cas, il est nécessaire de faire appel à des connaissances a priori sur le document pour bien définir le critère d'homogénéité utilisé.

Techniques de vectorisation

Même si l'on fait parfois appel à d'autres techniques, la représentation des formes fines rectilignes contenues dans un document sous forme de vecteurs est très fréquemment utilisée. Ce passage de la forme matricielle de l'image à une description sous forme de vecteurs se nomme vectorisation. Pour aboutir à ce type de représentation, différentes techniques sont utilisables et une foule d'articles traitent de ce problème. Les états de l'art les plus récents sont [Wenyin99, Tombre00a, Tombre00b, Song02b, Delalandre05]. La plupart du temps, deux problèmes doivent être résolus pour obtenir les vecteurs désirés : d'abord la transformation de l'image en un ensemble de courbes discrètes d'épaisseur unitaire, puis une approximation polygonale de ces courbes. Voici les principaux moyens d'obtenir ce type de représentations.

Squelettisation

De nombreux systèmes proposent d'effectuer, d'abord, une squelettisation de l'image pour obtenir des courbes ayant une épaisseur de 1 pixel en tentant de garder les propriétés des formes initiales. Pour cela, on dénombre trois grandes classes d'algorithmes [Tabbone05] :

- les algorithmes à critères topologiques basés sur une suppression itérative des points non essentiels le long des frontières de la forme par des opérations booléennes [Ahmed92, Lam92, Lam95],
- les algorithmes à critères métriques s'efforçant de conserver les points du centre de la forme, c'est à dire ceux à égale distance des frontières [DiBaga94, DiBaga96, Chang99],

• les algorithmes effectuant un codage des plages ou segments noirs et blancs de l'image balayée ligne par ligne, pour ensuite construire le squelette [Pavlidis86, Roosli95].

Après une squelettisation, la quantité d'informations à traiter dans les phases qui suivent se trouve réduite mais la squelettisation pose certains problèmes comme l'apparition de barbules, la perte d'information (épaisseur du trait) ou l'inadaptabilité au traitement des formes pleines. L'algorithme de squelettisation proposé dans [DiBaga96] semble le plus largement utilisé.

Suivi de traits

Pour localiser (et amincir) les formes fines présentes dans un document, certains travaux proposent de se déplacer de point en point en restant, soit à l'extérieur de l'objet (suivi de contours), soit à l'intérieur de l'objet (suivi de traits). Une ligne peut alors être vue comme un ensemble de pixels noirs entre deux régions blanches. Partant d'un point initial, il s'agit alors de progresser à l'intérieur de la ligne jusqu'à l'extrémité opposée. Pour effectuer cette progression, Paquet [Paquet90, Paquet91] définit comme direction locale de la ligne (figure 1.9), la direction de Freeman permettant le plus grand déplacement à l'intérieur de la ligne.

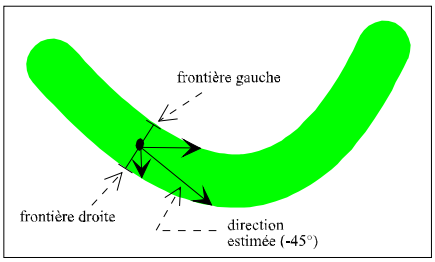


Figure 1.9 : Suivi de trait

L'algorithme de suivi estime, à chaque étape, les directions successives de la ligne et fournit la liste des points particuliers : jonctions (épaisseur anormale) ou extrémités. Paquet propose d'utiliser un pas de progression proportionnel à l'épaisseur moyenne de la ligne suivie. Cette méthode est utilisée par Ogier [Ogier94] pour l'analyse de plans cadastraux. L'équipe de Dori utilise une méthode similaire (Orthogonal Zig-Zag) permettant, en plus, la détection d'arcs de cercle [Dori95]. Cette technique a, de plus, été améliorée par Lui et Dori aboutissant à un algorithme intitulé SPV [Dori99]. D'autres systèmes, comme le système de lecture de dessins mécaniques ANON développé par Joseph et son équipe [Joseph89], utilisent une technique de ce type mais travaillent directement sur des images en niveaux de gris. Dans ce genre de méthodes, le processus est répété à partir du point courant tant qu'un critère de noirceur est vérifié. L'épaisseur du trait est évaluée après chaque étape par sondage suivant la direction perpendiculaire à celle de la ligne. Un deuxième seuil permet souvent une tolérance aux légères coupures du trait. Un autre problème devant être géré par ce type de méthodes concerne la gestion des croisements

et des jonctions ainsi que les éléments du dessin pour lesquels l'épaisseur estimée devient trop importante (formes pleines).

Suivi de contours

Ces méthodes ont la plupart du temps comme objectif d'obtenir les axes médians des formes [Chang99]. Quatre étapes sont souvent nécessaires : vectorisation des contours, appariement des vecteurs obtenus à partir de l'analyse de leur orientation, génération des axes médians et analyse des jonctions (remplissage des trous). Les deux premières étapes ont pour but d'obtenir des segments ("blocks"). Pour interconnecter les segments entre eux, des informations topologiques sont nécessaires [Fan98, Han94]. Certains proposent de prolonger les axes médians et de rechercher les points d'intersections entre ces derniers. Le principal défaut de ces techniques provient de la gestion des vecteurs non ou mal appariés et des différents paramètres ou heuristiques à utiliser.

L'autre possibilité est d'utiliser les contours afin d'accélérer le processus de vectorisation en obtenant les axes médians en une passe sans visiter tous les pixels de l'image [Baruch88]. Il faut alors suivre les deux contours opposés des formes simultanément et déterminer le milieu du segment reliant les 2 frontières. Abe [Abe86, Fukada84] propose une méthode de suivi de contours adaptée à l'analyse des organigrammes. Grâce à cette technique, les boucles fermées, utiles par la suite, sont extraites durant la vectorisation. Cugini [Cugini84] emploie une technique d'appariement de contours pour obtenir une représentation, sous forme de vecteurs, des projections orthogonales. Le problème de ce type d'algorithmes est leur mauvaise performance sur les formes complexes (nombreuses jonctions, présence de formes pleines, …).

D'autres systèmes couplent l'information « contour » avec l'information « squelette » pour améliorer la qualité des résultats [Leplumey95, Hori93, Tanigawa94]. Il ressort de ces études que la mise en correspondance des contours est moins sensible au bruit que la squelettisation mais que sa mise en place est beaucoup plus complexe.

Codage de plages et pavage

Dans le but d'obtenir un codage de l'image sous forme de vecteurs, les méthodes par «codage des plages» (Run Lenght Coding) balaient l'image ligne par ligne. Les sections horizontales (suites de pixels homogènes) noires et blanches de chaque ligne sont extraites puis codées par des paramètres relatifs à leur position et à leur longueur. Plusieurs algorithmes de codage fondés sur ce principe sont décrits dans la littérature [Ramachandra80, Pavlidis86]. L'algorithme de Pavlidis permet d'obtenir un codage des objets de l'image sous forme de graphes (figure 1.10 : Line Adjancy Graph). Les nœuds du graphe correspondent aux milieux des segments (ou sections) noirs. Les nœuds sont reliés par une arête s'ils correspondent à deux segments en contact situés sur deux lignes voisines. Ramachandra effectue un parcours du graphe et, par regroupement, selon des critères de pentes et de longueurs sur les sections, obtient un codage sous forme de vecteurs (figure 1.10).

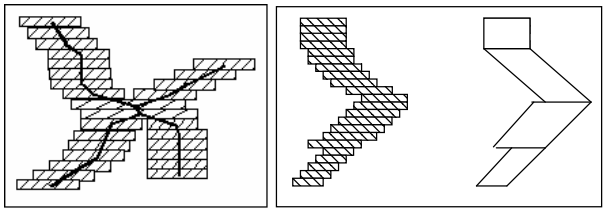


Figure 1.10 : Graphe d'adjacence des lignes (LAG) et vectorisation à partir du LAG

Ce type de technique a souvent été utilisé pour vectoriser des images de type plan cadastral [Madej91, Antoine91], car il s'adapte bien à la détection de zones hachurées.

Pour accélérer la phase de vectorisation, il est possible de sous-échantillonner l'image initiale à l'aide d'un quadrillage de dimensions adaptées à l'épaisseur des traits [Lin85]. Cette idée a été reprise et améliorée par Vaxivière [Vaxivière94a]. L'ajout d'un mécanisme de fusion et de division des mailles facilite ensuite le suivi des traits. Avec cette méthode, il est nécessaire de caractériser précisément les différents types d'intersections entre les traits. Arias propose une description hiérarchique des différents types d'intersections pouvant apparaître dans des plans téléphoniques [Arias94].

Approximation polygonale ou d'ordre supérieur

Dans tous les cas, il se pose ensuite le problème du choix de la méthode d'approximation : c'est à dire la méthode de sélection des points critiques : extrémités des vecteurs, … Ce problème souvent passé sous silence est pourtant sûrement aussi important que le précédent (méthode d'obtention d'une courbe d'épaisseur unitaire).

Une approximation se réalise à partir d'une entité de référence qui peut être le segment de droite, l'arc de cercle ou une courbe d'ordre supérieur. L'approximation polygonale est la plus fréquemment utilisée et concerne l'approximation à partir de segments de droite. Le calcul d'approximation est fondé, soit sur le calcul d'une erreur minimale pour un nombre de segments donnés, soit sur la recherche d'un nombre minimal de segments pour garantir la précision demandée.

Il existe bon nombre de méthodes permettant de résoudre ces deux types de problèmes [Kolesnikov03] : des méthodes optimales basées sur la programmation dynamique ou la théorie des graphes, des méthodes basées sur la reconnaissance de segments flous [Rennesson03]. Cependant, ces méthodes sont lentes et ne donnent pas toujours de très bons résultats. Durant le stage de Master 2 d'Aurélien Hublier [Hublier05], nous nous sommes intéressés à deux grandes familles de solutions :

- les méthodes fonctionnant par découpages récursifs [Rosin89] :

Le principe est de commencer à approximer grossièrement la courbe puis de l'affiner par découpages récursifs tant qu'un certain critère est vérifié. Généralement, la courbe est d'abord approximée à une droite entre deux points de la courbe. On cherche alors le point de la courbe se situant le plus loin de la droite suivant un certain critère. Si cette distance dépasse un certain seuil, on coupe la droite en ce point

et on recommence sur les deux segments obtenus sinon on arrête. Cette méthode est précise quant aux choix des points de coupure. Cependant elle a le désavantage d'être très sensible au bruit (figure 1.11).

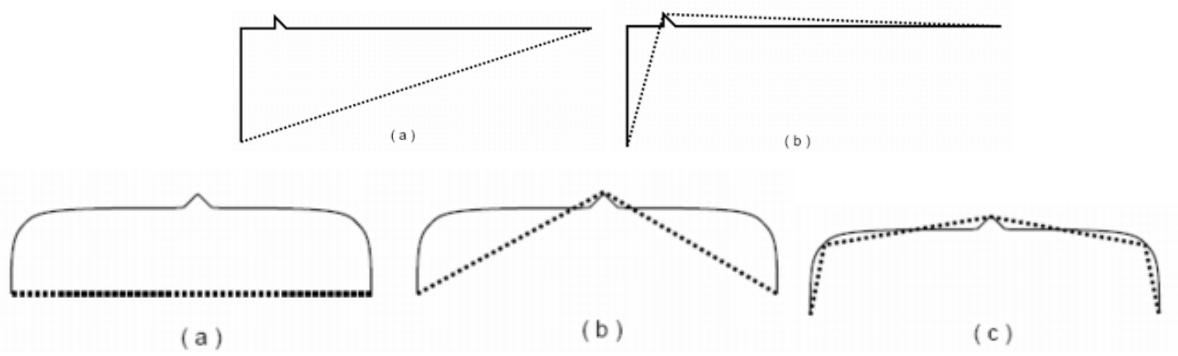


Figure 1.11 : L'influence du bruit : (a) résultat à la première étape, (b) résultat à la deuxième étape, (c) résultat final, une droite a été coupée à cause du bruit.

- les méthodes fonctionnant par fusions itératives :

Le principe est de parcourir la courbe du premier point au dernier point. Au fur et à mesure du parcours, les points rencontrés sont fusionnés en un même segment tant que l'ensemble de ces points vérifie un certain critère. Le dernier point vérifiant le critère correspond à l'extrémité du segment. On recommence alors ce processus à partir de ce point. Le principal défaut de cette méthode est que les points de coupure se situent généralement au delà des points de coupure souhaités (figure 1.12).

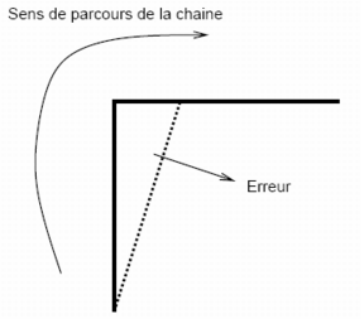


Figure 1.12 : Les méthodes de fusion génèrent une erreur sur la position du point de coupure

Pour ce type de méthodes, l'un des algorithmes les plus fréquemment utilisés a été proposé par Wall [Wall84]. Cette méthode d'approximation introduit un nouveau sommet lorsque l'erreur commise en remplaçant la courbe par une ligne droite devient trop importante (figure 1.13). La qualité du résultat obtenu est fonction d'un calcul d'aire et d'un seuil ε. L'erreur cumulée correspond à l'aire entre la courbe et le segment l'approximant à chaque étape du processus.

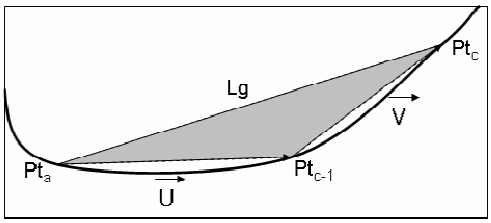


Figure 1.13 : Illustration du calcul d'aire entre la courbe et le segment l'approximant.

Afin de résoudre le problème décrit figure 1.12, nous proposons dans [Hublier05] d'effectuer un post-traitement qui repositionne correctement chaque point de contrôle fourni par l'algorithme initial. Nous montrons qu'effectuer un retour arrière dans la direction courante d'une longueur de ε pixels est pertinent (figure 1.14). Ce travail ouvre la voie à une réflexion plus poussée pour trouver des solutions encore plus précises.

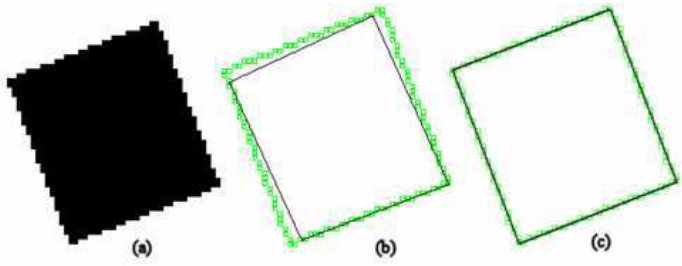


Figure 1.14 : (a) image d'origine, (b) approximation sans retour, (c) approximation avec retour, en clair : les pixels du contour détecté, en foncé : les segments déterminés par approximation polygonale.

Certaines méthodes d'approximation sont fondées sur la notion de courbure. Ce type de méthodes [Brault93] est souvent utilisé pour segmenter les signatures manuscrites ou les tracés en ligne [Rhee01]. Soit $S=(Pt_1, Pt_2, \ldots, Pt_n)$ une courbe constituée de n points successifs. La représentativité R d'un point Pt_i dépend de la variation d'angle entre le point sélectionné et les points voisins. Elle est calculée de la manière suivante et est illustrée sur la figure 1.15 :

$$R(Pt_i) = \sum_{k=m}^{m_{\max}} \cos(\theta_b(k)) * \cos(\theta_f(k))$$

Où m désigne l'éloignement du premier voisin pris en compte, θ_{max} l'angle maximum de $|\theta_b(k)|$ et de $|\theta_f(k)|$ et m_{max} l'entier supérieur à m tel que

$$\sup(\theta_b(m_{\max}), \theta_f(m_{\max})) > \theta_{\max} \text{ et } \sup(\theta_b(m_{\max-1}), \theta_f(m_{\max-1})) < \theta_{\max}.$$

Les points retenus sont ceux qui correspondent à des maximums locaux de la représentativité R.

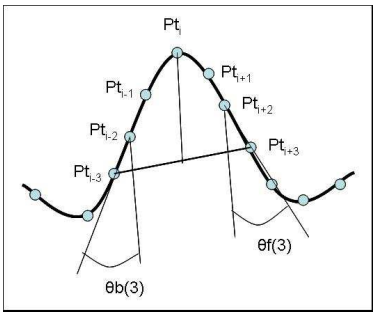


Figure 1.15 : Illustration du calcul de la représentativité du point Pt_i avec m=3.

L'utilisation de cette méthode dans le cadre de la thèse de Matthieu Wirotius [Wirotius04a, Wirotius05c] a démontré qu'il pouvait être très profitable de faire plus de liens entre les méthodes conçues initialement pour le traitement des documents en ligne et les méthodes conçues pour les images (et vice versa). La figure 1.16 montre sur deux exemples les points de contrôle retenus par cette méthode lors de l'approximation de signatures manuscrites en ligne.

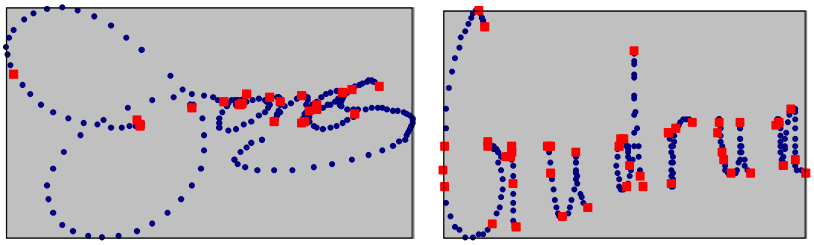


Figure 1.16: Illustration des points retenus par la méthode de Brault.

Après l'étape d'approximation polygonale, les formes fines de l'image sont représentées par une suite de points de contrôle correspondant aux extrémités des vecteurs avec parfois l'épaisseur moyenne du trait. Roosli a proposé d'intégrer 2 primitives (l'arc et le vecteur) dans un même temps pour approximer les différentes parties des dessins [Roosli95]. Il intègre de plus des contraintes de parallélisme, de perpendicularité et de concentricité pour obtenir des résultats plus « propres ». Dans le même ordre d'idées, bien que la vectorisation se situe dans les traitements bas niveau, certains auteurs proposent d'avoir recours à l'usage de connaissances a priori sur le type de documents pour obtenir de meilleurs résultats. Chhabra et al. [Arias95] ont par exemple développé un système très efficace mais utilisable uniquement pour la rétro conversion de plans de réseaux téléphoniques d'un type bien précis. Song [Song00, Song02a, Song02b] pense préférable d'intégrer des heuristiques pour corriger les erreurs au niveau des jonctions ou pour boucher les trous. Il propose également d'utiliser différents algorithmes de vectorisation en fonction du type d'éléments à représenter/approximer (texte, courbes, traits) mais on rejoint ici une nouvelle fois le fameux paradoxe du « reconnaître pour mieux segmenter et segmenter pour mieux reconnaître… »

La séparation courbes / droites est généralement effectuée durant cette étape à l'aide d'algorithmes capables de détecter les arcs, directement à partir du squelette ou de l'image binaire [Pikaz95, Song04]. La transformée de Hough est classiquement utilisée [BenYacoub95] mais il est également possible de travailler à partir des points de contrôle obtenus par approximation polygonale [Dosch00, Ramel98e] pour détecter les arcs et les courbes.

Il est à noter que quelle que soit la méthode utilisée, certains problèmes, comme l'influence du point de départ choisi, la séparation courbes/lignes droites, l'imprécision de position des points de contrôles (méthodes itératives), la sensibilité aux distorsions locales (méthodes récursives) restent des challenges. Je conseille la lecture des travaux de P. Rosin [Rosin89, Rosin03] sur l'évaluation des méthodes de représentation des courbes dans les images.

1.2.2. Signatures statistiques et caractérisation des contenus

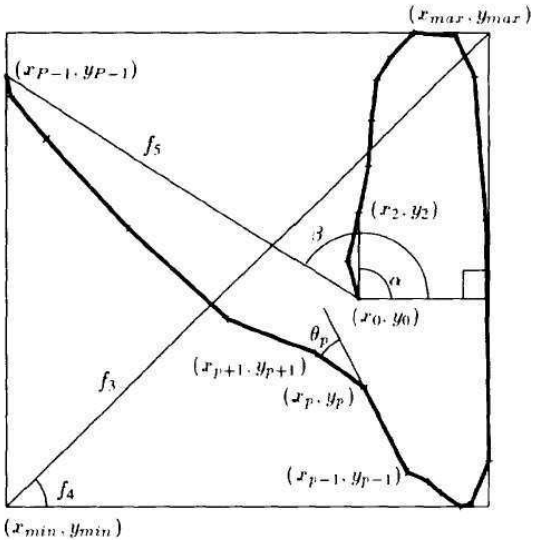
Signatures statistiques

De plus en plus de méthodes basées sur un calcul de signatures sont proposées pour identifier les EdC extraits d'images de documents [Ventura94, Dosch03, Wang06]. Les signatures peuvent être vues comme une collection de caractéristiques numériques extraites sur des formes pré-segmentées (le plus souvent les composantes connexes pour les symboles graphiques et les régions homogènes (paragraphes, images) pour les documents imprimés). Parmi les signatures les plus souvent utilisées, on peut citer :

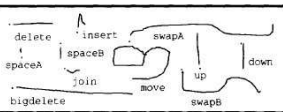
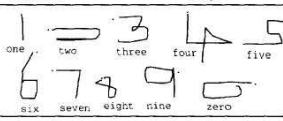
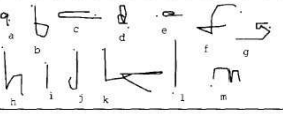
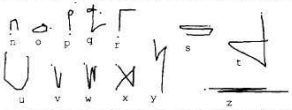
- Les profils horizontaux et verticaux des formes, les histogrammes des projections, les histogrammes des forces [Wendling02], …
- Les moments géométriques (Zernike, Hu, ...) [Khotazad90],
- Les transformées (Fourier [Zhang05], Fourier-Mellin [Adam00a], Radon [Tabbone02],…),
- Les signatures vectorielles [Dosch03].

Sur les tracés en ligne, beaucoup de techniques se basent simplement sur les coordonnées spatiales éventuellement échantillonnées pour caractériser les formes. L'étape de prétraitement (lissage) n'est quasiment jamais présente dans les architectures classiquement proposées (sauf pour la normalisation). La plupart du temps, l'extraction de caractéristiques est réalisée directement à partir de la liste initiale de points fournis par le dispositif de saisie et la reconnaissance est ensuite réalisée à partir de ces caractéristiques statistiques calculées à partir de ces points. Quelques auteurs préconisent l'utilisation de caractéristiques statistiques ou géométriques de plus haut niveau (orientation des traits, courbures, mesures géométriques, …) obtenues après extraction des parties stables [Fonseca01, Azé00, Schomeker93, Mouchère06]. Par exemple, la méthode implémentée dans la plupart des dispositifs (PDA, Palm) utilise un algorithme [Rubine91] fonctionnant par analyse d'une quinzaine de caractéristiques géométriques extraites à partir de la succession de points constituant un geste (voir

figure 1.17) et comparées à l'aide de classificateurs plus ou moins évolués (analyse discriminante, k-means, …).



(a)

Set Name	Gesture Classes	Number of Classes	Recognition Rate
Coleman		11	100.0%
Digits		10	98.5%
Let:a-m		13	99.2%
Let:n-z		13	98.4%
Letters	Union of Let:a-m and Let:n-z	26	97.1%

(b)

Figure 1.17 : Caractéristiques (a) et types de formes (b) utilisées dans [Rubine91]

Pour plus de détails sur les autres caractéristiques utilisées, le lecteur pourra consulter [Trier96 Schomaker99] la difficulté étant souvent de choisir sur quelles entités ou parties extraire ces caractéristiques (traits, tracés, caractères, mots).

Analyse de texture et multi-résolution

Les outils permettant d'analyser la texture des images de documents sont nombreux. Une définition possible de la texture est celle de la répétition périodique d'un motif donné. Dans le cas de la détection de texte, on cherche à mettre en évidence certaines caractéristiques attendues : horizontalité du motif, fréquence des transitions noir-blanc … Ces motifs peuvent être décrits par un ensemble de statistiques ou de propriétés locales appropriées que l'on va chercher à rendre invariantes à certains aléas : présence de bruit dans l'image, taille variable des caractères, interligne irrégulier, forte variation des niveaux de gris du texte…. Les approches les plus répandues sont celles qui s'appuient sur les propriétés fréquentielles des images (filtres de Gabor, ondelettes…). Parmi les plus pertinentes, on peut citer [Wu97] qui utilise des bancs de filtres de Gabor pour mettre en évidence la propriété d'horizontalité du texte et [Jain96, Eternad97] qui utilisent des ondelettes pour segmenter les images de documents composites. La plupart du temps une analyse locale du voisinage des pixels permet de construire une segmentation de l'image en régions de texture homogène.

Le concept de multi-résolution peut apporter de nombreux avantages lorsqu'il s'agit d'extraire ce type d'information d'une image. Dans ce cadre, les modèles pyramidaux [Jolion94a, Jolion94b] offrent plusieurs avantages tels que la parallélisation et l'augmentation de la robustesse du traitement lors de la

détection des contours, l'extraction de régions ou lors de l'analyse des textures. Cependant, les méthodes réduisant la résolution de l'image deviennent difficilement utilisables lorsque l'on veut obtenir directement une segmentation précise des images de documents, principalement à cause de la proximité entre les différents EdC. Les espacements deviennent indétectables lorsque l'on diminue la résolution et les zones de texte ne présentent, par exemple, aucune différence de texture significative. Ces techniques s'avèrent néanmoins intéressantes lorsque l'on cherche à caractériser globalement, à l'aide de différents indices, des zones particulières. Par exemple, certains auteurs comme [Cinque98, Lim99] effectuent une séparation texte/image par analyse multi-résolution. L'étude de l'image à différentes échelles permet d'apporter une information supplémentaire qui vient enrichir les connaissances utilisables lors de la phase de classement des différents EdC extraits.

1.2.3. Structuration des primitives et représentation

La manière de stocker et structurer les primitives de description ou signatures afin de pouvoir alimenter le système de reconnaissance est également d'une importance primordiale. Cette structuration dépend de l'approche retenue pour effectuer la reconnaissance des formes. Des formalismes de type vecteurs de descripteurs statistiques sont employés lorsqu'un classificateur statistique est utilisé et des structures relationnelles attribuées plus ou moins évoluées (listes, grammaires, graphes) sont construites lors de l'utilisation de classificateurs symboliques ou structurels.
L'étude de G. Nagy [Nagy00] vient compléter cette constatation en proposant une liste des primitives les plus couramment utilisées en analyse de documents imprimés et graphiques. On retrouve dans cette liste les éléments cités dans la section précédente :

- Pour les documents principalement textuels, on retrouve les pixels, les plages, les segments, les composantes connexes, les mots, les lignes, les blocs, …
- Pour les documents principalement graphiques, on retrouve les pixels, les plages, les segments, les arcs, les composantes connexes, les boucles, les jonctions, les extrémités, …

Lorsqu'une approche région est utilisée, bien souvent, les systèmes utilisent simplement des structures de listes pour stocker les primitives que sont les composantes connexes, les régions ou les blocs (mots, paragraphes, EdC, …) [Badawy02]. De même, avec une approche contour, de nombreux systèmes [Han94, Yu97, Llados01] emploient simplement des listes de points ou des listes de vecteurs. Notons que les primitives graphiques structurées comme les vecteurs et les arcs proviennent parfois directement de documents graphiques vectoriels (DXF) [Pasternak 93, Weindorf02, Fonseca04]. Dans le cadre de l'analyse de documents imprimés, les structures d'arbres sont majoritairement utilisées pour structurer les données graphiques extraites. Une approche intéressante est proposée dans [Jain98] qui commence par construire une représentation de l'image à l'aide d'un LAG (Line Adjacency Graph - voir section 1.2.1) pour ensuite extraire les composantes connexes puis les EdC de niveaux sémantiques plus élevés

tels que les dessins, les images, les tableaux, les séparateurs et le texte aboutissant, ainsi, à une structuration hiérarchique du document. Pour les documents graphiques, la majorité des systèmes [Messmer95, Llados01, Yan04] exploitent des descriptions sous forme de graphe attribué.

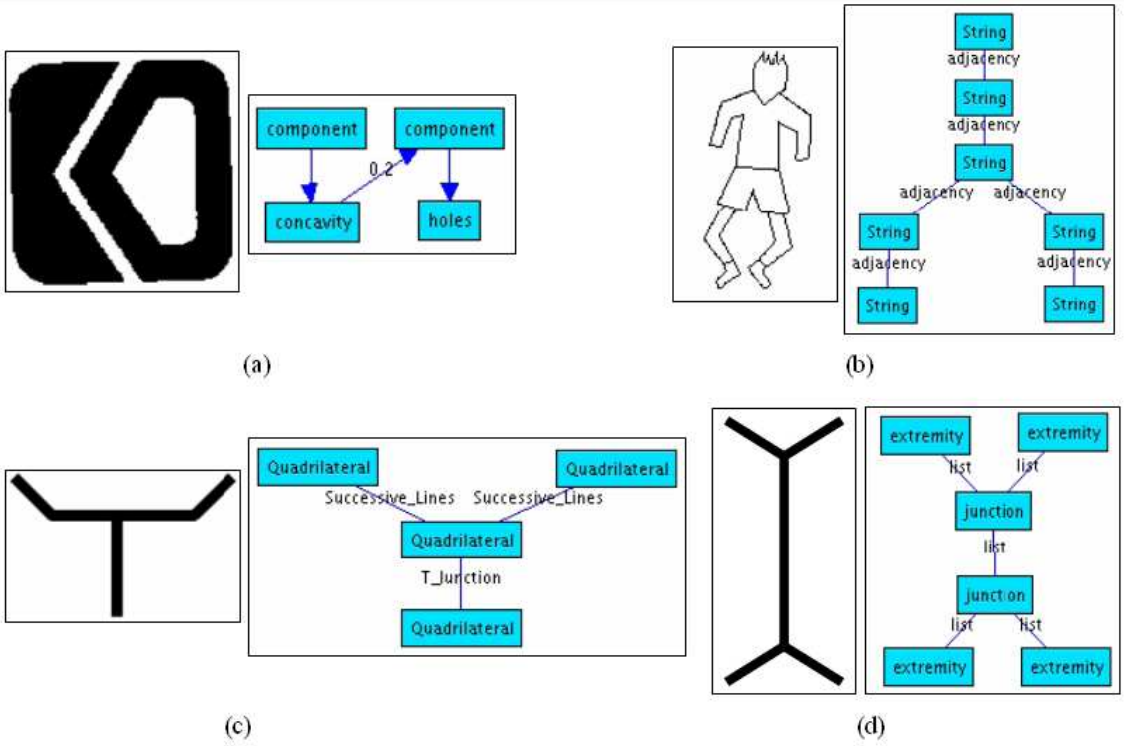


Figure 1.18 : Différentes formes de représentations structurelles extraites de [Delalandre05] : Région (a) voisinage/inclusion [Badawy02] (b) adjacence [Llados01] (c) contour [Ramel00] (d) squelette [Popel02]

Lors de l'usage de graphes, on distingue différents types de modélisation des relations entre primitives graphiques [Delalandre05]. La première solution consiste à exploiter les relations d'inclusion et/ou de voisinage entre composantes. Ce type de modèle est utilisé dans différents systèmes utilisant une approche région. La seconde solution consiste à utiliser les relations d'adjacence entre régions. On parle de graphe d'adjacence de régions (figure 1.18b). Ce type de modèle est utilisé dans de nombreux systèmes [Seong93, Llados01, Weindorf02]. Llados [Llados01] utilise ce formalisme pour associer les régions closes d'un squelette vectorisé et les caractériser par des chaînes (String). Les arcs décrivent alors des relations d'adjacence (Adjacency) entre les chaînes.

D'autres systèmes [Pasternak93, AhSoon01, Couasnon03, Weindorf02] ont recours à des langages à base de spécification de contraintes et de grammaire. Plusieurs systèmes [Joseph92, Wu93, Wenyin99, Song02a] exploitent une modélisation orientée objet. Les primitives graphiques correspondent alors aux objets du langage. L'utilisation d'un tel formalisme permet d'exploiter les relations d'héritage entre objets dans des buts d'extensibilité et de réutilisabilité. L'étude de ces différents modèles permet de constater différents points [Delalandre05]. D'abord, ceux-ci sont souvent peu ou mal formalisés dans la littérature, ensuite, chaque système utilise un modèle de représentation figé en fonction des besoins de

l'application visée. Ce modèle est alors défini en fonction de la base d'images utilisées et de l'expertise du concepteur du système.

1.3. Problèmes ouverts

Même si les nombreux travaux que l'on vient de voir ont permis d'importantes avancées en ce qui concerne les traitements de bas niveau, l'extraction et la structuration des primitives, des difficultés subsistent encore pour représenter de manière structurée le contenu des documents numériques (images ou tracés en ligne).

- Pour les images de documents comme pour les tracés en ligne, contrairement à ce que l'on pourrait croire, la détection de la structure physique (séparation des différents EdC) n'est pas un problème résolu. La localisation ou la segmentation des symboles dans les images de documents graphiques complexes est un problème très rarement traité car très difficile à résoudre. Preuve en est, l'organisation de différents concours ou campagnes d'évaluation qui se hasardent rarement sur ce terrain et préfèrent se limiter à l'évaluation d'algorithmes de vectorisation, de détection d'arcs de courbes ou de reconnaissance de symboles isolés. De même pour les documents imprimés anciens, dégradés ou complexes, il semble que les méthodes procédant par découpage (pavage, analyse de texture, …) soient moins sensibles au bruit que les méthodes procédant par regroupement (analyse de composantes connexes, filtrage, …). Cependant, elles ne fonctionnent que lorsque les images sont bien redressées et elles ne résolvent qu'en partie le problème de proximité entre les blocs, entre les mots et entre les caractères. De plus, elles sont difficilement applicables sur les documents ayant une mise en page fluctuante ou non rudimentaire, puisqu'elles nécessitent l'utilisation de connaissances a priori [Hadjar02] sur le document (nombre de colonnes, critères d'arrêt du découpage,…). Elles ne sont donc pas adaptées au traitement des documents qui ne respectent pas une charte éditoriale précise.

- La séparation Texte / Graphique pose des difficultés lorsque la taille des composantes connexes n'est pas significative (figure 1.19) : lorsqu'un caractère est connecté à une autre partie du dessin ou quand il s'agit de manuscrit dans les images de documents graphiques. Pour la localisation des zones graphiques, les seuils doivent correspondre à une taille légèrement supérieure au plus grand caractère contenu dans le document. Si des études statistiques peuvent permettre d'automatiser la sélection de ces seuils, ces méthodes supportent difficilement les variabilités de taille des caractères et de mise en page qui peuvent apparaître fréquemment dans les ouvrages anciens ou dans les documents graphiques. La mauvaise qualité des images (bruits, taches, …) pose aussi certains problèmes. De même, dans les documents graphiques, les caractères isolés sont très difficilement localisables et sont souvent confondus avec d'autres petits symboles. Par exemple, les pointillés ne doivent pas être confondus avec les caractères.

Figure 1.19 : Problèmes lors de la localisation

• Peu d'articles s'étendent sur l'extraction et le codage des courbes et autres formes non rectilignes, si nombreuses dans les tracés manuscrits, ainsi que sur leur mise en relation avec les autres parties du document (liaison entre une droite et une courbe) [Rosin95]. De plus, les problèmes posés par les algorithmes d'approximations polygonales (précision, robustesse, sensibilité, influence du point de départ) ont déjà été mentionnés. Ainsi, quelle que soit la technique de vectorisation, le résultat de l'approximation polygonale n'est pas toujours celui escompté (figure 1.20). Les points de contrôle fournis ne sont pas toujours représentatifs de la forme initiale comme on peut le voir sur la figure suivante.

Figure 1.20 : Problèmes lors de l'approximation

• Certains traitements bas niveau ne sont pas adaptés à certains types de formes. Par exemple, les formes pleines ainsi que certaines jonctions ou extrémités sont encore trop souvent ignorées (voire altérées durant la squelettisation) lorsque des méthodes structurelles sont utilisées (figure 1 .21).

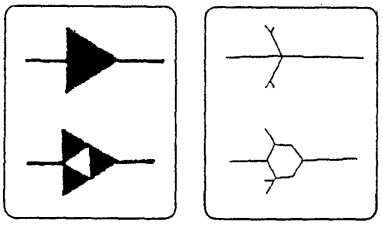


Figure 1.21 : Problèmes de représentation structurelle des formes pleines

• Mathieu Delalandre [Delalandre05] effectue une analyse détaillée des différentes méthodes d'extraction de primitives et de représentation du contenu de documents graphiques. Cette étude démontre l'engouement pour les méthodes structurelles lorsqu'il s'agit de traiter ce type de document même si, actuellement, certains chercheurs ont pris conscience de l'intérêt qu'il pouvait y avoir à

utiliser des approches plus statistiques. Malheureusement, si ces méthodes marchent bien sur des formes pré-segmentées, leur utilisation est quasiment impossible lorsqu'il devient nécessaire de travailler sur des images complètes contenant un ensemble de symboles interconnectés les uns avec les autres. Une nouvelle fois la solution idéale semble résider dans un couplage de ces deux types de méthodes. Matthieu Delalandre parle alors de multi-représentations mais l'implémentation qu'il propose reste très complexe car elle exploite en fait plusieurs représentations sous forme de graphe plutôt qu'une seule fédérant l'ensemble des informations statistiques et structurelles.

- Enfin, n'oublions pas de mentionner le problème que l'on retrouve pourtant dans toutes les méthodes d'extraction de primitives : les nombreux paramètres difficiles à comprendre et, donc, à régler par les utilisateurs non experts et les temps de calcul souvent prohibitifs qui sont à prendre en compte lors de l'évaluation des performances d'un système.

La section suivante décrit les différentes propositions produites dans le cadre de mes travaux de recherche, pour tenter d'apporter des solutions aux problèmes mentionnés précédemment.

1.4. Contributions

Les apports obtenus par la mise en place d'une représentation structurelle adaptée aux images de documents sont décrits dans la première partie de ce chapitre. Ensuite, différentes exploitations possibles de la représentation développée sont présentées afin de mieux comprendre et démontrer son intérêt.

1.4.1. Une représentation pour les images de documents et les tracés en ligne

La section précédente a mis l'accent sur les avancées et sur les problèmes restant ouverts concernant les méthodes d'extraction de primitives, de description, de segmentation, de caractérisation des EdC et de structuration des données extraites. Afin de résoudre certains des problèmes mentionnés, j'ai proposé dans la plupart de mes travaux d'avoir recours à une réflexion préalable poussée sur la ou les représentations (ici, le contenu des documents numériques) à mettre en place de manière à rester le plus générique possible. Cette transformation des données initiales me paraît indispensable afin de structurer les informations pour mieux décrire les données initiales et pour pouvoir les traiter plus facilement par la suite. Cette représentation doit posséder les caractéristiques suivantes :

- conservation de l'information intéressante contenue dans le document,
- réduction de la place nécessaire au stockage,
- simplification et adaptabilité du nouvel espace de représentation aux traitements à réaliser,
- généricité.

Il est difficile de trouver une représentation donnant entière satisfaction pour toutes ces contraintes. De manière pratique, on choisit un compromis en privilégiant les aspects importants du problème traité. Si l'on s'inspire des travaux faits dans le cadre de l'esquisse initiale de David Marr [Marr82], on voit rapidement que son défaut majeur proviendrait du fait qu'elle est uniquement basée sur les contours. Aussi, dès les origines de la détection de contours, l'approche duale, à savoir la détection de régions, a été envisagée [Lindenberg93] : en effet, si les contours délimitent les régions homogènes, il est tout aussi possible de détecter l'homogénéité de l'intensité lumineuse, plutôt que ses discontinuités. Lindeberg propose d'ailleurs une esquisse initiale basée sur la notion de taches plutôt que de contours.

L'utilisation des notions de taches et de contours constitue une étude intéressante en vue d'une représentation initiale des documents : les taches peuvent permettre d'extraire les principaux objets d'un document, lorsque ces derniers correspondent à des « bosses » en niveaux de gris. Néanmoins, l'information recueillie peut paraître incomplète sans l'information de type « contour », usuellement employée, pour caractériser la forme de ces taches. Dès lors, les descriptions obtenues paraissent complémentaires.

Ainsi, sur une image de document, une approche « région » permettrait de définir simplement des relations d'inclusion, de superposition, voire d'occlusion. Une approche de type « contour » permettrait quant à elle une bonne représentation des formes. De plus, une bonne représentation se doit de simplifier les données d'origine en un croquis pertinent. En ce sens, une information binaire constitue une simplification extrême, mais qui présente l'avantage d'être facilement utilisable. Nous en concluons que dans le cas des images de documents, l'information obtenue devra résider dans une description simple mais riche, construite à partir de primitives générales dérivant des notions de contours et de régions. Bien évidemment, l'information « niveaux de gris » ne doit pas être complètement négligée et pourra être utilisée dans certains cas.

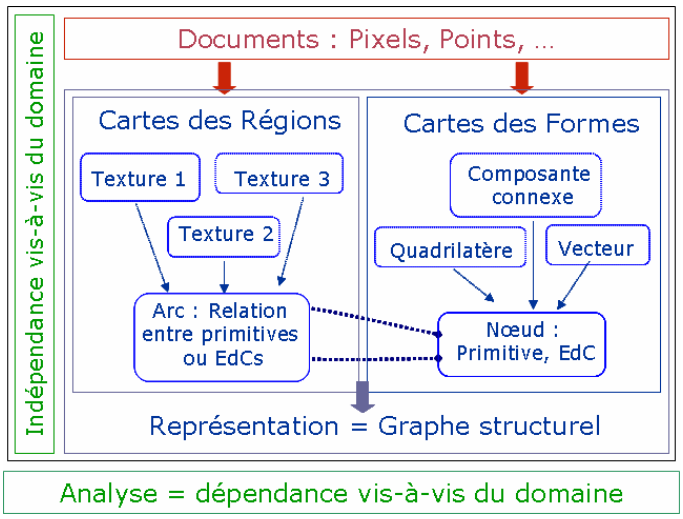


Figure 1.22 : Structure hiérarchique de la représentation et mode de construction

Afin de respecter toutes ces contraintes, la représentation que je propose se base sur l'exploitation duale d'informations « contour » et « région » (comme indiqué préalablement) et sur une organisation hiérarchique des primitives de description comme décrit figure 1.22. Différentes thèses ont été mises à contribution pour aboutir à une telle représentation : Matthieu Wirotius a cherché une représentation adaptée aux tracés en ligne ; Nicholas Journet travaille sur les aspects « région » par analyse de texture (en tenant compte des niveaux de gris) ; Rashid Jalal Qureshi travaille sur les aspects « formes » par vectorisation des contours afin de produire des graphes structurels (images binaires).
Le premier étage de cette représentation repose sur la construction de cartes mettant en évidence les différentes régions présentes dans le document et sur la construction de cartes fournissant des informations sur les contours des formes contenues. Je propose ensuite d'utiliser simultanément les informations fournies par ces deux types de cartes pour construire une représentation structurelle du contenu du document sous forme d'un graphe attribué.

Cartes des formes

Les composantes connexes fournissent une information pertinente au niveau des formes (bruits, graphiques, caractères d'un texte, symboles, …). Leur position, leur taille et le chevauchement de leurs rectangles circonscrits fournissent des informations précises sur la structure physique des documents. Elles correspondent à une information locale sur chaque forme contenue dans l'image. Les composantes connexes représentant les lettres d'un mot sont extrêmement proches même si dans les documents anciens, la mise en forme n'est pas irréprochable. Par exemple, la dernière lettre d'une ligne peut être plus proche d'une note en marge que de la lettre qui la précède.
Pour les dessins, les composantes connexes apportent également une information pertinente concernant certaines couches (texte, pointillés, liens entre symboles). Après une binarisation de l'image, pour obtenir des informations sur les composantes connexes, un suivi des contours est effectué permettant d'extraire à la fois les rectangles circonscrits de chacune des composantes connexes et les frontières des formes [Ramel05a]. La position et la taille des rectangles sont stockées dans une liste constituant la carte des formes (figure 1.30). En fonction de ses dimensions, un label peut être affecté à chaque forme comme par exemple :

- *Bruit* (composantes connexes de petite taille),
- *Graphique* (composantes connexes de grande taille),
- *Texte* (autres composantes connexes de taille moyenne).

Les seuils utilisés durant cette phase peuvent être choisis par l'utilisateur, par exemple, en fonction de la taille maximale et minimale des caractères dans les documents. De plus, il ne s'agit que d'un premier étiquetage qui peut évidemment évoluer par la suite.
Dans les images binaires, une forme est décrite de manière équivalente par son contour ou par les pixels de la région noire qui la constituent. Une information sur sa forme et sa direction peut être obtenue par

une étude locale de ses contours. Quant à l'épaisseur, qui correspond à la distance entre les deux frontières d'un trait, elle est surtout intéressante pour l'étude des formes fines (écriture et traits).
Une approximation polygonale des contours des formes constituant l'image initiale est effectuée afin d'obtenir une suite ordonnée (par le suivi des contours) de vecteurs. Cette étape permet d'éviter la squelettisation classiquement utilisée pour vectoriser les images de documents et les problèmes qui lui sont associés. Cette proposition correspond à la reprise de mes travaux de thèse [Ramel96], réalisée actuellement dans le cadre de la thèse de Rashid Jalal Qureshi, afin de les généraliser pour tous types de documents (graphiques et composites) [Qureshi05].

Chaque vecteur résultant de l'approximation polygonale des contours (figure 1.23) possède différents attributs utilisés, par la suite, durant une étape de fusion (de certains vecteurs) qui permet d'augmenter encore la robustesse de notre représentation des documents. Cette étape de fusion permet notamment de s'affranchir du problème de l'influence du point de départ lors des approximations polygonales des contours (un des problèmes mentionnés section 1.3).

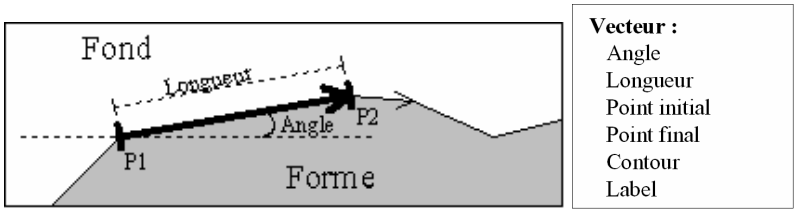


Figure 1.23 : Primitive Vecteur et attributs associés

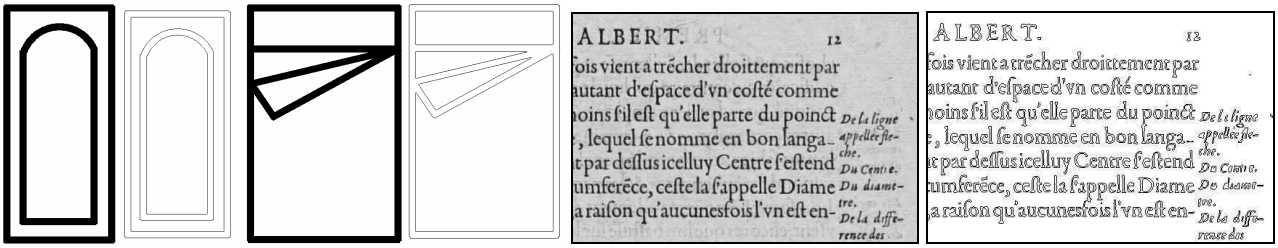


Figure 1.24: Représentation des formes graphiques et textuelles par des vecteurs

Par ce traitement, si l'approximation polygonale est de bonne qualité, toutes les informations acquises sur l'image sont conservées de manière très structurée comme décrit figure 1.24. Il est également possible de faire le lien entre toute forme (composante connexe) et ses contours représentés par des suites de vecteurs décrits par un ensemble d'attributs.
La représentation de l'image sous forme de liste de *Vecteurs* correspondant aux contours des formes permet de limiter la perte d'informations et donne une approximation très fidèle du document, mais elle n'est pas facilement exploitable et ne fournit pas suffisamment d'information sur la structure du document.

Une étape d'appariement des vecteurs pour obtenir des quadrilatères fait encore progresser notre représentation du document. C'est l'association de deux vecteurs constituant les frontières opposées d'un trait qui produit un Quadrilatère comme illustré figure 1.25.

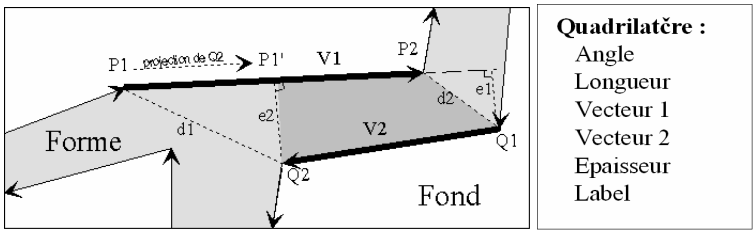


Figure 1.25 : Primitives Quadrilatères : construction et attributs

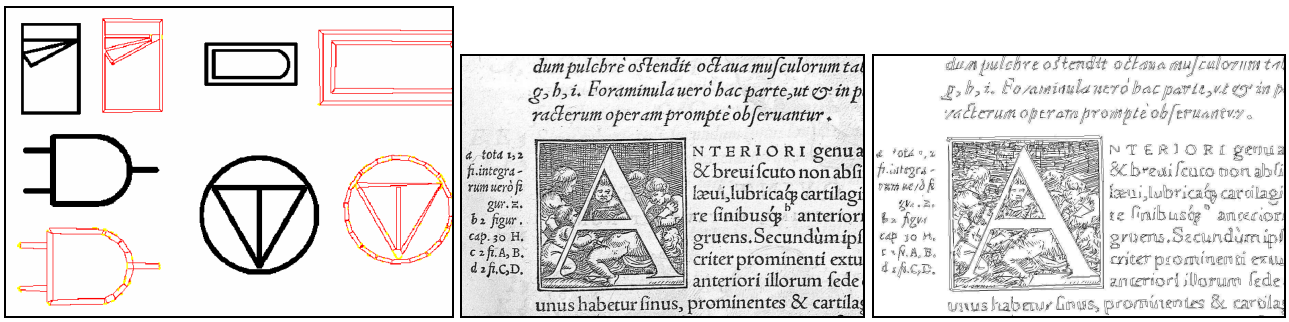


Figure 1.26: Représentation des formes graphiques et textuelles par des Quadrilatères

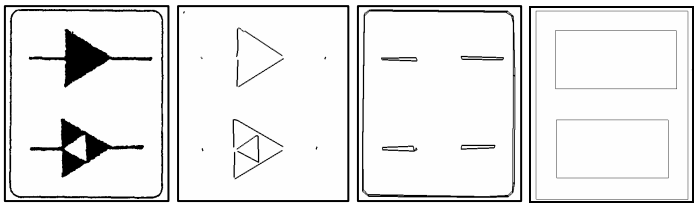


Figure 1.27 : Représentation des formes à l'aide de Vecteurs, Quadrilatères et Composantes connexes

L'ordonnancement des quadrilatères construits permet d'introduire, comme pour les vecteurs, une phase de fusion (conditionnée par les attributs associés à chaque quadrilatère) pour achever la mise en place d'une description des formes. Les quadrilatères ainsi ordonnés représentent de manière concise et robuste les traits présents dans l'image initiale (figure 1.26), tandis que les vecteurs non appariés représentent les frontières des formes pleines (figure 1.27). Ici aussi, la solution proposée résout les problèmes classiques des méthodes de vectorisation par suivi de contours notamment ceux mentionnés dans [Song02b] et [Tombre00a]. Pour avoir plus de détails sur les algorithmes impliqués dans cette phase, le lecteur pourra consulter [Ramel00a].

A la fin de cette phase, on obtient une description pertinente des formes contenues dans les images de documents imprimés ou graphiques (figure 1.27) via :

- La liste de toutes les composantes connexes présentes dans l'image carte des formes,
- Un ensemble de quadrilatères représentatifs des parties fines (segments) constituant chaque forme carte des Quadrilatères,

- Un ensemble de vecteurs décrivant les contours des formes pleines carte des Vecteurs.

Cartes des Régions

Afin de compléter l'approche « contour », dans le cadre de sa thèse, Nicholas Journet propose d'utiliser des informations de texture pour caractériser les images de documents [Journet05c, Journet06].

Le but des travaux de Nicholas Journet est de réaliser une caractérisation robuste du contenu des documents provenant d'horizons très différents, donc fortement hétérogènes entre eux. Afin de rester suffisamment générique, la méthode ne doit comporter ni seuil, ni modèle, ni structure explicite dans le processus d'analyse. Le processus global consiste donc, dans un premier temps, à caractériser précisément les contenus des pages en regroupant les régions de l'image comportant les mêmes caractéristiques de texture. Il est difficile d'exprimer une caractéristique qui soit générique d'un document à l'autre sans prendre en compte différentes échelles d'observation. Cette contrainte s'estompe avec l'introduction d'une approche multi-résolution. Ainsi, 4 cartes de textures sont calculées à 5 résolutions différentes. Chaque carte exprime le résultat d'une analyse spécifique (fréquence, orientation, …) sous la forme d'une valeur par pixel obtenue par analyse du voisinage local du pixel. Pour cela, excepté pour la carte du fond, la constitution des cartes est réalisée à l'aide de fenêtres glissantes de différentes tailles centrées sur le point traité.

Les cartes liées aux textures

Dans la littérature, la caractérisation des orientations est souvent réalisée à travers un filtre directionnel de type Gabor [Basa04, Eternad97]. Cependant, ce type de filtre nécessite un choix judicieux du banc de filtres à appliquer afin de mettre en évidence les réponses de la convolution à une fréquence et à une orientation. N. Journet a choisi un outil moins paramétrique basé sur la rose de directions proposée par S. Bres [Bres94]. La rose de directions est un diagramme polaire illustrant la fonction d'auto-corrélation d'une image. La figure 1.28 illustre la réponse de cette méthode à différents types d'images.

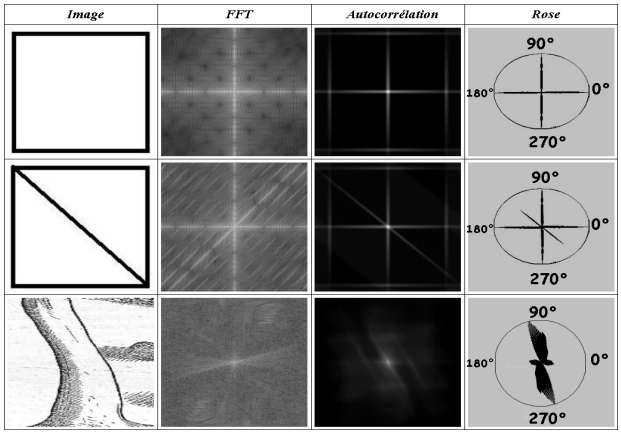


Figure 1.28 : Exemples de roses de directions [Journet05a]

A partir de ce diagramme, trois caractéristques liées aux orientations sont extraites permettant de constituer 3 cartes (figure 1.29) : la carte des orientations principales construite à partir de l'orientation du pic principal dans la rose de directions, la carte des amplitudes liée à l'amplitude du pic principal dans la rose, la carte des variances liées à l'épaisseur du pic principal dans la rose. Ces informations sont complétées par deux autres types de cartes que l'on retrouve plus classiquement en analyse de documents mais qui restent dans notre cas non paramétriques. Il s'agit, d'une part, de l'analyse des longueurs de plages (RLE), et d'autre part, d'une analyse du fond du document. L'analyse des plages de pixels noires fournit une information sur la compacité de l'information et permet de différencier les éléments composés par de nombreux traits (graphiques, fontes de petite taille) de ceux ayant une densité moins importante de traits (gros caractères, ..).

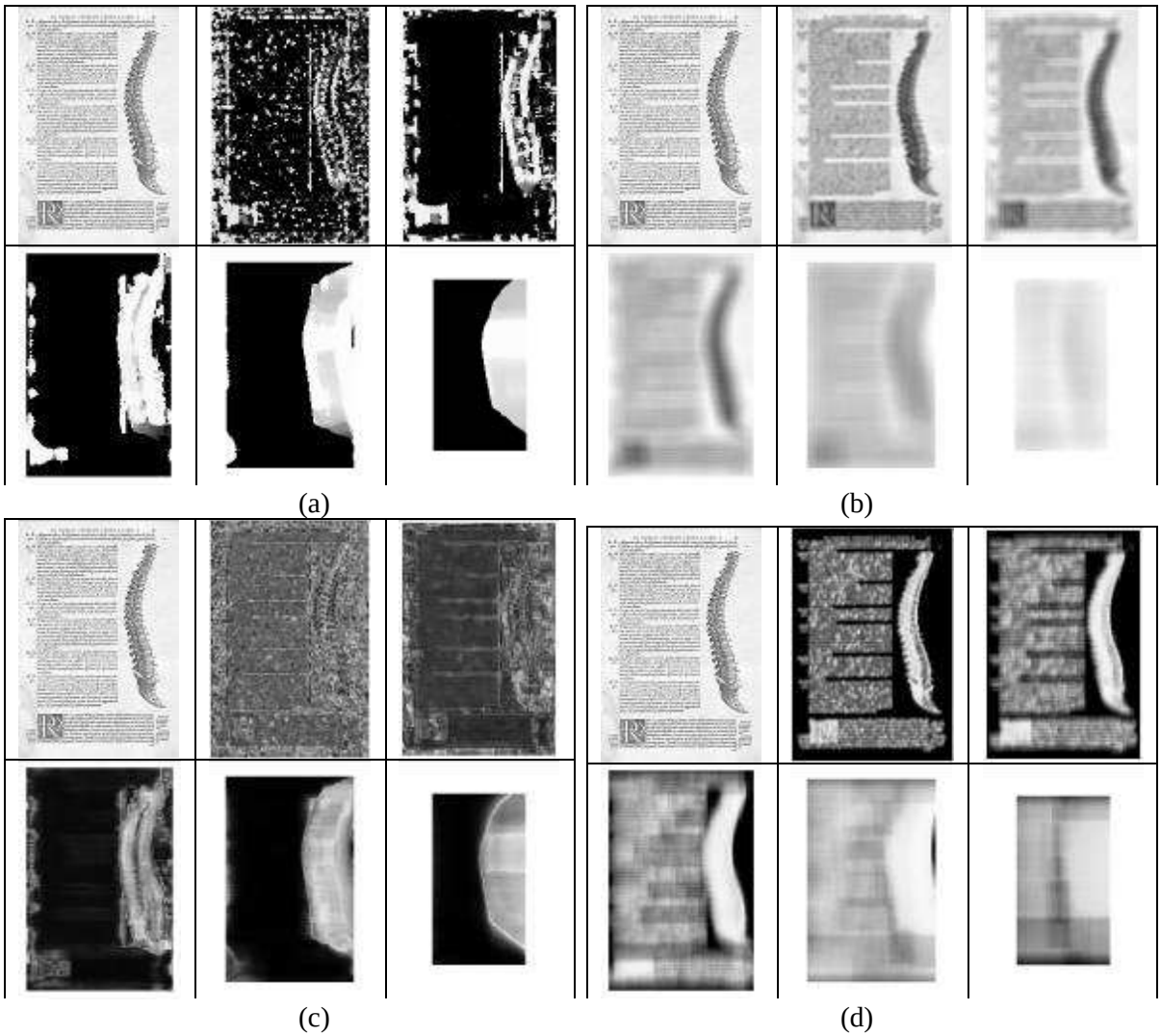


Figure 1.29 : Cartes des textures : (a) orientations, (b) amplitudes, (c) variances, (d) plages noires

Carte du fond

Normalement, un grand nombre de pixels clairs sont présents verticalement (*lgb_v*) ou horizontalement (*lgb_h*) dans les zones séparant deux formes (texte, graphiques, …). Par exemple, dans les images de documents anciens binarisés, un grand nombre de pixels blancs sont alignés horizontalement dans les

zones séparant deux paragraphes de texte. Par opposition, le nombre de pixels blancs que l'on est capable d'aligner dans un paragraphe entre deux lettres d'un mot ou entre deux mots d'une phrase ou entre deux lignes d'un même paragraphe est faible. Par conséquent, une valeur correspondant à la somme du nombre de pixels blancs successifs alignés horizontalement plus le nombre de pixels blancs successifs alignés verticalement est associée à chaque pixel de l'image. *C*omme le montre la figure 1.30, les niveaux de gris (*Ng(i,j)*) obtenus sont représentatifs de l'importance des séparations entre les formes. Ils constituent une carte du fond qui traduit les frontières (plus ou moins marquées) entre les différentes régions noires de l'image.

Notons que jusqu'à cette étape, aucune hypothèse sur le type de document n'a été faite et que cette approche peut être utilisée aussi bien sur les documents graphiques (dessins) que sur les documents imprimés comme le montre la figure 1.22.

Figure 1.30 : Superposition des cartes des formes et du fond sur des documents textuels et graphiques

Construction de graphes structurels

Une information est manquante dans les propositions faites jusqu'à présent pour décrire le contenu des documents. Il s'agit des informations concernant le voisinage ou le contexte associé à chaque élément ou primitive de description ; c'est ce qui fait la force des méthodes structurelles par rapport aux méthodes statistiques. Afin d'obtenir et d'organiser ce type d'informations, une représentation sous la forme d'un graphe paraît adaptée pour lier les différents objets (formes et primitives) qui constituent les documents. C'est aussi la plus apte à traduire les relations structurelles entre les primitives constituant les représentations. Pour l'instant, chaque type de primitives (Composantes connexes, Vecteurs et Quadrilatères) est traité indépendamment et on obtient 3 graphes structurels mais rien n'empêche, à terme, de prévoir une unification de ces graphes.

- Pour les Quadrilatères et les Vecteurs

Le premier objectif de la thèse de Rashid Jalal Qureshi [Qureshi05] consiste à mener une réflexion sur la meilleure manière de représenter les relations existant entre les primitives vectorielles (Vecteurs,

Quadrilatères) présentes dans la représentation d'un document à l'aide de graphes. Les principaux problèmes à résoudre sont :

- le choix des attributs à associer aux nœuds du graphe représentant les primitives structurelles
- le choix de la méthode de construction des arcs reliant les nœuds et des attributs à associer à ces arcs. Une piste intéressante, et encore inexplorée, serait d'utiliser certaines des cartes des textures proposées par Nicholas Journet, en plus de la proximité entre les objets pour décider de l'existence d'un arc ou non.

Actuellement, pour construire le graphe des Vecteurs et des Quadrilatères, une zone d'influence (un voisinage) est générée pour chacune des primitives. Des arcs décrivent la nature des liens existant entre la primitive en cours d'étude et celles appartenant à sa zone d'influence (rectangle pointillé). La figure 1.31 décrit le processus de construction de la partie correspondant au Quadrilatère 0.

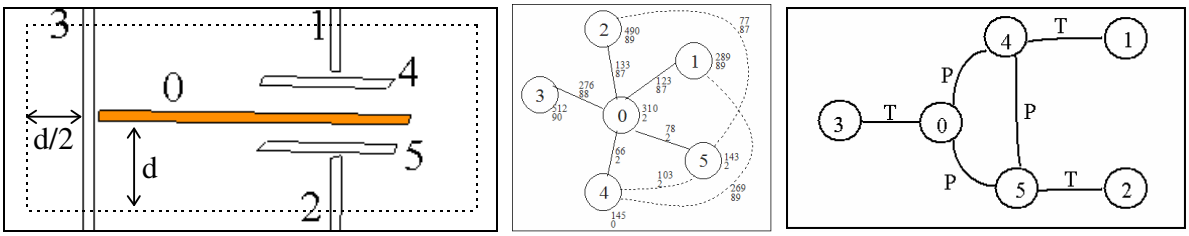


Figure 1.31 : Zone d'influence et graphes avec attributs numériques ou symboliques associés au Quadrilatère 0.

En ce qui concerne les attributs à associer aux arcs et aux nœuds du graphe, il existe différentes possibilités comme le montre la figure 1.31 (attributs statistiques et/ou symboliques). Ces différentes possibilités sont étudiées dans la thèse de Rashid Jalal Qureshi. Après la construction de l'ensemble des sous-graphes pour les quadrilatères, une représentation structurelle de la globalité du contenu de l'image est disponible.

- Pour les composantes connexes

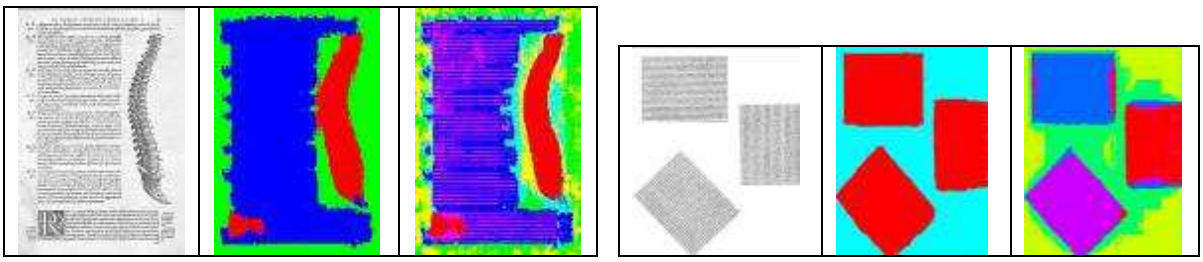
Une réflexion similaire doit être menée si l'on désire également construire un graphe mettant en relation les composantes connexes représentant les formes. On rejoint ici l'étude menée par Stéphane Leriche présentée section 1.4.2. Pour l'instant, seules les cartes du fond et des composantes connexes ont été utilisées conjointement pour construire un graphe des composantes connexes. Ces travaux démontrent l'intérêt de coupler des informations « région » (texture) et « contour » (formes) lors de la construction du graphe structurel comme indiqué figure 1.22.

1.4.2. Exemples d'exploitation de la représentation obtenue

Cartes des régions

Les cartes de textures donnent des indices assez précis sur l'information présente dans l'image traitée. L'objectif étant de catégoriser les EdC des documents. La pertinence des informations extraites par chaque carte (et la redondance entre cartes) a été vérifiée ainsi que leur capacité de caractérisation des

contenus. L'objectif est de regrouper les pixels de l'image correspondant à des zones homogènes ce qui revient à regrouper les vecteurs proches au sens d'une métrique. Ce problème est un problème de classification non supervisée pour lequel les étiquettes des points permettant de construire les classes ne sont pas connues a priori.



(a) (b)

Figure 1.32 : Classification en régions avec 3 et 6 classes [Journet06]

La figure 1.32 montre le type de résultats que l'on peut obtenir à l'aide de cette méthode. La première image (a) est une image mixte comportant du texte, des graphiques et une lettrine. L'image (b), quant à elle, montre que la méthode permet de détecter du texte dans différentes inclinaisons, sans paramétrisation.

Des études statistiques (ACP) ont montré une corrélation forte entre les cartes liées à une même caractéristique extraite à différentes résolutions. Cependant, cette redondance reste nécessaire lorsque l'on cherche à discriminer plus finement les EdC (par exemple distinguer les polices des caractères). Il s'agit ici des travaux de Nicholas Journet [Journet05a, Journet05b, Journet05c] qui se focalisent sur la caractérisation d'images de documents anciens par analyse de texture et n'ont pas pour objectif unique de produire une représentation générique pour les images de documents. Toutes les cartes (informations de texture) qu'il propose ne sont donc pas indispensables dans le cadre de la mise en place d'une représentation générique pour les images de documents.

Cartes des formes

Un autre travail a montré que l'utilisation conjointe de la carte des formes et de la carte du fond permet de segmenter finement les documents anciens. Il s'agit des travaux de Master 2 de S Leriche, menés en collaboration avec le CESR de Tours et qui se sont concrétisés par une publication dans la revue Traitement du Signal [Ramel05a]. Dans ce travail, la carte du fond et la carte des formes sont utilisées conjointement pour segmenter et labéliser les EdC dans les documents anciens Il est possible de partir de la liste des composantes connexes étiquetées *Texte* pour reconstruire les paragraphes par association de composantes connexes susceptibles d'être des caractères. Pour qu'une association soit réalisée, il faut que les deux composantes soient assez proches l'une de l'autre et que le segment reliant leurs centres de gravité (G_1 et G_2) ne traverse pas une frontière trop marquée dans la carte du fond (niveaux de gris faibles). Cette contrainte multicritère sur laquelle repose notre méthode est traduite sous forme d'un produit qui s'exprime par :

$$d(G_1, G_2) \times (256 - \underset{(i,j)\in[G_1,G_2]}{Min} [Ng(i,j)]) \leq Seuil$$

d désigne la distance euclidienne entre les centres de gravité des deux ensembles à associer. Lorsque ce critère est vérifié pour le plus proche voisin d'une composante, on crée une relation (arc) entre les 2 primitives (nœuds) sinon l'association est refusée. Un graphe des composantes est ainsi obtenu pour représenter chaque EdC (bloc de texte). La recherche des voisins pour chaque composante connexe peut se faire, successivement, horizontalement (fusion horizontale) puis verticalement (fusion verticale) de manière itérative et est stoppée lorsqu'aucune association ne peut avoir lieu. Il est possible d'utiliser des seuils adaptés pour chaque type d'association (horizontale et verticale). Les zones obtenues ne sont pas forcément rectangulaires puisqu'elles correspondent simplement à un rassemblement de composantes connexes reliées par des arcs.

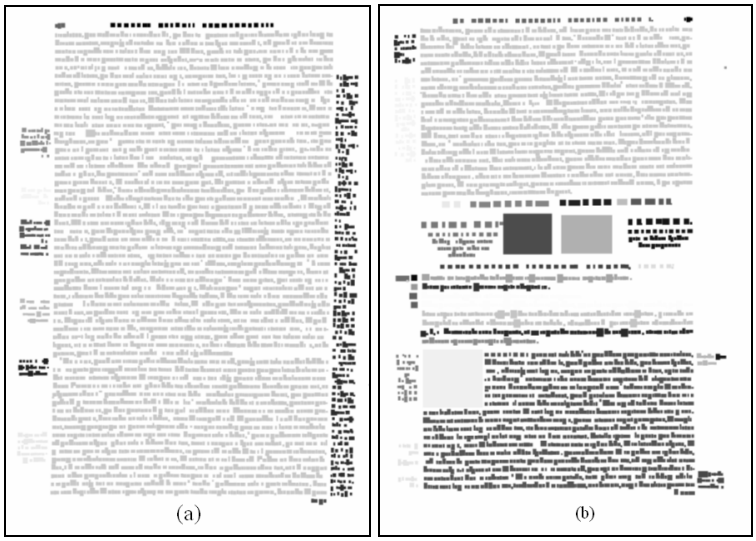


Figure 1.33 : Exemples de segmentations en zones.

Deux exemples de résultats obtenus sur des documents anciens sont fournis figure 1.33. Afin de visualiser la segmentation obtenue, chaque sous-graphe regroupant les formes d'une zone donnée correspond à un niveau de gris différent.

Représentation de tracés en ligne

Il n'est pas difficile de voir l'intérêt que pourrait avoir une représentation telle que celle proposée juste avant sur des documents en ligne. En effet, un graphe structurel peut être construit encore plus facilement par approximation polygonale, sur des tracés en ligne que sur des images bitmap. L'approximation fournit, comme en hors ligne, un ensemble ordonné de vecteurs représentatifs des tracés constituant le document initial. Une représentation à l'aide des primitives (Vecteurs) réduit la quantité de données à traiter et simplifie la gestion et l'utilisation des attributs associés aux différents tracés. On peut ainsi réduire (et/ou lisser) la quantité d'informations à traiter tout en conservant une représentation plus ou moins précise des formes produites (figure 1.34). Cependant, à chaque point peut

correspondre un ensemble important d'attributs comme la pression exercée sur le stylo, l'inclinaison du stylo, la couleur, … devant être géré par l'algorithme transformant les données initiales en un graphe. Une solution fédérative, applicable aussi bien sur les documents papier que sur les documents produits en ligne, est ainsi obtenue.

Cette vectorisation des tracés contribue également à la mise en place d'une nouvelle manière d'appréhender le traitement des documents en ligne en proposant une première étape de construction d'une représentation structurelle à partir des données initiales pour fournir plus de connaissances (ou en tout cas des données plus structurées) aux processus effectuant l'analyse et la reconnaissance.

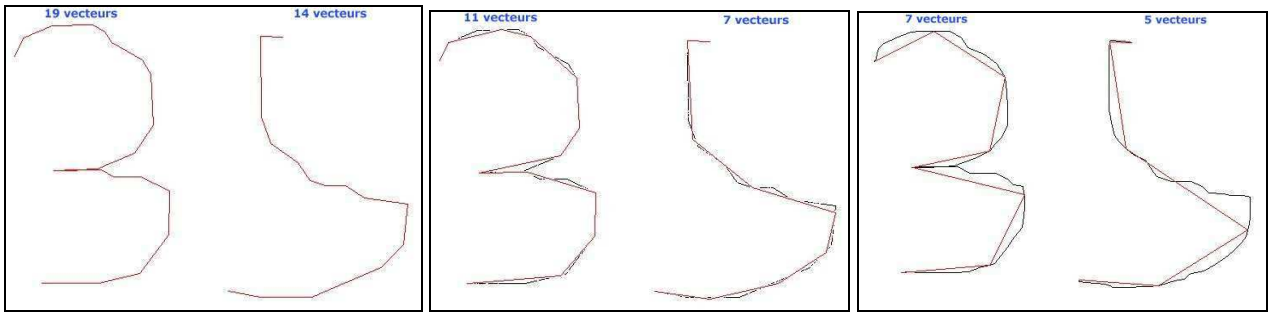


Figure 1.34 : Maîtrise du compromis qualité/quantité

Comme le montre la figure 1.35, un graphe structurel peut permettre d'effectuer une remise en forme (pour passer d'un brouillon griffonné à un dessin moins fluctuant, plus rectiligne) du dessin une fois certains tracés reconnus (intersections, courbes, lignes parallèles, perpendiculaires, …).

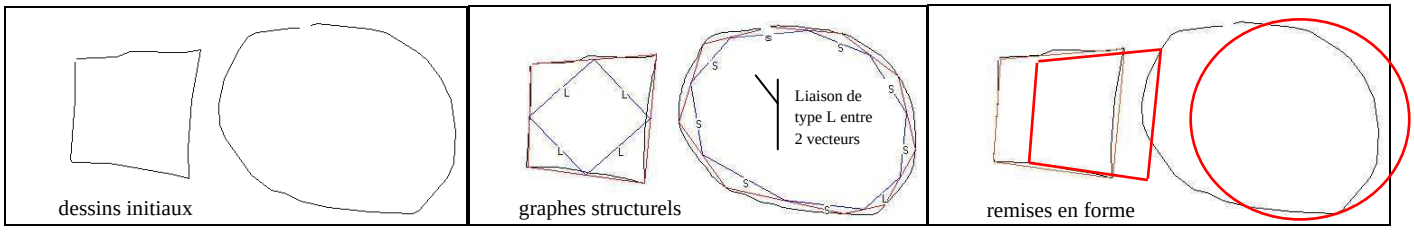


Figure 1.35 : Remise en forme de dessins

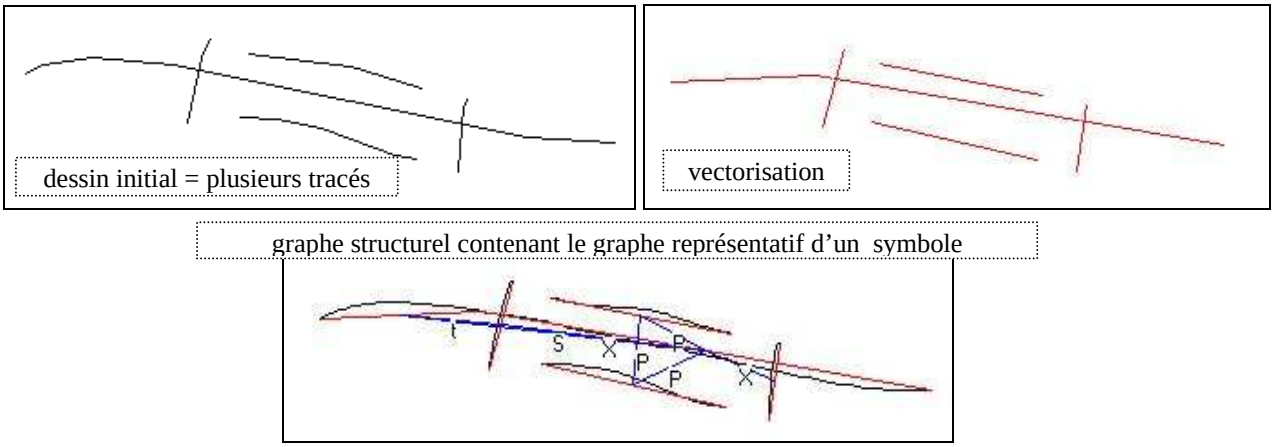


Figure 1.36 : Graphe structurel et reconnaissance d'un EdC (un symbole) composé de plusieurs tracés

La méthode structurelle proposée apporte également une solution simple à la problématique de la reconnaissance d'EdC composés de plusieurs tracés (figure 1.36). Il suffit de construire le graphe correspondant à l'ensemble des tracés élémentaires pour réaliser leur fusion et obtenir ainsi une représentation de l'EdC composé de plusieurs tracés puis de le reconnaître.

Recherche d'une représentation optimale pour les signatures manuscrites en ligne

La recherche d'une représentation performante adaptée aux tracés en ligne s'est poursuivie durant la thèse de Matthieu Wirotius concernant l'authentification de signatures manuscrites. Ce travail a eu lieu dans le cadre d'une bourse CIFRE avec la société ATOS ORIGIN et avait comme objectif l'authentification de scripteurs à partir de leurs signatures manuscrites réalisées sur PDA et tablet PC.
L'évaluation des différentes représentations proposées a été effectuée grâce à l'utilisation de l'algorithme DTW [Wirotius04ba]. Matthieu Wirotius a d'ailleurs proposé plusieurs améliorations à apporter à cet algorithme (détaillées plus loin). Il aborde aussi bien d'autres points qui ne seront pas détaillés ici [Wirotius04a].

Peu d'articles traitent du problème de la sélection de points représentatifs des tracés en ligne dans le but de faciliter, voire d'améliorer leur analyse durant la phase de reconnaissance. Habituellement, les points les plus intéressants sont supposés être ceux qui permettent d'effectuer une segmentation et d'étudier les différentes parties indépendamment [Lee04]. Selon nous, la réduction du nombre de points a deux principaux objectifs. Tout d'abord, éviter de conserver l'intégralité des points de la signature afin de minimiser la quantité de données à conserver et la durée du processus d'authentification. De plus, les données brutes sont très précises et de légères variations peuvent perturber la phase de comparaison. Par conséquent, une réduction du nombre de points peut permettre d'accroître la robustesse du système et lui donner une certaine invariance vis à vis des conditions et du matériel d'acquisition. La prise en compte des points stables de la signature peut permettre d'augmenter l'efficacité de la méthode d'authentification [Dimauro02].

La difficulté de mise en place d'une représentation vient de la difficulté de choix des critères devant être forcément respectés. Les critères que nous avons choisis sont la stabilité et les représentativités spatiale et dynamique des points ou des parties à conserver. Le but d'une approximation polygonale est justement de rechercher une forme globale qui coïncide au mieux avec la forme d'origine tout en limitant la perte d'information. Ainsi, celle-ci, appliquée à un tracé devrait donner un bon compromis entre précision et réduction du volume de données à conserver. Matthieu Wirotius a évalué la construction d'une représentation des signatures basée sur l'approximation polygonale proposée par Wall [Wall84]. Pour montrer la pertinence de cette méthode (*Wall*), des comparaisons avec d'autres méthodes de sélection de points ont été réalisées. La première consiste à choisir les points de manière aléatoire (*Aléatoire*), la seconde à utiliser un algorithme génétique (*AG*), la troisième est basée sur la méthode de Brault (*Brault* -

étude locale de la courbure) et la dernière repose sur l'analyse locale de la vitesse (*Vmin*) en chaque point.

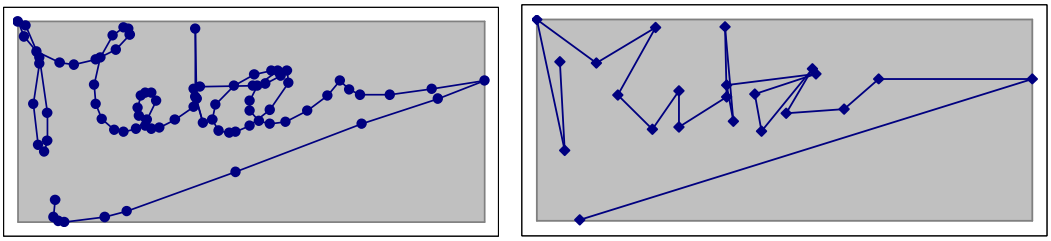


Figure 1.37: Représentation simplifiée des tracés par approximation polygonale.

La méthode de sélection des points de contrôle basée sur les minima de la vitesse instantanée, suivie de près par la méthode basée sur l'algorithme de Wall, donne les meilleurs résultats (meilleurs taux de reconnaissance – voir section 2.4.3). L'usage de ces représentations (figure 1.37) permet de simplifier la comparaison des signatures les unes avec les autres par élimination de l'information redondante, c'est-à-dire inutile et perturbante [Wirotius05c].

Ces différents exemples illustrent l'utilisabilité de la représentation, des descripteurs et des primitives proposés sur différents types de documents numériques. Ils témoignent de la possibilité de mise en place d'une représentation générique et des avantages que l'on peut obtenir en utilisant une telle structuration des informations. Il reste cependant à voir comment ces informations peuvent être utilisées durant la phase de reconnaissance des contenus.

Chapitre 2 :

Analyse et reconnaissance des contenus

Cette partie présente les stratégies d'analyse et les méthodes de reconnaissance qu'il est possible d'utiliser sur les images de documents ou les tracés en ligne. Nagy [Nagy95] proposait, il y a déjà 10 ans, de mettre de côté les recherches concernant les traitements bas niveau et l'extraction de primitives (inclinaison, filtrage, binarisation [Sauvola00, Kasturi90], vectorisation [Hilaire02, Kadonaga95, Song02b]) pour lesquels de nombreuses méthodes existent déjà, et de s'intéresser plus spécialement aux aspects coopération entre processus d'analyse des formes qui ont été beaucoup moins abordés. Il semble, en effet, que la manière d'associer les traitements est d'une importance cruciale et une étude poussée sur ce plan permettrait sans doute de franchir un cap.

Après quelques brefs rappels sur les principales architectures et méthodes utilisables pour reconnaître le contenu d'un document, ce chapitre présente les principaux axes de recherche que j'ai retenus dans cette problématique :

- les stratégies et les architectures des systèmes de reconnaissance notamment celles offrant une meilleure prise en compte du contexte ou un fort niveau d'interaction entre les algorithmes et les utilisateurs finaux,
- les méthodes de reconnaissance et leur adaptation afin de pouvoir prendre en compte efficacement différents types de primitives et d'attributs statistiques et symboliques lors du calcul de similarité entre EdC.

2.1. Stratégies d'analyse et architectures des systèmes

Un groupe de travail du GDR ISIS (GDR 10.2) animé par C. Garbay [Garbay96] a réalisé, il y a presque 10 ans déjà, un travail très intéressant sur les architectures des systèmes d'analyse d'images. Ce travail se poursuit actuellement dans le cadre d'un autre groupe de travail du GDR ISIS (GT SCATI). Les thèmes de recherche portés par ces GT se structurent autour de la notion de systèmes pour le traitement et l'analyse d'images. Il s'agit avant tout d'associer un caractère systémique à cette thématique et de

coupler les outils et les méthodes relevant du traitement d'images avec des mécanismes issus de la fusion de données, de l'IA, de l'ingénierie de la connaissance, de la cognition et de la perception visuelle, … Mes recherches ont été assez fortement influencées par ces travaux. Ils ont provoqué la volonté de proposer des méthodologies donnant une dimension nouvelle aux processus d'interprétation des documents et d'effectuer une modélisation des connaissances et des systèmes d'analyse en prenant explicitement en compte le rôle de l'utilisateur dans la boucle de traitements. Une attention particulière doit, à mon avis, être portée sur l'explicitation des connaissances utilisées afin d'envisager l'adaptabilité des traitements et des processus sur la base d'acquisition ou d'explicitation des contextes d'utilisation. Il s'agit aussi d'un souci de constitution de modèles génériques qui favorisent la reproductibilité et la réutilisation des schémas utilisés sur différents types de documents.

L'étape de classification ou de reconnaissance est généralement spécifique à un domaine d'application et utilise des règles et des connaissances a priori pour affecter un label aux différents EdC par sélection dans une liste d'étiquettes prédéfinies par le concepteur du système (liste de symboles, types de paragraphes,...) [Joseph92, Nagy93, Couasnon03]. Ces limitations apparaissent très restrictives lorsque l'on désire traiter des documents peu structurés ou lorsque l'on cherche à rester générique. Il est difficile de proposer un schéma général de l'architecture d'un système d'analyse d'images : la plupart des architectures comportent une librairie de traitements, un système de supervision et une base de connaissances et proposent un découpage en deux phases du processus [Saidali04, Bruce99, Clouard99, Adam01] : le traitement d'image et la reconnaissance. Il existe deux grandes familles d'approches relatives à ces traitements : statistiques-connexionnistes et syntaxiques-structurelles. Les premières décrivent les objets sous la forme de vecteurs de caractéristiques et les deuxièmes sous la forme d'un ensemble de primitives.

Comme tous les systèmes de reconnaissance de formes, la reconnaissance de contenu ou de structure de documents nécessite d'abord la création d'une base de connaissances à partir d'un apprentissage ou d'une modélisation des éléments à reconnaître. Dans une deuxième étape, il faut définir la méthode de reconnaissance en fonction de la complexité et de la variabilité des contenus et des descripteurs d'images choisis. Ces deux étapes sont en fait liées : c'est uniquement grâce à l'observation des différents contenus d'une classe de documents, que l'on peut en déduire la méthode de reconnaissance la plus adaptée et donc définir le modèle approprié. C'est la raison pour laquelle le choix de la méthode de reconnaissance semble indissociable du type de document, de la complexité et de la variabilité de sa structure. Un autre point, selon moi important, est que la reconnaissance de la fonction logique d'un objet (par exemple un bloc de texte) dépend de ses caractéristiques propres mais aussi de la fonction logique des objets voisins déjà reconnus. La reconnaissance ne peut donc pas s'effectuer élément par élément indépendamment les uns des autres. Elle nécessite la prise en compte des relations spatiales entre les différentes composantes de la structure. Cela a abouti à différentes stratégies de reconnaissance

auxquelles correspondent plusieurs catégories d'architectures logicielles. Les systèmes d'analyse de documents actuels peuvent être divisés en 3 catégories principales [Nagy00] : les systèmes guidés par les données (data-driven), les systèmes guidés par le modèle (model-driven) et les systèmes hybrides.

2.1.1. Analyse guidée par les données

Pour faire partie de cette catégorie, une méthode doit travailler principalement à partir des informations fournies par les pixels de l'image, dans le but de produire des éléments de niveau sémantique de plus en plus élevé tout en utilisant un minimum de connaissance a priori sur le modèle du document [Simon97].
Dans ce cas, comme nous l'avons déjà vu, les différentes parties d'un objet peuvent être dispatchées dans différentes couches (par exemple, un caractère touchant une autre partie du tracé sera dans la couche « formes fines » plutôt que dans la couche « petites composantes connexes ») ; ce qui complexifie énormément la phase de reconnaissance. Les incohérences entre les informations apportées par chaque couche ajoutent une difficulté supplémentaire sachant que la remise en cause de la segmentation en couches est quasiment impossible sans recommencer tout le traitement.
Ces approches se limitent souvent à l'analyse de documents peu bruités ou ne produisent pas des sorties de niveau sémantique suffisamment élevé. Comme on peut le voir figure 2.1a, les algorithmes de bas niveau (Px) utilisent uniquement l'image et quelques connaissances a priori sur le contenu des images pour produire la segmentation. Les traitements de niveau supérieur (Qx) utilisent la segmentation produite et des règles lexicales et syntaxiques pour effectuer l'interprétation [Bunke99, Llados02, Marinai05]. Parfois une analyse sémantique est effectuée pour valider les résultats obtenus [Ogier00, Dori92, Tombre92].
Il est regrettable qu'aucun lien n'existe entre les algorithmes de bas et de haut niveaux et, encore pire, qu'aucun lien n'existe entre les processus d'un même niveau ! Les critiques supplémentaires émises à l'égard de ce type de méthodes concernent les temps de calcul et la difficulté de choix des valeurs des nombreux paramètres à régler.

2.1.2. Analyse guidée par le modèle

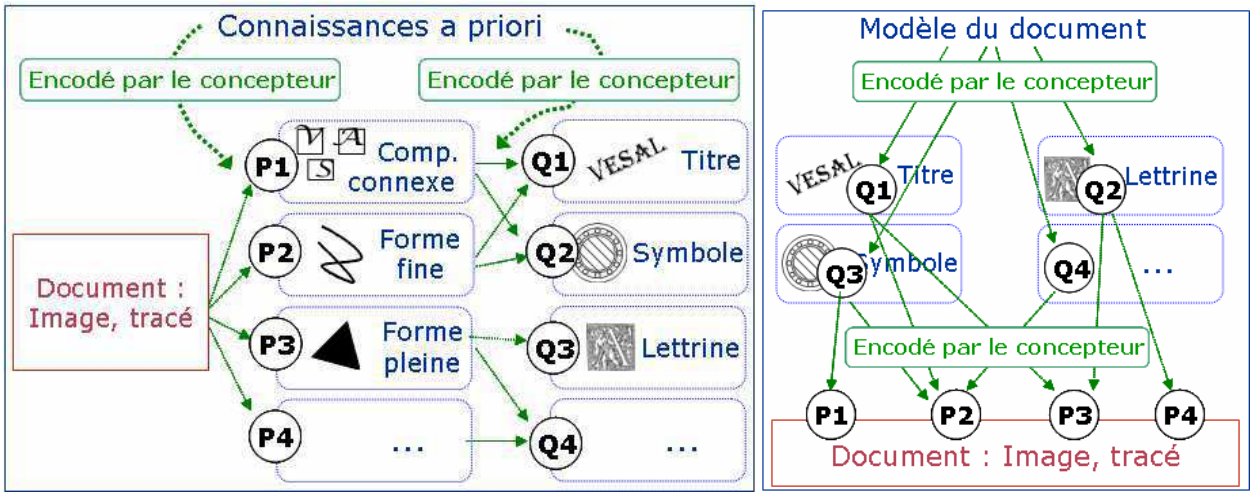
Avec cette approche, ce sont les connaissances (le modèle) qui sont utilisées pour guider la segmentation et l'analyse du contenu du document (figure 2.1b) [Joseph92] [Shimotsuji94] [DenHartog96] [Lee00]. Ainsi, l'analyse est achevée plutôt par découpages itératifs du contenu afin de retrouver les éléments spécifiés dans le modèle. Les EdC de haut niveau sont recherchés directement dans l'image (pas de représentation intermédiaire) et l'on peut mentionner deux façons de procéder :

- Recherche de configurations particulières de pixels, de textures, de boucles, d'espaces blancs à des localisations précises de l'image et correspondant à des EdC potentiels définis dans le modèle [Hajdar02, Shih89, Baird90].

- Extraction des composantes connexes et comparaison avec des modèles de formes à reconnaitre [Nagy95, Tabbone04, Adam00b].

Si l'image a pu être segmentée, c'est qu'elle est compatible avec le modèle de document. Le processus de reconnaissance est complètement différent de l'approche précédente, car on cherche dans le document, les informations décrites dans le modèle afin de valider leur existence. L'analyse des images est, donc, complètement contrainte par le modèle de connaissances a priori. La segmentation de l'image est dans ce cas très précise, mais elle peut échouer et conduire à un rejet complet de l'analyse quand certains éléments n'ont pas été trouvés dans l'image. Ce genre d'approche rend possible la mise en place d'une phase d'apprentissage des modèles de documents et conduit à une grande précision dans les structures trouvées mais donne lieu aussi à de nombreux rejets de documents qui diffèrent sensiblement du modèle. Le problème vient également de la spécificité des systèmes développés qui ne peuvent traiter qu'un type bien précis de documents.

L'idée de partir du modèle (connaissances lexicales, syntaxiques ou sémantiques) a fait son chemin et dans les années 90, les techniques d'interprétation de documents étaient souvent réalisées par recherche de couches non plus en fonction de critères géométriques (couche des traits forts, couche des traits fins, ...), c'est à dire, de traitements à mettre en place, mais plutôt sur des critères syntaxiques, c'est-à-dire en fonction du sens des informations obtenues en effectuant des regroupements judicieux des formes de l'image sur la base d'un modèle (sa structure logique par exemple). On ne parle alors plus de structure mais plutôt de syntaxe propre au document. Ces règles varient selon le type de documents et d'applications auxquels on s'intéresse : les bases de données d'un système de simulation de circuits électriques sont totalement différentes de celles d'un système de GED ou même de celles d'un système de CAO d'un autre type. Ici, la connaissance de ces différentes règles permet non seulement de vérifier la cohérence de la représentation obtenue mais elle guide aussi les traitements.



(a) (b)

Figure 2.1 : Analyse guidée par les données (a) ou guidée par le modèle (b)

Les systèmes de cette catégorie emploient, souvent de manière implicite, une liste de règles de conception ou de mise en page que le concepteur du système a consignée dans ses algorithmes et qui correspond à un domaine d'application précis voire même spécifiquement à la base d'images test utilisée. On trouve donc dans la littérature la description d'une multitude de systèmes dédiés, caractérisés par des objectifs différents ou par des hypothèses différentes faites sur les données d'entrée ou sur les données attendues en sortie.

Outre l'impossibilité de traiter des documents peu structurés, un autre point faible de ces techniques concerne le manque d'information concernant la gestion des régions qui ne correspondent à rien de spécifié dans le modèle.

La figure 2.1 résume l'architecture classique de ces deux types de systèmes d'analyse de documents. On voit au travers de cette figure la nécessité, dans tous les cas, de définir de manière statique un ensemble de traitements ainsi que l'ordre d'activation de ces traitements. Il semble, également, que l'on retrouve toujours au moins deux types d'algorithmes (niveaux d'analyse) : ceux liés à l'analyse d'image (Px) et ceux liés à la reconnaissance (Qx). Cependant, dans un cas, les algorithmes d'analyse d'images produisent des représentations intermédiaires utilisées par les algorithmes de reconnaissance et dans le deuxième cas, les algorithmes d'analyse d'images servent à valider les hypothèses produites à partir du modèle par les algorithmes de reconnaissance.

2.1.3. Intelligence Artificielle Distribuée (IAD), vision cognitive et approche psycho-visuelle

Nous avons déjà mentionné que lorsque l'on se réfère à l'Homme, les étapes ne se succèdent pas car elles sont étroitement liées : il existe un grand nombre d'interactions entre les différents processus de reconnaissance. La segmentation de l'image a besoin des connaissances a priori sur la structure des documents et inversement, l'étage de reconnaissance doit pouvoir remettre en cause la segmentation de l'image ou rechercher une information supplémentaire dans l'image pour affiner la décision finale. La reconnaissance des structures est donc un processus complexe qui nécessite l'analyse de différentes sources d'informations étroitement liées pour faire émerger la solution la plus probable. Les systèmes coopératifs émanant des recherches en intelligence artificielle (IA) semblent plus appropriés à la reconnaissance que les systèmes séquentiels, cependant, ils ont aussi beaucoup de limites et peuvent être parfois dépassés par de simples systèmes séquentiels adaptés à un problème particulier.

Le but de l'IA est, justement, l'acquisition par l'ordinateur de certaines compétences d'un être humain et de leur utilisation dans l'accomplissement d'une tâche nécessitant à la fois des connaissances, de l'expérience et une certaine capacité de raisonnement. Dans cette communauté, l'argumentaire suivant lequel, dans la réalité, les individus travaillent généralement en groupe, mettant en commun des savoirs et des savoir-faire afin de collaborer à la réalisation d'un même but, permet de justifier qu'une approche IA reproduise le fonctionnement de cette multiplicité d'experts co-oeuvrant : on parle alors de l'Intelligence Artificielle Distribuée.

Un système de résolution distribuée est conçu comme une communauté de systèmes travaillant de manière coordonnée ce qui pose alors de nouveaux problèmes théoriques [Germond00] : modélisation de la coopération entre les experts, organisation des données et des traitements (hiérarchiques ou non), nature de la structure de contrôle,...

Dans les systèmes multi-agents, chaque expert (agent) dispose d'une connaissance plus spécialisée mais parcellaire et utilise un mécanisme de raisonnement relativement simplifié. Cette simplification vient cependant compenser la complexification des interactions entre agents pendant la phase de résolution [Garbay98, Rosin04]. Chaque expert manipule généralement des données incomplètes du fait de sa vision partielle de l'environnement. Des données incertaines, fondées sur des informations variables propres à chaque expert, doivent être manipulées à un niveau local aussi bien qu'à un niveau global. Des données incorrectes peuvent être fournies par un agent, aussi il est nécessaire que ces données soient reconnues comme fausses à un niveau global.

L'interaction exploite alors la possibilité de mettre en jeu plusieurs points de vue sur une donnée ou sur la façon de la traiter : on parle de fusion dans le premier cas et de compétition dans le second. Ces deux procédés sont naturellement perçus comme complémentaires ; il s'agit dans les deux cas de multiplier les points de vue, qu'ils soient descriptifs ou opératoires, pour enrichir l'analyse et en exploiter les conflits éventuels. L'intérêt de ces approches est de gérer explicitement et de manière rationnelle les conflits de décision et la multiplicité des représentations.

La compétition de compétences consiste à produire des solutions (descriptions ou décisions) qui seront analysées par des techniques de fusion. Le principe est d'appliquer différents opérateurs sur un ensemble de données, puis de conserver le meilleur résultat ou un compromis des résultats obtenus.

La gestion des agents s'avère souvent laborieuse et d'autres approches ont vu le jour pour tenter d'apporter une réponse à cette difficulté. Les systèmes Blackboard (tableau noir) [Barr89] peuvent être considérés comme des méthodes de résolution de problèmes par construction incrémentale de solutions (évolution pas à pas vers la solution par résolution de sous-problèmes). Il s'agit de décrire l'état général d'un processus coopératif à un ensemble de spécialistes en train de résoudre un problème spécifique. Chaque spécialiste observe le Blackboard et est attentif à l'information qui pourrait l'intéresser (c'est-à-dire celle relative à son domaine). Aussitôt qu'une telle information apparaît, il exécute le travail prévu et transmet les résultats aux autres spécialistes. Une fois le travail terminé, chaque spécialiste écrit les résultats sur le tableau noir pour que les autres prennent connaissance des résultats. Ainsi, aucun spécialiste n'a besoin de connaître l'existence des autres spécialistes. Cela permet une intégration facile, à la fois des méthodes utilisant une approche d'analyse ascendante avec celles utilisant une approche descendante. L'analyse peut aussi bien être dirigée par les données que par les buts à atteindre.

Le système GRAPHEIN [Belaid94, Chenevoy92] est un bon exemple souvent utilisé pour illustrer ce type de mécanismes. Ce système a pour but de reconnaître la structure logique de documents imprimés. L'architecture logicielle du système se base sur un Blackboard qui permet de faire coopérer plusieurs

sources de connaissances par l'intermédiaire d'une gestion de tâches et de spécialistes. Celle-ci est composée de plusieurs modules :

- le modèle décrit les structures physiques et logiques de la classe de document, leurs attributs et leurs relations. Il est fortement lié à l'application.
- les spécialistes sont des programmes autonomes qui ont pour fonction d'effectuer, sur demande, une tâche spécifique de bas niveau (mesure de l'inclinaison, RLSA, capture et analyse des connexités ...) ou de niveau plus élevé (vérification d'hypothèses).
- les modules liés au contrôle sont assez complexes : le Blackboard est représenté sous forme d'un graphe d'hypothèses qui contient l'ensemble des solutions déjà examinées. Les spécialistes peuvent lire, créer, modifier ou supprimer des hypothèses stockées dans le graphe. Un événement particulier du Blackboard déclenche l'activation d'une tâche que d'autres spécialistes doivent effectuer. Il existe des tâches spécifiques pour l'analyse d'une page, d'une région, d'un bloc, de lignes, de mots ou de caractères.

Un sélecteur de tâches déduit la stratégie de contrôle à poursuivre à partir de mesures de la complexité ou de la plausibilité des hypothèses du Blackboard. Le choix des tâches se fait de manière opportuniste à partir des derniers changements (événements) opérés dans le Blackboard. Les hypothèses du Blackboard sont évaluées par un score de confiance qui tient compte de la probabilité de réalisation de l'hypothèse et de la complexité liée à son analyse.

Les recherches en vision cognitive proviennent du désir de certains chercheurs de combiner différentes disciplines telles que le traitement d'images, l'intelligence artificielle, l'apprentissage, les systèmes à base de connaissances et les sciences cognitives afin d'apporter des améliorations significatives aux systèmes de vision par ordinateur. Ainsi un système de vision cognitive doit obligatoirement mettre en jeu des mécanismes d'apprentissage (à partir d'expériences), des mécanismes d'auto-adaptation et des mécanismes de raisonnement. Grâce à ces différentes facultés, un système cognitif doit pouvoir détecter, localiser, reconnaitre et interpréter des scènes 2D ou 3D [Hudelot05]. Situées dans la mouvance de la vision active, les systèmes de vision cognitive mettent en œuvre des processus exploratoires capables d'analyser un environnement selon un point de vue et des objectifs précis. Les verrous scientifiques étudiés dans le cadre des recherches en vision cognitive concernent [Ecvision06] :

- la mise en place de techniques d'apprentissage progressif à partir d'expériences,
- la prise en compte des objectifs de l'utilisateur durant l'interprétation,
- l'identification des modules indispensables à un système de vision (de la détection à l'interprétation),
- les architectures des systèmes de vision de manière à les rendre plus interopérables (génériques)

Les approches psycho-visuelles ont pour objectif de rechercher une structure dans les documents par la simulation informatique du processus humain de lecture. Ces approches permettent d'obtenir une

caractérisation de la structure sans une segmentation préalable. Il s'agit de caractériser les différentes zones en fonction de leur pouvoir d'attraction par des mesures de complexité des formes, de lisibilité et de compacité indépendamment de l'alphabet utilisé [Eglin04]. Souvent, c'est l'analyse de la texture des images et l'extraction d'indices (dits « psycho-visuels ») qui permettent de trouver une organisation dans les documents correspondant ainsi le parcours visuel d'intérêt que pourrait réaliser l'observateur humain.

2.2. Méthodes de reconnaissance des contenus

Les travaux portant spécifiquement sur les techniques de reconnaissance des formes sont très nombreux et il est donc difficile de dresser un panorama exhaustif des méthodes existantes même en se focalisant sur certains aspects bien précis. Jusqu'à présent, mes travaux dans ce domaine sont assez restreints et se focalisent principalement sur la comparaison directe de tracés en ligne et la comparaison de graphes structurels représentant des images de documents. Aujourd'hui, il me semble pertinent d'essayer de coupler le plus possible les « approches structurelles » et les « approches statistiques » afin de tirer parti des avantages procurés par chacune de ces méthodes.

2.2.1. Méthodes statistiques

Les méthodes statistiques peuvent être classées selon différents critères : méthodes supervisées et non supervisées, paramétriques et non paramétriques, … Parmi les méthodes non supervisées les plus utilisées, on peut citer les méthodes de classifications hiérarchiques et les méthodes de partitionnements pour lesquelles le principal problème est de déterminer ou d'obtenir automatiquement le nombre de classes désiré. Avec les méthodes supervisées, une base d'apprentissage permet la détermination des lois de distribution des fonctions de densités des classes. Les classificateurs Bayesiens maximisent les probabilités a posteriori des classes et il est, ainsi, possible d'exploiter, d'une part, plusieurs modèles de formes pour chaque classe, et, d'autre part, des informations sur les caractéristiques discriminantes d'une forme par rapport aux autres. Cette phase d'apprentissage est beaucoup plus difficile à réaliser avec les méthodes structurelles. Depuis peu, les SVM (Suuport Vector Machine) subissent un engouement important car ils ont montré leur efficacité à plusieurs reprises même lorsque l'on dispose de petites bases d'apprentissage [Vapnik98]. Les treillis de Galois constituent un outil de reconnaissance particulier dont le principal atout, utilisé par S. Guillas dans [Guillas05], est de permettre une explicitation exploitable des connaissances utilisées (sous forme d'arbres) lors de la prise de décision.
Enfin terminons ce rapide tour d'horizon par l'approche la plus souvent utilisée, c'est-à-dire la méthode des K plus proches voisins qui allie simplicité de mise en œuvre et performances très honorables. Notons que certaines méthodes permettent le rejet alors que d'autres posent des difficultés lorsqu'il s'agit de déterminer un critère permettant de refuser une décision.

Dans le domaine de l'analyse de documents, les méthodes statistiques et neuronales ont plutôt été utilisées pour la reconnaissance de caractères ou de l'écriture manuscrite en ligne ou hors ligne. Les PDA utilisent par exemple des applications telles que Ms-Transcriber ou PenOffice [Penoffice06] pour permettre la saisie de texte manuscrit et de gestes élémentaires d'interaction réalisés à l'aide d'un stylo (figure 2.2). Il est à noter que, la plupart du temps, les formes reconnues ne peuvent correspondre qu'à un tracé unitaire (sans levée de stylo) et respectant une chronologie prédéfinie ; contrainte qui n'existe pas en hors ligne. De nombreux états de l'art concernent spécifiquement l'écriture manuscrite en ligne [Tappert90, Plamandon99], hors-ligne ou les deux simultanément [Plamandon00, Lorette99, Koerich05]. Les systèmes peuvent être classés en deux principales catégories : ceux qui s'intéressent à l'écriture dans sa globalité sans chercher à effectuer de segmentation (reconnaissance de mots) et ceux qui cherchent à segmenter les formes pour ensuite caractériser chacune des parties segmentées (caractères, graphèmes, traits). On parle alors de reconnaissance globale ou de reconnaissance locale. Les méthodes globales se basent classiquement sur des classificateurs statistiques tels que les réseaux de neurones, les SVM et les chaînes de Markov cachées [Marukata01]. Les méthodes locales utilisent plutôt des méthodes à base de prototypes (k plus proches voisins, distance élastique et DTW, …), des méthodes à base de réseaux de neurones et SVM et des méthodes hybrides combinant différents classificateurs. Avec ces méthodes, le nombre de classes à reconnaître est souvent restreint et il est difficile de faire apprendre de nouveaux modèles au système. Afin d'améliorer les performances de classification, les approches floues sont assez souvent utilisées [Anquetil97, Ragot03, Gagné06] ainsi que des méthodes proposant de fusionner ou de coupler différents types de classificateurs [Huang95, Lam97, Prevost03]

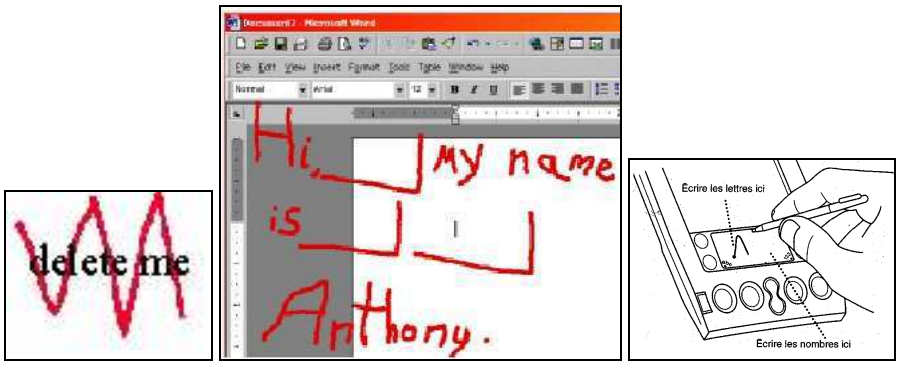
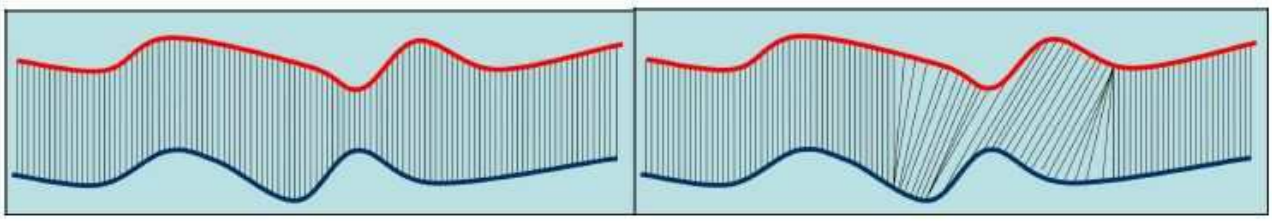


Figure 2.2 :: Gestes pour l'Interaction Homme-Document en ligne [PenOffice06]

2.2.2. Méthode de comparaison directe

La méthode Dynamic Time Warping (DTW) permet de comparer deux formes représentées par une suite ordonnée de points. Elle trouve, pour chaque élément d'une courbe, le meilleur élément correspondant dans l'autre courbe relativement à une certaine métrique [Plamondon88]. Ainsi, des points de chacune des deux courbes C et C' sont mis en correspondance avec une légère tolérance aux différences d'échantillonnage et de longueur (Figure 2.3).



(a) (b)

Figure. 2.3 : Illustration d'une mise en correspondance classique (a) et avec DTW (b) [Wirotius05].

La plupart du temps, comme sur la figure 2.3, la métrique utilisée est la distance spatiale euclidienne entre les points. Une fois la mise en correspondance effectuée, on peut calculer la distance entre les deux courbes en ajoutant la somme des distances entre couples de points correspondants. Les transformations autorisées dans la recherche de la correspondance sont l'allongement et le rétrécissement d'un signal relativement à l'autre. Le but de ces ajustements est de minimiser la différence entre les deux signaux. Il existe deux approches : asymétrique et symétrique. Dans le cas asymétrique, on cherche à établir une correspondance entre une courbe testée et un modèle alors que, dans le cas symétrique, on recherche en plus une correspondance entre le modèle et la courbe testée. Cette deuxième alternative qui tient compte de la réunion des deux appariements donne de meilleurs résultats [Hastie92]. Cette méthode de comparaison de courbes est très utilisée dans le domaine de l'authentification de signatures manuscrites en ligne et de la reconnaissance de la parole.

2.2.3. Méthodes structurelles

Si on veut prendre en compte une information plus contextuelle, une description des relations entre les objets ou les parties d'objets doit être utilisée en plus de leurs propriétés. Il paraît alors logique d'utiliser une approche structurelle. Les modèles descriptifs capables de représenter de tels objets sont les graphes et les structures syntaxiques (chaînes ou phrases inspirées de la théorie des langages). Les reconnaissances se font alors par calcul de distances entre chaînes, par appariement de graphes, par calcul de similarité entre graphes ou par analyse de la validité d'une phrase dans les grammaires correspondant aux modèles des formes (analyse syntaxique).

Concernant les documents imprimés, l'arbre est la structure de données la plus utilisée pour représenter les structures de documents. L'arbre peut représenter à la fois la structure physique et la structure logique d'un document. Suivant les auteurs, l'arbre ne code pas les mêmes informations et ne conserve pas toujours les informations spatiales et les relations logiques entre les objets. Dans certains arbres, les nœuds représentent les différents éléments physiques et logiques et les branches matérialisent les relations hiérarchiques entre ces éléments. Pour d'autres, chaque nœud code une structure alternative et les arcs matérialisent les hypothèses possibles, comme dans le cas du Gtree [Dengel89], reproduisant ainsi des arbres de décision. Dans la représentation classique d'une structure par des arborescences, l'opération de reconnaissance de structures consiste à mettre en relation l'arbre représentant la structure

physique construite avec l'arbre du modèle de structure logique ou à transformer par des règles le premier en second.

Les travaux relatifs à des approches syntaxiques proposent d'utiliser une description formelle de chaque classe de documents [Hadjar02]. Dans ce cas, des règles de composition représentant la structure logique générique sont représentées par une grammaire alors que les règles de présentation décrivent la structure physique. Les règles autorisent la représentation des relations spatiales mais peuvent aussi contenir des informations sur la typographie des caractères ou sur les alignements. L'analyseur syntaxique donne une liste de structures possibles avec un coût associé. L'un des systèmes les plus cités, ANON, développé au sein de l'équipe de Joseph [Joseph92], utilise une grammaire de type LR1 pour gérer les connaissances et reconnaître des symboles graphiques. Les règles ont pour but de contrôler l'activation des mécanismes de détection des primitives de bas niveau. Le système DMOS [Couasnon03] utilise également une grammaire pour décrire la structure type des documents à analyser à l'aide de règles de voisinage entre les objets. Cette technique est implémentée dans le logiciel FormuRead qui permet d'extraire automatiquement la structure de formulaires d'incorporation militaire du XIXe siècle. De la même façon, Cugini s'intéresse aux aspects mécaniques des dessins et essaie d'extraire des «form-features» par reconstruction à partir de primitives de base [Cugini90, Bottoni92].

En France, l'équipe de Karl Tombre a travaillé plusieurs années sur l'élaboration d'un système d'interprétation automatique de dessins techniques à partir d'un langage de description des formes caractéristiques. Une des premières versions, nommée Cellestin III, avait pour but de localiser et de reconstruire les blocs fonctionnels (vis, engrenages, ...) constituant le mécanisme [Tombre91, Vaxivière94b]. Le système a évolué jusqu'à donner naissance à Cellestin IV. Les connaissances utilisées concernaient non seulement les entités (cohérence interne de l'objet) et les liens entre entités (cohérence externe ou globale), mais aussi la sémantique du domaine puisque les dessins représentent des objets du monde réel, régis par des lois fondamentales venant de la mécanique, de la physique, ...

Dans le cadre du développement du système MDUS, l'équipe de Dori s'est longuement étendue sur l'analyse syntaxique des chaînes de cotations dans les projections orthogonales. Ces travaux ont abouti à la réalisation d'une méthode d'extraction, d'analyse et de vérification de cohérence des cotations (norme ANSI) modélisées à l'aide de graphes [Dori92, Weiss95].

Le graphe est également largement utilisé en informatique pour la représentation et la manipulation de données symboliques [Jolion01]. En analyse de documents, son exploitation est le plus souvent basée sur les méthodes d'appariement de graphes [Bunke99] ou de grammaire de graphes [Blostein96]. L'appariement de graphes se réfère au problème de trouver une correspondance entre les sommets de 2 graphes qui satisfait certaines contraintes ou critères d'optimalité. Les grammaires de graphes appliquent différentes règles afin de transformer les graphes extraits en graphes modèles. Il est possible d'utiliser des opérations de substitution, d'insertion, de suppression d'arcs et de sommets pour réaliser les transformations. On parle alors de distance d'édition [Milgram93]. Lorsque les graphes extraits

correspondent exactement aux graphes modèles, le problème de manipulation de graphes est considéré comme un problème de graphes exacts. Malheureusement, dans les applications d'analyse des documents, les graphes extraits sont souvent bruités [Conte06], le problème de reconnaissance devient un problème de graphes inexacts. Un des avantages de la représentation sous forme de graphes réside dans la possibilité de recherche de sous-graphes pour l'extraction de sous-structures dans un document [Llados01, Sanchez02]. Cet avantage est particulièrement adapté aux documents graphiques dans lesquels les parties graphiques sont en interconnexion. Un algorithme d'isomorphisme de graphe/sous-graphe, tolérant aux erreurs, formulé en termes de graphe d'adjacence de régions (RAG) a été proposé dans [Llados01] pour reconnaître des symboles dans des diagrammes dessinés à la main. La vectorisation de l'image est suivie d'une phase de construction d'une représentation des images sous forme de graphes à deux niveaux. Le graphe de premier niveau est généré à partir des points caractéristiques (points de jonction et extrémités) et des segments qui les relient. Le graphe de second niveau décrit les régions, c'est à dire les boucles minimales dans les graphes de premier niveau avec leurs relations d'adjacence. On obtient ainsi un RAG et une distance d'édition classique peut alors être utilisée pour calculer la similarité entre les chaînes de frontières des régions. Cependant, en cas de présence de bruit, des points caractéristiques supplémentaires peuvent apparaître, engendrant des sommets supplémentaires dans le graphe de premier niveau ce qui complexifie fortement l'analyse des RAG.

Le concept récent de graph-probing (sondage de graphes) présenté dans [Lopresti01] est une combinaison intéressante d'approches structurelles et statistiques. Différents types d'information, appelés sondes, sont extraits de la représentation structurelle d'une forme sous forme d'un graphe pour constituer un vecteur de caractéristiques statistiques. Il est alors possible d'utiliser une méthode de reconnaissance statistique classique. Les sondes utilisées peuvent concerner, par exemple, les occurrences d'un certain type de sommets, leur contenu et les spécifications de leur étiquette, les degrés d'entrée et de sortie des sommets, ... Bien que le sondage de graphes fournisse une mesure de similarité entre deux graphes ou entre un graphe et un sous-graphe, il ne donne pas d'information sur la mise en correspondance entre les deux graphes comparés.

2.3. Problèmes ouverts

Lorsque l'on se concentre sur les stratégies d'analyse et les méthodes de reconnaissance de contenu, les problèmes supposés appréhendables et restant ouverts concernent les points suivants :

- la localisation / reconnaissance des formes à l'intérieur de documents complets restent un challenge auquel peu de personnes s'attaquent hormis dans le cadre d'application très précise [Rendek06, Chhabra98, Cordella00]. Il semble que cette thématique va émerger dans un avenir proche sous la dénomination « spotting » (word spotting, symbol spotting, …). Actuellement, la plupart des classificateurs ne sont applicables que sur des symboles pré-segmentés. De plus, la plupart des représentations et des classificateurs utilisés actuellement imposent des contraintes fortes : par

exemple, les formes doivent être connexes (non ouvertes), elles ne doivent pas contenir de régions pleines, …

- L'utilisation des données extraites « durant l'analyse pour poursuivre l'analyse » est ardue. Pourtant la prise en compte du contexte permettrait de générer de nouvelles hypothèses et d'augmenter les possibilités de coopération entre les processus d'analyse. C'est ce que cherchent à faire les méthodes se basant sur les connaissances pour l'analyse mais celles-ci sont difficiles à mettre en œuvre et nécessitent la construction d'un modèle cohérent à partir de règles. La construction de règles est délicate et coûteuse en temps de développement car elles doivent être efficaces et compatibles entre elles. La grande majorité des travaux en reconnaissance de structures utilise des bases de connaissances fondées sur des règles ou sur une description formelle sous forme d'arbres, de graphes ou de grammaires. Cependant, il est difficile voire impossible d'adapter de tels systèmes. La grande variabilité des documents d'une même classe et la qualité aléatoire des images se traduisent par une complexification des modèles de structures. Les bases de connaissances deviennent alors très complexes et les règles peuvent se multiplier à l'infini. La difficulté se situe aussi au niveau de l'amorçage du processus lorsque le contexte sur lequel on se base est faible.

- Nous avons également vu que les systèmes coopératifs constituent des solutions théoriquement plus intéressantes que toutes les autres approches, car ils peuvent toujours s'adapter à n'importe quelles nouvelles situations. Cependant de tels systèmes, qui ont pourtant vu le jour il y a plus de 10 ans n'ont toujours pas donné lieu à de véritables évaluations à grande échelle ou à une comparaison avec des systèmes linéaires plus simples. Certaines équipes ont même abandonné cette approche au profit d'approches plus traditionnelles. En effet, un système coopératif est très difficile à implémenter et à faire fonctionner, les temps de développement de tels systèmes peuvent durer plusieurs années ce qui les rend inaccessibles. Ensuite, il est très difficile de maintenir un tel système et de corriger la moindre erreur d'analyse car celle-ci est le produit d'une suite d'interactions entre plusieurs agents ou spécialistes. Il faut donc retrouver l'erreur d'origine qui a conduit à une cascade de décisions provoquant l'erreur finale pour enfin corriger le programme spécialisé ou les règles qui dirigent la stratégie.

- Dans le domaine du CBIR [Smeulders00], on voit apparaître des systèmes se basant sur plus d'interactions homme-machine. La principale technique étudiée et utilisée jusqu'à présent correspond à la mise en place d'un bouclage de pertinence (relevence feedback) [Zhou03, Rendek06]. L'utilisateur peut juger et classer les réponses fournies par le système lors de ses requêtes de recherches successives. En fonction de ces retours, le système est alors capable d'adapter ou d'affiner ses réponses au fur et à mesure de l'interaction. Ce type de procédés est adapté à l'interrogation de grandes bases d'images mais ne permet pas de guider l'analyse d'une image. Dans ce dernier cas, il me semble que les remises en cause des réponses sont pratiquement irréalisables sans revoir la totalité de l'analyse quelle que soit la stratégie utilisée. Par conséquent, il semble n'exister aucun système convainquant qui permette la

correction d'erreurs avec reprise de l'interprétation. Il paraîtrait donc intéressant de faire intervenir les utilisateurs plus en amont [Bellet99] dans le processus d'analyse.

- L'usage de techniques d'apprentissage automatique, supervisé ou non supervisé, permettant d'inférer des connaissances à partir de l'observation d'un échantillon d'images constitue la solution indispensable au problème de conception de systèmes de reconnaissance. Malheureusement, Les techniques permettant d'apprendre les règles définissant la structure d'un type de documents ou même plus simplement d'identifier les caractéristiques discriminantes pour des EdC à reconnaître sont très peu étudiées. Si la littérature abonde sur les méthodes de reconnaissance des formes statistiques ou probabilistes, il n'en est pas de même lorsque l'on cherche à manipuler des données plus structurelles ou symboliques. Ce point me paraît d'ailleurs pouvoir améliorer significativement les performances de ce type de méthodes. Il explique en grande partie le fait, qu'actuellement, on ne propose que des classificateurs et des systèmes dédiés à un type bien précis de documents (manque de généricité).

2.4. Contributions

2.4.1. Analyses incrémentales et interactives

Il n'est pas raisonnable de limiter à une séquence d'étapes successives définies de manière rigide, le processus d'interprétation de documents puisque même l'homme n'est pas capable d'un tel exploit ! Il a besoin de voir un document à plusieurs niveaux pour l'appréhender totalement. Pour reproduire ce phénomène (qui traduit d'ailleurs la nécessité de la présence d'un contexte pour comprendre), les techniques d'extractions incrémentales semblent à privilégier car elles permettent d'enrichir le contexte et de limiter les erreurs quitte à augmenter le nombre d'étapes nécessaires à l'interprétation et/ou à demander l'aide de l'utilisateur. Les problèmes proviennent des nombreux niveaux de représentation nécessaires pour passer de l'image à une interprétation finale : phase lexicale, syntaxique, sémantique. Outre la mise en place d'une bonne représentation, les points cruciaux paraissent concerner :

- l'intégration de connaissances sur un domaine particulier dans un système générique d'analyse d'images de documents,
- l'enchaînement des différents algorithmes et les coopérations entre les différents niveaux d'analyse (lexicale, syntaxique voire sémantique),
- la possibilité de faire intervenir les utilisateurs finaux (non experts) le plus en amont possible dans le processus d'analyse automatique.

Analyse incrémentale : du plus simple au plus complexe

Pour l'analyse de la représentation structurelle lorsqu'il s'agit de documents graphiques, j'avais proposé dans le cadre de ma thèse d'utiliser des processus autonomes appelés spécialistes [Ramel96]. Pour faire coopérer les différents spécialistes, l'architecture utilisait un mécanisme de *"cycles perceptifs"*. Ceux-ci

étaient activés de manière séquentielle et progressivement ils contribuaient à simplifier la représentation du contenu des images. Les spécialistes utilisaient les primitives issues de la vectorisation pour communiquer. Ils se les appropriaient en modifiant les attributs, ils les associaient pour constituer des objets plus complexes (courbes, éléments mécaniques, ...). La représentation vectorielle pouvait être comparée à un système Blackboard contenant l'ensemble des informations utiles et facilement accessibles. Ce mécanisme est aussi le moyen que nous avions trouvé pour permettre aux spécialistes d'extraire seulement des données certaines et fiables appelées *les évidences.* Il permet de limiter les erreurs d'analyse ainsi que les étapes de résolution des incohérences difficiles à mettre en place. En effet, des données ambiguës à un instant donné, pour un spécialiste donné, peuvent constituer des évidences pour un autre spécialiste. De sorte que, lors d'une prochaine activation d'un spécialiste la représentation vectorielle aura été simplifiée et le nombre d'ambiguïtés réduit par rapport à l'état courant d'interprétation. Dans le système proposé, chaque spécialiste réalisait une tâche précise telle que, par exemple, la détection du texte, des courbes, des éléments mécaniques ou la reconstruction 3D du mécanisme dessiné (figure 2.4).

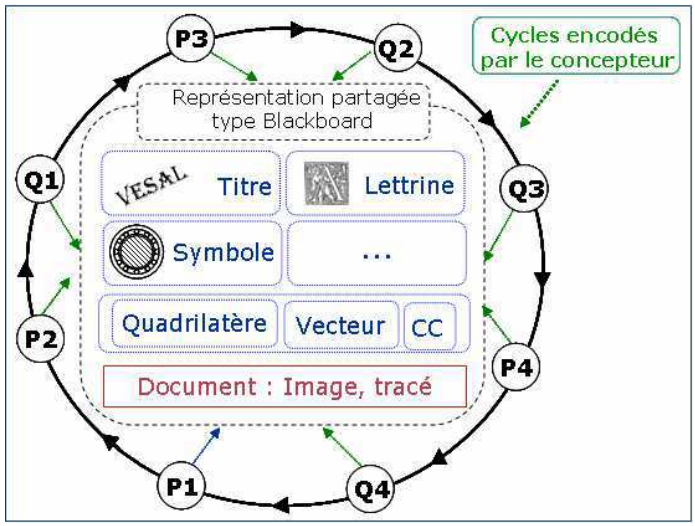


Fig. 2.4. Interprétation cyclique

L'interprétation est terminée lorsqu'un cycle complet n'apporte aucune information supplémentaire sur le contenu du document. Dans le cas contraire, un nouveau cycle d'activation des spécialistes est exécuté. Les paramètres utilisés par chaque traitement peuvent changer au cours du temps afin d'augmenter encore l'aspect incrémental du processus. Ces paramètres doivent être assez stricts au départ pour devenir moins restrictifs une fois le contexte suffisant pour éviter les erreurs.

Selon le type du document à analyser, l'expert (le concepteur) devait décider de la stratégie d'investigation la plus adéquate (définition de l'ordre d'activation des traitements et des paramètres associés). Par exemple, pour les schémas cinématiques, l'ordre d'activation qui avait été choisi était : Texte Formes pleines Lignes pointillées Courbes Annotations Symboles Reconstruction 3D.

Analyse interactive ou guidée par l'utilisateur

L'idéal est d'offrir à l'utilisateur (non expert) la possibilité de construire lui-même des chaînes de traitements permettant de faire évoluer progressivement le contenu de la représentation en fonction des caractéristiques des images à analyser et des objectifs visés. Le but est d'obtenir, à terme, un étiquetage le plus précis possible du contenu des documents de manière automatique par application d'un scénario construit de manière interactive par un utilisateur en fonction de ses besoins.

Afin de se rapprocher de cet objectif et sur la base des préconisations faites dans le chapitre précédent, concernant l'utilisation d'une représentation structurelle suffisamment générique, et avec l'aide de Stéphane Leriche [Ramel05a], une architecture logicielle permettant de réaliser une analyse interactive d'images de documents a été proposée. Un ensemble d'interfaces permet à l'utilisateur de construire à sa guise des scénarios d'analyse incrémentale provoquant l'évolution progressive du contenu de la représentation.

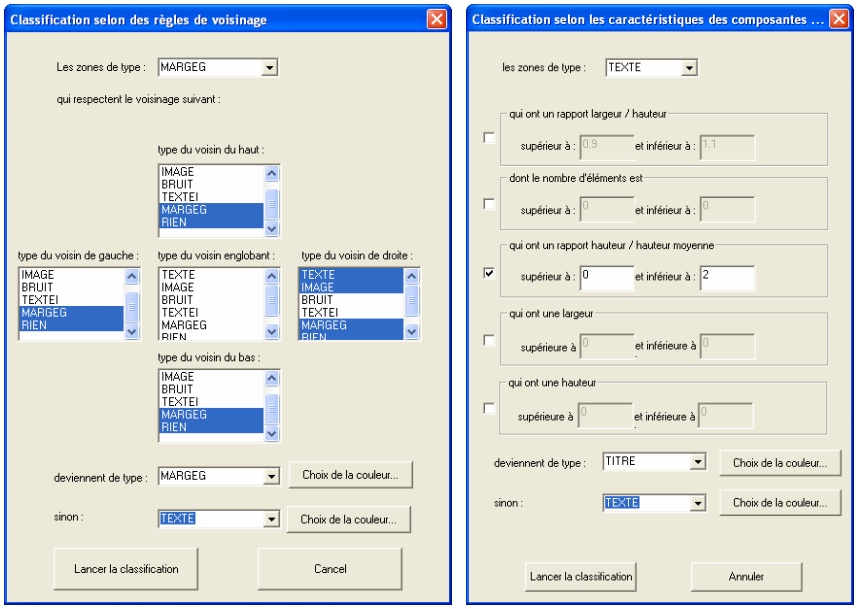
En fonction de ses besoins (extraction des parties graphiques, du texte, des lettrines, …) et en utilisant des interfaces conviviales, l'utilisateur (non expert en traitement d'images) peut construire un ensemble de règles d'étiquetage, de suppression et de transformations itératives des primitives et des EdC de manière à produire un modèle simplifié des documents à traiter. Ainsi, l'utilisateur peut extraire les informations et les métadonnées désirées sans se préoccuper des autres EdC présents dans les images. Les différentes séquences d'analyse (scénarios) ainsi obtenues peuvent être sauvegardées, modifiées et appliquées sur différentes bases d'images lors de traitements par lots. Nous avons proposé d'appeler cette nouvelle méthode d'analyse « Analyse guidée par l'utilisateur » [Ramel06a] par opposition aux méthodes guidées par les données et guidées par le modèle.

Nous avons d'abord expérimenté ce système sur le cas spécifique des images de documents anciens [Ramel05a]. Un prototype a été développé, nommé AGORA ; il permet à l'utilisateur de construire différentes séquences d'analyse. Ces scénarios provoquent une évolution graduelle de la représentation interne de l'image en cours d'interprétation en fonction du but poursuivi et des caractéristiques des images. Ce système permet une extraction précise des EdC composants les documents anciens par application d'un ou plusieurs scénarios conçus interactivement et simplement par les utilisateurs. Dans AGORA, les outils mis à disposition de l'utilisateur pour construire les scénarios sont :

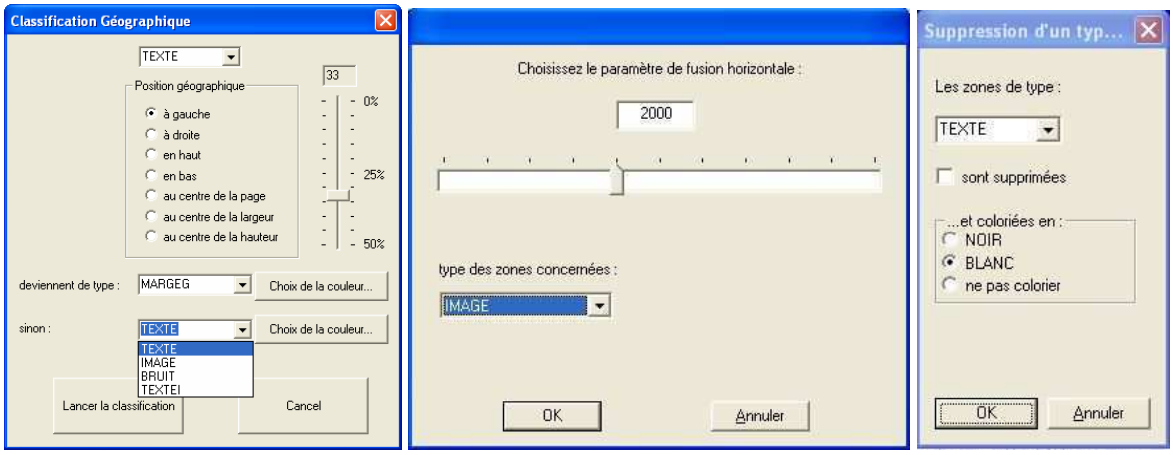
- Trois interfaces (au lieu d'un éditeur de règles non convivial ou, pire encore, d'un logiciel de développement) qui permettent à l'utilisateur de définir comment les labels doivent être attribués aux différents EdC au cours de l'analyse. Les interfaces permettent de définir les règles qui seront observées (exécutées) de manière séquentielle en fonction de la stratégie mise en place par l'utilisateur. Actuellement, les interfaces existantes permettant de faire évoluer les labels des EdC et des primitives contenus dans la représentation de l'image, concernent la position géographique des EdC, les relations de voisinage entre EdC et les propriétés internes (caractéristiques intrinsèques) des EdC.

En plus de ces interfaces, il est nécessaire de prévoir également des interfaces pour :

- La suppression d'un type spécifique d'EdC,
- La fusion/transformation/association d'EdC en fonction de leur label et de leurs relations de voisinage afin d'obtenir des EdC de niveau supérieur (par exemple, l'association de plusieurs caractères pour créer un mot, puis l'association de mots pour créer une ligne, …).



Règles de voisinage (a) Règles sur les caractéristiques intrinsèques (b)



Règles sur la position (c) Règles d'association (d) Règles de suppression (e)

Figure 2.5 : Interfaces de création des règles incluses dans les scénarios

Pour produire un scénario, l'utilisateur réalise les actions/règles successives devant être enregistrées sur une image typique grâce aux interfaces fournies (figure 2.5). Les règles sont appliquées en temps réel sur l'image et les résultats obtenus sont affichés directement de manière à ce que l'utilisateur puisse valider l'intérêt des règles qu'il produit. Les règles sont traduites sous forme littérale et affichées dans une liste visible en permanence au bas de la fenêtre d'AGORA (figure 2.6). L'utilisateur peut intervenir à tout

moment sur cette liste pour modifier l'ordre d'enchaînement ou supprimer des règles. Une fois le scénario considéré comme correct, il peut être sauvegardé pour application future sur une base conséquente d'images (un livre entier) lors d'un traitement par lot.

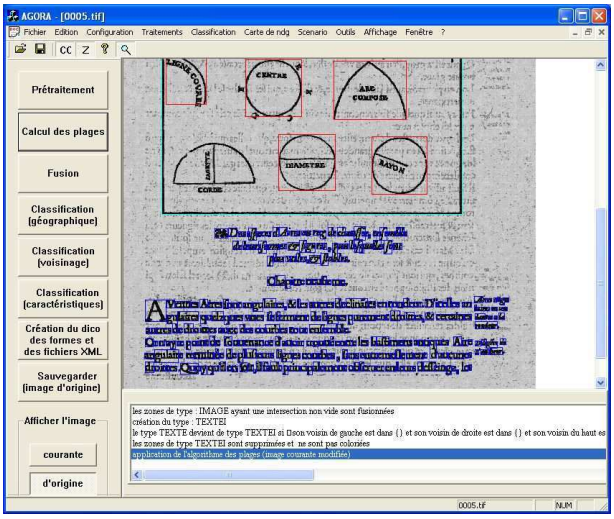


Figure 2.6: Le logiciel AGORA avec un scénario actif visible en bas de l'écran

L'analyse de la représentation de l'image organisée sous forme d'un graphe attribué (figure 2.7) est ainsi réalisée simplement par l'intermédiaire de cinq interfaces faciles à utiliser. Les algorithmes associés à ces différentes interfaces transforment la représentation initiale de l'image et il devient alors préférable de parler de représentation interne du contenu de l'image plutôt que d'esquisse initiale. Comme on peut le constater figure 2.7, les données extraites de l'image (EdC) sont clairement séparées des connaissances a priori (modèle du document = scénario) et des algorithmes de transformation/analyse (*Qx*). Les figures ci-dessous montrent clairement qu'une fois les algorithmes bas niveau (*Px*) exécutés pour obtenir l'esquisse initiale, plus aucune connaissance en analyse d'images n'est nécessaire. L'utilisateur doit "juste" construire un scénario cohérent à l'aide des interfaces fournies permettant d'extraire, de créer, de fusionner et de supprimer des EdC de manière incrémentale et interactive (figure 2.8).

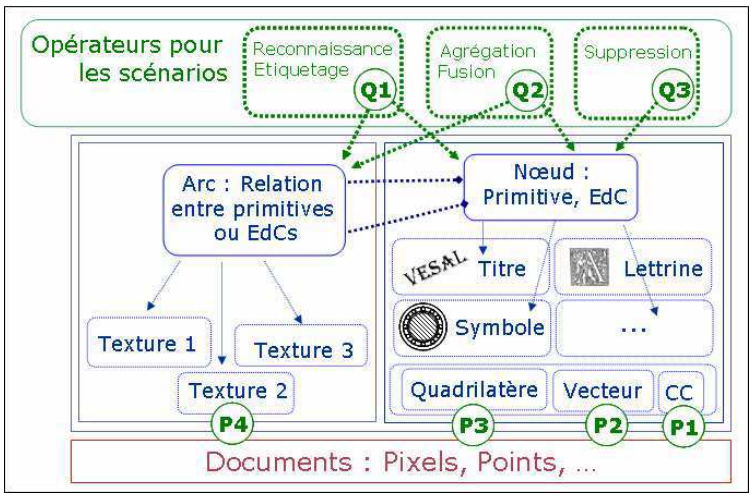


Figure 2.7 : Organisation de la représentation initiale

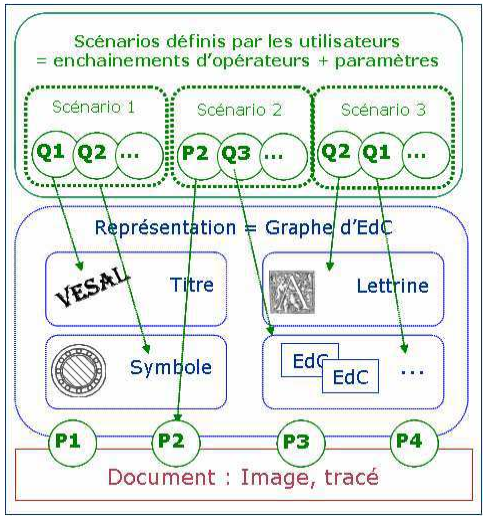


Figure 2.8. Analyse guidée par l'utilisateur

Durant un scénario, les algorithmes de fusion, de suppression, ou d'extraction peuvent être lancés autant de fois que nécessaire avec différents paramétrages. Progressivement, ils contribuent à simplifier la représentation et à faire progresser l'interprétation. Les EdC élémentaires peuvent être associés pour produire des EdC plus complexes. Par exemple, au début de l'analyse, la représentation de l'image ne contient que des composantes connexes, mais étape après étape des blocs de texte vont être localisés à l'aide de différentes règles jusqu'à obtenir des EdC de haut niveau (titres, légendes, paragraphes,…). La gestion du contexte est réalisée automatiquement et permet de tenir compte des EdC déjà extraits lors de la localisation de nouveaux EdC. Comme les algorithmes peuvent être activés plusieurs fois, l'utilisateur doit chercher à localiser en priorité les éléments facilement identifiables à un instant donné. Ensuite, les éléments plus difficiles à extraire sont alors plus facilement discernables. Par exemple, les lettrines sont plus facilement identifiables que les illustrations et peuvent donc constituer un pré-requis à la localisation d'autres entités.

En fonction du type de documents à analyser, l'utilisateur décide d'une stratégie d'investigation à suivre pour extraire de la manière la plus fiable possible les informations désirées. Pour cela, il choisit le type de règle à mettre en place et l'ordre d'activation de ces différentes règles. Par exemple, pour l'interprétation des partitions musicales, il peut être intéressant de suivre la séquence suivante : lignes horizontales lignes verticales lignes de portées texte têtes de notes queues de notes notes entières ….

2.4.2. Reconnaissance structurelle des éléments de contenu

L'utilisation d'une représentation structurelle m'a permis de porter une attention particulière envers les méthodes de reconnaissance structurelle. Ainsi, la thèse de Rashid Qureshi a pour but d'apporter des améliorations aux méthodes de reconnaissance de formes graphiques représentées à l'aide de graphes. Les résultats obtenus sont évalués dans le cadre du projet Technovision EPEIRES [EPEIRES05]. Nous proposons une nouvelle méthode pour déterminer le meilleur appariement possible entre deux représentations structurelles de symboles graphiques par mesure de similarité entre graphes [Qureshi06a, Qureshi06b]. Notre méthode décompose le processus en deux problèmes distincts : d'abord, la sélection d'appariements (*Mp*) entre sommets, et ensuite, la définition d'une mesure de similarité permettant de comparer deux graphes selon un appariement spécifique. Pour une meilleure compréhension, ces deux processus sont présentés, ci-dessous, dans l'ordre inverse de celui énoncé précédemment.

Mesure de similarité entre graphes

Pour comparer deux graphes avec des attributs numériques (afin de retrouver les avantages des méthodes statistiques) sur les sommets et sur les arcs, R.J. Qureshi propose une nouvelle mesure basée sur un calcul de distance entre les attributs associés aux sommets et un calcul de distance entre les

attributs associés aux arcs. La similarité entre 2 graphes attribués pour une mise en correspondance *Mp* donnée est définie à l'aide de la formule suivante :

$$Sc_{Mp} = \left[\sum_{i=1}^{m} (1 - \Delta V_i) + \sum_{j=1}^{n} (1 - \Delta E_j) - \left(\sum_{i=1}^{k} \omega_i + \sum_{j=1}^{\ell} \omega'_j \right) \right] \quad (1)$$

m est le nombre total de sommets appariés dans *Mp* et *n* le nombre total d'arcs entre ces nœuds. w_i et w'_j sont les poids associés aux pénalités qui peuvent être appliquées lorsqu'un nœud *i* ou respectivement un arc *j* est associé à plusieurs autres (*k* et *l* respectivement) dans l'appariement courant. Grâce à ce mécanisme de pénalités, nous autorisons l'association d'un sommet (ou un arc) d'un graphe à un ou plusieurs sommets (ou arcs) dans l'autre graphe. Les pénalités dépendent du nombre d'attributs utilisés lors des calculs de distances entre sommets (et entre arcs).

Dans la formule (1), ΔV_i correspond à la distance entre deux sommets appariés, normalisée entre 0 et 1.

$$\Delta V_i = \frac{\sum_{k=1}^{\delta} f_k\left(A_V^k , A_{V'}^k\right)}{\delta} \quad \text{et} \quad \Delta E_j = \frac{\sum_{k=1}^{\Omega} g_k\left(A_E^k , A_{E'}^k\right)}{\Omega}$$

La fonction f_k compare la valeur du $k^{\text{ième}}$ attribut des deux sommets appariés et retourne une valeur de similarité entre 0 et 1.

De même, ΔE_j correspond à la distance entre deux arcs normalisée entre 0 et 1. La fonction g_k est utilisée pour comparer les deux valeurs du $k^{\text{ième}}$ attribut des deux arcs. δ et Ω sont respectivement le nombre d'attributs associés à chaque nœud et à chaque arc du graphe.

Enfin, pour normaliser la mesure de similarité entre 0 et 1, on utilise :

$$\text{Sim}(V, V') = \frac{Sc_{Mp}}{\left[\delta\left(C(V) + C(V')\right) + \Omega\left(C(E) + C(E')\right)\right]}$$

C(x) est une fonction de cardinalité qui retourne le nombre de sommets ou d'arcs dans un graphe.

L'un des principaux intérêts de cette nouvelle formulation du calcul de similarité entre graphes réside dans sa généricité et sa simplicité de modification. Il est très facile de prendre en compte de nouvelles caractéristiques (attributs), tant statistiques que symboliques, concernant aussi bien les nœuds (EdC et primitives) que les arcs (relations). Pour cela, il suffît de définir les fonctions de similarité f_k et/ou g_k à associer à chacun des attributs pris en compte lors de la comparaison.

Recherche du meilleur appariement

La manière la plus simple de déterminer le meilleur appariement est de tester toutes les mises en correspondance possibles pour les deux graphes et de garder celle qui produit le meilleur score de similarité. Comme l'espace de recherche s'accroît de manière exponentielle quand la taille du graphe augmente, la demande en ressources de calcul devient rapidement excessive. C'est pourquoi, des algorithmes d'une complexité moindre, comme les procédures de séparation et d'évaluation ou les algorithmes gloutons, sont souvent utilisés pour éviter les recherches exhaustives malgré le risque de

trouver une solution sous-optimale. Dans notre système, pour trouver le meilleur appariement entre les deux graphes, un algorithme qui s'inspire de l'algorithme glouton présenté dans [Champin03] est utilisé. Il prend deux graphes en entrée et retourne la meilleure mise en correspondance (M_{best}) des sommets entre eux. La recherche débute avec une carte des mises en correspondance vide et, à chaque itération, un couple de sommets candidat (*v,v'*) est autorisé à entrer dans la carte des mises en correspondance courante (M_c) à condition qu'il génère le meilleur score de similarité. S'il y a plus d'un couple candidat qui fasse augmenter la fonction de score de manière égale, un score de second niveau est calculé sur la base de l'analyse des arcs associés aux nœuds candidats. Cette condition aide au choix des sommets à faire entrer dans l'appariement à l'aide d'une fonction *look_ahead* qui teste l'efficacité d'un couple candidat avant de l'autoriser à entrer dans M_c. Ce processus d'insertion itératif de couple de nœuds associés se poursuit jusqu'à ce que le score ait cessé d'augmenter et qu'il n'y ait plus, ni de nœuds, ni d'arcs à ajouter dans la carte des mises en correspondance courante.

Les expérimentations réalisées avec cette méthode sur les différentes bases d'images tests publiques se sont révélées très concluantes, donnant des résultats au moins aussi bons que les méthodes concurrentes [Qureshi06b]. Ces performances nous confortent donc dans l'idée de poursuivre nos recherches dans cette voie.

2.4.3. Fusion d'informations pour la reconnaissance

Mesure de dissimilarité entre signatures

Afin d'améliorer les performances des systèmes d'authentification de signatures en ligne, avec Matthieu Wirotius, nous avons proposé de combiner différents types d'information lors de l'utilisation de l'algorithme DTW. Cet algorithme utilise habituellement une distance euclidienne entre les points mis en correspondance dans les deux signatures pour évaluer la dissimilarité entre les 2 courbes.
Soit *S1* et *S2* deux signatures, *nbCorresp* le nombre de points en correspondance dans *S1*. On note Pt_i un point quelconque de *S1* et P'_i l'ensemble des points correspondants dans *S2*. La distance entre les deux signatures *S1* et *S2* est donnée par la formule suivante :

$$dist(S1,S2) = \sum_{i=1}^{nbCorresp} \sum_{P' \in P'_i} DistPt(Pt_i, P')$$

Matthieu Wirotius a apporté plusieurs améliorations pour cette phase de comparaison [Wirotius04b, Wirotius05b, Wirotius05c]. Dans un premier temps, il a choisi de normaliser la dissimilarité fournie par DTW en la divisant par le nombre de correspondances trouvées afin d'être indépendant du nombre de points de la signature. Le résultat peut alors être vu comme la moyenne des distances entre points mis en correspondance dans les deux signatures. Lorsqu'un point est mis en correspondance avec plusieurs autres points de la deuxième signature, toutes les distances entre ce point et ces correspondants sont ajoutées. Nous pensons qu'il est préférable de considérer uniquement la distance entre le premier couple

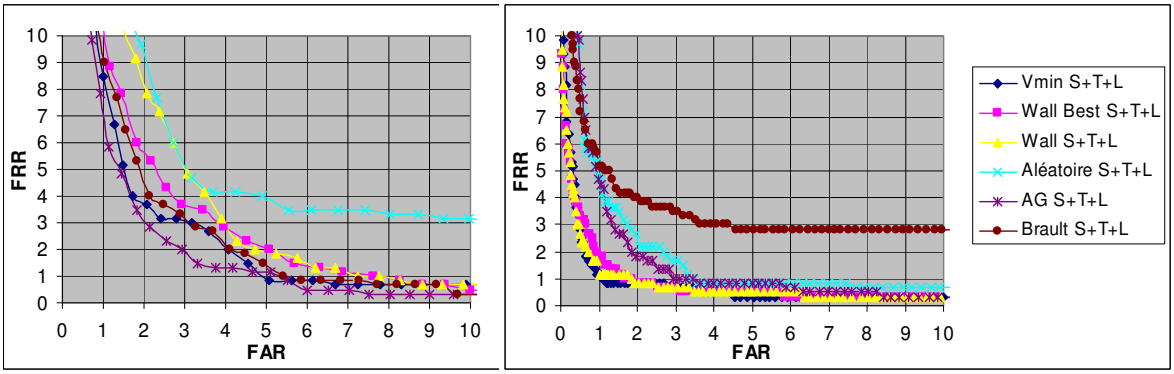
de points afin d'éviter un cumul d'erreurs. Afin de prendre en considération la dynamique de la signature, au lieu de calculer la dissimilarité uniquement à partir d'une distance euclidienne entre les coordonnées des points, Matthieu Wirotius propose d'utiliser l'instant d'acquisition (*DistT*) et l'abscisse curviligne (*DistC*) en chaque point en plus des coordonnées *x* et *y* (*DistS*) lors du calcul de la dissimilarité par DTW. La manière de calculer la dissimilarité a été adaptée afin de prendre en compte simultanément ces trois types d'information. La formule de calcul de la distance devient alors:

$$Dist(S_1, S_2) = \alpha \times DistS(S_1, S_2) + \beta \times DistT(S_1, S_2) + \gamma \times DistC(S_1, S_2)$$

α, β et γ sont des coefficients réels choisis de manière à ce que les différentes distances aient des influences comparables.

$$\begin{cases} \alpha + \beta + \gamma = 1 \\ \alpha, \beta, \gamma \in [0,1] \end{cases}$$

Comme on peut le voir sur la figure 2.9, la combinaison des 3 distances fournit les meilleures performances avec un EER (Equal Error Rate) de 1,1% et ce quelle que soit la méthode de selection des points représentatifs des signatures (*Aléatoire, Vmin, Wall, AG, Brault* – voir section 1.4.2).



(a) (b)

Figure 2.9 : Combinaison des distances temporelle et curviligne (a) - Combinaison des distances spatiale, temporelle et curviligne (b) – FAR = False Acceptance Rate et FRR = False Rejection Rate

	Distance Spatiale	Distance Temporelle	Distance Curviligne	Distance S+T	Distance L+T	Distance L+S	Distance S+T+L	Distance S+min(T,L)
EER	1,6%	3,2%	9,5%	1,3%	3,2%	1,3%	**1,1%**	**1,3%**

Table 2.1. Résultats des différentes combinaisons possibles

Ce résultat confirme l'hypothèse intuitive de complémentarité de ces 3 types d'informations. Si l'on compare les résultats avec ceux obtenus en utilisant uniquement la distance spatiale, on remarque un gain de performances de 30%. Le tableau 2.1 résume les résultats obtenus avec différentes combinaisons de ces 3 distances sur la base de test SVC [SVC04].

D'autres expérimentations ont été réalisées afin de voir s'il était judicieux de donner plus de poids à une information qu'à une autre. Des tests exhaustifs ont montré que la meilleure combinaison possible correspondait aux poids suivants :

$$d = 0{,}2 \times distS + 0{,}6 \times distT + 0{,}2 \times distL$$

Le EER passe alors de 1,1% à 0,9%, correspondant à un nouveau gain de 20%. On remarque également qu'aucune des distances, utilisée seule, ne produit de meilleurs résultats qu'une combinaison de distances et que, par conséquent, toutes les distances sont nécessaires. Il semble néanmoins que la distance temporelle soit la plus discriminante des 3 distances.

En ce qui concerne les performances obtenues avec les différentes représentations présentées en section 1.4.2 (approximation polygonale avec Wall, approximation avec Brault, vitesse minimum, algorithme génétique), le classement est conservé quelle que soit la manière de combiner les distances.

Comparaison des dynamiques de frappe au clavier

Ces travaux sur l'authentification de signatures m'ont permis d'acquérir des compétences dans le domaine de la biométrie en plein essor actuellement. Ils se sont concrétisés par l'obtention d'une seconde bourse CIFRE (pour Sylvain Hocquet) dont le sujet se focalise sur les problématiques de la fusion d'information pour l'authentification (classification à une classe). Dans ce cadre, nous nous intéressons à l'utilisation d'un dispositif disponible sur tous les ordinateurs : le clavier pour réaliser l'authentification d'une personne. Il s'agit d'étudier plus particulièrement la façon de saisir du texte d'un utilisateur appelée dynamique de frappe.

Pour extraire des données issues de la frappe d'un utilisateur, nous analysons les temps séparant les événements claviers. Un événement clavier peut être la pression ou le relâchement d'une touche. Pour un même couple de touches successives, nous pouvons extraire plusieurs temps différents comme illustré figure 2.10.

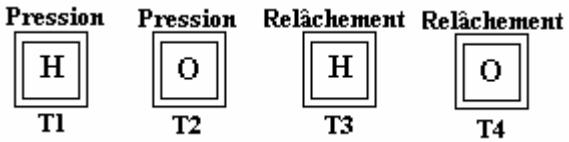


Figure 2.10 : Temps extraits au cours de la frappe du mot « HO »

Quatre types de temps sont disponibles :

- P-P (Press-Press) : temps entre les deux pressions des touches (T2-T1)
- P-R (Press-Release) : temps entre l'appui sur la touche et le moment où elle est relâchée (T3-T1)
- R-P (Release-Press) : temps entre la relâche d'une touche et l'appui sur la suivante (T3-T2)
- R-R (Release-Release) : temps entre la relâche de deux touches successives (T4-T3)

Sylvain Hocquet propose un système qui utilise simultanément trois méthodes offrant des performances équivalentes : une méthode basée sur l'analyse de la moyenne et de la variance des temps inter-touches, une méthode basée sur une comparaison des classements (rangs) des temps inter-touches et une méthode utilisant une discrétisation du temps en classes (temps courts, temps moyens, temps longs). Sur la base de test utilisée, les résultats des méthodes prises séparément donnent des performances de l'ordre de 12% pour le FRR, de 8% pour la FAR et de 10% pour le EER. Les travaux de Sylvain Hocquet démontrent qu'une fusion de ces trois méthodes à l'aide de différents opérateurs provoque une hausse conséquente des performances par rapport à chaque classificateur pris séparément [Hocquet05]. Nous retrouvons ici un résultat classique de la littérature. Ce résultat prouve en plus que même si les classificateurs travaillent à partir des mêmes données acquises, les caractéristiques construites par la suite sont suffisamment différentes pour justifier une fusion. Au niveau performance, l'opérateur « somme » est le plus performant avec un FAR de 4,5% et un FRR de 5,8%. Le EER est de 5% avec une nette diminution des taux d'erreur maximale par utilisateur. Ces taux d'erreurs ouvrent la voie à une utilisation de la dynamique de frappe à grande échelle. Dans sa thèse, Sylvain Hocquet montre également que l'étape de fusion permet d'apporter assez facilement plusieurs améliorations à un système d'authentification quel que soit son type. Nous avons observé que, pour certains utilisateurs, la fusion donne de moins bons résultats que la meilleure méthode prise séparément. Ainsi, la mise en place d'un système de poids associés à chaque méthode et adaptés à chaque utilisateur provoque encore un gain significatif pour les performances jusqu'à une valeur minimale du ERR égale à 3,6% [Hocquet06a, Hocquet06b]. La sélection automatique (apprentissage) des poids pour chaque utilisateur n'est pas triviale et constitue un travail encore en cours. Nous espérons, cependant, aboutir à d'excellents résultats rapidement [Hocquet06c].

Chapitre 3 :

Documents numériques et interactivité

Bien qu'ils se situent toujours dans le domaine des documents électroniques, les objectifs des travaux de recherche décrits dans ce chapitre diffèrent sensiblement de ceux présentés précédemment. Il ne s'agit plus de donner à l'ordinateur des capacités d'analyse et de reconnaissance automatique des contenus des documents, mais de proposer de nouvelles structurations et de nouveaux modes de descriptions des contenus afin d'améliorer les interactions entre les hommes et les documents informatiques. Dans les systèmes traditionnels de production et de consultation de documents orientés « texte », le mode de communication entre l'utilisateur et l'application est caractérisé par la présentation à l'écran, ou sur papier, d'une information principalement spatiale et statique. Actuellement, sur le Web, l'utilisateur effectue une lecture active via son parcours de navigation à travers l'information affichée. Les nouveaux systèmes doivent donc permettre certaines formes d'interaction avec les documents présentés. La nature dynamique des objets manipulés, tels que la vidéo et l'audio, ainsi que la définition de leur ordonnancement rendent plus complexe la réalisation d'outils de production de tels documents multimédias. Le principe statique du « wysiwyg », dans lequel l'information présentée à tout instant du processus d'édition correspond à l'information finale, ne peut s'appliquer à l'édition du scénario temporel des documents multimédias. Il faut alors distinguer deux étapes dans le processus de conception des documents multimédias : l'étape de spécification (ou d'édition) et l'étape de présentation (ou d'exécution).

De par ces évolutions, les recherches actuelles sur les plates-formes d'apprentissage à distance visent à définir des modèles pour la structuration, le déploiement, la gestion et l'utilisation de documents pédagogiques hypermédias [Chabert00, Khaled98, Nanard98]. Grâce aux hypermédias, l'apprenant peut s'affranchir de la linéarité des documents et accéder à une information dynamique, mais malheureusement, selon nous, pas encore suffisamment adaptée à ses besoins.
Ainsi, l'objectif recherché, durant mes travaux dans ce domaine, a été de concevoir un modèle d'environnement qui permette la génération dynamique de documents et d'interfaces adaptés à l'utilisateur lors d'activités pédagogiques. L'idée de base est d'utiliser les technologies XML pour

composer des documents à partir d'un rassemblement d'EdC. Ainsi, un document ou une interface pédagogique est un assemblage de briques élémentaires correspondant aux données multimédias (EdC) qui seront présentées de différentes façons, afin de fournir une information adaptée aux connaissances et au profil de l'apprenant.

Pour cela, mes recherches, réalisées majoritairement dans le cadre de la thèse de Sofiane Benadi [Bénadi01, Bénadi02a, Bénadi02b, Bénadi03], ont porté sur deux directions qui concernent la génération, la gestion et le traitement de documents hypermédias :

- l'analyse des besoins des producteurs de documents hypermédias (plus spécifiquement les enseignants) qui a amené à proposer une modélisation des données et des services adaptés aux hyperdocuments et à étudier les mécanismes de base pour la conception d'environnements d'édition et de stockage de documents numériques.
- l'étude des fonctions de présentations adaptatives de documents multimédias et de leur mise en œuvre dans les environnements de consultation (plus spécifiquement conçus pour les apprenants).

3.1. Documents numériques adaptatifs

3.1.1. Définitions

Le rôle de l'adaptativité est de permettre à l'utilisateur de minimiser l'effort consacré à l'exploration des capacités du système pour optimiser l'effort nécessaire à la résolution d'un problème. Edmonds [Edmonds81] distingue trois types de systèmes intégrant des dispositifs d'adaptation :

- un système ***adapté*** dans lequel l'adaptation est l'œuvre du concepteur lui-même. Le système prend en compte, une fois pour toutes, un profil d'utilisateur ou de groupes d'utilisateurs défini préalablement à sa mise en place. Les techniques d'adaptation sont appliquées durant la phase de conception du système (par "conception participative" par exemple) et l'adaptation ne peut, donc, pas être propre à chaque individu,
- un système ***adaptable*** est un système qui peut être modifié, à tout moment, sur demande explicite de l'utilisateur qui le consulte. Celui-ci peut choisir l'aspect physique des éléments de l'interface à l'aide, par exemple, de tableaux de bord. Il peut également demander la création de macro-commandes (comme sous Windows). Dans ce type de système, ce sont les utilisateurs qui saisissent leurs préférences et qui les enregistrent dans un modèle. Celui-ci n'est modifié, par la suite, que sur nouvelle demande explicite de l'utilisateur,
- un système ***adaptatif*** est un système qui s'adapte de lui-même, à chaque instant, aux besoins et aux habitudes de l'utilisateur. Des mécanismes de suivi (tracking) des comportements de l'utilisateur permettent de connaître les besoins de ce dernier en fonction de son environnement, de son état psychologique et de ses connaissances. La mise à jour du modèle utilisateur est réalisée par le système lui-même, par observation des actions et des réactions de l'utilisateur (interaction avec le système).

Les deux derniers types de systèmes permettent de prendre en compte les besoins et les habitudes de l'utilisateur. Mais, la différence entre l'adaptabilité et l'adaptativité réside dans le fait qu'un système adaptatif ne demande pas à l'utilisateur de se préoccuper des adaptations et de la mise à jour de ses préférences, qui doivent donc être déduites des actions qu'il entreprend. Dans de tels systèmes, l'utilisateur n'a pas à apprendre comment adapter son interface ou à mémoriser d'éventuels langages de spécifications pour opérer les modifications nécessaires à l'adaptation. Il peut concentrer son énergie sur le fonctionnement du logiciel et l'accomplissement de sa tâche en espérant que le système comprenne correctement ses objectifs et ses désirs et mémorise ses habitudes. Cette adaptativité, évoluant d'une façon autonome par rapport à l'utilisateur, présente des avantages manifestes mais aussi des inconvénients comme le risque de désorientation. Cela s'explique par le non respect de l'un des principes de base de l'ergonomie logicielle précisant que dans un système bien conçu les mêmes actions doivent toujours provoquer les mêmes réactions.

La plasticité des interfaces est une thématique de recherche qui a émergé récemment dans la communauté IHM française. Elle correspond à la capacité des interfaces à s'adapter à leur contexte d'usage dans le respect de leur utilisabilité [Calvary03]. Le contexte d'*usage* correspond au triplet : « *utilisateur, plate-forme et environnement* ». Dans ce triplet, l'*utilisateur* est un usager représentatif du public ciblé et est généralement décrit par ses capacités (physiques et cognitives), ses compétences métier et informatiques. La *plate-forme* désigne la structure matérielle et logicielle supportant l'interaction (un assistant personnel, un téléphone portable …). Sur ce type de matériel, la taille de l'écran, les dispositifs d'interaction disponibles, les capacités de calcul et de communication sont des éléments déterminants qui influent sur les modes d'interaction possibles. L'*environnement* se réfère à l'environnement physique dans lequel se situe l'interaction. Il est décrit par un ensemble d'informations externes à la tâche en cours mais susceptibles de l'influencer (la luminosité, le bruit, la localisation géographique, ...).

La plasticité apparaît ainsi comme une forme d'adaptation puisqu'elle mobilise un processus de type « Evénement Réaction » dans lequel l'évènement correspond à un changement survenu dans le contexte d'usage (utilisateur et/ou plate-forme et/ou environnement) et la réaction dénote les mesures mises en œuvre, en réponse, par le système et/ou l'utilisateur pour préserver l'utilisabilité du système. Les différences de ressources matérielles disponibles (taille de l'écran, présence ou non de clavier), lors d'une interaction personne-système, sont traitées au cas par cas lorsqu'elles sont peu nombreuses. Par contre, cela est impossible lorsque la multiplicité des contextes d'usage devient trop importante, l'utilisation d'interfaces plastiques devient alors incontournable.

3.1.2. Architectures des systèmes adaptatifs

L'adaptativité nécessite de disposer de mécanismes permettant de choisir les modifications nécessaires sur le document à partir d'observations effectuées sur le triplet : utilisateur, plate-forme, environnement. Pour cela, il est nécessaire que le système contienne une *modélisation interne* de chacune des parties de

ce triplet et dispose d'un mécanisme de suivi de leur évolution au cours du temps. Benyon [Benyon93] a proposé l'une des premières architectures informatiques permettant l'adaptation. Cette architecture comporte essentiellement trois modèles : *un modèle utilisateur*, *un modèle du domaine* et *un modèle d'interaction*. La majorité des systèmes actuels se sont fortement inspirés de cette architecture tout en essayant de lui ajouter ou d'améliorer certains de ses composants.

- Le modèle de l'utilisateur cherche à fournir une description la plus précise possible des différentes facettes (profil cognitif, compétences, aptitudes, préférences, …de l'utilisateur du système.
- Le modèle du domaine caractérise le contexte, il dépend du but et des objectifs visés par l'application. Il peut également dépendre des fonctionnalités que le logiciel offre pour réaliser les tâches courantes.
- Le modèle d'interaction est le cœur du système, il s'appuie sur les autres modèles pour produire des décisions de réaction telles que : comment présenter les informations et les interfaces ? Quel est le moment approprié pour fournir des indications ? Jusqu'où laisser l'utilisateur poursuivre sur une fausse piste ?

Le choix des adaptations à effectuer s'appuie souvent sur un mécanisme d'inférences utilisant la trace des actions de l'utilisateur. Cette trace peut être enregistrée dans un ou plusieurs historiques propres à chaque utilisateur et en se servant de capteurs plus ou moins évolués allant du clic souris à l'enregistrement vidéo. Les données captées permettent au système de construire un modèle utilisateur qui sert de base à la détermination des adaptations utiles.

On peut également noter des différences entre le triplet : *utilisateur, plate-forme et environnement,* proposé par les chercheurs travaillant sur la plasticité des IHM et l'architecture proposée par Benyon. Dans sa thèse, Sofiane Bénadi synthétise les deux propositions pour modéliser l'architecture d'un système adaptatif (figure 3.1) en tenant compte des travaux réalisés sur l'adaptativité mais aussi sur les recherches en plasticité plus récentes.

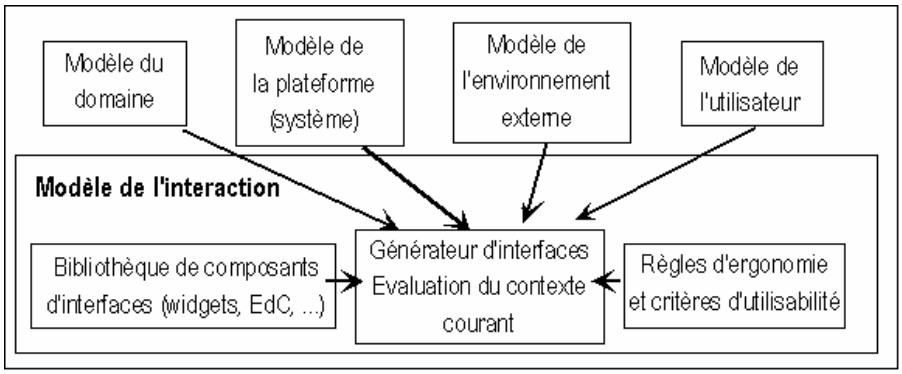


Figure 3.1 : Architecture d'un système adaptatif

3.1.3. Hypermédias adaptatifs

Depuis quelques années, les hypermédias ont ouvert un nouveau champ de recherche dans le domaine des systèmes adaptatifs. Cet engouement s'explique par le principal atout de ces systèmes : la non linéarité qui, malheureusement, est vite devenue un inconvénient majeur. Des études ont montré que l'usager pouvait se perdre rapidement dans l'hyper-espace. De ce fait, on a cherché à guider l'utilisateur dans son cheminement en fonction de ses objectifs, de ses connaissances et de ses préférences, en modifiant aussi bien, le contenu des pages que les liens entre les documents. De là est née l'idée de construire des systèmes nommés les « hypermédias adaptatifs » dont l'objectif principal est d'accroître les fonctionnalités du système en le personnalisant et de proposer de nouvelles méthodes d'accès à l'information. Etant donné la jeunesse de ce domaine de recherche, le concept de système hypermédia adaptatif n'a pas été encore complètement défini. Néanmoins, nous utiliserons la définition donnée par Peter Brusilovsky [Brusilovsky96] que nous avons traduite en français :

"Par systèmes hypermédias adaptatifs, nous désignons tout système hypermédia et hypertexte qui reflète certaines caractéristiques de l'utilisateur dans le modèle utilisateur et applique ce modèle pour adapter des aspects visibles et variés du système à l'utilisateur".

Selon cette définition, un hypermédia adaptatif doit satisfaire trois critères. Il doit être un système d'information dont l'interface est un hypermédia. Il doit contenir un modèle utilisateur et il doit être capable d'utiliser ce modèle pour adapter l'hypermédia. Toujours selon Brusilovsky [Brusilovsky98], l'hypermédia adaptatif est utile quand le système est appelé à être utilisé par des personnes ayant des connaissances différentes ou des buts différents et quand l'hyper-espace est étendu. Dans les systèmes actuels, l'objectif est toujours de filtrer l'information afin de limiter la dimension de l'hyper-espace utile. Selon nous, il faut ajouter à cette définition les aspects « environnement » et « plate-forme » qui peuvent également être vus comme des déclencheurs d'adaptation.

Il est possible d'adapter le contenu des pages, les liens présents dans celles-ci et la navigation. Deux classes différentes d'adaptation ont, donc, été définies et nous parlerons de support de navigation adaptatif et de présentation adaptative.

Adaptation de la navigation

L'objectif de l'adaptation de la navigation est d'aider l'utilisateur à trouver son chemin dans l'hyper-espace en agissant sur la façon dont les liens lui sont présentés.

Différentes techniques ont été développées au fil des années, le guidage direct, l'ordonnancement des liens (le tri), le masquage de liens, l'annotation des liens ou encore l'adaptation à la carte [Brusilovsky03]. Le guidage direct est la technique de base régulièrement utilisée puisqu'elle est simple à mettre en œuvre. Elle se base, la plupart du temps, sur l'ajout d'un lien "suivant" qui permet d'accéder à la page la

plus en adéquation avec les objectifs de l'utilisateur dans le contexte courant. Les problèmes sont : le risque de diminuer considérablement la capacité d'exploration et d'augmenter la passivité de l'usager puisque le système tend à devenir complètement linéaire.

L'ordonnancement des liens vise à afficher les liens suivant un ordre définissant leur intérêt ou leur qualité (comme dans "google"). L'inconvénient majeur de cette technique est que l'on ne peut pas l'utiliser avec des liens contextuels, c'est à dire ceux se situant à l'intérieur d'une phrase.

Le masquage de liens consiste à supprimer ceux qui sont en inadéquation avec les objectifs de l'utilisateur. Comme la méthode précédente, cela peut entraîner une certaine désorientation chez l'utilisateur.

L'annotation des liens donne plus de responsabilité à l'usager puisqu'on lui laisse le choix d'activer les liens mais en y adjoignant une explication sur le contenu de la page cible pour chacun. Ces annotations peuvent apparaître sous forme de bulles visibles dans le navigateur lors du survol des liens par la souris. Les annotations sont plus efficaces lorsqu'elles dépendent du contexte et du modèle de l'utilisateur. Pour aller encore plus loin, les cartes adaptatives permettent de fournir à l'utilisateur l'organisation de l'hyper-espace sous forme textuelle (arbre hiérarchique) ou sous forme graphique plus ou moins simplifiée en fonction de son profil.

Adaptation du contenu

Le but de la plupart des méthodes classiques d'adaptation de contenu (appelées aussi méthodes à explication additionnelle) est de cacher, à l'utilisateur, quelques parties d'informations à propos d'un concept particulier qui n'est pas pertinent pour l'utilisateur à un instant donné. Par exemple, des détails très pointus peuvent être masqués à un utilisateur ayant un niveau de connaissance limité dans le domaine. Inversement, des explications additionnelles peuvent être présentées à des novices. En terme général, en plus de la présentation de base, certaines catégories d'utilisateurs peuvent recevoir des informations supplémentaires conçues spécialement pour eux.

Afin d'améliorer la qualité de l'adaptation et de prendre en compte instantanément des nouvelles données, les recherches se sont orientées vers les hypermédias adaptatifs dynamiques. La principale caractéristique de ces systèmes est d'offrir un hyper-espace virtuel. Le système n'est plus constitué de pages et de liens prédéfinis. Il est construit dynamiquement. L'architecture de ces systèmes contient alors deux modules supplémentaires qui sont : une base d'objets multimédias ou d'objets d'interaction et le générateur de pages ou d'interfaces.

Le modèle du domaine, comme pour la dernière génération des hypermédias adaptatifs, est défini en fonction de l'architecture globale du système. Le modèle utilisateur permet de sélectionner les différents objets à présenter. La base d'objets contient tous les éléments en relation avec le domaine et pouvant être inclus dans le document à afficher. Le générateur de pages fonctionne de la même manière qu'un moteur SQL. En tenant compte des différents paramètres, il génère et affiche, d'une manière transparente pour l'utilisateur, les objets appropriés et structurés correctement dans la page, conformément à la requête

faite en fonction du contexte.

Brusilovsky [Brusilovsky96] propose un schéma qui résume les différents aspects des systèmes adaptatifs. Sofiane Bénadi a complété ce schéma : la figure 3.2 représente ce qui semble important de retenir sur l'adaptativité pour mieux comprendre certains choix.

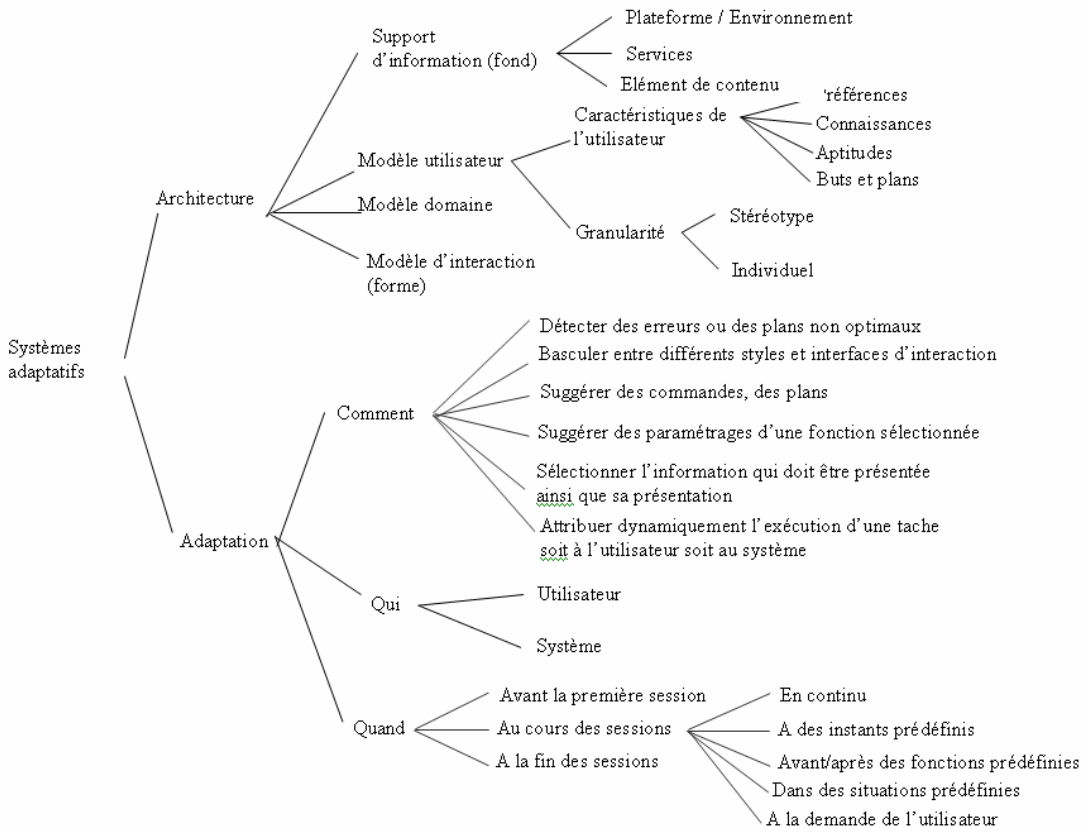


Figure 3.2 : Aspects fondamentaux des systèmes adaptatifs

3.2. Problèmes ouverts

Les principaux verrous scientifiques qui restent à lever dans cette thématique peuvent être répartis dans deux catégories. La première catégorie provient de l'ensemble des fonctions de base qu'un environnement de composition multimédia doit fournir aux auteurs :

- la spécification de la structure des documents doit être possible sans compétence en programmation informatique,
- les outils « auteurs » doivent permettre le passage simple et souple de la représentation mentale que se fait l'auteur du document à un véritable scénario d'affichage formalisé dans le langage de spécification,
- le langage de spécification doit être modulaire pour permettre à l'auteur de réduire la complexité de son document et aussi de réutiliser des parties déjà spécifiées,
- les environnements « auteurs » doivent permettre de modifier facilement les documents de façon à permettre un processus incrémental de création,

- les environnements « auteurs » doivent offrir des supports visuels permettant de visualiser le document au cours de sa spécification.

La seconde catégorie concerne la consultation et l'utilisation des documents. Coté lecteur, les points critiques peuvent selon nous s'exprimer de la manière suivante :

- comment structurer la navigation des lecteurs dans les différents documents sans altérer la liberté d'accès à l'information ? Un document en ligne se veut accessible à de nombreuses personnes, quels que soient leurs intérêts, leur situation géographique, le moment où elles sont disponibles pour la recevoir,
- comment avoir des informations sur le profil des utilisateurs ?
- comment faire en sorte que le contenu soit adapté "automatiquement" au profil de l'utilisateur ?
- comment faciliter le partage, la capitalisation et la réutilisation d'EdC (ressources) au sein de différents documents (description, méta-données, contraintes d'interopérabilité) ?

Jusqu'à présent, la plupart des travaux n'ont abordé ces problématiques que par l'un des deux points d'attaque possibles : le contenant ou le contenu. Ainsi, l'IEEE (groupe LTSC) a engagé des travaux de standardisation sur les architectures technologiques des systèmes de formation [LTSC02]. Ces travaux s'efforcent de modéliser le processus d'apprentissage de manière très générique et ne tiennent, par exemple, pas compte des aspects Auteurs (création, modification, gestion et adaptation des contenus). Concernant les contenus et leur structuration, les normes LOM [LOM97], Dublin-Core [Dublin-core98], EML [EML00], ADL-SCORM [Scorm98], IMS [IMS00], AICC [AICC98] et les travaux de S. Chabert-Ranwez [Chabert00] sur les documents hypermédias adaptatifs sont très intéressants mais oublient les contenants. Rappelons pourtant qu'au final, lors de l'utilisation (lorsque contenant et contenu se rejoignent inévitablement), les résultats restent décevants. Cela est probablement dû à l'histoire puisque dans les débuts de l'Enseignement Assisté par Ordinateur, il n'y avait aucune séparation entre contenu et contenant. Il semble, donc, important d'adapter les anciens modèles aux technologies actuellement utilisées.

3.3. Contributions

Nous avons tenté de répondre à la plupart des questionnements soulevés dans la section précédente lors de la thèse de S Benadi en considérant principalement l'exemple des systèmes de télé-enseignement qui nécessitent l'emploi d'interfaces et de documents numériques adaptatifs. Il nous semble que la plupart de ces difficultés (rencontrées lors de la mise en place d'une formation à distance) se trouvent dans un mauvais interfaçage entre le contenant (plate-forme pédagogique) et les contenus proposés (ensemble des activités pédagogiques multimédias). Une nouvelle fois, une réflexion sur la manière de structurer et de décrire les EdC entrant en jeu dans la conception des documents numériques et interfaces adaptatives parait nécessaire pour faire émerger de nouvelles solutions.

3.3.1. Structuration des documents numériques adaptatifs

La structuration que nous proposons a été conçue dans le cadre applicatif du télé-enseignement et utilise donc des dénominations propres à ce domaine mais il nous semble assez facile d'étendre nos propositions à l'ensemble des environnements nécessitant l'utilisation d'hyperdocuments adaptatifs.

Concept d'Activité Pédagogique Médiatisée

Nous avons choisi de définir la notion d'Activité Pédagogique Multimédia ou Médiatisée (APM). Cette entité constitue le cœur du modèle proposé ; elle a comme objectif principal de faciliter la création, la gestion et la réalisation (consultation) d'activités pédagogiques. Selon ce modèle, un enseignant (formateur) ou un apprenant ne pourra effectuer « son travail » que dans le cadre d'une APM. C'est aux enseignants (auteurs) de concevoir les APM selon leur désir, leurs objectifs, leurs envies et leurs habitudes. Pour cela, ils travaillent à partir d'outils que le système leur propose, d'outils personnels et de données/contenus personnels mis à leur disposition par le système. Bien que chaque enseignant ait un degré de liberté immense dans la mise en place de ses APM, des types et des modèles d'APM ont été prédéfinis afin de faciliter leur tâche, afin de structurer et d'harmoniser les contenus pédagogiques au sein des formations. Ce point est, également, crucial pour des questions d'interopérabilité.

L'APM constitue l'entité pédagogique de base (document ou interface) gérée par le système. Une APM répond à un objectif pédagogique précis et constitue une entité complètement autonome reposant à la fois sur des supports d'information (EdC pédagogiques) et sur des services logiciels (chat, forum, streaming, …) nécessaires pour un déroulement correct. En plus, à chaque APM peuvent être associées différentes caractéristiques descriptives (méta-données) ainsi que des règles de structuration qui définissent la manière d'adapter sa présentation en fonction du contexte (profil de l'apprenant). Ces règles sont définies par l'auteur lors de la construction de l'APM.

Concept de Briques Elémentaires

Nous définissons une brique élémentaire (BE) comme étant l'unité minimale (EdC) entrant dans la composition d'une APM (document) et pouvant être gérée par le système. Il peut s'agir d'un élément textuel, d'une image, d'un son, d'une vidéo, d'une interface de saisie, etc. Elle correspond la plupart du temps à un fichier multimédia (pour les données passives) ou à un "programme exécutable" (quand il s'agit d'un élément d'interaction). Les enseignants (auteurs) peuvent produire leurs propres BE ou utiliser des BE mises à disposition par d'autres enseignants ou par le système lui-même. Ces briques de base ne sont pas forcément associées à une seule activité, mais au contraire, sont conçues pour être réutilisées dans différentes activités. Ainsi, une APM utilise de manière évolutive les Briques Elémentaires qui sont nécessaires et suffisantes à son exécution.

Gestion de l'adaptativité

Modèle de l'utilisateur

La modélisation de l'utilisateur vise la construction d'un modèle cognitif à partir de l'observation du comportement d'un apprenant vis à vis de l'interface d'un système informatique [Self, 1988]. Au delà de la simple prise en compte des actions locales de l'utilisateur lors de la réalisation de son travail, cette modélisation doit pouvoir fournir des explications sur le raisonnement de ce dernier en identifiant les connaissances sous-jacentes. Dans une perspective d'enseignement, un modèle de l'apprenant est un modèle qui permet une inférence de ce que "connaît" (au sens large) l'apprenant au fur et à mesure de l'interaction avec le système. Ainsi, classiquement, le modèle de l'apprenant ne comprend que des informations sur les connaissances acquises par l'apprenant durant ses activités pédagogiques. Il se construit de façon incrémentale en gardant une trace des activités effectuées et de la qualité des actions de l'apprenant.

Pour répondre à nos objectifs, il nous a paru important d'élargir le modèle de l'apprenant au-delà de la dimension "connaissances acquises" (ce qu'il sait) en y intégrant des indications sur ce qu'il est (comportement, aptitudes) et sur ce qu'il aime (préférences). Nous avons choisi de réaliser la modélisation de l'utilisateur grâce à des attributs et à des règles d'évolution de ces paramètres. L'apprenant n'est pas vu comme un puits vide de connaissance qu'il faudrait remplir mais comme un humain aux facettes multiples [Prévot97]. C'est pourquoi nous proposons un modèle de l'apprenant comportant trois parties :

- le niveau de connaissances de l'apprenant ⇒ *Ce qu'il sait et veut apprendre*
- l'aptitude de l'apprenant ⇒ *Ce qu'il est*
- les préférences de l'apprenant ⇒ *Ce qu'il aime et désire*

C'est probablement les deux dernières parties du profil qui sont les plus difficiles à mettre en œuvre et à utiliser. Les questions à résoudre sont :

- faut-il que le système, au fil des actions, détecte (et comment ?) ce qu'aime un apprenant et fasse évoluer les attributs associés ? ou faut-il faire confiance à une auto-déclaration ?
- comment obtenir des informations sur ce qu'est l'apprenant [Dufresne01] ?

Pour essayer de répondre à ces questions, nous avons proposé d'avancer par étapes successives et nous avons, avec l'aide de chercheurs en SHS, mis en place un ensemble d'attributs permettant de caractériser chacune de ces parties [Benadi02a]. Les valeurs associées aux différents attributs devront évoluer au fur et à mesure des actions (réponses, mode de navigation, comportements, …) de l'apprenant au cours des activités pédagogiques pour ainsi enrichir son modèle et s'adapter au maximum à ses caractéristiques personnelles.

Adaptation de la navigation et du contenu

Dans notre modèle, chaque enseignant (formateur) propose des activités portant sur un ensemble de connaissances élémentaires assurant la couverture cognitive de son cours. Ces Eléments de Connaissance (EC) peuvent être communs à plusieurs disciplines. Pour chaque apprenant, c'est le taux d'assimilation mesuré de chacun de ces EC qui constitue le moyen choisi pour implémenter le suivi pédagogique. Durant la mise en place d'un cours, chaque enseignant a la possibilité de dresser la liste des EC qu'il désire aborder. Il peut associer un ou plusieurs EC à différentes APM. Il peut définir des règles faisant évoluer le taux d'assimilation des EC en fonction des actions faites par un apprenant à l'intérieur d'une APM (lecture, réponse à un exercice, …).

Les hyperdocuments conduisent à des parcours peu structurés, aléatoires et que la consultation libre rend parfois anarchiques. Et ce d'autant plus que l'outil informatique est privilégié au contenu du document. Dans notre environnement, l'historique de chaque apprenant mémorise les APM déjà consultées et les EC déjà abordés. Le formateur a, ainsi, connaissance de la manière dont est utilisé son cours. En outre, grâce à un système de pré-requis, il intervient au niveau de la macro-navigation (adaptation de la navigation) en conditionnant l'accès à certaines APM, créant ainsi un parcours pédagogiquement structuré, fonction des objectifs pédagogiques associés à chaque apprenant. Lorsqu'un apprenant tente de consulter une APM alors qu'il n'a pas acquis les EC pré-requis définis par le formateur, le système lui indique les chemins pédagogiques qu'il peut emprunter afin d'assimiler les concepts nécessaires à la bonne compréhension du contenu de l'APM à laquelle il désirait accéder. Une fois leurs besoins ponctuels identifiés et satisfaits, les apprenants peuvent ensuite retrouver une autonomie en accédant à l'ensemble des APM ou à des sites extérieurs conseillés par l'enseignant.

Au niveau de la micro-navigation (adaptation du contenu), chaque APM est à même d'auto-adapter sa présentation en fonction du contenu du modèle de l'apprenant et des spécifications données par l'auteur lors de la construction de l'APM.

3.3.2. Structuration des environnements hypermédias adaptatifs

Une structuration multi-niveaux

Pour ne négliger, ni le contenant, ni le contenu, nous proposons [Benadi00b, Ramel02b] un découpage des systèmes d'informations en trois niveaux (couches). Support, Structure et Sémantique (figure 3.3). On peut noter que l'on retrouve ici en des termes légèrement différents la notion de structure physique, logique et sémantique des documents numériques rencontrés dans le chapitre précédent.

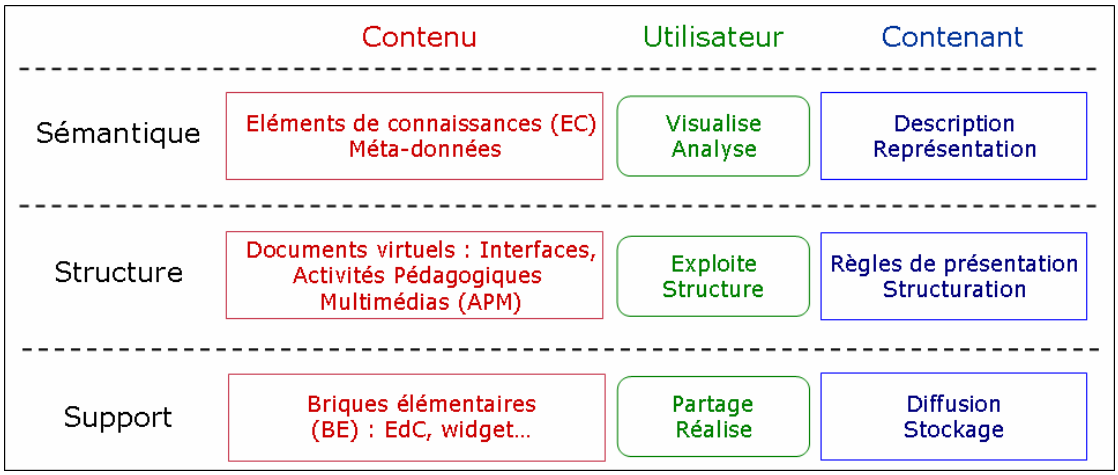


Figure 3.3. Structuration horizontale du système

Le niveau Support

Le niveau Support est le plus proche de la machine, donc, le plus basique. C'est sur lui que reposent toutes les activités et toutes les fonctionnalités de la plate-forme de télé-enseignement. Ce niveau concerne la création, le stockage et la gestion des Briques Elémentaires (EdC) que constituent les images, les textes, les vidéos et autres supports d'informations stockés sous forme de fichiers. Ces ressources doivent pouvoir (au niveau supérieur) être partagées par différents enseignants et, ainsi, intervenir dans différentes activités pédagogiques. Les règles d'assemblage et les styles de présentation de ces BE ne sont, par contre, pas gérés au niveau Support mais au niveau Structure.

Le niveau Structure

Ce niveau constitue le cœur d'un environnement pédagogique : c'est ici que sont définies les activités pédagogiques, les formes de présentations, les types d'interaction, de collaboration, de coopération et d'organisation (enchaînements, relations) des BE en fonction des objectifs pédagogiques. Il s'agit de structurer les ressources pédagogiques et les outils logiciels, de les paramétrer de manière à rendre possible (au niveau du contenant) la génération automatique d'activités pédagogiques (APM) respectant des règles et/ou des objectifs pédagogiques précis. Selon les profils, une forme de présentation (hiérarchique, séquentielle, métaphorique, conceptuelle, …) plus qu'une autre, peut favoriser l'acquisition des connaissances par l'apprenant [Meyer01]. La diversité de ces structures de présentation peut servir, non seulement à accéder aux informations de façon personnalisée, mais également, au renforcement de l'ancrage en mémoire. Le choix de la forme de présentation peut être effectué dynamiquement par le système ou explicitement précisé par l'apprenant, toujours dans le but d'obtenir une information la plus personnalisée possible.

La structure et la forme des activités pédagogiques proposées incombent principalement aux enseignants (auteurs) qui les conçoivent à l'intérieur d'outils « auteurs » en exploitant pour cela les BE mises à leur disposition par le niveau 1 (Support) de la plate-forme. Encore plus qu'ailleurs, à l'intérieur d'une APM (micro-navigation) ou entre les APM (macro-navigation), la navigation libre ne doit être possible que si

elle a été étudiée. Dans notre modélisation, nous considérons que la définition des patrons de navigation est à la charge de l'enseignant (auteur) qui crée les APM. Il s'agit là d'une lourde tâche qui se traduit par la mise en place de règles définissant les styles de présentation en fonction des profils des apprenants.
Il nous semble, également, qu'aujourd'hui très peu de systèmes offrent des fonctionnalités suffisantes de macro-navigation (mise en relation de différentes activités pédagogiques ou de parties d'activités pédagogiques). La mise en place du niveau Sémantique prévu dans notre modélisation apporte une réponse à ce problème non trivial.

Le niveau Sémantique

Le niveau Sémantique favorise la transparence en rapprochant l'espace des informations (enfouies dans le contenu) de l'espace cognitif de l'utilisateur (qu'il soit enseignant ou apprenant). Grâce à la définition des EC (éléments de connaissances), le niveau sémantique fournit une image claire de l'organisation des connaissances à acquérir ou à transmettre dans les différentes disciplines. Il devient aisé de produire une vue structurée des contenus (carte des concepts) afin de réduire la désorientation de l'apprenant et la difficulté que celui-ci éprouve à explorer de façon cohérente cet espace complexe. La désorientation ressentie dans les systèmes hypermédias est souvent liée à l'incapacité qu'a l'apprenant, qui passe d'une information à une autre, de se rappeler d'où il vient et ce qu'il cherchait à faire. Ces éléments sont liés à la macro-navigation qui, dans notre cas, est conçue par les enseignants (formateurs) et est contrôlée par le système (pré-requis sur les APM via les taux d'assimilation des EC). L'exploitation des informations du niveau Sémantique (EC, taux d'assimilation, …) permet de fournir des outils aux apprenants comme, par exemple, une vue globale des éléments de connaissances insérés dans la formation, vue structurée et donc structurante, faisant le lien entre les concepts transversaux à plusieurs disciplines.

3.3.3. Présentations adaptatives

La présentation (ou exécution) d'un document multimédia consiste en la restitution des objets « média » qu'il contient en respectant d'une part les contraintes internes représentées par les relations de composition entre ces objets (micro-navigation), et d'autre part les contraintes externes imposées par le contexte et par la disponibilité des ressources (macro-navigation). La prise en compte de ces contraintes soulève de nombreux problèmes dans la réalisation des fonctions de présentation comme la prise en compte du contexte courant, la gestion de la navigation et la synchronisation entre objets et documents. Pour gérer ces aspects, nous avons réalisé un environnement pouvant accueillir nos contenus adaptatifs (APM et BE). Ce système nommé OWASIS (Outil Web pour l'ApprentiSsage à dIStance) a été installé dans plusieurs établissements et départements de l'INSA de Lyon et a été intégré dans le campus numérique INSA-V (intranet pédagogique). Les entités gérées par ce système correspondent exactement aux notions d'APM, d'EC et de BE tels que nous les avons définis.

La plate-forme OWASIS est un prototype de LMS (Learning Management System). Cet environnement « full Web » est destiné à être régulièrement alimenté par des enseignants qui créent des APM et utilisé par les apprenants inscrits dans différentes formations. Cet outil devient, alors, le fil conducteur pédagogique tout au long de leur parcours. OWASIS a, également, vocation à favoriser la communication entre les enseignants et les enseignés : réflexion sur la forme et le contenu des cours, ….

Lien entre contenant et contenu

Il est indispensable que chaque APM fournisse les moyens d'instaurer une communication avec l'environnement qui l'accueille (contenant). Pour cela la figure 3.4 résume notre proposition

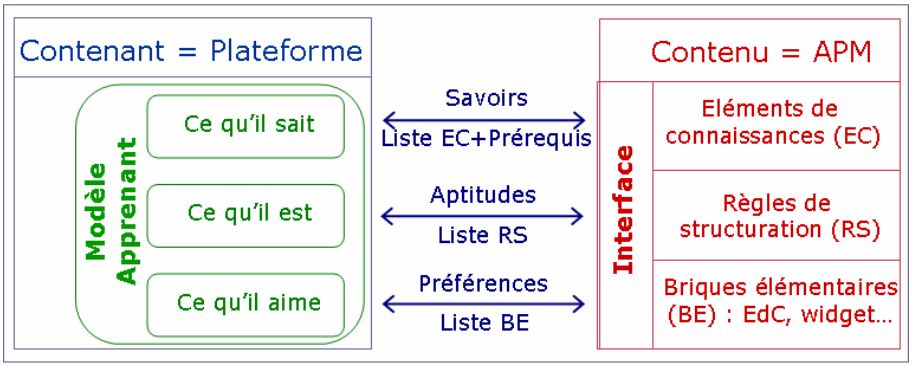


Figure 3.4 : Modélisation de la structure d'une APM et de ses interfaces

Une couche (interface) ajoutée à chaque APM doit pouvoir, par des mécanismes d'introspection (examen de son état et de sa structure), fournir :

- au niveau 1, la liste des Briques Elémentaires qu'elle utilise ou englobe,
- au niveau 2, la liste des règles définissant les structurations possibles de son contenu (pour l'adaptation en fonction du profil de l'apprenant),
- au niveau 3, la liste des connaissances qu'elle aborde ainsi que la liste des pré-requis nécessaires à sa compréhension.

A l'inverse, l'interface de l'APM doit, également, par des mécanismes d'intercession [Bourguin00] (modification de son état, de sa structure, de son comportement) permettre :

- au niveau 1, la suppression, la modification, l'ajout de BE et le réglage de leurs paramètres,
- au niveau 2, la modification des structures et des apparences prévues en fonction des règles qui ont été définies lors de sa création concernant son auto-adaptation en fonction de ce qu'aime, sait et souhaite savoir l'apprenant,
- au niveau 3, les relations existant entre son contenu et le contenu des autres APM.

Une plate-forme de télé-enseignement doit également fournir des fonctionnalités pour le suivi des apprenants. Celles-ci existent dans notre modèle puisque l'interface que nous ajoutons aux APM n'interagit principalement qu'avec le modèle des apprenants (qui est géré par la plate-forme). Il est à

noter que le modèle de l'apprenant repose, également, sur un découpage en 3 niveaux comme nous l'avons vu précédemment.

Mise en œuvre

Dans [Benadi2004], nous avons défini des DTDs (Document Type Definition) qui décrivent, d'une part, la structure que doivent respecter les Activités Pédagogiques Multimédias (APM), et d'autre part, les briques élémentaires (BE) telles que nous les avons présentées dans le chapitre précédent. Une DTD a, également, été définie pour spécifier comment construire le modèle des apprenants à l'intérieur du contenant (plate-forme LMS). Ce sont ces trois DTD qui contribuent à l'interopérabilité et à l'adaptativité en créant un lien entre les contenus et le contenant.

Dans les systèmes pédagogiques adaptatifs, le modèle du domaine est la composante qui permet à l'ordinateur de «connaître » ce qui va être enseigné à l'apprenant. Grâce à notre proposition, ce modèle devient dynamique puisque ce sont les auteurs qui le créent, le gèrent, et le font évoluer grâce à un méta-modèle du domaine (EC +APM + BE). La figure 3.5 résume la structure de ce méta-modèle du domaine.

A un instant donné, le modèle du domaine représente le savoir sélectionné par l'enseignant pour construire une formation donnée. Il a en charge de décrire du mieux possible la structure générale de cette dernière, aussi bien au niveau macroscopique (macro-navigation) qu'au niveau microscopique (micro-navigation). Comme le montre la figure 3.5, la macro-navigation est régie par les pré-requis mis en place sur les différentes APM proposées dans la formation. Ces pré-requis peuvent évoluer grâce aux interventions du formateur et ont préalablement été mis en place par l'auteur du cours. La micro-navigation (navigation à l'intérieur d'une APM et de ses interfaces) est, quant à elle, complètement à la charge de l'auteur qui a construit l'activité pédagogique : il a défini différents styles de présentations potentielles. Ces présentations sont définies à l'aide de règles de structuration faisant le lien entre le profil de l'apprenant et les contenus de l'APM devant être affichés ou filtrés.

Lors de l'utilisation d'OWASIS, sa structuration en trois niveaux est transparente aux utilisateurs. Ceux-ci passent d'un niveau à l'autre, aussi bien pendant la réalisation d'une activité pédagogique (côté apprenant), que lors de la création, ou lors du suivi (côté enseignant). L'environnement offre à tout moment la possibilité de faire le lien entre les activités pédagogiques (APM – niveau 2) et les concepts/connaissances associés (EC – niveau 3). Les usagers peuvent visualiser l'ensemble des éléments de connaissance abordés dans l'APM ou dans leur formation de manière plus globale. Cela améliore, en outre, la lisibilité et la cohérence entre les activités créées par les différents enseignants grâce à une vision globale du contenu (utile d'ailleurs aussi bien aux enseignants qu'aux apprenants). Les fonctions de suivi pédagogique interviennent, également, dans les différents niveaux de notre structuration. L'historique de chaque apprenant résume à la fois les activités réalisées, les EC et les taux d'assimilation associés.

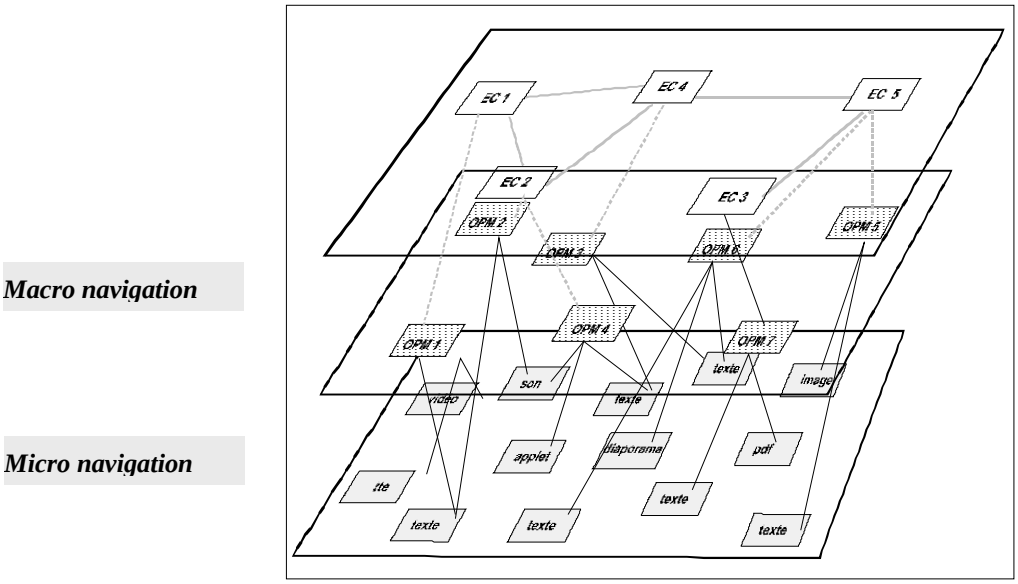


Figure 3.5 : Méta- modèle du domaine et structuration en 3 niveaux

Une activité pédagogique de type « Cours/Chapitre » peut se décliner sous plusieurs formes de présentation en fonction du profil de l'apprenant. La figure 3.6 contient un affichage sous forme sommaire (figure 3.6a) et un affichage sous forme intégrale (figure 3.6b) d'une même APM exemple.

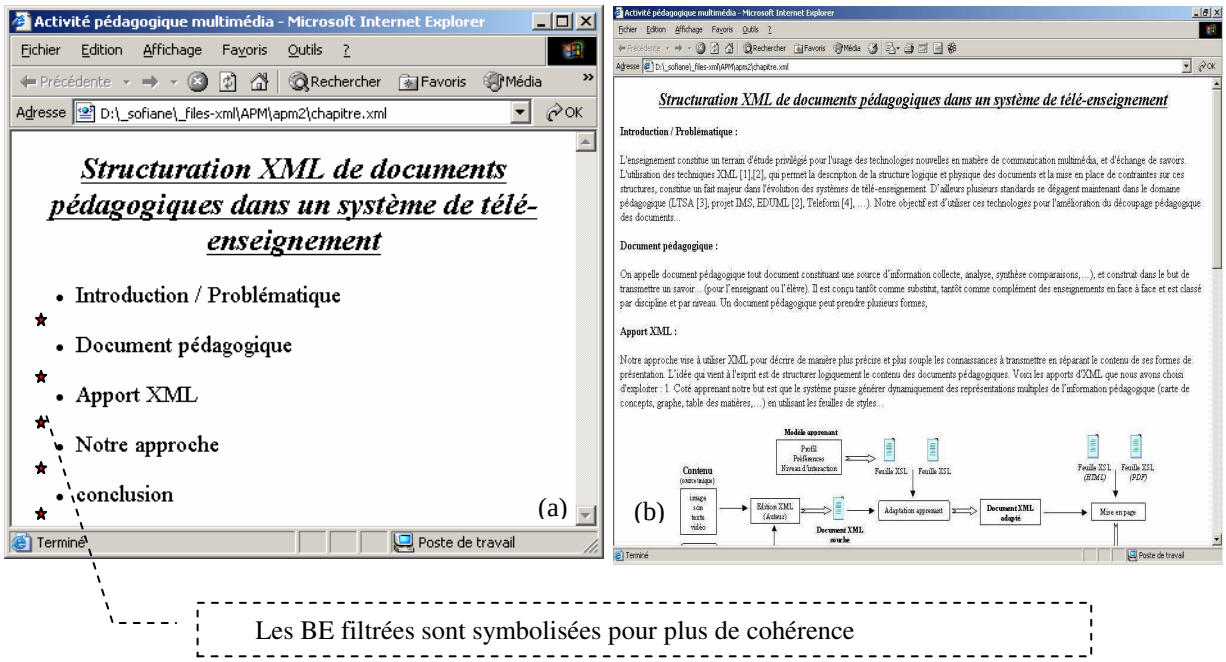


Figure 3.6 : Exemple de deux présentations différentes d'une APM

Cette différence de présentation est simplement liée au fait que le profil de l'apprenant spécifie dans un cas, une préférence pour un affichage présentant uniquement les BE de type « Titre », et dans l'autre cas, une préférence pour un affichage intégral de toutes les BE sans filtrage. Les différentes présentations sont construites dynamiquement, par transformation, en fonction du contenu du profil de l'apprenant. L'application d'une feuille de style XSL (construite à partir des données stockées dans un profil

apprenant) sur le document XML constituant l'APM, produit un document pédagogique adapté au niveau de connaissances et aux préférences de l'apprenant. Grâce à XSLT [XSLT02], il est possible de réaliser deux opérations : la sélection de l'information désirée (BE) et la présentation de cette information selon le formatage désiré. La figure 3.7 résume les étapes nécessaires à la construction d'une interface / d'un document adapté en fonction du profil d'un apprenant.

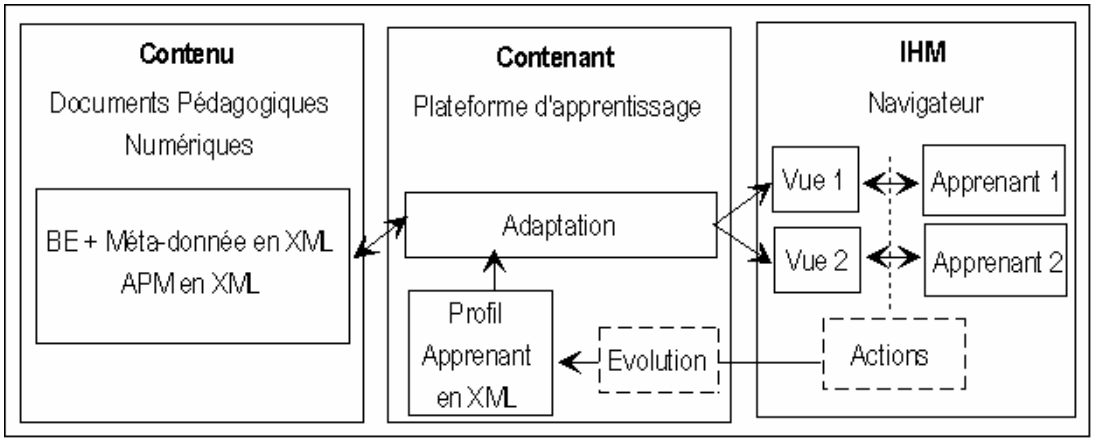


Figure 3.7 : La construction d'une activité pédagogique adaptée

Ces travaux, réalisés au sein d'un laboratoire pluridisciplinaire (ICTT est un laboratoire mixte Sciences Humaines et Sociales et Sciences pour l'Ingénieur) m'ont offert l'opportunité de donner une dimension nouvelle à mon expérience de travail en équipe. Ils m'ont aussi montré que des collaborations avec des chercheurs dans d'autres disciplines sont, dans certains cas, d'une importance primordiale.

Conclusion générale

Bilan

Par ce tour d'horizon de mes principales activités de recherche dans les domaines de l'analyse d'images de documents, de la reconnaissance des formes et des documents (interfaces) adaptatifs, j'espère avoir montré l'intérêt qu'il peut y avoir de porter une vision plus globale sur la représentation et le traitement des documents numériques. Nous avons vu que, dans tous les cas, une étude des différents modes de représentation et de structuration des contenus est primordial. Nous avons noté, qu'à chaque fois, cette étude peut être produite à trois niveaux différents (physique, logique, sémantique) et que l'un des principaux problèmes est de ne pas succomber à la tentation de produire « un système » trop spécifique à un domaine d'application donné (un type de documents) comme c'est encore trop souvent le cas (manque de généricité).

.

Concernant la description du contenu des images de documents, le principal problème réside dans la mise en place d'algorithmes de construction automatique d'une représentation abstraite à partir des données directement fournies par les dispositifs d'acquisition en hors ligne (images) et en ligne (suite de points horodatés). Dans ce cadre, mes contributions se sont focalisées sur l'utilisation conjointe d'informations contour (par vectorisation avec Matthieu Wirotius et Rashid Jalal Qureshi) et région (cartes des textures avec Nicholas Journet). Je propose ensuite d'effectuer une structuration des informations extraites sous forme de graphes structurels grâce auxquels il est possible de construire une description des contenus des documents pouvant couvrir de nombreux domaines. Les éléments constituant le document sont alors représentés par les nœuds du graphe tandis que les arcs du graphe décrivent les relations (topologiques, temporelles, d'imbrication, …) existant entre ces nœuds. Des attributs peuvent être ajoutés aux arcs et aux nœuds afin de les caractériser aussi finement que nécessaire, physiquement (forme) et sémantiquement (fond), de manière statistique ou dynamique au cours des traitements.

Outre la possibilité de décrire les relations entre objets ou parties d'objets, l'avantage de ce type de représentation réside dans les nombreux outils, déjà existants, permettant de rechercher (localisation et reconnaissance) des parties d'objets ou des EdC à l'intérieur d'ensembles plus importants (recherche de symboles dans un document complet par exemple). Forcément, cet avantage est compensé par certains

aspects négatifs qui concernent, d'une part, la difficulté de mise en place d'une étape d'apprentissage (des modèles) et, d'autre part, la faible tolérance aux légères différences (comparaison inexacte et mesure de similarité). Pour pallier ces problèmes, je pense qu'il est nécessaire de combiner les méthodes structurelles avec les méthodes statistiques. C'est ce que nous effectuons dans le cadre des travaux de thèse de Rashid Jalal Qureshi qui propose différentes méthodes permettant de mesurer des similarités entre des graphes attribués.

Nous avons indiqué qu'une représentation structurelle des contenus des documents paraissait satisfaisante et générique puisqu'adaptable à la plupart des types de documents numériques (hypermédias, images de documents graphiques, composites et tracés en ligne). Nous avons également noté, lors de nos références au système visuel humain, qu'une « contextualisation » des traitements d'analyse peut être efficace. De ce point de vue, le traitement d'un document doit correspondre à des transformations successives appliquées aux informations acquises et mises en commun via une ou plusieurs représentations. Le problème de l'interprétation (reconnaissance) réside alors dans l'identification d'un ensemble de traitements spécialisés mais interdépendants permettant d'obtenir des applications plus génériques ou tout au moins plus facilement adaptables en séparant bien les données initiales, les connaissances a priori et les algorithmes de reconnaissance. Pour obtenir cette faculté d'analyse incrémentale, nous pensons qu'il est possible de faire en sorte que les processus d'analyse et de reconnaissance s'efforcent d'extraire les nouvelles connaissances en prenant un minimum de risque. Les nouvelles connaissances peuvent alors être utilisées pour résoudre les problèmes les plus complexes à partir de données non ambiguës (analyse contextuelle). Enfin, nous préconisons d'introduire l'utilisateur final au plus tôt dans la chaîne des traitements en lui permettant de guider l'analyse sans pour autant avoir de fortes compétences en traitement d'images ou en reconnaissance des formes (analyse guidée par l'utilisateur proposée avec Stéphane Leriche).

Pour ce qui est de la production et de la consultation de documents adaptatifs, les notions de transformations successives et de granularité des contenus restent globalement identiques à celles utilisées lors de l'analyse des contenus. Un document numérique peut à nouveau être vu à différents niveaux. Au niveau le plus bas, les documents numériques peuvent être vus comme un assemblage d'objets élémentaires sur un support physique (niveau support). Ensuite, il est possible de s'intéresser à l'enchaînement et aux liens existant entre les EdC (niveau structure). Enfin, on peut s'intéresser au sens et aux messages véhiculés par le document (niveau sémantique). Avec Sofiane Bénadi, nous avons tiré partie de cette structuration en trois niveaux pour simplifier et améliorer la production et la consultation de documents hypermédias adaptatifs dans le cadre du télé-enseignement.

Perspectives

La « petite » expérience que j'ai acquise au cours des années qui ont succédé à ma thèse me laisse aujourd'hui penser qu'il est préférable (et plus simple !) de se fixer des objectifs raisonnables à court terme plutôt que d'essayer d'énoncer de grandes idées. Je ne tenterai donc pas, dans cette partie, de prévoir quels seront les futurs axes de recherches des différentes communautés de recherche travaillant sur les documents numériques mais je préciserai, plus simplement, les problématiques sur lesquelles j'aimerais travailler dans un avenir proche sachant qu'il reste encore beaucoup de choses à faire et, qu'aujourd'hui encore, la notion de document numérique est en constante évolution.

Je désire bien évidemment poursuivre les travaux, déjà bien entamés, qui concernent la mise en place d'une représentation générique capable de décrire précisément le contenu des images de documents, qu'ils soient majoritairement textuels ou graphiques. L'idée de combiner, à l'aide de graphes structurels, des informations région (texture) avec des informations contour (vecteurs) me semble pertinente. Reste à proposer des méthodes permettant la construction d'un unique graphe rassemblant simultanément toutes les informations disponibles au sein d'une même structure. Actuellement, les cartes de textures, les vecteurs représentant les contours des formes pleines, les quadrilatères représentant les formes fines et les composantes connexes apportant une information sur les différentes formes contenues dans l'image sont étudiées indépendamment les uns des autres lors de la phase de reconnaissance ce qui restreint le champ d'application de la représentation proposée.

Une fois cette représentation construite, encore faut-il être capable de proposer des algorithmes (de reconnaissance et d'analyse) capables de l'utiliser efficacement. La mise en place du prototype AGORA, utilisé depuis plus de deux ans par le CESR pour indexer leurs ouvrages anciens, constitue la première brique et la première preuve de l'utilisabilité de la représentation structurelle proposée. Le procédé d'analyse interactive guidée par l'utilisateur rendue possible à l'aide des scénarios qu'utilisent AGORA me paraît également intéressant mais n'a pour l'instant fait ses preuves que sur un type bien précis de documents (les documents imprimés anciens). De plus, il n'utilise qu'une partie des informations disponibles dans la représentation. J'aimerais donc avoir l'opportunité d'étendre cette approche à d'autres types d'images de documents, notamment aux documents graphiques. Pour cela, il sera nécessaire d'adapter (en généralisation) les algorithmes de traitements pouvant être intégrés dans les scénarios et probablement aussi de revoir certains aspects de la représentation utilisée. Ces travaux ayant des points communs avec les travaux réalisés en vision cognitive, j'aimerais également étudier les points communs et les différences entre ces deux approches.

Le travail entamé dans la thèse de Rashid Jalal Qureshi visant à apporter des améliorations aux méthodes de reconnaissance structurelle en permettant d'utiliser plus facilement des mesures de similarité entre

graphes me passionne également. Combiner au maximum méthodes structurelles et méthodes statistiques me semble une voie très prometteuse. L'idée de mettre en place des mécanismes d'apprentissage automatique des modèles adaptés aux méthodes structurelles (génération automatique de graphes modèles) va dans le même sens et mérite, à mon avis, d'être plus approfondie. Pour cela, j'espère avoir l'opportunité de continuer, voire d'augmenter, mes collaborations avec d'autres chercheurs ayant des compétences complémentaires aux miennes. J'ai d'ailleurs la chance d'appartenir à une équipe de recherche comportant à la fois des spécialistes en traitement d'images et des spécialistes en reconnaissance de formes : il serait dommage de ne pas profiter, encore plus qu'actuellement, de cette complémentarité.

La représentation et le traitement des documents numériques sont de vastes domaines de recherche qui continuent à nous lancer des défis passionnants. Je souhaite que mon travail puisse apporter une petite contribution à la résolution de certains de ces défis. J'espère, en tout cas, avoir provoqué, chez certains lecteurs, l'envie d'utiliser des méthodes proches de celles proposées dans ce rapport afin d'améliorer encore leurs performances.

Enfin, je tiens aussi à mentionner que, durant toutes ces années de recherche, j'ai beaucoup apprécié et beaucoup appris au cours des différents séminaires et journées thématiques organisés par les groupes de recherche (GDR, GRCE, AFRIF, …) ou les consortiums de projets (ACI Madonne, projet Technovision EPEIRES, …). Je remercie donc les chercheurs qui se sont engagés et qui s'engagent encore actuellement pour que de telles rencontres existent. J'espère qu'elles persisteront et j'espère trouver, prochainement, encore plus de temps à consacrer à cette animation scientifique.

Références

[Abe86] Abe K., Azumatani Y., Mukouda M. and Suzuky S., Discrimination of symbols, lines, and characters in flowchart recognition. Proceedings of the 8th International Conference on Pattern Recognition, Paris (France), October 27-31, 1986. Vol. 2, p1071-1074.

[Adam00a] Adam S., Ogier J. M., Cariou C., Mullot R., Gardes J., Lecourtier Y., Fourier Mellin Based Invariants for the Recognition of Multi-Oriented and Multi-Scaled Shapes : Application to Engineering Drawing Analysis. Invariants for Pattern Recognition and Classification. Dans la série Machine Perception & Artificial Intelligence, World Scientific Publishing, 2000.

[Adam00b] Adam S., Ogier JM., Cariou C., Mullot R., Labiche J., Gardes J., Symbol and Character recognition: application to engineering drawings. International Journal of Document Analysis and Recognition. 2000. Vol 3(2). p89-101.

[Adam01] Adam S., Interprétation de documents techniques : des outils à leur intégration dans un système à base de connaissances. Thèse soutenue à Rouen le 11/12/2001. 206p.

[Ahmed02] Ahmed M., Ward R., A Rotation Invariant Rule Based Thinning Algorithm for Character Recognition. Pattern Analysis and Machine Intelligence (PAMI), 2002. Vol 24(12), p1672–1678.

[AhSoon01] Ah-Soon C., Tombre K., Architectural Symbol recognition using a network of constrainst. Pattern Recognition Letters, 2001. Vol 22(2). p231-248

[Aicc98] AICC, Aviation Industry Computer Based Training Committee, 1998. Disponible sur : <http://aicc.org> (consulté le 26/05/04).

[Akindele93] Akindele O., Belaid A., Page Segmentation by Segment Tracing. 2nd International Conference on Document Analysis and Recognition, 1993, p341-344.

[Akpotsui97] Akpotsui E., Quint V., Roisin C., Type modelling for document transformation in structured editing systems. Mathematical and Computer Modelling, 1997. Vol 25(4): p1-19.

[Andersen03] Andersen T, Zhang W., Features for Neural Net Based Region Identification of Newspaper Documents. 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, UK. IEEE Computer Society 2003, p403-407.

[Andre99] Andre J., Richy H., Paper-less editing and proofreading of electronic documents Conference EuroTex99. (Heidelberg, Germany), September 1999, p3-16.

[Anquetil97] Anquetil E., Lorette G., On-line Handwriting Character Recognition System Based on Hierarchical Qualitative Fuzzy Modeling. Progress in Handwriting Recognition, World Scientific, 1997, p109-116.

[Antoine91] Antoine, D., CIPLAN : A ruled-based system with original features for understanding French plats. First International Conference on Document Analysis and Recognition, Saint-Malo (France), 1991, p647-655.

[Antonaco98] Antonacopoulos A., Page Segmentation Using the Description of the Background. Computer Vision and Image Understanding, Special Issue on Document Image Understanding and Retrieval, 1998, Vol. 70(3), p350-369.

[Arias94] Arias J.F., Prasad A., Kasturi R., Chhabra A., Interpretation of telephone central office equipment drawings. Proceedings of the 12th International Conference on Pattern Recognition, Jérusalem (Israël), october 9-13, 1994. Vol. 2, p310-314.

[Arias95] Arias J F., Chan Lai P., Surya S., Kasturi R., Chhabra AK., Interpretation of telephone system manhole drawings. Pattern Recognition Letters. Vol. 16(4): 1995, p355-369

[Azé00] Azé J, Richy H., Lorette G., Interface-stylo de correction en ligne de documents électroniques : application aux pages Web. Colloque International Francophone sur l'Ecrit et le Document. CIFED'2000, p261-269.

[Badawy02] Badawy O., Kamel M.. Shape Representation Using ConcavityGraphs. International Conference on Pattern Recognition (ICPR), 2002, p461–464.

[Baird90] Baird H., Jones S.E., and Fortune S.J., Image segmentation by shape-directed covers. In Proceedings of International Conference on Pattern Recognition, Atlantic City, NJ, June 1990, p820-825.

[Baird92] Baird H., Background structure in document images. In Advances in Structural and Syntactical Pattern Recognition, ED. H. Bunke. 1992, p253-269.

[Barr89] Barr A., Cohen PR., Feigenbaum EA., The handbook of artificial intelligence, Vol. 1-4. Addison Wesley Publisher, 1989, ISBN : 0201118106.

[Baruch88] Baruch O., Line Thinning by Line Following. Pattern Recognition Letters, Vol. 8(4), 1988, p 271–276.

[Basa04] Sabari S., Nishikanta J., Gabor filters for document analysis in indian bilingual documents. Proceedings International Conference on Intelligent Sensing and Information Processing, 2004, p123-126.

[Belaid97] Belaïd A., Analyse du document : de l'image à la représentation par les normes de codage, Document numérique, 1997. Vol. 1(1), p21-38.

[Bellet99] Bellet F., Garbay C., Une Interface Homme-Machine pour un Système de Vision de Bas Niveau, Revue d'Interaction Homme-Machine, Vol. 1(2), 1999, p37-60.

[BenYacoub95] Benyacoub S., Jolion JM., Characterizing the hierarchical Hough transform through a polygonal approximation algorithm, Pattern Recognition Letters. Vol. 16(4), 1995, p 389-397.

[Benyon93] Benyon D., Adaptive systems: A solution to Usability Problems. User modeling and user adapted interaction, 1993, vol. 1(3), p1-22.

[Blasselle97] Blasselle B., Histoire du livre, Collection, Decouvertes, Editeur Gallimard. 1997. Numéro 321. 160 pages.

[Blostein96] Blostein D., Fahmy H., Grbavec A., Issues in the Practical Use of Graph Rewriting. In Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science (LNCS 1073) , 1996, p38–55.

[Boatto92] Boatto, L., Consorti V., Detection and separation of symbols connected to graphics in line drawings. 11th International Conference on Pattern Recognition, The Hague (Netherlands), august 30-september 3, 1992. Vol. 2, p545-548.

[Bonhomme98] Bonhomme S., Transformation de documents structurés, une combinaison des approches automatique. Thèse de doctorat en informatique de l'Université Joseph Fourier, 1998, Grenoble.

[Bottoni92] Bottoni P., Mussio P., Protti M., Metareasoning as tool for pattern recognition. 11th International Conference on Pattern Recognition, The Hague (Netherlands), august 30-september 3, 1992, p285-289.

[Bourguin00] Bourguin G., Un support informatique à l'activité coopérative fondé sur la théorie de l'activité : le projet DARE. Thèse de doctorat en informatique, Université de Lille, 2000, 217 pages.

[Brault93] Brault J., Plamondon R., Segmenting handwritten signatures at their perceptually important points, IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), 1993, Vol. 15(9), p 953-957.

[Bres94] Bres. S., Contributions à la quantification des critères de transparence et d'anisotropie par une approche globale. Thèse de doctorat de l'INSA de Lyon, 1994.

[Breuel03] Breuel TM., An Algorithm for Finding Maximal Whitespace Rectangles at Arbitrary Orientations for Document Layout Analysis. International Conference for Document

Analysis and Recognition (ICDAR 2003). Vol.1, p66-70.

[Bruce99] Bruce KB., Draper A., Bins J., Adore: Adaptive object recognition. Proceedings of the International Conference on Vision Systems, Las Palmas, Spain, 1999. p522--537,

[Brusilovsky98] Brusilovsky P., Methods and techniques of adaptive hypermedia. Adaptive Hypertext and Hypermedia, P. Brusilovsky, A. Kobsa and J. Vassileva eds. Dordrecht: Kluwer Academic Publishers, 1998, p1-43.

[Brusliovsky96] Brusilovsky P., Adaptative Hypermedia: An At-tempt and Generalize, Multimedia, Hypermedia, and Virtual Reality, P. Brusilovsky, P. Kommers, N. Streitz Eds., Berlin: Springer-Verlag, 1996, p288-304.

[Brusilovsky03] Brusilovsky P., Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. British Journal of Educational Technology, 2003, Vol. 34(4), p487-497.

[Bullier95] Bullier J., Nowak LG., Parallel versus serial processing : new vistas on the distributed organization of the visual system. Current Opinion in Neurobiology, 1995, Vol. 5, p497-503.

[Bunke99] Bunke H., Error correcting graph matching. On the influence of the underlying cost function. IEEE Transaction on Pattern Analysis and Machine Intelligence. 1999. Vol. 21(9), p917-922.

[Calvary03] Calvary G., Coutaz J, Thevenin D, Limbourg Q., Bouillon L., Vanderdonckt J., An Unifying Reference Framework for multi-target user interfaces. Interacting with Computers, Vol. 15(3) , 2003. p289-308

[Cao01] Cao R. and Tan C.L. Separation of Overlapping Text from Graphics. 6th International Conference on Document Analysis and Recognition, Seattle (USA), 2001, p44-48.

[Chabert00] Chabert-Ranwez S., Composition Automatique de Documents Hypermédia Adaptatifs à partir d'Ontologies et de Requêtes Intentionnelles de l'Utilisateur, Thèse de doctorat, Université de Montpellier 2, décembre 2000. 187 pages.

[Champin03] Champin PA., Solnon C., Measuring the Similarity of Labelled Graphs. Proceedings of the 5th International Conference on Case-Based Reasoning, Eds. K.D. Ashley and D.G Bridge, Springer-Verlag. 2003, p80-95.

[Chang99] Chang F., Lu YC., Palvidis T., Feature Analysis Using Line Sweep Thinning Algorithm, Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 21(2). 1999, p145-158.

[Chatti06] Chatti N., Calabreto S., Pinon JM., Managing concurrent XML structures: The Multi-Structured Document building process. To appear in 17th Annual Information Resources Management Association International Conference, Washington D.C. (USA), May 21-24, 2006.

[Chen95] Chen S., Haralick RM., Recursive Erosion, Dilation, Opening and Closing Transforms. IEEE Transaction on Image Processing, Vol 4(3), 1995, p335-345.

[Chenevoy92] Chenevoy Y., Reconnaissance structurelle de documents imprimés : études et réalisations. Thèse de doctorat INPL, Centre de Recherche en Informatique de Nancy, décembre 1992.

[Chhabra98] Chhabra A., Graphic Symbol Recognition: An Overview, Graphics Recognition -- Algorithms and Systems, Eds. K. Tombre and A. Chhabra, Lecture Notes incomputer sciences (LNCS 1389), Springer Verlag, 1998, p68-79.

[Cinque98] Cinque L., Lombardi L., Manzini G., A multi-resolution approach for page segmentation. Pattern Recognition Letters, Vol 19(2), 1998, p 217-225.

[Clouard99] Clouard R., Elmoataz A., Porquet C., Revenu M., Borg : A knowledge-based system for automatic generation of image processing programs, IEEE Transaction on

Pattern Analysis and Machine Intelligence, Vol. 21(2), 1999, p128-144.

[Conte06] Conte D., Foggia P., Jolion JM. and Vento M., A Graph-based, Multiresolution Algorithm for Tracking Objects in Presence of Occlusions, Pattern Recognition, Vol. 39(4), 2006, p562-572.

[Cordella00] Cordella LP., Vento M., Symbol recognition in documents : a collection of techniques. International Journal of Document Analysis and Recognition. 2000. Vol 3(2). p73-88.

[Couasnon03] Couasnon B., Leplumey I., A generic recognition system for making archives documents accessible to public, Third International Conference on Document and Analysis and Recognition (ICDAR'03), Edinburgh, August 2003, p 228–232.

[Cugini84] Cugini U., Ferri G., Mussio P., Protti M., Pattern directed restoration and vectorization of digitized engineering drawings. Computer and Graphics, 1984, Vol. 8(4), p337-350.

[Cugini90] Cugini U., Falcidieno B., Mussio P., Exploiting knowledge in CAD/CAM architecture. Intelligent CAD II, H. Yoshikawa Editor. Amsterdam : Elsevier Science Publishers, 1990. p133-148.

[Delalandre05] Delalandre M., Analyse des Documents Graphiques : une Approche par Reconstruction d'Objets. Thèse de Doctorat en Informatique, Université de Rouen, France, 2005.

[Dengel89] Dengel A., Barth G., ANASTASIL: A Hybrid Knowledge-Based System for Document Layout Analysis. 11th International Joined Conference for Artificial Intelligence (IJCAI), Detroid, MI, 1989, p1249-1254.

[Denhartog96] DenHartog JE., TenKate TK., Gerbrands JJ., Knowledge based interpretation of utility maps. Computer Vision and Image Understanding. 1996, Vol. 63(1), p105-117.

[Desseilligny95] Pierrot-Desseilligny M., Le Men H., Stamon G., Characters string recognition on maps, a method for high level reconstruction. In : Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal (Canada), august 14-16, 1995. Vol. 1, p249-252.

[DiBaja94] DiBaja GS., Well-Shaped, Stable, and Reversible Skeletons from the (3,4)-Distance Transform. Journal of Visual Communication and Image Representation, Vol. 5(1), 1994, p107–115.

[DiBaja96] DiBaja GS. Representing Shape by Line Patterns. International Workshop on Syntatical and Structural Pattern Recognition, 1996. p230-239.

[Dimauro02] Dimauro G., Impedovo S., Modugno R., Pirlo G., Sarcinella L., Analysis of Stability in Hand-Written Dynamic Signatures, 8th International Workshop on Frontiers in Handwriting recognition (IWFHR'02), Ontario (Canada) , 2002, p259-263.

[Dori92] Dori, D., Dimensioning analysis: Towards automatic understanding of engineering drawings. Communication of the A.C.M., 1992, Vol. 35(10), p92-103.

[Dori95] Dori D., Tombre K., Form engineering drawings to 3D CAD models : are we ready now ? Computer Aided Design, 1995, Vol. 29(4), p243-254.

[Dori96] Dori D., Liu W., Vector-Based Segmentation of Text Connected to Graphics in Engineering Drawings, Advances in Structural and Syntactical Pattern Recognition, eds. P. Perner, P. Wang, and A. Rosenfeld, Lecture Notes in Computer Science, (LNCS 1121), Springer Verlag, 1996, p322-331.

[Dori99] Dori D., Sparse Pixel Vectorisation : An Algorithm and its Performance Evaluation. Pattern Analysis and Machine Intelligence (PAMI), vol. 21,(3), 1999, p202–215,

[Dosch00] Dosch P., Masini G., Tombre K., Improving Arc Detection in Graphics Recognition. International Conference Pattern Recognition, 2000. p2243-2246.

[Dosch03] Dosch P., Lladós J., Vectorial Signatures for Symbol Discrimination. International

Workshop on Graphic Recognition, 2003, p154-165.

[Dublincore98] DUBLIN-CORE. 1998. Disponible sur : <http://www.dublincore.org> (consulté le 26/05/2004).

[Dufresne01] Dufresne A. Conception d'une interface adaptée aux activités de l'éducation à distance : ExploraGraph. Sciences et Techniques Éducatives, 2001, Vol. 8(3-4), p301-320.

[Duperthuy97] Duperthuy C., Pyramide des contrastes et masques binaires 3x3 pour une esquisse initiale généralisée , Thèse de Doctorat de l'INSA de Lyon, 1997.

[Ecvision06] Groupe de recherche ECVision : http://www.ecvision.org/ (consulté le 10/05/2006)

[Edmonds81] Edmonds EA., Adaptive Man-Computer Interfaces. In: Computing skills in the user interface, Coobs M.J & Alty J.L Eds. Computer and People series, NewYork: Academic Press, 1981, p389-426.

[Eglin04] Eglin V., Bres S., Analysis and interpretation of visual saliency for document functional labelling. 2004. International Journal on Document Ananlysis and Recognition, Vol 7(1), p28-43.

[Eml00] EML. Educational Modelling Language, 2000 [en ligne] Disponible sur : http://eml.ou.nl (consulté le 27/05/04).

[Epeires05] http://www.epeires.org (consulté le 27/06/06).

[Eternad97] Etemad K, Doermann D., Chellappa R., Multiscale Segmentation of Unstructured Document Pages Using Soft Decision Integration, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19(1), January 1997, p92-96.

[Fan98] Fan KC., Chen DF., Wen MG., Skeletonization of Binary Images with Nonuniform Width via Block Decomposition and Contour Vector Matching. Pattern Recognition, Vol. 31(7), 1998, p823–838.

[Fletsher88] Fletsher LA., Kasturi RA., Robust algorithm for text string separation from mixed text/graphics images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, Vol. 10(6), p 910-918.

[Fonseca01] Fonseca MJ.,Jorge JA., Experimental Evaluation of an On-line Scribble Recognizer. Pattern Recognition Letters, Vol 22. 2001, p1311-1319

[Fonseca04] Fonseca M.J., Barroso B., Ribeiro P. & Jorge J.A. Retrieving Vector Graphics Using Sketches. In Symposium on Smart Graphics (SG), volume 3031 of Lecture Notes in Computer Science (LNCS) , 2004, p66–76.

[Fukada84] Fukada Y., A Primary Algorithm for the Understanding of Logic Circuit Diagrams. Pattern Recognition, Vol. 17(1), 1984, p125-134.

[Gagné06] Gagné C., Parizeau M., Genetic Engineering of Hierarchical Fuzzy Regional Representations for Handwritten Character Recognition, International Journal of Document Analysis and Recognition, A paraître 2006.

[Gao95] Gao J., Tang L. Liu W., Tang Z., Segmentation and Recognition of Dimension Texts in Engineering Drawings, International Conference on Document Analysis and Recognition, Montreal, Canada, 1995. p528-531.

[Garbay96] Garbay C., Interaction système/environnement pour l'interprétation des signaux et des images. Rapport de synthèse des travaux de l'action inter PRC 10.2 GDR PRC ISIS et CHM. 1996.

[Garbay98] Garbay C., Bellet F., Boucher A., Des agents situés pour l'interprétation de scènes, Revue d'Intelligence Artificielle, 1998, Vol 12(1), p11-36.

[Germond00] Germond L., Dojat M., Taylor, C., Garbay C., A cooperative framework for segmentation of MRI brain scans, Artificial Intelligence in Medicine, Vol 20(1), 2000, p.77-94.

[Gloger92] Gloger, JM., Use of the Hough Transform to separate merged Text/Graphics in Forms. In : Proceedings of the 11th International Conference on Pattern Recognition,

The Hague (Netherlands), august 30-september 3, 1992. Vol. 2, p268-271.

[Guillas05] Guillas S., Bertet K, and Ogier JM., Concept lattices: a tool for primitives selection ?. Sixth IAPR International Workshop on Graphics Recognition, Hong Kong, Chine, 2005.

[Hadjar01] Hadjar K, Hitz O, Ingold R., Newpaper page decomposition using Split and merge approach. 5th International Conference one Document Analysis and Recognition, 2001. p1186-1191.

[Hadjar02] Hadjar K, Hitz O., Robadey L, Ingold R. Configuration REcognition Model for Complex Transfers Methods Engineering: 2(CREM). Proceedings of the 5th International Workshop one Document Analysis Systems, 2002. p469-479.

[Han94] Han CC., Fahn KC., Skeleton Generation of Engineering Drawings via Contour Matching. Pattern Recognition, Vol. 27(2), 1994, p261–275.

[Hastie92] Hastie T., Kishon E., Clark M., Fan J., A Model for Signature Verification, Rapport technique AT&T Bell Laboratories, 1992.

[Hilaire02] Hilaire X., Tombre K., Improving the accuracy of skeleton based vectorization. Graphics recognition – algorithms and applications. In Lecture Notes in Computer Sciences. LNCS 2390. Springer Verlag. 2002, p273-288.

[Hori92] Hori O., Okazaki A., High Quality Vectorization Based on a Generic Object Model. In H. S. Baird, H. Bunke, and K. Yamamoto, editors, Structured Document Image Analysis, Springer-Verlag, Heidelberg, 1992. p325–339.

[Hori93] Hori O., Tanigawa S., Raster-to-Vector Conversion by Line Fitting Based on Contours and Skeletons. In Proceedings of 2nd International Conference on Document Analysis and Recognition, Tsukuba (Japan) , 1993, p353–358.

[Huang95] Huang YS., Suen CY., A method of combining multiple experts for the recognition of unconstrained handwritten numerals. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1995; Vol. 17(1). p90-94.

[Hudelot05] Hudelot C., Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms. Thèse de doctorat en informatique. Université de Nice. 2005. 283 pages.

[Ims00] IMS. Global Learning Consortium. 2000. Disponible sur : <http://imsproject.org> (consulté le 26/05/04).

[Jaekyu95] Jaekyu Ha, Haralick RM., Phillips IT., Recursive x-y cut using bounding boxes of connected components. 3rd International Conference on Document Analysis and Recognition, 1995, p952--955.

[Jain98] Jain AK., Yu B., Document representation and its application to page decomposition, IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. 20, 1998, p294–307.

[Janssen97] Janssen RDT., Vossepoel AM., Adaptive Vectorization of Line Drawing Images. Computer Vision and Image Understanding, Vol. 65(1) , Jan. 1997, p38–56.

[Jolion94a] Jolion JM, Analyse multi-résolution du contraste dans les images numériques, Traitement du Signal, Vol. 11(3), 1994, p245-255.

[Jolion94b] Jolion JM, Rosenfeld A., A Pyramid Framework for Early Vision, Kluwer, 1994.

[Jolion01] Jolion JM, Graph based Representations, Walter Kropatsch and Mario Vento eds., Proc. of GbR'01, 3rd IAPR Int. Workshop on Graph based Representations, Cuen Editor, Italy ISBN:88 7146 579-2. 2001.

[Joseph89] Joseph SH., Processing of engineering line drawings for automatic input to CAD. Pattern Recognition, 1989, Vol. 22(1), p1-11.

[Joseph92] Joseph SH., Pridmore TP., Knowledge directed interpretation of mechanical engineering drawings. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, Vol. 14(9), p928-940.

[Julia96] Julia L, Faure C, Un éditeur interactif de tableaux dessinés à main levée. Traitement du Signal, numéro spécial sur l'Ecrit et le Document, 1996, Vol. 12(6), p620-626.

[Kadonaga95] Kadonaga, T., Abe K., Comparison of methods for detecting corner points from digital curves. Lecture Notes in Computer Science (LNCS 1072). Graphics Recognition, Methods and Applications. R. Kasturi and K. Tombre Eds. 1995. p23-34.

[Kasturi90] Kasturi R., A system for interpretation of line drawings. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, Vol. 12(10), p978-991.

[Kasturi92] Kasturi R., An Overview of techniques for graphics recognition. In : Structured Document Analysis. Baird HS., Bunke H., Yamamoto K. Eds. Berlin (Allemagne), Springer Verlag, 1992. p285-324.

[Khaled98] Khaled A., Pettenati M., Rekik Y., Vanoirbeck C., Medit: structured document for web based teaching. Conference Internationale sur le Document Electronique. 15-17 Avril, Rabat, Maroc, 1998, p203-212.

[Khotazad90] Khotazad A., Hong YH., Invariant Image recognition by Zernike Moments, IEEE Transactions on Pattern Analysis and Machine Inteligence, Vol 12(5), May 1990. p.489 – 497.

[Kise98] Kise K., Sato A., Iwata M. Segmentation of page images using the area Voronoi diagram. Computer Vision and Image Understanding Special issue on document image understanding and retrieval. Vol. 70(3), 1998. p370–382.

[Koerich05] Koerich AL., Sabourin R., Suen CY., Recognition and verification of unconstrained handwritten words, IEEE Transaction on Pattern Ananysis and Machine Intelligence, October 2005, p1509-1522.

[Kolesnikov03] Kolesnikov A, Pasi F., Polygonal Approximation of Closed Contours. Scandinavian Conference on Image Analysis, (Halmstad) Sweden, Lecture Notes in Computer Science 2749, Springer-Verlag, 2003, p778-785.

[Lam92] Lam L., Lee SW.. Suen CY., Thinning Methodologies - A Comprehensive Survey. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 14(9), 1992, p869–887.

[Lam95] Lam L., Suen CY., An Evaluation of Parallel Thinning Algorithms for Character Recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 17(9) , 1995, p914–919.

[Lam97] Lam L., Huang YS., Suen CY., Combination of Multiple Classifier Decisions for Optical Character Recognition, Handbook of Character Recognition and Document Image Analysis, Ed. H. Bunke and P.S.P. Wang, World Scientific, 1997, p79-101.

[Lebourgeois99] Lebourgeois F., Emptoz H., Document Analysis in Gray level and typography extraction using Character Pattern redundancies, 5th International Conference on Document Analysis and Recognition., Bengalore India. 1999. p177-180.

[Lee00] Lee K., Choy Y., Cho S., Geometric structure analysis of document images: a knowledge-based approach, IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 22, 2000, p1224–1240.

[Lee04] Lee J., Yoon H.S., Soh J., Tae Chun B., Koo Chung Y., Using geometric extrema for segment-to-segment characteristics comparison in on-line signature verification, Pattern Recognition, Vol. 37, 2004, p93-103.

[LeMaitre00] Le Maître J., Charlet J., Garbay, C., Le document multimédia en Sciences du Traitement de l'Information, École thématique "Documents & Évolution" du Groupement De Recherche "Information, Interaction, Intelligence", Tome I, Cépaduès, 2000.

[Leplumey95] Leplumey I., CamillerapJ. and Queguiner C., Kalman filter contributions towards document segmentation. 3rd International Conference on Document Analysis and

Recognition, Montreal (Canada), august 14-16, 1995. Vol. 2, p765-769.

[Lim99] Lim T., Bo Y., Weihua H., Qian W., Zheng Z., Text/Graphics Separation using Agent-based Pyramid Operations. International Conference on Document Ananlysis Recognition, 1999: p169-172.

[Lin85] Lin, X., Shimotsuji, S., Minoh, M., Sakai, T., Efficient diagram understanding with characteristic pattern detection. Computer Vision, Graphics and Image Processing, 1985, Vol. 30, p84-106.

[Lindeberg93] Lindeberg T., Detecting Salient Blob-Like Image Structures and Their Scales With a Scale - Space Primal Sketch - A Method for Focus of Attention. International Journal of Computer Vision, December 1993, Vol. 11(3). p283-318.

[Llados01] Lladós J., Martí E., Villanueva J., Symbol Recognition by Error-Tolerant Subgraph Matching between Region Adjacency Graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23(10). 2001. p1137–1143

[Llados02] Liados J., Valveny E., Sanchez G., Marti E., Symbol recognition : current advances and perspectives. Lecture Notes in Computer Xciences. Springer Verlag. Vol. 2390. 2002, p104-127.

[Lom97] LOM. Learning Object Metadata IEEE 1484.12. 1997. Disponible sur : <http://ltsc.ieee.org/wg12/index.html> (consulté le 26/05/04).

[Long00] Long AC, Landay JA, Rowe LA, and Michiels J., Visual Similarity of Pen Gestures. Human Factors in Computer Systems. SIGCHI, 2000. p.360-367.

[Long99] Long AC, Landay JA, Rowe LA, Implications for a Gesture Design Tool. Human Factors in Computing Systems. CHI-ACM. Pittsburgh, PA, 1999, p.40-47.

[Lopresti01] Lopresti, D., Wilfong, G., Applications of Graph Probing to Web Document Analysis. Proceedings of the First International Workshop on Web Document Analysis, Seattle, WA, 2001, p51-54

[Lorette99] Lorette G., Handwriting Recognition or Reading? What is the situation at the Dawn of the 3rd Millenium?, International Journal on Document Analysis and Recognition, Vol. 2(1) , 1999. p2-12.

[Ltsc02] LTSC. Learning Technology Standards Committee. 2002. Disponible sur : <http://grouper.ieee.org/groups/ltsc/index.html> (consulté le 26/05/04)

[Madej91] Madej, D., An intelligent map to CAD conversion system. First International Conference on Document Analysis and Recognition, Saint-Malo (France), september 30-october 2, 1991. p603-610.

[Marinai05] Marinai S., Gori M., Giovanni S:, Artificial Neural Networks for Document Analysis and Recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol.27(1). 2005. p23-35.

[Marr82] Marr, D. Vision : A computional investigation into human representation and processing of visual information. New-York (USA) : W.H. Freeman and company, 1982. 397 pages.

[Marukatat01] Marukatat S., Artières T, Gallinari P, Dorizzi B. Sentence recognition through hybrid neuro-markovian modelling International Conference on Document Analysis and Recognition (ICDAR), Vol. I, 2001. p.731-735,

[Messmer95] Messmer B.T., Efficient Graph Matching Algorithms for Preprocessed Model Graphs. PhD thesis, Bern University, Switzerland, 1995.

[Meyer01] Meyer C., Un Environnement d'Apprentissage Fondé sur les Métaphores, les Hypermédias et les Cartes de Concepts. Application aux réseaux informatiques. Thèse de Doctorat d'Informatique, Ecole Centrale de Lyon, Décembre 2001, 260 pages.

[Michard98] Michard A., XML, Langage et Applications. Paris : Eyrolles, 1998, 376p., ISBN: 2-212-09206-7.

[Milgram93] Milgram M. Reconnaissance des formes - Méthodes numériques et connexionnistes. Edition Armand Colin, 1993.

[Mitchell01] Mitchell PE., Yan H., Newspaper Document Analysis Featuring Connected Line Segmentation. International Conference on Document Analysis and Recognition, 10-13 September 2001, Seattle, WA, USA, p1181-1185.

[Monagan93] Monagan G., Roosli M., Appropriate Base Representation Using a Run Graph. International Conference on Document Analysis and Recognition, Tsukuba, Japan, 1993, p623-626.

[Moran95] Moran TP, Chiu P, van Melle W, Kurtenbach G,. Implicit structures for pen-based systems within a freeform interaction paradigm. Conference on Computer Human Interaction. ACM, New York, 1995. p487-494.

[Mouchere06] Mouchère H., Anquetil E., and Ragot N., Writer Style Adaptation of On-line Handwriting Recognizers: A Fuzzy Mechanism Approach, International Journal of Pattern Recognition and Artificial Intelligence, special issue on IGS, World Scientific, To appear.

[Nagy84] Nagy G., Seth S., Hierarchical representation of optically scanned documents. 7th International Conference on Pattern Recognition (ICPR), 1984, p347-349.

[Nagy93] Nagy G., Seth S., Krishnamoorthy M., and Viswanathan M.. Syntactic Segmentation and Labeling of Digitized Pages from Technical Journals. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15(7), 1993. p737-747.

[Nagy95] Nagy, G., Document image analysis: what is missing ? In : Proceedings of the 8th International Conference on Image Analysis and Processing, San-Remo (Italy), 1995. p577-587.

[Nagy00] Nagy G., Twenty years of document image analysis in PAMI, IEEE Transaction on Pattern Analysis Machine Intelligence. Vol. 22. 2000, p38–62.

[Nanard98] Nanard J., Kahn P., Pushing reuse in hypermedia design: Golden rules, design patterns and generic templates". 8th Conference on Hypertexts HTX'98, ACM Press, Pittsburg, PS, June, 1998, p11-20.

[Ogier94] Ogier JM., Mullot R., Labiche J., Lecourtier Y., Intégration d'outils bas niveau dans une stratégie d'interprétation de documents. 9e Congrès Reconnaissance de Formes et Intelligence Artificielle, Paris (France), 11-14 janvier, 1994. Vol. 1, p533-544.

[Ogier00] Ogier, JM., Mullot, R., Labiche J., Lecourtier Y., Semantic coherency : the basis of an image interpretation device – application to the cadastral map interpretation. IEEE Transaction on Systems, Man and Cybernetics. 2000, Vol. 30(2). p237-244.

[OGorman93] O'Gorman L., The Document Spectrum for Page Layout Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15(11) , 1993. p1162-1173.

[OGorman95] O'Gorman L., Kasturi R., Document Image Analysis. IEEE Computer Society Press, Los Alamitos, CA, 1995.

[Paquet90] Paquet T., Vallee T., Lecourtier Y., Extraction de primitives par suivi de traits dans l'image binarisée d'un mot manuscrit. Colloque National sur l'Ecrit et le Document, Bigre. 1990, N° 68, p179-188.

[Paquet91] Paquet, T., Lecourtier Y., Reconnaissance de l'écriture manuscrite sur des chèques. In : Proceedings du 8e Congrès Reconnaissance des Formes et Intelligence Artificielle, Lyon-Villeurbanne (France), 25-29 novembre, 1991. Vol. 2, p687-693.

[Pasternak93] Pasternak, B., Neumann, B. ADIK : An adaptable drawing interpretation kernel. International Joint Conference on artificial Intelligence. Avignon, 1993. Vol. 1. p531-540.

[Pavlidis86] Pavlidis T., A vectorizer and feature extractor for document recognition. Computer Vision, Graphics, and Image Processing, Vol. 35, 1986. p111-127.

[Pavlidis91] Pavlidis T., Zhou J., Page Segmentation by White Streams, 10th International Conference on Pattern recognition, Saint-Malo, France, 1991, p945-953.

[Penoffice06] http://www.phatware.com/penoffice/ (consulté le 01/06/2005)

[Pikaz95] Pikaz A., Dinstein I., Optimal polygonal approximation of digital curves. Pattern Recognition, Vol. 28(3), 1995. p373-379.

[Plamandon88] Plamondon R., Parizeau M., Signature verification from position, velocity and acceleration signals : A comparative Study, 9th International Conference on Pattern Recognition (ICPR'88), Rome (Italie), Vol. I, 1988, p260-265.

[Plamandon99] Plamondon R., Lopresti DP., Schomaker LRB., Srihari R. On-line handwriting recognition. In: J.G. Webster (Ed.). Wiley Encyclopedia of Electrical & Electronics Engineering, New York: Wiley, ISBN 0-471-13946-7. p123-146.

[Plamandon00] Plamondon R., Srihari SN., On-line and Off-line Handwriting Recognition, IEEE Transactions on Pattern Analysis and Recognition, Vol. 22(1), , January, 2000, p63-84.

[Popel02] Popel DV., Compact Graph Model of Handwritten Images : Integration into Authentification and Recognition. In Conference on Structural and Syntactical Pattern Recognition (SSPR), LNCS 2396, 2002, p272–280.

[Prévot97] Prevot P. Les nouvelles technologies éducatives et leurs usages. Chapitre 18 de l'ouvrage "Connaissances et savoir-faire en entreprise. Intégration et capitalisation". JM Fouet, Edition Hermès, 1997, ISBN : 2-86601-627-0, p343-380.

[Prevost03] Prevost L., Michel-Sendis C., Moises A., Oudot L., Milgram M., Combining model-based and discriminative classifiers : application to handwritten character recognition 7th International Conference on Document Analysis and Recognition (ICDAR'03) 2003. Volume 1, p31-36.

[Ragot03] Ragot N., MÉLIDIS : Reconnaissance de formes par modélisation mixte intrinsèque/discriminante à base de systèmes d'inférence floue hiérarchisés, Thèse de l'Université de Rennes 1, Octobre 2003.

[Ramachandra80] Ramachandran K., Coding method for vector representation of engineering drawings. Proceedings of the IEEE, 1980, Vol. 68(7), p813-817.

[Rendek06] Rendek J., Lamiroy B., Tombre K., A Few Step Towards On-the-Fly Symbol Recognition with Relevance Feedback. 7th International Workshop on Document Analysis Systems, Nelson (New Zealand), Lecture Notes in Computer Science, LNCS 3872, Springer Verlag, february 2006, p604-615.

[Rennesson03] Rennesson-Debled J, Remy JL, Rouyer J, Segmentation of discrete curves into fuzzy segments, 9th International Workshop on Combinatorial Image Analysis, Electronic Notes in Discrete Mathematics, Vol. 12, Palermo, Italy, May 2003.

[Rhee01] Rhee TH., Cho SJ., Kim JH., On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries. International Conference on Document Analysis and Recognition (ICDAR'01) , 2001, p645-649.

[Roisin99] Roisin C., Documents multimédia structurés, Habilitation à diriger des recherches, INP Grenoble, septembre 1999.

[Roosli95] Röösli M., Monagan G., A high quality vectorization combining local quality measures and global constraints. Third international Conference on Document Analysis and Recognition, Vol. 1. IEEE Computer Society, Washington. 1995, p.243-248.

[Rosin89] Rosin PL., West GA., Segmentation of Edges into Lines and Arcs. Image and Vision Computing, Vol. 7(2) , May 1989, p109–114.

[Rosin95] Rosin PL, West GA., Nonparametric Segmentation of Curves into Various Representations. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. 17(12): 1995, p1140-1153.

[Rosin03] Rosin PL., Assessing the behaviour of polygonal approximation algorithms. Pattern Recognition Vol. 36(2): 2003, p505-518.•

[Rosin04] Rosin PL., Agent-based computer vision. Pattern Recognition Vol. 37(4), 2004. p627-629.

[Rubine91] Rubine D., Specifying gestures by example. Computer Graphics. Vol 25(4). 1991. p329-337.

[Saidali04] Saidali Y., Adam S., Ogier J.M., Trupin E., Labiche J., Knowledge Representation and Acquisition for Engineering Document Analysis. In Wokshop in Graphics Recognition (GREC), Lecture Notes in Computer Science (LNCS 3088), 2004, p25–36.

[Sanchez02] Sánchez G., Lladós J., Tombre K., An error-correction graph grammar to recognize textured symbols. In D. Blostein and Y.-B. Kwon, editors, Graphics recognition---Algorithms and applications, Lecture Notes in Computer Science (LNCS 2390), Springer Verlag, 2002. p128-138.

[Sauvola00] Sauvola J., Pietikainen M., Adaptive Document Image Binarisation. Pattern Recognition Vol. 33, 2000, p225-236.

[Schilit98] Schilit BN., Golovchinsky G., Price MN., Beyond paper: supporting active reading with free form digital ink annotations. SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, California, United States, April 18-23, 1998, p249-256.

[Schomaker93] Schomaker L., Using Stroke or Character-based Self-organizing Maps in the Recognition of On-line, Connected Cursive Script. Pattern Recognition, Vol. 26(3) , 1993, p443-450.

[Schomaker99] Schomaker L., Segers E., Finding features used in the human reading of cursive handwriting International Journal on Document Analysis and Recognition, Vol. 2, 1999, p13-18.

[Scorm98] ADL-SCORM. Advanced Distributed Learning - Sharable Content Object Reference Model, 1998. Disponible sur : <http://adlnet.org> (consulté le 26/05/04).

[Self88] Self J., Bypassing the intractable problem of student modelling. International Conference on Intelligent Tutoring Systems, Canada, 1988, p18-24.

[Seong93] Seong DS., Kim HS. Park KH., Incremental Clustering of Attributed Graphs. Transactions on Systems, Man and Cybernetics (TSMC), Vol. 23(5), 1993, p1399–1411.

[Shih89] Shih C. Kasturi R., Extraction of graphic primitives from images of paper based line drawings. Machine Vision and Applications, 1989, Vol. 2, p103-113.

[Shimotsuji94] Shimotsuji S. Hori O., Asano M., Robust drawing recognition based on model-guided segmentation. IAPR Workshop on Document Analysis Systems. Kasserslautern (Allemagne) 1994. p337-348.

[Simon97] Simon A., Pret JC., Johnson AP., A Fast Algorithm for Bottom-Up Document Layout Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19(3), 1997, p273-277.

[Smeulders00] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, 2000 , p1349-1380.

[Song00] Song J., Su F., Chen J., Cai S., A Knowledge-Aided Line Network Oriented Vectorisation Method for engineering Drawings. Pattern Analysis and Applications, 2000, Vlo. 3: p142–152.

[Song02a] Song J., Su F., Tai C., Cai S., An Object-Oriented Progressive- Simplification based vectorization System for Engineering Drawings : Model, Algorithm and Performance. Transaction on Pattern Analysis and Machine Intelligence (PAMI),

Vol. 24(8), 2002, p1048–1060.

[Song02b] Song J., Cai M., Lyu MR., Cai S., Graphics Recognition from Binary Images : One Step or Two Steps. International Conference on Pattern Recognition (ICPR), Vol. 3, p.135-138.

[SVC04] Signature Verification Competition (SVC) of International Conference on Biometric Authentication (ICBA 2004). http://www.cs.ust.hk/svc2004/

[Tabbone02] Tabbone A., Wendling L., Technical Symbols Recognition Using the Two-Dimensional Radon Transform. International Conference on Pattern Recognition (ICPR) Vol. 3, 2002: p200-203.

[Tabbone04] Tabbone A, Wendling L., Tombre K. Matching of graphical Symbols in line-drawing images using angular signature information. International Journal on Document Analysis and Recognition, 2003, Vol. 6(2), p115-125.

[Tabbone05] Tabbone A, Quelques contributions à la reconnaissance de formes dans les documents graphiques. Habilitation à diriger des recherches, Université Nancy 2, Decembre 2005.

[Tanigawa94] Tanigawa S., Hori, O., Shimotsuji S., Precise line detection from an engineering drawing using a figure fitting method based on contours and skeletons. 12th International Conference on Pattern Recognition, Jérusalem (Israël), 9-13 octobre, 1994. Vol. 2, p356-360.

[Tappert90] Tappart CC., Suen CY., Wakahara T., The state of the art in on-line handwriting recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, Vol. 12(8), p789–808.

[Tombre91] Tombre K., Vaxivière P., Structure, syntax et semantics in technical document recognition. First International Conference on Document Analysis and Recognition, Saint-Malo (France), september 30-october 2, 1991. Vol. 1, p61-69.

[Tombre92] Tombre, K., Technical drawing Recognition and understanding : From Pixel to semantics. IAPR Workshop on Machine Vision and Application. Tokyo (Japon), 7-9 december, 1992. p393-401.

[Tombre00a] Tombre K., Ah Soon C., Dosch P., Masini G. , Tabbone S., Stable and Robust Vectorization : How to Make the Right Choices. Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science (LNCS 1941), 2000, p3–18.

[Tombre00b] Tombre K., Tabbone A. Vectorization in Graphics Recognition : To Thin or not to Thin. International Conference on Pattern Recognition (ICPR) , 2000, Vol 2, p91–96.

[Tombre02] Tombre K., Tabbone S., Pélissier L., Lamiroy B, Dosch P., Text/Graphics Separation IAPR International Workshop on Document Analysis Systems, Princeton (NJ, USA), LNCS 2423, Springer-Verlag, Aug. 2002. p200–211.

[Tran02] Tran-Thuong T., Roisin C., Media content model for Authoring and Presenting Multimedia Document, Web Document Analysis: Challenges and Opportunities, World Scientific - Series in Machine Perception and Artificial Intelligence, 2002.

[Trier96] Trier O.D., Jain A.K., Taxt T., Feature extraction methods for character recognition - A survey, Pattern Recognition. Vol 29, 1996, p641-662.

[Tsujimoto92] Tsujimoto S., Asada H., Major Components of a Complete Text Reading System, Proceeding of IEEE, July 1992, Vol. 80(7), p1133-1149.

[Vapnik98] Vapnik VN., Statistical Learning Theory. New York, Wiley edition, 1998.

[Vaxivière94a] Vaxivière P., Une vectorisation par maillage du dessin et séparation en calques, adaptée aux plans mécaniques. 3e Colloque National sur l'Ecrit et le Document, Rouen (France), 6-8 juillet 1994. p205-214.

[Vaxivière94b] Vaxivière P., Tombre K., Knowledge organisation and interpretation process in engineering drawing interpretation. International Workshop on Document Analysis

System, Kaiserslautern (Allemagne), 1994. p313-321.

[Ventura94] Ventura AD., Schettini R., Graphic Symbol Recognition using a Signature Technique. 12th International Conference on Pattern Recognition, Jerusalem (Israel). Vol. 2, 1994, p533-535.

[Wall84] Wall K.,Danielsson P., A fast sequential method for polygonal approximation of digitized curves. Computer Vision, Graphics and Image Processing, 1984, Vol. 28, p220-221.

[Wang89] Wang D. Srihari SN., Classification of Newspaper Image Blocks Using Texture Analysis. Computer Vision, Graphics and Image Processing, 1989, Vol. 47, p327-352.

[Wang01] Wang Y., Phillips IT., Haralick R., Using Area Voronoi Tessellation to Segment Characters Connected to Graphics. 4th IAPR International Workshop on Graphics Recognition, Kingston, Ontario (Canada), 2001, p147-153.

[Wang06] Wang Y., Phillips IT., Haralick RM., Document zone content classification and its performance evaluation. Pattern Recognition, 2006. Vol 39. p57-73.

[Weindorf02] Weindorf M., Structure Based Interpretation of Unstructured Vector Maps. Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science (LNCS 2390), 2002, p190–199.

[Weiss95] Weiss M. Dori D., A scheme for 3D object reconstruction from dimensioned orthographic views. 3rd International Conference on Document Analysis and Recognition, Montreal (Canada), august 14-16, 1995. Vol. 1, p335-338.

[Wendling02] Wendling L, Tabbone A., Matsakis P., Fast and robust recognition of orbit and sinus drawings using histograms of forces. Pattern Recognition Letters, Vol. 23(14): 2002, p1687-1693.

[Wenyin99] Wenyin L., Dori D., From Raster to Vectors : Extracting Visual Information from Line Drawings. Pattern Analysis and Applications (PAA), 1999, Vol. 2(2), p10–21.

[Wong82] Wong, KY., Casey RG., Wahl FM., Document Analysis System, IBM Journal of Research and Development, Nov. 1982, Vol. 26(6), p647-656.

[Wu93] Wu W., Sakauchi M., A Multipurpose Drawing Understanding System with Flexible Object-Oriented Framework. International Conference on Document Analysis and Recognition (ICDAR) , 1993, p870–873.

[Wu97] Wu V., Manmatha R., Riseman EM., Finding Text in Images, Proc. of the Second ACM International Conference on Digital Libraries, Philadelphia, PA (USA), 1997, p23-26.

[XSLT02] XSL Transformations (XSLT) http://www.w3.org/Style/XSL/ (consulté le 01/06/04)

[Yu97] Yu Y., Samal A., Seth S., A System for Recognizing a Large Class of Engineering Drawing. Transaction on Pattern Analysis and Machine Intelligence (PAMI), 1997, Vol. 19(8), p868–890.

[Zhang05] Zhang D, Lu G., Study and evaluation of different Fourier methods for image retrieval. Image, Vision and Computer. 2005, 23(1), p33-49.

[Zhou03] Zhou XS, Huang TS., Relevance feedback in image retrieval: A comprehensive review. Multimedia System, Vol. 8(6) , 2003, p536--544.

ANNEXE

Sélection d'articles représentatifs de mes activités de recherche

[Ramel06a] J.Y. Ramel, S. Leriche, ML Demonet, S. Busson. User-driven Page Layout Analysis of historical printed Books. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition Accepté, à paraitre 2006.

[Qureshi06a] Qureshi RJ., Ramel JY, Cardot H., Graphic Symbols Recognition Using Flexible Matching of Attributed Graphs. The Sixth IASTED International Conference VIIP 2006. August 28-30, 2006 Palma de Mallorca, Spain. A paraître août 2006.

[Hocquet06c] Hocquet S., Ramel JY, Cardot H., User specific parameters in one-class problems. International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong. p449-452.

[Journet06] Journet N., Eglin V., Ramel J.Y., Mullot R., Dedicated texture based tools for characterisation of old books. 2nd IEEE International Conference on Document Image Analysis for Libraries. Lyon, France, April 27-28, 2006. p60-70.

[Wirotius04a] Wirotius M., Ramel JY., Vincent N., Selection of Points for On-Line Signature Comparison. IWFHR. Tokyo (Japon). p503-508. Octobre 2004.

[Benadi04] Benadi S., Ramel JY, Prévot P.,AHXEL: Adaptive Hyperdocument using XML for E-Learning. 5th IEEE Int. Conf. on Information Technology Based Higher Education and Training: ITHET'04, Istambul. Turquie. 31 Mai - 2 Juin, 2004. CD-ROM

[Ramel00a] J.Y. Ramel, N. Vincent, H. Emptoz. A structural representation for understanding line drawing images. International Journal on Document Analysis and Recognition. Special issue on Graphics Recognition, Volume 3 n°2 - Décembre 2000. p58-66.

Propositions pour la représentation et l'analyse de documents numériques

Résumé : Les travaux présentés dans ce document concernent à la fois l'analyse et la reconnaissance d'images de documents et les interfaces homme-document multimédia. Les documents numériques peuvent être vus, dans les deux cas, comme un assemblage d'objets élémentaires sur un support physique (niveau support). A un niveau légèrement plus élevé, il est également possible de s'intéresser à l'enchaînement et aux liens existants entre les différents éléments de contenus d'un document (niveau structure). Enfin, le sens ou le message véhiculé par le document (niveau sémantique) ne doit pas être négligé lorsque l'on cherche à analyser son contenu qu'il s'agisse d'une image ou d'une interface. Trois problématiques principales sont développées en partant de cette vision commune. La première traite de la mise en place de représentations abstraites permettant de décrire le contenu des images de documents et des tracés en ligne. Différents algorithmes de construction automatique de ces représentations abstraites à partir des données directement fournies par les dispositifs d'acquisition en hors ligne (images) et en ligne (suite de points horodatés) sont présentés. Ensuite, différentes architectures d'analyse et méthodes de reconnaissance utilisables sur ces représentations sont proposés pour effectuer l'interprétation du contenu des documents. Enfin, la structuration en trois niveaux des documents numériques est à nouveau exploitée pour produire des interfaces adaptatives de façon à mieux pouvoir formater et présenter les informations demandées par les utilisateurs notamment lorsque ces derniers utilisent des hyperdocuments dans le cadre du télé-enseignement. A chaque fois, une attention particulière est portée sur la modélisation des données et des traitements de manière à obtenir des environnements suffisamment génériques.

Mots clés : Analyse d'images de document, Traitement d'images, Reconnaissance de formes, Interaction Homme-Machine, Télé-enseignement.

Propositions for the representation and the analysis of electronic documents

Abstract: The research works presented in this document deal simultaneously with document images analysis and human-computer interaction. In both subjects, electronic documents could be seen as collection of elementary objects gathered on a physical support (support level). At a higher level, it is also possible to be interested in the relations and in the links between the different elements of content constituting the documents (structure level). Finally, the sense or the message carried by the document should not be ignored when we try to analyse its content regarding as well the images as the interfaces (semantic level). Three main problematics are developed by using this three level structure. The first one concerns the construction of representations for the description of the content of the document images and of on-line strokes. Different algorithms for the automatic construction of these representations starting with the data provided by the acquisition devices (images for off-line documents and set of ordered points for on-line documents) are discussed. Next, the analysis strategies and recognition methods available to work on such representations in order to achieve the interpretation of the contents of the documents are presented. Finally, we explain how the three level structure of the electronic documents can be exploited in order to generate adaptive interfaces. The objective is to format and display the information needed by the end users according to their profile in particular when they are using hyperdocuments during e-learning sessions. In all the cases a particular attention is turned to the modelling of the data and processing to obtain generic frameworks.

Keywords : Document image analysis, Image processing, Pattern recognition, Human-Machine Interaction, E-learning.