

La reconnaissance d'images de documents : Un panorama

Document images recognition: A survey

Éric Trupin

Université de Rouen Laboratoire, PSI Place Emile Blondel, 76821 Mont-Saint-Aignan Cedex
Tél: 02 35 14 67 65 Fax: 02 35 14 66 18
Email: Eric.Trupin@univ-rouen.fr

Manuscrit reçu le 19 mai 2004

Résumé et mots clés

Cet article propose de dresser un panorama des outils et techniques mis en place pour la reconnaissance d'images de documents. Depuis les pré-traitements jusqu'à l'extraction de la structure logique, il balaie les différents moyens de reconnaissance et de modélisation des contenus pour ensuite discuter des connaissances exploitables pour construire un système d'interprétation permettant d'extraire tout ou partie des informations contenues dans l'image, sachant que ces informations n'ont de sens que dans un contexte bien défini et suivant une intention exprimée.

Interprétation de documents, modèle, connaissance, utilisateurs.

Abstract and key words

This article states of the art about tools and techniques for images documents analysis and recognition. From common used processings to the logical layout extraction, it surveys the different means employed for modelling and recognition. Then it discusses about the knowledge to be exploited for designing an understanding system able to extract all, or parts of, information from documents images. This information only makes sense in a well-defined context and according to an expressed intention.

Document understanding, modelling, knowledge, users.

1. Introduction générale

L'Analyse d'Image de Documents est un thème de recherche du domaine du traitement d'images numériques qui avait pour objectif principal de convertir des images de documents en vue de la modification, l'archivage, la recherche, la réutilisation et la transmission de l'information que ces images contiennent. Cela fait près de quarante ans que ce thème se développe depuis l'apparition des premiers systèmes de lecture optique de caractères¹. Notons que des brevets avaient déjà été proposés au 19^{ème} siècle pour concevoir des aides à la lecture pour les non-voyants ou des entrées automatiques pour le télégraphe [Nagy-2]. Les premiers systèmes de lecture des adresses postales furent installés en 1965. La lecture automatique de l'écriture manuscrite et contrainte dans les formulaires commença à fonctionner dans les années 80. Ceci fût fortement conditionné par les progrès importants effectués en matière de possibilité de traitement par ordinateur, la progression spectaculaire des capacités de stockage de documents numériques et la réduction de leur coût. Mais au lieu de permettre de vider les classeurs et les étagères de leur contenu papier pour le passage à un monde sans papier, on peut constater aujourd'hui que si les documents électroniques abondent déjà, il y a de plus en plus de documents papier.

Indépendamment de la manière d'utiliser les images de documents, nous pouvons classer ces images en fonction de leur contenu et ainsi les répartir en deux catégories : naturelle ou symbolique. Les portraits, les empreintes digitales, les photographies aériennes, les images satellites et les radiographies décrivent des objets ou des scènes naturelles. Les adresses postales, les articles scientifiques, les journaux et magazines, les formulaires administratifs ou de vente par correspondance, les partitions musicales, les dessins techniques, les plans et les cartes géographiques représentent des objets symboliques. Ainsi Nagy [Nagy-2] propose la définition suivante : «L'Analyse d'Images de Documents est une théorie et une pratique de reconstruction de la structure symbolique des images numériques directement produites par l'ordinateur ou simplement numérisées à partir du papier». Il est bien entendu possible d'étendre cette définition à tout enregistrement multimédia mais nous choisissons dans cet article² de circonscrire la notion de document à des objets créés exclusivement pour transporter des informations codés symboliquement.

Traditionnellement, on distingue les documents à forte teneur textuelle de ceux étant essentiellement composés de graphiques³. Les documents structurés sont composés de colonnes, de paragraphes, de lignes de texte, de mots, de caractères, de figures et de schémas. Un logiciel de lecture optique de caractères

convertit alors les images de mots et de caractères en code ASCII ou Unicode. Mais cette transformation du document en une simple chaîne de symboles n'est pas suffisante car le document transporte d'autres informations, souvent essentielles, provenant de l'usage de conventions de présentation des documents, comme le choix et l'enrichissement des fontes utilisées, l'arrangement bi-dimensionnel des tableaux ou la présentation des formules mathématiques. Pour complètement analyser le document, il s'agit donc d'extraire et d'interpréter tous ces codages⁴ subtiles qui dépendent de l'usage du document (une communauté d'intérêt, un métier,...). C'est l'objet de l'Interprétation⁵, ou la Compréhension, d'Image de Documents qui est un processus différent de l'Analyse d'Image de Documents selon [Tang Y.Y.]⁶. Alors que celle-ci extrait la structure physique d'une image de document, la Compréhension d'Image de Document fait correspondre cette structure physique à une structure logique en considérant les relations entre les objets dans le document spécifique et ce en utilisant la connaissance du format de présentation du document. Ceci s'effectue en considérant un modèle ou une structure générique de document qui permet d'étiqueter logiquement voire même sémantiquement ce que le document contient.

Les dessins techniques, les cartes géographiques, les plans, les partitions musicales, les diagrammes, les schémas électroniques, les organigrammes sont des exemples de documents graphiques. Ils sont segmentés en entités élémentaires comme les traits, les courbes, les points de rebroussement ou de croisement pour que des objets de plus haut niveau, comme les symboles, puissent ensuite être interprétés. Les cartes nécessitent de plus une séparation des couleurs et une reconstruction des liens entre les symboles utilisés et les toponymes associés. Ainsi, comme pour les documents structurés, une connaissance spécifique du domaine d'usage du document est essentielle pour une interprétation correcte du contenu.

C'est pourquoi différentes approches dirigées par les modèles ont été développées. Les modèles ont par exemple été utilisés pour l'interprétation de formules mathématiques [Kosmala], [Zannibi-1], des formulaires [Couasnon], des factures [Cesarini-2], des tableaux [Watanabe], des schémas mécaniques [Ablameyko-1], des logos [Chang], des partitions musicales [Armand], des notices bibliographiques [Parmentier], des tables des matières [Belaid-2], des documents juridiques [Ingold-1]. Ces modèles de connaissance doivent cependant offrir une représentation de la connaissance permettant d'orienter l'Analyse et la Compréhension d'image de Documents car les résultats dépendent de l'application visée et des connaissances spécifiques au type, ou à la classe de documents [Ingold-2].

1. En anglais OCR pour Optical Character Recognition.

2. Cet article traite des documents de contenu symbolique.

3. Dans cet article, nous distinguerons les documents structurés des documents techniques même si ces derniers respectent une structuration certaine.

4. Cette information sur l'information est aujourd'hui appelée *Metadata* ou *Métadonnée*, mais ce terme peut revêtir différentes significations selon le domaine concerné : les bibliothèques, la recherche sur Internet, la programmation...

5. Understanding.

6. Nous adopterons cette position.

Ces connaissances sont soit descriptives, soit opératoires au sens de [Newell] et [Adam-3], et elles doivent être correctement décrites. Les premières fournissent un modèle de contenu du document en décrivant les différentes entités manipulées et les liens entre elles quelque soit leur type (physique ou logique). Les secondes décrivent les entrées, les sorties et les paramètres des algorithmes utilisés, ainsi que les lois de réglage de ces paramètres en fonction d'un contexte et de contraintes données. L'acquisition de ces connaissances par le système d'interprétation fait intervenir un spécialiste par domaine d'usage (un bibliothécaire pour [Parmentier] et un juriste pour [Ingold-1]) et un spécialiste du traitement d'images dans tous les cas. A ces deux acteurs, s'ajoute bien sûr l'utilisateur du système. Nous nous plaçons donc dans une logique d'interaction des acteurs avec le système depuis sa conception jusqu'à son utilisation, qui nécessite la définition d'Interfaces Homme-Machine entre ces acteurs et le système.

Nous proposons dans cet article de faire une synthèse des différents travaux relevant de l'analyse et de la compréhension de documents, ce que nous appelons plus simplement la reconnaissance d'images de document. Nous nous intéresserons à la fois aux documents à dominante textuelle et aux documents, ou parties de documents, graphiques. Dans la section 2, nous présenterons les outils, les méthodes et des systèmes de la littérature liés à la reconnaissance d'images de document. La section 3 présentera, en relation avec les modèles de documents décrits dans le chapitre 2, les connaissances dont un système d'interprétation de documents devrait être doté. La section 4 conclura.

2. La reconnaissance d'images de documents

2.1. Introduction

L'architecture d'un système de reconnaissance d'images de documents est décrite schématiquement par la figure 1. Elle décompose le traitement global du document en différentes étapes qui elles-mêmes peuvent se décomposer en sous-étapes. Elle montre que le document est composé de régions homogènes telles que du texte, du graphique ou des formules. On pourrait ajouter d'autres types de régions comme des tableaux, des photographies... Chacune d'entre-elles est soumise à un traitement spécialisé dépendant de son contenu. Toutes ces régions s'organisent de manière hiérarchique pour composer une structure. Cette figure s'applique de manière générale aux documents structurés mais aussi aux documents graphiques qui ne sont composés que de régions graphiques.

Cependant on peut noter quelques différences entre la reconnaissance d'images de documents structurés et techniques

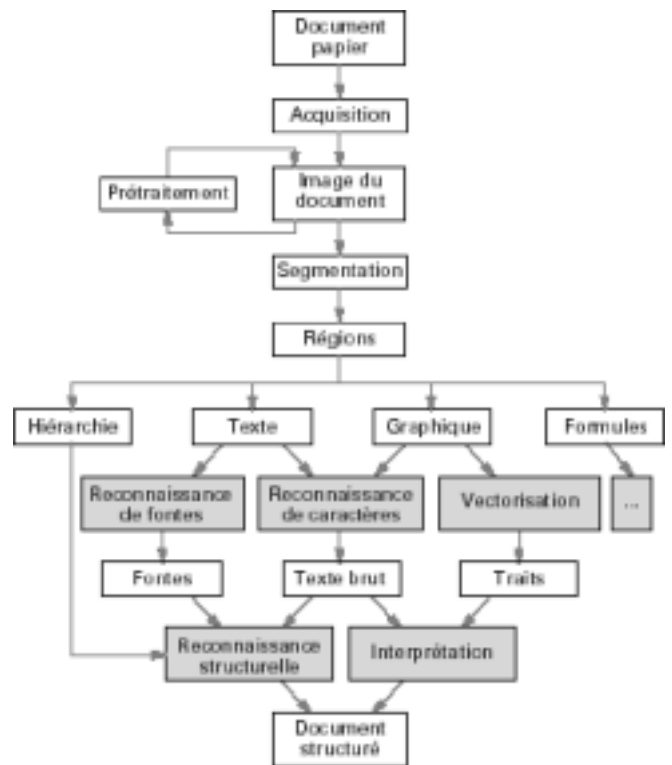


Figure 1. Architecture d'un système de reconnaissance de documents (tiré de [Ingold-2]).

comme le montre le tableau 1 qui distingue cinq niveaux de traitement selon les objets manipulés : les pixels, les primitives, les structures, le document et le corpus ou la collection de documents. Dans la suite de cette section, nous tentons d'illustrer des parties importantes de ces cinq niveaux.

2.2. Acquisition des images de documents

L'acquisition des images numériques de documents s'effectue généralement à l'aide d'un scanner, périphérique devenu très courant depuis l'explosion de la bureautique dans les années 80. La qualité des images produites dépend du contenu des documents. Ainsi, l'ajustement particulier du contraste, le choix d'utiliser des images binaires s'appliquent souvent à des images contenant des symboles graphiques ou du texte. Les photographies doivent être produites en niveau de gris. La couleur peut également être utilisée pour des images de magazines par exemple. Mais tous les pré-traitements de ces images pour les rendre interprétables automatiquement sont généralement indépendants du type de document et tous ceux procédant au filtrage du bruit, à la binarisation, à la détection de trait et à la segmentation, peuvent être appliqués de la même manière, qu'il s'agisse de textes imprimés ou manuscrits, de dessins en « fil de fer », de dessins techniques ou de cartes. Les images en niveau de gris ou couleur nécessitent des traitements plus spécialisés.

Tableau 1. Niveaux de Traitement selon la prédominance de contenu du document (tiré de [Nagy-2]).

Niveau du traitement	Document structuré	Document graphique
Pixels	Prétraitement Représentation Réduction de bruit Binarisation Détection de l'inclinaison Segmentation de zones informatives Segmentation de caractères Reconnaissance de la fonte et de la langue	Prétraitement Représentation Réduction de bruit Binarisation Squelettisation Vectorisation
Primitives	Reconnaissance de glyphe Composantes connexes Lignes, ponctuation et diacritiques Mots	Reconnaissance de primitives Segments de droites et de courbes Jonctions et nœuds Caractères
Structures	Reconnaissance de texte Segmentation en mots Reconstruction des lignes de texte Analyse de tableau Contexte morphologique Contexte lexical Syntaxe, sémantique	Reconnaissance de structure Champs texte Légendes Attribution d'étiquette Dimensions Symboles graphiques Caractéristiques de surface et de texture
Documents	Analyse de la structure de page Séparation texte/graphique Analyse des composantes physiques Analyse des composantes logiques Composantes fonctionnelles (étiquetage) Compression	Interprétation Reconnaissance d'objet Analyse de la connexité Séparation en couches CAO/SIG Extraction des attributs de base de données Compression
Corpus	Recherche d'Information Classification de documents et indexation Recherche Sécurité, authentification	Base de données, CAO, SIG Validation Recherche Mise à jour

Lorsque l'image du document est segmentée en composantes constitutives, d'autres traitements sont nécessaires.

2.3. Taxonomie des documents

Outre le fait que les documents puissent être à prédominance textuelle ou graphique (structuré versus graphique), il existe une grande diversité de documents. Celle-ci est illustrée par la figure 2.

Les documents structurés se distinguent également entre eux par les différentes langues et écritures utilisées. L'écriture latine comporte un alphabet restreint et les caractères sont généralement bien isolés. L'écriture arabe (figure 5) comporte de nombreuses ligatures compliquant la localisation des caractères et il

existe une variété de forme des caractères selon leur place dans le mot. Quant aux écritures asiatiques (figures 3, 4), elles sont généralement constituées de milliers de pictogrammes complexifiant d'autant les structures des textes. Il existe également de nombreuses fontes dactylographiées ou imprimées⁷ qui élargissent la variété typographique des documents. Notons également que les documents structurés peuvent comporter des zones graphiques telles que des logos ou des illustrations. Dans ce cas, il s'agit de séparer le texte de ces graphiques qui seront traités selon les techniques propres aux documents graphiques (cf. section 2.9). On remarque également sur cette figure la diversité des structures rencontrées: texte plein, journal, article scientifique, annuaire, table des matières ou formulaire... Toutes ces

7. Chacun peut créer sa fonte pour personnaliser ses documents.

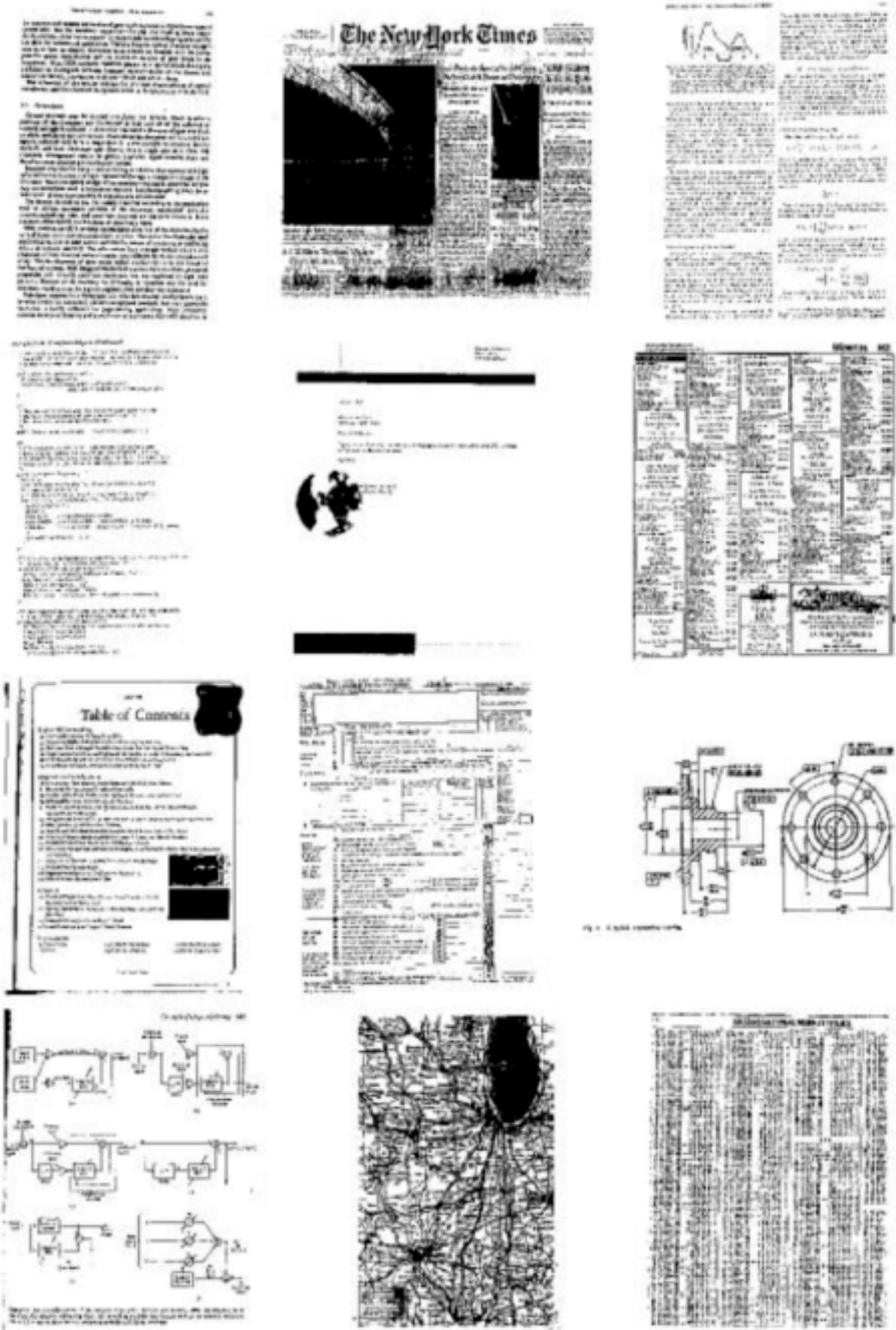


Figure 2. Taxonomie des documents (tiré de [Nagy-2]).

catégories sont caractéristiques et impliquent un schéma de traitement différent dépendant du contenu présenté. Parmi les documents graphiques, on en rencontre différentes catégories se distinguant par les éléments rencontrés et les conventions associées à leur représentation : les dessins techniques, les schémas électriques, les plans de cadastre, les plans d'architecte, les cartes... Ils comportent aussi un ensemble de données textuelles, comme des légendes ou des cotes, dépendantes du domaine pour leur association avec les données graphiques.

On constate donc que les documents sont très variés dans leur présentation allant du simple document textuel, tel un roman, qui suit une structure linéaire jusqu'à des documents très complexes tels qu'une revue technique automobile, par exemple, qui comportent des informations textuelles, divers schémas mécaniques et électriques, et présentent même des liens texte/graphique au sein du document lui-même.

2.4. Pré-traitement

Les pré-traitements consistent en des transformations d'images pour faciliter l'extraction d'information à reconnaître dans le document. La table 1 présente les différents pré-traitements rencontrés et nous présentons les principales opérations que sont le redressement, le filtrage et la binarisation.

2.4.1. Le redressement

Beaucoup d'algorithmes ont été développés pour détecter l'inclinaison globale du document et opérer un redressement de son image. En effet, lors de l'étape d'acquisition au moyen d'un scanner, il est fréquent que le document à saisir ne soit pas parfaitement disposé et il en résulte que l'image obtenue ne présente pas des lignes parfaitement horizontales ou verticales. Il convient de détecter cette inclinaison et alors de procéder au redressement de l'image au moyen d'une transformation géométrique⁸. Cependant cette opération n'a pas toujours de sens dans le cas d'un document technique. Différentes techniques classiques [Kasturi & O'Gorman] reposent sur l'utilisation de la Transformée de Hough [Srihari-1] [Illingworth], des profils de projection [Akiyama] [Baird-1] ou des plus proches voisins [Hashizume], mais nous pouvons aussi citer les travaux de [Aghajan] assimilant l'ensemble des lignes de texte à la propagation d'un front d'ondes et de [Chaudhuri] sur les textes indiens d'écriture Devanagari ou Bangla.

2.4.2. La binarisation

La binarisation permet de convertir l'image numérique en niveau de gris issue du scanner en une image bi-modale dans

8. Il est à noter que cette transformation peut générer des distorsions du fait de la nature matricielle du tableau de pixels et généralement une correction d'inclinaison de seulement quelques degrés est acceptable pour tenter par la suite de reconnaître le contenu du document.

laquelle la forme, assimilée à l'information utile, et le fond sont séparés. Lorsque les documents papier sont de bonne qualité et de fond blanc, un simple seuillage global suffit après analyse de l'histogramme des niveaux de gris [Otsu]. Mais lorsque le fond est texturé ou si le document est dégradé (pliures, taches ...), une analyse plus fine est nécessaire. C'est ainsi qu'une grande variété de méthode est proposée. Si la dimension des caractères ou l'épaisseur des traits sont *a priori* connus, des approches de type contour peuvent être employées [Trier-1] [White] et d'autres permettent de sélectionner un modèle de ligne [Jang]. Certaines approches s'appliquent à calculer des seuils locaux et à réaliser une binarisation adaptative [Sauvola] [Yang Y.] et d'autres sont des variantes pour la modélisation du bruit [Don] ou l'utilisation d'ondelettes [Tan C.L.]. Pour des comparaisons complètes des méthodes de binarisation, on peut se référer aux synthèses présentées dans [Sahoo] ou [O'Gorman-3] pour les techniques globales ou multi-seuils et dans [Trier-2] ou [Kamel] pour les techniques adaptatives.

2.4.3. Le filtrage

Le filtrage a pour effet de réduire le bruit de l'image afin de rendre plus robuste l'étape de reconnaissance. Des traitements morphologiques généralement employés en traitement d'images [Haralick-1] peuvent spécifiquement être utilisés pour les documents. D'autres techniques sont basées sur l'utilisation d'opérateur de convolution ou opèrent des filtrages locaux. On pourra se reporter à [O'Gorman-1] et [Haralick-2].

2.5. La segmentation

La segmentation a pour but de découper l'image du document en régions ou blocs homogènes (texte, graphique, photographies...) puis en lignes de texte et en caractères à l'intérieur de ceux-ci. Ce processus permet d'extraire la structure physique du document qui peut être représentée de diverses manières, indépendamment ou conjointement à la structure logique du document. Des paramètres de présentation du document ont été utilisés dans [Kreich], [Niyogi], [Saitoh] [Lin C.C.] et [Ishitani]. Ces paramètres correspondent aux dimensions et aux distances entre les objets contenus dans le document tels que les caractères, les mots, les lignes ou les régions. Bien que cette méthode de représentation fournisse une information utile, elle ne reflète pas complètement les relations spatiales entre les composantes physiques du document. La structure physique du document peut être mieux représentée par un arbre hiérarchique dérivé d'un ensemble de règles [Tsujiimoto-1] [Yamashita] [Fisher]. Une telle représentation décrit les relations spatiales entre les composantes physiques du document de manière hiérarchique. L'inconvénient des représentations à base de règles réside dans le fait que les règles peuvent devenir arbitraires. D'autre représentations basées sur des grammaires formelles offrent l'avantage de limiter les types de production qui peuvent être utilisés. Dans ces représentations, un document est considéré comme

une séquence ou une chaîne de caractéristiques des composantes physiques.

2.5.1. La segmentation en blocs

Les différentes approches de décomposition en régions sont soit descendantes⁹, soit montantes¹⁰ voire hybrides. Les premières divisent de manière itérative l'image du document en régions de plus en plus petites, et stoppent ce processus lorsqu'un critère d'arrêt est vérifié. Les approches montantes démarrent au niveau pixel de l'image en regroupant les pixels en composantes connexes, telles que des caractères, qui sont ensuite regroupées pour former des mots, des lignes ou des régions. Les approches hybrides peuvent être vues comme un mélange des deux approches précédentes. L'algorithme Docstrum de [O'Gorman-2], celui de [Kise] basé sur les diagrammes de Voronoi, le RLSA¹¹ de [Wahl], l'algorithme de [Jain-2] et l'algorithme de [Fletcher] sont typiques des approches montantes. L'algorithme de découpe récursive en X-Y de [Nagy-1] et celui de [Baird-2] sont des approches descendantes. [Pavlidis-1] a proposé une approche hybride utilisant une stratégie de type « split and merge ». Un état de l'art sur les algorithmes de segmentation de page peut être trouvé dans [Kasturi & O'Gorman] et dans [Jain-2].

La plupart des algorithmes mentionnés ci-dessus ne créent pas de description hiérarchique ou ne permettent pas à l'utilisateur de renseigner la structure du document. De plus, ils ne possèdent pas de méthodes d'estimation des paramètres de l'algorithme à partir de données de « vérité terrain¹² ». D'autre part, certains travaux décrivent des algorithmes d'analyse de la structure physique des documents utilisant des méthodes à base de grammaire. [Kopec] propose par exemple un algorithme de segmentation d'une colonne de texte qui est modélisée en utilisant une grammaire régulière stochastique. Cependant la page est supposée avoir été préalablement segmentée en colonnes et il n'y a aucune procédure d'estimation des paramètres pour le modèle. [Tokuyasu] propose une approche de segmentation de page basée sur des grammaires régulières pour décrire la structure du document en termes de rectangles obtenus par subdivisions verticales et horizontales de l'image. [Krishnamoorthy] décrit également un algorithme de segmentation hiérarchique d'une page de document qui construit un arbre dans lequel chaque nœud représente un rectangle.

Une étude comparative de cinq algorithmes d'analyse de la structure physique de document peut être trouvée dans [Mao-1]. De même, [Liang] propose une métrique pour analyser la performance de tels algorithmes. Il présente notamment une méthode d'estimation de paramètres optimaux pour ces algorithmes. [Yanikoglu] décrit un environnement *Pink Panther* permettant d'évaluer la segmentation de page de document.

9. top-down.

10. bottom-up.

11. Run Length Smearing Algorithm ou Run Length Smoothing Algorithm.

12. ou fonds de vérité terrain (*groundtruth* en anglais).

2.5.3. La segmentation en caractères

La segmentation en caractères consiste à extraire chaque caractère des mots et des phrases. C'est une étape cruciale si l'on désire de bons résultats de reconnaissance. L'approche classique consiste à construire un histogramme de projection verticale pour chaque ligne de texte et à déterminer ainsi la position des caractères. Cependant c'est une approche assez sensible au bruit et surtout à l'inclinaison des caractères italiques ou lorsque les caractères sont collés ou dégradés. Généralement, un seuillage, parfois adaptatif, d'histogramme est effectué pour s'affranchir de ces problèmes. D'autre part, l'étape de reconnaissance peut provoquer une nouvelle segmentation des caractères si les résultats ne sont pas satisfaisants. On illustre ainsi le paradoxe segmentation/reconnaissance pour lequel on ne peut reconnaître une forme sans l'avoir segmentée et inversement. [Tsujimoto-2] présente en détail cette étape de segmentation en caractères et [Casey] survole toutes les approches proposées depuis 1959 pour segmenter les caractères collés ou fragmentés.

2.6. La reconnaissance du type d'écriture, de la langue et des fontes

Le fait de préalablement identifier le type d'écriture (latine, orientale ...) utilisé dans le document permet de réduire la variété des symboles à reconnaître. D'autre part, la reconnaissance de la langue (anglais, français, allemand ...) utilisée permet de s'adapter au contexte afin de pouvoir utiliser les modèles adéquats. Finalement, reconnaître la fonte utilisée réduit également la diversité de représentation graphique d'un même symbole et permet d'adapter généralement le système en vue d'une reconnaissance mono-fonte, ou permet de sélectionner le logiciel de reconnaissance de caractères adapté. [Hochberg] parvient à classifier les types d'écriture en utilisant des modèles à partir des symboles textuels (ou composantes connexes) apparaissant fréquemment dans les documents. [Spitz] sépare d'abord les écritures orientales des écritures latines selon la distribution verticale des concavités par rapport à la ligne de base, puis distingue différentes langues orientales par analyse d'histogrammes de densité de pixels et distingue les langues latines selon la distribution des hampes et des jambages dans les mots. [Tan T.N.] utilise les fonctions de Gabor. Quant aux fontes, elles peuvent se décliner en un nombre quasiment infini rendant le problème de leur reconnaissance très complexe. [Zramdini] obtient jusqu'à 97 % de bonne identification sur la base de 340 fontes différentes en analysant des profils de projection verticale et la distribution des composantes connexes dans les mots.

2.7. La reconnaissance des caractères

Nombreux ont été les travaux consacrés à la reconnaissance des caractères imprimés, dactylographiés ou manuscrits isolés. Selon la figure 1, cette étape apparaît comme une étape ulté-

rieure à la segmentation en blocs. Or ceci n'est dû qu'à la chronologie de la recherche dans le domaine où initialement les chercheurs travaillant sur la décomposition de page ne disposaient pas de logiciels de reconnaissance de caractères. Mais ceux travaillant sur l'OCR ne disposaient pas de la décomposition automatique de page. Or il s'avère que la reconnaissance de caractères peut précéder, suivre ou être co-occurrence à la décomposition de page [Nagy-2].

Les algorithmes de reconnaissance de caractères fonctionnent selon deux étapes successives : l'extraction de caractéristiques et la classification. Ils font correspondre l'imagette d'un caractère à une classe donnée en utilisant un algorithme de classification basé sur des caractéristiques extraites de l'imagette et sur les relations entre elles. Cependant il existe souvent une troisième étape qui permet de corriger des erreurs de reconnaissance en utilisant des connaissances contextuelles.

L'extraction de caractéristiques consiste à analyser un certain nombre d'attributs prédéfinis dont la mesure peut être imparfaite. En effet, elles sont obtenues par des mesures métriques ou topologiques qui sont très sensibles au bruit. On impose à ces caractéristiques de permettre de distinguer des formes proches mais différentes, de tolérer les variations normales des caractères et d'être peu sensibles à la variation de l'échelle, de la courbure ... [Bélaïd-1]. Ainsi se pose le problème du choix et de la définition de ces caractéristiques. Une classe de caractères pouvant être représentée par un prototype (une forme idéale) ou un ensemble d'exemples, la sélection des caractéristiques tente donc de trouver les attributs de formes caractéristiques de chaque classe. Sans exhaustivité, des attributs de formes globaux, comme le nombre d'occlusions, les concavités et l'allongement relatif des extrémités du caractère, et des attributs de forme locaux, comme les positions relatives d'extrémités de traits, les croisements de traits et les changements de direction des traits, sont couramment utilisés.

Des méthodes de type « template matching » ou « elastic matching » ont été parmi les premières à être utilisées [Burr] pour la classification de caractères latins. D'autres ont utilisé les propriétés de l'invariance à la rotation à l'échelle et à la translation des moments de Zernike [Khotanzad] [Adam-3]. [Stentiford] cherchait à sélectionner les meilleures configurations de N-uplets de pixels correspondant à une forme de caractère donnée et à appliquer cette méthode aux adresses postales. De même [Stringa] procédait à une succession de réduction de la taille des caractères de codes postaux pour comparer les formes de caractères à des modèles de référence. Mais dès 1987, l'idée de combiner diverses techniques en vue d'une reconnaissance de caractères multi-fonte émergea. Pour ceci, [Kahan] utilisait une représentation structurelle des caractères¹³. Ensuite les méthodes neuronales sont apparues avec un gain notable du taux de reconnaissance [Avi-Itzhak]. Peu de caractères prototypes suffisaient pour entraîner le réseau de neurones à correc-

tement reconnaître les caractères^{14,15}, mais en édulcorant les problèmes de confusion comme entre le « I », le « l » et le « 1 » par exemple. D'autres travaux comme [Ennaji] considéraient par contre ces problèmes de confusion. En 1993, [Pavlidis-2] utilisa une méthode structurelle sur des images en niveau de gris pour reconnaître les codes postaux. Sa méthode approximait des « lignes de crête » à partir de l'analyse d'un indice de réflexion des surfaces et assemblait ces « lignes de crête » en un graphe qu'il tentait de mettre en correspondance avec des graphes de référence [Rocha-1] [Rocha-2]. [Ho] développa une méthode de combinaison de quatre classifieurs basée sur les techniques de « Borda Count » et de modèles de régression ou de sélection dynamique. Ceci permit d'améliorer sensiblement le taux d'erreur de reconnaissance.

La classification des caractères orientaux (figures 3, 4, 5) est plus difficile que pour les caractères latins du fait du nombre et de la complexité des idéogrammes, mais il y a une moindre variété de fontes.

Des approches structurelles ont été développées pour les caractères chinois en les représentant sous forme de graphes. Ainsi des méthodes de correspondance de graphe¹⁶ furent utilisées [Huang]. Pour la reconnaissance de l'arabe [Al-Yousefi] utilisa également ce type d'approche basée cette fois sur des profils de projection pour séparer les diacritiques et en appliquant un classifieur bayésien sur des caractéristiques extraites de ces projections. [Gu], [Wang-1] et [Wang-2] utilisèrent une autre approche basée sur les arbres de décision. Vue la faible redondance de formes dans l'écriture coréenne, [Lee K.H.] utilisa une méthode syntaxique basée sur les segments de lignes et construite sur un ensemble de 377 règles de production. On peut également citer les travaux de [Murthy] et de [Karnik] sur la reconnaissance des caractères indiens.

Les techniques de traitement contextuel utilisent des connaissances au niveau mot pour corriger les erreurs commises par l'OCR. Elles utilisent les informations concernant les autres caractères reconnus dans le mot ainsi que la connaissance du texte dans lequel le mot apparaît. Typiquement, la connaissance du texte se matérialise sous la forme d'un lexique ou d'une liste de mots qui apparaissent dans le texte. Par exemple, un logiciel de reconnaissance de caractères peut ne pas être capable de distinguer avec fiabilité le *u* du *v* en seconde position du mot *qualité*. Le traitement contextuel déterminera que *u* est correct car le mot *qvalité* n'existe pas. Ainsi [Suen] utilisaient les « N-Grams » pour effectuer une correction des erreurs d'OCR. [Shinghal-1] [Shinghal-2] ont appliqué l'algorithme de Viterbi pour trouver le chemin le plus probable dans un treillis de mots candidats. [Hull-1] et [Hull-2] combinaient ces deux approches. Puis l'application de ces techniques contextuelles s'est étendue à des vocabulaires de plus en plus larges pouvant contenir des

13. Line Adjency Graph (LAG).

14. 94 prototypes en apprentissage et 1 072 000 en test pour une fonte !

15. 1128 prototypes en apprentissage et 347 712 en test pour des caractères de fontes et de tailles différentes !!!...

16. Graph matching.

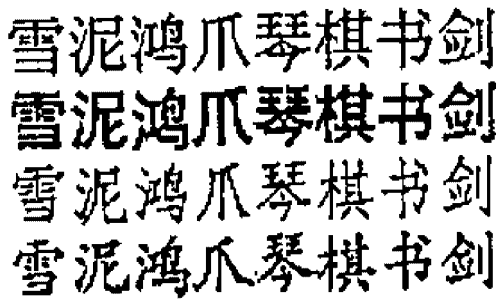


Figure 3. Fontes chinoises.

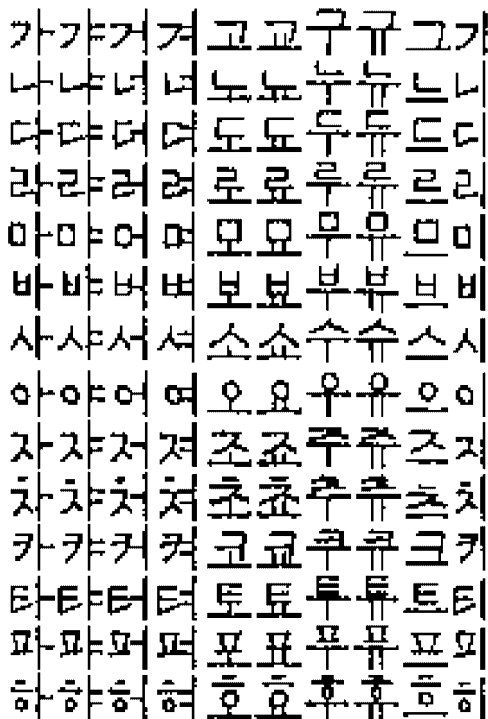


Figure 4. Fontes coréennes.

وذكر راديو الجزائر أيضاً أن مصادمات
وقعت في بلدة خنشلة الواقعة شرق
الجزائر.

Figure 5. Écriture arabe.

abréviations, des noms propres, etc... [Sinha]. Les Modèles de Markov Cachés¹⁷ furent utilisés pour détecter des mots dans des images dégradées de documents [Kuo]. De même [Hull-3] appliqua des grammaires statistiques qui pouvaient être adaptées à différentes langues.

Diverses études ont été entreprises sur l'évaluation des performances d'un OCR pour l'anglais. [Rice-1] [Nartker] [Rice-2] ont régulièrement produit une revue des produits du commerce existants pour la reconnaissance des documents imprimés. On peut y lire que si les documents sont de bonne qualité, le taux de reconnaissance atteint les 99,77 % et chute dès que les documents sont de qualité médiocre (entre 89 % et 97 %). [Sawaki] a évalué la reconnaissance de caractères dans des documents comportant des graphiques et [Wilson] présente un rapport sur l'utilisation et l'évaluation de l'OCR pour des applications de lecture automatique de formulaires.

Ainsi de nombreuses méthodes ont été développées pour cette étape de reconnaissance des caractères et actuellement les méthodes neuronales et celles à base de modèles de Markov cachés semblent s'être imposées du fait de leur grande capacité d'apprentissage. Ces dernières sont d'ailleurs plus communément employées pour la reconnaissance de l'écriture manuscrite.

2.8. Les tableaux et les formulaires

Les tableaux et les formulaires sont à l'intersection entre des documents structurés et des documents graphiques. Les premiers sont un moyen de présenter l'information tandis que les seconds sont un moyen de collecter de l'information. Ainsi, soit les cellules sont imprimées en même temps que leur contenu soit elles le sont avant. Toutefois, les étapes de traitement de ces éléments sont les mêmes. [Watanabe] présente des méthodes d'analyse et de classification de différentes représentations tabulaires à l'aide d'un arbre de classification. Chaque feuille dans cet arbre correspond à un arbre de structure globale ou spécifique. Les arbres de structure sont composés d'une liste de blocs successifs. L'arbre de classification est utilisé pour déterminer si un formulaire donné appartient à l'ensemble de ceux déjà analysés. Si ce n'est pas le cas, l'arbre de classification est enrichi de ce nouveau formulaire. Ainsi le système proposé permet d'acquérir dynamiquement des descriptions de nouveaux formulaires. [Yu-2] applique une méthode ascendante pour extraire la structure du tracé des cellules du tableau et les contenus des cellules. Des modèles sont construits à partir de formulaires vierges et mis en correspondance avec des formulaires remplis à analyser. Pour ceci, il utilise des graphes de bloc adjacents¹⁸. [Tang Y.T.] effectue une analyse multi-résolution en ondelettes pour extraire les tracés des tableaux. [Cesarini-2] présente un système permettant de traiter les formulaires ainsi que les factures en représentant les différentes zones informatives (logos, champs d'instruction ou d'information ...) sous la forme de nœuds d'un graphe dont les arcs expriment leurs positions relatives. Un réseau de neurones permet de reconnaître les logos et les champs d'instruction tandis que les modèles de formulaire sont construits grâce à une interface interactive. Des états de l'art sont présentés dans [Lopresti-2] et [Zannibi-2], et une éva-

17. Hidden Markov Models : HMM.

18. BAG = Block Adjacency Graph.

luation des performances des systèmes d'analyse de tableaux ou de formulaires a été décrite dans [Hu].

2.9. La reconnaissance des graphiques

La reconnaissance et l'interprétation des graphiques est un champ disciplinaire important et difficile du domaine de la reconnaissance d'image de documents car les graphiques peuvent se trouver mêlés ou liés à du texte, comme pour les diagrammes, les logos dans les en-têtes de lettres, et les lignes séparant les éléments d'un tableau ou plus simplement les subsections de texte dans un document. Les applications sont nombreuses et vont de la simple vectorisation et de la reconnaissance de primitives graphiques à l'analyse et l'interprétation de dessins techniques, de diagrammes, de cartes, de plans de cadastre ou de plans d'architecte [Kasturi-2] [Tombre-1]. Même s'il existe des systèmes dans le commerce, ceux-ci sont généralement limités en fonctionnalité et en performance [Bunke-2] pour la séparation texte/graphique et la vectorisation par exemple, et les techniques utilisées sont rarement explicitées. Aussi ces systèmes ne sont-ils généralement que de simples aides à l'utilisateur pour une édition manuelle ou des assistants pour une semi-automatisation de l'interprétation [Tombre-2]. On retrouve un ensemble de traitements communs avec l'analyse de documents textuels ou dits structurés comme les pré-traitements, la segmentation ou l'extraction de caractéristiques (cf. subsections précédentes). Mais généralement on a d'abord une étape de séparation texte/graphique puis la vectorisation et enfin la reconnaissance de symboles ou autres éléments contenus dans le document.

2.9.1. La séparation texte/graphique/image

La séparation texte/graphique/image reste actuellement un problème ouvert car très complexe. Elle concerne la segmentation physique des primitives pour éviter d'effectuer des traitements inappropriés sur les différentes composantes de l'image (OCR sur un graphique, vectorisation sur du texte...). Pour extraire des images dans un document hétérogène, certains auteurs proposent des méthodes de segmentation à base de texture [Chetverikov] [Etemad] [Jain-1]. [Aubert] préconise d'intégrer la segmentation à la binarisation par une mesure d'homogénéité des zones identifiées. D'autres méthodes sont basées sur le RLSA de [Wahl] en effectuant des dilatations linéaires sur l'image du document pour ensuite filtrer les composantes selon des critères géométriques, d'excentricité et de densité. Dans une variante, [Lu] propose d'extraire le texte en effaçant tous les pixels qui ne composent pas de longues composantes linéaires, qui ne sont pas dans des régions à forte densité de traits ou au contraire qui appartiennent à des régions à forte densité de pixels, ou qui forment une composante de caractéristiques très différentes de ses voisines. [Gloger] utilise la transformée de Hough pour séparer les lignes horizontales et verticales du texte

dans les formulaires. Toutefois pour les documents à dominante graphique et dans lesquels le texte ne représente que des légendes ou des cotes de mesure par exemple, ces méthodes ne sont pas efficaces. Il convient alors d'utiliser d'autres méthodes telles que le filtrage morphologique directionnel permettant de localiser toutes les formes linéaires et de les séparer du texte et autres formes assimilables à du texte. Ces méthodes ont été appliquées à des cartes [Luo-1] [Luo-2] mais leur extension à des documents plus complexes reste difficile. Il existe également des méthodes de recherche explicite des lignes basées soit sur une transformée de distance [Kaneko], soit sur une vectorisation de l'image de document [Dori-3]. Enfin, d'autres méthodes sont basées sur l'analyse des composantes connexes et consiste à filtrer celles-ci par un ensemble de règles de taille, de forme, etc... On peut citer l'algorithme de [Fletcher]. Cette méthode comprend notamment une étape de regroupement des caractères reconnus en chaînes en trouvant les alignements entre les rectangles englobant chaque caractère. Cependant elle ne permet pas d'extraire le texte qui « touche » le graphique. Pour extraire ces caractères, [Tombre-4] propose de tenir compte de la largeur moyenne des caractères et de l'espace moyen entre eux pour étendre la zone de recherche dans laquelle des caractères sont susceptibles de compléter les chaînes de caractères déjà extraites. C'est une méthode similaire à [Cao]. Une autre approche possible est d'utiliser les diagrammes de Voronoi [Wang Y.]. Il existe également des méthodes de reconnaissance de caractères basées sur l'utilisation de la transformée de Fourier-Mellin qui sont capables de reconnaître des caractères même lorsque ceux-ci touchent le graphique [Adam-1] [Cariou].

2.9.2. La vectorisation

De nombreuses méthodes de vectorisation permettant la conversion de données image en vecteurs ou en arcs, ont été développées et même commercialisées mais elles souffrent généralement d'un manque de précision et de robustesse [Tombre-3]. Initialement cette opération était considérée comme une application directe de la squelettisation de l'image binaire, suivie d'une approximation polygonale des chaînes de points du squelette, mais la squelettisation déforme les jonctions et les extrémités des lignes. Aussi, un certain nombre de pré et post-traitements ont été développés pour pallier cet inconvénient. Les techniques à base de squelettisation permettent de réduire l'épaisseur des traits à 1 pixel. [Lam] présente un état de l'art sur certains algorithmes de squelettisation. [Arcelli] utilise une transformée de distance [Borgefors] pour extraire des lignes de crêtes. [Desseilligny] amène un nouveau concept, appelé « veinerization », qui est un cadre théorique permettant d'intégrer la majeure partie des techniques de squelettisation. Cette approche générique, basée sur la construction d'un graphe contenant toutes les informations topologiques nécessaires à chaque méthode de squelettisation, permet à l'utilisateur de paramétrer le contexte et les particularités du squelette qu'il désire extraire.

[Zhu] utilise des modèles statistiques d'alignements et de jonctions pour émettre des hypothèses d'axe médian et retient la plus probable. [Tabbone] décompose l'image en primitives graphiques 2D. [Hilaire] a étudié une amélioration du positionnement des vecteurs et de leurs jonctions par le biais de méthodes statistiques.

Les méthodes d'approximation polygonale permettent d'approcher une liste de points par un ensemble de segments de droites sur la base d'une fonction de coût. Ainsi l'algorithme dit « de la corde » [Ramer] consiste à approcher récursivement la liste de points par un ensemble de segments contigus en utilisant un critère d'arrêt défini par un seuil de distance d'approximation [Leu]. Cette technique a été améliorée par [Lowe] et [Rosin]. D'autres méthodes diffèrent par la définition de la fonction de coût : critère surfacique [Wall], étude de la variabilité des listes de points dans une bande de référence [Leung]. Ensuite, des post-traitements peuvent être appliqués pour améliorer la précision des vecteurs trouvés ou fusionner des vecteurs proches [Röösl] [Chen].

En complément, il existe également des méthodes d'approximation elliptique permettant de reconstruire certains objets contenus dans le document graphique. Soit cette approximation s'effectue à partir d'une liste de points [Rosin] [Pei] [Dosch-1], soit elle consiste à détecter directement dans l'image la présence de cercles ou d'ellipses à l'aide de la Transformée de Hough [Thomas] [Davies].

D'autres techniques reposent sur des outils d'appariement de contours [Han] selon des critères de parallélisme et de distance. Certaines reconstituent l'axe central de la forme à partir d'un algorithme de suivi de trait [Paquet] [Chouinard]. De même, [Dori-2] a développé une méthode appelée « Orthogonal Zig Zag » améliorée ensuite pour éviter les effets de perte de connexité [Dori-4]. [Vaxivière-2] utilise des techniques à base d'appariement de mailles qui consiste à trouver des correspondances entre une base de 48 formes de référence intégrées dans des mailles et l'image à analyser. [Boatto] applique un codage de type « run length »¹⁹ horizontal et vertical des traits pour mettre en évidence des points anguleux, des lignes et leurs inter-sections ensuite. De même, [Chhabra] détecte des lignes horizontales dans un environnement bruité. Pour finir, des méthodes à base de transformées ont été largement utilisées, notamment les transformées de Hough [Duda] [Illingworth] [Leavers] [Kang] et de Radon [Copeland] [Murphy] [Lipari] pour leur application au réseau routier [Lopez] ou pour la détection de cercle ou d'ellipse [Yuen] [Yoo].

2.9.3. La reconnaissance de symboles

La reconnaissance de symboles est une des applications de la reconnaissance de formes dans des images de documents. L'architecture, la cartographie, l'électronique, et d'autres

domaines d'ingénierie utilisent des notations graphiques dépendantes de leur domaine pour illustrer leurs documents. L'interprétation automatique de tels documents nécessite des traitements capables de reconnaître les alphabets correspondants de symboles. Ainsi [Llados-4] définit-il un symbole comme une entité graphique possédant une signification particulière dans le contexte d'un domaine applicatif spécifique. De plus, suivant l'application on peut trouver différents types de symboles selon leurs propriétés visuelles : de simples formes binaires composées de lignes (schémas électroniques, plans d'architecte), une combinaison de lignes et de formes pleines (partitions musicales), des formes en niveaux de gris ou en couleur (logos), des « silhouettes » (symboles géographiques)... Dans les années passées, la plupart des travaux cherchaient à convertir de manière automatique les images de documents graphiques en un format capable d'être lu et interprété par un système de CAO²⁰. Ceci a dirigé beaucoup d'études sur la conversion des données image en vecteurs²¹. Cependant ceci ne peut être le but final qui devrait plutôt viser à fournir une description sémantique et une interprétation complète ou limitée des graphiques en termes d'entités composites et de relations entre elles.

Les stratégies de reconnaissance de symboles utilisées dans différentes applications sont fortement influencées par les propriétés particulières du domaine, par le but du processus d'interprétation et par la stratégie globale employée durant l'analyse du document, de la même manière qu'elles peuvent influencer les autres étapes de l'analyse, telles que l'extraction de primitives et de caractéristiques, la segmentation et l'interprétation sémantique. Ainsi, beaucoup d'applications ont mené au développement de diverses approches pour la reconnaissance de symboles. Traditionnellement, on distingue les approches statistiques des approches structurelles²².

Avec une approche statistique, et comme pour la reconnaissance de caractères, chaque forme est représentée par un vecteur de caractéristiques dans un espace à N dimensions. Ainsi, il suffit de partitionner cet espace en différentes classes à condition bien sûr d'avoir correctement sélectionné les caractéristiques ainsi que les méthodes de partitionnement. [Trier-3] présente un état de l'art sur les méthodes d'extraction de primitives pour l'application à la reconnaissance de caractères, mais seulement un sous-ensemble de ces primitives est utilisable pour les symboles : celles basées directement sur les pixels de l'image, les caractéristiques géométriques (centre de gravité, axes d'inertie, excentricité, surface, périmètre, intersubsections de lignes, occlusions, profils de projection, ...), les moments géométriques et les transformées d'image. Les premières présentent l'inconvénient d'être sensibles à la rotation et au bruit. Les deuxièmes restent très sensibles au bruit et des méthodes de pré-traitement

19. Littéralement « Longueur de parcours » pour illustrer des séquences de pixels identiques.

20. Conception Assistée par Ordinateur.

21. Raster-to-Vector.

22. Cette dichotomie est quelque peu abusive. Il faut plutôt comprendre la distinction entre les approches structurelles de celles qui ne le sont pas.

doivent rendre plus robuste l'extraction de ces primitives. Les troisièmes ont des relations avec les caractéristiques géométriques et une certaine invariance aux transformations affines peut être obtenue [Kim S.H.] [Cheng] [Adam-2]. Enfin, les dernières changent l'espace de représentation de l'image pour extraire de nouvelles signatures. On peut citer la transformée de Fourier [Kim S.H.] [Yu-1], la transformée de Fourier-Mellin [Adam-2] ou d'autres plus particulières présentées par [Doermann]. Notons que l'un des problèmes des méthodes statistiques réside dans le fait qu'elles ne peuvent s'appliquer correctement que si le symbole à reconnaître est correctement segmenté ou idéalement, si il correspond à une masse connexe. Une fois l'ensemble des primitives choisi, le vecteur de caractéristiques de chaque symbole est mis en correspondance avec des classes de symbole prédéfinies, toutes les méthodes de mise en correspondance utilisées nécessitant un apprentissage préalable de ces classes. La manière la plus simple de partitionner l'espace des caractéristiques, et donc de séparer les classes, est basée sur une mesure de similarité des vecteurs de caractéristiques. En général une simple fonction de calcul de distance avec les vecteurs représentatifs des classes suffit à décider de la classe d'appartenance d'un symbole par rapport au représentant le plus proche. Une variante consiste à considérer k représentants par classe, c'est la méthode des k-plus proches voisins (k-ppv)²³. Les arbres de décision ont également été utilisés de même que les réseaux de neurones qui sont dotés d'une grande capacité d'apprentissage. Le tableau 2 résume les différentes approches utilisées selon les primitives extraites.

Tableau 2. Les approches de reconnaissance statistique de symbole : intersection entre le choix des primitives extraites et les méthodes de classification (tiré de [Lladós-2])

	Image	Géométrique	Moments	Transformées
Distance	[De Stefano] [Suda] [Yu-3]	[Arias] [Armand] [Furuta] [Lee H.J.] [Lin X.] [Parker] [Soffer]	[Kim S.H.]	[Doermann] [Kim S.H.]
k-ppv		[Samet]	[Adam-2]	[Adam-2]
Arbre de décision		[Jorge] [Okazaki]		[Yu-1]
Réseau de neurones	[Anquetil] [Cesarini-1] [Cesarini-3] [Reiher] [Yadid-Pecht] [Yang D.]	[Francesconi] [Miyao]	[Cheng]	

23. k-NN : k-Nearest Neighbors.

Avec une approche structurale, les symboles sont représentés à l'aide d'une décomposition hiérarchique de primitives géométriques (régions, lignes, points, arcs) et de leurs relations entre elles. Pour chaque symbole, un modèle de forme idéale est construit à partir de ces primitives. L'étape de reconnaissance consiste alors à comparer l'image du symbole avec tous les modèles et à lui assigner la classe du modèle lui correspondant le mieux. Beaucoup d'approches structurales sont basées sur une représentation des symboles à base de graphe [Groen] [Habacha] [Jiang] [Kuner-1] [Kuner-2] [Lee S.] [Lee S.W.] [Messmer] [Segen]. Ceci permet d'associer respectivement les noeuds et les arcs du graphe à des points et des lignes de l'image, ce qui est une description très naturelle des symboles. La reconnaissance d'un symbole s'effectue alors en cherchant le meilleur isomorphisme de sous-graphe entre l'image d'entrée et tous les modèles. Avec ce type d'approche, la segmentation et la reconnaissance des symboles sont combinées en une seule passe car les symboles peuvent être identifiés comme des sous-graphes de l'image entière. Mais à cause du bruit, de la distorsion et des erreurs de vectorisation, les isomorphismes de sous-graphes doivent être tolérants aux erreurs [Lladós-3]. C'est pourquoi des opérations d'édition, avec un coût associé à chaque opération, ont été définies pour transformer le modèle idéal de symbole en une représentation sous forme de graphe de l'image. Ces méthodes peuvent reconnaître des symboles déformés comme ceux tracés à la main, mais elles sont très coûteuses en temps de calcul.

Les grammaires formelles sont à la base d'autres approches syntaxiques. On peut citer [Bunke-1], [Fahmy], [Fahn], [Kiyko], [Lavirotte]. A l'aide de règles de production, elles permettent de générer toutes les formes valides d'un symbole à partir des primitives de base qui les composent. Du fait de la structure bidimensionnelle des symboles graphiques, les grammaires de graphe sont les plus souvent utilisées. L'image d'entrée est représentée par un ensemble de primitives qui est analysé par la grammaire qui teste s'il peut être généré à l'aide des règles de production. Ces grammaires ont été appliquées aux dessins techniques [Collin], [Dori-1], [Min] et à la reconnaissance de symboles texturés [Sanchez].

Certaines approches utilisent la mise en correspondance déformable de modèles²⁴ avec une représentation structurale [Burr], [Valveny]. Elles sont adaptées pour traiter les problèmes de distorsion de forme des symboles tracés à la main en cherchant à minimiser une fonction d'énergie pour trouver la meilleure représentation déformée d'un modèle de symbole qui ressemble à l'image en entrée.

D'autres approches codent la structure des symboles en utilisant un ensemble de règles définissant les contraintes géométriques entre les primitives composant les symboles. Ces règles sont alors appliquées sur l'image pour y trouver des symboles [Bley] [Myers] [Vaxivière-1]. [Pasternak] utilise une architecture de type « tableau noir²⁵ » pour guider l'application de ces règles.

24. deformable template matching.

25. Blackboard.

Les modèles de Markov cachés sont également utilisés pour représenter des modèles structurels si l'on considère un symbole comme une séquence d'états qui génèrent les caractéristiques extraites de l'image. L'étape de reconnaissance consiste alors à chercher la séquence d'états possédant la plus haute probabilité de générer l'image en entrée [Muller] [Chang] [Kosmala].

2.9.4. Quelques domaines d'applications

Un ensemble de domaines d'application peut être identifié pour la reconnaissance des graphiques. Bien qu'ils partagent un certain nombre de méthodes, les connaissances liées à chaque domaine ont engendré des développements particuliers.

Les schémas électriques et logiques : Historiquement, c'est une des premières applications de la reconnaissance de symboles et il existe une littérature abondante sur le sujet [Groen] [Habacha] [Jiang] [Kim S.H.] [Kiyko] [Kuner-2] [Lee S.]. Du fait de la standardisation des notations employées basées sur des occlusions, les schémas électroniques présentent l'avantage d'utiliser des symboles qui peuvent facilement être segmentés avant d'être classifiés, et il suffit de détecter leurs connexions linéaires pour les isoler.

Les dessins techniques : Dans ce domaine, les notations ne sont pas standardisées et les entités graphiques ont une signification particulière selon le contexte de l'application. [Ablameyko-1] recense quatre types d'entités graphiques : les arcs et les lignes, les lignes pointillées, les hachures et les dimensions. Ces primitives peuvent représenter une information angulaire, une partie d'un schéma mécanique, un axe de symétrie... Les symboles sont formés de ces primitives et leur reconnaissance est liée à la connaissance du domaine pour leur donner une signification propre au document et permettre ainsi une conversion dans un format acceptable pour un système de conception assisté par ordinateur. La reconnaissance des primitives de bas niveau a été largement étudiée et par exemple, le problème de la détection des arcs de cercle peut se faire en combinant diverses approches de la littérature [Dosch-2]. La détection de formes hachurées est un autre problème qui nécessite de regrouper des lignes ayant la même inclinaison [Ablameyko-2] [Antoine] [Kasturi-1]. Mais pour reconnaître des symboles de haut niveau tels que des parties mécaniques, la reconnaissance automatique ne peut se faire sans des techniques interactives de dialogue avec l'utilisateur, ni sans l'acquisition de connaissances du domaine [Lladós-2].

Les partitions musicales : La reconnaissance des symboles dans des partitions musicales est facilitée par la notation standardisée du domaine. Ainsi, un ensemble de techniques très spécifiques et seulement applicables à cette catégorie de document ont été développées. Généralement, le processus d'interprétation se déroule en trois étapes principales [Armand] [Fahmy] [Randriamahefa] [Yadid-Pecht]. D'abord les lignes de portée sont extraites par des techniques de projection ou d'analyse de parcours pour pouvoir extraire les différents symboles. Puis la reconnaissance des notes est effectuée par un réseau de neu-

rones par exemple. Dans cette application, il n'y a qu'un nombre fini de symboles à reconnaître. Finalement l'interprétation complète de la partition est formulée à l'aide d'une grammaire [Couasnon2].

Les cartes : La reconnaissance des symboles est au cœur du problème de la conversion des cartes pour un Système d'Information Géographique²⁶. Cependant il existe différents types de cartes incluant un géo-référencement. Les plans de cadastre sont peut-être ceux qui induisent le moins de difficultés car la reconnaissance des symboles se ramène à détecter des formes polygonales et hachurées et des OCR peuvent extraire de l'information présente sur le plan [Antoine] [Boatto] [Lladós-1] [Madej]. Les plans de réseaux (téléphone, électricité, gaz...) sont un autre type de cartes généralement binaires comportant des lignes et des petits symboles composés de primitives de base [Adam-2] [Hartog]. Pour ces plans, les approches développées sont très dépendantes du domaine d'application. Finalement les cartes géographiques constituent la catégorie la plus complexe avec des objets linéaires représentant les courbes de niveau, les axes routiers, et des symboles dont la signification est renseignée par la légende [De Stefano] [Myers] [Reiher] [Samet]. De plus, la couleur a un rôle informatif très important.

Les plans d'architecte : Un grand nombre de méthodes a été développé pour la reconnaissance de symboles dans les plans d'architecte permettant ainsi une conversion de l'image vers un système de CAO pour une nouvelle édition du plan ou pour une visualisation en 3D [Aoki] [Valveny] [Ah-Soon] [Park]. Deux structures de symboles peuvent être recensées dans ces plans : les symboles prototypés (porte, fenêtre ...) et les symboles texturés (mur hachuré, escalier ...). La principale difficulté de l'interprétation réside dans le fait que les symboles ne peuvent pas être segmentés avant d'être reconnus. La reconnaissance s'effectue en analysant l'image du document dans sa globalité, ainsi c'est un traitement lourd en temps de calcul.

La reconnaissance des logos : Les logos sont des exemples de symboles graphiques mais leur reconnaissance diffère du fait que celle-ci a pour objectif la classification de documents ou la recherche par le contenu dans une base de documents [Chang] [Cortelazzo] [Doermann] [Francesconi] [Soffer]. De tels systèmes de reconnaissance sont généralement basés sur l'extraction de signatures de l'image (codage des contours, moments invariants, étiquetage de composante connexe ...) qu'il s'agit de mettre en correspondance avec une base de modèles en utilisant différents types de distance, ou de reconnaître à l'aide d'un réseau de neurones [Cesarini-2].

Autres applications : Il existe un grand nombre d'autres applications de la reconnaissance des symboles mais nous n'en citons que trois d'entre elles. La reconnaissance des formules mathématiques se situe à la frontière de la reconnaissance de

26. SIG.

caractères et de la reconnaissance de symboles car elles sont composées de ces deux entités. De par sa structure, l'interprétation d'une formule n'est pas possible avec une approche de type OCR. La littérature présente des approches [Lavirotte] [Lee H.J.] [Ramel] qui utilisent un vecteur de caractéristiques pour reconnaître les différents symboles pour ensuite valider la structure de la formule par une approche syntaxique. La reconnaissance de symboles peut également être utilisée comme un outil dans les IHM²⁷ [Jorge] [Landay] [Wilfong]. Dans ce cadre, la reconnaissance est effectuée en ligne dans le but d'utiliser des raccourcis dans un environnement de type PDA²⁸ et d'ainsi permettre à l'utilisateur diverses opérations ou de dessiner dans une application de DAO²⁹. La reconnaissance des tracés combine alors l'information spatio-temporelle du tracé avec des techniques de mise en correspondance ou à base de modèles de Markov cachés. Pour finir, une application récente de la reconnaissance de symboles est l'indexation des documents Web par le contenu [Lopresti-1]. La possibilité de faire une requête graphique devrait permettre d'améliorer la recherche d'information sur le Web.

2.9.5. Discussion

Cette section soulève les problèmes non encore résolus ou non encore explorés pour le développement d'un système de reconnaissance de graphiques robuste et fiable. La séparation texte/graphique comporte encore des défauts concernant par exemple les tirets et autres formes allongées qui peuvent être assimilées à des caractères (le «I» ou le «l») en tenant compte de la multi-orientation possible des caractères dans un document graphique. A l'inverse, les caractères de petite taille peuvent être confondus avec des éléments de graphique. Les caractères collés au graphique reste également un problème récurrent s'il n'y a pas la connaissance *a priori* de ce genre de cas à traiter. La vectorisation est un problème partiellement résolu qui achoppe dans la détection robuste des points de jonctions entre segments. Les méthodes à base de squelettisation sont très sensibles au bruit et donc fragiles dans ce sens contrairement aux méthodes d'appariement de contours. D'autres méthodes tendent à détecter trop de points de jonctions. Quant aux symboles, ils inspirent de nombreuses perspectives au niveau de leur segmentation, leur représentation et de leur reconnaissance.

2.10. La structure logique du document

Reconstituer le contenu du document ne se limite pas à la reconnaissance des caractères ou des symboles contenus dans le document, il s'agit de générer une représentation de haut niveau sous la forme d'un document structuré et selon une forme adé-

quate pour l'application visée. [Ingold-2] présente l'exemple d'une page (figure 6) tirée d'un livre scientifique. Interpréter le document nécessite de connaître la structure logique du document, c'est-à-dire son organisation hiérarchique en chapitres, sections et paragraphes, d'identifier les définitions, les énoncés d'exercices, les descriptions d'expériences, les formules, etc... La figure 7 reflète visuellement cette structure au niveau de la page tandis que la figure 8 illustre la structure qui en découle.

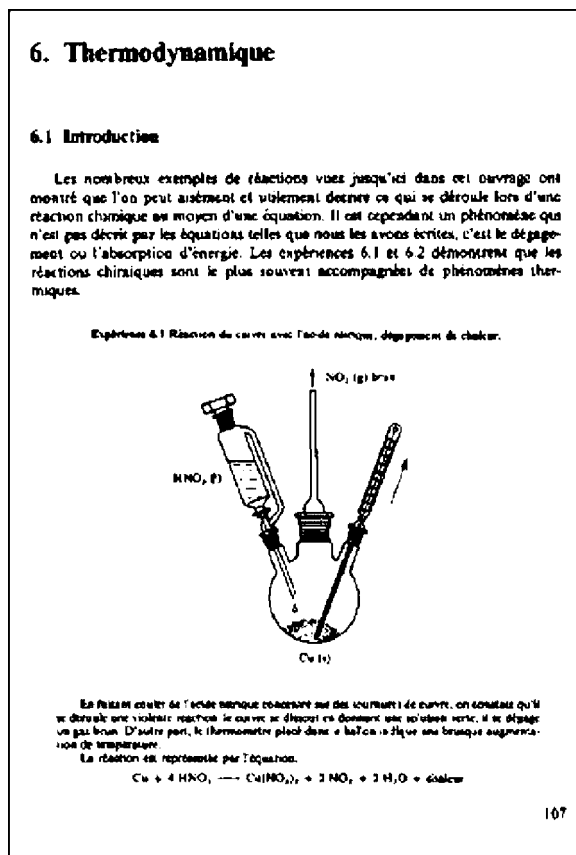


Figure 6. Une image de page numérisée.

Ainsi la détermination de l'organisation logique des entités élémentaires ou composites du document est l'objectif de la reconnaissance structurelle de documents³⁰. Il s'agit par exemple de :

- détecter les titres, les légendes de figures, les citations, les références bibliographiques, etc...
- construire l'organisation hiérarchique du document en chapitres, sections, sous-sections, etc...
- établir les liens entre les figures ou les notes de bas de page avec le texte

Cependant, il convient d'extraire au préalable la structure physique du document (cf. section 2.5.1) et d'étiqueter les blocs extraits pour ensuite déterminer les liens entre eux, de même

27. IHM : Interface Homme-Machine.

28. PDA : Personal Digital Assistant.

29. DAO : Dessin Assisté par Ordinateur.

30. Document understanding : Compréhension ou Interprétation de document.

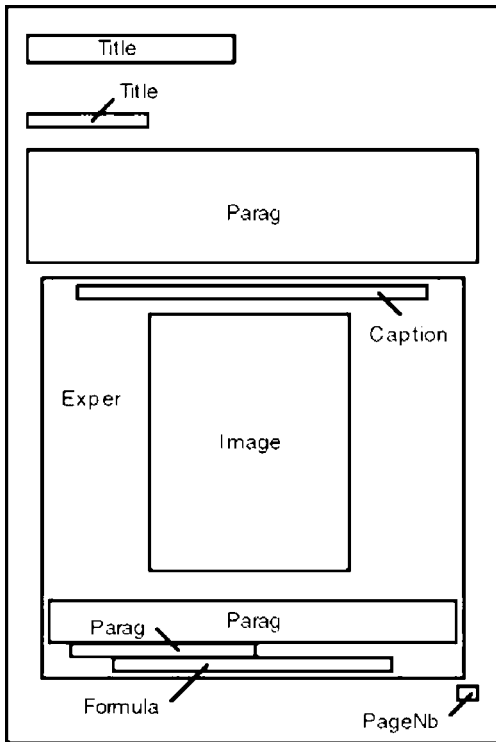


Figure 7. Etiquetage logique de la page.

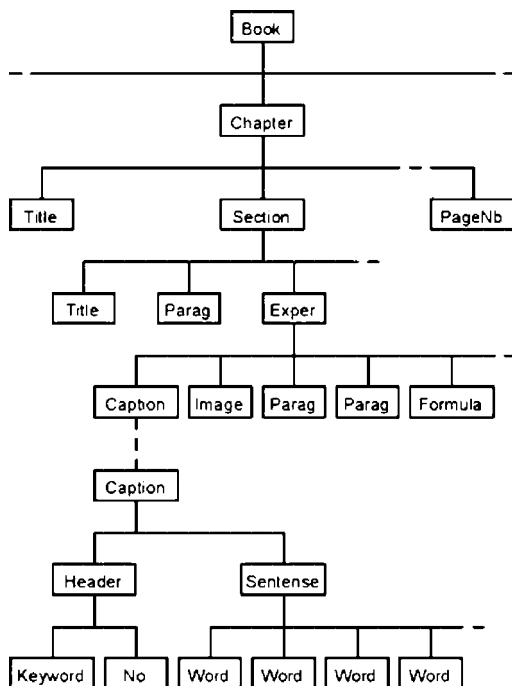


Figure 8. Structure logique du document.

que certains fragments de texte contenus à l'intérieur d'un bloc (citation, définition, date, référence...), les tableaux, les formules mathématiques... Cet étiquetage des différents compo-

sants du document est depuis son apparition généralement fait à l'aide du langage de structuration XML³¹.

2.11. Analyse et représentation des structures de documents : quelques systèmes

La structure logique du document peut être vue comme l'éti-quetage logique de composantes physiques du document. Ces étiquettes logiques sont généralement issues d'un ensemble de règles [Kreich], [Lin C.C.], [Ishitani], [Kim J.], [Summers]. Avec ce type de représentation, il n'y a pas de description des relations sémantiques entre les composantes logiques. Pour illustrer ces relations, les structures logiques de document doivent être représentées par des arbres qui sont également issus d'un ensemble de règles [Niyogi], [Tsujiimoto-1], [Yamashita], [Fisher], [Derrien Peden] ou de grammaires formelles comme dans [Ingold-1], [Conway], [Krishnamoorthy]. Dans ce cas, le document est considéré comme une phrase qui peut être soit une chaîne d'étiquettes logiques, soit une suite d'observations de caractéristiques liées aux composantes physiques du document. Généralement les règles de grammaire sont déterministes et il est alors difficile de lever certaines ambiguïtés si l'extraction des composantes physiques n'est pas robuste et présente des erreurs. Aussi il est possible d'agréments la grammaire d'attributs de coût pour construire la structure logique présentant un moindre coût [Tateisi].

Représentation à base de règles : [Tsujiimoto-1] représente les structures logiques et physiques sous la forme d'arbres et considère que la reconnaissance structurelle du document consiste en une transformation de l'arbre de structure physique en un arbre de structure logique en appliquant des règles de transformation génériques et en considérant des blocs de séparation virtuels. Il suppose que chaque nouvel article dans un document de type journal doit commencer par un titre et les blocs extraits dans la structure physique sont alors répartis en deux catégories selon leurs propriétés physiques (en-tête ou corps de texte). Considérant que chaque bloc correspond à une classe logique, des étiquettes logiques (résumé, sous-titre, paragraphe, entête, légende...) sont assignées aux blocs à l'aide d'un ensemble de règles. Testé sur 106 images de document, cet algorithme a permis de reconstruire la structure logique de 94 documents. Les erreurs sont dues à des erreurs de segmentation, à l'insuffisance de certaines règles de transformation et au fait que certaines pages ne présentent pas une structure hiérarchique.

[Kreich] décrit un environnement expérimental appelé SODA³² pour une analyse de document basée sur des modèles. Une approche montante est d'abord utilisée pour regrouper les composantes connexes en blocs de texte, puis les lignes de texte et les mots en sont extraits. Une reconnaissance de caractères et de graphiques est effectuée. Une base de connaissances contient la

31. XML : eXtended Markup Language.

32. System for Office Document Analysis.

connaissance du domaine, de la structure physique et de la structure logique. Les objets du document sont mis en correspondance avec les informations physique et logique de cette base, en utilisant une distance de Hamming afin de calculer une mesure de confiance qui doit être supérieure à un seuil pour une correspondance correcte. Le système est testé sur un document seulement et il n'y a aucune évaluation quantitative des résultats. [Fisher] présente un système à base de règles pour reconnaître les structures physique et logique de l'image d'un document sans aucune information *a priori* concernant le contenu ou le format du document. Le système extrait automatiquement la structure physique du document et la transforme en une structure logique. Trois types de règles sont utilisés selon des critères de position, de format et de composition textuelle. Ce système peut reconstruire les paragraphes coupés par le processus de mise en page, déterminer l'ordre logique de lecture des blocs de texte et les résultats sont stockés à l'aide du langage de balise MIF³³. Les régions de texte et de non-texte sont supposées déjà identifiées. Tout d'abord, les mots sont regroupés en paragraphes ou en colonnes. Ensuite la dimension et la position des colonnes sont calculées. Les structures physique et logique sont finalement déterminées à l'aide de règles appropriées. Aucun résultat expérimental n'est présenté mais la performance de l'algorithme dépend fortement du processus de séparation texte/non-texte.

[Yamashita] propose une méthode basée sur les modèles pour l'analyse de la structure logique. Le modèle est un arbre structuré qui définit l'information concernant l'arrangement géométrique des objets du document. Les objets physiques sont des chaînes de caractères, des lignes et des images. Des séparateurs horizontaux et verticaux (longues zones blanches et lignes noires) sont détectés en considérant les objets extraits et des étiquettes sont assignées aux chaînes de caractères grâce à une méthode de relaxation. Les différentes classes d'étiquette sont : en-tête, titre, auteur, affiliation, résumé, corps de texte, numéro de page, colonne, pied de page, bloc et figure. Cette technique a été appliquée à 77 pages de garde de brevets japonais et 59 structures logiques ont été correctement déterminées. Les erreurs sont dues à une mauvaise reconnaissance du numéro de page, à l'inclinaison, à des tâches et à des chaînes de caractères connectées.

[Derrien Peden] propose un système à base de frame pour analyser les structures physique et logique. La structure physique est obtenue en trois étapes. Tout d'abord les blocs de texte et les colonnes sont extraites par une découpe récursive en X-Y du document. Ensuite les lignes sont extraites à l'aide de règles spécifiques. Finalement, les zones physiques sont obtenues en analysant leurs propriétés topologiques. L'ordre logique de lecture est obtenu par un parcours en profondeur de l'arbre de la structure physique. La reconnaissance de la structure logique est menée en deux étapes : les paragraphes présentant les mêmes caractéristiques sont regroupés dans la même catégorie, et les étiquettes logiques sont assignées à chaque classe en fonction de règles d'attribution. Le système utilise une base de connais-

sance qui contient des modèles de structures physique et logique. Aucune évaluation du système n'est indiquée.

[Saitoh] présente un système pour la segmentation de document, la classification des zones extraites et la détermination de l'ordre de lecture. Cette approche est indépendante de la forme des blocs physiques et elle est robuste par rapport à l'inclinaison. Tout d'abord, les composantes connexes sont extraites puis classifiées. Ensuite elles sont fusionnées en lignes qui sont regroupées en zones. Ces zones sont alors classifiées en corps de texte, légende, entête... Un arbre de structure est généré à partir de ces zones et l'ordre de lecture est déterminé en parcourant cet arbre. Le test de cette technique a porté sur 393 documents japonais et anglais. Pour caractériser la performance de cette approche, [Saitoh] mesure le nombre moyen par document de corrections effectuées par un opérateur.

[Niyogi] présente le système DeLoS³⁴ pour la déduction de la structure logique de document. Dans ce système, un modèle est construit grâce à une structure de contrôle à base de règles et à un schéma de représentation de connaissance hiérarchique multi-niveau. Dans ce schéma, la connaissance sur les structures physique et logique de différents types de documents est stockée dans une base. Il inclut trois niveaux de règles pour la connaissance, le contrôle et la stratégie. Les règles de contrôle veillent à l'application des règles de connaissance et les règles stratégiques déterminent l'usage des règles de contrôle. Une image de document est d'abord segmentée en utilisant une approche montante et les blocs segmentés sont ensuite classifiés. Ensuite ces blocs forment l'entrée du système DeLoS et une structure logique est déduite. DeLoS a été testé sur 44 pages de journal et les résultats ont été estimés en termes de précision de classification de blocs, de fusion de blocs et d'extraction de l'ordre de lecture.

[Summers] décrit un algorithme pour la déduction automatique de la structure logique de document à partir de la structure physique générique. Cet algorithme considère la segmentation de texte en zones et la classification de ces zones en composantes logiques. La structure logique est obtenue en calculant une distance entre un élément physique et des prototypes prédéfinis. Le système a été testé sur 196 pages de rapports techniques en informatique et le système prend en entrée les blocs de texte. Deux mesures de précision ont été utilisées pour évaluer la performance du système qui affiche 85 % de précision.

[Lin C.C.] propose une méthode d'analyse de la structure logique de pages de livres en utilisant l'information *a priori* du sommaire, c'est-à-dire une description précise de la structure logique du livre entier. Les lignes de texte sont d'abord extraites et la reconnaissance de caractères opère sur chacune de ces lignes. Tous les éléments du document (numéros de page, titres, dessins, texte, etc...) sont alors analysés et mis en correspondance avec l'information issue du sommaire.

[Dengel-1] décrit le système DAVOS qui est à la fois capable d'apprendre et d'extraire une structure logique de document. Ce système apprend des concepts de structure de document en

33. Maker Interchange Format.

34. Derivation of the Logical Structure.

déteçant différentes valeurs d'attribut sur les objets du document. Un GTree³⁵ est construit grâce à des techniques d'apprentissage basées sur un arbre de décision. Ayant entraîné le système sur 40 lettres, le test du système a consisté à classer 40 autres lettres inconnues.

[Ishitani] propose un système d'analyse de structure logique de document basé sur «l'émergence». Le système inclut cinq modules interactifs : l'analyse de la typographie, la reconnaissance d'objets, la segmentation d'objets, le regroupement d'objets et la modification d'objets. L'interaction réside dans la configuration interactive du système qui amène à une analyse robuste du document. L'image du document est d'abord segmentée en lignes de texte qui sont ensuite classées en différents types selon certaines règles dédiées. Ces lignes sont ensuite regroupées et classées en composantes logiques selon des règles heuristiques. Les objets incorrectement segmentés peuvent être modifiés en vérifiant une certaine cohérence entre les objets. Dès qu'une structure du document est créée, «l'émergence» de la structure correcte est induite par des bouclages entre différents niveaux grâce aux modules interactifs. [Ishitani1999] présente un taux de 96,3 % d'extraction correcte des objets logiques sur une base de 150 documents.

[Srihari-2] propose une méthode basée sur la théorie de l'information pour l'interprétation d'adresses postales. L'entropie de Shannon est utilisée pour caractériser les composantes de l'adresse et leur interaction. Les composantes logiques considérées sont : la ville, l'abréviation de la ville, le code postal, le nom de la rue, etc... Les tests ont porté sur des adresses postales américaines.

[Kim J.] propose un module d'étiquetage logique à base de règles dans le système MARS³⁶ pour extraire les bibliographies de la base MEDLINE³⁷. Il déduit les règles à partir des résultats de l'analyse de la structure physique de journaux médicaux et de caractéristiques extraites de la reconnaissance de caractères. Le système a été testé sur une base de 11000 articles extraits de plus de 1000 journaux et les résultats ont montré plus de 96 % de précision.

Représentation à base de grammaires : [Ingold-1] propose une méthode de reconnaissance de la structure logique d'un document en décrivant de manière formelle chaque classe de document à l'aide de règles de composition et de présentation. La technique a été appliquée à des documents juridiques³⁸. Les règles de composition définissent la structure logique générique et les règles de présentation définissent les caractéristiques physiques des entités logiques à reconnaître. Les règles de composition sont décrites par un formalisme inspiré des grammaires formelles. Ce formalisme est une adaptation de la notation

EBNF³⁹. La description du document décrit un graphe d'analyse dont les arcs sont étiquetés avec les classes des entités à reconnaître. Le «successeur» d'une entité est l'étiquette logique de l'entité qui sera évaluée après l'entité courante. L'«alternative» spécifie des remplacements possibles pour l'entité courante. L'analyse du document est effectuée par la recherche d'un chemin dans le graphe d'analyse sous la contrainte que les attributs typographiques d'une entité de ce chemin doivent correspondre à ceux de l'objet correspondant du document. Dans la présentation de la méthode, les zones physiques sont supposées avoir déjà été préalablement segmentées et la reconnaissance de caractères déjà effectuée. Aucun résultat expérimental n'est disponible.

[Conway] utilise des grammaires de page associées à des techniques d'analyse de page pour reconnaître la structure logique du document à partir de la structure physique. La structure physique est décrite par un ensemble de règles grammaticales. Chacune d'entre elles est une chaîne de composantes liées par une relation de voisinage. Les différentes relations sont du type «sous», «à gauche de», «sur», «sur la gauche» et «proche de». Une grammaire déterministe est également utilisée pour décrire la structure logique. La grammaire utilisée pour la structure physique prend en compte des informations telles que la taille de la fonte, le style, l'alignement et l'indentation. Aucun résultat quantitatif n'est présenté.

[Krishnamoorthy] propose une méthode de reconnaissance de la structure logique du document en appliquant récursivement des grammaires aux profils de projection horizontaux et verticaux de la page. Il y a quatre étapes dans le processus d'analyse. La première seuille l'histogramme de projection pour extraire des «atomes» qui sont regroupés en «molécules» dans une deuxième étape. La troisième permet d'assigner des étiquettes logiques aux «molécules» et dans la dernière étape des entités de même type sont fusionnées. Cette approche ne fait pas la distinction entre la structure physique et la structure logique.

Représentation à base de modèles stochastiques :

[Tateisi1994] considère l'analyse de la structure logique d'un document comme un problème d'analyse syntaxique et stochastique. Le document est modélisé comme une chaîne de lignes de texte et d'objets graphiques. Une étape de pré-traitement permet de segmenter et de classer les lignes de texte et les objets graphiques, et la chaîne est analysée à l'aide d'une grammaire stochastique régulière avec des attributs. Les caractères contenus dans les lignes sont reconnus et la taille de leur fonte est déterminée. Chaque grammaire étant associée à une fonction de coût, l'analyse consiste à retenir les résultats en fonction de ce coût. Les tests ont porté sur 70 pages de livres ou de magazines japonais avec près de 85 % de bonne reconnaissance.

[Brugger] décrit un modèle de structure logique de document basé sur une représentation statistique de motifs⁴⁰ dans une

35. Geometric Tree.

36. Medical Article Record System.

37. MEDLINE est une base de données bibliographiques qui couvre tous les domaines médicaux de l'année 1966 à nos jours : plus de 11 millions de références issues de 4 300 périodiques, principalement en langue anglaise.

38. Recueil systématique des lois fédérales suisses

39. EBNF: Extended Backus-Naur Form.

40. patterns.

classe de document. Cette approche consiste à combiner des modèles structurels avec des données statistiques. Il propose ainsi un modèle de n-gram généralisé dans lequel il représente les fréquences des différentes relations d'une structure hiérarchique. Pour chaque nœud de l'arbre, cinq types de relations de voisinage sont considérés. En utilisant les fréquences relatives de bi-grams et des tri-grams, il détermine l'étiquetage qui maximise la plausibilité globale de la structure arborescente au moyen d'une heuristique. L'atout majeur de ce modèle est sa parfaite adéquation à l'apprentissage incrémental. A chaque reconnaissance validée par l'utilisateur, le système peut mettre à jour les fréquences de chaque relation, ce qui a en principe l'effet d'améliorer la précision du modèle.

Représentation à base de patterns : [Robadey] propose une méthode faisant appel à un modèle bidimensionnel de motifs. Le principe consiste à caractériser chaque entité physique à l'aide d'une configuration qui comprend les propriétés physiques de cette entité (taille, fonte dominante) ainsi que de ses voisins (distances). Dans le modèle, chaque classe est représentée par un sélecteur de caractéristiques pertinentes et un ensemble de motifs qui constitue le répertoire des configurations connues. La reconnaissance consiste alors à comparer la configuration d'une entité donnée avec les motifs de chaque classe. Cette approche est dotée d'un mécanisme d'apprentissage incrémental et semble adaptée à traiter l'étiquetage logique de documents à structure bidimensionnel complexes tels que les journaux. Après apprentissage manuel, le système est capable de classer plus de 90 % des entités en ne commettant que 2 % d'erreurs.

Un autre système : Légèrement en marge des systèmes précédents, [Hitz] propose le système XMillum (eXtensible Markup ILLuminator) pour permettre la visualisation de données issues de la reconnaissance de documents. Il se fixe également comme objectif la création d'applications interactives de reconnaissance structurelle de documents. XMillum est basé sur une transformation XSLT de données au format XML issues de la reconnaissance. C'est le langage de transformation XSLT qui permet de définir comment les données doivent être représentées dans XMillum pour leur visualisation et comment les utilisateurs peuvent interagir avec ces données pour créer des applications. Il s'agit ainsi d'un cadre modulaire permettant de réaliser des scripts de résolution de tâches en interprétation de documents. Ce système intègre notamment :

- des modules d'affichage pour une visualisation rapide des résultats de traitement ;
- des modules d'interaction pour une correction/validation des résultats de l'interprétation ;
- des modules d'outils pour créer des dialogues avec l'utilisateur ou bien pour fournir des vues différentes sur les données en mettant en évidence des zones d'intérêt.

L'utilisation des transformations XSLT pour créer des vues différentes du même document suivant la tâche, permet également de gérer les différents profils utilisateur.

2.12. Évaluation de performances des systèmes

Cette section se propose de discuter de la difficulté d'évaluer les systèmes présentés dans la section précédente, mais également de tous les systèmes, car nous n'en avons indiqués que quelques-uns. Une liste complète des systèmes présentés dans la littérature nécessiterait un document à elle-seule. D'autre part, cette section peut également s'appliquer aux systèmes d'interprétation de documents techniques qui ont été présentés dans la section 2.9.4.

[Haralick1-3] indiquait dès 1994, qu'il est clair que beaucoup des systèmes présentés dans le paragraphe précédent donnent des résultats propres à la technique employée et sur des documents dédiés à leur utilisation. Mais dès que l'on désirera élargir la diversité des documents à traiter, ces systèmes devront avoir un fonctionnement proche de la perfection. Ceci veut dire qu'ils devront être validés sur des bases de plusieurs milliers de documents et pour ceci il faudrait également disposer de mesures de performances appropriées tant pour l'extraction de la structure physique que pour la construction de la structure logique. Il faut pour cela disposer de bases conséquentes de données d'images de documents avec des structures physiques et logiques correctement étiquetées pour chaque image. D'autre part, bien que des niveaux de performance importants soient nécessaires, très peu des méthodes présentées cherchent à optimiser leurs performances sur une base donnée. Ceci signifie que les modèles concernant la dégradation de document et les modèles pour décrire les variations de format de document ne sont pas toujours formulés de manière suffisamment riche pour permettre leur utilisation. Pourtant des modèles sur la dégradation avaient été publiés [Baird-3] [Kanungo] mais aucun d'entre eux n'a été validé.

Depuis, [Mao-2] présente un état de l'art dans lequel il discute de l'évaluation de performances des algorithmes développés sur la base de différents critères : métrique de performance, données expérimentales, spécification de la « vérité terrain », analyse des erreurs et évaluation comparative.

Une métrique est nécessaire pour évaluer les performances d'un algorithme donné. Elle est fonction de la base de données, de la « vérité terrain » et des paramètres de l'algorithme. Elle n'est pas obligatoirement unique afin de permettre de sélectionner la métrique adaptée à une analyse de performance particulière. [Krishnamoorthy] propose une métrique basée sur le pourcentage de zones correctement étiquetées. [Saitoh] utilise trois critères pour illustrer leurs résultats. [Niyogi] fait appel à trois métriques pour la classification des blocs, la fusion des blocs et la précision de l'ordre de lecture trouvé. [Lin C.C.] utilise deux types d'erreur d'étiquetage et un taux d'identification. Mais un point commun à toutes ces métriques est le manque de définition formelle. Elles correspondent plutôt à des évaluations intuitives.

[Yamashita] décrit une métrique basée sur une fonction de coût pour sélectionner le résultat de moindre coût. [Kreich] utilise une métrique basée sur la distance de Hamming pour calculer

une mesure de confiance des correspondances entre une base de structures physique et logique de document et un document donné. [Summers] définit des métriques de précision pour évaluer la performance de son algorithme. Ces métriques sont déjà plus formelles et présentent ainsi moins d'ambiguïté dans leur interprétation.

Une évaluation basée sur des bases de grande taille est nécessaire pour évaluer objectivement les performances des algorithmes et une « vérité terrain » est indispensable pour quantifier les résultats expérimentaux. Ainsi quelques auteurs ont utilisé des bases de test relativement importantes allant jusqu'à une centaine d'images.

Une évaluation comparative de performances est nécessaire pour pouvoir comparer deux algorithmes sur le même jeu de données. Pourtant, pour la plupart des algorithmes, il n'existe pas d'évaluation comparative, sauf pour [Dengel-1] qui a effectué une évaluation comparative de son algorithme selon le choix d'une approche montante ou descendante.

[Mao-2] résume ainsi les limitations et les besoins des travaux sur l'analyse des structures de document :

1. La plupart des travaux présentés ne sont pas basés sur des modèles formels pour les pages de document. Pourtant un tel usage présenterait plusieurs avantages :

- On pourrait utiliser un modèle qui a un niveau de complexité approprié pour une classe donnée de documents.
- Dès qu'un modèle a été choisi pour une classe de documents donnée, des exemples de cette classe pourraient être utilisés pour estimer les paramètres du modèle.
- Des modèles formels pourraient à la fois être utilisés pour l'analyse et la synthèse de documents. En effet, un modèle pourrait être validé en étant utilisé pour créer des images synthétiques de document qui pourraient être comparées à des images réelles pour une classe donnée.

2. La plupart des travaux sur la structure logique de document supposent que la structure physique a déjà été extraite.

3. La plupart des travaux utilisent des modèles déterministes qui sont très sensibles au bruit (photocopies, fax, etc...), ce qui peut mener à des résultats ambigus ou faux.

4. Parfois, une évaluation quantitative a complètement été négligée. Bien que les algorithmes présentés sont basés explicitement ou implicitement sur des modèles de document, peu d'entre eux fournissent une définition formelle de ces modèles. C'est ce qui rend difficile la caractérisation de la relation entre les modèles et la performance des algorithmes. De plus, les paramètres des algorithmes sont généralement choisis empiriquement. La génération d'images synthétiques pourrait permettre de simuler le fonctionnement du système et d'effectuer des expérimentations pour évaluer l'algorithme et notamment pour détecter ses faiblesses. De même, l'extraction de la structure physique peut produire des résultats erronés. Aussi, des modèles stochastiques, représentés par des grammaires stochastiques pourraient être utilisés pour pallier ce problème, en associant des probabilités aux résultats fournis par la structure physique et au bruit apparaissant sur le document.

Mais peut-on vraiment concevoir des modèles formels de documents ?

2.13. Conclusion

Cette section marque un bilan sur la présentation faite des techniques liées à la reconnaissance des images de documents. Si celui-ci est considéré en fonction de ce qui est abouti et ce qui ne l'est pas, on peut donc distinguer les problèmes résolus de ceux qui ne le sont pas. Les pré-traitements sont une pièce maîtresse dans les chaînes de traitement. On peut constater par exemple que les techniques de binarisation tendent à offrir des résultats très corrects, mais beaucoup d'images de documents ne peuvent être binarisées sans une analyse fine des niveaux de gris. Faire l'économie d'une binarisation correcte induit inévitablement des problèmes liés à la formation de l'image à traiter qui sera alors remplie de défauts et de bruit auxquels les traitements de plus haut niveau ne pourront remédier. Les techniques pour reconnaître les objets d'un document ont déjà atteint un certain niveau de maturité. Par exemple, les taux d'erreurs sur la reconnaissance des caractères isolés, des fontes, des codes barre⁴¹ est si faible qu'il est difficile de conduire des travaux significatifs dans ces domaines. Ceci semble également être le cas pour l'extraction de primitives graphiques et pour la vectorisation de dessins et de cartes. De même, les méthodes dirigées par les modèles pour l'interprétation des tableaux et des formulaires semblent abouties⁴². Des méthodes interactives appropriées existent pour spécifier leur modèle de lecture à partir de formulaires vierges ou remplis. La validation du contenu des cellules peut être effectuée par des vérifications croisées ou en invoquant des méthodes adaptées qui accèdent à des bases de données pour « recouper » l'information. Mais il n'y a pas de théorie qui ait pu émerger de ces traitements de masse. Peut-être est-il temps d'aller plus loin qu'une simple interprétation de la disposition géométrique de l'information⁴³.

D'autres points méritent une attention particulière. Pouvoir directement exploiter des formats d'image compressée permettrait d'améliorer la vitesse de traitement de ces images. L'exploitation de la couleur a été très certainement mise à l'écart dans le domaine et il faudrait transposer les études faites pour les images en niveaux de gris à la couleur. On est loin de pouvoir convertir des schémas techniques complexes ou des cartes en un format directement exploitable par un système de CAO. Il faudrait s'appuyer sur les limites des systèmes commerciaux existants et réfléchir à comment les parties automatisées d'un système d'interprétation automatique pourraient profiter pleine-

41. Les caissières de votre hyper-marché utilisent cette technologie avec leur « douchette »...

42. Qui n'a pas rempli un formulaire papier composé de champs pré-casés pour s'abonner à une revue, pour le recensement Insee, pour une commande par correspondance ou même pour obtenir des papiers officiels, etc...

43. DIA-oriented interpretation have been discovered for almost all the 26³ TLAs (three letters acronyms). Will FLAs be next ? [Nagy-2].

ment de l'interaction avec un utilisateur. Les techniques de reconnaissance de caractères devraient viser à encore diminuer les taux d'erreur quelque soit le support (fax, photocopie, etc...). Mais peut-être faut-il orienter les travaux à venir vers l'apprentissage et l'adaptativité. La reconnaissance statique n'est pas suffisante pour construire des systèmes dynamiques et adaptatifs. Il faut pouvoir exploiter les retours d'expérience des utilisateurs du système qui reflètent les erreurs spécifiques à l'application. [Nagy-2] propose que la fédération de chercheurs sur les méthodes syntaxiques, statistiques et structurelles s'associent à la communauté s'intéressant aux techniques d'apprentissage pour produire de nouvelles solutions.

L'étiquetage des contenus de document a entamé un virage avec l'apparition du langage XML qui joue un rôle important dans la représentation des données pour des applications de reconnaissance d'images de documents. Extraire des balises XML nécessite de considérer les problèmes liés à l'interface entre la reconnaissance d'images de documents et les systèmes de gestion de bases de données, le work-flow et les modèles formels. Les mêmes considérations peuvent s'appliquer entre les dessins et les standards d'échange de données pour la CAO ou les cartes et les SIG⁴⁴.

Une meilleure évaluation des algorithmes demeure un des objectifs à atteindre. Quels sont les documents les plus représentatifs pour imposer une méthodologie collégiale ? Il n'y a aucune idée directrice dans ce domaine aujourd'hui, surtout devant la prolifération de documents sur Internet.

Dans tous les cas, et à mesure que la maturité des outils avance, la maîtrise de ceux-ci devient de plus en plus difficile car ils sont de plus en plus nombreux et complexes, et parfois encore trop confidentiels ou simplement issus de développements académiques.

Finalement, au terme de cette section, on ne peut que constater la richesse en terme de techniques et de méthodes développées pour le domaine non encore complètement résolu de la reconnaissance d'images de document. Elle recouvre largement les domaines de la vision avec les problèmes liés à l'acquisition des images et à leur traitement. De même, elle emploie et vise à développer tout un ensemble de techniques de sélection de caractéristiques, de classification et de modélisation par exemple. Ce n'est pas un domaine de recherche se limitant à la numérisation de documents et au simple lancement d'un logiciel de reconnaissance de caractères comme certains aiment à le croire. Il est complexe et fait appel à des techniques proches de l'intelligence artificielle et de la modélisation de processus cognitifs afin d'aboutir à l'interprétation sémantique des documents pour qu'ils puissent être ré-utilisés dans tout contexte informatique avec une interaction humaine. Il suffit de constater l'énorme effort scientifique produit ces vingt dernières années pour aboutir à des solutions robustes. Bien sûr, il reste du chemin à parcourir...

3. Modèles de document et connaissances

Dans la section précédente, nous avons présenté différents systèmes utilisant implicitement ou explicitement des modèles de documents. Ceux-ci sont généralement composés d'informations contextuelles, dépendantes du type de document analysé, que certains auteurs qualifient de connaissances. Ce peut être une base de caractères de référence pour l'OCR, une base de caractéristiques pour les différentes fontes susceptibles d'apparaître dans le document ou des dictionnaires linguistiques. Mais un modèle de document doit représenter de manière générique les caractéristiques d'une classe de documents de manière à faciliter l'analyse de documents spécifiques appartenant à cette classe. Il décrit les contraintes sur les structures possibles au sein de cette classe. Généralement, il peut intégrer différentes formes de connaissances de base [Ingold-2] :

- des règles syntaxiques qui expriment les structures de (dé)composition admissibles pour les entités logiques ;
- des règles graphiques qui expriment de manière absolue ou relative le positionnement des objets dans la page ;
- des règles typographiques qui décrivent les caractéristiques propres au texte ;
- des contraintes sémantiques liées au domaine d'usage du document et/ou des connaissances linguistiques ;
- des données statistiques.

La section précédente illustre d'ailleurs les différentes approches utilisées pour représenter les modèles de structuration des documents : modèles à base de règles ou de grammaire, modèles stochastiques ou encore modèles à base de motifs. De plus, un système de Reconnaissance d'Images de Documents doit viser à extraire du document toutes ou partie des informations liées au domaine d'usage que le concepteur du document y a fait figurer, en fonction des attentes de l'utilisateur du système. Le domaine d'usage caractérise le niveau sémantique des informations dans le document. Aussi le processus d'interprétation peut-il être vu comme l'instanciation d'un modèle issu du domaine ou d'une partie de ce domaine. Cette instanciation est réalisée en utilisant des procédures de traitement d'images et de reconnaissance des formes, elles-mêmes guidées par une stratégie d'interprétation qui doit être déduite d'une requête initiale et d'éventuelles coopérations avec l'utilisateur. Pour cela, le système doit être doté d'une grande quantité de connaissances diverses provenant des experts du domaine pour concevoir les modèles de données à extraire et des traiteurs d'image pour décrire les outils nécessaires à l'extraction des informations contenues dans l'image et la façon de les utiliser. La catégorisation des différentes formes de connaissances présentée par [Ingold-2] semble donc insuffisante, si l'on se place au niveau du système d'interprétation d'images de documents et non seulement au niveau de l'interprétation d'un document, et il est intéressant de s'inspirer des études menées dans le domaine de

44. Système d'Information Géographique.

l'Ingénierie des Connaissances. Dans ce contexte, les travaux de [Newell] pour le développement de systèmes experts de seconde génération constituent une référence. [Adam-3] montre d'ailleurs quelles connaissances sont présentes dans les systèmes de vision et dans les systèmes d'interprétation de documents techniques.

3.1. Quelles connaissances ?

En terme de catégorisation des connaissances mises en jeu dans les systèmes d'interprétation de documents, [Sandakly] propose de les structurer en descriptives et opératoires. S'inspirant de cette approche et des travaux de [Newell], [Adam] présente une classification selon trois niveaux :

- Les connaissances descriptives ;
- Les connaissances opératoires ;
- Les connaissances stratégiques.

3.1.1. Les connaissances descriptives

Les connaissances descriptives ont pour but de fournir un modèle du contenu du document à interpréter. Elles représentent l'ensemble des connaissances symboliques et numériques décrivant les différentes entités manipulées par le système. Elles se décomposent en connaissances descriptives du domaine d'usage et en connaissances descriptives physiques. En effet, pour interpréter un document, la connaissance de l'association entre une entité logique et sa représentation physique n'est pas suffisante car celle-ci respecte des règles précises quant au domaine d'usage, en terme de fonctionnalité. Si le domaine d'usage correspond à un domaine précis comme celui de l'ingénierie par exemple, on parlera de domaine métier. D'autre part, les connaissances descriptives physiques correspondent à la connaissance a priori que le système a de la représentation graphique des objets contenus dans le document (une liste de points de contour ou de squelette, un graphe, un vecteur de caractéristiques...). Cette dichotomie fait apparaître qu'il y a deux acteurs qui peuvent intervenir dans l'acquisition de connaissances par le système : l'expert du domaine (plan de réseau de France Télécom pour [Adam-3], le juriste connaissant les règles de présentation des Recueils Systématiques pour les Lois Fédérales pour [Ingold-1] etc...) et le traiteur d'image. Si l'on souhaite qu'un système puisse véritablement être générique, il faut disposer d'une Interface Homme-Machine pour acquérir ces connaissances qu'il s'agit d'externaliser pour ne pas rendre le système dédié à une seule application.

3.1.2. Les connaissances opératoires

Pour appliquer une stratégie de traitement en fonction d'une requête initiale, un système d'interprétation doit disposer d'une boîte à outils de traitement d'images et de reconnaissance de formes, mais aussi certaines connaissances sur ces outils. Au niveau du domaine d'usage, les connaissances opératoires

concernent les méthodes d'association d'entités selon leur fonctionnalité. Elles consistent à décrire l'efficacité des méthodes d'associations d'objets en fonction des relations exprimées dans les connaissances descriptives du domaine. La liste suivante présente les différentes connaissances opératoires qu'un système d'interprétation d'images idéal devrait intégrer selon [Adam-3] :

- Des connaissances sur les entrées, les sorties et les paramètres des algorithmes ;
- Des connaissances sur la loi de réglage des paramètres en fonction d'un contexte, de contraintes et/ou de son historique, ce qui exige :
- Des connaissances sur les contextes possibles, et leur caractérisation ;
- Des connaissances sur les contraintes applicables à un algorithme (temps, qualité...) ;
- Des connaissances sur l'historique des exécutions de l'algorithme en fonction des contextes et des contraintes ;
- Des connaissances sur l'évaluation de la qualité d'un résultat de traitement.

3.1.3. Les connaissances stratégiques

Pour rendre un système dynamiquement adaptable, les traitements intégrés et modélisés dans le système doivent être utilisés de différentes manières selon le contexte du traitement en cours et l'application considérée. La stratégie, c'est-à-dire l'enchaînement des tâches à effectuer, revêt donc un aspect fondamental pour parvenir à satisfaire un but d'interprétation sous des contraintes données. Il est nécessaire, pour que le système soit à la fois compréhensible et adaptable, de modéliser les différentes stratégies utilisables. [Adam-3] relève un certain nombre de connaissances du système concernant les stratégies :

- Des connaissances sur les tâches possibles du système. Elles regroupent l'ensemble des stratégies possibles d'utilisation du système dépendantes des attentes de l'utilisateur ;
- Des connaissances sur les contraintes possibles à appliquer à une tâche. L'utilisateur peut imposer une recherche dans une zone donnée, de respecter un certain délai... ;
- Des méthodes sur la méthode de décomposition d'une tâche en sous-tâches. Elles consistent à savoir lire le modèle de connaissances descriptives pour déterminer les données à extraire ;
- Des connaissances sur la planification des tâches (compétition, fusion, collaboration...) ;
- Des connaissances sur le rôle de l'utilisateur : quand et comment son intervention peut-elle être sollicitée ?
- Des connaissances sur les méthodologies d'acquisition de connaissances, pour prendre en compte de nouveaux objets, de nouvelles méthodes ou de nouvelles stratégies ;

3.2. Conclusion

Dans cette section, nous avons présenté une liste des différents types de connaissances qui peuvent être mises en jeu dans un processus d'interprétation d'images de documents. Nous avons

relevé trois principaux types de connaissances : descriptives, opératoires et stratégiques. Il existe bien évidemment des liens entre ces types de connaissances. Le choix effectué pour modéliser les traitements est tributaire des données puisque les traitements vont les manipuler. Cette dualité données/traitement, de même que l'interaction avec les connaissances stratégiques, rend très complexe la représentation des connaissances. Nous avons également fait apparaître différents acteurs pour la conception d'un système de reconnaissance de documents : un expert du domaine d'usage lorsque celui-ci est spécifié, un traiteur d'image et un utilisateur. Les deux premiers permettent au système d'acquérir toutes les connaissances nécessaires à son bon fonctionnement. Le dernier doit pouvoir exprimer une intention ou un but par rapport au contenu d'un document. Aussi faut-il assortir le système d'Interfaces Homme-Machine pour instaurer un dialogue entre chaque acteur et le système qui pourra devenir l'intermédiaire du dialogue entre un expert, ou un traiteur d'image, et un utilisateur si celui-ci désire acquérir une nouvelle compétence. On parle alors d'interface Homme-Machine-Homme.

D'autre part, si l'on considère les connaissances de base présentées par [Ingold-2], nous constatons que les règles syntaxiques, graphiques, typographiques et les données statistiques correspondent à des connaissances descriptives tandis que les contraintes sémantiques correspondent plutôt à des connaissances opératoires. Manquent alors les connaissances stratégiques.

Finalement, si l'on désire qu'un système ne soit plus dédié à une seule application, mais généralisable à d'autres, nous avons indiqué qu'il fallait rendre le système capable d'externaliser ses connaissances et non les figer dans le code de l'application, afin de ne pas contraindre le système. Ceci reste un défi majeur pour concevoir des systèmes génériques ou plus modestement multi-usage, c'est-à-dire capables de s'adapter à plusieurs domaines d'usage.

En effet, un système d'interprétation de documents a pour objectif de mettre en œuvre les outils et les techniques algorithmiques nécessaires pour extraire tout ou partie des informations contenues dans l'image, sachant que ces informations n'ont de sens que dans un contexte bien défini et suivant une intention exprimée [Garbay]. Ainsi, à partir d'une requête utilisateur, de l'interaction avec ce même utilisateur et des connaissances mises en jeu, le système doit-il permettre de mettre en œuvre une stratégie d'interprétation. Il doit pour cela utiliser une bibliothèque d'outils, une Interface Homme-Machine (IHM), et des connaissances. Mais si certains systèmes proposent à l'utilisateur de définir un but et un contexte à l'aide d'une interface graphique paramétrable, ils cherchent à le faire en assistant complètement l'utilisateur. Le dialogue machine-utilisateur ne donne pas à l'utilisateur la capacité d'expression de son intention au sujet de l'image et des objets particuliers qu'il y recherche, même si le dialogue est adapté dynamiquement par des règles *a priori*. Généralement l'utilisateur ne peut pas évaluer le processus d'interprétation. D'autre part les systèmes font appel à des modèles *a priori* trop rigides et donc dédiés à un métier ou à un type de document.

4. Conclusions et Perspectives

Tous les développements décrits dans la section 2 peuvent se rapporter au modèle de système d'interprétation de documents qui est illustré par la figure 9.

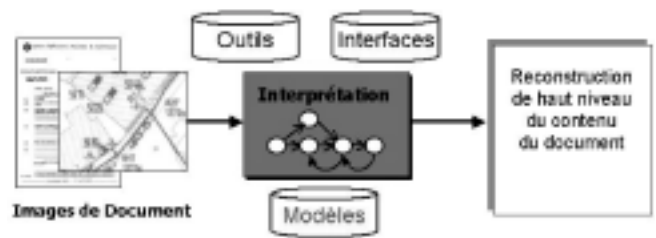


Figure 9. Un modèle de système d'interprétation de documents.

Nous disposons en entrée d'une image de document qu'il s'agit d'analyser pour en extraire de l'information. Ceci se fait en utilisant des outils, en appliquant des modèles et au-travers d'Interfaces Homme-Machine.

Vus les travaux décrits, une perspective est désormais de concevoir des systèmes adaptables. En effet, si beaucoup de systèmes de la littérature semblent robustes, ils le sont pour une application dédiée. A notre avis, ceci est dû au fait qu'ils n'ont pas été développés en utilisant une méthodologie de conception liée à la connaissance. Plus précisément, deux problèmes principaux méritent d'être soulignés. D'une part, les concepteurs de systèmes ne tiennent pas compte du fait que beaucoup de connaissances doivent être séparées (ou externalisées cf. section 3) des programmes si l'objectif est de concevoir un système d'interprétation générique. D'autre part, un autre point dur qui explique le manque d'adaptabilité des systèmes, et qui est bien sûr lié au premier problème, concerne l'étape d'acquisition de connaissances, dont l'objectif est de configurer le système, mais surtout de pouvoir re-configurer celui-ci en terme d'adaptation à une nouvelle application.

De ce point de vue, l'interaction Homme-Machine n'est jamais prise en compte. C'est pourquoi, nous pensons qu'aucun système d'interprétation existant ne peut ni être rapidement adapté à une nouvelle application, ni être initialisé par un utilisateur novice (ou naïf) en interprétation de document.

Ceci doit nous amener à réfléchir à la notion de connaissance. Mais ce concept couvre différents aspects, et selon les points de vue, il peut correspondre à des données observées, des informations structurées, un mécanisme de raisonnement ou à des processus d'analyse. De plus, la connaissance peut provenir d'un opérateur humain, de systèmes d'informations, de bases de don-

nées, d'expériences ou de documents par exemple. En tous cas, l'ensemble des connaissances relie les informations et les savoirs [Dupoirier & Ermine], mais aussi et surtout les acteurs pour lesquels la connaissance est indispensable pour préparer une décision et pour agir. De ce fait, ils sont tout à la fois porteurs, destinataires et utilisateurs de cette connaissance en vue de l'action, et pour l'action.

- D'une part, la question est de savoir comment exprimer la connaissance, l'identifier, la représenter et la partager.

- D'autre part, comment les acteurs peuvent-ils acquérir la connaissance et se l'approprier ?

Il faut tout au-moins rendre accessible l'information à la bonne personne et au bon moment. Dans le contexte des documents, cette information est ce que les systèmes et/ou acteurs, producteurs et consommateurs de documents, s'échangent. L'information écrite ou graphique, médiatise une pensée, se matérialise et laisse une trace sur un support physique ou numérique. Le document que vous lisez est une trace de mémoire et de capitalisation qui se compose de texte et d'images fixes. C'est aussi une trace dans la chaîne de valeur « données-informations-connaissances » [Gandon]. C'est également le point de vue de [Dengel-2] qui écrit que l'interprétation de documents est un processus qui transforme une(des) donnée(s) en de l'information en appliquant de la connaissance.

D'autre part, deux types de connaissances co-existent selon [Nonak]: les connaissances explicites et les connaissances tacites. Les premières sont saisies et stockées dans des archives, des systèmes de gestion de document, des bases de données ou des bibliothèques. Elles sont accessibles, transférables et formalisées. Les secondes sont très suggestives. Elles comportent des schémas mentaux, des expertises et des compétences dans des contextes spécifiques, et on peut très difficilement les formaliser. Elles ne sont pas explicitées. Par contre, l'utilisateur peut construire des schémas mentaux, c'est-à-dire des connaissances tacites qui lui sont propres, à force d'utilisation du système et donc d'apprentissage également. De cette manière, il pourra (ré)agir selon son intention en fonction de l'information que lui transmet le système. Mais cette information n'existe que dans le cas où l'utilisateur possède ces connaissances. Sinon, ce ne sont que des données.

Ainsi, c'est un grand défi de combiner ces sources de connaissances de manière à ce qu'une information puisse être automatiquement extraite d'un volume important de documents, pour être ensuite dirigée là où elle est nécessaire [Dengel-2].

Ces réflexions nous amènent donc à compléter le schéma de la figure 9. La figure 10 n'intègre pas directement la connaissance mais simplement les zones d'influence entre les acteurs et le système.

Les acteurs, chacun relié à son propre réseau, service ou organisation, sont :

- Le spécialiste du domaine d'usage, ou le concepteur du document. Celui-ci détient la connaissance lié au domaine d'usage, ou au métier, et qui produit des documents « métier » pour l'utilisateur. Il dialogue avec le concepteur dans une phase de spécification du système.

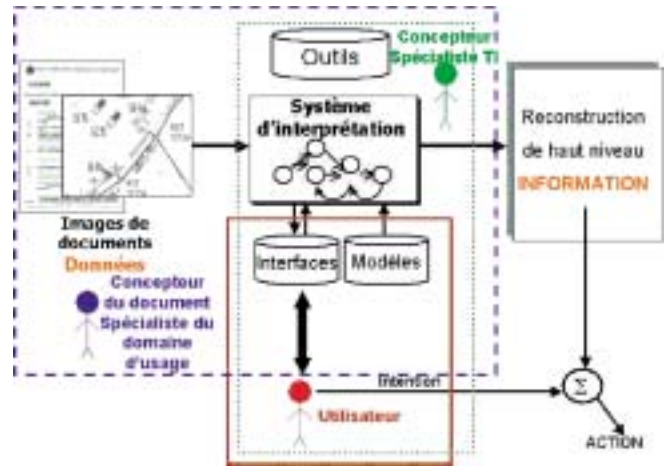


Figure 10. Un complément au modèle de système d'interprétation de documents.

- Le concepteur du système ou le spécialiste en Traitement d'Images. Il intègre les connaissances descriptives, opératoires et stratégiques liées à l'interprétation d'images.

- L'utilisateur peut être novice. Il est en interaction avec le système et peut rendre compte au concepteur des défauts du système comme un retour d'expérience. Il est muni d'une intention sur le document pour déclencher une action.

Il est très difficile d'illustrer la connaissance, si ce n'est par le biais du flux des informations que s'échangent les acteurs et le système. Car s'il y a information, il y a connaissance. C'est pourquoi, nous avons choisi de simplement formaliser ce flux comme l'illustration d'un dialogue, d'une interaction ou d'un apprentissage, et ceci de manière orientée par des flèches dans la figure 11.

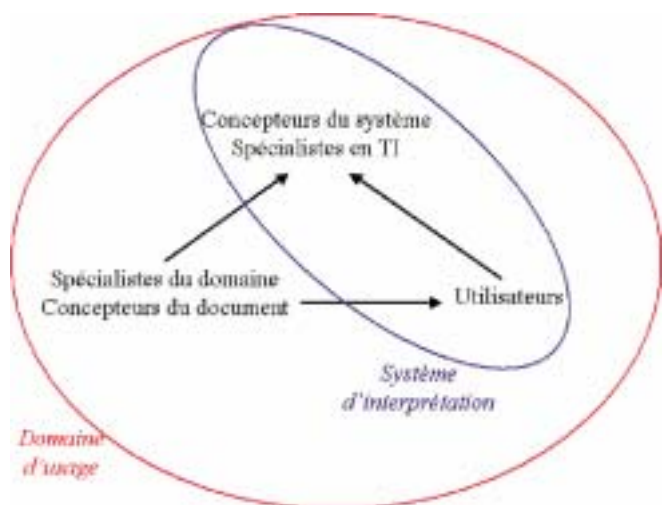


Figure 11. Flux d'information entre les acteurs et le système, dans un domaine d'usage ou métier.

Sur cette figure, le concepteur du système, spécialiste en Traitement d'Image, et l'utilisateur sont inclus dans le système d'interprétation. Ils interagissent avec lui et tous les acteurs sont liés au domaine d'usage.

Pour conclure, la conception d'un système d'interprétation de documents doit s'inscrire dans une démarche pluridisciplinaire car les documents sont liés à leur domaine d'usage. Cet usage peut s'inscrire dans une pratique professionnelle et donc correspondre à un métier.

En effet, il ressort des travaux menés ces dernières années que les points durs qui ont émergé lors de la conception de systèmes assistés d'analyse de documents « métier » constituaient de réels enjeux scientifiques. Nos conclusions placent l'utilisateur, socialement situé dans l'organisation, au centre du dispositif et offrent plus particulièrement des perspectives à moyen terme sur les points suivants :

- Il est nécessaire de concevoir une méthodologie ancrée dans une démarche pluridisciplinaire la plus générique possible pour l'analyse des interactions entre un corpus de documents et des communautés d'intérêt constituées au sein d'organisations. Nous pouvons utiliser le terme de sphère d'activité pour évoquer ces communautés d'intérêt,

- et il faut doter un système d'interprétation de fortes capacités d'interaction avec l'utilisateur (dialogue et apprentissage).

Dans ces conclusions, le document n'est plus considéré en tant que tel, mais du point de vue des stratégies d'acteur mises en place autour de lui, et au sein d'une organisation. Ceci conduit à une meilleure connaissance :

- des pratiques langagières en fonction des normes et pratiques sociales en vigueur,

- de la gestion des flux documentaires, autrement dit, des cycles de vie des documents,

- des contraintes propres au système d'information considéré qui font du document un élément déclenchant dans un contexte donné au sein d'une sphère d'activité,

- des sources considérées, autrement dit, de la typologie des corpus en fonction de leur degré d'interaction avec l'organisation considérée.

Cette démarche pluridisciplinaire doit considérer les traces que peuvent laisser les contraintes organisationnelles qui pèsent sur le document dans son intégralité, c'est-à-dire sur ses formes comme sur ses contenus : traces linguistiques, cognitives, interactionnelles entre producteur et utilisateur tout au long de son cycle de vie.

Elle peut ambitionner de faire émerger de nouveaux usages grâce à des interfaces faciles à faire évoluer et en analysant méthodiquement les glissements entre les comportements prévus par les concepteurs et les comportements réels des utilisateurs.

Nous lançons donc un appel à tous les acteurs de la communauté scientifique liés au traitement des images, à la reconnaissance des formes et aux sciences cognitives à se lancer dans cette voie de recherche pluridisciplinaire.

Références

- [Ablameyko-1] ABLAMEYKO S. (1997), « An introduction to interpretation of graphic images », *SPIE Optical Engineering Press*, 182 pages, 1997
- [Ablameyko-2] ABLAMEYKO S., BEREISHIK V., FRANSKEVITCH O., HOMENKO N., PARAMONOVA N. (1998), « Knowledge-based recognition of crosshatched areas in engineering drawings », *Lecture Notes in Computer Science*, Springer Verlag, Vol.1451: pp460-467, 1998
- [Adam-1] ADAM S., OGIER J.M., CARIO C., MULLOT R., LABICHE J., GARDES J. (2000), « Symbol and character recognition: application to engineering drawings », *Int. Journal on Document Analysis and Recognition*, Vol.3 (N.2) : pp 89-101, 2000.
- [Adam-2] ADAM S., OGIER J.M., CARIOU C., GARDES J., MULLOT R., LECOURTIER Y. (2000), « Combination of invariant pattern recognition primitives on technical documents », *Lecture Notes in Computer Science*, Vol.1451: pp 238--245, 2000.
- [Adam-3] ADAM S. (2001), « Interprétation de documents techniques : des outils à leur intégration dans un système à base de connaissances », *Thèse de doctorat de l'Université de Rouen*, 2001.
- [Aghajan] AGHAJAN H.K., KAILATH T. (1994), « Slide: Subspace-based line detection », *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.16 (N.11): pp. 1057-1073, 1994.
- [Ah-Soon] AH-SOON C., TOMBRE K. (2001), « Architectural symbol recognition using a network of constraints », *Pattern Recognition Letters*, Vol.22: pp. 231-248, 2001.
- [Akiyama] AKIYAMA T., HAGITA N (1990), « Automated entry system for printed documents », *Pattern Recognition*, Vol.2 (N.11): pp. 1141-1154, 1990.
- [Al-Yousefi] AL-YOUSEFI H., UDPA S.S. (1992), « Recognition of arabic characters », *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.14 (N.8): pp. 853-857, 1992.
- [Anquetil] ANQUETIL E., COUASNON B., DAMBREVILLE F. (2000), « A symbol classifier able to reject wrong shapes for document recognition systems », *Lecture Notes in Computer Science*, Vol.1941: pp. 209-218, 2000.
- [Antoine] ANTOINE D., COLLIN S., TOMBRE K. (1992), « Analysis of technical documents. », *Structured document image analysis*, Baird H.S - Bunke H. - Yamamoto K. editors, Springer Verlag, pp.385-402, 1992.
- [Aoki] AOKI Y., ARAI A., ODAKA H. (1996), « A prototype system for interpreting hand-sketched floor plans », *Proc. 13th Int. Conf. on Pattern Recognition*, pp. 747-751, Vienna, Austria, 1996.
- [Arcelli] ARCELLI C., SANNITI DI BAJA G. (1989), « A one-pass two-operation process to detect the skeletal pixels on the 4-distance transform », *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.11(N.4): pp. 411-414, 1989.
- [Arias] ARIAS J., LAI C., SURYA S., KASTURI R., CHHABRA A. (1995), « Interpretation of telephone system manhole drawings », *Pattern Recognition Letters*, Vol.16: pp. 355-369, 1995.
- [Armand] ARMAND J.P. (1993), « Musical score recognition: a hierarchical and recursive approach », *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, pp. 906-909, Tsukuba, Japan, 1993.
- [Aubert] AUBERT M. (1991), « Système de binarisation optimale de documents », *Thèse de doctorat de l'Institut National Polytechnique de Grenoble*, 1991.
- [Avi-Itzhak] AVI-ITZHAK H.I., DIEP T.A., GARLAND H. (1995), « High accuracy optical character recognition using neural networks with centroid dithering », *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17(N.2): pp. 418-431, 1995.
- [Baird-1] BAIRD H.S. (1987), « The Skew Angle of Printed Documents. », *Proc. Conf. Society of Photographic Scientists and Engineers*, pp 14-21, 1987.
- [Baird-2] BAIRD H.S., JONES S.E., FORTUNE S.J. (1990), « Image segmentation by shapes-directed covers. », *Proc. Int. Conf. on Pattern Recognition*, pp 820-825, Atlantic City, NJ, USA, June 1990.

- [Baird-3] BAIRD H.S. (1993), «Document image defect models and their uses », *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, Tsukuba, Japan, pp. 62-67, 1993.
- [Belaid-1] BELAID A., BELAID Y. (1992), «Reconnaissance de formes. Méthodes et applications», *InterEditions*, Paris, ISBN 2-7296-0399-8, 1992.
- [Belaid-2] BELAID A., TOUSSAINT Y. (2000), «Une méthode d'étiquetage morphosyntaxique pour la reconnaissance des tables des matières », *Conf. Int. Francophone sur l'Ecrit et le Document*, pp. 51-60, Lyon, France, 2000.
- [Bley] BLEY H. (1984), «Segmentation and preprocessing of electrical schematics using picture graphs», *Computer Vision, Graphics and Image Processing*, Vol.28: pp. 271-288, 1984.
- [Boatto] BOATTO L., and al. (1992), «An interpretation system for land register maps», *IEEE Computer magazine*, Vol.25(N.7): pp. 25-33, 1992.
- [Borgefors] BORGEFORS G. (1986), «Distance transform in digital images», *Computer Vision Graphics and Image Processing*, Vol.34: pp. 344-371, 1986.
- [Brugger] BRUGGER R., ZRAMDINI A., INGOLD R. (1997), «Modeling documents for structure recognition using generalized n-gram», *Proc. Int. Conf. on Document Analysis and Recognition*, pp. 56-60, Ulm, Germany, 1997.
- [Bunke-1] BUNKE H. (1982), «Attributed programmed graph grammars and their application to schematic diagram interpretation», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.4(N.6): pp. 574-582, 1982.
- [Bunke-2] BUNKE H., WANG P.S.P. (1997), «Handbook of character recognition and document image analysis», *World scientific*, 1997.
- [Burr] BURR, D.J. (1981), «Elastic matching of line drawings», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.3(N.6): pp. 708-712, 1981.
- [Cao] CAO R., TAN C.L. (2002), «Text/graphics separation in maps», *Lecture Notes in Computer Science*, Vol.2390: pp. 167-177, 2002.
- [Cariou] CARIOU C., OGIER J.M., ADAM S., MULLOT R., LECOURTIER Y., GARDES J. (1999), «A multiscale and multi-orientation recognition technique applied to document interpretation: application to the French Telephone network maps», *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol.13(N.8): pp. 1201-1218, 1999.
- [Casey] CASEY R.G., LECOLINET E. (1996), «A survey of methods and strategies in character segmentation», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.12(N.12): pp. 1190-1195, 1995.
- [Cesarini-1] CESARINI F., FRANCESCO E., GORI M., MARINAI S., SHENG J.Q., SODA G. (1997), «A neural-based architecture for spot-noisy logo recognition», *Proc. 4th Int. Conf. on Document Analysis and Recognition*, pp. 175-179, Ulm, Germany, 1997.
- [Cesarini-2] CESARINI F., GORI M., MARINAI S., SODA G. (1998), «INFORMys: a flexible invoice-like form reader system», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20(N.7): pp. 730-745, 1998.
- [Cesarini-3] CESARINI F., GORI M., MARINAI S., SODA G. (1996), «A hybrid system for locating and recognizing low level graphic items», *Lecture Notes in Computer Science*, Vol.1072: pp. 89-96, 1996.
- [Chang] CHANG M.T., CHEN S.Y. (2001), «Deformed trademark retrieval based on 2D pseudo-hidden Markov model», *Pattern Recognition*, Vol.34: pp. 953-967, 2001.
- [Chaudhuri] CHAUDHURI B.B., PAL U. (1997), «Skew angle detection of digitized documents», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19(N.2): pp. 182-186, 1997.
- [Chen] CHEN Y., LANGRANA N.A., KAS A.K. (1996), «Perfecting vectorized mechanical drawings», *Computer Vision and Image Understanding*, Vol.63(N.2): pp. 273-286, 1996.
- [Cheng] CHENG T., KHAN J., LIU H. YUN Y.Y. (1993), «A symbol recognition system», *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, pp. 918-921, Tsukuba, Japan, 1993.
- [Chetverikov] CHETVERIKOV D., LIANG J., KÖMÜVES J., HARALICK R.M. (1996), «Zone classification using texture features», *Proc. 13th Int. Conference on Pattern Recognition*, pp. 676-680, Vienna, Austria, 1996.
- [Chhabra] CHHABRA A.K., MISRA V., ARIA J. (1996), «Detection of horizontal lines in noisy run length encoded images: the FAST method», *Lecture Notes in Computer Science*, Vol.1389: pp. 314-321, 1996.
- [Chouinard] CHOUINARD C., PLAMONDON R. (1992), «Thinning and segmentation handwritten characters by line following», *Machine Vision and Applications*, Vol.5: pp. 185-197, 1992.
- [Collin] COLLIN S., COLNET D. (1994), «Syntactic analysis of technical drawing dimensions», *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol.8(N.5): pp. 1131-1148, 1994.
- [Conway] CONWAY A. (1993), «Page grammars and page parsing», *Proc. Int. Conf. on Document Analysis and Recognition*, Tsukuba, Japan, pp. 761-764, 1993.
- [Copeland] COPELAND A.C., RAVICHANDRAN G., TRIVEDI M.M. (1994), «Localized Radon transform based detection of linear features in noisy images», *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 664-667, Seattle, WA, USA, 1994.
- [Cortelazzo] CORTELAZZO G., MIAN G., VEZZI G., ZAMPERONI P. (1994), «Trademark shapes description by string matching techniques», *Pattern Recognition*, Vol.27: pp. 1005-1018, 1994.
- [Couasnon] COUASNON B. (2001) «DMOS: A generic document recognition method, application to an automatic generator of musical scores, mathematical formulae and table structure recognition systems», *Proc. 6th Int. Conf. on document Analysis and Recognition*, pp. 215-220, Seattle, WA, USA, 2001.
- [Davies] DAVIES E.R. (1988), «A modified Hough scheme for general circle location», *Pattern Recognition Letters*, Vol.7: pp. 37-43, 1988.
- [De Stefano] DE STEFANO C., TORTORELLA F., VENTO M. (1995), «An entropy based method for extracting robust binary templates», *Machine Vision and Applications*, Vol.8: pp. 173-178, 1995.
- [Dengel-1] DENGEL A.R., DUBIEL F. (1996), «Computer understanding of document structure», *Int. Journal of Imaging Systems and Technology*, Vol.7: pp. 271-278, 1996.
- [Dengel-2] DENGEL A.R. (2003), «Making document work: challenges for document understanding», *Int. Conf. on Document Analysis and Recognition*, pp. 1026-1035, Edinbourg, Scotland, 2003.
- [Derrien Peden] DERRIEN PEDEN D. (1991), «Frame-based system for macro-typographical structure analysis in scientific papers», *Proc. 1st Int. Conf. on Document Analysis and Recognition*, Saint-Malo, France, pp. 311-319, 1991.
- [Desselligny] DESSEILLIGNY M.P., STAMON G., SUEN C.Y. (1998), «Veinization: a new shape description for flexible skeletonization», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20(N.5): pp. 505-520, 1998.
- [Doermann] DOERMANN D., RIVLIN E., WEISS I. (1996), «Applying algebraic and differential invariants for logo recognition», *Machine Vision and Applications*, Vol.9: pp. 73-86, 1996.
- [Don] DON H.S. (2001), «A noise attribute thresholding method for document image binarization», *Int. Journal on Document Analysis and Recognition*, Vol.4(N.2): pp. 131-138, 2001.
- [Dori-1] DORI D. (1989), «A syntactic/geometric approach to recognition of dimensions in engineering machine drawings», *Computer Vision, Graphics and Image Processing*, Vol.47(N.3): pp. 271-291, 1989.
- [Dori-2] DORI D., LIANG Y., CHAI I. (1993), «Sparse pixel recognition of primitives in engineering drawings», *Machine Vision and Applications*, Vol.6: pp. 69-82, 1993.
- [Dori-3] DORI D., WENYIN L. (1996), «Vector-based segmentation of text connected to graphics in engineering drawings», *Lecture Notes in Computer Science*, Vol.1121: pp. 322-331, 1996.
- [Dori-4] DORI D., LIU W. (1999), «Sparse pixel vectorization: an algorithm and its performance evaluation», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.21(N.3): pp. 202-215, 1999.
- [Dosh-1] DOSH P. (2000), «Un environnement pour la reconstruction 3D d'édifices à partir de plans d'architecte», *Thèse de doctorat de l'Université H.Poincaré*, juin 2000.

- [Dosch-2] DOSCH P., MASINI G., TOMBRE K. (2000), «Improving arc detection in graphics recognition», *Proc. 15th Int. Conf. on Pattern Recognition*, pp. 243-246, Barcelona, Spain, 2000.
- [Duda] DUDA R.O., HART P.E. (1972), «Use of the Hough transformation to detect lines and curves in pictures», *Graphics and Image Processing*, Vol.15: pp. 11-15, 1972.
- [Dupoirier & Ermine] DUPOIRIER G., ERMINE J.L. Coordonnateurs (1999), «Gestion des documents et gestion des connaissances (préface)», *Document Numérique*, Hermès, Vol.3(N.3-4), 1999.
- [Ennaji] ENNAJI A. (1993), «Classification et parallélisme en reconnaissance de caractères imprimés multi-fontes», *Thèse de doctorat de l'Université de Rouen*, 1993.
- [Etemad] ETEMAD K., DOERMANN D., CHELLAPPA R. (1994), «Page segmentation using decision integration and wavelet packets», *Proc. 12th Int. Conf. on Pattern Recognition*, pp. 345-349, Jerusalem, Israël, 1994.
- [Fahmy] FAHMY H., BLOSTEIN D. (1993), «A graph grammar programming style for recognition of music notation», *Machine Vision and Applications*, Vol.6: pp. 83-99, 1993.
- [Fahn] FAHN C.S., WANG J.F., LEE J.Y. (1989), «A topology-based component extractor for understanding electronic circuit diagrams», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.2(N.11): pp. 1140-1157, 1989.
- [Fisher] FISHER J.L. (1991), «Logical structure descriptions of segmented document images», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 302-310, Saint-Malo, France, September 1991.
- [Fletcher] FLETCHER L.A., KASTURI R. (1988), «A robust algorithm for text string separation from mixed text/graphics images», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.10: pp. 910-918, 1988.
- [Francesconi] FRANCESCONI E., FRASCONI P., GORI M., MARINAI S., SHENG J.Q., SODA G., SPERDUTI A. (1998), «Logo recognition by recursive neural network», *Lecture Notes in Computer Science*, Vol.1389: pp. 104-117, 1998.
- [Furuta] FURUTA M., KASE N., EMORI S. (1984), «Segmentation and recognition of symbols for handwritten piping and instrument diagram», *Proc. 7th Int. Conf. on Pattern Recognition*, pp. 626-629, 1984.
- [Gandon] GANDON F. (2002), «Ontology engineering: a survey and a return on experience», *INRIA Research Report - RR 4396*, <http://www.inria.fr/rrrt/rr-4396.html>, 2002.
- [Garbay] GARBAY C., MAETRE H. (1998), «Interaction systèmes/environnement pour l'interprétation des signaux et des images», *Rapport de synthèse de l'action inter PRC 10.2 du PRC ISIS & CHM*, 1998.
- [Gloger] GLOGER J.M. (1992), «Use of Hough transform to separate merged text/graphics in forms», *Proc. 11th Int. Conf. on Pattern Recognition*, pp. 268-271, La Haye, Pays-Bas 1992.
- [Groen] GROEN F., SANDERSON A.C., SCHLAG F. (1985), «Symbol recognition in electrical diagrams using probabilistic graph matching», *Pattern Recognition Letters*, Vol.3: pp. 343-350, 1985.
- [Gu] GU Y.X., WANG Q.R., SUEN C.Y. (1983), «Application to a multi-layer decision tree in computer recognition of chinese characters», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.5(N.1): pp. 83-88, 1983.
- [Habacha] HABACHA A.H. (1991), «Structural recognition of distrubed symbols using discrete relaxation», *Proc. 1st Int. Conf. on Document Analysis and Recognition*, pp. 170-178, Saint-Malo, France, 1991.
- [Han] HAN C.C., FAN K.C. (1994), «Skeleton generation of engineering drawinds via contour matching», *Pattern Recognition*, Vol.27: pp. 261-275, 1994.
- [Haralick-1] HARALICK R.M., STERNBERG S.R., ZHUANG X. (1987), «Image analysis using mathematical morphology», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.9: pp. 532-550, 1987.
- [Haralick-2] HARALICK R.M., SHAPIRO L.G. (1992), «Computer and Robot vision», *Addison-Wesley, Reading mass*, 1992.
- [Haralick-3] HARALICK R.M. (1994), «Document Image Understanding: Geometric and Logical Layout», *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 385-390, 1994.
- [Hartog] DEN HARTOG J.E., TEN KATE T.K., GERBRANDS J.J. (1996), «Knowledge-based interpretation of utility maps», *Computer vision and image understanding*, Vol.63(N.1): pp. 105-117, 1996.
- [Hashizume] HASHIZUME A., YEH P.S., ROSENFELD A. (1986), «A method of detecting the orientation of aligned components», *Pattern Recognition Letters*, Vol.4: pp. 125-132, April 1986.
- [Hilaire] HILAIRE X., TOMBRE K. (2001) «Improving the Accuracy of Skeleton-Based Vectorization», *4th IAPR International Workshop on Graphics Recognition*, pp. 381-394, Kingston, Ontario, Canada, 2001.
- [Hitz] HITZ O., INGOLD R. (2000), «Vizualisation of document recognition results using xml technology», *Colloque International sur le Document Electronique*, pp. 207-215, Lyon, France, 2000.
- [Ho] HO T.K., HULL J.J., SRIHARI S.N. (1994), «Decision combination in multiple classifier systems», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16(N.1): pp. 66-75, 1994.
- [Hochberg] HOCHBERG J., KELLY P., THOMAS T., KERNS L. (1997), «Automatic script identification from document images using cluster-based templates», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(N.2): pp. 176-181, 1997.
- [Hu] HU J., KASHI R., LOPRESTI D.P., WILFONG G.T. (2001), «Evaluating the performance of table processing algorithms», *Int. Journal on Document Analysis and Recognition*, Vol.4: pp. 140-153, 2001.
- [Huang] HUANG X., GU J., WU Y. (1993), «A constrained approach to multifont chinese character recognition», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15(N.8): pp. 838-843, 1993.
- [Hull-1] HULL J.J., SRIHARI S.N. (1982), «Experiments in text recognition with binary n-Gram and Viterbi algorithms», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.4(N.5): pp. 520-529, 1982.
- [Hull-2] HULL J.J., SRIHARI S.N., CHOUDHARI R. (1983), «An integrated algorithm for text recognition: comparison with a cascaded algorithm», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.5(N.4): pp. 384-395, 1983.
- [Hull-3] HULL J.J. (1996), «Incorporating language syntax in visual text recognition with a statistical mode», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.18(N.12): pp. 1251-1256, 1996.
- [Illingworth] ILLINGWORTH J., KITTLER J. (1988), «A survey of the Hough transform», *Computer Graphics and Image Processing*, Vol.44: pp. 87-116, 1988.
- [Ingold-1] INGOLD R., ARMANGIL D. (1991), «A top-down document analysis method for logical structure recognition», *Proc. 1st Int. Conf. on Document Analysis and Recognition ICDAR*, Saint-Malo, France, pp. 41-49, 1991.
- [Ingold-2] INGOLD R. (2002), «Analyse et reconnaissance d'images de documents», *Techniques de l'Ingénieur 2002*, Référence H 7 020, Août 2002.
- [Ishitani] ISHITANI Y. (1999), «Logical structure analysis of document images based on emergent computation», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 189-192, Bangalore, India, September 1999.
- [Jang] JANG J.H., HONG K.S. (1999), «Binarization of noisy gray-scale character images by thin line modeling», *Pattern Recognition*, Vol32(N.5): pp. 743-752, 1999.
- [Jain-1] JAIN A.K., CHEN Y. (1994), «Address block location using color and texture analysis», *Computer Cision, Graphics and Image Processing*, Vol.60(N.2): pp. 179-190, 1994.
- [Jain-2] JAIN A.K., YU B. (1998), «Document representation and its application to page decomposition», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20: pp. 294-308, 1998.
- [Jiang] JIANG X., MUNGER A., BUNKE H. (2000), «Synthesis of representative graphical symbols by computing generalized median graph», *Lecture Notes in Computer Science*, Vol.1941: pp. 183-192, 2000.
- [Jorge] JORGE J.A., FONSECA M.J. (2000), «A simple approach to recognize geometric shapes interactively», *Lecture Notes in Computer Science*, Vol.1941: pp. 266-274, 2000.

- [Kahan] KAHAN S., PAVLIDIS T., BAIRD H.S. (1987), «On the recognition of printed character of any font and size», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.9(N.2): pp. 274-288, 1987.
- [Kamel] KAMEL M., ZHAO A. (1993), «Extraction of binary character/graphics images from grayscale document images», *Graphical models and image processing*, Vol.55(N.3): pp. 203-217, 1993.
- [Kaneko] KANEKO T. (1992), «Line structure extraction from line-drawing images», *Pattern Recognition*, Vol.25(N.9): pp. 963-973, 1992.
- [Kang] KANG C.W., PARK R.H., LEE K.H. (1991), «Extraction of straight line segments using rotation transformation: generalized Hough transform», *Pattern Recognition*, Vol.24(N.7): pp. 633-641, 1991.
- [Kanungo] KANUNGO T., HARALICK R.M., PHILLIPS I.T. (1993), «Global and local document degradation models», *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, Tsukuba, Japan, pp. 730-734, 1993.
- [Karnik] KARNIK R.P. (1999), «Identifying Devnagari characters», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 669-672, 1999.
- [Kasturi-1] KASTURI R., BOW S.T., EL-MASRI W., SHAH J., GATTIKER J.R., MOKATE U.B. (1990), «A system for interpretation of lines drawings», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12(N.10): pp. 978-992, 1990.
- [Kasturi-2] KASTURI R., TOMBRE K. (1996), «Graphics recognition - Methods and applications», *Volume 1072 of Lecture Notes in Computer Science*, Springer Verlag, May 1996.
- [Kasturi & O'Gorman] KASTURI R., O'GORMAN L. (1997), «Document Image Analysis», *IEEE Computer Society Press*, ISBN 0-8186-7802-X, 1997.
- [Khotanzad] KHOTANZAD A., HONG Y.H. (1990), «Invariant image recognition by Zernike moments», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12(N.5): pp. 489-497, 1990.
- [Kim J] KIM J., LE D.X., THOMA G.R. (2001), «Automated labeling in document images», *Proc. 8th Conf. on Document Recognition and Retrieval*, San Jose, CA, USA, pp. 111-122, 2001.
- [Kim S.H] KIM S.H., SUH J.W., KIM J.H. (1993), «Recognition of logic diagrams by identifying loops and rectilinear polylines», *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, Tsukuba, Japan, pp. 349-352, 1993.
- [Kise] KISE K., SATO A., IWATA M. (1998), «Segmentation of page images using the area Voronoi diagram», *Computer Vision and Image Understanding CVIU*, Vol 70: pp. 370-382, 1998.
- [Kiyko] KIYKO V.M. (1995), «Recognition of objects in images of paper based line drawings», *Proc. 3rd Int. Conf. on Document Analysis and Recognition*, pp. 970-973, Montréal, Canada, 1995.
- [Kopec] KOPEC G.E., CHOU P.A. (1994), «Document image decoding using Markov source models», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16: pp. 602-617, 1994.
- [Kosmala] KOSMALA A., LAVIROTTE S., POTTIER L., RIGOLL G. (1999), «On-line handwritten formula recognition using hidden Markov models and context dependent graph grammars», *Proc. 5th Int. Conf. on Document Analysis and Recognition*, pp. 107-110, Bangalore, India, 1999.
- [Kreich] KREICH J., LUHN A., MADERLECHNER G. (1991), «An experimental environment for model based document analysis», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 50-58, Saint-Malo, France, September 1991.
- [Krishnamoorthy] KRISHNAMOORTHY M. and NAGY G., SETH S., VISWANATHAN M. (1993), «Syntactic segmentation and labelling of digitized pages from technical journals», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15: pp. 737-747, 1993.
- [Kuner-1] KUNER P., UEBERREITER B. (1986), «Knowledge-based pattern recognition in disturbed line images using graph theory, optimization and predicate calculus», *Proc. 8th Int. Conf. on Pattern Recognition*, pp. 240-243, Paris, France, 1986.
- [Kuner-2] KUNER P., UEBERREITER B. (1986), «Pattern recognition by graph matching, combinatorial versus continuous optimization», *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol.2(N.3): pp. 527-542, 1988.
- [Kuo] KUO S., AGAZZI O.E. (1994), «Keyword spotting in poorly printed documents», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16(N.8): pp. 842-847, 1994.
- [Lam] LAM L., LEE S.W., SUEN C.Y. (1992), «Thinning methodologies - A comprehensive survey», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14(N.9): pp. 869-885, 1992.
- [Landay] LANDAY J.A., MYERS B.A. (2001), «Sketching interfaces: toward more human interface design», *IEEE Computer*, Vol.34(N.3): pp. 56-64, 2001.
- [Lavirotte] LAVIROTTE S., POTTIER L. (1997), «Optical formula recognition», *Proc. 4th Int. Conf. on Document Analysis and Recognition*, pp. 357-361, Ulm, Germany, 1997.
- [Leavers] LEAVERS V.F. (1993), «Survey - Which Hough Transform», *Computer Vision, Graphics and Image Processing*, Vol.58: pp. 250-264, 1993.
- [Lee H.J] LEE H.J., LEE M.C. (1994), «Understanding mathematical expressions using procedure-oriented transformation», *Pattern Recognition*, Vol.27(N.3): pp. 447-457, 1994.
- [Lee K.H] LEE K.H., EOM K.B., KASHYAP T.L. (1992), «Character recognition based on programmed attribute-dependent grammar», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14(N.11): pp. 1122-1128, 1992.
- [Lee S.] LEE S. (1992), «Recognizing handwritten electrical circuit symbols with attributed graph matching», *Structured Document Analysis*, Baird H.S. - Bunke H. - Yamamoto K. editors, Springer Verlag, pp. 340-358, 1992.
- [Lee S.W] LEE S.W., KIM J.H., GROEN F.C.A. (1990), «Translation-, rotation- and scale-invariant recognition of hand-drawn symbols in schematic diagrams», *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol.4(N.1): pp. 1-25, 1990.
- [Leu] LEU J.G., CHEN L. (1988), «Polygonal approximation of 2D shapes through boundary merging», *Pattern Recognition Letters*, Vol.7(N.4): pp. 231-238, 1988.
- [Leung] LEUNG M.K., HANG Y.H. (1990), «Dynamic two-step algorithm in curve fitting», *Pattern Recognition*, Vol.23(N.1): pp. 69-79, 1990.
- [Liang] LIANG J.S., PHILIPPS I.T., HARALICK M. (2001), «Performance evaluation of document structure extraction algorithms», *Computer Vision and Image Understanding*, Vol.84: pp. 144-159, 2001.
- [Lin X] LIN X., SHIMOTSUJI S., MINOH M., SAKAI T. (1985), «Efficient diagram understanding with characteristic pattern detection», *Computer Vision, Graphics and Image Processing*, Vol.30: pp. 84-106, 1985.
- [Lin C.C] LIN C.C., NIWA Y., NARITA S. (1997), «Logical structure analysis of book document images using contents information», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 1048-1054, Ulm, Germany, August 1997.
- [Lipari] LIPARI C.A., TRIVEDI M.M., HARLOW C.A. (1989), «Geometric modeling and recognition of elongated regions in aerial images», *IEEE Transactions on Systems, Man and Cybernetics*, Vol.19(N.6): pp. 1600-1612, 1989.
- [Lladós-1] LLADOS J., MARTI E., LOPEZ-KRAHE J. (1999), «A Hough-based method for hatched pattern detection in maps and diagrams», *Proc. 5th Int. Conf. on Document Analysis and Recognition*, pp. 479-482, Bangalore, India, 1999.
- [Lladós-2] LLADOS J., VALVENY E., SANCHEZ G., MARTI E. (2001), «Symbol recognition: current advances and perspectives», *Lecture Notes in Computer Science*, Vol.2390: pp. 104-127, 2001.
- [Lladós-3] LLADOS J., MARTI E., VILLANUEVA J. (2001), «Symbol recognition by error-tolerant subgraph matching between region adjacency graphs», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23: pp. 1137-1143, 2001.
- [Lladós-4] LLADOS J., VALVENY E., SANCHEZ G. (2003), «A case study of pattern recognition: symbol recognition in graphic documents», *Invited talk of Proc. 3rd Int. Workshop on Pattern Recognition in Information Systems*, pp. 1-11, 2003.
- [Lopez] LOPEZ KRAHE J., POUSSSET P. (1988), «The application of the Hough transform to the detection of roads», *Signal Processing IV*:

- Theories and Application*, Lacoume Chehikian Martin Malbos Eds, pp. 1673-1676, 1988.
- [Lopresti-1] LOPRESTI D., ZHOU J. (1998), «Document analysis and World wide web», *Document Analysis Systems*, Hull J.J. - Taylor S.L. editors, World Scientific, pp. 479-497, 1998.
- [Lopresti-2] LOPRESTI D., NAGY G. (2000), «A tabular survey of automated table processing», *Lecture Notes in Computer Science*, Vol.1941, pp. 93-120, 2000.
- [Lowe] LOWE D. (1987), «Three-dimensional object recognition from single two-dimensional images», *Artificial Intelligence*, Vol.31: pp. 355-395, 1987.
- [Lu] LU Z. (1998), «Detection of text regions from digital engineering drawings», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.20(N.4): pp. 431-439, 1998.
- [Luo-1] LUO H., DINSTEIN I. (1994), «Using directional mathematical morphology for separation of character strings from text/graphics images», *Proc. Int. Workshop on Syntactic and Structural Pattern Recognition*, pp. 372-381, Nahariya, Israël, 1994.
- [Luo-2] LUO H., KASTURI R. (1998), «Improved directional morphological operations for separation of characters from maps/graphics», *Lecture Notes in Computer Science*, Vol1389: pp. 35-47, 1998.
- [Madej] MADEJ D. (1991), «An intelligent map-to-CAD conversion system», *Proc. 1st Int. Conf. on Document Analysis and Recognition*, pp. 602-610, Saint-Malo, France, 1991.
- [Mao-1] MAO S., KANUNGO T. (2001), «Empirical performance evaluation methodology and its application to page segmentation algorithms», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.23: pp. 242-256, 2001.
- [Mao-2] MAO S., ROSENFELD A., KANUNGO T. (2003), «Document structure analysis algorithms: a literature survey», *Proc. SPIE Electronic Imaging*, Vol. 5010: pp. 197-207, January 2003.
- [Messmer] MESSMER B.T., BUNKE H. (1995), «Automatic learning and recognition of graphical symbols in engineering drawings», *Lecture Notes in Computer Science*, Vol.1072: pp. 123-134, 1995.
- [Min] MIN W., TANG Z., TANG L. (1993), «Using web grammar to recognize dimensions in engineering drawings», *Pattern Recognition*, Vol.26(N.9): pp. 1407-1416, 1993.
- [Murphy] MURPHY L.M. (1986), «Linear feature detection and enhancement in noisy images via the Radon transform», *Pattern Recognition Letters*, Vol.4(N.4): pp. 279-284, 1986.
- [Murthy] MURTHY B.K., DESHPANDE W.R. (1998), «Optical character recognition for indian languages», *Proc. Indian Conf. on Computer, Vision, Graphics and Image Processing*, New Delhi, India, 1998.
- [Miyao] MIYAO H., NAKANO Y. (1996), «Note symbol extraction for printed piano scores using neural network», *IEICE Trans. on Information and Systems*, Vol.E79-D(N.5): pp. 548-543, 1996.
- [Muller] MULLER S., RIGOLL G. (2000), «Engineering drawing database retrieval using statistical pattern spotting techniques», *Lecture Notes in Computer Science*, Springer Verlag, Vol.1941: pp. 246-255, 2000.
- [Myers] MYERS G., MULGAONKAR P., CHEN C., DE CURTIS J., CHEN E. (1996), «Verification-based approach for automated text and feature extraction from raster-scanned maps», *Lecture Notes in Computer Science*, Vol.1072: pp. 190-203, 1996.
- [Nagy-1] NAGY G., SETH S., VISWANATHAN M. (1992), «A prototype document image analysis system for technical journals», *Computer*, Vol.25: pp. 10-22, 1992.
- [Nagy-2] NAGY G. (2000), «Twenty years of document image analysis in PAMI», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.22(N.1): pp. 38-62, 2000.
- [Nartker] NARTKER T.A., RICE S.V., KANAI J. (1994), «OCR accuracy», *Technical report, Information Science Research institute of Nevada*, Las Vegas, 1994.
- [Newell] NEWELL A. (1982), «The knowledge level», *Artificial Intelligence*, Vol.19(N.2): pp. 87-127, 1982.
- [Niyogi] NIYOGI D., SRIHARI S.N. (1995), «Knowledge-based derivation of document logical structure», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 472-475, Montréal, Québec, Canada, August 1995.
- [Nonak] NONAK I. (1994), «Dynamic theory of organisation knowledge creation», *Organizational Science*, Vol.5(N.1): pp. 14-37, 1994.
- [O'Gorman-1] O'GORMAN L. (1992), «Image and document processing techniques for the righthpages electronic library system», *Int. Conf on Pattern Recognition ICPR*, Alamitos, California, USA, pp. 260-263, 1992.
- [O'Gorman-2] O'GORMAN L. (1993), «The document spectrum for page layout analysis», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.15: pp. 1162-1173, 1993.
- [O'Gorman-3] O'GORMAN L. (1994), «Binarization and multithresholding of document images using connectivity», *Graphical Models and Image Processing*, Vol.56(N.6): pp. 494-506, 1994.
- [Okazaki] OKAZAKI A., KONDO T., MORI K., TSUNEKAWA S., KAWAMOTO E. (1988), «A automatic circuit diagram reader with loop-structure-based symbol recognition», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.10(N.3): pp. 331-341, 1988.
- [Otsu] OTSU N. (1979), «A threshold selection method from gray-level histograms», *IEEE Tran. Systems, Man and Cybernetics*, Vol.9(N.1): pp. 62-66, 1979.
- [Paquet] PAQUET T., LECOURTIER Y. (1993), «Recognition of handwritten sentences using a restricted lexicon», *Pattern Recognition*, Vol.26(N.3): pp. 391-407, 1993.
- [Park] PARK Y., KWON Y. (2002), «A effective vector extraction method on architectural imaging using drawing characteristics», *Lecture Notes in Computer Science*, Springer Verlag, Vol.2390: pp. 299-309, 2002.
- [Parker] PARKER J.R., PIVOVAROV J., ROYKO D. (2000), «Vector templates for symbol recognition», *Proc. 15th Int. Conf. on Pettern Recognition*, pp. 602-605, Barcelona, Spain, 2000.
- [Parmentier] PARMENTIER F. (1998), «Spécificité d'une architecture émergente fondée sur le raisonnement par analogie. Applications aux références bibliographiques», *Thèse de doctorat de l'Université Henri Poincaré - Nancy I*, 1998.
- [Pasternak] PASTERNAK B. (1994), «Processing imprecise and structural distorted line drawings by an adaptable drawing interpretation system», *Document Analysis System*, Dengel A. - Spitz L. editors] World Scientific, pp. 349-365, 1994.
- [Pavlidis-1] PAVLIDIS T., ZHOU J. (1992), «Page segmentation and classification», *Graphical models and image processing*, Vol.54: pp. 484-496, 1992.
- [Pavlidis-2] PAVLIDIS T., WANG L. (1993), «Direct gray-scale extraction of features for character recognition», *IEEE Transactions on Pattern nalysis and Machine Intelligence*, Vol.15(N.10): pp. 1053-1067, 1993.
- [Pei] PEI S.C., HORNG J.H. (1995), «Fitting digital curve using circular arcs», *Pattern Recognition*, Vol.28(N.1): pp. 107-116, 1995.
- [Ramel] RAMEL, J.Y., BOISSIER, G., EMPTOZ, H. (2000), «A structural representation adapted to handwritten symbol recognition», *Lecture Notes in Computer Science*, Springer Verlag, Vol.1941: pp. 228-237, 2000.
- [Ramer] RAMER U. (1972), «A iterative procedure for the polygonal approximation of the plane curves», *Computer, Graphics and Image Processing*, Vol.1: pp. 244-256, 1972.
- [Randriamahefa] RANDRIAMAHEFA R., COCQUEREZ J., FLUHR C., PPIN F., PHILIPP S. (1993), «Printed music recognition», *Proc. 2nd Int. Conf. on Document Analysis and Recognition*, pp. 898-901, Tsukuba, Japan, 1993.
- [Reiher] REIHER E., LI Y., DONNE V.D., LALONDE M., HAYNE C., ZHU C. (1996), «A system for efficient and robust map symbol recognition», *Proc. 13th Int. Conf. on Pattern Recognition*, pp. 783-787, Vienna, Austria, 1996.
- [Rice-1] RICE S.V., KANAI J., NARTKER T.A. (1992), «A report on the accuracy of OCR devices», *Technical report, Information Science Research institute of Nevada*, Las Vegas, 1992.
- [Rice-2] RICE S.V., JENKINS F.R., NARTKER T.A. (1996), «The fifth annual test of OCR occuracy», *Technical report, Information Science Research institute of Nevada*, Las Vegas, 1996.
- [Robadey] ROBADEY L. (2001), «2(Crem): une méthode de reconnais-

- sance structurelle de documents complexes basée sur les patterns bidimensionnels», *Thèse de doctorat, Faculté des Sciences de l'Université de Fribourg, Suisse, Décembre 2001.*
- [Rocha-1] ROCHA J., PAVLIDIS T. (1994), «A shape analysis model with applications to a character recognition system», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.16(N.4): pp. 393-404, 1994.
- [Rocha-2] ROCHA J., PAVLIDIS T. (1995), «Character recognition without segmentation», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17(N.9): pp. 903-909 1995.
- [Röösli] RÖÖSLI P.L., MONAGAN G. (1996), «Adding geometric constraints to the vectorization of line drawings», *Lecture Notes in Computer Science*, Springer Verlag, Vol.1072: pp. 49-56, 1996.
- [Rosin] ROSIN P.L. (1997), «Techniques for assessing polygonal approximation of curves», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(N.6): pp. 659-666 1997.
- [Sahoo] SAHOO P.K., and al. (1988), «A survey of thresholding techniques», *Computer Vision, Graphics and Image Processing*, Vol.41(N.2): pp. 233-260, 1988.
- [Saitoh] SAITOH T., TACHIKAWA M., YAMAAI T. (1993), «Document image segmentation and text area ordering», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 323-329, Tsukuba Science City, Japon 1993.
- [Samet] SAMET H., SOFFER A. (1996), «MARCO: Map retrieval by content», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.18(N.8): pp. 783-797, 1996.
- [Sanchez] SANCHEZ G., LLADOS J. (2001), «A graph grammar to recognize textured symbols», *Proc. 6th Int. Conf. on Document Analysis and Recognition*, pp. 465-469, Seattle, WA, USA, 2001.
- [Sandakly] SANDAKLY F. (1995), «Contribution à la mise en œuvre d'une architecture à base de connaissances pour l'interprétation de scènes 2D et 3D», *Thèse de l'Université de Nice, Sophia Antipolis*, 1995.
- [Sauvola] SAUVOLA J., PIETIKAINEN M. (2000), «Adaptive document image binarization», *Pattern Recognition*, pp. 225-236, 2000.
- [Sawaki] SAWAKI M., HAGITA K. (1998), «Text-line extraction and character recognition of document headlines with graphical design using complimentary similarity measure», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20: pp. 1103-1109, 1998.
- [Segen] SEGEN J. (1989), «From features to symbols: learning relational models of shape», *From Pixels to Features*, Simon J.C. editor, Elsevier Science Publishers, pp. 237-248, 1989.
- [Shinghal-1] SHINGHAL R., TOUSSAINT G.T. (1979), «Experiments in text recognition with the modified Viterbi algorithm», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.1(N.2): pp. 184-192, 1979.
- [Shinghal-2] SHINGHAL R., TOUSSAINT G.T. (1980), «The sensitivity of the modified Viterbi algorithm to the source statistics», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.2(N.2): pp. 1181-1184, 1980.
- [Sinha] SINHA R.M.K., PRASADA B., HOULE G.H., SABOURIN M. (1993), «Hybrid contextual text recognition with string matching», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15(N.9): pp. 915-923, 1993.
- [Soffer] SOFFER A., SAMET H. (1998), «Using negative shape features for logo similarity matching», *Proc. 14th Int. Conf. on Pattern Recognition*, pp. 571-573, Brisbane, Australia, 1998.
- [Spitz] SPITZ A.L. (1997), «Determination of the script and language content of document images», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(N.3): pp. 235-245, 1997.
- [Srihari-1] SRIHARI S.N., GOVINDARAJU V. (1989), «Analysis of textual images using the Hough transform», *Machine Vision and Applications*, Vol.2: pp. 141-153, 1989.
- [Srihari-2] SRIHARI S.N., YANG W., GOVINDARAJU V. (1999), «Information theoretic analysis of postal address fields for automatic address interpretation», *Proc. Int. Conf. on Document Analysis and Recognition*, pp. 309-312, Bangalore, India, 1999.
- [Stentiford] STENTIFORD F.W.M. (1985), «Automatic feature design for optical character recognition using an evolutionary search procedure», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.7(N.3): pp. 349-354, 1985.
- [Stringa] STRINGA L. (1990), «A new set of constraint-free character recognition grammars», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12(N.12): pp. 1210-1216, 1990.
- [Suda] SUDA P., BRIDOUX C., KAMMERER, MADERLECHNER G. (1997), «Logo and word matching using a general approach to signal registration», *Proc. 4th Int. Conf. on Document Analysis and Recognition*, pp. 61-65, Ulm, Germany, 1997.
- [Suen] SUEN C.Y. (1979), «N-gram statistics for natural language understanding and text processing», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.1(N.2): pp. 164-172, 1979.
- [Summers] SUMMERS K. (1995), «Near-wordless document structure classification», *Proc. Int. Conf. on Document Analysis and Recognition*, Montréal, Québec, Canada, pp. 462-465, 1995.
- [Tabbone] TABBONE S., WENDLING L. (2000), «Décomposition graphique sous forme de primitives 2D», *Actes du deuxième colloque international francophone sur l'écrit et le document*, pp. 131-140, Lyon, France, juillet 2000.
- [Tan T.N] TAN T.N. (1998), «Rotation invariant texture features and their use in automatic script identification», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20(N.7): pp. 751-756, 1998.
- [Tan C.L] TAN C.L., CAO R., SHEN P. (2002), «Restoration of archival documents using a wavelet technique», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24(N.10): pp. 1399-1404, 2002.
- [Tang Y.T] TANG Y.T., MA H., LIU Y., XI D. (1997), «Multiresolution analysis in extraction of references lines from documents with gray level background», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(N.8): pp. 921-925, 1997.
- [Tang Y.Y] TANG Y.Y., SUEN C.Y., YAN C.D., CHERIET M. (1991), «Document analysis and understanding: a brief survey», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, Saint-Malo, France, pp. 17-31, 30 septembre - 2 octobre 1991.
- [Tateisi] TATEISI Y., ITOH N. (1994), «Using stochastic syntactic analysis for extracting a logical structure from a document image», *Proc. Int. Conf. on Pattern Recognition*, Jerusalem, Israël, pp. 391-394, 1994.
- [Thomas] THOMAS S.M., CHAN Y.T. (1989), «A simple approach for the estimation of circular arc center and its radius», *Computer Vision, Graphics and Image Processing*, Vol.45: pp. 362-370, 1989.
- [Tokuyasu] TOKUYASU T.A., CHOU P.A. (2001), «Turbo recognition: a statistical approach to layout analysis», *Proc. Conf. on Document Recognition and Retrieval*, pp. 123-129, San José, CA, USA, January 2001.
- [Tombre-1] TOMBRE K., CHHABRA A.K. (1998), «Graphics recognition - Algorithms and systems», *Volume 1389 of Lectures Notes in Computer Science*, Springer Verlag, April 1998.
- [Tombre-2] TOMBRE K. (1998), «Ten Years of Research in the Analysis of Graphics Documents, Achievements and Open Problems», *Portuguese Conference on Pattern Recognition, 1998. Image Processing and Image Understanding*, 1998.
- [Tombre-3] TOMBRE K., TABBONE S. (2000), «Vectorization in graphics recognition: to thin or not to thin», *Proc. 15th Int. Conf. on Pattern Recognition*, pp. 91-96, Barcelona, Spain, 2000.
- [Tombre-4] TOMBRE K., TABBONE S., PÉLISSIER L., LAMIROY B., DOSH P. (2002), «Text/graphics separation revisited», *Lecture Notes in Computer Science*, Vol.2423: pp. 200-211, 2002.
- [Trier-1] TRIER O.D., TAXT T. (1995), «Improvement of integrated function algorithm for binarization of document images», *Pattern Recognition Letters*, Vol.16: pp. 277-283, 1995.
- [Trier-2] TRIER O.D., TAXT T. (1995), «Evaluation of binarization methods for document images», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17(N.3): pp. 312-315, 1995.
- [Trier-3] TRIER O.D., JAIN A.K., TAXT T. (1996), «Feature extraction methods for character recognition - A survey», *Pattern Recognition*, Vol.29(N.4): pp. 641-662, 1996.
- [Tsumimoto-1] TSUJIMOTO S., ASADA H. (1990), «Understanding multi-articled documents», *Proc. Int. Conf. on Pattern Recognition*, pp. 551-556, Atlantic City, NJ, USA, 1990.

- [Tsujiimoto-2] TSUJIMOTO S., ASADA H. (1992), «Major components of a complete text reading system», *Proc. IEEE*, vol80(n.7): pp. 1133-1149, 1992.
- [Valveny] VALVENY E., MARTI E. (2000), «Hand-drawn symbol recognition in graphic documents using deformable template matching and a bayesian framework», *Proc. 15th Int. Conf. on Pattern Recognition*, pp. 239-242, Barcelona, Spain, 2000.
- [Vaxivière-1] VAXIVIÈRE P., TOMBRE K. (1995), "CELESTIN: CAD conversion of mechanical drawings», *Computer*, Vol.25(N.7): pp. 46-54, 1992.
- [Vaxivière-2] VAXIVIÈRE P., TOMBRE K. (1995), « Knowledge organization and interpretation process in engineering drawing interpretation», *Document Analysis System*, Spitz and Dengel Eds, World Scientific, pp. 307-317, 1995.
- [Wahl] WAHL F., WONG K., CASEY R. (1982), «Block segmentation and text extraction in mixed text/image documents», *Graphics Models and Image Processing*, Vol.20: pp. 375-390, 1982.
- [Wall] WALL K., DANIELSSON P.E. (1984), «A fast sequential method for polygonal approximation of digitized curves», *Computer Vision, Graphics and Image Processing*, Vol.28: pp. 220-227, 1984.
- [Wang-1] WANG Q.R., SUEN C.Y. (1986), « Analysis and design of a decision tree based on entropy reduction and its application to large character set recognition», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.6(N.4): pp. 406-417, 1986.
- [Wang-2] WANG Q.R., SUEN C.Y. (1987), «Large tree classifier with heuristic search and global training», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.9(N.1): pp. 91-102, 1987.
- [Wang Y] WANG Y., PHILLIPS I.T., HARALICK R. (2001), «Using area Voronoi tessellation to segment characters connected to graphics», *Proc. 4th Int. Workshop on Graphics Recognition*, pp. 147-153, Kingston, Ontario, Canada, 2001.
- [Watanabe] WATANABE T., LUO Q., SUGIE N. (1995), «Layout recognition of multi-kinds of table-form documents», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17(N.4): pp. 432-445, 1995.
- [White] WHITE J.M., ROHRER G.D. (1983), «Image thresholding for optical character recognition and other applications requiring character image extraction», *IBM Journal of research and development*, Vol.27(N.4): pp. 400-411, 1983.
- [Wilfong] WILFONG G., SINDEN F., RUEDISUELI L. (1996), «On-line recognition of handwritten symbols», *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.18(N.9): pp. 935-940, 1996.
- [Wilson] WILSON C.L., GEIST J., GARRIS M.D., CHELLAPA R. (1996), «Design, integration and evaluation of form-based handprint and OCR systems», *Technical report, NISTIR5932, National Institute of Standards & Technology*, <http://www.itl.nist.gov/iad/894.03/pubs.html>, 1983.
- [Yadid-Pecht] YADID-PECHT O., GERNER M., DVIR L., BRUTMAN E., SHIMONY U. (1996), «Recognition of handwritten musical notes by a modified neocognitron», *Machine Vision and Applications*, Vol.9: pp. 65-72, 1996.
- [Yamashita] YAMASHITA A., AMANO T., TAKAHASHI I., TOYOKAWA K. (1991), «A model based layout understanding method for the document recognition system», *Proc. Int. Conf. on Document Analysis and Recognition ICDAR*, pp. 130-138, Saint-Malo, France, September 1991.
- [Yang D] YANG D., WEBSTER J.L., RENDELL L.A., GARRET J.H., SHAW D.S. (1993), «Management of graphical symbols in a CAD environment: a neural network approach», *Proc. Int. Conf. on Tools with Artificial Intelligence*, pp. 272-279, Boston, Massachusetts, USA, 1993.
- [Yang Y] YANG Y., HAN H. (2000), «An adaptive logical method for binarization of degraded document images», *Pattern Recognition*, vol.33: pp. 787-807, 2000.
- [Yanikoglu] YANIKOGLU B.A., VINCENT L. (1998), «Pink panther: a complete environment for ground-truthing and benchmarking document page segmentation», *Pattern Recognition*, vol31, pp. 1191-1204, 1998.
- [Yoo] YOO J.H., SETHI I.K. (1993), «An ellipse detection method from the polar and pole definition of conics», *Pattern Recognition*, Vol.26(N.2): pp. 307-315, 1993.
- [Yu-1] YU B. (1995), «Automatic understanding of symbol-connected diagrams», *Proc. 3rd Int. Conf. on Document Analysis and Recognition*, pp. 803-806, 1995.
- [Yu-2] YU B., JAIN A.K. (1996), «A generic system for form dropout», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.18(N.11): pp. 1127-1134, 1996.
- [Yu-3] YU B., SAMAL A., SETH C. (1997), «A system for recognizing a large class of engineering drawings», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19(N.8): pp. 868-890, 1997.
- [Yuen] YUEN S.Y.K., LAM T.S.L., LEUNG N.K.D. (1993), «Connective Hough transform», *Image and Vision Computing*, Vol.11: pp. 295-301, 1993.
- [Zannibi-1] ZANNIBI R., BLOSTEIN D., CORDY J.R. (2002), «Recognizing Mathematical Expressions Using Tree Transformation», *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.24(N.11): pp. 1455-1467, 2002.
- [Zannibi-2] ZANNIBI R., BLOSTEIN D., CORDY J.R. (2004), «A survey of table recognition. Models, observations, transformation and inferences», *Int. Journal on Document Analysis and Recognition*, Vol.7(N.1): pp. 1-16, 2004.
- [Zhu] ZHU (1999), «Stochastic jump - Diffusion process for computing medial axes in Markov random fields», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.21(N.11): pp. 1158-1169, 1999.
- [Zramdini] ZRAMDINI A., INGOLD R. (1998), «Optical font identification using typographic features», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20(N.8): pp. 877-882, 1998.



Éric Trupin

Eric Trupin est né en 1963 à Boulogne Sur Mer. Après avoir passé son doctorat en 1993 à l'Université de Rouen, il y fut nommé Maître de Conférences la même année. Il a soutenu une Habilitation à Diriger des Recherches en 2003. Ses activités concernent l'interprétation d'images de documents structurés et techniques au sein du laboratoire PSI. Ses intérêts portent sur l'analyse des structures de documents, la reconnaissance de symboles, la modélisation et l'acquisition de connaissance pour l'implémentation de plate-formes génériques d'interprétation d'images. Ils visent également à s'appliquer à la fouille, à la classification et au tri automatique de documents.



