



IRISA • Campus universitaire de Beaulieu • 35042 Rennes Cedex France • Tél. : +33 2 99 84 71 00 • Télécopie : +33 2 99 84 71 71 • Internet : [www.iris.fr](http://www.iris.fr)

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

## *Projet Imadoc*

*Interprétation et Reconnaissance d'Images et de Documents*

*Rennes*

————— THÈME 3B —————





## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>3</b>
<b>2</b>	<b>Fondements scientifiques</b>	<b>3</b>
2.1	Vision précoce . . . . .	3
2.2	DMOS, une méthode générique de reconnaissance de documents structurés . . . . .	6
2.3	Reconnaissance des formes par systèmes d'inférence floue . . . . .	9
2.4	Modèle et système de perception et d'interprétation (SPI) . . . . .	10
<b>3</b>	<b>Domaines d'applications</b>	<b>13</b>
3.1	Analyse de documents papier . . . . .	13
3.2	Reconnaisseurs d'écriture manuscrite . . . . .	13
3.3	Interfaces orientées <i>stylet</i> . . . . .	14
<b>4</b>	<b>Logiciels</b>	<b>15</b>
4.1	Resif : Reconnaissance d'Écriture Manuscrite en-ligne par Systèmes d'Inférence Floue	15
4.2	DocRead : générateur automatique de systèmes de reconnaissance de documents structurés . . . . .	15
4.3	Bibliothèque d'extraction d'indices visuels . . . . .	16
<b>5</b>	<b>Résultats nouveaux</b>	<b>16</b>
5.1	Annotations pour l'accès aux documents d'archives . . . . .	16
5.2	Détection des pages dans des registres paroissiaux et d'état civil . . . . .	17
5.3	Application de la méthode DMOS aux registres matricules . . . . .	18
5.4	Application de la méthode DMOS à des décrets de naturalisation . . . . .	19
5.5	Application de la méthode DMOS à la reconnaissance de terrain de tennis dans une séquence vidéo . . . . .	21
5.6	Méthode générique de segmentation de symboles . . . . .	21
5.7	Bibliothèque virtuelle de typographie . . . . .	23
5.8	Système de classification hybride interprétable par construction automatique de systèmes d'inférence floue . . . . .	23
5.9	Embarquement du système ResifCar sur des téléphones mobiles . . . . .	25
5.10	Intégration et modélisation de connaissances lexicales pour la reconnaissance en-ligne de l'écriture manuscrite . . . . .	26
5.11	Système générique SPI, interface orientée stylet, saisie de formules mathématiques sur machine de poche . . . . .	28
5.12	Graphe de voisinage basé sur l'utilisation des distances discrètes . . . . .	30
<b>6</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>32</b>
6.1	Embarquement du système ResifCar sur des téléphones mobiles . . . . .	32
6.2	Annotation collective de documents d'archives . . . . .	32
6.3	Accès en ligne aux documents d'archives . . . . .	32

---

<b>7</b>	<b>Actions régionales, nationales et internationales</b>	<b>33</b>
7.1	Actions régionales . . . . .	33
7.2	Actions nationales . . . . .	33
7.2.1	ICL . . . . .	33
7.2.2	RTP . . . . .	33
7.2.3	GdR . . . . .	34
7.2.4	Sociétés savantes . . . . .	34
7.2.5	Listes de discussion . . . . .	34
7.3	Actions internationales . . . . .	34
<b>8</b>	<b>Diffusion de résultats</b>	<b>35</b>
8.1	Animation de la Communauté scientifique . . . . .	35
8.1.1	Comité éditorial de journaux . . . . .	35
8.1.2	Comités de programme de conférences . . . . .	35
8.2	Enseignement . . . . .	35
8.3	Participation à des colloques, séminaires, invitations . . . . .	36
<b>9</b>	<b>Bibliographie</b>	<b>37</b>

# 1 Composition de l'équipe

## Responsable scientifique

Jean Camillerapp [professeur Insa]

## Assistante de projet

Édith Blin [TR Inria]

## Personnel Inria

Jacques André [DR]

Bertrand Coüasnon [CR en détachement depuis le 01/09/2002]

## Personnel Insa

Éric Anquetil [maître de conférences]

Bertrand Coüasnon [maître de conférences jusqu'au 31/08/2002]

Ivan Leplumey [maître de conférences]

Yann Ricquebourg [maître de conférences depuis le 1/09/2002]

## Personnel Université de Rennes 1

Guy Lorette [professeur]

Charles Quéguiner [maître de conférences]

Hélène Richy [maître de conférences]

Laurent Pasquer [ATER jusqu'au 31/08/2002]

## Chercheurs doctorants

Nicolas Ragot [boursier université]

Sabine Carbonnel [boursière ministère depuis le 1/10/2002]

## Ingénieurs Experts

Nicolas Briec [sur contrat région Mégalis]

Laurent Pasquer [sur contrat département des Yvelines depuis le 1/10/2002]

Grégory Maitrallain [sur contrat département des Yvelines depuis le 1/10/2002]

# 2 Fondements scientifiques

## 2.1 Vision précoce

**Mots clés** : vision précoce, binarisation, squelettisation, filtrage de Kalman.

### Résumé :

*Afin de réduire les volumes de stockage et les temps de traitement, les documents sont généralement numérisés avec la densité la plus faible possible (souvent 200 dpi). Dans ces conditions, les éléments du tracé n'ont que quelques pixels d'épaisseur. Pour répondre aux exigences des procédures de reconnaissance, les traitements d'image doivent être capables de fournir une binarisation fiable des petits éléments, une extraction du squelette à partir d'images en niveaux de gris, et une détection des segments de droite même en présence de recouvrement et d'intersection.*

### Binarisation adaptative

Les documents sont généralement numérisés en limite de résolution et certains éléments très importants pour la reconnaissance peuvent n'occuper que quelques pixels. Ainsi, par exemple, la non détection d'une toute petite zone blanche peut masquer la présence, pourtant très informante, de la boucle d'un *e*.

Une numérisation en niveaux de gris des images de document permet de mettre en œuvre des algorithmes de binarisation adaptative, souvent basés sur des coopérations entre la détection des contours et le grossissement des régions. Ceux-ci donnent une interprétation locale des valeurs des niveaux de gris qui séparent bien les différents objets dans le document (cf figures 1 et 2).

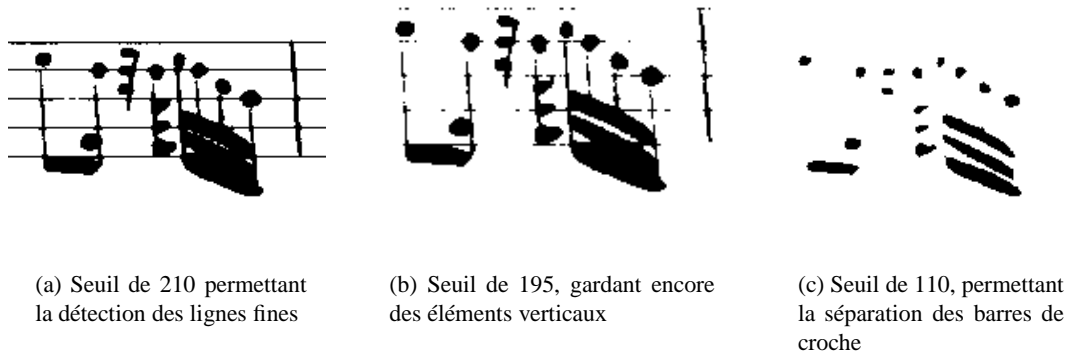


FIG. 1 – Binarisation avec un seuil fixe de l'image en niveaux de gris de la figure 2

Dans ces conditions, en limite de résolution, les mesures des caractéristiques des objets dans des images binaires sont sensibles au bruit provenant de l'échantillonnage spatial. L'interprétation locale des niveaux de gris permet de positionner la transition entre les objets et le fond avec une précision meilleure que le pixel, qualifiée de *sub-pixelle* par la suite. En augmentant ainsi la résolution spatiale, on améliore la pertinence des paramètres liés aux indices visuels.

### Squelettisation d'images en niveaux de gris

Destinée au départ à l'affinage d'images de ligne, les techniques de squelettisation se sont par la suite développées en oubliant un peu cet objectif. Dans les images de documents, il nous semble important de revenir à la notion primitive d'axe médian pour modéliser certaines parties du tracé. Il faut alors distinguer (figure 3) les zones réellement squelettisables ou zones régulières qui correspondent à un tracé simple et les zones singulières localisées dans les portions du tracé qui comportent des recouvrements et dans lesquelles la notion théorique d'axe médian n'a pas d'intérêt pratique [10].

Afin de limiter le bruit de binarisation évoqué ci-dessus, on peut obtenir une localisation sub-pixelle des points de cet axe en travaillant à partir d'images en niveaux de gris.

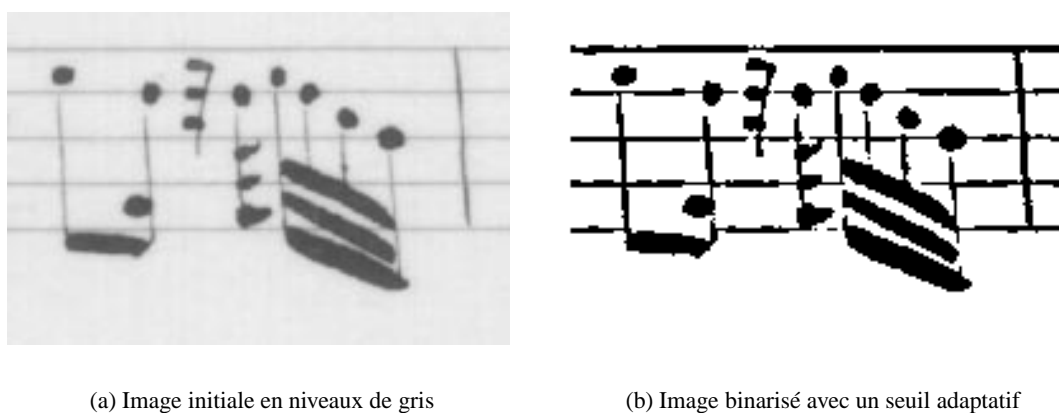


FIG. 2 – Binarisation adaptative

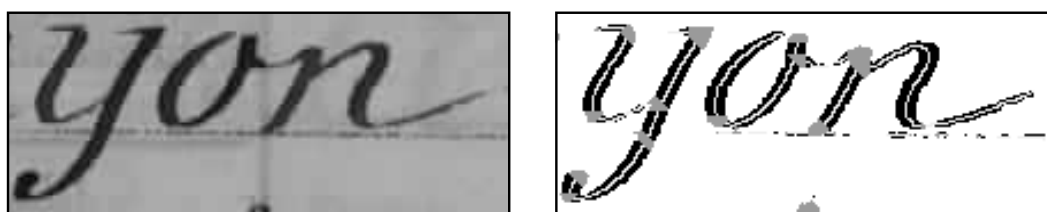


FIG. 3 – Traitement d'une image issue d'un document ancien, le squelette apparaît en blanc au milieu du tracé et les régions singulières sont en gris foncé

### Segmentation par filtrage de Kalman

Les structures linéaires tiennent une grande place dans les documents mais elles sont susceptibles d'interférer entre elles ou avec le tracé lui-même. Cependant la simplicité de leur structure permet d'envisager de conduire simultanément la segmentation et la reconnaissance.

Le filtrage de Kalman est une technique d'identification des paramètres d'un modèle à partir d'une suite ordonnée de mesures. Dans le cas des tracés rectilignes, le modèle se réduit à l'épaisseur du trait, sa pente et l'équivalent de l'ordonnée à l'origine. La mesure se déduit de la position et de l'épaisseur des empan noirs <sup>1</sup> dans une direction approximativement orthogonale au tracé.

Le filtrage de Kalman calcule également la matrice de covariance de son estimation des paramètres du modèle. Cette matrice permet d'évaluer d'une part la vraisemblance de l'affectation d'une mesure à un segment et d'autre part la vraisemblance de la poursuite d'une hypothèse perturbée par la présence d'un autre objet [4].

La véritable difficulté de cette approche ne réside pas dans les équations de Kalman, mais dans le processus d'association des empan extraits de l'image avec les empan prédits lorsque les structures linéaires se coupent ou se chevauchent (figure 4). Un même empan de l'image peut être le résultat du recouvrement de plusieurs tracés. C'est à ce niveau que l'information apportée par les matrices de

<sup>1</sup>Ensemble de pixels noirs consécutifs selon une des quatre directions : horizontale, verticale ou diagonales.

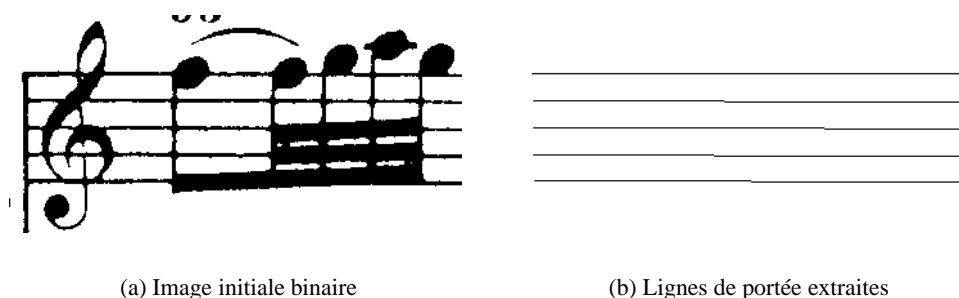


FIG. 4 – Exemple d'utilisation du filtrage de Kalman pour l'extraction des lignes de portées traversant des barres de croche

covariance est d'une grande importance.

Dans cette approche nous pouvons également utiliser une localisation des extrémités des empanns fournie par la binarisation adaptative.

## 2.2 DMOS, une méthode générique de reconnaissance de documents structurés

**Mots clés :** analyse structurelle, grammaire, Definite Clause Grammar (DCG), segmentation, gestion des connaissances a priori, reconnaissance de documents, partitions musicales, formules mathématiques, tableaux, formulaires.

**Résumé :** *Afin d'éviter de réécrire complètement un nouveau système de reconnaissance à chaque nouveau type de documents, nous proposons une méthode générique de reconnaissance de document : DMOS. Cette méthode est constituée du langage EPF permettant de modéliser grammaticalement la connaissance a priori, et d'un analyseur associé capable de remettre en cause, au cours de l'analyse, la structure analysée tout en gérant correctement le bruit très présent en reconnaissance de documents.*

La conception d'un système de reconnaissance de documents est longue et difficile. Or nous constatons dans l'état de l'art que l'adaptation d'un système de reconnaissance à un nouveau type de documents est impossible. Elle nécessite en fait à chaque fois une réécriture complète du système, ce qui correspond à une perte d'énergie très importante. La conception d'un système générique de reconnaissance de documents constitue un défi important, car il permettrait d'éviter ce travail coûteux de réécriture.

En outre, l'augmentation de la qualité de la reconnaissance, nécessaire dans un contexte de traitement industriel de documents, peut se faire notamment en résolvant les problèmes de segmentation que l'on rencontre souvent dans les documents à traiter.

Nous pouvons finalement constater que ces problèmes de reconnaissance de documents se situent autour de la gestion des connaissances *a priori*. En effet :

- un système générique doit permettre de ne modifier que ce qui est nécessaire lors de l'adaptation à un nouveau type de document, c'est-à-dire la connaissance *a priori* sur le document ;



- pour résoudre les problèmes de segmentation, il faut introduire de la connaissance *a priori* dans le système.

Nous avons donc proposé, pour des documents à forte structure dans lesquels des règles d'écriture peuvent être connues, la méthode DMOS (Description avec MODification de la Segmentation), constituée de :

- un langage grammatical de description de documents, EPF (*Enhanced Position Formalism*), permettant de modéliser la connaissance *a priori* ;
- l'analyseur associé autorisant une modification en cours d'analyse de la structure analysée. Cette modification permet d'introduire le contexte (niveau symbolique) dans la phase de segmentation (niveau numérique), afin d'améliorer la reconnaissance.

### Le langage EPF

Un certain nombre de formalismes grammaticaux permettant de décrire des objets bidimensionnels ont été proposés. Cependant, soit ils offrent une expressivité trop faible (grammaires d'arbres, *Web grammars*), soit ils ont une syntaxe trop compliquée (grammaires Plex, grammaires de graphes), ce qui rend très difficile la mise en œuvre de connaissances complexes. En outre, aucun ne permet d'introduire la connaissance formalisée *a priori*, dans la phase de segmentation.

Nous avons donc défini et développé le langage EPF[16], permettant de décrire un document structuré aussi bien au niveau graphique que syntaxique. Ce langage peut être vu comme une extension bidimensionnelle des grammaires classiques, où au lieu d'être des caractères, les terminaux sont des segments ou des matrices de pixels. Cette extension est également faite à l'aide d'un certain nombre d'opérateurs spécifiques dont voici quelques exemples :

**Opérateur de position** (argument de AT):

```
A && AT(pos) && B
```

signifie qu'on trouve A, et à la position `pos` relativement à A, on trouve B,

où `&&` est la concaténation dans la grammaire, A et B représentent un terminal ou un non-terminal.

**Opérateur de factorisation** (`##`, en association avec les opérateurs de position):

```
A && (AT(pos1) && B ##
      AT(pos2) && C)
```

signifie : A && AT(pos1) && B et A && AT(pos2) && C

Grâce à cette syntaxe, il est possible, par exemple de décrire un groupe de notes simplifié (des croches reliées par une seule barre de groupe) :

```
groupeDeNote ::= barreDeGroupe &&
              (AT(extremiteGauche) && noteGr ##
               notesAuMilieu ##
               AT(extremiteDroite) && noteGr).
```

où `::=` est le constructeur d'une règle grammaticale. Le concepteur de la grammaire peut définir à la demande des opérateurs de position, de la même manière qu'il peut le faire pour des non-terminaux.

**Opérateurs de référence multiple** (`--->` et `<---`): Afin de pouvoir faire référence plusieurs fois à une même instance du terminal ou du non-terminal A, nous proposons de sauvegarder cette instance à l'aide de `--->`. Cette sauvegarde permet ensuite de faire référence (à l'aide de `<---`) à A autant de fois que nécessaire. Si cet opérateur n'existait pas, une deuxième référence à A ne représenterait pas le même élément, mais une nouvelle instance. Il est ainsi possible de décrire ce qu'est un rectangle :

```
rectangle ::=
  (segV ---> segCoteGauche) &&
  AT(extremiteHaute) && segH &&
  AT(extremiteDroite) && segV &&
  AT(extremiteBasse) && segH &&
  AT(extremiteGauche) &&
  (segV <--- segCoteGauche).
```

### Analyseur associé

Le langage EPF décrit ci-dessus permet de définir grammaticalement le document à reconnaître. À partir de cette grammaire, nous produisons automatiquement un analyseur qui possède des caractéristiques spécifiques à l'analyse de documents bidimensionnels. Nous pouvons souligner les trois principales caractéristiques de l'analyseur que nous avons développé (à deux dimensions), par rapport à un analyseur classique (à une dimension) pour les langages formels :

- remise en cause de la structure analysée en cours d'analyse pour effectuer des segmentations contextuelles. C'est-à-dire que l'analyseur utilise le contexte pour modifier la structure analysée, pendant l'analyse, afin que cette structure soit dans l'état dans lequel elle aurait du être s'il n'y avait pas eu de problème de segmentation ;
- détection de l'élément suivant à analyser. En effet pour les analyseurs classiques, l'élément suivant est simplement celui qui est en tête de la chaîne analysée alors qu'en deux dimensions, l'élément suivant peut être n'importe où dans l'image, donc n'importe où dans la structure analysée ;
- gestion correcte du bruit. Contrairement aux analyseurs classiques où la structure analysée est peu bruitée, en reconnaissance de documents il est nécessaire que l'analyseur soit capable de reconnaître le maximum d'informations dans un flux très bruité. Nous pouvons considérer que la gestion du bruit correspond à trouver l'élément suivant, malgré le bruit.

En conclusion, nous sommes arrivés, grâce à la définition et la mise en œuvre du formalisme EPF et de l'analyseur associé, à concevoir un système générique de reconnaissance de documents structurés.

La création d'un nouveau système de reconnaissance adapté à un nouveau type de document s'obtient alors par simple compilation d'une description du document réalisée avec le langage EPF. Éventuellement, un nouvel apprentissage automatique du classifieur servant à reconnaître les symboles terminaux (représentés par des matrices de pixels), peut être nécessaire pour de nouveaux symboles.

La méthode DMOS a ainsi permis de développer des systèmes de reconnaissance de partitions musicales, de tableaux, de formulaires et de formules mathématiques, en définissant uniquement une description du document.

### 2.3 Reconnaissance des formes par systèmes d'inférence floue

**Mots clés :** modélisation explicite des connaissances, systèmes d'inférence floue, apprentissage, classification.

**Résumé :** *Dans le domaine de la reconnaissance des formes, la modélisation de systèmes complexes engendre généralement l'interaction de plusieurs processus interdépendants. Face à la complexité d'interaction des différents maillons qui composent un tel système, il est important de pouvoir maîtriser chacun des concepts mis en jeu. Dans cet objectif, les systèmes d'inférence floue (SIF) permettent la conception de systèmes de reconnaissance interprétables basés notamment sur une modélisation explicite des connaissances.*

Pour la plupart des problèmes réels de reconnaissance des formes, la conception automatique d'une modélisation explicite constitue un problème particulièrement complexe par suite de la grande variabilité observée tant à l'intérieur de chaque classe de formes (variabilité intra-classes) qu'entre classes de formes différentes (variabilité inter-classes). La plupart des approches s'appuient sur la capacité d'apprentissage automatique des méthodes de type *stochastique* ou *connexionniste* en se basant respectivement, soit sur une modélisation de l'étendue de la variabilité des formes, soit sur une discrimination directe des classes. Ces approches conduisent souvent à des systèmes de reconnaissance *opaques* (type boîte noire) pour lesquels il est très difficile d'avoir une interprétation des processus de décision mis en œuvre. C'est pourquoi, même si ces approches permettent la réalisation de systèmes affichant assez rapidement des performances intéressantes, ces systèmes se révèlent peu évolutifs et il est difficile d'améliorer leurs performances.

Dans l'optique de pouvoir réaliser des systèmes interprétables, nous avons mis au point une méthodologie de modélisation basée sur la génération automatique de *systèmes d'inférence floue* [2]. Plus précisément, les règles floues utilisées répondent aux hypothèses suivantes :

- les entrées  $x = (x_1, \dots, x_n)$  sont non floues et définies dans l'espace à  $n$  dimensions appelé espace des attributs ou encore espace des entrées ;
- les sous-ensembles flous de sortie  $B_{ik}$  sont des singletons  $\{b_{ik}\}$  ;
- les conclusions des règles ( $Y_k = b_{ik}$ ) sont donc précises :  $Y_k$  représente le degré d'appartenance de la forme présentée en entrée, vis-à-vis de la règle  $R_i$  et de la classe  $k$ .

Considérons maintenant  $N_r$  règles floues formant un SIF permettant de caractériser ou de classer les formes considérées dans  $C$  classes. Les conditions ( $x_j$  est  $M_{ij}$ ) sont interprétées comme le degré d'appartenance de l'observation  $x_j$  au sous-ensemble flou  $M_{ij}$ , c'est-à-dire  $\mu_{M_{ij}}(x_j)$ .

L'expression linguistique d'une règle floue associée à la caractérisation d'une classe  $k$  se formalise d'une manière générale de la façon suivante :

$R_i$  : **SI**  $x_1$  est  $M_{i1}$  **et**  $x_2$  est  $M_{i2}$  **et** ... **et**  $x_n$  est  $M_{in}$   
**ALORS** la donnée appartient à la classe  $k$  **et** non aux autres classes.

Ce qui se traduit dans le formalisme classique des règles floues par :

$R_i$  : **SI**  $x_1$  est  $M_{i1}$  **et**  $x_2$  est  $M_{i2}$  **et** ... **et**  $x_n$  est  $M_{in}$   
**ALORS**  $Y_1 = \delta_{k1}$  **et**  $Y_2 = \delta_{k2}$  **et** ... **et**  $Y_k = \delta_{kk}$  **et** ... **et**  $Y_c = \delta_{kc}$ ,  
 où  $\delta_{ij}$  désigne le symbole de Kronecker.

L'inférence floue est alors définie de la manière suivante :

$$\mu_{B'_k}(y) = \perp_{i=1}^{N_r} I(\beta_i, \mu_{B_{ik}}(y)) \text{ avec } \beta_i = \mathbf{T}_{j=1}^n \mu_{M_{ij}}(x_j),$$

où

- $\mathbf{T}$  est une  $T$ -norme symbolisée par le *et*, utilisée dans la partie prémisse des règles ;
- $\beta_i$  représente la valeur d'activation de la règle  $R_i$  ;
- $I$  est l'implication floue [BM95] ;
- $\perp$  représente l'opérateur d'agrégation ;
- et  $N_r$  est le nombre total de règles.

Le résultat obtenu en sortie du SIF mesure alors le degré d'appartenance à chacune des classes de la forme présentée en entrée.

Les systèmes d'inférence floue permettent de concilier une modélisation de nature numérique qui caractérise la plupart des problèmes réels de reconnaissance des formes, avec une modélisation robuste et qualitative des connaissances extraites.

La génération automatique des SIF repose sur une analyse non supervisée des données d'apprentissage. Pour ce faire, nous utilisons les concepts récents de la classification possibiliste [2] afin d'extraire et de qualifier un ensemble de propriétés pertinentes. La classification non supervisée de type possibiliste permet de caractériser de manière précise les propriétés intrinsèques de chaque prototype modélisé. Ces propriétés sont alors représentées par des sous-ensembles flous  $M_{ij}$ , directement par projection sur les différents axes de l'espace de représentation. Ces différents sous-ensembles flous sont ensuite rassemblés pour former des règles floues dites prototypes qui vont modéliser de manière "lisible" ces propriétés pertinentes.

L'approche méthodologique proposée a été mise en œuvre pour la conception d'un système complet de reconnaissance en-ligne d'écriture manuscrite nommé Resif (cf. paragraphe 4.1).

## 2.4 Modèle et système de perception et d'interprétation (SPI)

**Mots clés :** perception, interprétation, cycles perceptifs, réseaux de neurones.

### Résumé :

*Les méthodes classiques de reconnaissance des formes concernent la reconnaissance d'objets isolés. Cette reconnaissance est un processus purement statique. Cependant, le gestaltisme a mis en évidence le fait que la perception humaine est influencée par les objets qui se trouvent dans le voisinage immédiat de l'objet à reconnaître, c'est-à-dire par son contexte spatial. De plus, les recherches en sciences cognitives ont montré que la perception et l'interprétation visuelle chez l'Homme reposent sur un processus de construction dynamique d'une représentation mentale [Kan79] [EK00]. Ce processus consiste à organiser et à intégrer dynamiquement des données et des connaissances situées à plusieurs niveaux de représentation, en se fondant sur des critères d'organisation perceptive et de*

---

[BM95] B. BOUCHON-MEUNIER, *La logique floue et ses applications*, Addison-Wesley, 1995.

[Kan79] G. KANIZA, *Computing Perceptual Organization in Computer Vision*, New York: Praeger, 1979.

[EK00] M. W. EYSENCK, M. KEANE, *Cognitive Psychology, A Student's Handbook*, Psychology Press, 2000.

cohérence globale afin d'aboutir à différentes interprétations possibles. Ce n'est que depuis quelques années que les problèmes d'automatisation des phénomènes de perceptions sensorielles sont devenus un enjeu de recherche important, notamment pour la perception des signaux et des images [SB94] [BS99]. Il s'agit de doter l'ordinateur de capacités de perception auditive ou visuelle analogues à celles de l'être humain.

Nous avons étudié une nouvelle méthode de reconnaissance de formes *dynamique et multi-contextuelle* ainsi qu'un *système de perception et d'interprétation* de formes en contexte (SPI) [9].

Le *modèle générique* développé repose sur une organisation modulaire et hiérarchique nommée *colonne*, constituée d'un empilement de  $N$  niveaux de représentation. Chacun des niveaux comporte des interactions intra-niveau et inter-niveaux avec les niveaux voisins inférieurs et supérieurs (cf. figure 5). L'interaction entre plusieurs niveaux de représentation, l'entrée de données contextuelles sur plusieurs niveaux et la fusion de données en contexte se font d'une part par un mode de représentation unifié sous forme de couches de neurones et d'autre part par une succession de *cycles perceptifs* qui modifient de façon itérative les activations des neurones de chacune des couches. Il s'agit donc d'un réseau *dynamique et récurrent*. De plus, compte tenu de son architecture, ce système est intrinsèquement parallèle et réparti. Un mécanisme de *focalisation d'attention* est également mis en œuvre dans le système. Ce mécanisme consiste en une sélection dynamique des nœuds-hypothèses et des liens correspondants à partir des résultats d'une phase de pré-reconnaissance. Ceci conduit à ne considérer qu'un faisceau restreint dans l'ensemble du treillis d'hypothèses.

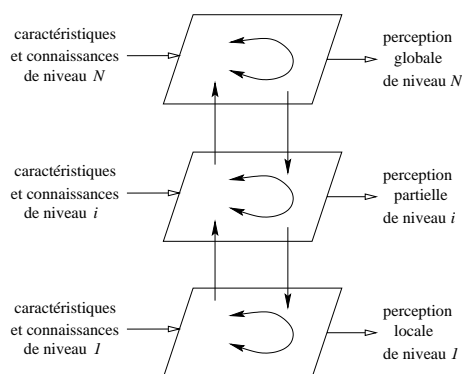


FIG. 5 – Modèle générique d'une colonne de perception et d'interprétation

Chaque hypothèse d'existence de primitives, de sous-formes ou de formes est modélisée par un nœud-hypothèse dans un graphe valué ou dans un réseau par un neurone avec une valeur d'activation associée.

L'introduction de connaissances contextuelles dans les différents niveaux de représentation se fait en trois temps: création de nœuds représentatifs des connaissances disponibles *a priori*, introduction d'activations initiales pour ces nœuds, création de liens activateurs entre les nœuds modélisant des connaissances compatibles entre elles et de liens inhibiteurs entre les nœuds représentant des informations incompatibles entre elles.

[SB94] S. SARKAR, K. L. BOYER, *Computing Perceptual Organization in Computer Vision*, World Scientific, 1994.

[BS99] K. L. BOYER, S. SARKAR (éditeurs), *Perceptual Organization in Computer Vision*, Special Issue of *Computer Vision and Image Understanding*, vol. 76, n°1, Academic Press Ed., 1999.

Les données issues de la forme placée à l'entrée du système (les sensations visuelles) sont hiérarchisées et entrées dans le système (cf. figure 5) à des niveaux différents selon leur nature respective. Elles sont perçues et interprétées en fonction de leurs contextes respectifs. À chacun des niveaux de représentation correspond donc une interprétation possible sur la sortie correspondante.

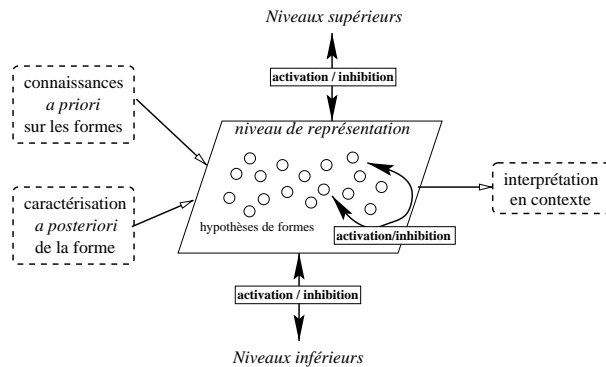


FIG. 6 – Schéma d'un niveau de représentation

À chaque niveau (cf. figure 6), le fonctionnement du système comporte trois phases successives : l'extraction et l'entrée des informations issues de la forme à percevoir (*les sensations visuelles*) relatives à chaque niveau de représentation, la fusion de ces informations avec les connaissances a priori (contextes intra et inter-niveaux), une interprétation en sortie du niveau de représentation considéré.

Les cycles perceptifs consistent en une succession de trois processus qui se répètent :

- un processus de relaxation à l'intérieur de chaque niveau qui augmente les activations latérales des informations compatibles entre elles et qui accroît les inhibitions latérales des informations incompatibles entre elles (mécanisme de perception en contexte),
- un processus ascendant qui consiste en une augmentation des valeurs d'activation ou d'inhibition associées aux nœuds du graphe d'un des niveaux de représentation en fonction de celles de la couche de niveau inférieur (mécanisme d'induction),
- un processus descendant qui propage les informations des niveaux les plus élevés vers les niveaux inférieurs également selon un processus d'activation et d'inhibition (mécanisme de déduction).

Ces mécanismes d'activation, d'inhibition et de fusion de connaissances servent à gérer la cohérence de l'ensemble des informations disponibles, chacune d'elle étant placée en contexte. Il s'agit dès lors d'une interprétation multi-contextuelle. Ceci se traduit, *in fine*, par une interprétation. Cette interprétation, qui émerge au dessus de toutes les autres s'impose alors comme une *évidence*. Elle résulte de la *fusion cohérente* de l'ensemble des connaissances disponibles. L'interprétation est partielle si elle correspond à un niveau de représentation intermédiaire et elle est globale si elle intervient au niveau le plus élevé de la colonne.

Le fonctionnement du système possède trois types de critères d'arrêt. S'il apparaît, au niveau de représentation le plus élevé dans la colonne, un neurone dont l'activation est très supérieure à celle de tous les autres, il correspond à l'interprétation finale de la forme observée. Si aucune valeur d'activation ne dépasse les autres, la forme observée est rejetée comme étant inconnue. Si après un nombre de cycles prédéterminé il existe, dans un niveau de représentation donné, un neurone dont l'activation est

très supérieure à celle de tous les autres neurones, celui-ci correspond à une interprétation partielle. Ceci permet notamment au système de généraliser et de proposer une interprétation probable d'une forme qui n'a pas été apprise, mais dont tous les éléments constitutifs sont cohérents entre eux. Une première version de ce système a déjà été développée et a été appliquée à la reconnaissance de mots d'écriture cursive manuscrite [9]. Cette année un système SPI générique a vu le jour (section 5.11).

### 3 Domaines d'applications

#### 3.1 Analyse de documents papier

Les domaines d'applications de nos recherches dans le contexte de l'analyse de documents papier sont très nombreux. En effet, l'approche générique que nous avons développée (méthode DMOS) permet une adaptation relativement rapide à de nouveaux types de documents. Les documents sur lesquels nous avons déjà travaillé sont les suivants :

- partitions d'orchestre à portées polyphoniques ;
- formules mathématiques ;
- tableaux/formulaires, avec compréhension de l'organisation hiérarchique ;
- formulaires anciens altérés.

La méthode DMOS s'applique à de nombreux types de documents structurés : formules chimiques, cartouches de plans, pages de journaux, factures, bons de livraison, etc.

En outre, la description grammaticale d'un type de document (grâce au langage EPF) peut être générale afin de reconnaître une classe de documents (les cartouches de plans par exemple) ou bien spécifique pour compenser une absence d'informations (un type de cartouche précis mais très souvent altéré).

Bien entendu, il est possible de fusionner des descriptions grammaticales de documents. Ainsi, nous pouvons très simplement construire un système capable de traiter des tableaux contenant des formules mathématiques.

Les utilisations de ces systèmes de reconnaissances sont également nombreuses (avec l'intégration de la reconnaissance de l'écriture manuscrite) :

- rétro-conversion de documents papier pour éviter une saisie manuelle afin d'obtenir une version électronique exploitable. Cette version électronique peut ensuite servir par exemple à une réédition, à la production de Braille, etc.
- identification de types de documents pour effectuer, par exemple, une gestion automatique de fax ;
- génération automatique d'index et d'annotations pour permettre ensuite l'accès à des images de documents (section 5.1) ;
- localisation de zones particulières dans une image de documents pour en faciliter également l'accès (section 5.1).

#### 3.2 Reconnaisseurs d'écriture manuscrite

Les applications en reconnaissance d'écriture manuscrite connaissent un nouvel essor à la fois dans les domaines du traitement automatique de document papier (*reconnaissance hors-ligne*) et dans

les nouvelles modalités d'interactions homme machine basées sur l'utilisation d'un stylet couplé à un écran tactile (*reconnaissance en-ligne*).

En *reconnaissance hors-ligne*, les besoins industriels sont aujourd'hui considérables, et par conséquent l'attente de systèmes de reconnaissance robustes et fiables est très grande. L'étendue des applications est vaste ; elle recouvre notamment les problèmes de traitement automatique de l'écriture contenue dans tout type de document papier, notamment les bordereaux de livraison, les feuilles de sécurité sociale, ou encore les télécopies. Dans ce domaine, nos efforts ont porté plus particulièrement sur la garantie d'une grande fiabilité des résultats en cas de traitement automatique effectif de l'écriture pour éviter tout risque d'erreur.

En *reconnaissance en-ligne* de l'écriture manuscrite, un marché très important s'est ouvert suite au développement économique récent des produits suivants :

- ordinateurs de poche (Palm, PocketPC),
- ordinateurs tablettes (ordinateurs sans clavier munis d'un écran LCD tactile),
- téléphones mobiles de nouvelle génération (« smart phone ») qui intègrent les modalités d'interaction des ordinateurs de poche (écran LCD tactile) et leurs principales fonctionnalités (agenda, éditeur de texte, accès à Internet, etc.).

Le logiciel ResifCar est aujourd'hui intégrable à ce type de produit (cf. paragraphes 4.1 et 5.9).

### 3.3 Interfaces orientées *stylet*

Les interfaces utilisant un *stylet* constituent une nouvelle modalité d'interaction entre l'homme et la machine. Elles sont utilisées sur des produits possédant des écrans tactiles ou des tablettes à digitaliser. Ces interfaces offrent à l'utilisateur la possibilité d'effectuer à main levée à l'aide d'un stylet :

- des saisies naturelles (il peut directement écrire des mots, des caractères ou des symboles sur l'écran tactile) ;
- des dessins de graphe, de croquis, d'organigrammes, etc. ;
- ou encore des commandes gestuelles intuitives (suppression, insertion, déplacement, etc.).

Ces différents tracés sont alors reconnus, interprétés et exécutés par la machine en fonction de leur nature et de leur contexte.

L'étendue des applications basées sur des interfaces orientées stylet est très vaste. On peut tout d'abord mentionner les produits dont la communication, pour des questions d'encombrement et de convivialité, s'appuie uniquement sur un écran tactile : ordinateur de poche (Palm, PocketPC), téléphone mobile de nouvelle génération (« smart phone »), ordinateur tablette. On peut de plus citer les applications qui, de par leur nature, profiteront pleinement des avantages d'une communication basée sur une saisie de symboles ou des graphiques à main levée : CAO, édition électronique, édition musicale, édition de formule mathématique, etc.

Dans ce domaine, nous avons conçu et réalisé plusieurs prototypes qui ont pour objectif de mettre en évidence les possibilités et la convivialité de ces nouvelles modalités d'interaction homme-machine : éditeur interactif de partitions musicales, éditeur de graphes dessinés à main levée, éditeur de page WEB basé sur la reconnaissance de gestes graphiques.



## 4 Logiciels

### 4.1 Resif : Reconnaissance d'Écriture Manuscrite en-ligne par Systèmes d'Inférence Floue

**Participant** : Éric Anquetil.

**Mots clés** : système de reconnaissance, écriture manuscrite, plate-forme mobile, PocketPC, smartphone.

Resif est un ensemble de composants logiciels, développés dans l'équipe IMADOC, pour l'analyse, la modélisation et la reconnaissance automatique de l'écriture manuscrite en-ligne. Les principes du système de reconnaissance sont basés sur une expertise spécifique de l'écriture manuscrite en s'appuyant sur une modélisation par Systèmes d'Inférence Floue (cf. section 2.3). Les quatre composants logiciels principaux sont :

- ResifApp : système de modélisation automatique par apprentissage des caractères manuscrits ;
- ResifCar : système d'analyse et de reconnaissance de caractères isolés (lettres, chiffres, symboles spéciaux) ;
- ResifMot : système d'analyse et de reconnaissance de mots et de lettres manuscrits ;
- ResifSeg : système de segmentation automatique supervisée de mots en lettres.

Ces logiciels ont été déposés à l'APP (auteurs : E. Anquetil et G. Lorette). Un transfert industriel avec la société PurpleLabs<sup>2</sup> (cf. section 6.1) a porté sur le logiciel de reconnaissance de caractères isolés ResifCar. Celui-ci a été embarqué sur des téléphones mobiles de nouvelle génération (smartphone). Il permet à l'utilisateur d'écrire directement des caractères cursifs (lettres, chiffres, symboles) sur l'écran tactile du téléphone, en conservant une écriture naturelle.

Une licence d'exploitation commerciale est actuellement en cours (cf. section 6.1). Ce portage illustre notamment les possibilités d'intégration de la technologie Resif sur des plate-formes mobiles à capacité limitée : processeurs cadencés à 13 Mhz et occupation mémoire inférieure à 200 Ko en Rom et 50 Ko en Ram.

### 4.2 DocRead : générateur automatique de systèmes de reconnaissance de documents structurés

**Participant** : Bertrand Coüasnon.

**Mots clés** : reconnaissance, documents structurés, partitions musicales, formules mathématiques, tableaux, formulaires, archives.

DocRead est un générateur automatique de systèmes de reconnaissance de documents structurés, développé grâce à la méthode DMOS. Il est constitué d'un compilateur du langage EPF (permettant de décrire un document à l'aide d'une grammaire), d'un module d'analyse lié à ce langage, d'un module de vision précoce (binarisation et extraction de segments) et d'un classifieur ayant des capacités de rejet.

---

<sup>2</sup>PurpleLabs : cf. <http://www.purplelabs.com>

Ce générateur nous permet une adaptation rapide à un nouveau type de document. En effet, il faut simplement définir une nouvelle grammaire (à l'aide d'EPF) qui décrit le nouveau type de document et, si nécessaire, il faut effectuer un nouvel apprentissage du classifieur pour lui permettre de reconnaître de nouveaux symboles. Le système de reconnaissance adapté est alors automatiquement produit par compilation.

Grâce à ce générateur, nous avons défini un certain nombre de systèmes de reconnaissance :

- ScoRead : prototype de reconnaissance de partitions musicales ;
- MathRead : prototype de reconnaissance de formules mathématiques ;
- TennisRead : prototype de reconnaissance de terrain de tennis dans des vidéo (section 5.5);
- TabRead : prototype de reconnaissance de structures tabulaires ;
- DecRead : prototype de reconnaissance de décrets de naturalisation du XIX<sup>e</sup> siècle. Ce sont des formulaires uniquement manuscrits (section 5.4);
- FormuRead : logiciel qui permet d'extraire automatiquement la structure de formulaires d'incorporation militaire du XIX<sup>e</sup> siècle, malgré leur dégradation. Ce logiciel a été testé sur 60 223 pages des Archives de la Mayenne et a montré ses très bonnes capacités (section 5.3).

### 4.3 Bibliothèque d'extraction d'indices visuels

**Participant** : Jean Camillerapp.

**Mots clés** : Vision précoce, binarisation adaptative, vectorisation, filtrage de Kalman, squelettisation, multi-résolution.

Il s'agit de la bibliothèque sur laquelle s'appuie le logiciel DocRead décrit dans la section 4.2.

Elle comporte différentes méthodes de binarisation adaptative. Celles-ci utilisent soit des déterminations locales du seuil soit des coopérations contours-régions. Certaines de ces méthodes peuvent fournir des résultats avec une résolution sub-pixelle.

Elle réalise l'extraction des segments rectilignes dans des images complexes en s'appuyant sur le filtrage de Kalman décrit dans la section 2.1.

Elle comporte également une méthode de squelettisation à partir d'images en niveaux de gris.

La résolution choisie pour la numérisation des documents n'est pas forcément celle qui est la mieux adaptée à la reconnaissance de toutes les structures. C'est pourquoi la bibliothèque fournit également des possibilités de traitement en multi-résolution.

Bien qu'initialement optimisée pour le traitement des images de documents, cette bibliothèque peut être utilisée dans d'autres contextes de traitement d'images. Elle a ainsi été essayée pour repérer les limites d'un court de tennis dans des séquences d'images vidéo.

## 5 Résultats nouveaux

### 5.1 Annotations pour l'accès aux documents d'archives

**Participants** : Nicolas Briec, Jean Camillerapp, Bertrand Coüason, Ivan Leplumey, Hélène Richy.

La représentation numérique des documents facilite considérablement leur consultation en permet-

tant, d'une part l'accès à distance et d'autre part l'accès simultané des lecteurs. C'est pourquoi dans de nombreux domaines, et en particulier dans celui des archives, on assiste actuellement à des campagnes de numérisation des documents papier. Mais cela ne suffit pas, il faut également construire des outils qui permettent au lecteur de sélectionner les documents susceptibles de l'intéresser.

Afin de faciliter la sélection des documents, il faut adjoindre aux images de ceux-ci des informations complémentaires à partir desquelles il sera possible de faire une indexation (cette indexation permet d'éviter le feuilletage manuel d'une quantité gigantesque de documents pour retrouver un document particulier). Ces informations, appelées annotations, peuvent être textuelles (date, nom, mots-clés...) ou géométriques (découpage d'une page en actes, position de certaines zones : noms, lettrines, signatures...) et sont stockées grâce aux métadonnées.

Un même document peut être annoté selon deux modes complémentaires.

### **Annotations automatiques**

Des annotations peuvent être produites de manière automatique en utilisant des techniques de reconnaissance de documents comme la méthode DMOS. Ainsi, cette méthode a été appliquée sur différents documents anciens manuscrits pour détecter des pages dans des doubles pages de registres d'état civil (section 5.2), pour détecter des cases dans des formulaires de recrutement militaire du XIX<sup>e</sup> siècle (section 5.3) et pour détecter de noms et numéros manuscrits dans des décrets de naturalisation également du XIX<sup>e</sup> siècle (section 5.4).

À partir des images, des annotations XML sont produites et elles facilitent l'accès au documents. Ainsi, grâce au langage de transformation XML, il est possible de constituer automatiquement des pages d'index d'images limitées, par exemple, à la case contenant le nom. Ces index rapidement consultables permettent au lecteur de sélectionner les documents qu'il veut voir in-extenso.

### **Annotations collectives**

Dans certains cas, les annotations produites automatiquement par le système peuvent être complétées par des annotations manuelles. Le réseau permet de mettre en commun les efforts de lecture de tous et rend ainsi les documents de plus en plus accessibles. Chaque lecteur peut enrichir les annotations par les informations qu'il trouve pertinentes. Celles-ci sont réalisées, à travers le réseau, par les utilisateurs au fur et à mesure de leurs consultations en s'appuyant sur les annotations automatiques. Ces nouvelles annotations permettent à d'autres lecteurs de sélectionner plus facilement les documents les plus en adéquation avec leurs recherches.

Un prototype de plate-forme d'annotations collectives intégrant des annotations automatiques sur des registres d'état civil est en cours de finalisation (section 6.2).

## **5.2 Détection des pages dans des registres paroissiaux et d'état civil**

**Participants :** Jean Camillerapp, Ivan Leplumey.

La numérisation des registres d'état civil et des registres paroissiaux est effectuée en double-page. La mauvaise qualité de ces documents rend relativement difficile leur annotation automatique. Cependant, la détection automatique de la position des pages dans l'image a pu être réalisée afin d'augmenter l'ergonomie de la consultation.

Le système développé s'appuie sur la détection des segments par filtrage de Kalman en l'appliquant dans une image multi-résolution : à une faible résolution, les segments détectés correspondent à la reliure ou aux bords de la page. Les hypothèses de page sont ensuite validées par la corrélation des positions sur la séquence des doubles pages de l'ensemble d'un registre.

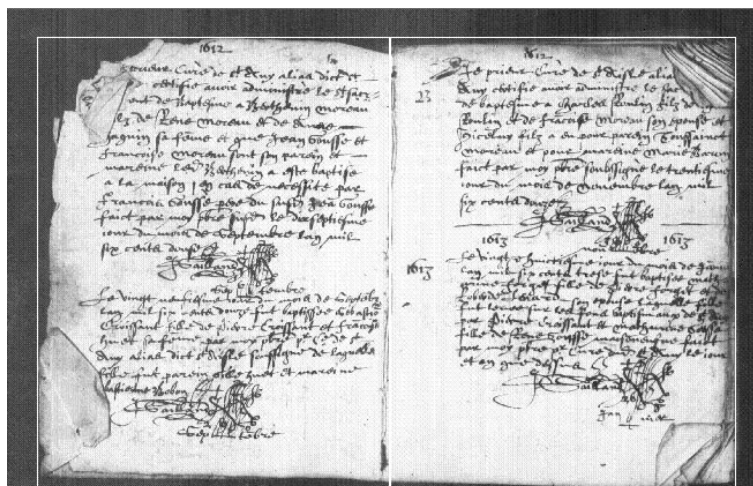


FIG. 7 – Registre d'état civil et l'annotation produite pour la détection des pages

Le système a été testé sur un ensemble de registres d'une commune de la Mayenne contenant 1407 double-pages, correspondant à des données de 1612 à 1890. Le système a procédé au découpage de l'intégralité des 1407 double-pages en simple-pages sans rencontrer d'échec (figure 7).

### 5.3 Application de la méthode DMOS aux registres matricules

**Participants** : Nicolas Briec, Jean Camillerapp, Bertrand Couasnon.

Ces registres matricules sont des documents d'incorporation militaire du XIX<sup>e</sup> siècle qui sont constitués sur des formulaires préimprimés. Si la structure de chaque fiche est restée stable pendant une quarantaine d'année, en revanche la hauteur des cases varie de 1 à 2 cm selon les éditions. De plus certaines cases se sont avérées trop petites à l'usage et les secrétaires ont collé de petites feuilles annexes (paperolles ou retombes) qui masquent largement la structure du document.

Grâce à la méthode DMOS, une grammaire EPF décrivant des tableaux de manière générale a pu être définie afin de produire un système de reconnaissance adapté aux structures en tableau, comportant un nombre indéterminé de lignes ou de colonnes de taille inconnue a priori. Pour obtenir des performances satisfaisantes avec ce système, il faut posséder des documents de bonne qualité dans lesquels la structure est nettement apparente.

Dans l'étude sur les registres matricules, il a fallu compenser les dégradations importantes du document par une description grammaticale en EPF un peu plus précise de la structure. Cette description indique notamment que le document comporte deux colonnes, qui sont elles-mêmes redécomposées en cases. En revanche, il n'est pas nécessaire de décrire exactement la taille de ces différentes parties. Grâce à cette grammaire un système de reconnaissance adapté aux registre matricules a été produit (figure 8).

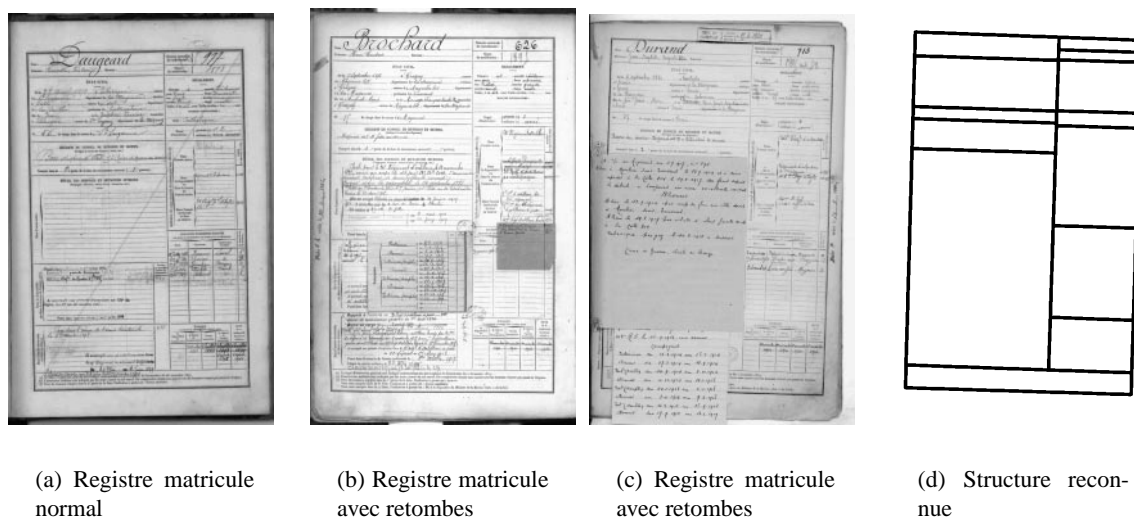


FIG. 8 – Exemple d’annotation automatique de registres matricules

Dans un premier temps ce système détecte les images qui ne correspondent pas au modèle décrit, mettant ainsi en évidence des pages telles que les tables d’index ou les pages mal numérisées. Dans un second temps ce système vérifie que la taille des cases détectées reste constante pour un même registre ce qui garantit une bonne reconnaissance.

Certaines cases contiennent des informations d’ordre médical, que l’on ne doit pas rendre publiques car elles sont protégées pendant 150 ans. Grâce à la détection des cases, il devient possible d’augmenter la collection de registres que le public est admis à consulter en ne lui présentant que la partie publique de chaque formulaire.

Une seule description EPF a été utilisée pour annoter une collection complète de 60 223 formulaires [21]. Sur cet ensemble uniquement 0,4% des images (259) sont rejetées dans la première passe (documents non traitables, même manuellement, car trop abimés), ensuite 0,4% des images (256) n’ont pas été traitées tandis que le taux de reconnaissance de la structure sur les images restantes est de 99,6%, même en présence de *retombes*. Il est important de noter que dans les images validées il n’y a aucune erreur de localisation de la structure, ce qui est primordial dans un contexte industriel.

La détection des cases nous permet actuellement de travailler sur la reconnaissance de champs manuscrits (comme le nom) pour réaliser une indexation automatique de ces registres matricules, dans le cadre d’un contrat avec les Archives des Yvelines (section 6.3).

#### 5.4 Application de la méthode DMOS à des décrets de naturalisation

**Participant :** Bertrand Couïasnon.

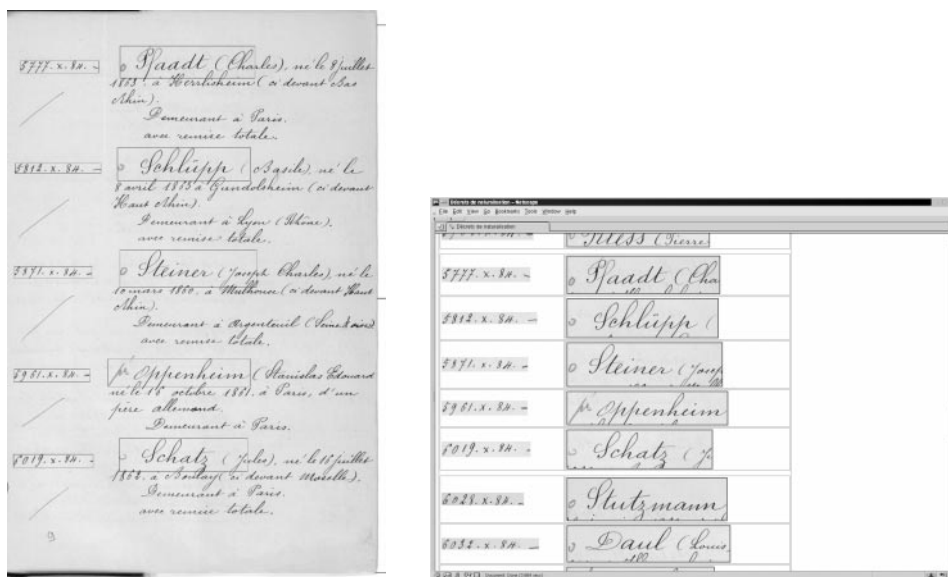
Le service historique des Archives Nationales souhaite permettre l’accès du public aux images des décrets de naturalisation de la fin du XIX<sup>e</sup> et du début du XX<sup>e</sup> siècle. En effet, ce sont des documents uniques qui sont les seuls à pouvoir justifier de la nationalité française de certaines personnes. Un dé-

cret contient en moyenne une dizaine de pages, généralement manuscrites, organisées en paragraphes, où chaque paragraphe concerne la naturalisation d'une seule personne. En outre, le patronyme de la personne concernée est placé en début de paragraphe.

La recherche d'une personne dans ce type de document est difficile car il faut feuilleter l'ensemble des décrets. Il n'existe pas de table récapitulative permettant de savoir pour un patronyme donné quel est le décret qui le concerne. Or la demande du public est très forte pour retrouver une personne dans l'ensemble de ces décrets, et la constitution manuelle de tables récapitulatives n'est pas envisageable car trop fastidieuse et trop coûteuse.

Par rapport aux registres matricules, la structure est beaucoup plus faible puisque constituée uniquement par l'organisation en paragraphe de l'écriture manuscrite. Grâce à la généralité de la méthode DMOS, il a été possible de définir une grammaire EPF décrivant l'organisation de ces documents en lignes de texte manuscrit, paragraphes et colonnes, en utilisant uniquement les composantes connexes calculées dans l'image.

Cette description a permis par compilation de produire un système de reconnaissance capable de détecter la position du nom et d'un numéro de dossier (figure 9a). A partir de la structure XML produite, par transformation avec XML, le système génère une page XML contenant les imagerie du numéro de dossier et du nom, présentés sous forme de tableau. Ceci permet de feuilleter beaucoup plus rapidement un ensemble de décrets pour retrouver une personne et d'accéder à la page la concernant (figure 9b).



(a) Décret de naturalisation avec la position des numéros de dossier et des noms détectés

(b) Tableau d'imagerie pour accélérer la consultation

FIG. 9 – Exemple d'annotation automatique de décret de naturalisation

Les premiers tests ont porté sur 234 pages de décrets, qui contiennent 1918 noms ou numéros. Seul

un numéro n'a pas été détecté, ce qui représente un taux de détection des noms ou numéros de plus de 99,9%. En revanche, le système a produit 203 fausses reconnaissances de noms ou numéros soit 10,6%. Ces fausses détections ne sont pas très gênantes dans ce contexte applicatif, puisque l'utilisateur aura simplement un peu plus d'images à consulter. Pour diminuer ces fausses détections il faudrait être capable de reconnaître les inscriptions manuscrites.

Nous avons montré ainsi que la méthode DMOS pouvait également être utilisée sur des documents ayant une structure plus faible sans présence de filets, et contenant uniquement du manuscrit.

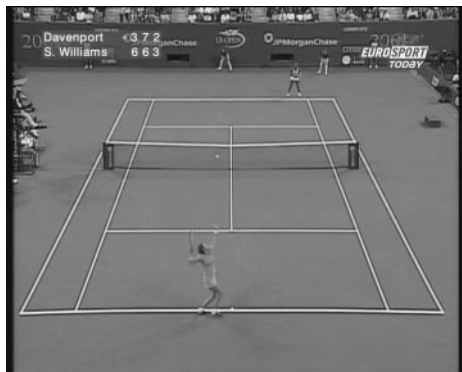
### 5.5 Application de la méthode DMOS à la reconnaissance de terrain de tennis dans une séquence vidéo

**Participants** : Bertrand Couasnon, Jean Camillerapp.

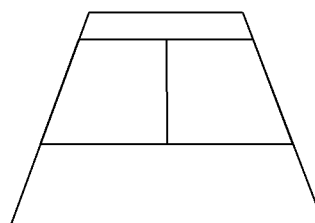
Afin de valider sur d'autres images la généralité de la méthode DMOS, et son extension à des images en niveaux de gris, nous avons cherché à reconnaître les limites d'un court de tennis sur des images vidéo.

Pour cela nous avons défini deux descriptions de l'agencement des lignes délimitant le court : l'une pour le terrain vue du fond et l'autre pour le terrain vue de côté.

Ce formalisme fournit une solution élégante à la localisation du terrain dans chaque image d'une séquence vidéo et donne ainsi des repères spatiaux à une étude sur les déplacements des joueurs dans les différentes phases du jeu (figure 10).



(a) Image du court



(b) Limites détectées du terrain de jeu en simple

FIG. 10 – Exemple de reconnaissance de la structure d'un terrain de tennis

### 5.6 Méthode générique de segmentation de symboles

**Participants** : Bertrand Couasnon, Yann Ricquebourg.

Pour la reconnaissance des documents fortement structurés (partitions musicales, formules mathé-

matiques, formulaires de tableaux...), le système DMOS développé dans le projet Imadoc est constitué d'un formalisme grammatical et d'un analyseur associé (cf. section 2.2). Cet analyseur permet d'introduire le contexte (modélisé par la grammaire descriptive du type de document) formulant un *a priori* sur le symbole à retrouver dans la phase de segmentation ce qui peut permettre, par exemple, de séparer deux symboles qui se touchent et ainsi les reconnaître. Cette approche s'appuie sur la segmentation de l'image sous-jacente qui fournit les « terminaux » lexicaux de l'analyse grammaticale. Cependant dans le système existant il est nécessaire de définir une technique de segmentation/reconnaissance spécifique à chaque symbole, ce qui gêne sa généralisation et son adaptation à de nouveaux documents.

Les premiers objectifs de ce travail sont donc de développer une méthode générique de segmentation pouvant s'adapter à n'importe quel symbole, qui de plus permette de traiter les cas difficiles où la segmentation initiale n'a pas isolé le symbole à reconnaître (sur- ou sous-segmentation). Nous explorons les possibilités de baser cette adaptation sur une modélisation classique par apprentissage statistique : une analyse en composantes principales, dégageant un modèle général et ses modes de variation (figure 11a,b).

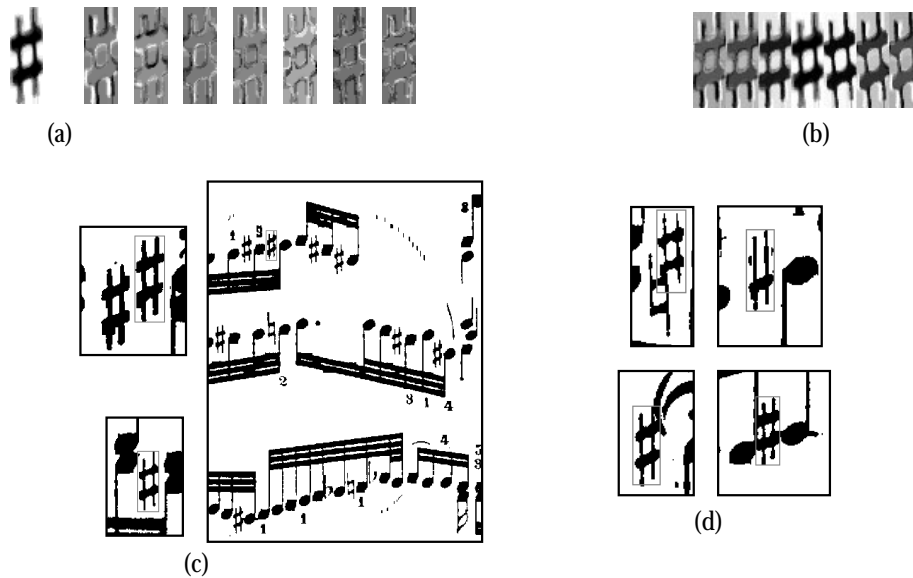


FIG. 11 – Résultats de détection/resegmentation grâce à l'analyse en composantes principales d'un ensemble d'apprentissage de dièses. (a) Modèle dégagé et premiers vecteurs propres associés. (b) Illustration d'une plage du mode de déformation permis par le premier vecteur propre. (c) Détection/reconnaissance dans trois environnements bruités. (d) Détection/reconnaissance malgré des difficultés (dégradations, connexité)

Les principaux intérêts de cette technique, éprouvée dans des contextes voisins (reconnaisances d'imagettes de visages, de signalisation routière verticale, etc.) sont :

- d'être générique. En effet, l'apprentissage est basé sur une population de vecteurs qui peuvent donc être des pixels d'images de symboles mais aussi d'autres caractéristiques propres au problème.
- de produire, par le jeu des vecteurs propres extraits dans la transformation statistique (figure 11a), un ensemble de caractéristiques discriminantes hiérarchisées par ordre de représentativité (des



facteurs communs aux détails rares). L'expérience montre communément que l'on peut alors se réduire aux seuls premiers vecteurs propres et alléger la phase de reconnaissance, tout en conservant une très bonne couverture de l'ensemble à caractériser.

- de permettre une re-segmentation : associée à un algorithme d'optimisation, on peut extraire la zone de l'image correspondant le mieux aux modèles appris (figure 11c). Ainsi, avec un tel outil, dans les cas délicats où un simple découpage en composantes connexes n'est pas satisfaisant pour conduire à une reconnaissance (car fournissant une sur- ou sous-segmentation de l'objet à identifier) on peut réviser la segmentation et en proposer une autre isolant la partie reconnue (figure 11d).

Ce mécanisme pourra avantageusement intégrer la chaîne d'analyse DMOS, lui conférant alors la faculté de revoir la segmentation initiale et de traiter ainsi les zones d'échec de la reconnaissance.

## 5.7 Bibliothèque virtuelle de typographie

**Participant** : Jacques André.

Si de nombreux ouvrages anciens sont souvent cités dans des bibliographies (par exemple en histoire de la typographie), il est en revanche fréquent qu'il soit très difficile de trouver ces ouvrages, ou des rééditions. Une version expérimentale est en cours de définition pour une « bibliothèque virtuelle de typographie » dont la collecte du fond a déjà démarré en relation avec l'Institut Culturel du Livre (Lyon), l'Institut de documentation et communication (Vannes) et la bibliothèque de l'Irisa. Voir <http://www.irisa.fr/faqtypo/BiViTy.html>.

Il nous semble intéressant de définir un formalisme permettant la recherche sur le web d'ouvrages (bibliothèques où ils se trouvent, rééditions – parfois électroniques –, accessibilité, etc.) à partir de métadonnées (notices des bibliothèques, auteurs, sites, etc.). Le formalisme prévu est celui des structures de traits typées.

## 5.8 Système de classification hybride interprétable par construction automatique de systèmes d'inférence floue

**Participants** : Nicolas Ragot, Éric Anquetil, Guy Lorette.

Étant donné les nombreuses applications de l'équipe susceptibles d'utiliser un système de reconnaissance de formes, nous travaillons à l'élaboration d'une nouvelle architecture hybride pour la classification [?, 23]. Celle-ci s'attache notamment à résoudre les objectifs suivants :

- généralité : une automatisation complète de la phase d'apprentissage permet d'extraire et de mettre en forme l'information pour le traitement de différents problèmes de classification ;
- compacité : afin de pouvoir embarquer le système sur des plate-formes aux ressources limitées (PocketPC, smartphone, etc.), le nombre de paramètres nécessaires à la modélisation doit être aussi réduit que possible ;
- interprétabilité : cette propriété permet à un expert d'effectuer *a posteriori* des « corrections » sur le système en fonction d'une application donnée afin d'optimiser les performances brutes (taux de reconnaissance), la fiabilité (rejet), l'utilisation des ressources (temps de calculs et utilisation mémoire) (cf. section 5.9, [20]) ;

- performances en reconnaissance : le respect des contraintes précédentes ne doit pas se faire au détriment des performances qui restent l'objectif principal dans des applications complexes comme la reconnaissance de symboles manuscrits.

Afin de satisfaire ces objectifs, le classifieur proposé utilise une architecture hybride reposant sur une hiérarchisation des connaissances sur deux niveaux. Dans chacun d'eux, les informations modélisées sont représentées par des sous-ensembles flous. Il s'agit en effet d'un outil de représentation idéal pour synthétiser une information imprécise et sujette à la variabilité. Les sous-ensembles flous sont extraits par apprentissage automatique et sont exploités dans des règles floues agrégées dans des systèmes d'inférence flous (SIF). Ce formalisme permet une gestion robuste (qualitative) des données numériques présentées en entrée tout en conservant une lisibilité des connaissances exploitées (cf. section 2.3).

Le premier niveau du système effectue une modélisation intrinsèque des classes afin de déterminer des éléments de similarité propres à chaque classe. Les connaissances sont extraites par apprentissage non supervisé [Kri94] pour être ensuite formalisées en SIF. Ce premier niveau permet d'établir une pré-classification sur laquelle va s'appuyer le niveau de modélisation suivant.

Le second niveau se focalise sur l'établissement d'éléments de séparabilité pour chaque classe d'après les confusions mises en évidence par le premier niveau. Cette modélisation inter-classes de nature discriminante est réalisée à l'aide d'arbres de décision flous spécifiques [RA01b,RA01a]. Ceux-ci déterminent à partir de zones d'attention précises issues de la pré-classification, les frontières entre les classes entrant en confusion. Ils sont formalisés par des SIF ce qui garantit après la phase d'apprentissage une modélisation homogène des deux niveaux qui sont donc interdépendants (modélisation hiérarchique) et complémentaires (modélisation intrinsèque et discriminante).

Grâce à ce formalisme homogène, le processus de classification final s'appuie sur l'ensemble des connaissances du système et bénéficie ainsi de leur interdépendance et de leur complémentarité. La décision finale est issue de la fusion entre le résultat de l'exploitation des connaissances intrinsèques (pré-classification du premier niveau) et des connaissances discriminantes (deuxième niveau de classification) en respectant leur organisation hiérarchique.

Des expérimentations ont été conduites sur différents problèmes de classification et comparés à différentes autres approches. Le tableau 1 présente les résultats sur des benchmarks classiques. Les conditions expérimentales sont identiques à celles présentées dans l'article de Lim et Loh [LLS00] dont les résultats de trois classificateurs parmi les meilleurs ont été extraits (LVQ, RBFN et QUEST). SH désigne le Système Hybride proposé.

Sur ces deux problèmes, le système proposé obtient des performances en terme de taux de reconnaissance qui se situent parmi les meilleurs sur les 33 recensés dans [LLS00]. Cette aptitude du système à s'adapter à différents problèmes de classification valide son aspect générique. Celui-ci est encore

---

[Kri94] R. KRISHNAPURAM, « Generation of membership functions via possibilistic clustering », in : *IEEE World congress on computational intelligence*, p. 902–908, 1994.

[RA01b] N. RAGOT, E. ANQUETIL, « A New Hybrid Learning Method for Fuzzy Decision Trees », in : *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, 3, 2001.

[RA01a] N. RAGOT, E. ANQUETIL, « Modélisation Automatique des Connaissances par Systèmes d'Inférence Floue Hiérarchisés », in : *Actes du congrès LFA'01, Rencontres Francophones sur la Logique Floue et ses Applications*, p. 105–111, 2001.

[LLS00] T.-S. LIM, W.-Y. LOH, Y.-S. SHIH, « A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms », *Machine Learning* 40, 3, 2000, p. 203–228.

	Formes d'ondes	Images satellites
SH	82,4 %	89,6 %
LVQ	83,0 %	90,2 %
RBFN	84,9 %	87,9 %
Quest	82,3 %	84,6 %

TAB. 1 – Taux de reconnaissance de différents classifieurs sur deux benchmarks classiques

confirmé dans le tableau 2 où le problème de la reconnaissance en-ligne de chiffres manuscrits est traité. La base utilisée pour cette série d'expérimentations regroupe environ 4 000 chiffres saisis par 400 scripteurs différents. Le tableau présente les performances en terme de taux de reconnaissance (% reco.) et de nombre de paramètres utilisés pour la modélisation (nb param.). La comparaison s'effectue avec différentes autres approches classiques (MLP, RBFN, SVM) ainsi qu'avec un système dédié à la reconnaissance de l'écriture manuscrite (RESIFCar) développé dans l'équipe et récemment porté sur des téléphones mobiles (cf. section 5.9, [20]).

	nb param.	% reco.
SH	12 150	94,7
MLP	8 100	93,5
SVM	114 075	94,6
RBFN	8 600	93,6
ResifCar	6 084	94,3

TAB. 2 – Performances (nombre de paramètres et taux de reconnaissance) de différents algorithmes de classification sur la reconnaissance en-ligne de chiffres manuscrits

Trois de ces approches obtiennent des performances en reconnaissance voisines entre elles mais supérieures aux deux autres. Parmi ces trois systèmes de classification, les SVM bénéficient d'un apprentissage entièrement automatique mais nécessitent un grand nombre de paramètres. Le système hybride présenté (SH) permet par son processus de modélisation automatique (sans optimisation *a posteriori* et sans gestion du rejet) d'obtenir des performances équivalentes aux SVM mais avec 10 fois moins de paramètres. Enfin, le système RESIFCar, dédié à la reconnaissance de caractères manuscrit, a bénéficié d'une optimisation *a posteriori* par un expert ce qui a permis de réduire de manière significative le nombre de paramètres, tout en conservant des performances équivalentes en reconnaissance.

## 5.9 Embarquement du système ResifCar sur des téléphones mobiles

**Participants** : Éric Anquetil, Laurent Pasquer.

Le système de reconnaissance de caractères ResifCar a été embarqué dans des téléphones mobiles de nouvelle génération (smartphone). Ce travail, débuté en décembre 2001, s'est effectué en collaboration avec la société PurpleLabs qui conçoit des téléphones pour en vendre le design à des fabricants de mobiles. ResifCar est aujourd'hui sous licence d'exploitation commerciale. La première version de

téléphone embarquant la technologie ResifCar est aujourd'hui commercialisée par la société italienne Enteos. Ce mobile regroupe les fonctionnalités d'un PDA avec celles d'un téléphone.

L'intégration de ResifCar sur des téléphones mobiles offre à l'utilisateur la possibilité d'écrire directement ses commandes ou messages sur l'écran tactile du téléphone, en conservant son écriture naturelle et sans période d'apprentissage.

ResifCar est basé sur une modélisation des caractères par SIF hiérarchisés. Cette modélisation repose sur les fondements théoriques de la logique floue décrits brièvement au paragraphe 2.3, et s'appuie sur les connaissances structurelles de l'écriture manuscrite résumées au paragraphe 3.2. Elle est compacte, robuste et explicite (c'est à dire interprétable).

Les fondements scientifiques associés à ResifCar ont permis d'optimiser le système [20] afin de l'embarquer sur des processeurs de type ARM7 TDMI cadencés à 13 Mhz. Il occupe moins de 50 Ko en RAM et moins de 200 Ko en RAM. En terme de temps de calcul (temps de reconnaissance d'un caractère sur le processeur à 13 Mhz), un facteur de l'ordre de vingt a été gagné avec l'optimisation du système : le temps de reconnaissance moyen d'un caractère est aujourd'hui de l'ordre de 0,5 s contre environ 11 s avant optimisation.

Dans sa version embarquée, ResifCar permet la reconnaissance des 26 lettres minuscules cursives de l'alphabet latin, des chiffres et de certains symboles spéciaux (@, &, ...) Le système est capable d'absorber une grande variété de styles d'écriture en offrant la reconnaissance de plusieurs allographes<sup>3</sup> d'un même symbole.

L'évaluation du système a porté sur plus de 300 scripteurs (200 utilisés en phase d'apprentissage et 100 autres en phase de test). Les taux de reconnaissance moyens obtenus par le système embarqué sont de l'ordre de 92 % en première position et de 97 % dans les trois premières positions.

Une version du système ResifCar a aussi été portée cette année sur Pocket PC (agenda de poche).

Nos travaux de recherche se poursuivent aujourd'hui autour de deux voies principales : l'optimisation du système ResifMot pour aller vers une reconnaissance embarquée de mots manuscrits cursifs liés et la généralisation du principe de modélisation de l'écriture utilisée dans Resif pour étudier son adaptation aux autres types de caractères et plus particulièrement aux caractères chinois.

## 5.10 Intégration et modélisation de connaissances lexicales pour la reconnaissance en-ligne de l'écriture manuscrite

**Participants** : Sabine Carbonnel, Éric Anquetil, Laurent Pasquer.

Le système de reconnaissance ResifCar permet de reconnaître des caractères isolés (cf. section 5.9) ; il est complété par le système ResifMot qui gère la reconnaissance de mots manuscrits liés. Celui-ci, opère une segmentation automatique du mot en lettres. Les erreurs de segmentation ou de reconnaissance génèrent des ambiguïtés : les propositions brutes du système ne correspondent pas alors forcément à des mots. Pour identifier le mot qui a été écrit on a alors besoin d'intégrer des connaissances lexicales qui vont s'appuyer sur des dictionnaires allant de 1 000 à 100 000 mots selon le contexte de la reconnaissance.

Les objectifs d'intégration du système de reconnaissance de mots sur des plates-formes à capacités réduites (ordinateurs de poche, téléphones portables, ...) induisent des contraintes de temps de

<sup>3</sup>Allographes : formes graphiques différentes associées à une même classe de lettre.

reconnaissance et d'encombrement mémoire qui nécessitent de porter une attention particulière à l'exploitation de ces connaissances lexicales.

Une optimisation de l'intégration des connaissances lexicales a été réalisée cette année en phase de post-traitement (après la phase de reconnaissance) en utilisant un lexique (liste de mots) contenant jusqu'à 25 000 mots. L'utilisation de ces connaissances permet de corriger les erreurs de reconnaissance ou de segmentation en proposant des solutions lexicalement correctes. La recherche de ces solutions s'effectue en estimant le meilleur appariement des propositions du reconnaisseur avec les mots du lexique.

Le lexique a été organisé avec l'objectif de répondre au mieux aux contraintes d'utilisation et d'accès, mais également de temps et d'encombrement.

- *Organisation du lexique en sous-ensembles* : l'idée est de regrouper les mots graphiquement proches (au niveau de leur forme manuscrite). Pour cela, nous avons utilisé des caractéristiques de l'écriture manuscrite (notamment les traits graphiques les plus robustes vis-à-vis de la variabilité des styles d'écriture) pour définir la silhouette d'un mot. Les mots de silhouettes semblables sont regroupés dans un sous-ensemble indexé par la silhouette qui les représente. Cependant certains mots peuvent avoir plusieurs silhouettes et être donc présents dans plusieurs sous-ensembles, ce qui introduit une certaine redondance (utile lors de l'utilisation du lexique). Cette organisation permet de guider la recherche en se focalisant uniquement sur les sous-ensembles adéquats. Cette phase de réduction de lexique permet un gain de temps important lors de la phase d'appariement des propositions avec les mots du lexique, car seuls les sous-ensembles contenant des mots ressemblants sont retenues.
- *Organisation des sous-ensembles* : chaque sous-ensemble est structuré sous la forme d'un arbre lexical ce qui permet de factoriser les préfixes communs des mots. Ceci induit une diminution du nombre de calculs de comparaison (gain de temps) et une compression du dictionnaire (gain de place).

Les connaissances lexicales sont utilisées de la façon suivante :

- à partir des hypothèses fournies par le module de reconnaissance de ResifMot, sélection du (ou des) sous-ensemble(s) du lexique de silhouettes semblables ;
- comparaisons à l'aide d'une distance d'édition spécifique, adaptée à l'écriture manuscrite, des hypothèses avec les mots du lexique réduit ;
- et enfin, proposition d'une liste ordonnée de mots identifiés.

Nos travaux s'inspirent de la distance d'édition définie dans [?]. (cette distance est une extension de la distance de Damereau-Levenshtein pour l'écriture manuscrite). Nous l'avons adaptée au reconnaisseur ResifMot. Elle est basée sur six opérations d'édition sur les lettres : suppression, insertion, substitution, division, fusion et substitution de paires. Ces opérations permettent d'évaluer au mieux la notion de proximité entre les hypothèses du reconnaisseur et les mots du lexique en s'appuyant sur les erreurs possibles de reconnaissance et de segmentation.

Nous avons comparé l'encombrement mémoire, le temps de reconnaissance et le taux de reconnaissance pour un lexique de 6 915 mots, entre l'organisation décrite ci-dessus (sous-ensembles puis compression) et une simple compression (pas de partitions, mais compression du lexique complet sous

forme d'arbre lexical). Les résultats observés mettent en valeur un gain de temps : le temps de calcul moyen par mot est divisé par sept et une réduction du taux d'erreurs de reconnaissance de 12 % est obtenue ; mais c'est au détriment de l'encombrement mémoire : la taille du lexique est multipliée par 1,5 (ceci est dû au fait que la compression d'un lexique est plus efficace que la compression des partitions de ce même lexique, mais également que des redondances de mots sont introduites).

### 5.11 Système générique SPI, interface orientée stylet, saisie de formules mathématiques sur machine de poche

**Participants** : Laurent Pasquer, Éric Anquetil, Guy Lorette.

Nous avons développé en Java une version générique du système SPI sous la forme d'une boîte à outils nommée : DigiCarbon. Cette boîte à outils facilite la construction d'interfaces orientées stylet permettant l'édition et la manipulation de documents structurés simples.

Elle s'appuie, en entrée, sur une représentation précise des documents que l'on veut éditer et manipuler. Cette représentation des documents nécessaire pour DigiCarbon est composée de deux descriptions complémentaires au format XML :

- d'une part, la description graphique des différents symboles et gestes possibles,
- d'autre part, la description grammaticale de la structure du document.

La première description détaille précisément comment l'utilisateur doit tracer les différents symboles et gestes possibles. Cette description permettra au système d'identifier correctement les symboles que l'utilisateur tracera *via* l'interface stylet. La figure 12 présente un exemple de description d'un symbole.

```
<primitive nom="plus">
  <segment nom="verticale" angle="-90±30"/ >

  <segment nom="horizontale" angle="0±30"/ >
</primitive>
```

FIG. 12 – Exemple de description du symbole +

La seconde description décrit la structure du document. Elle permet à DigiCarbon de prédire quels sont les symboles que l'utilisateur peut tracer (et où ils doivent être situés) ainsi que les gestes de manipulations possibles dans le document courant. La figure 13 présente un la description d'une formule de type somme.

Grâce à cette double description, DigiCarbon va pouvoir, à tout moment, prédire les symboles et gestes que l'utilisateur peut tracer dans le document, ainsi que le nouveau document résultant de ces modifications.

Pour construire une véritable interface orientée stylet, qui permette l'édition et la manipulation de documents structurés, il suffit donc de décrire ces documents dans le formalisme XML utilisé par DigiCarbon. Ensuite, l'interface proprement dite s'appuiera sur DigiCarbon qui se chargera d'analyser, pendant leur tracé, les actions que l'utilisateur veut effectuer. L'exécution de ces actions se fera ensuite à la fois dans la structure manipulée par DigiCarbon (il faut qu'elle reste à jour), ainsi que du côté

```

<forme nom="somme">
  <composition>
    <formes-obligatoires>
      <ref-forme nom-forme="sigma" nom="sigma"/>
    </formes-obligatoires>
    <formes-facultatatives>
      <ref-forme nom-forme="formule" nom="dedans"/>
      <ref-positionnement nom="à_droite" liste="dedans,sigma"/>
      <ref-forme nom-forme="formule" nom="en_haut"/>
      <ref-positionnement nom="pos_au_dessus" liste="en_haut,sigma"/>
      <ref-forme nom-forme="formule" nom="en_bas"/>
      <ref-positionnement nom="pos_en_dessous" liste="en_bas,sigma"/>
    </formes-facultatatives>
  </composition>
</forme>

```

FIG. 13 – Exemple de description d'une formule de type somme ( $\Sigma$ )

de l'interface qui exploite DigiCarbon. L'interface peut aussi refuser ou modifier une action que la grammaire accepte<sup>4</sup>.

Un prototype de calculette a été construit en Java, et son portage sur machine de poche est en voie d'achèvement. Les documents possibles sont composés des symboles et formes suivants :

- les dix chiffres et la virgule,
- les quatre opérations de base : + - / \* ,
- la possibilité de construire des fractions et des sommes,
- un opérateur pour calculer le résultat numérique : = .

Quant aux gestes, ils sont pour l'instant très simples :

- gestes pour annuler/refaire,
- geste de sélection d'objet(s),
- geste de suppression de la sélection.

L'utilisateur peut ainsi écrire et modifier des formules simples de manière conviviale à l'aide du stylet. La figure 14 montre un exemple de formule tracée dans cette interface.

Nous sommes actuellement en train de terminer le portage de cette calculette sur machine de poche. Pour faire ce portage nous avons décidé de nous appuyer sur la machine virtuelle SuperWaba<sup>5</sup>.

Conçu pour fonctionner dans des environnements réduits, SuperWaba n'est pas entièrement compatible avec Java. Notre portage est toutefois en bonne voie d'aboutissement. Le principal souci qui nous a été évité est le portage de la phase d'analyse de la description XML.

Il est en effet exclu de faire cette analyse sur la machine de poche car, d'une part, cela pénaliserait l'utilisateur en prenant une place conséquente en mémoire (à la fois la description et l'analyseur XML), et d'autre part cela le pénaliserait en interactivité car l'interface aurait à analyser cette structure à chaque démarrage.

<sup>4</sup>Sur l'exemple de partitions musicales, les calculs des temps dans chaque mesure ne sont pas décrits dans la grammaire : seule l'interface d'exploitation les manipule.

<sup>5</sup>SuperWaba (<http://www.superwaba.org>) est une machine virtuelle Java libre et ouverte destinée aux machines de poches (PalmOS et PocketPC).

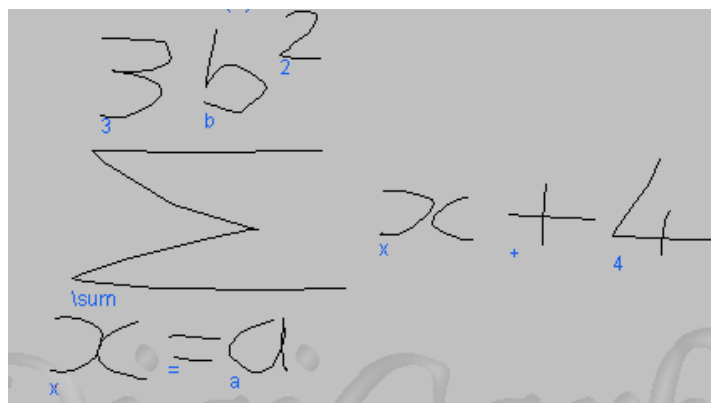


FIG. 14 – Capture de l'interface calculette

Notre principe ayant été dès le départ de séparer l'analyse de la description XML de son utilisation, nous imposons que la phase d'analyse soit effectuée sur une machine puissante. Ainsi, cette analyse peut se faire de manière plus poussée, aucune contrainte de temps ni de mémoire n'entrant en ligne de compte.

Lorsque le portage de cette interface sera terminé et validé, l'interface pourra facilement être adaptée à d'autres types de documents structurés simples. Pour ce faire, il suffira d'analyser la description XML de chacun de ces documents pour ensuite transférer ces résultats à l'interface.

Nous pensons notamment développer des interfaces d'édition de formules mathématiques plus complexes, de mini-partitions musicales et de graphes valués. Ceci montre le caractère générique de l'approche SPI implémentée dans la boîte à outils DigiCarbon.

## 5.12 Graphe de voisinage basé sur l'utilisation des distances discrètes

**Participants :** Ivan Leplumey, Charles Quéguiner.

Les distances discrètes sont largement utilisées en traitement d'images : morphologie mathématique, extraction rapide de squelette, reconstruction... Nous utilisons ces distances pour établir un graphe des positions relatives entre les composantes connexes d'une image. Ce travail [22] s'insère dans un projet plus vaste d'interprétation de documents écrits (textes, partitions musicales, formules mathématiques...)

Le fait de qualifier au mieux la relation de voisinage entre deux composantes nous semble un élément primordial sur lequel il est possible de s'appuyer pour interpréter une image. La qualification proposée consiste à obtenir les liens entre deux composantes proches **A** et **B**. Une structure de graphe est la structure la plus naturelle pour gérer un tel ensemble de liens.

La construction du graphe se découpe en trois phases (cf. Figure 15) :

- calcul des composantes connexes ;
- calcul d'une image de distances (avec propagation conjointes d'attributs) ;
- calcul du graphe.



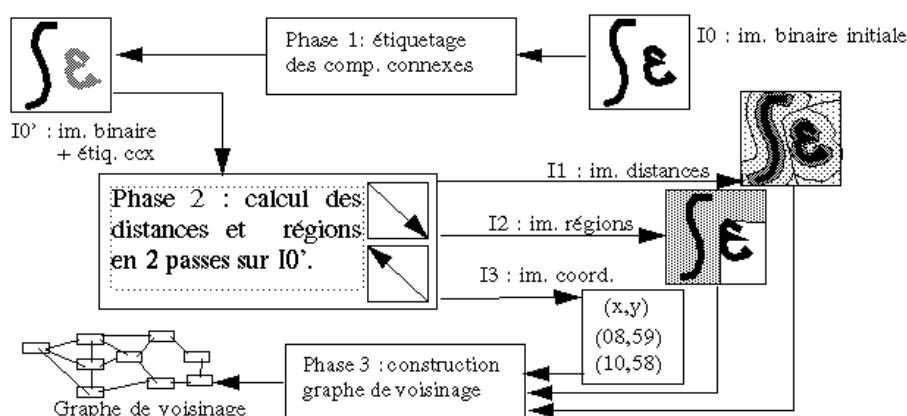


FIG. 15 – Schéma général de l'extraction, en trois phases, du graphe de voisinage

Le calcul des distances est illustré par la figure 16 B ; tandis que le graphe est illustré par un graphe de Delaunay (cf. Figure 17 A).

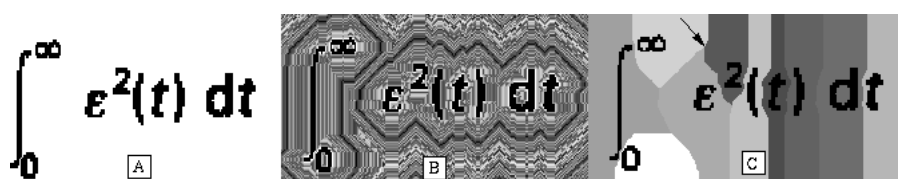


FIG. 16 – **A**-Image originale (formule mathématique). **B**- Image des distances ( $d_1$  : city block) entre les composantes connexes. **C**- Pavage de Voronoï de l'image d'origine en appliquant la distance discrète  $d_1$  (4-connexe).

Pour tenir compte de la forme spécifique des documents écrits (hauteur relative des composantes connexes fines), nous représentons les liens gérés dans le graphe par un doublet de points **a** ( $x_1, y_1$ ) et **b** ( $x_2, y_2$ ), où le point **a** (resp.**b**) appartient au contour de la composante **A** (resp. **B**). Ayant défini une mesure de distance discrète, ces deux points **a** et **b** sont choisis de telle sorte que la distance de **a** vers **b** soit minimum.

Ceci conduit à l'obtention d'un graphe "entre points contours" (cf. figure 17 B), liant entre elles toutes les composantes dont les zones d'influences (voir pavage de Voronoï figure 16 C) sont connexes. Le couple de points (**a**, **b**) à l'origine du lien entre 2 composantes est conservé comme un attribut de ce lien.

L'implémentation naturelle est relativement gourmande en espace mémoire (nombreuses structures matricielles à gérer).

Nous proposons une nouvelle implémentation ayant des besoins moins importants en taille mémoire et en temps de calcul ; celle-ci en regroupant les phases 2 et 3, calcule à la volée un graphe de voisinage, à l'aide d'une mémorisation locale sur deux lignes adjacentes.

La construction du graphe en une seule passe (NW→SE) sur l'image donne des résultats satisfai-

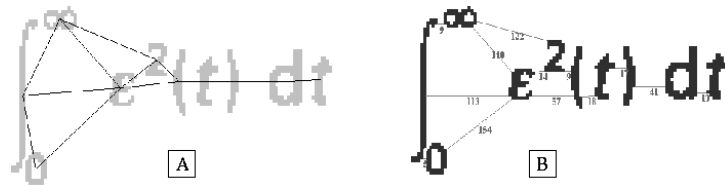


FIG. 17 – **A-** Graphe de Delaunay (lien entre composantes voisines). **B-** Graphe entre les points contours des différentes composantes.

sants pour notre problème de positionnement relatif des composantes les unes par rapport aux autres. Cependant, étant donné le très faible coût de cette implémentation, il nous est apparu intéressant de faire une seconde passe (SE→NW) et de faire coopérer les deux graphes obtenus afin de ne conserver que les liens les plus pertinents.

## 6 Contrats industriels (nationaux, européens et internationaux)

### 6.1 Embarquement du système ResifCar sur des téléphones mobiles

**Participants** : Eric Anquetil, Laurent Pasquer.

- Référence de la licence d'exploitation : INSA G223,
- Période : décembre 2001 à décembre 2003.

ResifCar est un système de reconnaissance de caractères manuscrits isolés (cf. paragraphes 4.1 et 5.9). Il a été embarqué dans des téléphones mobiles de nouvelle génération «smartphone».

Le contrat de collaboration avec la société PurpleLabs (Réf. INSA OE18) qui conçoit des nouveaux téléphones mobiles pour en vendre le design à des fabricants s'est prolongé par une licence d'exploitation commerciale signée en décembre 2001 (Réf. INSA G223).

La première version de téléphone embarquant la technologie ResifCar est aujourd'hui commercialisée par la société italienne Enteos. Ce mobile regroupe les fonctionnalités d'un PDA avec celles d'un téléphone.

### 6.2 Annotation collective de documents d'archives

**Participants** : Bertrand Couïasnon, Ivan Leplumey, Hélène Richy, Nicolas Briec.

Les relations avec les Archives départementales de la Mayenne et celles d'Ille-et-Vilaine, nous ont permis de présenter à la région un projet fédérateur sur un travail coopératif d'annotation des documents des archives à travers le web.

- Numéro de la convention : INSA 1E23,
- Période : octobre 2001 à décembre 2002.

L'équipe IMADOC et les Archives de la Mayenne et de l'Ille-et-Vilaine travaillent au sein d'un projet Mégalis, piloté par les régions Bretagne et Pays de la Loire. Le travail effectué dans le cadre de

ce contrat est décrit dans la section 5.1.

### 6.3 Accès en ligne aux documents d'archives

**Participants** : Bertrand Couïasnon, Eric Anquetil, Laurent Pasquer, Jean Camillerapp, Ivan Leplumey, Grégory Maitrallain.

- Référence du contrat : INSA G222, INRIA 1 02 C 0602 00 31406 01 1
- Période : Juillet 2002 à juin 2004.

Le département des Yvelines entend faciliter l'accès aux archives conservées par les Archives Départementales des Yvelines. Pour cela il souhaite permettre leur consultation soit par internet soit par intranet.

Dans ce but le département est en train de faire numériser les cahiers de recensement de la population (26 mètres linéaires) et les registres de matricules (55 mètres linéaires) [25]. Pour que le public puisse facilement avoir accès à ces documents numérisés, il faut les indexer. Cela nécessite un important travail de reconnaissance de la structure des documents et de reconnaissance de certaines mentions manuscrites.

Pour la reconnaissance de la structure des registres matricules, Imadoc va pouvoir repartir du travail fait pour les Archives de La Mayenne. Cette analyse sera complétée par une reconnaissance du nom manuscrit.

La reconnaissance des registres de recensement, nécessite d'une part une adaptation de la méthode DMOS à ce nouveau type de documents et d'autre part des recherches sur la reconnaissance de l'écriture manuscrite.

Parallèlement à ces travaux sur les documents, Imadoc va travailler sur la définition d'une interface stylet destinée à faciliter le dialogue entre le public et le système d'information.

## 7 Actions régionales, nationales et internationales

### 7.1 Actions régionales

Jacques André fait partie du conseil scientifique du *Musée des transmissions et de l'informatique* de l'Armée de terre (ESAT, Rennes).

Jean Camillerapp fait partie des coordinateurs régionaux d'expertise des conventions CIFRE auprès de l'ANVAR.

Guy Lorette fait partie des conseils scientifiques des laboratoires PSI (Université-INSA de Rouen), L3I (Université de La Rochelle) et du laboratoire d'informatique (E3I-Université de Tours).

## 7.2 Actions nationales

### 7.2.1 ICL

Jacques André collabore avec l'ICL (*Institut Culturel du Livre*, Lyon) où il a initialisé la page « Musées européens de typographie<sup>6</sup> » et prépare une *Bibliothèque virtuelle de typographie*<sup>7</sup>.

### 7.2.2 RTP

Participation du projet Imadoc aux Réseaux Thématiques Pluridisciplinaires :

- RTP-DOC (STIC) :
  - AS « Numérisation et valorisation des collections » : Jacques André, Bertrand Couïasnon
  - AS « Modèle(s) de publication sur le web » : participation de Jacques André et gestion de l'AS par l'Irisa.

### 7.2.3 GdR

Eric Anquetil, Bertrand Couïasnon, Guy Lorette et Laurent Pasquer participent au groupe de travail « 5.2-Écrit », thème 5 : Communication, du GDR-PRC I<sup>3</sup> (Information, Interaction, Intelligence). Laurent Pasquer et Guy Lorette ont fait une présentation au colloque « Informatique Mobile » organisé par le groupe de travail « 4.2-Mobilité et ubiquité » lors des assises I<sup>3</sup> (Nancy, décembre 2002).

### 7.2.4 Sociétés savantes

Jacques André est membre du conseil d'administration de l'association GUTenberg.

Jean Camillerapp, Guy Lorette, Eric Anquetil et Bertrand Couïasnon participent aux activités de l'association GRCE : « Groupe de Recherche en Communication Écrite ».

Guy Lorette est membre :

- du GRCE et vice-président en charge des relations internationales ;
- de l'AFRIF : « Association Française de Reconnaissance et d'Interprétation de Formes », de son CA et représentant à l'IAPR ;
- de l'ASTI.

### 7.2.5 Listes de discussion

Jacques André est responsable de la liste *typographie* hébergée à l'Irisa ([sympa@irisa.fr](mailto:sympa@irisa.fr)) et archivée sous l'adresse : <https://www.irisa.fr/wws/info/typographie>.

---

<sup>6</sup><http://histoire.enssib.fr/icl-forum/>

<sup>7</sup><http://www.irisa.fr/faqtypo/BiViTy.html>

### 7.3 Actions internationales

Jacques André participe à la version française<sup>8</sup> du standard Unicode (norme ISO-10646).

Guy Lorette est responsable des relations internationales de l'IFSIC, membre de la CAI (commission des affaires internationales) de l'Université de Rennes 1 ainsi que le représentant français au *Governing Board* de l'IAPR (*International Association of Pattern Recognition*) et membre du *Constitutions and Bylaws Committee* de cette association.

Guy Lorette a été membre du jury ou rapporteur pour des jurys de thèses de l'HKBU (Hong-Kong) et de l'ENIS (Sfax, Tunisie).

## 8 Diffusion de résultats

### 8.1 Animation de la Communauté scientifique

#### 8.1.1 Comité éditorial de journaux

Jacques André est :

- rédacteur en chef de la revue *Document numérique* des éditions Hermès, qui a publié en 2002 :
  - « Nouvelles écritures » (vol. 5, n°1-2, 208 pages)
  - « Espaces numériques d'information et de coopération » (vol. 5, n°3-4, 230 pages)
  - « Dossiers numériques » (vol. 6, n°1-2, 176 pages)
  - « Unicode : écriture du monde ? » (vol. 6, n°3-4, 248 pages) ;
- rédacteur en chef des *Cahiers GUTenberg* ;
- responsable de la rubrique « Documents numériques » de l'encyclopédie *Techniques de l'ingénieur – Informatique* ; cette rubrique paraît aussi désormais tous les six mois sous forme d'un CD ;
- membre de l'*Editorial board* de la *Typography series* de Addison-Wesley.

Guy Lorette est membre du comité de rédaction des revues :

- *Pattern Recognition*,
- *IJDAR—International Journal of Document Analysis and Recognition*.

#### 8.1.2 Comités de programme de conférences

- Jacques André a été ou est membre du comité de programme de CFD (Tunis, 2002), CIDE'6 (Caen, 2003), Didapro (Paris, 2003), Colloque Histoire de l'Informatique (Grenoble, 2002) et est expert auprès de *European Science Foundation Conferences*.
- Jean Camillerapp a été membre du comité de programme de CFD (Tunis, 2002).
- Guy Lorette a été ou est membre des comités de programme de : RFIA'2002 (Angers, 2002), IWFHR'8 (Niagara on the Lake, 2002), CFD (Tunis, 2002).

---

<sup>8</sup><http://hapax.i.quebec.com>

- Bertrand Couïasnon est membre du comité de programme de WEDELMUSIC (3rd International Conference on Web Delivering of Music) (Leeds, 2003).
- Éric Anquetil et Guy Lorette sont membres du comité de candidature pour organiser IWFHR'10 en France (en 2006) et cette candidature a été acceptée lors de IWFHR'8.

## 8.2 Enseignement

L'équipe étant en grande partie constituée par des enseignants chercheurs, ceux-ci sont très impliqués dans les activités d'enseignement, mais comme la plupart ne sont pas rattachées à ce thème de recherche, elles ne sont pas citées ici.

### Cours spécifiques :

- Jacques André assure un enseignement de typographie numérique en 5<sup>e</sup> année des ERBA (École Régionale des Beaux-Arts) de Rennes et d'Angers et est membre du Diplôme supérieur de création typographique à l'École Supérieure Estienne à Paris.
- Guy Lorette participe au DEA d'informatique, à l'IFSIC, Université de Rennes 1, (cours perception, reconnaissance et interprétation d'images).
- Hélène Richy est responsable du DESS Méthodes Informatiques et Technologies de l'Information et de la Communication (Mitic) à l'IFSIC, Université de Rennes 1. Ce DESS accueille une vingtaine d'étudiants depuis septembre 2001.  
Hélène Richy assure un enseignement sur les documents numériques en DESS Méthodes Informatiques et Technologies de l'Information et de la Communication, à l'IFSIC, Université de Rennes 1.

### Conférences :

- Jacques André a donné une conférence « Règles typographiques vues comme méthodes d'ingénierie du document électronique », module DN, ENST-Bretagne, Brest.
- Éric Anquetil et Guy Lorette ont participé à l'enseignement en 3<sup>e</sup> année de la filière « Systèmes d'information statistique » de l'ENSAI (cours : reconnaissance à l'aide de réseaux de neurones, reconnaissance de l'écrit manuscrit).
- Guy Lorette a donné une conférence sur la reconnaissance de l'écrit manuscrit en DESS Méthodes Informatiques et Technologies de l'Information et de la Communication, à l'IFSIC, Université de Rennes 1.
- Hélène Richy a donné une conférence sur XML en 5<sup>e</sup> année INSA.
- Hélène Richy a donné un cours sur les documents numériques en DESS NTIM à l'Université de Cocody (Abijan, Côte d'Ivoire) du 22 au 27 avril 2002.

### Stages effectués au sein du projet :

- Fabrice Boyer, Sabine Carbonnel et Yohann Nicolas (DEA d'informatique, IFSIC).
- Fabien Lescellière-Dumilly, Éric Morillon et Nicolas Bunel (stage 4<sup>e</sup> année INSA).
- Eimear Vallely (stage Erasmus, Belfast University, Irlande).

### 8.3 Participation à des colloques, séminaires, invitations

- Jacques André a donné un exposé « Codage des caractères, d'Ascii à Unicode » lors de la Journée sur les multilinguisme et multiscripts, Mediadix, Saint-Cloud, juin 2002.
- Éric Anquetil a donné une conférence « Intégration du système de reconnaissance de caractères manuscrits ResifCar dans des téléphones mobiles » lors de la Journée sur la Commande et le Diagnostic : Applications industrielles de la logique floue et des réseaux de neurones, conférence organisée par JESSICA OUEST, MEITO/ INSA/Pays de la Loire Innovation, Rennes, janvier 2002.
- Bertrand Couïasnon a donné une conférence sur l'accès aux documents numérisés, pour le Ministère de la culture aux Archives départementales de la Mayenne à Laval (avril 2002) et une sur le « Traitement des registres matricules du XIX<sup>e</sup> siècle aux Archives de la Mayenne » lors de la journée Valorisation des documents et numérisation des collections » (ISDN/ENS, Lyon, mars 2002) [27].
- Bertrand Couïasnon, Jean Camillerapp et Ivan Leplumey ont présenté leurs travaux aux journées « Stockage, traitement et diffusion des bases d'images numériques » organisées par les Archives départementales de la Mayenne, Laval, mars 2002 [26].
- Bertrand Couïasnon et Jacques André ont donné des exposés lors du tutoriel « Numérisation et patrimoine » de la conférence CFD, Hammamet (Tunisie), octobre 2002.

## 9 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] J. ANDRÉ, M.-A. CHABIN (éditeurs), *Les documents anciens*, Numéro spécial de *Document numérique*, vol. 3, n°1-2, Hermès, 1999.
- [2] E. ANQUETIL, G. LORETTE, « Automatic generation of hierarchical fuzzy classification systems based on explicit fuzzy rules deduced from possibilistic clustering: application to on-line handwritten character recognition », in : *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, (IPMU'96), p. 259–264, Grenade, Espagne, 1996.
- [3] B. COÛASNON, J. CAMILLERAPP, « A way to separate knowledge from program in structured document analysis: application to optical music recognition », in : *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, (ICDAR'95), p. 1092–1097, Montréal, Canada, 1995.
- [4] V. P. D'ANDECY, J. CAMILLERAPP, I. LEPLUMEY, « Analyse de partitions musicales », *Traitement du Signal* 12(6), 1996, p. 653–661.
- [5] G. LORETTE, E. ANQUETIL, « Théorie des catastrophes, géométrie différentielle et segmentation de l'écriture cursive », in : *Actes du 4ème Colloque National sur l'Écrit et le Document*, (CNED'96), p. 1–6, Nantes, France, 1996.
- [6] G. LORETTE, J.-P. CRETTEZ, « Reconnaissance de l'écriture manuscrite », in : *Traité informatique*, Techniques de l'Ingénieur, 1998, ch. H 1358, p. 1–15.
- [7] G. LORETTE, Y. LECOURTIER, « Is recognition and interpretation of handwritten texts a scene analysis problem? », in : *Proceedings of the 3rd International Workshop in Frontiers in Handwriting Recognition*, (IWFHR3), p. 184–196, Buffalo, USA, 1993.
- [8] G. LORETTE, « Handwriting recognition or reading? What is the situation at the dawn of the 3<sup>rd</sup> millennium? », *International Journal on Document Analysis and Recognition* 2, 1, 1999, p. 2–12.

- [9] L. PASQUER, ÉRIC ANQUETIL, G. LORETTE, « Coherent Knowledge Source Integration through Perceptual Cycle Framework for Handwriting Recognition », in : *Advances in Handwriting Recognition*, S.-W. Lee (éditeur), World Scientific, 1999, ch. 5 in PART2 : On-Line Handwriting Recognition, p. 59–68.
- [10] J. PETTIER, J. CAMILLERAPP, « Script representation by a generalized skeleton », in : *Proceedings of the 2nd International Conference on Document Analysis and Recognition, (ICDAR'93)*, p. 850–853, Tsukuba, Japon, 1993.

### Livres et monographies

- [11] J. ANDRÉ, G. DUPOIRIER (éditeurs), *Documents numériques*, Techniques de l'Ingénieur, 2002, (existe en CD ; voir aussi : <http://www.techniques-ingenieur.fr>).
- [12] J. ANDRÉ, H. HUDRISIER (éditeurs), *Unicode, écriture du monde*, Lavoisier-Hermès, 2002, <http://www.hermes-science.com>.

### Articles et chapitres de livre

- [13] J. ANDRÉ, « Digital Forms of Typography », in : *A History of Writing - from Hieroglyph to Multimedia*, A.-M. Christin (éditeur), Flammarion, 2002, p. 384–385.
- [14] J. ANDRÉ, « L'échelle arithmétique des corps », in : *Le Romain du roi – la typographie au service de l'État, 1702-2002*, J. Mosley et S. de Turckheim-Pey (éditeurs), Musée de l'Imprimerie, Lyon, 2002, p. 79–80.
- [15] J. ANDRÉ, « Le codage des caractères, de Chappe à Unicode », *Document Numérique* 6, 3-4, 2002, p. 13–49.
- [16] B. COÛASNON, J. CAMILLERAPP, « Une méthode générique de rétroconversion de documents pour la constitution de dossiers numériques », *Document Numérique* 6, 1-2, 2002, p. 129–144.
- [17] P. GARCIA, B. COÛASNON, « Using a Generic Document Recognition Method for Mathematical Formulae Recognition », in : *Graphics Recognition: Algorithms and Applications, LNCS, 2390*, Springer-Verlag, 2002, p. 236–244.
- [18] H. RICHY, « Métadonnées et documents numériques », in : *Traité informatique*, H7155, Techniques de l'Ingénieur, 2002, p. 1–14, disponible sur CD et à : <http://www.teching.com/>.
- [19] H. RICHY, « Traitement électronique des documents anciens », *éc/artS#3, spécial Colloque\_Montréal, Textualités et nouvelles technologies*, 02\_03, 2002, p. 391–395.

### Communications à des congrès, colloques, etc.

- [20] E. ANQUETIL, H. BOUCHEREAU, « Integration of an On-line Handwriting Recognition System in a Smart Phone Device », in : *Proceedings of the 16th IAPR International Conference on Pattern Recognition (ICPR 2002)*, p. 192–195, Quebec, 2002.
- [21] B. COÛASNON, J. CAMILLERAPP, « DMOS, une méthode générique de reconnaissance de documents : évaluation sur 60 000 formulaires du XIX<sup>e</sup> siècle », in : *Actes du Colloque International Francophone sur l'Écrit et le Document (CIFED'02)*, Hammamet, octobre 2002.
- [22] I. LEPLUMEY, C. QUÉGUINER, « Un graphe de voisinage basé sur l'utilisation des distances discrètes », in : *Actes du Colloque International Francophone sur l'Écrit et le Document (CIFED'02)*, Hammamet, octobre 2002.



- 
- [23] N. RAGOT, E. ANQUETIL, « Combinaison hiérarchique de systèmes d'inférence floue : application à la reconnaissance en-ligne de chiffres manuscrits », in : *Actes du congrès CIFED'02, Conférence Fédérative sur l'Écrit et le Document*, Hammamet, octobre 2002.

## Divers

- [24] J. ANDRÉ, « Abréviations et sigles d'Unicode, *Document Numérique* », 6, 3-4, 2002.
- [25] B. COÜASNON, J. CAMILLERAPP, « Spécifications pour la numérisation des tables de recensements civils et des registres matricules aux Archives départementales des Yvelines », in : *Rapport interne pour les Archives départementales des Yvelines*, novembre 2002.
- [26] B. COÜASNON, I. LEPLUMEY, J. CAMILLERAPP, « Consultation à distance des documents numérisés », in : *Journées « Stockage, traitement et diffusion des bases d'images numériques »*, Archives départementales de la Mayenne, Laval, mars 2002.
- [27] B. COÜASNON, I. LEPLUMEY, J. CAMILLERAPP, « Le traitement des registres matricules aux Archives de la Mayenne », in : *Journées « Valorisation des documents et numérisation des collections »*, ISDN/ENS, Lyon, mars 2002.