



Collaboration entre approches statistiques et structurelle pour la reconnaissance de caractères anciens

Denis Arrivault, Noël Richard, Christine Fernandez-Maloigne, Philippe Bouyer

► To cite this version:

Denis Arrivault, Noël Richard, Christine Fernandez-Maloigne, Philippe Bouyer. Collaboration entre approches statistiques et structurelle pour la reconnaissance de caractères anciens. Jun 2004. sic_00001216

HAL Id: sic_00001216

https://archivesic.ccsd.cnrs.fr/sic_00001216

Submitted on 7 Dec 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collaboration entre approches statistique et structurelle pour la reconnaissance de caractères anciens

Denis Arrivault^{1,2} – Noël Richard¹ – Christine Fernandez-Maloigne¹ – Philippe Bouyer²

¹ Laboratoire SIC - CNRS - FRE 2731

SP2MI, Boulevard Marie et Pierre Curie - F-86 962 FUTUROSCOPE CEDEX

² RC-SOFT, Domaine de la Combe - BP39 - F-16 710 Saint-Yriex

denis.arrivault@etu.univ-poitiers.fr

philippe.bouyer@rcsoft.fr

Résumé : *Dans le cadre de cet article nous cherchons à faire coopérer plusieurs types d'informations pour construire un système de reconnaissance de caractères issus de manuscrits anciens. Nous avons choisi d'exploiter, d'une part une approche statistique classique basée sur des descripteurs de formes (Zernike, Fourier) et, d'autre part une approche structurelle exploitant une modélisation du caractère avec un graphe d'attributs associé à une modélisation des classes par un graphe aléatoire. Compte tenu du nombre variable d'individus et des évolutions dans le temps des écrits, une approche purement statistique ne peut suffire pour obtenir des résultats satisfaisants. Nous proposons une stratégie de mise en coopération basée sur la gestion du rejet.*

Mots-clés : caractères anciens, reconnaissance statistique, reconnaissance structurelle, graphes aléatoires, coopération.

1 Introduction

Ce travail rentre dans le cadre des systèmes d'aide à la reconnaissance des caractères anciens. Il est partie intégrante du projet Eureka COROC¹, initié par la société RC-Soft, qui s'articule essentiellement autour des manuscrits écrits en grec anciens (époques médiévales et antérieures) et des documents hiéroglyphiques. Dans cet article nous nous contenterons d'exposer nos travaux portant sur le grec ; le hiéroglyphe faisant intervenir un degré de complexité supérieur lié au très grand nombre de pictogrammes de références (classes).

Cependant, l'idée sous-jacente à notre propos est que, compte tenu des variations et des variantes de ces caractères, notamment dans la complexité des hiéroglyphes, il ne peut être imaginé la résolution de ce problème par une seule méthode. C'est pourquoi, notre point de vue réside dans le choix de méthodes susceptibles de fournir des informations différentes. Le terme méthode implique ici l'association d'attributs avec un outil de classification.

Combiner plusieurs méthodes implique de disposer d'informations à redondance limitée. Dans notre cas, nous avons choisi deux gammes d'informations différentes : une gamme centrée autour de descripteurs statistiques et une seconde basée sur une description structurelle du symbole. Ce que nous appelons descripteurs statistiques s'inscrit dans les

techniques d'identification classiques en reconnaissance des formes et par extension pour les systèmes d'OCR. Nous avons choisi d'utiliser dans cette catégorie différents attributs de description de forme (Hu, Zernike, ...).

La seconde gamme de méthodes vise à exploiter des informations de structure sur le symbole. La technique choisie est celle de l'extraction d'un graphe décrivant le squelette du symbole. Cette technique qui permet de décrire plus finement la composition d'un symbole apporte une information plus complète mais aussi plus complexe. L'enjeu de la coopération des deux approches est donc d'optimiser le coût combinatoire de décision, mais aussi de pouvoir confirmer ou infirmer des hypothèses de décision.

Nous développerons notre propos selon le schéma suivant. Dans un premier temps nous présenterons les objectifs du projet COROC dans lequel s'inscrit ce travail et nous exposerons brièvement le schéma global de notre système. Dans un second temps, nous présenterons le développement de l'approche statistique qui doit fournir, un certain nombre d'hypothèses de décision. Par la suite nous présenterons la méthode complémentaire, basée sur une description structurelle du caractère. Avant de conclure et de présenter quelques résultats, nous reviendrons sur la manière d'associer ces deux catégories d'informations.

2 Objectifs

L'objectif du projet COROC est de concevoir un logiciel faiblement supervisé qui, à partir d'un thésaurus minimal, est capable de s'autoévaluer pour faire intervenir l'expert et optimiser ses paramètres. Le logiciel traitera les documents anciens comme le grec, à la base de nos tests, dont la figure 1 donne un exemple.

Nous n'aborderons pas ici la problématique extrêmement complexe de la segmentation qui fait partie d'un autre volet du projet COROC. Le thésaurus est, pour les résultats de cet article, constitué exclusivement de caractères bien segmentés (isolés et sans occlusion).

Comme nous l'avons déjà indiqué, du fait de la grande variabilité des caractères et de la diversité des tailles de thésaurus à traiter, il nous paraît indispensable de combiner plusieurs approches de reconnaissance. Les méthodes statistiques sont

¹Cognitive Optical Recognition Old Characters

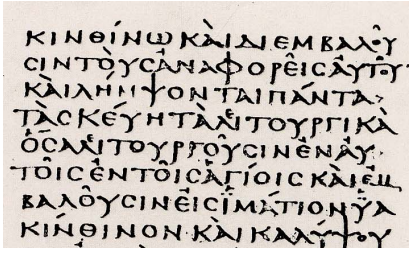


FIG. 1 – Exemple de document grec

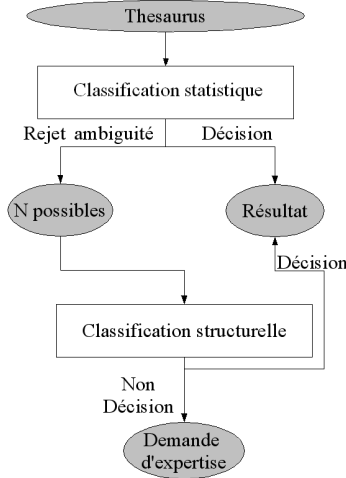


FIG. 2 – Schéma fonctionnel de mise en coopération

relativement simples et peu coûteuses en temps de calcul (surtout pour l'approche paramétrique ; l'approche non paramétrique ne sera, pour cette raison, utilisée que pour les thésaurus de faible taille). A l'inverse les méthodes structurales sont plus complexes et plus coûteuses en temps de calcul. Nous avons donc fait le choix, a priori, de combiner les deux approches en cascade afin de n'avoir recours à la classification structurale que sur des cas rejetés par les approches statistiques. La figure 2 donne le schéma global de mise en coopération, nous y reviendrons dans la partie 5.

3 L'approche Statistique

3.1 Principe

Depuis les débuts de la reconnaissance des formes, de nombreuses méthodes exploitant une description statistique d'attributs ont été développées. L'objet de ces approches est de pouvoir décrire des comportements spécifiques à partir de modèles simples à utiliser. Le schéma général d'exploitation est de prendre un ou plusieurs attributs de description de forme et de décrire l'organisation des individus dans l'espace des attributs au travers de modèles.

Parmi les différents descripteurs de forme, nous avons choisi de conserver les descripteurs de Fourier et Zernike pour avoir à la fois une informations sur la forme (Zernike) et sur le contour (Fourier). Nous rappellerons tout d'abord comment sont obtenus ces descripteurs puis nous expliciterons brièvement les principes de la classification paramétrique et non paramétrique.

3.2 Les descripteurs de forme

Les deux descripteurs de forme utilisés permettent d'établir un vecteur d'attributs invariants selon les principales transformations (homothétie, rotation). La première méthode décrit le contour externe du symbole (descripteur de Fourier) alors que l'autre méthode permet de décrire la forme du symbole en tant que région (Moments de Zernike). Dans une application classique de reconnaissance de caractères, un seul de ces descripteurs aurait pu suffire. Cependant dans la perspective d'un vaste thésaurus inconnu a priori (comme pour l'égyptien), le fait d'exploiter contour extérieur d'un coté et forme complète de l'autre nous apparaît plus judicieux.

3.2.1 Description du contour : descripteurs de Fourier

Le principe est d'échantillonner le contour sur un nombre de points préfixé afin de calculer la Transformée de Fourier Discrète. Les coefficients fréquentiels normalisés ainsi obtenus forment les descripteurs. Il est alors possible de calculer des descripteurs selon la courbure ou selon la distance au centre de gravité. D.Zhang et G.Lu ont montré que cette seconde solution est plus performante ([ZHA 02]).

Le contour étant normalisé à M points, la distance au centre de gravité est établie pour chaque point (eq.1). De la DFT (eq.2), seul le module est conservé au vu de son invariance en rotation. Le vecteur d'attribut conservé a ses coefficients normalisés par $|F_0|$ pour s'affranchir des changements d'échelle (eq.3).

$$R(t) = \sqrt{(x(t) - x_c(t))^2 + (y(t) - y_c(t))^2} \quad (1)$$

$$F_i = \frac{1}{M} \sum_{t=0}^{M-1} R(t) \exp\left(\frac{-j2\pi it}{M}\right) \quad i = 0, 1, \dots, M-1 \quad (2)$$

$$f_R = \left\{ \frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{\frac{M}{2}}|}{|F_0|} \right\} \quad (3)$$

3.2.2 Description de la forme : moments de Zernike

Les moments A_{nm} de Zernike correspondent à la projection de la forme $f(x, y)$ sur une base ZP de fonctions orthogonales $V_{nm}(x, y)$ (eq.4)

$$A_{nm} = \frac{n+1}{\pi} \sum_{(x,y) \in I} V_{nm}(x, y)^* f(x, y) \quad (4)$$

* définissant le complexe conjugué. La base ZP est définie sur le cercle unité par :

$$ZP = \{V_{nm}(x, y) \mid (x^2 + y^2) \leq 1\} \quad (5)$$

où le polynôme complexe V_{nm} d'ordre n et de répétition m est défini avec $n \in \mathbb{N}^+$ et $m \in \mathbb{N}$ tel que $n - |m|$ soit pair et $|m| \leq n$ (eq.6).

$$V_{nm}(x, y) = R_{nm}(r) e^{jm\theta} \text{ en coordonnées polaires} \quad (6)$$

avec :

$$R_{nm}(r) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)! r^{n-2s}}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \quad (7)$$

Les moments sont invariants par rotation, translation et changement d'échelle (après normalisation de la taille de la forme). De plus, grâce à l'exploitation d'une base de fonctions orthogonales, ces moments sont peu corrélés ([KHO 90], [CHO 03]).

3.3 Classification

Nous utilisons deux types de classification suivant la taille du thésaurus.

Pour les thésaurus suffisamment fournis (plus de 10 éléments par classe), nous avons utilisé une classification paramétrique après réduction des données par une analyse en composantes principales. Les distributions sont supposées gaussiennes (eq.8) et nous utilisons la loi de Bayes pour obtenir les probabilités a posteriori (eq.9).

$$p(X|\omega_i) = \frac{1}{(2\pi)^{-\frac{d}{2}}|\Sigma_i|^{-1}} e^{-\frac{1}{2}(X-\bar{X}_i)^T \Sigma_i^{-1}(X-\bar{X}_i)} \quad (8)$$

où \bar{X}_i est le vecteur moyen de la classe et Σ_i est la matrice de variance-covariance de la classe.

L'estimation de $p(\omega_k|X)$ se fait alors en utilisant la loi de Bayes :

$$p(\omega_i|X) = \frac{p(\omega_i)p(\omega_i|X)}{p(X)} \quad (9)$$

Pour les thésaurus plus petits nous utilisons une classification non paramétrique de type k plus proches voisins. La probabilité a posteriori s'estime alors comme dans l'équation 10.

$$\hat{p}(\omega_i|X) = \frac{\hat{p}(X, \omega_i)}{\sum_{j=1}^C \hat{p}(X, \omega_j)} = \frac{k_i}{k} \quad (10)$$

La prise de décision dans ces deux approches est effectuée classiquement suivant la règle de Bayes en minimisant la probabilité a posteriori.

4 La classification structurelle

D'un point de vue général, la classification structurelle utilise une décomposition du caractère en objets primitifs et permet de décrire une partie de l'organisation de ces objets les uns par rapport aux autres. Plusieurs types de représentation permettent de stocker ces structures, soit au travers de cartes topologiques, soit par l'intermédiaire de graphes d'adjacences. L'intérêt de la première représentation est de décrire complètement l'organisation spatiale du caractère, cependant il n'existe pas à ce jour de formalisme satisfaisant pour la mesure de similarité entre cartes topologiques [DAM 04]. Nous utiliserons donc une structure de graphe d'adjacences en essayant d'inclure un maximum d'informations topologiques dans les attributs.

Le graphe d'adjacences d'un symbole se construit par la description des relations entre les points d'intérêt caractérisant le symbole (figure 3). La notion de point d'intérêt dans notre cas correspond aux points terminaux des segments et aux points d'intersection des segments dans le squelette extrait du symbole.

Pour permettre la comparaison de deux graphes, un certain nombre d'attributs sont associés aux noeuds et aux arcs. La structure de représentation est présentée en première partie



FIG. 3 – Exemple de squelettes et de points caractéristiques

de cette section. Chaque classe du thésaurus étant composée de plusieurs graphes d'attributs, ils seront fusionnés dans un même graphe dont la structure sera celle d'un graphe dit "modèle", décrit par des attributs aléatoires. Ces graphes aléatoires sont détaillés dans la deuxième partie. Enfin la troisième partie s'attachera à décrire le problème d'appariement et de classification.

4.1 Modélisation du caractère : graphe d'attributs

La construction du graphe d'attributs à partir d'un caractère se fait en trois étapes. L'étape initiale est la construction du squelette du symbole, se succèdent ensuite l'extraction des points d'intérêt et le calcul des attributs du graphe.

Le squelette est une représentation synthétique de la forme ([BLU 67]) qui fournit des segments de largeur unitaire. Parmi les diverses méthodes existantes pour extraire un squelette ([LAM 92]), seules celles capables de travailler en 8-connextité sont susceptibles de fournir correctement les points d'intersection entre deux segments [ZHA 94].

La figure 3 présente quelques exemples de caractères et leurs squelettes avec la méthode de Zhang & Wang. Les points caractéristiques détectés sont en noir sur le squelette gris.

Une fois détectés dans le squelette, les points caractéristiques sont assimilés aux noeuds dans la structure de graphes d'attributs. Les arcs représentent les segments qui les relient dans le squelette.

Le choix des caractéristiques pour les noeuds et les arcs est essentiellement topologique : on cherche à décrire la topologie du squelette. Ainsi, les attributs liés aux noeuds représentent les coordonnées polaires du point dans un repère centré sur le centre de gravité et ayant pour axe principal la droite de moindres carrés. Nous avons gardé le sinus et le cosinus (plutôt que l'angle, pour des raisons calculatoires) ainsi que la distance au centre, ce qui fait trois attributs normalisés entre 0 et 1 (fig.4).

Les attributs associés aux arcs sont la courbure (rapport entre la longueur du segment et la distance séparant les extrémités), ainsi que le rapport entre la longueur du trait et la longueur totale du squelette. Ces deux attributs sont normalisés entre 0 et 1. La modélisation autorise l'intégration des boucles (segments dont les noeuds terminaux sont identiques). Dans ce cas particulier, la courbure est initialisée à 1.

De manière plus formelle ([LU 91]), un graphe d'attributs est décrit de la manière suivante :

Définissons :

- $\Sigma_v = \{v_k | k = 1, \dots, n\}$ un ensemble de sommets,

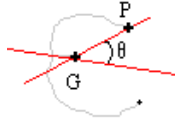


FIG. 4 – Attributs d'un noeud



FIG. 5 – Appariement de graphes

- $\Sigma_e = \{e_{ij} | i, j \in \{1, \dots, n\}\}$ un ensemble d'arcs,
- $\gamma_v : \Sigma_v \rightarrow \Pi_v$ un interpréteur de sommets assignant n_{av} attributs à un sommet,
- $\gamma_e : \Sigma_e \rightarrow \Pi_e$ un interpréteur d'arcs assignant n_{ae} attributs à un arc.

Π_v représente l'ensemble des couples (attribut, valeur) possibles pour les sommets et Π_e l'ensemble des couples (attribut, valeur) possibles pour les arcs. Dans notre cas $n_{av} = 3$, $n_{ae} = 2$ et tous les attributs prennent leurs valeurs dans $[0, 1]$. Un graphe d'attributs se définit, alors, comme une paire (V, A) où $V = (\Sigma_v, \gamma_v)$ est appelé ensemble des sommets attribués et $A = (\Sigma_e, \gamma_e)$ est appelé l'ensemble des arcs attribués.

4.2 Modélisation de la base d'apprentissage : graphe aléatoire

Le graphe aléatoire a pour fonction de représenter un modèle de classe.

Un graphe aléatoire se définit de la façon suivante ([SER 99], [KIM 01]) :

- $\Sigma_\omega = \{\omega_k | k = 1, \dots, n\}$ un ensemble de sommets,
- $\Sigma_\epsilon = \{\epsilon_{ij} | i, j \in \{1, \dots, n\}\}$ un ensemble d'arcs,
- $\gamma_\omega : \Sigma_\omega \rightarrow \Omega_\omega$ un interpréteur de sommets aléatoires,
- $\gamma_\epsilon : \Sigma_\epsilon \rightarrow \Omega_\epsilon$ un interpréteur d'arcs aléatoires,
- P : une loi de probabilité conjointe pour les sommets et les arcs aléatoires.

Ω_ω et Ω_ϵ représentant deux ensembles de variables aléatoires. Dans notre cas les ensembles sont constitués de, respectivement, trois et deux variables aléatoires à valeurs dans $[0, 1]$.

Un graphe aléatoire est donc un triplet (W, B, P) où $W = (\Sigma_\omega, \gamma_\omega)$ et $B = (\Sigma_\epsilon, \gamma_\epsilon)$.

On dit que le graphe d'attributs $G = (V, A)$ est une réalisation du graphe aléatoire $R = (W, B, P)$ s'il existe un monomorphisme $\mu : G \rightarrow R$ tel que si pour ω_i et ϵ_{kl} de R , $a_i = \gamma_v(\mu^{-1}(\omega_i))$ et $b_j = \gamma_e(\mu^{-1}(\epsilon_{kl}))$, alors la probabilité que G soit une réalisation de R par μ est donnée par :

$$\begin{aligned} P_R(G|\mu) &= P_R \left(\bigwedge_{i=1}^n (\alpha_i = a_i) \wedge \bigwedge_{j=1}^m (\beta_j = b_j) \right) \\ &= P(a_1, \dots, a_n, b_1, \dots, b_m) \end{aligned} \quad (11)$$

où les $\alpha_i = \gamma_\omega(\omega_i)$ sont les sommets aléatoires de R et les $\beta_j = \gamma_\epsilon(\epsilon_{kl})$ les arcs aléatoires de R .

En pratique, il est difficile d'estimer la probabilité conjointe P sans faire de supposition sur les indépendances probabilistiques entre arcs et sommets. En faisant les trois hypothèses suivantes :

1. les sommets aléatoires α_i sont mutuellement indépendants.
2. un arc aléatoire β_j est indépendant de tous les sommets exceptés les deux qu'il relie.
3. les distributions conditionnelles $\{\beta_j | \alpha_i = a_i, \forall i\}$ sont mutuellement indépendantes.

on peut montrer ([LU 91]) que l'équation 11 s'exprime par :

$$\begin{aligned} P_R(G|\mu) &= \prod_{i=1}^n P_R(\alpha_i = a_i) \\ &\times \prod_{j=1}^m P_R(\beta_j = b_j | \alpha_k = a_k, \alpha_l = a_l) \end{aligned} \quad (12)$$

où α_k et α_l sont les sommets aléatoires que relie β_j .

Chaque classe du thésaurus est modélisée par un graphe aléatoire dont la structure est définie par un modèle (choisi aujourd'hui par expertise). Les distributions des variables aléatoires du graphe sont évaluées à partir des attributs des graphes simples de la classe et d'une hypothèse de normalité.

4.3 Appariement et Classification

Dans la phase d'apprentissage des graphes aléatoires comme dans la phase de reconnaissance, il nous faut trouver un monomorphisme entre un graphe simple (à apprendre ou à reconnaître) et le modèle du graphe aléatoire. Pour simplifier le processus, nous avons opté pour un appariement de graphes simples en prenant pour attributs du graphe modèle la moyenne des réalisations des variables aléatoires. Le problème devient ainsi un problème classique d'appariement de graphes d'attributs pouvant avoir des structures différentes. Un bon algorithme pour résoudre ce problème est présenté par Chipman & Ranganath ([RAN 92]).

Il s'agit d'un algorithme d'appariement de graphes par relaxation flou qui permet d'apparier des graphes n'ayant pas la même structure. Nous construisons un graphe d'appariement où les sommets correspondent à un couple de sommet à apparier dans les graphes de départ et les arcs à un couple d'arcs. Les poids affectés sont des mesures de similitudes entre les sommets et entre les arcs apparés. On itère ces poids avec une règle de relaxation. Nous cherchons ensuite une clique de poids maximum qui respecte la structure des graphes de départ.

Un inconvénient de cet algorithme est de ne pouvoir apparier que des arcs existants réellement dans les graphes de départ. C'est pourquoi nous avons ajouté une possibilité de construction d'arcs virtuels, par addition simple d'arcs réels, pour pouvoir faire correspondre des arcs réels avec des arcs résultant de l'addition de plusieurs arcs adjacents.

Dans la figure 5, l'appariement simple donne la combinaison des noeuds suivante ($g1$ désigne le squelette de gauche et $g2$ celui de droite) : $\{(g1 : 0, g2 : 0), (g1 : 1, g2 : 2), (g1 : 2, g2 : 3), (g1 : 3, g2 : 6)\}$. Avec une possibilité d'appariement d'arcs virtuels sur une profondeur de 2 (c'est à dire en autorisant l'addition de deux arcs

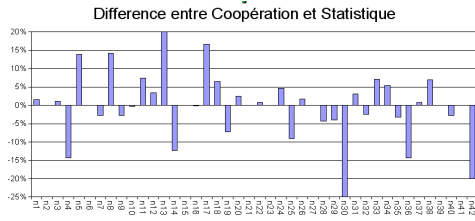


FIG. 6 – Différences de détections entre classification coopérative et statistique

adjacents), on obtient la combinaison $\{(g1 : 0, g2 : 0), (g1 : 1, g2 : 2), (g1 : 2, g2 : 3), (g1 : 3, g2 : 7)\}$ qui est plus réaliste.

Pour la reconnaissance, une fois l'appariement effectué, on calcule la probabilité à priori $P(X|\omega_i) = P(G|\mu)$ à l'aide de l'équation 12. La probabilité a posteriori $P(\omega_i|X)$ découle de la loi de Bayes (eq.9).

5 Mise en coopération

La mise en coopération des deux approches de classification se fait suivant le schéma de la figure 2.

La taille du thésaurus détermine le choix du classifieur statistique à utiliser. L'approche paramétrique sera privilégiée pour les thésaurus de grande taille (avec un minimum d'éléments par classe pour pouvoir effectuer une modélisation gaussienne correcte). Le classifieur statistique prend ensuite une décision de reconnaissance ou rejette la décision si la probabilité a posteriori de l'hypothèse la plus probable est trop faible [DUB01]. Si ω_i est l'hypothèse la plus probable en fin de classification statistique et si le critère de l'équation 13 est satisfait, alors la décision est rejetée .

$$P(\omega_i|X) \leq (1 - a) \quad (13)$$

où a est un seuil à déterminer par apprentissage ou par boucle d'expertise.

Dans les tests présentés ici, nous n'avons pas mis en place de rejet en distance car la classification était totalement supervisée.

En cas de rejet, une reconnaissance structurelle est lancée sur les N classes les plus probables. N est déterminé suivant le même seuil a que dans l'équation 13. Seuls sont conservés les classes les plus probables dont la somme des probabilités a posteriori est supérieure à $(1 - a)$. Cette restriction est motivée par le coût de calcul de la classification structurelle. Nous verrons dans par la suite qu'elle est peut-être un peu trop restrictive.

6 Discussion, résultats

Deux séries de tests ont été réalisés. Nous avons d'abord cherché à valider l'approche avec une base de caractères grecs dessinés à la main par quatre scripteurs différents. Elle possède 42 classes et 17641 caractères dont 15641 ont servi à l'apprentissage et 2000 à la reconnaissance. Les résultats obtenus ne sont pas discriminants entre approche statistique seule et approche coopérative (9,4% d'erreur pour la première contre 9,2% pour la deuxième). Cependant on peut

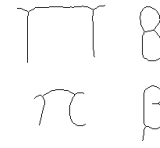


FIG. 7 – Squelettes confondus par l'approche coopérative (vérité en bas).

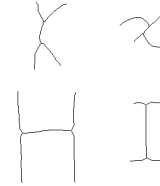


FIG. 8 – Squelettes confondus par l'approche statistique (vérité en bas).

constater une réelle différence entre les deux approches. La figure 6 donne les différences en pourcentage de bonnes détections entre l'approche par coopération et l'approche statistique ($\Delta(\%coop - \%stat)$). Si globalement les résultats sont les mêmes, on voit bien pour les classes $n13$ et $n30$ par exemple que le comportement des deux classifications est différent (meilleur pour la coopération sur $n13$, inversement pour $n30$).

La figure 7 donne deux exemples ($n30$ et $n4$) de squelettes de caractères confondus par l'approche coopérative que l'approche statistique simple avait réussi à distinguer (les caractères reconnus sont en haut, la vérité est en bas). Tandis que la figure 8 donne deux exemples ($n13$ et $n17$) que l'approche statistique avait confondu et que la coopération a reconnu.

Dans un deuxième temps nous avons réalisé une série de tests plus détaillés sur une base de caractères grecs résultant d'une segmentation et d'un apprentissage de 5 pages manuscrites d'époques et de scripteurs différents. Elle est composée de 1240 éléments répartis en 37 classes de façon très inhomogène. Ces tests ont été réalisés suivant une méthode de validation croisée. Le thésaurus est découpé en K groupes aléatoires. Chaque groupe est utilisé tour à tour pour la reconnaissance par rapport aux autres groupes qui servent de thésaurus. On obtient ainsi une erreur moyenne de détection et un écart type sur les groupes. La figure 9 donne un récapitulatif des résultats comparatifs de chaque méthode

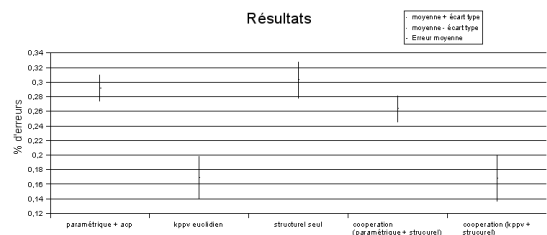


FIG. 9 – Résultats en % d'erreur sur la base de grec ancien

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
1	100%																						
2	80%	100%																					
3	10%	10%	100%																				
4	10%	10%	10%	100%																			
5	10%	10%	10%	10%	100%																		
6	10%	10%	10%	10%	10%	100%																	
7	10%	10%	10%	10%	10%	10%	100%																
8	10%	10%	10%	10%	10%	10%	10%	100%															
9	10%	10%	10%	10%	10%	10%	10%	10%	100%														
10	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%													
11	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%												
12	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%											
13	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%										
14	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%									
15	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%								
16	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%							
17	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%						
18	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%					
19	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%				
20	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%			
21	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%		
22	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%	

FIG. 10 – Matrice de confusion (vérité en ligne) pour la classification non paramétrique seule sur la base de grec ancien

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
1	100%																						
2	80%	100%																					
3	10%	10%	100%																				
4	10%	10%	10%	100%																			
5	10%	10%	10%	10%	100%																		
6	10%	10%	10%	10%	10%	100%																	
7	10%	10%	10%	10%	10%	10%	100%																
8	10%	10%	10%	10%	10%	10%	10%	100%															
9	10%	10%	10%	10%	10%	10%	10%	10%	100%														
10	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%													
11	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%												
12	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%											
13	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%										
14	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%									
15	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%								
16	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%							
17	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%						
18	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%					
19	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%				
20	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%			
21	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%		
22	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	100%	

FIG. 11 – Matrice de confusion (vérité en ligne) pour la classification coopérative sur la base de grec ancien

de classification avec et sans coopération. Les figures 10 et 11 donnent les matrices de confusion des approches non paramétrique seule et de l’approche coopérative sur les 22 premières classes.

Sur cette série de tests, comme sur la première, on constate que l’approche coopérative n’est pas redondante avec la classification non paramétrique puisqu’elle n’entame pas sa performance et que ses résultats sont sensiblement différents. Cependant, la classification structurale est encore trop peu performante pour être vraiment complémentaire dans la coopération. D’une part, l’approche par graphes aléatoires utilisant des modèles gaussiens pour les distributions des attributs est clairement limitative dans le cas des faibles thésaurus. D’autre part les attributs des graphes ne semblent pas encore suffisamment pertinents d’un point de vue topologique.

7 Conclusion et perspectives

Nous avons développé dans cet article une approche de classification de caractères anciens par mise en coopération d’approches structurales et statistiques. La particularité du projet COROC, traitant à la fois des bases de grande et petite taille, oblige à trouver des méthodes complémentaires et capables de s’organiser suivant certains critères.

Notre proposition actuelle de coopération a un sens mais n’est pas encore assez discriminante. Les raisons sont de deux ordres essentiellement. La première raison est que le classifieur structurel seul est insuffisamment performant. Cela est certainement dû aux hypothèses d’indépendances probabilistes et aux attributs choisis. C’est pourquoi nous voulons, d’une part, nous affranchir de cet aspect probabiliste par une approche structurale floue et, d’autre part, travailler sur des structures de graphes conservant mieux les informations topologiques. La seconde raison est que la coopération demande à être optimisée. L’idée est de mettre en place une structure de bouclage permettant au logiciel d’adapter ses paramètres internes (tels que les seuils de rejet ou les paramètres d’appariement) en fonction du thésaurus du moment.

8 Bibliographie

Références

[BLU 67] BLUM H., A Transformation for Extracting New Descriptions of Shape, *Models for the Perception of Speech and Visual Form*, MIT Press, 1967, pp. 362-380.

[CHO 03] CHONG C., RAVEENDRAN P., MUKUNDAN R., A comparative analysis of algorithms for fast computation of Zernike moments, *PR*, vol. 36, n° 3, 2003, pp. 731-742.

[DAM 04] DAMIAND G., BERTRAND Y., FIORIO C., Topological Model for 2D Image Representation : Definition and Optimal Extraction Algorithm, *Computer Vision and Image Understanding*, vol. 93, n° 2, 2004, pp. 111-154.

[DUB01] *Diagnostic, Intelligence Artificielle et Reconnaissance des Formes*, Chapitre Diagnostic et reconnaissance des formes, pp. 107-140, Hermès Science, 2001.

[GON 92] GONZALEZ R., WOODS R., *Digital Image Processing*, Addison-Wesley, 1992.

[KHO 90] KHOTANZAD A., HONG Y., Invariant Image Recognition by Zernike Moments, *PAMI*, vol. 12, n° 5, 1990, pp. 489-497.

[KIM 01] KIM H., KIM J., Hierarchical random graph representation of handwritten characters and its application to Hangul recognition, *PR*, vol. 34, n° 2, 2001, pp. 187-201.

[LAM 92] LAM L., LEE S.-W., SUEN C. Y., Thinning Methodologies - A Comprehensive Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, n° 9, 1992, pp. 869-885.

[LU 91] LU S., REN Y., SUEN C., Hierarchical attributed graph representation and recognition of handwritten chinese characters, *PR*, vol. 24, n° 7, 1991, pp. 617-632.

[RAN 92] RANGANATH H., CHIPMAN L., A Fuzzy Relaxation Algorithm for Matching Imperfectly Segmented Images to Models, *IEEE Southeastcon'92*, vol. 1, April 1992, pp. 128-136.

[SER 99] SERRATOSA F., ALQUÉZAR R., A. S., Function-Described Graphs for structural pattern recognition, rapport, 1999, Universitat Rovira i Virgili, Tarragona, Spain.

[TRI 96] TRIER O., JAIN A., TAXT T., Feature-Extraction Methods for Character-Recognition : A Survey, *PR*, vol. 29, n° 4, 1996, pp. 641-662.

[WON 87] WONG A., Structural Pattern Recognition : A Random Graph Approach, *PRTA*, vol. F30, 1987, pp. 323-345.

[ZHA 94] ZHANG Y. Y., WANG P. S. P., A New Parallel Thinning Methodology, *PRAI*, vol. 8, 1994, pp. 999-1011.

[ZHA 01] ZHANG D., LU G., A comparative study on shape retrieval using Fourier descriptors with different shape signatures, *ICIMADE*, 2001, pp. 1-9.

[ZHA 02] ZHANG D., LU G., A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval, *JVCIR*, vol. 14, n° 1, 2002, pp. 39-57.