Habilitation à Diriger des Recherches de l'Université

Pierre et Marie Curie (Paris VI)

ANALYSE D'IMAGES DE DOCUMENTS NON CONTRAINTS : DE LA STRUCTURATION A LA RECONNAISSANCE

Laurence LIKFORMAN-SULEM

Mémoire de synthèse

16 Juillet 2008

TABLE DES MATIERES

MEMOIRE

PRESENTATION GENERALE DES TRAVAUX

Cette section est consacrée à la description de mes travaux de recherche réalisés depuis l'obtention de ma thèse en 1989. Celle ci concernait l'authentification de caractères par système expert. L'ensemble des travaux rassemblés ici portent principalement sur trois thèmes de recherche dans le domaine de l'analyse de l'écrit et du document : (A) la structuration de documents manuscrits, anciens et modernes, et plus précisément l'extraction de la structure physique dans ces images de documents, (B) la recherche d'information dans les images de documents de type télécopie ou articles de revue qui correspond à une reconnaissance partielle de la structure logique de ces documents, (C) les modèles stochastiques appliqués à la reconnaissance des caractères et mots imprimés et manuscrits.

Le sujet (A) est orienté principalement sur l'extraction de lignes de textes et d'éléments graphiques (longs traits,…) dans les images. Quinze articles ont été publiés dans ce domaine dont 6 dans des revues à comité de lecture, 1 chapitre d'ouvrage et 8 dans des comptes rendus de congrès. Ce sujet de recherche a fait l'objet d'une coopération avec des partenaires académiques (Irisa, IRHT, Université de Reims) dans le cadre du projet Philectre (1995-1998). Cet axe de recherche inclut des travaux sur la structuration perceptive des alignements, le suivi de traits par filtrage de Kalman dans le cadre de la réalisation d'un poste de travail pour les chercheurs en sciences humaines. Les critères perceptifs utilisés sont relatifs à la théorie de la Gestalt et nous les avons introduits dans des algorithmes ascendants de regroupement de composantes (dont la transformée de Hough).

Le sujet (B) est orienté sur l'extraction de noms propres dans des images de documents non contraints tels que les télécopies et articles de revues d'origines variées. Cette recherche fait appel à la fois à des traitements d'image et des traitements linguistiques. Un article de revue est issu de ce travail ainsi que cinq articles dans des comptes rendus de congrès. Cette recherche a fait l'objet d'un projet européen de recherche du programme Eurekâ (Majordome, no 2340, 2000-2003) dont j'étais coordinatrice pour la partie française, et d'une coopération avec des partenaires académiques et industriels.
Cette recherche présente l'extraction de la structure physique dans les images de télécopies ainsi que les modes de combinaison des caractéristiques spatiales et textuelles pour extraire la structure logique recherchée. Plusieurs types de combinaison ont été réalisés, notamment par réseau neuronal.

Le sujet (C) est une recherche amorcée dans les années 2000. Elle concerne la reconnaissance de caractères et de mots par modèles stochastiques : modèles de Markov cachés et Réseaux Bayésiens Dynamiques. Une partie de cette recherche a été réalisée dans le cadre du projet Eurekâ Majordome. J'ai encadré deux thèses de Doctorat qui ont été soutenues en 2004 et 2007 respectivement. La thèse de R. El-Hajj (2007) a fait l'objet d'une coopération avec l'Université de Balamand (Prof. C. Mokbel). Ces recherches ont donné lieu à la publication de 3 articles de revues et 9 articles de conférences.

A. STRUCTURATION DE DOCUMENTS MANUSCRITS

A.1 Introduction

L'extraction de la structure physique (colonnes, lignes ou mots) est une des premières étapes d'un système de reconnaissance d'écriture, imprimée ou manuscrite. Ces systèmes de reconnaissance sont opérationnels pour la reconnaissance des blocs adresses et des chèques bancaires et bientôt pour les courriers manuscrits entrant. Les systèmes d'analyse de documents, en particulier ceux qui ont pour objectif d'extraire la structure logique (titre, corps de texte, date,...), reposent aussi sur la structuration préalable en blocs physiques. L'extraction de la structure physique peut être aussi la première étape d'un système d'authentification (reconnaître un scripteur, un style), ou pour l'établissement de liens texte/image dans les manuscrits (cf. projet Philectre) et à plus long terme pour l'interrogation de bases de données d'images de documents à partir du contenu. Si la structuration physique est souvent considérée comme un prétraitement pour une tâche de plus haut niveau, la qualité de ce traitement conditionne les performances globales du système réalisant cette tâche.

Dans les documents manuscrits, la structure physique recherchée se limite à une hiérarchie simple : les lignes de texte, parfois les mots qui la composent. Cependant dans les documents calligraphiés, comme par exemple les manuscrits médiévaux, la structure des colonnes ou des blocs en marge est aussi recherchée. La séparation du texte des autres composantes du document (images, figures, lettrines, grands traits) est supposée effectuée au préalable.

Les stratégies d'extraction de lignes pour l'analyse des documents imprimés se divisent en deux catégories : descendantes et ascendantes. Les méthodes descendantes divisent le document pour aboutir aux lignes : ce sont les méthodes basées sur les profils de projections [Nagy et Seth, 1984]. Les méthodes ascendantes partent des pixels ou des composantes connexes de l'image et les fusionnent pour former des lignes. On peut citer le lissage en séquence de plage [Wong et Casey, 1982], la transformation de Hough [Fletcher et Kasturi, 1988], le regroupement d'unités (composantes connexes, minima locaux,...) par proximité [Meynieux et al., 1989; Feldbach, 2001]. Une approche mixte [Viard-Gaudin et Barba, 1992] consiste à localiser les lignes à partir d'une représentation multi-résolution du texte. Ces méthodes ont été appliquées à des textes manuscrits dont la mise en page n'est pas trop éloignée de celle du document imprimé : lignes proches de l'horizontale, bien espacées [Paquet et al., 1989; Downton et Leedham, 1990; Cohen et al., 1991; Shapiro et al., 1993; Govindaraju et al., 1994; Zahour et al., 2004]. D'autres méthodes comme [Oztop et al. 1999] utilisent les forces d'attraction-répulsion issues des pixels pour trouver une ligne de base passant dans la zone centrale de l'écriture. Les méthodes stochastiques basées sur les modèles de Markov cachés [Tseng et Lee, 1999] divisent l'image en cellules qui correspondent à des états : les chemins optimaux encadrant les lignes, traversent l'image en largeur en évitant les pixels d'écriture et conservant une direction la plus horizontale possible. Les champs de Markov [Nicolas et al., 2005] ont été récemment appliqués pour

étiqueter les pixels en tant qu'interlignes, mots, rature et fond : ceci constitue une première étape pour l'extraction des lignes dans un contexte très bruité.

Les méthodes énoncées ci-dessus diffèrent également dans la manière de représenter les lignes (chemins, clusters, lignes de base, unités chaînées) et sur les unités qu'elles traitent (pixels, composantes connexes, minima, ...). Un état de l'art sur la représentation et l'extraction des lignes est publié dans [Likforman-Sulem et al., 2007].

Nous avons étudié la structuration de documents manuscrits non contraints (brouillons, manuscrits d'auteur, enveloppes postales). Ces manuscrits sont caractérisés par des lignes de longueurs différentes, plus ou moins espacées et fluctuantes. Les difficultés majeures sont l'imbrication des lignes, le chevauchement de composantes (composantes appartenant à plusieurs lignes de texte du fait de la présence de hampes et de jambages) et la fragmentation des caractères (due à la binarisation ou à l'inhomogénéité de l'encre). Nous avons proposé deux approches à stratégie ascendante pour la structuration en lignes dont la perception est indépendante du contenu linguistique. La première est décrite dans [Likforman-Sulem et Faure 1994b 1995] et consiste en un groupement perceptif à partir de composantes ancrages. La deuxième, décrite dans [Likforman-Sulem et al 1995] [Likforman-Sulem et Faure 1996] introduit des critères perceptifs dans la transformée de Hough. Ces deux méthodes utilisent une stratégie d'hypothèse-validation. Ces méthodes ont été appliquées au sein du projet Philectre [Likforman-Sulem et al. 1997][Robert et al. 1997] pour réaliser le couplage texte/image dans un mode collaboratif. La structuration en lignes de texte présuppose que l'image soit nettoyée des éléments graphiques tels que les grands traits ou les ratures. Ceci est réalisé par une technique de suivi par filtrage de Kalman qui a été appliquée aux manuscrits d'auteur [Likforman-Sulem 1998].

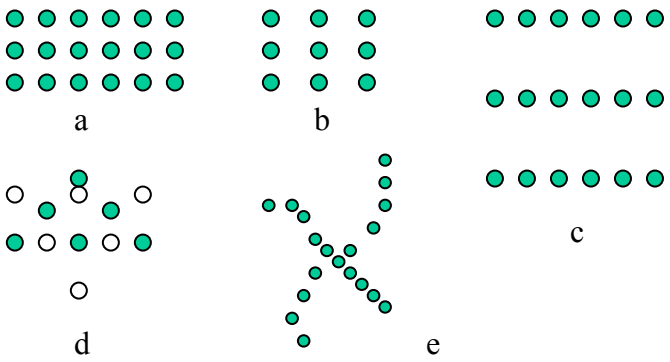


Figure 1. *Principes d'organisation perceptive : les éléments de a) sont groupés indifféremment en lignes ou en colonnes. Par contre les éléments de b) sont groupés suivant les colonnes, ceux de c) en lignes par principe de proximité. Le principe de similarité permet de voir 2 triangles dans d). On perçoit les 2 branches de la forme e) par principe de continuité de direction.*

A.2 Extraction d'alignements et organisation perceptive

Le processus de structuration en lignes que nous avons développé [Likforman-Sulem et Faure, 1995] s'appuie sur les mécanismes perceptifs qui permettent à l'être humain de voir des lignes de texte, notamment à distance, indépendamment de la lecture proprement dite. Il utilise certains principes de la physiologie de la vision et des lois d'organisation perceptive de la théorie de la Forme (Gestalt) (Fig. 1). A partir d'un ensemble d'excitations élémentaires produites par des unités d'écriture orientées dans une direction (points d'ancrage), les composantes de l'image sont organisées suivant des structures linéaires pouvant présenter de légères courbures ou des fluctuations autour d'une direction principale.

A.2.1 Description générale

La stratégie itérative de construction des alignements est décrite en Figure 2. La première étape consiste à appliquer des masques sur les composantes connexes de l'image pour sélectionner celles qui ont une direction fiable. A partir des points d'ancrage, les alignements sont construits en regroupant les composantes connexes voisines dans un voisinage dépendant d'un seuil de proximité S. Les alignements obtenus constituent des hypothèses qui sont évaluées du point de vue de leur qualité et des conflits possibles. Le seuil de proximité S est incrémenté à chaque itération. Ainsi les composantes les plus proches (donc les plus fiables) sont groupées en début de formation des alignements. Et les composantes les plus éloignées ont groupées plus tardivement quand on dispose de plus de connaissances sur les alignements.

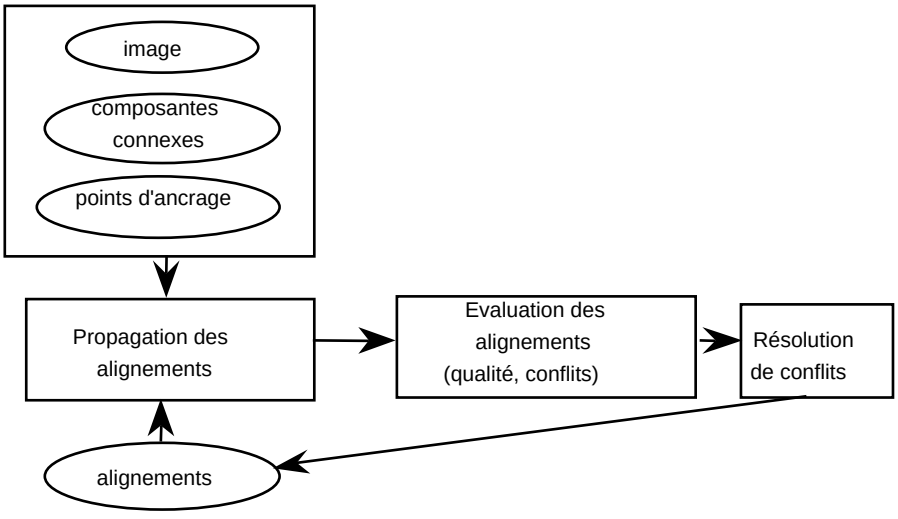


Figure 2. *Stratégie de construction des alignements*

A.2.2 Détection des points d'ancrage

Les points d'ancrage sont des composantes connexes ayant une direction privilégiée. Pour les repérer, nous nous sommes inspirés du système visuel humain, notamment des propriétés des cellules simples de l'aire 17 du cortex visuel [Busert et Imbert, 1987]. En effet, la configuration des champs récepteurs de certaines cellules les rendent sensibles à l'orientation des segments de lignes. Par analogie avec les champs récepteurs, nous définissons 4 masques correspondants à une discrétisation de l'espace en 4 directions (0°, 45°, 90°, 135°). Chaque masque m est carré et contient une zone d'excitation z^+_m sensible à la densité d'écriture dans la zone (Fig. 3).

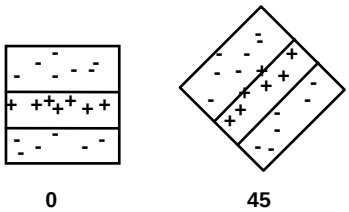


Figure 3. *Masques de détection d'orientation suivant 2 directions (0°, 45°).*
La zone d'excitation z^+_m de chaque masque est signalée par les signes +

Soit une composante connexe i, et S_i son rectangle englobant. La réponse d'une composante connexe i au masque m est :

$$R_m(i) = \frac{P(i, z_m^+)}{P(i, S_i)}$$ avec P(i, S) le nombre de pixels de la composante i dans la zone image S

Les points d'ancrage sont les composantes ayant une réponse suffisante (>= 80%) et qui correspondent à des mots ou parties de mots de forme allongée et compacte.

A.2.3 Groupement

Les points d'ancrage sont à l'origine des alignements. A un point d'ancrage, on associe les composantes voisines situées de chaque côté pour former un groupement.

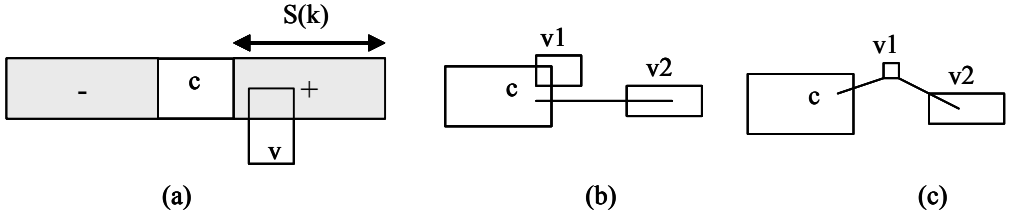


Figure 4. *Principes perceptif de groupement entre deux composantes : (a) proximité, (b) continuité de direction (c) similarité.*

La zone à droite du point d'ancrage est appelée le 'voisinage positif', la zone à sa gauche le 'voisinage négatif'. Le groupement est effectué selon trois principes perceptifs issus de la théorie de la Forme (Gestalt) et qui permettent à l'œil humain de percevoir des groupements d'éléments. Ces principes sont la proximité, la continuité de direction et la similarité, rappelés dans [Guillaume, 1979].

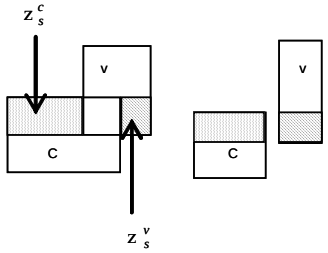


Figure 5. *Zones de superposition entre deux composantes c et v*

Par principe de proximité, une composante connexe v sera groupée à la composante c (Fig. 4-a) si v est suffisamment proche de c. La longueur S(k) du voisinage dépend de l'itération k et la hauteur du voisinage dépend de celle de c et de la direction de l'alignement (horizontale, verticale ou orientée en diagonale).
Le principe de continuité de direction consiste à vérifier que les composantes à grouper sont bien en vis à vis (Fig. 4b). On définit d'abord les zones de superposition z^i_S, i= c ou v, par les parties en vis à vis des boîtes englobant c et v, à l'exclusion de la zone d'intersection (Fig. 5).
Puis on définit les densités relatives d'écriture Ri dans une zone de superposition par :

$$R_i = \frac{P(i, z_S^i)}{P(i, S_i)} \qquad i = v,c$$

$P(i,z_s)$ et $P(i,S_i)$ étant respectivement le nombre de pixels de la composante i dans la zone zs et la zone occupée par son rectangle englobant. Si aucun des deux rapports (pour chacune des composantes v et c), n'est supérieur à un seuil, v n'est pas accepté comme voisin de c, et un autre voisin dans le voisinage de c est recherché. Il faut en effet que suffisamment de points d'écriture soient dans la zone de superposition.

Le principe de similarité consiste à rechercher d'autres voisins si les composantes groupées ne sont pas similaires. Les accents par exemple sont de taille beaucoup plus petite que les mots. Dans la figure 4-c, v2 est un autre voisin de c. v2 est groupé à c (et v1) car on ne peut trouver de voisin à v1.
Une fois, les premières composantes groupées autour du point d'ancrage, le groupement s'effectue ensuite de part et d'autre des composantes déjà groupées.

A.2.4 Analyses locale et globale des conflits

Lors du groupement des composantes, peut apparaitre une composante appartenant déjà à un autre alignement. La composante est dite ambiguë et il y a conflit d'alignements. Les différents types de conflit sont le croisement et les embranchements en Y ou en U. L'analyse globale est utilisée pour les conflits de type croisement. L'analyse locale est utilisée pour les cas d'embranchements.

En ce qui concerne les croisements, le conflit se résout en comparant le facteur de qualité (global) P des alignements. P est défini de la façon suivante:

$$P = P_1 * P_2 \quad \text{avec } P_1 = \frac{1}{1+\frac{1}{NC\text{-}2+\frac{MD}{NC}}} \quad \text{si } NC > 1 \;,\; P_1 = 0 \text{ sinon} \qquad \text{et } P_2 = \frac{1}{1+\frac{DD}{NC}}$$

avec *NC* le nombre de composantes de l'alignement, *MD* le nombre de points d'ancrage de l'alignement ayant la même orientation que celle de l'alignement, et *DD* le nombre de points d'ancrage ayant une orientation différente. Le facteur de qualité P est d'autant plus élevé que l'alignement contient des composantes ancrages ayant l'orientation de l'alignement.
La composante ambiguë est affectée à l'alignement de meilleure qualité et celui ci est conservé, tandis que le deuxième alignement est éliminé.

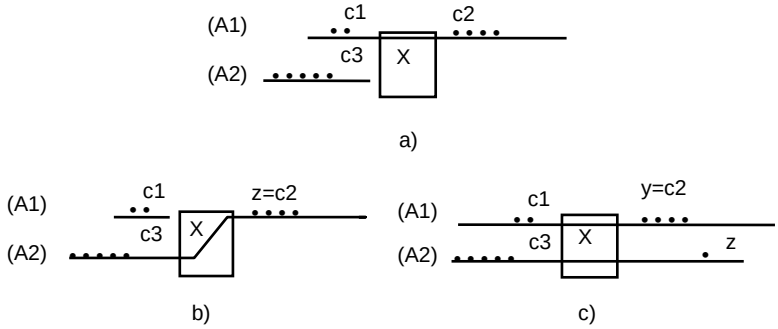


Figure 6. *(a) Conflit de type embranchement en Y. (b) La composante ambiguë X est affectée à un des alignements (c) X est conservée dans les 2 alignements (composante de chevauchement).*

En ce qui concerne les embranchements, l'analyse locale consiste à chercher le meilleur chemin pour les alignements concernés en formulant des hypothèses sur les voisins au delà de la composante conflictuelle. Soit l'embranchement Y de la figure 6-a. Dans la figure 6-b, c1 n'a pas de voisin d'ordre supérieur, contrairement à c3, X est alors affectée de manière non ambiguë à l'alignement A2. Dans la figure 6-c, les voisins d'ordre supérieur des composantes c1 de l'alignement A1 et c3 de l'alignement A2 sont respectivement y et z. Dans ce cas, X restera attribuée aux 2 alignements A1 et A2 car il s'agit d'une composante de chevauchement.

Cette composante de chevauchement est découpée en deux parties. Le point de coupure est déterminé à partir du profil vertical de projection horizontale des pixels de la composante et de l'analyse locale de la configuration des deux alignements (stage de DEA de A. Wang [Wang 1994]).

A.2.5 Applications

Cette méthode a été appliquée à la structuration d'enveloppes postales, de brouillons [Likforman-Sulem et Faure 1995] et de documents anciens dans le projet Philectre [Likforman-Sulem 2003](cf. Section A.5). Nous avons utilisé une base restreinte de 35 blocs adresses et de 26 brouillons. Pour l'évaluation de la tâche de structuration, nous considérons qu'une ligne est détectée quand toutes les composantes connexes de la ligne d'écriture ont été affectées à un seul alignement. La ligne peut cependant être incomplète : composantes restant non affectées ou signes diacritiques affectés à un alignement voisin. Nous considérons qu'une ligne est non détectée quand ses composantes ne sont affectées à aucun alignement. Une ligne est détectée de manière erronée quand un ensemble des composantes de cette ligne, appartient à un alignement qui correspond à plusieurs lignes d'écriture. Une ligne est détectée de manière fragmentée quand à cette ligne correspondent plusieurs alignements qui mis bout à bout reconstituent la ligne d'écriture.

Les résultats sont donnés en Table 1. Sur les enveloppes, quelques lignes courtes n'ont pas été détectées ou sont restées incomplètes. En effet, les alignements de trop faible qualité sont détruits. Dans les brouillons, les lignes ont tendance a être détectées en plusieurs alignements, qui mis bout à bout reconstituent des lignes véritables. Ceci est dû à la présence de nombreux points diacritiques. Le taux de lignes trouvées fragmentées est de 7.5 %. D'autres lignes n'ont pas été détectées, ou ont été détectées partiellement car des parties d'alignements sont parfois détruits par erreur à une étape du traitement. Les erreurs sont souvent le fait de lignes en biais qui ont une direction intermédiaire à celle des directions choisies pour déterminer les orientations des points d'ancrage.

Table 1. Résultats de la structuration en lignes

	détection	non détection	erreur
enveloppes	96,3 %	0.9 %	2,8%
brouillons	82 %	7,5 %	5,7 %

A.2.6 Conclusion

La méthode d'extraction d'alignements que nous avons proposée s'appuie sur un groupement ascendant perceptif des composantes connexes de l'image. Lors du groupement des conflits d'alignements apparaissent du fait de la proximité des lignes de texte et de la présence de composantes de chevauchement. Des règles de résolution de conflits sont appliquées. Lorsque les alignements sont de même direction, le choix d'un des alignements pour une composante

connexe, est celui qui assure la meilleure continuité, au sens de l'organisation perceptive. Cette continuité est vérifiée soit à partir de la recherche de prolongements lors de l'analyse locale des embranchements, soit à partir de la position relative de la composante dans les alignements lors de l'analyse globale. Lorsque les alignements sont de directions différentes, le choix d'un des alignements pour la composante au carrefour des deux, est basé sur la qualité perceptive des alignements.

La méthode par groupement perceptif a été reprise par [Koch et al. 2005] pour l'extraction de champs numériques dans les courriers et par [Noi Bai et al 2008] pour la recherche de lignes incurvées.

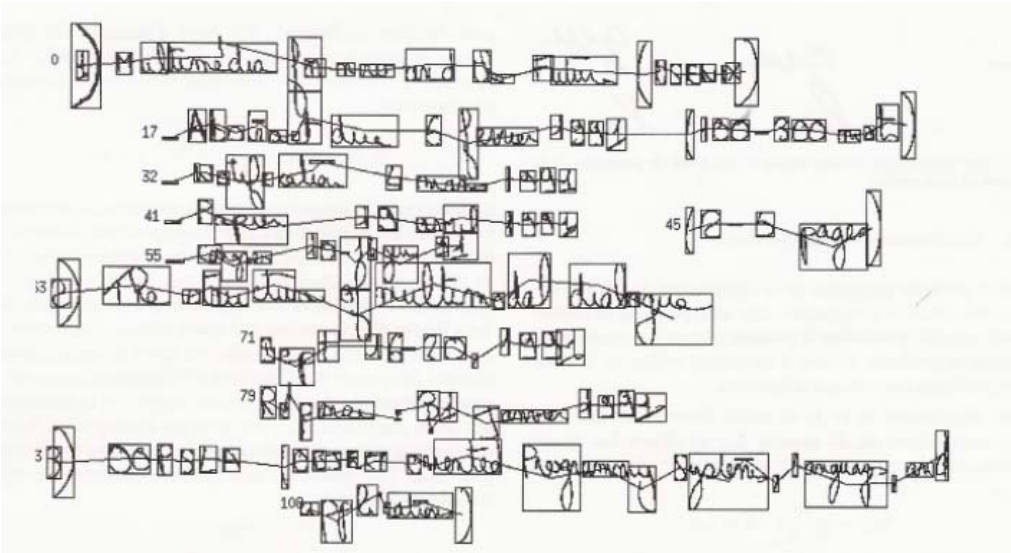


Figure 7. *Structuration en ligne d'un brouillon*

A.3 Extraction d'alignements par transformée de Hough

La transformée de Hough [Hough 1962] est une méthode très utilisée pour la détection d'alignements en traitement des images et pour le traitement des documents. Dans ce dernier domaine, elle a été appliquée à la détermination de l'angle d'inclinaison [Hinds et al., 1990], à l'extraction de chaînes de caractères dans les documents techniques [Fletcher et Kasturi, 1988], à l'extraction des lignes dans les textes imprimés [Srihari et Govindaraju, 1989] ou les textes manuscrits simplifiés [Shapiro et al., 1993]. La transformée de Hough est aussi utilisée pour déterminer l'orientation de traits dans les mots ou caractères manuscrits [Vincent et al., 1992] [Ruiz-Pinales, 2001] [Touj et al., 2002].

Cette deuxième méthode de structuration en lignes a été développée lors de la thèse professionnelle d'A. Hanimyan (Hanimyan 1995) et publiée dans [Likforman-Sulem et al. 1995]. Nous nous sommes inspirés de la méthode de [Fletcher et Kasturi, 1998] qui extrait les chaînes de caractères imprimés dans les documents graphiques, pour trouver une méthode adaptée aux documents manuscrits complexes (brouillons, manuscrits d'auteurs). La méthode utilise une stratégie d'hypothèse-validation (Fig. 8). Les hypothèses d'alignements sont créées dans le *domaine de Hough*. Les hypothèses sont ensuite validées ou invalidées dans le *domaine de l'Image* sur des critères perceptifs.

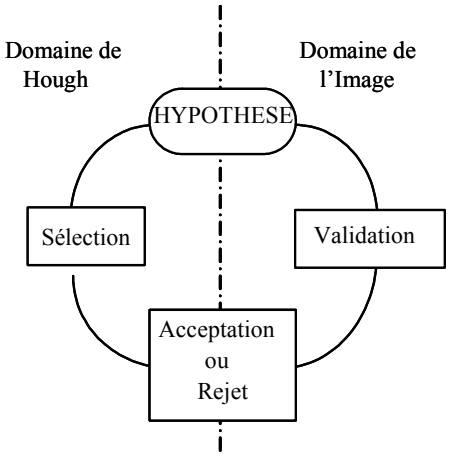


Figure 8. *Stratégie de Sélection - Validation d'hypothèses.*

A.3.1 Sélection dans le domaine de Hough

Les unités de base sont les centres de gravité des composantes connexes de l'image. La transformation de Hough est appliquée sur ces unités et à chaque unité correspond à plusieurs cellules (ρ_i, θ_i), dans l'espace quantifié de Hough (ρ, θ), où ρ_i, et θ_i sont les coordonnées polaires des droites passant par cette unité dans l'espace cartésien (Fig. 9). Quand des unités sont alignées dans l'espace cartésien sur une droite, la cellule correspondant à cette droite dans l'espace de Hough a une valeur élevée.

A chaque itération, on repère la cellule (ρ_0, θ_0) de valeur la plus élevée. Elle correspond à une colinéarité exacte des composantes connexes. Cependant, les unités des lignes de texte ne sont pas exactement colinéaires, et les lignes fluctuent.
Pour tolérer ces variations, nous définissons une structure d'alignement par l'ensemble des unités de la cellule (ρ_0, θ_0) et celles des cellules voisines (ρ_0-i, θ_0) (ρ_0+i, θ_0), i=[1, f_{clus}]. Ceci revient à grouper dans un alignement les unités sur la droite (ρ_0, θ_0) et celles sur des droites parallèles et très proches. Le voisinage de cellules autour de (ρ_0, θ_0) dépend du facteur de groupement f_{clus} . Celui ci est défini en fonction de la hauteur moyenne des composantes connexes dans un voisinage très proche de la cellule (ρ_0, θ_0). Si H_{moy} est la taille moyenne des composantes appartenant aux cellules (ρ_0-δ, θ_0) (ρ_0+ δ , θ_0), avec δ=5, alors :

$f_{clus} = H_{moy}/R$ si $85° \leq \theta_0 \leq 95°$
$f_{clus} = 0.5*H_{moy}/R$ si $0° \leq \theta_0 < 85°$ ou si $95° \leq \theta_0 < 180°$

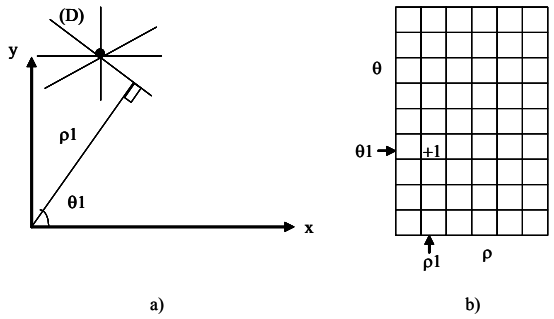


Figure 9. (a) *Unité dans l'espace cartésien par laquelle passe plusieurs droites dont la droite D repérée en cordonnées polaires par (ρ1,θ1).(b) Cette droite incrémente la cellule correspondante dans l'espace de Hough*

Cette tolérance n'est pas toujours suffisante. Du fait des fluctuations possibles dans les positions des unités, la première hypothèse sélectionnée n'est pas toujours optimale. Une deuxième hypothèse (ρ_1, θ_1) est recherchée dans un voisinage rectangulaire de la cellule (ρ_0, θ_0). Ce voisinage dans l'espace de Hough est défini par x0, x1, z0, z1 (Fig. 10) avec :
$x0=\rho_0 - f_{clus}$, ; $z0=\theta_0-3$
$x1=\rho_0 + f_{clus}$, ; $z1=\theta_0+3$
(ρ_1, θ_1) est la cellule de valeur la plus élevée dans ce voisinage après (ρ_0, θ_0) et correspond à une rotation de la droite (ρ_0, θ_0) d'au maximum de 3 degrés (Fig. 10).
Soit n0 (resp. n1) est le nombre de composantes dans la cellule (ρ_0, θ_0) (resp. (ρ_1, θ_1)). Si les ($n1/n0 > 0.65$) et la structure d'alignement (ρ_1, θ_1) comprend plus de composantes que la structure (ρ_0, θ_0), alors la structure d'alignement (ρ_1, θ_1) est choisie à la place de (ρ_0, θ_0) comme meilleure hypothèse issue de l'espace de Hough.

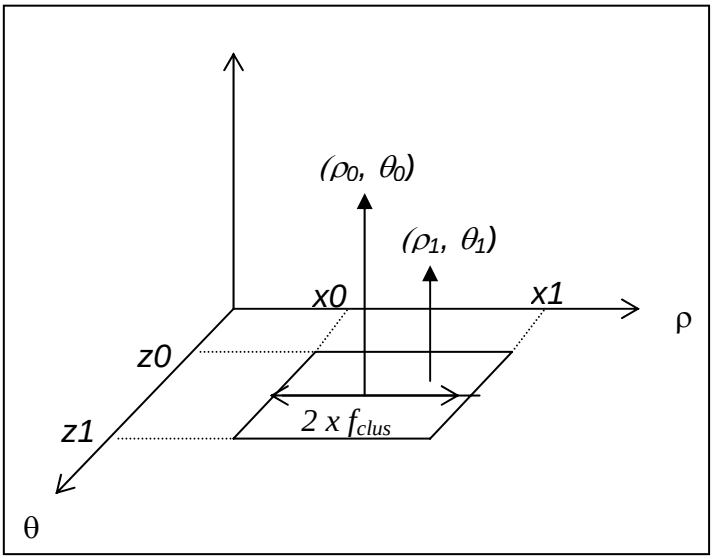


Figure 10. *Hypothèse d'alignement (ρ_1, θ_1) dans le voisinage de l'hypothèse initiale (ρ_0, θ_0) (espace de Hough)*

A.3.2 Validation dans le Domaine de l'Image

L'étape précédente de sélection émet la meilleure hypothèse d'alignement possible pour une itération donnée. L'alignement (ρ,θ) contient des composantes colinéaires dans l'image, mais elles correspondent ou pas à des lignes de texte. En particulier, si un texte disposé horizontalement est plus haut que large, les alignements les plus forts (i.e. les plus longs) croiseront les lignes d'écriture. Il est donc nécessaire de valider les hypothèses d'alignement correspondant à des lignes de texte et d'invalider les autres.

Nous utilisons comme précédemment (Section A.2) un critère perceptif pour valider les alignements. Les alignements perçus sont ceux qui correspondent au critère de proximité. Ce critère est transposé ici de la manière suivante. Les composantes sont d'abord ordonnées le long de (ρ, θ). Puis les distances entre composantes voisines sont évaluées (plusieurs types de distance peuvent être utilisées : distance bord à bord, distance bord à point (centre de gravité projeté), ou encore distance point à point). Ces composantes voisines sont dites *internes*. Puis pour chaque composante, ses composantes voisines hors alignement sont recherchées. Ce sont les composantes voisines *externes*. La distance utilisée est la moyenne des deux espacements entre la composante considérée et la précédente dans l'alignement, et la composante et sa suivante. Si globalement, le nombre de composantes voisines externes est supérieur à celui des composantes internes, alors l'alignement est invalidé, sinon il est validé. L'invalidation d'un alignement implique sa disparition de l'espace de Hough et ses composantes sont susceptibles de participer à d'autres alignements. S'il est validé, toutes ses composantes sont retirées de l'espace de Hough de manière à ce qu'elles ne participent plus à aucun autre alignement.

Nous avons également ajouté dans [Likforman-Sulem et Faure 1996] un mécanisme de détection de rupture dans la dimension des espacements entre composantes à l'intérieur d'un alignement, pour traiter le cas où les documents disposés sur deux colonnes ou sur deux blocs de directions différentes.

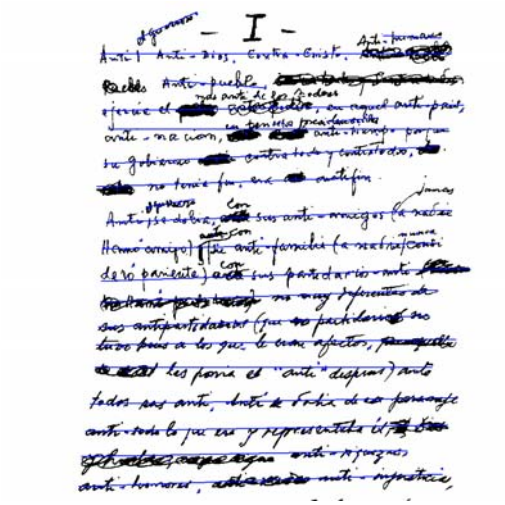


Figure 11. *Extraction des lignes sur un manuscrit d'auteur (Asturias)*

A.3.3 Résultats

Des tests ont été effectués sur une trentaine de documents manuscrits (Fig. 11) dont certains sont disposés sur plusieurs blocs de même direction ou de directions différentes. Les résultats sont les suivants (Table 3).

Table 3. Résultats de la structuration en lignes

Détection	Détection partielle	Non détection	Erreurs
65 %	22 %	9 %	4 %

Une ligne de texte est dite détectée quand toutes ses composantes font partie d'un seul alignement ou de deux alignements de directions proches. Ce dernier cas provient du fait qu'un écart a été trouvé dans l'alignement de départ, correspondant à un écart inter-mot. L'alignement est alors découpé en plusieurs parties. Une fusion sera alors nécessaire, à l'aide par exemple d'une interface appropriée.

La détection d'une ligne est dite partielle quand une ou plusieurs composantes manquent dans l'alignement trouvé. Là encore, une procédure soit automatique, soit réalisée au travers d'une interface peut être engagée pour affecter les composantes isolées à un alignement existant.

Les alignements courts ne sont parfois pas détectés. La notion de court ou long dépend du nombre de composantes et non de l'encombrement spatial effectif. Ce problème a été résolu par

[Louloudis et al. 2006] qui ont repris cette méthode mais découpent les longues composantes en y créant artificiellement des coupures.

A.4 Discussion

Les deux méthodes d'extraction de lignes présentées précédemment sont itératives et utilisent des critères perceptifs. L'avantage d'une stratégie itérative est de former les alignements les plus fiables en début de processus. Par exemple les composantes qui composent ces alignements sont écartées de la suite du processus de Hough, ce qui restreint les choix pour former les alignements restants plus délicats à trouver (par exemple les annotations). Pour la méthode par groupement, les débuts d'alignements fiables permettent de mieux régler les conflits quand ils surviennent car on a une connaissance sur la direction, la position des alignements en conflit.

Une comparaison qualitative dans entre ces deux méthodes publiée dans [Likforman-Sulem, 2003] a été réalisée lors du projet Philectre (cf. Section A.4) sur des manuscrits anciens (Manuscrits de Flaubert, Lettres de Rémission du XVIème siècle [Fekete et Dufournaud, 1999]). La figure 12 (gauche) donne un exemple d'extraction de lignes par groupement. Les lignes extraites sont repérées par un chaînage de composantes. Certaines grandes composantes ont été éliminées avant l'extraction des lignes mais restituées sur la figure à des fins de visualisation. Sur la figure 12 (droite), les lignes sont obtenues par transformation de Hough.
Nous avons constaté que la méthode par groupement est plus rapide car les voisinages de chaque composante sont restreints à une petite zone de l'image. Par contre, l'espace de Hough est plus lent à construire et il est balayé entièrement à chaque itération. De plus, la résolution des angles est très fine (1°), plus fine que celle de la méthode par groupement (qui quantifie l'espace des angles en quatre secteurs). Ceci permet une meilleure adaptation aux orientations individuelles des lignes.
La méthode basée sur Hough accorde aussi plus d'importance au critère global qu'au critère local. En effet, le premier critère de sélection est basé sur le nombre de composantes de l'alignement. C'est le contraire pour la méthode par groupement qui favorise les critères locaux (disposition des voisins autour des composantes en conflit). Une erreur de décision pour les composantes en conflit à un stade précoce de formation des alignements, peut entraîner des erreurs en cascade.

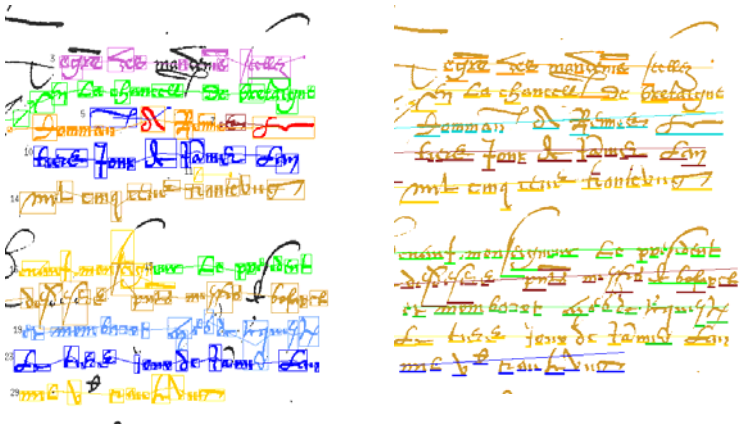


Figure 12. *Résultat de l'extraction automatique de lignes sur un extrait de Lettres de Rémission (XVIème siècle). Groupement perceptif (gauche). Méthode utilisant la transformée de Hough (droite)*

A.5 Le Projet Philectre : édition et consultation de manuscrits

A.5.1 Introduction

Les méthodes de structuration développées précédemment ont été appliquées sur des manuscrits anciens lors du projet Philectre : PHILologie élECTRoniquE [Cerquiglini et Lebrave, 1994]. L'interface conçue pour ce projet et pour les chercheurs en sciences littéraires est présenté dans [Robert et al., 1997] : elle a pour vocation l'édition électronique et la consultation des manuscrits. Le sujet d'étude est les dossiers génétiques [Flaubert, Ponge] avec pour objectif d'établir leur transcription diplomatique (transcription qui respecte la mise en page du manuscrit). Les dossiers génétiques comprennent l'ensemble des manuscrits (brouillons, copies) établis avant l'édition imprimée. D'autres objectifs du projet sont le développement d'outils pour l'édition hypertextuelle, le classement des manuscrits pour reconstituer l'ordre génétique, la mise en correspondance des variantes du texte.

Figure 14. *Fenêtre principale de l'application.(a) La partie supérieure contient les représentations iconiques des documents. (b) La partie médiane contient l'image d'un manuscrit. (c) Trois régions (en bleu) ont été sélectionnées par trois types de marqueurs. (d) La partie inférieure contient la transcription diplomatique (e) du manuscrit.*

A.5.2 Sélection de zones

L'interface Philectre est présentée en Figure 14. Deux approches permettent de sélectionner les zones de l'image (mots, lignes) du manuscrit en vue de la transcription. La première correspond à l'utilisation d'outils interactifs de sélection de zones : les marqueurs. Ces marqueurs interactifs sont au nombre de quatre: la ligne, le rectangle, le stabilo, le lasso. Ils permettent de déterminer la partie du texte à transcrire en agissant directement sur l'image du manuscrit. Par leur intermédiaire, l'utilisateur dépose une trace graphique sur ou autour du texte à sélectionner. La création de transcriptions, sur un mode uniquement interactif, peut rapidement s'avérer lourde et fastidieuse dès lors que le document à traiter atteint une certaine taille. Cette remarque nous amène donc à une seconde approche de sélection, selon un mode collaboratif semi-automatique. Le système propose à l'utilisateur un ensemble de solutions (les lignes de texte trouvées par l'extraction automatique des lignes) que l'utilisateur valide ou corrige interactivement. La fin du projet Philectre n'a pas permis d'intégrer le code développé pour l'extraction des lignes dans l'interface développée par L. Robert [Robert et al. 1997]. Les résultats de l'extraction des lignes par la méthode de groupement (cf. Section A.2) et pour un échantillon de documents, ont cependant été intégrés à l'éditeur structuré *Thot* développé à l'IRISA [Gusnard et al 1999].

A.5.3 Extraction d'éléments graphiques dans les images de manuscrits

Les méthodes de structuration de documents nécessitent des images débarrassées des éléments graphiques non textuels. Les documents étudiés lors du projet Philectre reflètent la genèse d'une œuvre et contiennent de nombreux éléments graphiques. Ce sont les ratures portant sur une ligne ou des paragraphes, les signes de renvoi et d'entourage. De grands traits (appelés croix de St André) sont parfois apposés sur la page par l'auteur après recopie du brouillon. La fonction et la signification des ratures sont analysées dans [de Biasi, 1994] en fonction de leur localisation, de leur étendue, de l'objet sur lequel elles portent. L'autre avantage d'éliminer ces traits est que l'extraction et la suppression de ces éléments améliore la visibilité du document à l'écran pour les activités de déchiffrage.

La méthode d'extraction d'éléments graphiques dans que nous avons développée dans [Likforman-Sulem 1998] utilise la technique du filtrage de Kalman et est utilisée au travers d'une interface homme-machine. La méthode de Kalman de suivi est généralement utilisée pour le suivi de d'objets dans des séquences d'images. Elle a aussi été utilisée avec succès sur des images isolées pour le suivi de routes [Veran, 1993] et dans le domaine du document imprimé, pour la détection des segments (portées, barres) [Poulain d'Andecy et al., 1994]. Le modèle de filtre développé pour suivre les traits longilignes (ratures) et présenté ci dessous.

A.5.3.1 Modélisation

Le filtrage de Kalman consiste à suivre un objet pas à pas à partir de son modèle, du modèle de son évolution, de mesures ou observations faites dans l'image ainsi que de la modélisation des

erreurs pouvant affecter son observation dans l'image et son évolution. L'objet à suivre est ici un trait. Il est représenté par un vecteur d'état $\mathbf{X}_k$ caché dont on cherche l'estimation $\mathbf{X}_k^+$ à chaque itération k. La concaténation des vecteurs $\mathbf{X}_k^+$ correspond au trait suivi. Le filtre de Kalman réalise son estimation à partir de la *prédiction* du vecteur d'état, notée $\mathbf{X}_k^-$, ainsi que de son *observation* $\mathbf{Z}_k$. Les observations sont la partie observable du vecteur d'état et correspondent ici à des séquences de plages noires trouvées dans l'image. Pour les observations, on a de même la prédiction de l'observation notée $\mathbf{Z}_k^-$ et l'observation réelle notée $\mathbf{Z}_k^+$. Les composantes du vecteur d'état $\mathbf{X}_k$ sont :

- les coordonnées de la position centrale de la rature : x_k , y_k
- la pente de la rature par rapport à la verticale : p_k
- l'épaisseur de la rature : e_k

Ce vecteur est indexé par k qui correspond au déplacement en ligne ou en colonne par rapport au point de départ de la rature.

Dans notre problème l'évolution des traits orientés dans le sens de la hauteur est régie par :

$$\begin{aligned} x_k &= x_{k-1} + p_{k-1} + v_1 \\ y_k &= y_{k-1} - 1 \\ p_k &= p_{k-1} + v_2 \\ e_k &= e_{k-1} + v_3 \end{aligned}$$

où le vecteur $\mathbf{V}=(v_1, v_2, v_3)$ est un bruit de modèle supposé blanc gaussien, indépendant de l'étape k, de moyenne nulle et de matrice de covariance Q. Le paramètre y_k est entièrement déterministe car nous recherchons les observations ligne par ligne.

Le vecteur d'observation $\mathbf{Z}_k=(x_c(k), y_c(k), e_c(k))$ est recherché dans l'image autour de l'observation prédite. Il correspond à une séquence de pixels noirs de point central (x_c (k), y_c (k)), et de largeur (qui correspond à l'épaisseur du trait) : $e_c(k)$. Une observation $\mathbf{Z}_k$ est liée au vecteur d'état $\mathbf{X}_k$ par l'équation :

$$\begin{aligned} x_c(k) &= x_k + w_1 \\ y_c(k) &= y_k \\ e_c(k) &= e_k + w_2 \end{aligned}$$

où W=(w1,w2) est un bruit d'observation supposé blanc gaussien, de moyenne nulle et de matrice de covariance R .

A.5.3.2 Algorithme de filtrage

A partir de l'estimation du vecteur d'état précédent $\mathbf{X}_{k-1}^{+}$ et de la matrice de covariance d'erreur d'estimation correspondante notée P_{k-1}^{+}, l'algorithme de filtrage estime $\mathbf{X}_{k}^{+}$ en trois étapes : prédiction, mise en correspondance et mise à jour. Les deux étapes de prédiction et de mise à jour étant des étapes classiques relatives au filtre de Kalman, nous détaillons ci dessous l'étape de mise en correspondance spécifique à notre problématique. Il s'agit de sélectionner autour du point prédit d'observation $\mathbf{Z}_{k}^{-}$ la séquence de plage $\mathbf{Z}_{k}^{+}$ qui correspond à un trait de rature ou à une superposition rature écriture. Une fois l'observation sélectionnée, on la valide si elle correspond à une rature ou on la rejette si elle correspond à une superposition d'éléments. Les cas d'échec lors de la mise en correspondance sont donc ceux où il n'y a pas d'observation (trait interrompu) ou lorsqu'il y a superposition d'éléments.
On recherche une observation à l'intérieur d'une zone située sur la ligne du point d'observation prédit $\mathbf{Z}_{k}^{-}$ et proche de ce point. Les observations sont recherchées sur une distance $l \leq S_l * e_i$. Le paramètre global e_i est l'épaisseur de la rature supposée constante au cours du suivi. Ce paramètre est recherché lors de l'initialisation.

Soit L la longueur de la séquence trouvée dans la zone de recherche. L'observation est validée si $L \leq S_L * e_i$. Si deux séquences sont trouvées dans la zone de recherche, la plus proche du point prédit est conservée. Des valeurs typiques pour les seuils S_l et S_L sont 2 et 0.5 .
La sélection ainsi que la validation sont contraintes par le paramètre global e_i. En effet, il faut éviter que le filtre modifie fortement le paramètre d'épaisseur du vecteur d'état. Sinon les observations issues de l'intersection du trait avec de l'écriture ne seraient plus rejetées.

En cas d'échec lors de la mise en correspondance, le vecteur d'état est mis à jour à partir du vecteur d'état prédit, ce qui permet de suivre les traits même en cas d'interruption. La position estimée est alors position prédite et la pente estimée est la pente moyenne entre la pente initiale et la pente prédite pour permettre de légères incurvations.

La position initiale du vecteur d'état est donnée manuellement (au travers d'une interface) pour désigner le début du trait à suivre. On clique à la souris sur deux points de la rature, le premier en début de trait. On initialise ainsi le vecteur d'état (position et pente). La largeur de la séquence de plage au premier point cliqué donne le paramètre d'épaisseur du vecteur d'état qui est aussi le paramètre global e_i .

Pour suivre les traits de renvoi ou d'entourage, des adaptations sont nécessaires pour permettre les changements importants de direction lors du suivi. Lors de ces changements, il faut aussi commuter entre un modèle à progression constante en y (comme présenté précédemment) et un

modèle à progression constante en x si le trait s'incurve dans la direction horizontale. Les critères utilisés pour déterminer un changement d'évolution sont la pente et son signe.

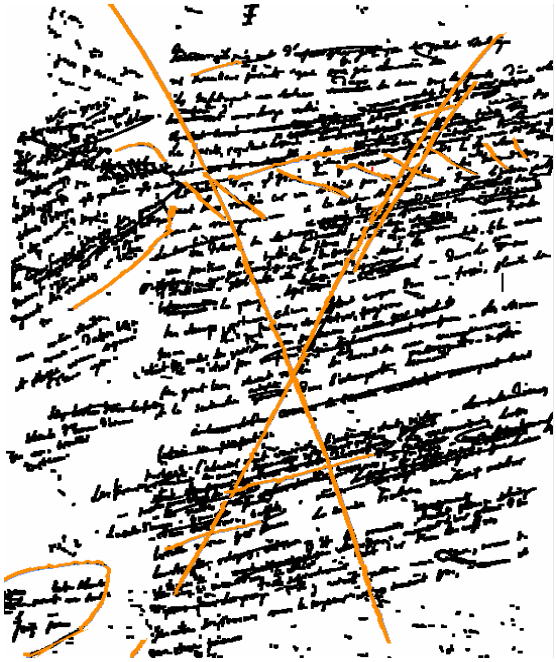


Figure 13. *Extraction des longs traits dans un manuscrit de Flaubert*

A.5.3.3 Conclusion

Les documents numérisés à partir de photocopies noir et blanc, ont une résolution allant de 150 à 200 dpi. Certaines images ont été ensuite réduites puis dilatées en vue d'accentuer le contraste trait-écriture et d'homogénéiser l'épaisseur des traits. La méthode de suivi a été appliquée à une dizaine de manuscrits d'auteurs (Flaubert, Ponge) comme celui présenté en figure 13. Le résultat du suivi apparaît en orange pour chacun des éléments graphiques préalablement désignés à la souris. Sur l'écran, l'utilisateur voit progresser pas à pas le suivi le long des traits.

Le filtre arrête de suivre un élément quand il trouve des points appartenant au fond pendant une courte séquence. Dans ce cas, si un élément a été suivi de manière incomplète, on relance le filtre sur la partie restante à l'aide de l'interface. La plupart des éléments trouvés sur cette figure ont été suivi en une fois, à l'exception des deux grandes ratures obliques qui ont été suivies en deux fois.

En ce qui concerne les traits courbes, ceux ci sont d'autant mieux suivis que la courbure est douce. En effet pour commuter entre les modèles, il faut disposer d'observations indiquant un changement de direction. Si le changement de direction est trop brutal (≥90°), les observations correspondantes risquent d'être rejetées et le changement non détecté. Ce type de traits est généralement trouvé en plusieurs parties.

Une fois le suivi réalisé, on construit une image éclaircie où les points appartenant aux éléments graphiques suivis et qui ne superposent pas avec de l'écriture, sont retirés (les points restants sont ceux où l'observation n'a pas été validée à cause d'une épaisseur trop élevée de l'observation). Cependant le fait de travailler sur une image au départ réduite, rend celle ci non lisible à l'écran. Une possibilité serait de travailler avec plusieurs images de résolutions différentes. Le suivi serait effectué à basse résolution et l'image affichée et éclaircie le serait à haute résolution pour permettre son déchiffrement.

A.6 Conclusion

La structuration des documents manuscrits est plus complexe que celle des documents imprimés car il n'y a pas de régularité dans l'espacement et l'inclinaison des lignes. Les lignes de texte présentent de nombreuses autres irrégularités telles les composantes connectées à travers plusieurs lignes, ainsi que des hampes et des jambages, qui emplissent l'espace inter-ligne. Deux méthodes d'extraction de lignes ont été présentées. Celles ci sont adaptées aux documents de type brouillons et à certains manuscrits d'auteur.

La première méthode est basée sur le groupement itératif des composantes connexes mais inclut des critères de groupement locaux et d'évaluation globale des alignements formés issus de la théorie de la Forme (Gestalt). Ces critères permettent de s'adapter à la recherche de lignes dans différentes directions et à la fluctuation des lignes. La deuxième méthode est basée sur la transformée de Hough et est appliquée sur les centres de gravité des composantes connexes de l'image. La méthode inclut des critères de voisinage dans le domaine de l'image et dans le domaine de Hough qui permettent de rejeter les alignements non pertinents perceptivement et de s'adapter aux fluctuations des lignes. Cette méthode permet d'extraire des lignes de différentes orientations qui coexistent dans un voisinage proche sur la page comme les lignes du corps principal et les annotations en biais. Une comparaison qualitative des deux méthodes a été présentée.

La dernière partie présente la comparaison de ces méthodes réalisée lors du projet Philectre de développement d'outils à l'usage des philologues. La sélection des lignes, interactive ou semi-automatique, a pour objectif d'établir la correspondance entre l'image du manuscrit et sa transcription diplomatique. Pour améliorer le déchiffrement du texte, ou avant d'extraire les lignes, l'image du document doit être nettoyée de ses éléments non textuels. Nous avons proposé une méthode d'extraction des éléments graphiques tels que les croix de St André et les traits d'entourage sur un mode collaboratif. La technique est basée sur le filtrage de Kalman. Le point central du trait de rature est suivi ligne à ligne malgré les interruptions, à partir d'un point et d'une direction initiale fournis par l'utilisateur.

La partie A concerne la structuration d'un document en lignes qui sont des éléments de sa structure *physique*. Et nous avons eu pour cadre d'application des documents manuscrits. Dans la partie B, nous allons rechercher des éléments relatifs à la structure *logique*. Les documents

qui seront étudiés sont des documents imprimés et des documents mixtes (incluant à la fois des éléments imprimés et manuscrits).

Bibliographie (sujet A)

publications issues du sujet A

Lhuillier L. (1991), *Structuration de documents manuscrits*, mémoire de DEA IARFA de l'Université de Paris VI.

Faure C., Likforman-Sulem L. (1993), Traitement automatique de l'écrit : structuration perceptive et catégorisation, *Textuel*, no 25, pp 37-54.

Hanimyan A. (1994), *Utilisation de la transformée de Hough pour la détection des alignements dans les textes manuscrits non contraints*, Thèse Professionnelle de Mastère.

Likforman-Sulem L., Faure C. (1993**),** Extracting text lines in handwritten documents by perceptual grouping, *6-th International Conference on Handwriting and Drawing*, Paris, pp. 192-194.

Likforman-Sulem L., Faure C., (1994a), Extracting text lines in handwritten documents by perceptual grouping, in *Advances in handwriting and drawing : a multidisciplinary approach* (C. Faure, P. Keuss, G. Lorette, A. Winter, eds), Europia.

Likforman-Sulem L., Faure C. (1994b), Une méthode de résolution de conflits d'alignements pour la segmentation des documents manuscrits, *Actes de CNED 94*, Rouen, pp. 265-272.

Lecolinet E., Likforman-Sulem L. (1994), Handwriting Analysis: segmentation and recognition, *IEE European workshop on handwriting analysis and recognition*, Bruxelles, pp. 17/1- 17/8.

Wang A. (1994), *Détection des lignes dans les textes manuscrits non contraints*, Mémoire de DEA IARFA de l'Université de ParisVI, 1994.

Likforman-Sulem L., Faure C. (1995), Une méthode de résolution des conflits d'alignements pour la segmentation des documents manuscrits, *Traitement du Signal*, Vol. 12, no 6, pp. 541-549.

Likforman-Sulem L., Hanimyan A., Faure C. (1995)**,** A Hough Based Algorithm for Extracting Text Lines in Handwritten Documents, *Actes de Int. Conf. On Document Analysis and Recognition ICDAR'95*, Montréal, pp. 774-777.

Likforman-Sulem L., C. Faure (1996), Structuration de manuscrits pour l'édition électronique, *Actes de CNED 96*, Nantes, pp. 267-273.

Likforman-Sulem L., Robert L., Lecolinet E., Lebrave J-L., Cerquiglini B. (1997) Edition hypertextuelle et consultation de manuscrits, *Hypertextes et Hypermédias*, Vol. 1, no 2-3-4, pp. 299-310.

Robert L., Likforman-Sulem L., Lecolinet E. (1997), Image and Text Coupling for Creating Electronic Books from Manuscripts, *Actes de ICDAR'97*, Ulm, pp. 823-826.

Likforman-Sulem L. (1998), Extraction d'éléments graphiques dans les images de manuscrits, *Colloque International Francophone sur l'Ecrit et le Document (CIFED'98)*, Québec, pp. 223-232

Lecolinet E., Role F., Likforman-Sulem L., Lebrave J-L, Robert L. (1998), An integrated reading and editing environment for scholarly research on literary works and their

Louloudis G., Gatos B., Pratikakis I., Halatsis K., A Block-Based Hough Transform Mapping for Text Line Detection in Handwritten Documents, IWFHR 06, La Baule.

Gusnard De Ventabert, André J., Richy H., Likforman-Sulem L. (1999), Représentation et exploitation électroniques de documents anciens, *Document Numérique*, Vol 3, no1-2, pp. 57-73.

Likforman-Sulem L. (2003), Apport du traitement des images à la numérisation des documents manuscrits anciens, *Document Numérique*, Hermès, Vol 7, no 3-4, pp. 13-26.

Likforman-Sulem L., Zahour A ., Taconet B., (2007) Text Line Segmentation of Historical Documents: A Survey, *International Journal on Document Analysis and Recognition*, DOI 10.1007/s10032-006-0023-z, Springer, 2006.

Autres Références

André J., Richy H. (1997), Hypertextes ou documents structurés ? étude de cas en critique génétique, *4ème Conférence Hypertextes et Hypermédias*, Saint Denis, France, pp. 13-26.

Buser P., Imbert M. (1987), Vision, Herman, pp 404-436.

Biasi De P-M (1996), "Qu'est-ce qu'une rature ?", 5ème Colloque CIGADA Ratures et repentirs, 1-3 dec 1994, B. Rougé (ed), Publications de l'Université de Pau, pp. 17-48.

Cerquiglini B., Lebrave J-L., Philectre (1994), Un projet de recherche pluri-disciplinaire en philologie électronique, *Gis de la cognition*, CNRS.

Cohen E., J. Hull, S. Srihari, (1991), Understanding handwritten text in a structured environment : determining zip codes from addresses, *Int. Journal of Pattern Recognition and AI*, Vol. 5, No 1 & 2, June, World Scientific, 1991, p. 221-264.

Downton A. C, Leedham C. (1990), Preprocessing and presorting of envelope images for automatic sorting using OCR, *Pattern Recognition*, Vol. 23, No 3-4, pp. 347-362.

Fekete J-D, Dufournaud N.(1999), Analyse historique de sources manuscrites : application de TEI à un corpus de lettres de rémission du XVIe siècle , *Document numérique*, Hermès, Vol. 3, no 1-2, pp. 117-134.

Feldbach M., Tönnies K.D. (2001), Line detection and segmentation in Historical Church registers, Proc. of ICDAR'01, Seattle, pp. 743-747.

Fletcher L. A., Kasturi R. (1988), Text string segmentation from mixed text/graphics images, *IEEE PAMI*, Vol 10, No 3, pp. 910-918.

Guillaume P.(1979), La psychologie de la Forme, Flammarion, chapitre III.

Govindaraju V., R. Srihari, S. Srihari (1994), Handwritten text recognition, *Actes de Document Analysis Systems DAS 94*, Kaiserlautern, p. 157-171.

Hinds S. C., Fisher J., D'Amato D.(1990) A document skew detection method using run-length encoding and the Hough transform, *Proceedings of the 10th IAPR*, Atlantic City, pp 464-468.

Hough, P.V.C. Method and means for recognizing complex patterns. *United States Patent*, n° 3 069 654, (1962).

Koch G., Heutte L., Paquet T. (2005), Automatic extraction of numerical sequences in handwritten incoming mail documents, Pattern Recognition, Vol. 26, p. 1118-1127.

Meynieux E., S. Seisen, K. Tombre (1989), Bilevel Information Recognition and Coding in Office Paper Documents, *8th Int. Conf. on Pattern Recognition*, Rome, p. 442-445.

Nicolas S, Kessentini Y., Paquet T., Heutte L. (2005) Handwritten document segmentation using hidden Markov random fields, *8th International Conference on Document Analysis and Recognition, ICDAR'2005*, Seoul, Corée, pp. 212-216, 2005.

Nagy G., S. Seth (1984), Hierarchical Representation of optically scanned documents, *7th Int. Conf. on Pattern Recognition*, Montréal, p. 347-349.

Noi Bai, N., Nam K., Song Y. (2008), Extracting curved text lines using the chain composition and the expanded grouping method, SPIE Document Recognition and Retrieval XV, San Jose,

Oztop E., Mulayim A. Y., Atalay V., Yarman-Vural F. (1999), Repulsive attractive network for baseline extraction on document images, *Signal Processing*, 75:1-10.

Paquet Th., R. Mullot, R. Trupin, K. Romeo, Y. Lecourtier (1989), Un algorithme rapide de détection des mots d'un texte manuscrit, *Congrès AFCET-RFIA*, Paris, p. 1501-1510.

Poulain d'Andecy V., Camillerapp J., Leplumey I. (1994), Analyse de partitions musicales, *Actes de CNED'94*, Rouen, pp. 223-232.

Ruiz-Pinales J. (2002), *Reconnaissance hors ligne de l'écriture cursive par utilisation de modèles perceptifs et neuronaux*, Thèse de Doctorat de l'ENST, septembre 2002.

Seni G., Cohen E. (1994) External word segmentation of off-line handwritten documents, *Pattern Recognition,* Vol 27, No 1, pp 41-52.

Shapiro V., G. Gluhchev G., V. Sgurev (1993), Handwritten document image segmentation and analysis, *Pattern Recognition Letters*, No 14, pp. 71-78.

Srihari S. , Govindaraju V. (1989) Analysis of textual images using the Hough transform, in *Machine Vision and Applications*, 2, pp 141-153.

Touj S., Ben Amara N., Amiri H. (2002), Reconnaissance hors ligne de caractères arabes isolés manuscrits, *actes de CIFED 02*, Hammamet.

Tseng Y.H., Lee H.J. (1999), Recognition-based handwritten Chinese character segmentation using a probabilistic Viterbi algorithm, *Pattern Recognition Letters*, 20(8):791-806.

Veran J-P (1993), *Suivi de routes dans une image aérienne par filtrage de Kalman*, Rapport ENST 93 D007.

Viard-Gaudin C. , D. Barba (1992), Extraction robuste et structuration des informations par une approche multirésolution pour la localisation du bloc adresse sur les objets postaux plats, Actes de CNED'92, *Bigre no 80*, pp. 48-56.

Vincent N., Dargenton P., Emptoz H. (1992), Utilisation de la transformée de Hough dans la reconnaissance de l'écriture manuscrite, *Actes de CNED'92,* Nancy, pp 294-301.

Wong K., R. Casey, F. Wahl (1982), Document analysis system, *I.B.M. Journal of Research and Development*, 26, no 6.

Zahour A, Taconet B., Ramdane S. (2004), Contribution à la segmentation de textes manuscrits anciens, *Proc. of CIFED 2004*, La Rochelle.

B. EXTRACTION DE NOMS PROPRES DANS LES DOCUMENTS DEGRADES

B.1 Introduction

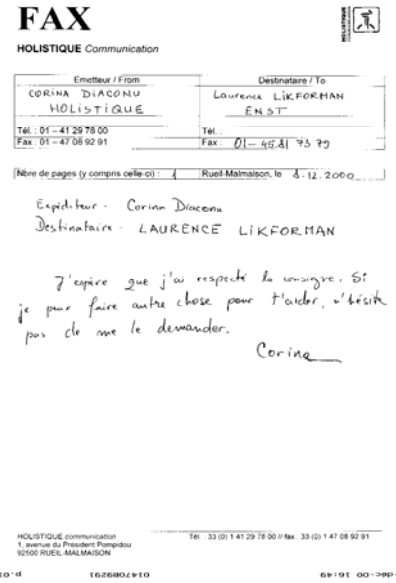
Cette recherche a été effectuée dans le cadre du projet Eurekâ 2340-Majordome financé par le MINEFI (convention no 01.2.93.0268). Le projet Majordome consiste à développer des outils pour les messageries dites unifiées. Ces messageries proposent différents services (messages vocaux, SMS, fax, mèls, agenda, carnet d'adresses, signets, etc.). Si on traite de façon intelligente les informations (filtrages, résumés, reconnaissance de l'expéditeur ou de l'appelant), on peut alors réduire le flux d'informations délivré à l'utilisateur. Ces fonctionnalités sont utiles aux travailleurs nomades de l'entreprise qui ont ainsi accès à une information condensée de leurs messages. Et pour les travailleurs plus sédentaires, ces informations permettent de mieux trier et gérer les messages reçus.

Dans une messagerie classique qui reçoit des documents (télécopies, fichiers attachés,...) les documents sont indexés par le numéro de téléphone ou de fax de l'émetteur du message. Ce numéro n'est pas toujours un numéro personnel, ce qui rend l'indexation par numéro peu précise. D'autre part, les numéros des nouveaux clients, encore inconnus du système de messagerie ne peuvent être mis en correspondance avec le nom de l'émetteur. La reconnaissance des informations présentes dans les messages textuels, comme le nom de l'émetteur permet leur indexation. Cette information synthétique permet aussi de hiérarchiser les messages lors d'une consultation distante. Pour la tâche de routage, la reconnaissance du nom associé au destinataire permet de diriger précisément le message en cas de numéro de réception unique.

Reconnaître automatiquement les différentes parties d'un document (télécopie, lettre) est relatif à l'extraction de la structure logique. Pour cela il est tout d'abord nécessaire d'extraire la structure physique du document (blocs, lignes), puis d'attribuer une fonction à ces blocs (date, signature, coordonnées de l'envoyeur, de l'expéditeur). Le domaine de la bureautique est un domaine d'application privilégié pour l'extraction de la structure logique car celle ci permet l'archivage et l'indexation des documents. L'extraction de la structure logique s'applique notamment aux lettres d'affaires et aux factures. Plus récemment, l'augmentation du nombre de documents pdf circulant sur Internet, fait émerger des techniques d'extraction de structures sur ces documents à partir de leur contenu textuel et de la mise en page. Ceci permet leur conversion au format html, leur réédition ou la recherche d'information [Hadjar et al. 2004].

Les systèmes d'étiquetage logique reposent généralement sur un apprentissage de modèles représentatifs des différentes classes que le système aura à traiter. Cet apprentissage est réalisé de manière supervisé dans [Baumann et al. 1997] à partir de documents exemples. Les objets logiques sont repérés à partir de la position des blocs physiques correspondants. De même Cesarini dans [Cesarini et al. 1998],] repère les montants et numéros de facture à partir de la position physique (position absolue et/ou relative) des blocs correspondants après avoir reconnu

le logo de l'organisme émetteur. L'apprentissage peut être aussi réalisé de manière incrémentale comme dans [Liang et Doermann 2002 et 2003] où le modèle logique des documents, représenté sous forme de graphe, est affiné au fur à mesure que l'on présente les exemples d'une classe de document. Cependant, c'est l'utilisateur qui vérifie ici si le système a fait une erreur et si le modèle doit être modifié. [Klink et al., 2001] construit un ensemble de règles floues définies manuellement pour interpréter les lettres d'affaires.


FAX

HOLISTIQUE *Communication*

Emetteur / From	Destinataire / To
CORINA DIACONU HOLISTIQUE	Laurence LIKFORMAN ENST
Tél. : 01 – 41 29 78 00 Fax : 01 – 47 08 92 91	Tél. : Fax : 01 – 45 81 73 79

Nbre de pages (y compris celle-ci) : 1 Rueil-Malmaison, le 8.12.2000

Expéditeur : Corina Diaconu
Destinataire : LAURENCE LIKFORMAN

J'espère que j'ai respecté la consigne. Si je peux faire autre chose pour t'aider, n'hésite pas de me le demander.

Corina

HOLISTIQUE *communication*
1, avenue du Président Pompidou
92500 RUEIL-MALMAISON

Tél : 33 (0) 1 41 29 78 00 // fax : 33 (0) 1 47 08 92 91

Figure 1. *Exemple de télécopie reçue sur la plate-forme*

Nous avons traité dans Majordome les documents dégradés de type télécopies et notre recherche a porté sur l'extraction de la structure physique et sur certains éléments de la structure logique. L'expérience a montré que les noms recherchés peuvent apparaître dans toutes les zones de l'image. Nous proposons une méthode qui utilise des conventions de communication génériques et les chaînes de caractères, plutôt que les positions. En ce qui concerne la structure physique, le stage de B. Cuenca [Cuenca 1999] a permis de construire un classifieur pour la discrimination des zones manuscrites et imprimées par sélection de caractéristiques spécifiques et classification par réseau neuronal. Ce travail a été publié dans [Likforman-Sulem et Cuenca 1999]. En ce qui concerne la structure logique, nous avons défini dans [Likforman-Sulem 2001] le concept de *Paire Logique* qui s'appuie sur des critères perceptifs et qui permet de réaliser un ancrage dans la structure logique. Ce concept est adapté aux documents non contraints de structures variées. Nous avons élaboré dans [Vaillant et al. 2002][Likforman-Sulem et al. 2003] une méthode d'analyse mixte qui combine des caractéristiques issues de deux analyses : analyse sur l'image du document et analyse textuelle sur les chaînes de caractères issues de la reconnaissance. Nous avons également étendu cette approche à la recherche des noms d'auteurs dans les images d'articles de revues pendant le stage d'A. de Bodard, travail qui a été publié dans [Likforman-Sulem et al. 2006].

D’autre part, ce projet a nécessité la construction de deux bases de données : une base d’images de télécopies dont les champs expéditeurs et destinataires ont été étiquetés et une base de caractères majuscules, la base ENST-FAX-CHAR de caractères issus de télécopies [Godeau et al. 2002].

B.2. Construction de bases de données

B.2.1 Base d’images de télécopies

La construction d’une base de données d’images de télécopies a été nécessaire pour cette étude. Une plate-forme constituée d’un PC équipé du logiciel Winphone, et d’un modem a permis de récupérer les télécopies en différents formats (Gif, Tiff, Bmp).

150 images de télécopies ont été collectées sur la plate-forme en sollicitant des volontaires situés dans des organismes ou entreprises diverses, afin que les pages d’en-tête reçues aient des mises en page les plus variées possible. Le nom du destinataire à inscrire dans le champ ‘destinataire’ est préalablement sélectionné parmi une liste de 10 noms pour simuler un environnement de travail (PME par exemple) ou familial.

Un tiers seulement des demandes de participation à la base ont abouti à l’envoi effectif d’une télécopie. Les consignes sont diversement suivies. De pages d’en-tête papier reçues au Laboratoire ont été numérisées par scanner pour compléter la base.

Un étiquetage des noms propres expéditeur et destinataire, ainsi que celui des champs correspondants (Expéditeur, Destinataire, Nom -le cas échéant) a été réalisé en indiquant la position dans l’image (boite englobante), la chaîne de caractères, et le type d’écriture (imprimé ou manuscrit).

B.2.2 Base de caractères ENST-FAX -CHAR

De la base de données d’images de télécopies réalisée précédemment, nous avons extrait une base de caractères manuscrits isolés (stage de J. Godeau, [Godeau, 2002]). Cette base reflète la variabilité de l’écriture manuscrite bâton (Fig. 2). La résolution faible de ces images (qualité télécopie) en font une base difficile pour les algorithmes de la reconnaissance des caractères.

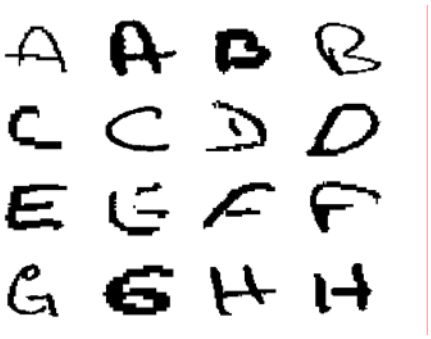


Figure 2. *Exemples de caractères de la base ENST-FAX-CHAR*

Une interface de saisie a été mise au point. Les images de télécopies (format tif) sont affichées et les caractères choisis découpés à la souris. Les imagettes sont enregistrées en format image pbm. Environ 6000 caractères manuscrits de type majuscule ont été découpés (environ 270 exemples par classe). La base ENST-FAX-CHAR [Godeau et al. 2002] est disponible pour la communauté scientifique et est actuellement utilisée par le CTA (thèses de S. Chevallier et M. Lemaître).

B.3 Extraction des noms propres

La recherche de noms propres dans les images de télécopie se heurte à difficultés spécifiques. Ces documents sont mixtes, incluant des informations manuscrites et/ou imprimées. Il est difficile de se baser sur des modèles appris au préalable car le nombre de mises en page possibles est très grand. La mise en page d'une télécopie peut être celle d'une lettre ou peut contenir des champs. La télécopie peut contenir des tableaux, être sur une ou plusieurs colonnes... D'autre part, la faible qualité des images de télécopie entraîne des erreurs de reconnaissance de caractères par OCR (Optical Character Recognition). Plusieurs langages peuvent être utilisés sur une même télécopie et il existe un certain nombre de synonymes (par exemple pour désigner un envoyeur), mais aussi des homonymes qui véhiculent des informations différentes (ex 'de'). Il n'y a parfois pas de mot clé pour repérer le nom cherché, qui n'est présent que dans la signature.

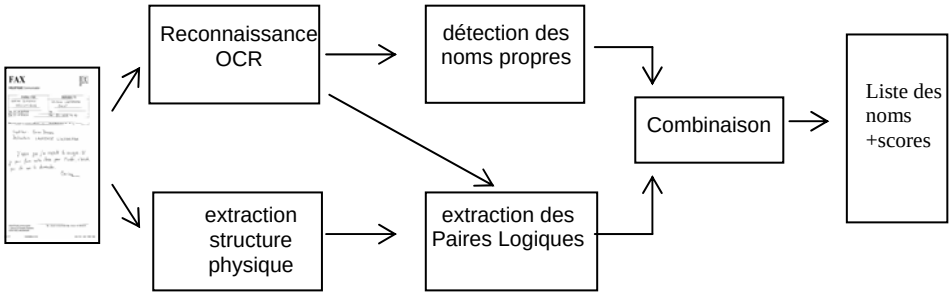


Figure 3. *Analyse mixte image et textuelle des images de télécopie*

L'analyse d'images de télécopies a été étudiée dans la littérature pour la tâche de routage. Les auteurs de [Lii et al. 1995] sélectionnent les champs en fonction de critères spatiaux uniquement, puis appliquent la reconnaissance OCR sur ces champs. Alam [Alam et al. 2000] recherche tout d'abord les chaînes de caractères correspondant aux noms propres, puis sélectionne le nom du destinataire sur des critères de position dans l'image. Le système de Viola [Viola et al. 2004] extrait un grand nombre de caractéristiques sur chaque mot (plus de 2000, incluant les occurrences et cooccurrences de mots) et c'est un classifieur de type AdaBoost qui sélectionne les caractéristiques utiles à la tâche d'extraction du nom du destinataire. Des bases de noms clients sont très souvent utilisées dans les tâches de routage et la difficulté est de réaliser rapidement l'appariement entre ce qui reconnu et le nom dans la base.

Nous avons ensuite développé une analyse mixte (Fig. 3), spatiale et textuelle, qui détecte les noms propres (expéditeur/destinataire) dans les images de télécopies. L'objectif de l'analyse spatiale est de se focaliser sur les zones de l'image où les noms propres relatifs aux expéditeurs sont le plus susceptibles de se trouver. L'analyse textuelle complète l'analyse spatiale en examinant les chaînes de caractères issues d'un OCR. Dans le cadre de documents dégradés ces analyses complémentaires permettent de pallier aux défauts de chaque analyse prise isolément. Ces analyses nécessitent l'extraction préalable d'éléments de la structure physique et logique (les Paires Logiques) du document (voir Section B.3.2).

B.3.1 Extraction et analyse de la structure physique

La structure physique est composée d'éléments de la taille des mots, appelés *pseudo mots*. Ce niveau d'analyse permet d'éviter de fusionner des composantes manuscrites et imprimées appartenant à la même ligne de texte. Cependant, des groupes de mots peuvent être fusionnés si leur espacement est faible. Les pseudo mots sont extraits par l'algorithme RLSA [Wong et al. 1982] puis classés en deux classes, imprimé ou manuscrit. L'extraction des pseudo mots et leur classement est exposé dans [Likforman-Sulem et Cuenca 1999]. Le classement se fait par réseau neuronal à partir de trois critères relatifs à la disposition des composantes connexes dans le pseudo mot : nombre de composantes connexes, différence de hauteur entre la plus petite et la plus grande composante, différence moyenne de positionnement entre deux composantes consécutives. La classification dans la classe 'manuscrit' (resp. ' imprimé') s'obtient en comparant la sortie du réseau correspondant à cette classe au seuil θ_h (resp. θ_p). Pour favoriser la bonne classification dans la classe 'manuscrit', on choisit $\theta_h < \theta_p$ et on obtient un taux de bonne classification de 97.4 % pour les pseudo mots manuscrits, et de 79.8 % pour les pseudo mots imprimés.

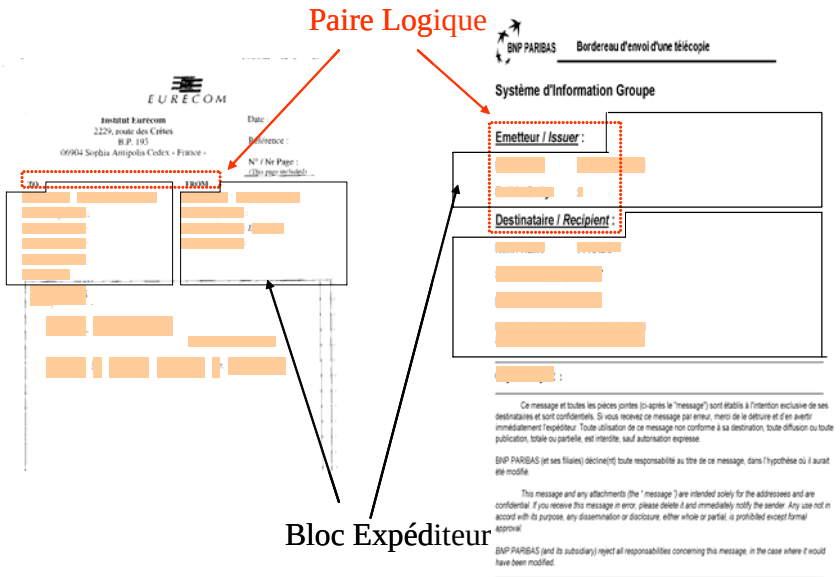


Figure 4. *Paires Logiques et blocs Expéditeur associés*

B.3.2 Extraction des Paires Logiques

La recherche de noms propres repose habituellement sur l'analyse textuelle seule. Dans le cas de documents dégradés, les chaînes produites par OCR peuvent contenir des erreurs sur la transcription des noms. D'autre part, nous recherchons plutôt les noms propres du bloc expéditeur. Nous utilisons donc la structure logique du document pour extraire ces noms à partir de la transcription mais aussi de l'image.

Nous avons défini dans [Likforman-Sulem, 2001] une méthode de construction d'objets logiques de plus haut niveau à partir des blocs physiques et de la détection (textuelle) de mots clés. Des critères perceptifs sont utilisés pour valider les hypothèses émises. Des critères perceptifs similaires ont été utilisés dans [Holstege et al. 1991] pour grouper des objets physiques en objets de plus haut niveau. De même [Faure, 1999] combine des critères spatiaux et la détection de mots clés pour la détection des en-têtes de revues scientifiques. Notre analyse suppose que les champs principaux sont alignés (verticalement ou horizontalement). Ceci est vérifié dans la plupart des télécopies de type «en-tête» qui sont structurés en champs. Néanmoins, la base d'images comporte aussi des images de type «lettre». Les critères perceptifs permettent de ne pas supposer un modèle prédéfini pour extraire la structure logique correspondant aux champs (expéditeur/destinataire) principaux.

Les champs (expéditeur/destinataire) sont repérés à l'aide d'un dictionnaire. Cependant toutes sortes de mots homonymes sont aussi repérés, ainsi que les mots proches (au sens de la distance de Levenshtein) du vocabulaire considéré. La sélection des blocs pertinents est réalisée par sélection d'hypothèses : les hypothèses sont sélectionnées par la contrainte d'alignement (horizontal ou vertical). Les paires de blocs sélectionnées constituent des *Paires Logiques*. Chaque bloc de la Paire Logique définit une zone de recherche (bloc expéditeur ou bloc destinataire). Dans cette zone, les pseudo mots les plus proches du champ expéditeur constituent les hypothèses d'objets logiques «nom expéditeur». La figure 4 montre deux exemples de Paires Logiques formées de deux blocs physiques (contenant sur un des exemples les mots clés 'from' et 'to'). Ces blocs sont alignés horizontalement, ce qui permet d'émettre l'hypothèse que ce sont bien des intitulés de champs. Puis à chaque hypothèse correspondent des blocs expéditeur/destinataire dont la taille et la forme s'adaptent à la configuration locale.

B.3.3 Analyse textuelle

La reconnaissance des entités nommées (noms propres) constitue l'une des tâches de l'extraction automatique d'information dans les textes (cf. séries des conférences MUC-Message Understanding Conference [MUC, 1998]). Les entités nommées comprennent les noms de personnes, mais aussi les noms de lieux ou d'organisations. La reconnaissance des entités nommées inclut généralement une première étape d'analyse syntaxique du texte (*parsing*) nécessitant un texte sans erreurs de transcription ni de ponctuation.

Selon les applications et les données disponibles, les méthodes et les stratégies suivies peuvent être extrêmement variées (cf. [TAL 1998]) : règles de marquage, modèles HMMs. Elles peuvent aussi être de type supervisé ou non supervisé. [Bikel et al 1997] classe les mots en différentes catégories par un apprentissage supervisé et un classifieur de type HMM. Dans

[Collins et Singer 1999], des règles de classement des mots (présélectionnés par une analyse syntaxique) sont construites automatiquement à partir de très peu d'exemples étiquetés permettant d'amorcer le processus non supervisé. Les caractéristiques extraites sur les mots sont, dans ces deux travaux, liées à la typographie et à la présence ou non de marqueurs (Mr, company, etc...) dans le voisinage immédiat du mot à classer.

Dans les documents dégradés, tels que les télécopies (ou les documents scannés à partir de photocopies), des erreurs de transcription existent au niveau des mots et de la ponctuation. Dans les télécopies, la ponctuation est souvent très réduite. L'objectif de l'analyse textuelle consiste à marquer dans la version transcrite par OCR, les mots susceptibles d'être des noms propres relatifs à l'identité d'une personne. Cette analyse présentée dans [Vaillant et al. 2002] [Likforman-Sulem et al. 2003] utilise des dictionnaires (prénoms, mots communs) et des règles locales de marquage. Des structures de données spécifiques, les arbres binaires de recherche équilibrés [Velski-Landis 62], sont utilisées pour représenter ces dictionnaires de taille importante et accélérer la comparaison.

Plusieurs propriétés peuvent fournir des indices permettant de repérer des noms propres, en particulier le nom de l'expéditeur. Nous distinguons les indices *internes* des indices *externes*. Les indices (ou caractéristiques) internes sont relatifs au mot isolé (majuscule en début de mot ; présence dans un dictionnaire de prénoms ; absence dans un dictionnaire de mots communs). Les indices externes sont relatifs au contexte local du mot, (présence de marqueurs d'identité ('M.', 'Mme', 'Mlle', 'Professeur', 'Docteur', etc.) ; présence, dans le contexte local du mot, d'un prénom répertorié dans un dictionnaire de prénoms)

Les indices fournis par ces différentes analyses peuvent être pondérés et combinés entre eux, mais également avec les indices fournis par les analyses purement visuelles de l'image, afin de se renforcer mutuellement et de fournir l'identification la plus sûre possible des entités recherchées.

B.3.4 Combinaisons linéaire et non linéaire

La première étape des traitements textuels consiste en une analyse du fichier résultant du processus OCR de reconnaissance [Xerox, 1994] sur un document, analyse qui en extrait un flot de chaînes annotées avec leurs coordonnées spatiales. L'étape de fusion de ces données hétérogènes, de type spatial et textuel, commence par l'établissement d'un tableau, dans lequel on regroupe, pour chaque mot, les résultats des deux analyses. Les caractéristiques prises en compte sont les suivantes :

(**f1**) le fait que le mot soit dans un bloc hypothèse expéditeur (extrait lors de l'analyse spatiale) est un indice fort ; cependant, plusieurs hypothèses étant générées, il y a également dans ces blocs physiques des mots qui ne sont pas des noms propres ; la caractéristique f1 est égale à un si le mot est à l'intérieur d'un bloc expéditeur.

(**f2**) le fait que le mot soit un précédé d'un *marqueur d'identité* du type 'M', 'Mme' ou 'Mlle' indique qu'il y a de fortes chances que l'ensemble de ce mot et du ou des mots suivants constitue un nom propre ; ou bien le fait que le mot soit un *prénom* répertorié indique

également qu'il y a de fortes chances pour que l'ensemble de ce mot et du mot suivant constitue un nom propre; mais ce n'est pas forcément un des noms recherchés mais des noms apparaissant dans des adresse ou le nom d'institutions ; (ex. 'Marie Curie' dans 'Université Pierre et Marie Curie'); le fait que le mot soit une *initiale* et soit *suivi par un mot en majuscules* va dans le même sens. Un ensemble de règles de marquage donne la valeur un à la caractéristique f2 si le mot fait partie d'une séquence de mots telle que celles énoncées ci dessus.

(**f3**) le fait que l'item repéré sur la page soit identifié par l'algorithme de discrimination comme un item *manuscrit* va également plutôt dans le bon sens, car beaucoup de pages d'en-tête de télécopie pré-imprimées contiennent des champs « en blanc » qui sont remplis à la main par les utilisateurs ; cependant la confiance que l'on peut accorder au résultat de la discrimination imprimé/manuscrit ne doit pas être surévaluée. f3 est égale à un si le mot a été classé comme manuscrit.

(**f4**) le fait que le mot commence par une *majuscule* ou soit entièrement en *capitales* est un indice, mais un indice faible. En effet tous les sigles, tous les mots en début de phrase, commencent également par une majuscule, ainsi que beaucoup d'items isolés répandus dans les pages d'en-tête de télécopie ;

(**f5**) le fait que le mot ne figure pas dans une *liste de mots communs* de la langue dans laquelle le document est rédigé a plutôt une corrélation positive avec le fait qu'il soit un nom propre (mais : comme nous l'avons fait observer, il y a des noms propres qui sont également des noms communs : 'Vaillant') ;

Les caractéristiques f_1-f_5 sont tout d'abord combinées linéairement pour obtenir, pour chaque mot, un score reflétant la possibilité d'être un nom propre expéditeur. Les poids w_i de la combinaison sont déterminés par apprentissage d'un réseau de neurones sans couche cachée avec 5 cellules en entrée et 2 en sortie. La règle d'apprentissage de Widrow-Hoff est appliquée sur l'ensemble des mots extraits de huit télécopies. Si le score du mot est supérieur à un seuil prédéfini (WAT: *word acceptance threshold*), le mot est classé comme nom propre expéditeur.

Après convergence de la règle (25 itérations sur l'ensemble d'apprentissage), le score $score_f$ d'un mot est la sortie du réseau associée à la classe nom expéditeur. Ce score correspond à la combinaison linéaire suivante des entrées du réseau (caractéristiques f_1-f_5) :

$$score_f(mot) = \sum_{i=1}^{5} w_i \times f_i = 3.84 \times f_1 + 1.53 \times f_2 + 0.04 \times f_3 + 1.28 \times f_4 + 0.54 \times f_5 + 0.79$$

L'importance de l'analyse de type image est soulignée par le fort poids devant la caractéristique image f_1. L'ensemble des mots marqués par une règle de marquage (lors de la recherche de la valeur de f2) sont regroupés en une seule expression, dont le score est égal à la somme des scores. L'emploi de la règle de Widrow-Hoff est nécessaire car les classes ne sont pas linéairement séparables. Le calcul d'un score a l'avantage de pouvoir classer les mots suivant leur rang. On peut ainsi proposer plusieurs mots (N premiers choix-top N) et faire fonctionner le

système d'extraction sous différents points opérationnels suivant que l'on cherche à favoriser plutôt le rappel ou la précision.

Nous avons aussi cherché la fonction de combinaison pour les documents exclusivement imprimés. En effet, les mots manuscrits ne sont pas transcrits par l'OCR, et ne bénéficient pas de l'analyse textuelle. Pour évaluer l'apport de la combinaison des deux approches s, textuelle et image, un nouvel ensemble de caractéristiques g_1-g_5 est défini:

g_1 =1 si le mot est dans un bloc expéditeur (0 sinon).
g_2 =1 si le mot est un prénom
g_3 =1 si le mot commence par une initiale (ex : I. ou Ph.)
g_4 =1 si le mot est en lettres capitales ou commence par une lettre capitale
g_5 =1 si le mot n'est pas dans le dictionnaire

A la différence de l'ensemble f_1-f_5, il n'y a pas de règle de marquage pour rechercher des séquences type de noms. Car celles ci peuvent être sujettes à erreur, par exemple si un prénom n'appartient pas au dictionnaire, ou si une erreur OCR s'est glissée dans le prénom, la séquence Prénom+Nom n'est pas détectée. Les caractéristiques g_2 et g_3 servent à repérer les noms propres mais sur des mots isolés.

Comme précédemment, un apprentissage du réseau permet de déterminer les poids de la combinaison linéaire. Après apprentissage et convergence de la règle, on obtient la combinaison suivante pour l'ensemble g_1-g_5 de caractéristiques :

$$score_g(mot) = 3.62 \times g_1 + 1.59 \times g_2 + 3.52 \times g_3 + 1.57 \times g_4 + 0.11 \times g_5 + 0.36$$

Cette combinaison, comme la précédente, donne de l'importance à la caractéristique de type image (g_1) qui détecte les mots dans le bloc expéditeur. Le fait qu'un mot ne soit pas dans le dictionnaire a aussi moins d'importance que le fait d'être en lettres capitales. Ceci permet d'extraire les noms qui sont aussi dans le dictionnaire des mots communs.

Nous avons ensuite cherché à améliorer la classification des mots (nom expéditeur-pas nom expéditeur) par une combinaison non linéaire des caractéristiques g_1-g_5. Celle ci est réalisée par un perceptron multicouche (MLP) à une couche cachée (de 3 cellules). La convergence nécessite ici 200 itérations du fait de l'existence d'une couche cachée et le score attribué à chaque mot est la sortie de la cellule de sortie dédiée à la classe « nom expéditeur ». La classification des mots est suivie d'un post-traitement qui revient sur l'image pour regrouper un mot ayant un haut score avec ses voisins. En effet, certains voisins comme les initiales reçoivent généralement un score assez bas. Le groupement forme une expression dont le score global est soit la somme des scores de chaque mot, soit la valeur maximale des scores de l'expression. Pour les télécopies, la règle est celle du score maximum car les expressions contiennent peu de mots.

B.3.5 Résultats

La combinaison des critères image et textuels a été testée sur la base des 150 images de télécopies collectées (cf. Section B.2.1). Celle ci inclut des télécopies de type en-tête, mais aussi des télécopies de type lettre où les noms propres apparaissent dans la signature ou dans le bloc adresse. Sur cette base, l'analyse combinée spatiale et textuelle extrait les mots (ou les blocs physiques si l'écriture dans le bloc est classée comme manuscrite) présentant les meilleurs scores.

Figure 5. *Expressions et scores obtenus par combinaison non linéaire des caractéristiques g1-g5.*

La figure 5 montre les expressions classées « nom expéditeur » obtenues sur un exemple bruité (fax scanné à partir de la sortie papier). Le score des expressions est calculé par la combinaison non linéaire des caractéristiques g1-g5. Le nom de l'expéditeur a été trouvé dans l'expression de score le plus élevé. Le nom de l'expéditeur est aussi présent dans la signature : il a également été extrait mais avec un score moins élevé. Les expressions restantes, de score suffisamment élevé, contiennent des mots en capitales et notamment le nom de la société expéditrice France Telecom (notons que « France » appartient au dictionnaire des prénoms), et des mots isolés en capitale absents du dictionnaire car mal reconnus par l'OCR.

Nous avons également testé l'influence respective des indices spatiaux et textuels pour l'ensemble des caractéristiques f1-f5. Trois systèmes ont été comparés par les taux de rappel et de précision. Le système (I) utilise seulement les caractéristiques image (f1) et (f3). Le système (T) utilise seulement les caractéristiques textuelles (f2) (f4) (f5). Le système (I+T) utilise l'ensemble des caractéristiques f1-f5. Le système (I+T+) est le système (I+T) testé sur les documents ne contenant que du texte imprimé. Les performances sont données en Table 1. Ces performances ont été calculées avec un seuil d'acceptation WAT=2 pour les systèmes (I) et (T)

et avec un seuil d'acceptation WAT=3 pour les systèmes (I+T). D'autre part, les systèmes (I) et (T) ont été re-entrainés sur leurs sous-ensembles respectifs de caractéristiques.

Table 1. Performances de la combinaison linéaire. Les systèmes (I+T) et (I+T+) utilisent l'ensemble des caractéristiques f1-f5. Les systèmes (I) et (T) une partie seulement.

système	rappel	précision
(I)	42.05 %	50 %
(T)	53.35 %	34.87 %
(I+T)	57.98 %	45.18 %
(I+T+)	64.66	45.86 %

Le système (I) est le meilleur du point de vue de la précision car il ne propose que les noms dans la zone correspondant au bloc expéditeur. Par contre, s'il ne trouve pas de bloc expéditeur (absence de champ, erreurs OCR), il ne peut pas compenser par l'analyse textuelle ce qui fait qu'il a le moins bon rappel. Le système (T) a un meilleur rappel car il peut trouver des noms n'importe où dans le document, même ceux qui ne correspondent pas au nom de l'expéditeur et sa précision est donc plus faible.

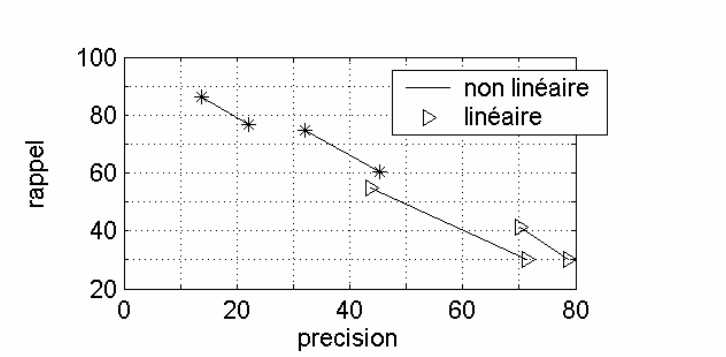


Figure 6. *Points de fonctionnement obtenus pour la combinaison linéaire et la combinaison non linéaire des caractéristiques g1-g5 sur les documents imprimés.*

Le système (I+T) bénéficie des deux analyses et son taux de rappel est encore meilleur avec une précision un peu plus faible que celle du système (I). Le système (I-T+) a le meilleur rappel pour une précision équivalente au système (I-T) car la tâche est plus aisée si on ne traite que les documents imprimés.

Nous avons aussi comparé la combinaison linéaire et la combinaison non linéaire pour l'ensemble des caractéristiques g1-g5. Les différents points de fonctionnement (Fig. 6) sont obtenus en faisant varier le seuil d'acceptation de la combinaison linéaire de WAT=2 à WAT=3 (chaque ligne correspond à une valeur fixée de WAT) et la valeur de N de N=1 à N=3 qui correspond à ne garder que les N expressions de score les plus élevés. Pour la combinaison non linéaire, les scores obtenus en sortie du réseau sont plus faibles et le seuil d'acceptation varie de 0.7 à 0.9.

Le fonctionnement du système linéaire privilégie la précision tandis que le système non linéaire privilégie le rappel. Dans une zone intermédiaire (précision autour de 45%, rappel entre 50 et 60%) les deux systèmes fonctionnent de manière équivalente.

B.4 Extension aux articles de revue

Cette approche a été adaptée à l'extraction de noms dans les articles de revue scannés de la base UW (University of Washington) [Askilrud et Haralick 1993]. Ce travail a été réalisé dans le cadre du stage d'Aliette de Bodard [de Bodard 2005]. Il s'agit d'extraire les noms d'auteurs dans des documents provenant de revues scientifiques de domaines très variés (physique, médecine, informatique, sociologie,..). La notion de Paire Logique définie pour les télécopies par les blocs contenant les champs principaux est remplacée ici par les blocs contenant des éléments d'adresse. Le dictionnaire des mots courants est remplacé par son équivalent anglophone. La règle de combinaison des caractéristiques est la combinaison non linéaire de l'ensemble g1-g5. Le score global d'une expression, obtenue par regroupement des mots proches d'un mot de score élevé, est la valeur maximale des scores des mots l'expression. Car les expressions extraites des articles contiennent plus de mots que celles extraites des télécopies.

Les résultats sont meilleurs pour les pages d'articles que pour les facsimile : 92.22 % de rappel pour 41.33% de précision. La structure d'un article est effectivement plus formalisée que celle d'un fax mais les articles contiennent beaucoup plus de mots et beaucoup de noms propres (citations, section référence,...). Par contre, les mots écrits en lettres capitales sont principalement les titres dans les articles tandis que les facsimile en contiennent naturellement plus (champs, intitulés et contenus).

Comme précédemment, nous n'avons pas inclus de modèle a priori, car les articles sont d'origine très variée. Certains articles débutent même parfois en milieu ou bas de page, à la suite d'un article précédent. Les erreurs sont essentiellement dues aux noms propres qui font aussi partie du dictionnaire des noms courants (noms de famille asiatiques tels que 'Pang', 'Fang', noms de lieux tels que 'Austin'), aux titres écrit en lettres capitales et qui contiennent des mots techniques (qui ne font pas partie du dictionnaire).

B.5 Conclusion

Nous avons abordé dans cette partie plusieurs aspects fondamentaux d'un système d'analyse de documents: bases de données, structuration physique et logique. L'objectif est d'extraire une information clé comme le nom de l'expéditeur sans utiliser de modèle a priori de documents. Pour cela, nous nous sommes appuyés sur deux analyses complémentaires, de type image et de type textuel. L'analyse de type image est basée sur des conventions générales de présentation (existence de paires de champs, alignement des champs). L'analyse textuelle est une analyse de bas niveau qui extrait sur chaque mot des caractéristiques locales liées à la typographie et à la présence du mot dans des dictionnaires. Ces deux analyses ont été appliquées aux documents dégradés mixtes de type télécopie qui contiennent de l'écriture manuscrite et imprimée et qui produisent des transcriptions OCR bruitées. Nous avons comparé la fusion de ces deux analyses par des classifieurs linéaires et non linéaires de type neuronaux.

Cette approche assez générale a été facilement transposée à l'extraction des noms d'auteurs dans les images d'articles de revues issues elles aussi de documents bruités (numérisation de photocopies). Les résulats ont été présentés sur la base d'évaluation UW (University of Washington).

L'aspect reconnaissance des caractères a été abordé de manière transparente lors de ces travaux, car celle ci a été réalisée par un OCR du commerce. La reconnaissance des caractères et des mots manuscrits fait l'objet de la partie C.

Bibliographie (sujet B)

publications issues du sujet B

Cuenca B. (1998), *Extraction automatique de l'écriture manuscrite et accès au contenu de télécopies*, Mémoire de DEA IARFA, Paris VI, avril-septembre1998.

Likforman-Sulem L., Cuenca B. (1999), Facsimile processing for a messaging server », DEXA'99, tenth int'l workshop on database and expert systems applications, Florence (Italie), p. 539–543.

Likforman-Sulem L. (2001), Name block location in facsimile images using spatial/visual cues, *ICDAR'01, sixth int'l conference on document analysis and recognition*, Seattle (USA), pp. 680-684.

Vaillant P., Likforman-Sulem L., Yvon F. (2002), Exploitation d'informations spatiales et textuelles en analyse de documents : le cas des télécopies, *Conférence CFD-CIFED'02*, Hammamet, Tunisie

Vaillant P. (2002), *Outils linguistiques pour l'analyse des télécopies*, rapport de recherche, septembre 2002, ENST.

Godeau J., Likforman-Sulem L., Sigelle M (2002). ENST fax character database, CD-ROM set, 2002.

Likforman-Sulem L., Vaillant P., Yvon F. (2003), Proper names extraction from fax images using textual and image features, Actes de *ICDAR'03*, Edimbourg, pp. 545-549.

Azzabou N. (2003), *Extraction des noms propres dans les images de télécopies par réseau neuronal*, Mémoire de stage post-mastère, décembre 2002- décembre 2003.

Azzabou N., Likforman-Sulem L. (2004), Neural Network-based proper name extraction in fax images, Icpr'04, Cambridge.

Likforman-Sulem L., Chollet G., Vaillant P., Azzabou N., Blouet R., Chollet G., Renouard S., Mostefa D. (2004), *Reconnaissance de noms propres et vérification d'identité dans un système de messagerie*, convention Minefi no 01.2.93.0268, Rapport final, 100 pages.

de Bodard de la Jacopière A. (2005), *Reconnaissance des noms propres dans les documents dégradés*, rapport de stage d'option scientifique, option mathématiques appliquées, Ecole Polytechnique, juillet 2005

de Bodard de la Jacopière A., Likforman-Sulem L. (2006), Author Name recognition in degraded journal images, *IS&T/SPIE Conference on Document Recognition and Retrieval XIII*, San Jose, Janvier 2006.

Likforman-Sulem L., Vaillant P., de Bodard de la Jacopière A. (2006), Automatic Name Extraction from Degraded Document Images, *Pattern Analysis and Applications*, DOI 10.1007/s10044-006-0038-6, Springer, 2006.

Autres Références

Alam H., Hartono R., Sugono Y., Tran T. (2000), FaxAssist : an automatic routing of unconstrained fax to email location, *IS&T/SPIE conference on document recognition and retrieval*, San José (USA).

Askilrud E.S., Haralick R.M. (1993), *A Quick Guide to UW English Document Image Database I*, Department of Electrical Engineering, Department of Computer Science/Software Engineering, University of Washington.

Baumann S., Ali M., Dengel A., Jäger T., Malburg M., Weigel A., Wenzel C. (1997), Message extraction from printed documents: a complete solution, *ICDAR'97*, Ulm (Allemagne).

Bikel D. *et al.* (1997), Nymble a high-performance learning Name-finder, *5th Conf. on Applied Natural. Language Processing*, Washington.

Cesarini F., M. Gori, S. Marinai, G. Soda (1998), INFORMys : a flexible invoice-like form reader system, *IEEE PAMI*, Vol 20, no 7, pp. 730-745

Collins M., Singer Y. (1999), Unsupervised models for named entity recognition, *Joint SIGDAT conference on empirical methods in natural language processing and very large corpora*, University of Maryland (USA).

Daciuk J., Mihov S., Watson B., Watson R. (2000), Incremental construction of minimal acyclic finite-state automata, *Computational Linguistics*, vol. 26, n°1.

Fan K-C., Wang L-S, Tu Y-T. (1998), Classification of machine printed and handwritten texts using character block layout variance, *Pattern recognition*, Vol. 31, no 9, pp. 1275-1284.

Faure C. (1999), Preattentive Reading and selective Attention for Document Image Analysis, *Proc. of ICDAR 99*, Bengalore, pp. 577-580

Hadjar K., Rigamonti M., Lalanne D., Ingold R., Xed: a new tool for eXtracting hidden structures from Electronic Documents, *Proc. of DIAL*, 2004.

Holstege M., Inn Y., Tokuda L. (1991), Visual Parsing: An aid to text understanding, *Proc. of RIAO'91*, Barcelone, pp. 175-193.

Klink S., Kieninger T. (2001), Rule-based document structure understanding with a fuzzy combination of layout and textual features, *IJDAR*, Vol 4, pp. 18-26.

Liang J., Doermann D. (2002), Logical labeling of document images using layout graph matching with adaptive learning, *Document Analysis Systems*, D. Lopresti, J. Hu and R. Kashi (eds), pp. 224-235.

Liang J., Doermann D. (2003), Content features for logical document labeling, *IST/SPIE Document Recognition and Retrieval X*, Santa Clara, pp. 189-196.

Lii J., Srihari S. N.(1995), Location of name and address on fax cover pages, *Proc. of ICDAR'95*, Montréal (Canada).

MUC (1998), *Proceedings of the Message Understanding Conference (MUC-4-7)*, Morgan Kaufman, San Mateo, USA, 1992-98.

TAL (1998), Traitement automatique des noms propres, numéro spécial de la revue TAL – *Traitement Automatique des Langues*, vol. 41, n°3, Paris, Hermès.

Vel'skii, G.M. A., Landis E.M. (1962), An algorithm for the organization of information, *Soviet Mathematics Doklady*, vol 3, pp. 1259-1263..

Viola P, Rinker J, Law M (2004) Automatic fax routing, *Proc. of DAS 2004*, pp 484–495.

Xerox (1994), ScanWorX API Release Notes, Xerox Imaging Systems.

C. APPROCHES STOCHASTIQUES POUR LA RECONNAISSANCE DE CARACTERES ET DE MOTS

C.1 Introduction

Les travaux précédents concernaient la structuration et l'interprétation d'images de documents. Nous avions alors utilisé un système OCR (Optical Character Recognition) pour produire la transcription du texte des images de télécopies et des articles de revues. Si pour les manuscrits d'auteur, la reconnaissance des caractères et des mots est un objectif à long terme, la reconnaissance de caractères imprimés et manuscrits est déjà intégrée dans certains systèmes industriels (OCR, reconnaissance de formulaires, de codes postaux, de montants de chèques...). Ces systèmes utilisent largement les modèles de Markov cachés (HMMs). Leur application à la reconnaissance des caractères a fait l'objet de deux stages de DEA [Recht, 1999; Godeau, 2002] et ainsi que de la thèse de K. Hallouli [Hallouli 2004] [Hallouli et al. 02]. La reconnaissance de mots cursifs par HMMs a fait l'objet de la thèse de R. El-Hajj [El-Hajj 2007] et d'une coopération avec l'Université de Balamand (Prof. C. Mokbel). Un système très performant a été développé pour la reconnaissance de mots arabes : ce système a gagné la compétition ICDAR 05 (Seoul).

Les Réseaux Bayésiens Dynamiques (DBNs) sont une extension des HMMs. Ils sont notamment utilisés pour le suivi de visages [Wang et al. 2004] et en traitement de la parole [Zweig, 1998; Daoudi et al. 2000]. Pour améliorer la reconnaissance en environnement bruité (téléphone, bruit ambiant), les auteurs de [Daoudi et al. 2000] font l'hypothèse que le bruit n'affecte pas les sous bandes fréquentielles avec la même intensité. Le formalisme des réseaux Bayésiens leur permet de décomposer le signal de parole en différents flux d'observations, chaque flux agissant dans des sous-bandes fréquentielles différentes.
Dans le domaine de l'écrit et du document, les travaux existants concernent l'authentification de signatures [Xiao et Leedham2002], la reconnaissance de caractères on-line [Cho et Kim 2004][Sicard et al 2007], l'analyse de la structure de documents [Souafi 2002][Souafi et al. 2002]. Les réseaux Bayésiens permettent de modéliser les dépendances statistiques entre traits (positions), ou les dépendances entre blocs fonctionnels.

Nous avons proposé dans [Hallouli et al. 2003][Likforman-Sulem et Sigelle 2007a] [Likforman-Sulem et Sigelle 2008] l'utilisation des réseaux Bayésiens pour la reconnaissance de caractères. La modélisation des caractères par Réseaux Bayésiens Dynamiques permet d'observer conjointement deux flux d'observations : celui des colonnes de pixels, et celui des lignes de pixels. Nous obtenons ainsi une modélisation qui couple deux modèles de Markov et ne fait plus l'hypothèse d'indépendance des lignes ou des colonnes entre elles. ce travail a té initié lors de la thèse de Doctorat de K. Hallouli [Hallouli, 2004]. Nous avons poursuivi ces travaux dans [Likforman-Sulem et Sigelle 2007b] en démontrant la robustesse des Réseaux Bayésiens

Dynamiques pour la reconnaissance des caractères dégradés par des coupures (caractères de documents anciens par exemple). Les DBNs sont plus robustes que les SVMs pour ce type de dégradation du fait qu'au moins un des flux d'observation est non dégradé à un instant donné et du fait du caractère auto-régressif (donc prédictif) de certains modèles couplés.

C.2 Modèles de Markov Cachés

Les approches stochastiques, telles que les modèles de Markov cachés (HMMs), sont largement utilisées pour la reconnaissance de la parole et de l'écrit pour leur capacité à s'adapter aux distorsions élastiques temporelles ou spatiales [Rabiner 1989][Elms et al. 1998] [Arica et Yarman-Vural 2004][Vinciarelli et al. 2004][Hallouli 2002]. Cependant ces modèles sont mono-dimensionnels. Une adaptation doit donc être réalisée pour les images, par nature bi-dimensionnelles : celles ci sont converties en séquences 1D d'observations le long d'une direction. Une séquence admissible d'observations est par exemple la suite des colonnes de pixels de l'image balayée de gauche à droite. D'autres séquences peuvent être obtenues en extrayant des vecteurs de caractéristiques de plus haut niveau dans des fenêtres glissantes, ou dans des secteurs angulaires [Anigbogu et Belaid 1995]. La modélisation HMM permet de regrouper en états les observations voisines dans la séquence, et similaires. Le formalisme HMM inclut des algorithmes puissants d'apprentissage (Baum-Welch) et de décodage (décodage de Viterbi) [Rabiner 89].

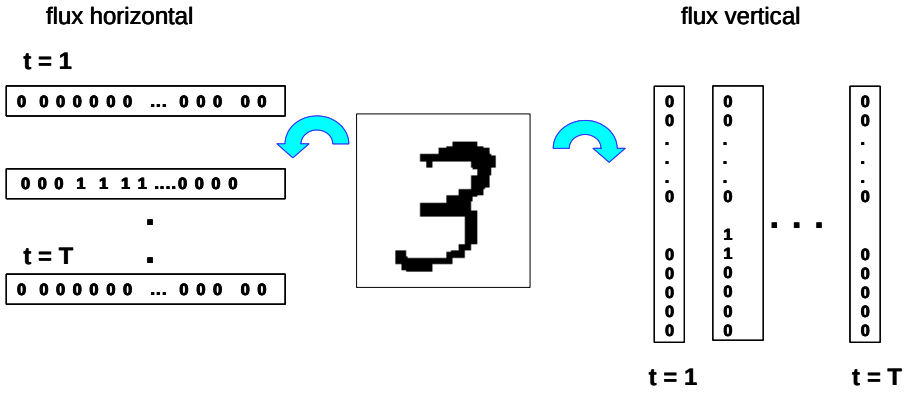


Figure 1. *Séquences d'observations lignes/colonnes extraites sur les images*

C.2.1 Application à la reconnaissance de caractères

Une topologie classique des modèles HMMs est le modèle gauche-droite (dit aussi de Bakis): le degré de transition avant est de deux états (cela permet de sauter un état si une observation n'est pas présente) et sans degré arrière. Nous avons cherché lors du stage de DEA de J. Godeau [Godeau, 2002], à exploiter la symétrie des caractères en autorisant les transitions avant et arrière entre états, ces degrés de transition étant des paramètres du modèle adaptés à chaque classe de caractère. Les paramètres 'degrés avant' et 'degré arrière' sont déterminés sur une base de

validation. Les résultats obtenus sur la base de caractères ENST-FAX-CHAR ont montré l'amélioration obtenue avec ces modèles par rapport aux modèles n'autorisant que les degrés avant. Les performances obtenues (84 % de bonne reconnaissance) sont cependant équivalentes à celles obtenues par un réseau neuronal de type MLP à une couche cachée.

Une autre approche (thèse de K. Hallouli [Hallouli 2002]) que nous avons menée conjointement, consiste à conserver la modélisation gauche-droite mais à associer deux modèles HMMs observant le caractère selon deux flux différents.

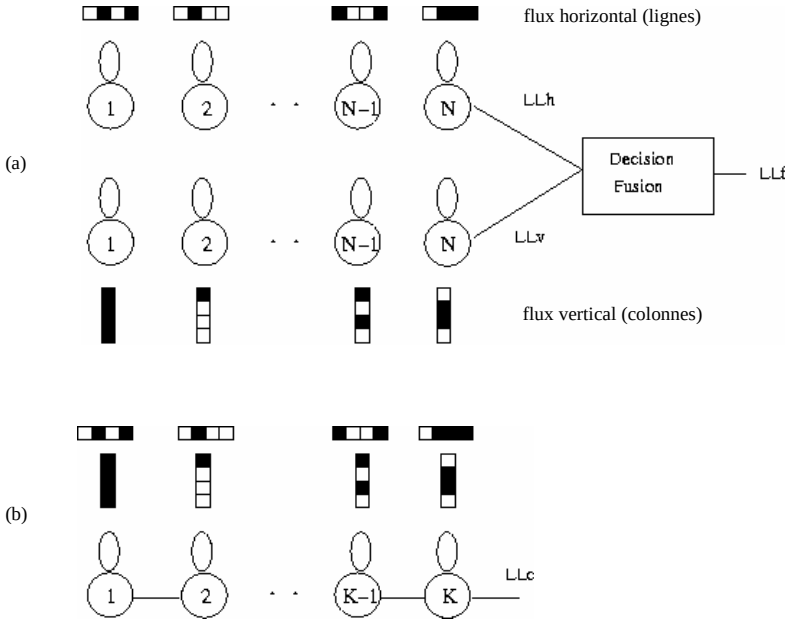


Figure 2. *Modèles HMMs utilisant (a) la fusion de scores: les log-vraisemblances LLh et LLv sont combinées pour obtenir le score LLf (b) la fusion de données: la log-vraisemblance LLc est obtenue par un seul HMM qui fusionne les deux vecteurs d'observation du flux vertical et du flux horizontal.*

Ces HMMs sont nommés *HMM vertical* et *HMM horizontal* et observent les colonnes et les lignes de pixels respectivement. Le HMM vertical prend pour séquence d'entrée les colonnes de pixels (une colonne constitue un vecteur d'observation) balayées de gauche à droite tandis que le HMM horizontal prend pour séquence d'entrée les lignes de pixels balayées de haut en bas (Fig. 1). Le caractère est normalisé dans un carré pour que les séquences horizontales et verticales soient de même longueur.

Plusieurs modèles de fusion sont réalisés: fusion au niveau de la décision, et fusion au niveau des données (Fig. 2). La fusion des scores est obtenue par combinaison linéaire des deux log-vraisemblances en favorisant le flux vertical. Elle suppose que les deux classifieurs, horizontal et vertical, sont indépendants. La fusion de données inclut les deux flux dans un même vecteur d'observation et un même classifieur. Les dépendances entre observations sont modélisées dans les matrices de covariances des lois d'observation (Gaussiennes).

Ces modèles ont été appliqués dans [Hallouli et al. 2002 et 2003] à la reconnaissance des caractères mono-police imprimés dégradés de la base UW-English Document Database [UW, 1993] [Baird, 1982]. Ils ont aussi été appliqués dans [Hallouli 2004] à la reconnaissance des chiffres manuscrits de la base MNIST [LeCun, 1998]. Nous avons adapté la boîte à outils HTK [Young, 2001] développée originellement pour la modélisation et la reconnaissance de la parole, aux images de caractères.

Dans un HMM mono-flux, les observations peuvent être continues ou discrètes. Dans le cas continu, la densité de probabilité d'observation est modélisée par une loi gaussienne. Dans le cas discret, une quantification vectorielle affecte une valeur discrète (qui est un indice correspondant à un prototype du dictionnaire) à chaque observation. Pour ces deux modélisations (discrète ou continue), le nombre d'états possibles pour les variables cachées est déterminant. Une base de validation, indépendante de la base de test, permet de déterminer le nombre d'états optimal pour chaque type de caractère.

Dans le cas discret (étudié lors du stage de DEA [Recht, 1999]), la quantification vectorielle utilise la distance SIHD-Shift Invariant Hamming Distance [Elms et al. 1998] et un algorithme de type k-moyennes pour construire le dictionnaire. La distance SIHD est invariante aux translations mais sensible aux dilatations. La taille du dictionnaire influe sur les performances [Hallouli, 2004] : en faisant varier le dictionnaire de 16 à 120 vecteurs, les performances se stabilisent à partir d'un dictionnaire de taille 80. Cette taille est corrélée à la longueur de la séquence d'observations (ici de longueur 50). Cependant les résultats obtenus par quantification vectorielle (Table 1) sont moins bons que ceux obtenus pour des observations continues (Table 2). Dans tous les cas, discret ou continu, le HMM vertical a de meilleures performances que le HMM horizontal. Ceci est dû au fait que pour les formes des caractères que nous étudions, la présence des traits verticaux est prédominante.

Table 1. Performance pour les HMMs discrets mono-flux en fonction de la taille du dictionnaire utilisé pour la quantification vectorielle.

	taux de reconnaissance (%)		
	majuscules (dict=80)	minuscules (dict=80)	chiffres (dict=16)
HMM vertical	92.4	91.8	78.8
HMM horizontal	89.8	88.8	73.5

Table 2. Performances pour les HMMs continus, mono-flux et fusionnés.

	taux de reconnaissance (%)			
	majuscules	minuscules	majuscules+minuscules	chiffres
HMM vertical	98.8	96.9	91.8	96.6
HMM horizontal	94.4	95.5	87.6	93.6
Fusion de scores	98.9	98.6	93.3	96.9
Fusion de données	99.8	99.8	93.8	97.7

Nos expériences dans le cas continu montrent que le couplage des deux analyses verticales et horizontales par fusion de scores et par fusion de données sont meilleures que chacun des HMMs mono-flux. La fusion de données étant plus performante que celle des scores, ceci nous a

encouragés à développer des modèles utilisant la fusion des données. Cette fusion a été réalisée par les réseaux Bayésiens (cf. Section C.3).

C.2.2 Application à la reconnaissance de mots

Les systèmes de reconnaissance de mots cursifs se divisent en deux grandes catégories : les systèmes globaux qui traitent les images comme un tout et ne cherchent pas à segmenter les mots, et les systèmes analytiques. Les systèmes analytiques se distinguent par le fait que les modèles de mots sont construits à partir des modèles des lettres constituantes. L'avantage de cette approche est que l'on peut modéliser un nouveau mot par concaténation de modèles de lettres. Les systèmes analytiques se divisent aussi en deux catégories : les systèmes à segmentation implicite et les systèmes à segmentation explicite. Ces deux dernières approches sont en réalité bien différentes. Les systèmes à segmentation explicite cherchent à découper les mots en caractères ou en entités plus fines avant la reconnaissance de ces unités [Miled et al. 1997]. La segmentation implicite est utilisée dans la plupart des modèles HMMs. Dans ces systèmes, la segmentation en caractères est fournie conjointement avec la reconnaissance [Bazzi et al. 1999][Choisy 2002][Khorsheed 2003].

L'écriture arabe est particulièrement difficile à segmenter du fait de la présence de nombreuses boucles, hampes et jambages de taille importante [Ben Amara et Bouslama 2003]. L'approche choisie dans la thèse de R. El-Hajj [El-Hajj et al. 2007] et publiée dans [El-Hajj et al. 2005][El-Hajj et al. 2008] est analytique à segmentation implicite. Cette recherche est motivée d'une part pour les applications postales potentielles (tri du courrier) et d'autre part du fait de l'existence d'une grande base de données publique, la base IFN/ENIT [Pechwitz et al. 2002] de noms de villes Tunisiennes.

Le système proposé ne nécessite pas de segmentation en caractères pour l'apprentissage et la reconnaissance. Les modèles de lettres sont appris par apprentissage croisé à partir des séquences d'observation issues des images de mots et de leur transcription. Le décodage par l'algorithme de Viterbi fournit une segmentation implicite conjointement à la reconnaissance. L'apprentissage et le décodage utilisent le moteur HCM [Mokbel et al., 2002]. La reconnaissance est d'autre part dirigée par un dictionnaire. Lors de la reconnaissance, la séquence d'observations est confrontée aux modèles de mots appartenant à un dictionnaire. Le mot reconnu est celui obtenu par maximum de vraisemblance.

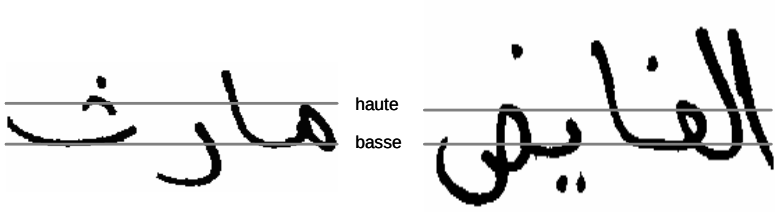


Figure 3. *Lignes de base extraites sur les mots*

C.2.2.1 Système de référence

Le premier système développé, dit système de référence, est décrit dans [El-Hajj et al. 2005]. Ce système utilise une fenêtre glissante qui balaye l'image des mots de droite à gauche. Les deux lignes de base qui encadrent le corps de l'écriture (Fig.3) sont extraites à partir du profil vertical de projection, obtenu par projection des pixels selon l'axe horizontal. Un ensemble de 28 caractéristiques est extrait dans chaque fenêtre, divisée en nc=21 cellules. Ces caractéristiques sont les densités, nombre de transitions N/B entre cellules, la position du centre de gravité de l'écriture, les configurations locales de concavités (Fig. 4). Les fenêtres sont de largeur w=8 et décalées de $\delta=w/2$. Les valeurs optimales des paramètres nc et w sont déterminées sur une base de validation ([El-Hajj et al. 2007]).

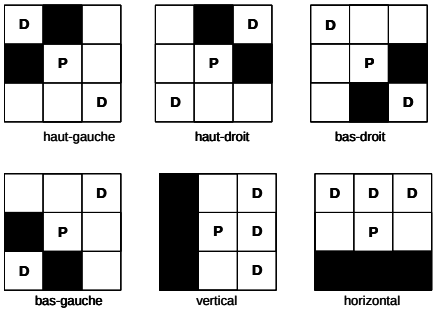


Figure 4. *Analyse des concavités locales : un point P du fond peut être dans une des ces 6 configurations. Un pixel D peut être quelconque (noir ou blanc)*

L'ensemble des caractéristiques du système de référence est séparé en deux groupes suivant que les caractéristiques dépendent ou non de la position des lignes de base.

caractéristiques indépendantes des lignes de base :
f1 : densité des pixels d'écriture
f2 : transitions N/B entre cellules
f3 : caractéristique dérivative (centre de gravité)
f4-f11 : densité des pixels d'écriture dans chaque colonne de la fenêtre
f17-f22 : nombre de pixels fond selon les 6 configurations locales (haut-gauche, haut-droit, bas-droit, bas-gauche, verticale et horizontale)

caractéristiques dépendantes des lignes de base :
f_{12} *:* distance normalisée (en y) du centre de gravité des pixels d'écriture à la ligne de base basse.
f_{13}-f_{14} *:* densité des pixels au-dessus et en-dessous de la ligne de base basse.
f_{15} *:* transitions N/B pour les cellules au-dessus de la ligne de base basse
f_{16} *:* position du centre de gravité: zone centrale, au-dessus de la ligne de base haute, sous la ligne de base basse.
f_{23}-f_{28} *:* nombre de pixels fond dans la zone centrale selon les 6 configurations locales (haut-gauche, haut-droit, bas-droit, bas-gauche, verticale et horizontale)

La modélisation HMM d'un mot consiste à concaténer les modèles de lettres, en ajoutant éventuellement un modèle espace (symbolisé par #) entre deux pseudo-mots. Chaque modèle de caractère consiste en un modèle HMM gauche-droite à 4 états. La densité de probabilité des observations en chaque état suit une loi mélange de 3 Gaussiennes de matrices de covariance diagonales. Ces paramètres (nombre d'états, nombre de lois du mélange,...) sont obtenus sur une base intermédiaire de validation.

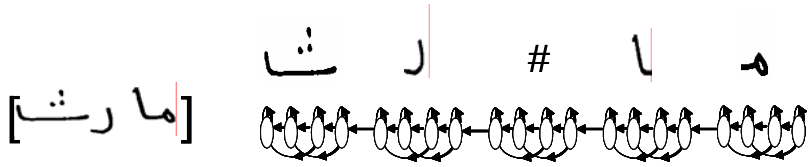


Figure 5. *Modélisation d'un mot obtenue par concaténation des modèles de lettres et du modèle espace (#).*

La base de données IFN/ENIT [Pechwitz et al. 2002] contient 946 noms de villes différents et 26 459 images de mots, ce qui en fait actuellement la plus grande base publique pour l'écriture arabe. Elle est divisée en 4 ensembles nommés a, b, c et d. Le protocole pour tester un système sur cette base est la validation croisée : entraîner sur 3 ensembles et tester sur le quatrième. Le système de référence est évalué sur cette base. Les résultats obtenus sont données dans la Table 3.

Table 3. Performances du système de référence sur la base IFN/ENIT

Apprentissage	**Test**	**Reconnaissance**
b, c, d	a	**83.04 %**
a, c, d	b	**83.10 %**
a, b, d	c	**82.47 %**
a, b, c	d	**83.31 %**

Le système de référence a participé à la compétition ICDAR 05 sur la reconnaissance de mots cursifs arabes. Tous les systèmes en compétition ont été testés par l'IFN (Institut für Nachrichtentechnik, Braunschweig) à partir d'exécutables fournis par les participants. Sur le nouvel ensemble de mots *e* de 3000 images collecté pour cette compétition, le système a obtenu le meilleur taux de reconnaissance : 75.93 %, devançant légèrement le système de l'IFN [Pechwitz et al, 2003].

C.2.2.2 Système combiné

Une technique pour améliorer un système de reconnaissance est de le combiner au niveau de la décision à d'autres systèmes complémentaires (c-à-d ne commettant pas les mêmes erreurs). Les architectures de combinaisons possibles sont les architectures parallèles, et les architectures séquentielles. Dans l'approche parallèle, la plus simple, chaque classifieur émet une ou plusieurs décisions pondérées par un score ou rangées suivant l'ordre décroissant et ces décisions sont combinées par un module de fusion qui les réordonne. Les règles de combinaison sont par exemple la règle somme (des log-vraisemblances ou des scores), la règle produit, la

règle moyenne, la règle du Borda Count, etc...[Kittler et al. 1998] [Duin, 2002][Ho et al. 1994]. Dans l'approche séquentielle, un seul classifieur opère en même temps. Un classifieur de la séquence a soit la tâche de réduire le nombre de classes pour un objet à classer (approche de type réduction de classe), soit la tâche d'estimer la classe ou de rejeter l'objet (approche de type réévaluation). Dans ce dernier cas, le classifieur suivant qui utilise un autre ensemble de caractéristiques et/ou un autre mécanisme de décision, essayera d'estimer la classe de l'objet rejeté à l'étape précédente.

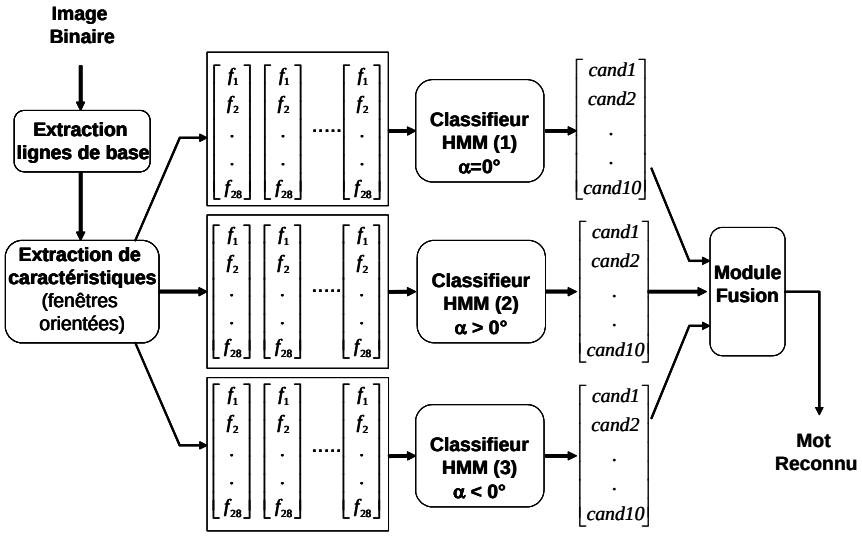


Figure 6. Combinaison du système de référence et des systèmes orientés au niveau décision

Le système combiné présenté dans [El-Hajj et al. 2006] consiste à associer le système de référence avec deux autres classifieurs de même type (HMMs, à fenêtre glissante) au niveau de la décision (Fig. 6). Ces deux classifieurs supplémentaires utilisent des fenêtres orientées avec un angle α par rapport à la verticale, l'une vers la droite (α >0) et l'autre vers la gauche (α >0) (Fig. 7). Mais le nombre et type de caractéristiques restent inchangé. Le classifieur combiné peut ainsi mieux prendre en compte l'inclinaison des traits, ou les points diacritiques décalés. Plusieurs techniques de combinaison ont été étudiées dans [El-Hajj 2007]: somme des log-vraisemblances, vote à la pluralité, combinaison par sélection de classifieur. Ce dernier type de combinaison utilise un réseau de neurones qui désigne le classifieur estimant au mieux la classe d'un mot test à partir des dix meilleures log-vraisemblances produites par les trois classifieurs.

La figure 8 donne les performances des HMMs individuels pour différentes inclinaisons de fenêtres allant de -45° à +45°. Les fenêtres de faible inclinaison (0°, 5°, 10°, 15°) à gauche ou à droite donnent les meilleurs résultats pour les systèmes individuels. La figure 9 donne le résultat du système combiné (somme des log-vraisemblances) pour des angles de fenêtres symétriques, et quand on introduit une dissymétrie dans l'orientation de la fenêtre inclinée à gauche avec celle inclinée à droite.

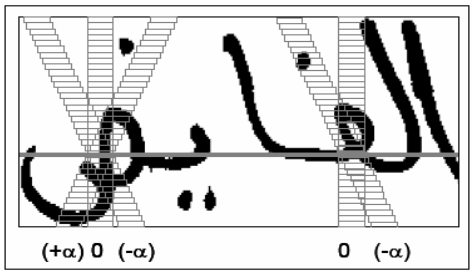


Figure 7. Croisement des fenêtres inclinées sur la ligne d'écriture (ici symétriquement)

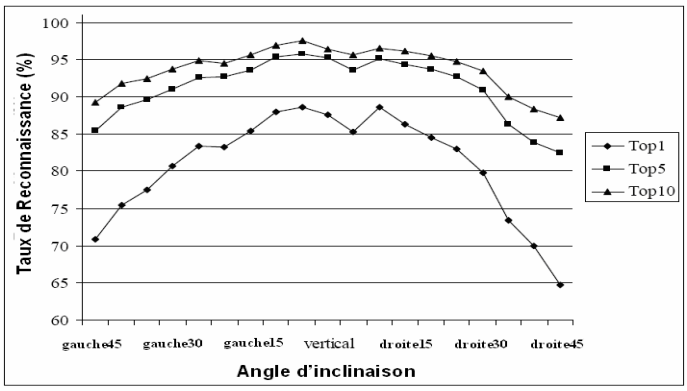


Figure 8. Performance des systèmes HMMs individuels (vertical, orienté à gauche, orienté à droite) en fonction de l'angle d'inclinaison

Les performances sont meilleures que pour n'importe quel HMM individuel et le système combiné est opérationnel pour une large gamme d'angles. Par exemple, dans le cas où les angles sont symétriques (courbe Top1), les performances ont bonnes pour les angles entre 10° à 30°. Pour la deuxième courbe (gauche+5) qui correspond à une combinaison d'angles dissymétriques, les performances sont bonnes pour les systèmes utilisant les triplets d'angles (-20°, 0°, 15°), (-25°, 0°,20°), (-30°, 0°, 25°), (-35°, 0°, 30°), les angles étant donnés dans l'ordre suivant: orientation gauche, orientation verticale et orientation droite.

Le système combiné a également participé à la deuxième compétition de reconnaissance de mots lors d'ICDAR 07 (Brésil). De nouveaux ensembles de données *f* et *s* ont été utilisés par les organisateurs. Le système combiné a obtenu la deuxième place sur l'ensemble de données *s* après le système présenté par Siemens-Allemagne.

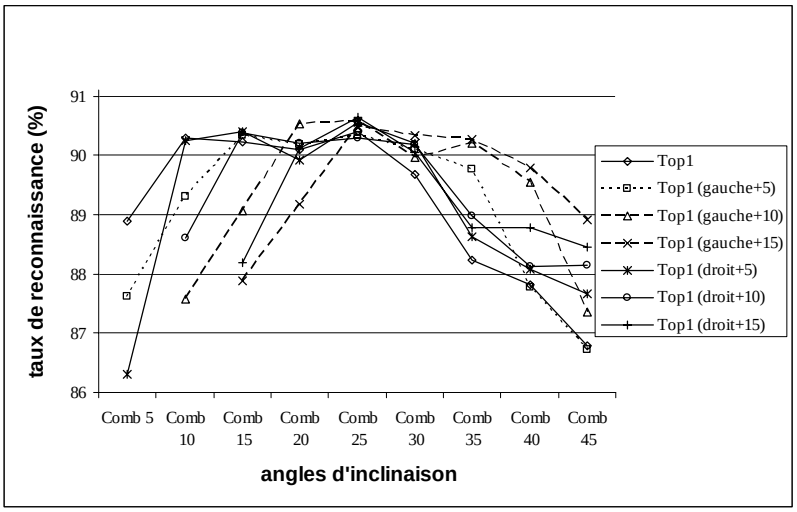


Figure 9. Performance du système combiné (somme des log-vraisemblances) en fonction de l'angle d'inclinaison et de la dissymétrie dans les orientations des fenêtres.

C.2.3 Conclusion

Les HMMs offrent un formalisme adapté à la reconnaissance de l'écrit. Nous avons proposé des modélisations au niveau caractère et au niveau mot. Les classifieurs HMM obtiennent de bonnes performances sur les tâches de reconnaissance de caractères imprimés dégradés et sur la reconnaissance des mots cursifs arabes. Ces performances sont encore améliorées si on combine plusieurs classifieurs HMMs au niveau décision ou au niveau des données.

Les modèles HMMs font l'hypothèse d'indépendance des observations conditionnellement aux états. Pour étendre le cadre des HMMs et modéliser les dépendances entre observations, nous proposons en partie C des modèles basés sur les réseaux Bayésiens Dynamiques.

C.3 Réseaux Bayésiens

C.3.1 Introduction

Les réseaux Bayésiens appartiennent à la classe des modèles graphiques, unissant la théorie des probabilités avec la théorie des graphes. Les premiers travaux sont dus à Pearl [Pearl, 1988] avec l'objectif initial d'intégrer la notion d'incertitude dans le raisonnement pour les systèmes experts : les premières applications l'ont été dans le domaine du diagnostic médical [Becker et al., 1996], puis dans les domaines de la sécurité (détection des fraudes, fiabilité), de l'ingénierie des connaissances et de la défense (système Hugin de contrôle de sous marins).

Les modèles de Markov font l'hypothèse que les observations sont indépendantes conditionnellement aux états cachés, ce qui n'est pas toujours réaliste pour les images. Des extensions des HMMs permettant de mieux prendre en compte l'aspect bi-dimensionnel des images ont ainsi été proposées avec les modèles pseudo-2D (ou planar HMMs) [Gilloux 1994]. Plus récemment, des modèles 2D à base de champs de Markov ont été développés [Park et Lee, 1998 ; Belaid et Saon 1997 ; Geoffrois et al. 2004]. Les modèles probabilistes s'appuyant sur les réseaux Bayésiens sont apparus dans le domaine de la reconnaissance de l'écriture en-ligne [Cho et Kim 2003][Sicard et al. 2006], l'analyse de documents [Souafi et al. 2002] et l'authentification de signatures [Xiao et Leedham, 2002]. Dans [Cho et Kim 2003], les points caractéristiques des traits (points terminaux, points milieux) sont les variables d'un réseau Bayésien statique. Les dépendances de positions entre traits d'un caractère sont modélisées par une relation graphique entre points terminaux adjacents. La structure logique des tables des matières d'articles est analysée dans [Souafi et al. 2002] par des réseaux Bayésiens qui utilisent les attributs typographiques des blocs physiques et de leurs voisins. Ces attributs sont des variables observées dont les dépendances sont exprimées par la structure du réseau, appris en utilisant un algorithme génétique.

C.3.2 Formalisme

Un réseau Bayésien (ou BN - *Bayesian Network*) statique est un graphe acyclique orienté dont les noeuds sont des variables aléatoires qui ont des valeurs discrètes ou continues. Formellement un réseau Bayésien statique est défini par la donnée de deux éléments :

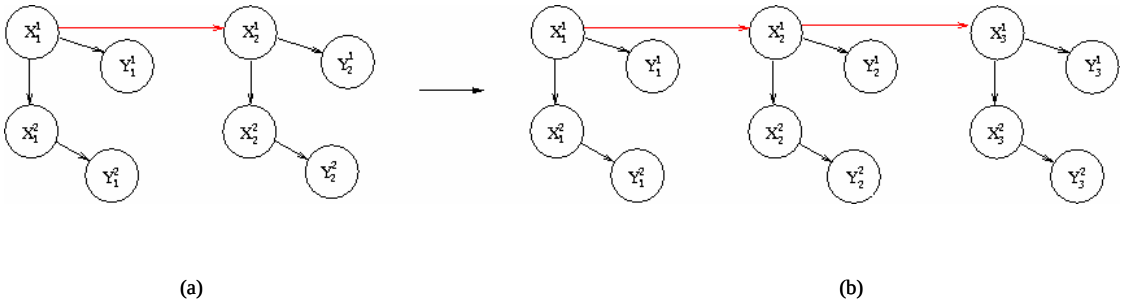
- le graphe G : c'est un graphe causal, orienté, acyclique dont les noeuds sont des variables d'intérêt du domaine, les arcs des relations de dépendance entre ces variables. L'ensemble des noeuds et des arcs forme la structure graphique du réseau Bayésien.

- la paramétrisation θ : c'est l'ensemble des distributions conditionnelles en chaque nœud. Les distributions conditionnelles sont soit des CPTs (*Conditional Probability Tables*) si le nœud et ses parents représentent des variables discrètes, soit des CPDs (*Conditional Probability Distributions)* si le nœud représente un variable continue. Les distributions de type continu suivent généralement une loi Gaussienne.

Soient un ensemble de variables aléatoires $X=(X_1, X_2, \ldots X_N)$ et $P(X)$ sa distribution jointe de probabilité. Celle ci peut s'écrire sous forme factorisée :

$$P(X_1, X_2, X_3, ..., X_N) = \prod_{1 \le i \le N} P(X_i | Pa(X_i))$$

où Pa(Xi) est l'ensemble des causes (parents) de X_i dans le graphe G. La sémantique des indépendances conditionnelles d'un réseau Bayésien implique qu'une variable est indépendante de toutes les autres variables du réseau, connaissant ses parents, à l'exception de ses descendants. Les Réseaux Bayésiens Dynamiques (ou DBNs-*Dynamic Bayesian Networks*) [Murphy, 2002] sont une extension des réseaux Bayésiens statiques aux processus évoluant au cours du temps, et pour des instants discrets $t >= 1$.



(a) (b)

Figure 10. (a). Réseau Bayésien Dynamique défini sur deux pas de temps. (b) Réseau déroulé sur trois pas de temps pour s'adapter aux séquences d'observations Y ($\{Y^1\}$ et $\{Y^2\}$) de longueur trois.

Les modèles de réseaux que nous présentons plus loin (cf. Section C.3.3.2) incluent des variables cachées notées X et des variables observées notées Y. Les variables observées sont issues de deux flux d'observations de même longueur T (cf. Section C.2.1): le flux vertical $\{Y_t^1\}_{1 \le t \le T}$ des colonnes et le flux horizontal $\{Y_t^2\}_{1 \le t \le T}$ des lignes. A chaque instant t, deux variables sont observées, l'une provenant du flux colonnes, l'autre du flux lignes.

On peut représenter un DBN par les deux premiers pas de temps si on suppose que le processus est Markovien du premier ordre et stationnaire : les parents d'une variable donnée appartiennent au même pas de temps ou au pas de temps précédent, et les paramètres sont liés (les probabilités conditionnelles associées sont indépendantes de l'instant t). La figure 10 représente un DBN sur deux pas de temps et ce DBN est déroulé sur trois pas de temps pour s'adapter à la séquence d'observations de longueur trois. Un HMM est un cas particulier de DBN, qui inclut une variable d'état et une variable d'observation à chaque pas de temps.

Des algorithmes généraux d'apprentissage et de décodage existent pour les réseaux Bayésiens. L'apprentissage consiste à estimer les tables de probabilité CPTs et les distributions conditionnelles CPDs. L'apprentissage est effectué classe par classe à partir de séquences d'observations. Dans le cas où le réseau ne contient pas de variable cachée, les tables sont estimées par comptage (fréquences de chaque configuration d'une variable et de ses parents).

S'il existe des variables cachées, comme dans nos modèles (variables d'états), on utilise classiquement pour ces estimations l'algorithme EM ou un algorithme de descente de gradient. Le décodage consiste à calculer la vraisemblance des séquences d'observation et la meilleure séquence d'états suivant un modèle. Le décodage (et aussi l'apprentissage) utilise le mécanisme d'inférence [Pearl 1988] qui consiste à propager les informations (connaissances sur les variables observées) dans le réseau et à en déduire l'état des nœuds cachés.

L'inférence s'appuie sur la construction d'un arbre de jonction issu du graphe original *G*. Une fois l'arbre de jonction construit, l'algorithme d'inférence est une généralisation de l'algorithme forward-backward utilisé pour les HMMs [Jensen 1996 ; Lauritzen et Spiegelhalter 1988]. Il utilise notamment des variables π et λ jouant un rôle similaire aux variables α et β des HMMs.

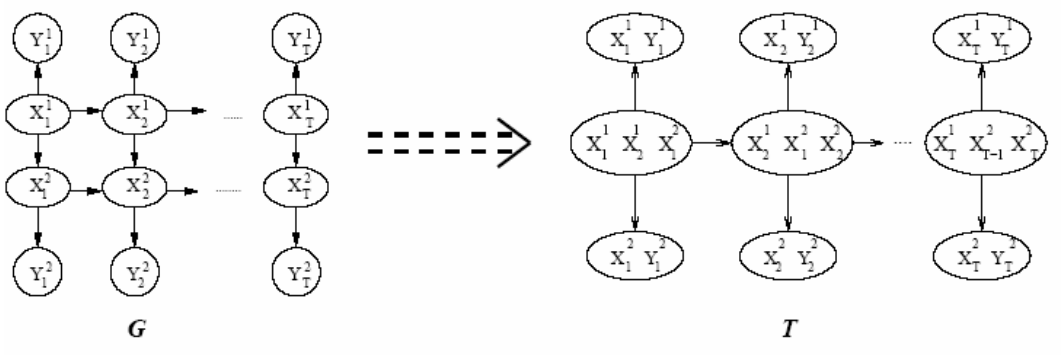


Figure 11. *Réseau Bayésien et Arbre de Jonction correspondant*

L'arbre de jonction est construit en trois étapes :
- moralisation (ajout d'arcs entre nœuds parents) et élimination des directions d'arcs.
- triangularisation : ajout d'arcs pour éliminer les cliques d'ordre 4 ou plus.
- regroupement des nœuds en cliques et construction de l'arbre de jonction avec les propriétés suivantes : si une variable est dans deux cliques distinctes *C1* et *C2*, toutes les cliques sur le chemin dans l'arbre entre *C1* et *C2* contiennent cette variable; il existe au moins une clique dans l'arbre qui contient à la fois un nœud et tous ses parents.

La figure 11 est un exemple d'arbre de jonction *T* issu du graphe original *G*. L'inférence a lieu ensuite dans un arbre, l'arbre de jonction, au lieu du graphe initial.

C.3.3 Modélisation de caractères par Réseaux Bayésiens Dynamiques

L'utilisation de deux flux d'observations pour reconnaître des caractères est énoncée dans [Elms et al. 1998] et [Wang et al. 2002]. Dans ces travaux, la classification des caractères s'effectue en couplant deux classifieurs, l'un examinant les lignes de pixels, l'autre les colonnes. Nos travaux précédents sur les caractères imprimés [Hallouli 2003 et 2004] montrent cependant qu'un seul HMM opérant sur les deux flux lignes et colonnes, est meilleur que la combinaison au niveau décision de deux HMMs opérant sur chacun des flux.
D'autre part l'idée de coupler en un seul classifieur deux modèles HMMs a été proposée initialement par [Brand 1997] pour modéliser les processus qui interagissent (processus ayant une même origine mais enregistrés par plusieurs capteurs : signal vocal et mouvement des lèvres par exemple): dans les architectures de type 'coupled HMMs', un état caché du pas de temps *t* est lié à tous les autres états cachés du pas de temps *t+1*. Les modèles obtenus sont symétriques.

Nos modèles consistent aussi à coupler deux modèles HMM en un seul classifieur et de les appliquer à la reconnaissance de caractères. Ces modèles sont différents d'une fusion de classifieurs au niveau décision. Ils sont aussi différents d'un modèle de fusion où les deux flux seraient regroupés dans un seul vecteur d'observations : la prise en compte des interactions entre les deux flux nécessiterait de très grandes matrices de covariance.

Nous proposons plusieurs types de couplage qui modélisent les interactions entre parties distantes de l'image. L'architecture d'un modèle couplé consiste en deux modèles HMMs dont les états associés à chaque flux sont reliés par des arcs. Les modèles obtenus ne sont pas forcément symétriques et les dépendances entre les observations des différents flux sont prises en compte par les matrices de transitions entre états. Si l'on considère que les états représentent des configurations similaires d'observations, ces dépendances expriment les probabilités d'avoir des configurations jointes d'observations suivant les deux flux : par exemple une colonne de pixels noirs avec une ligne de pixels noirs.

La recherche de la meilleure structure de réseau (emplacement des liens et orientation) à partir des données est l'objet de l'*apprentissage structurel* [Heckerman, 1999]. Cette recherche automatique est délicate quand le réseau contient beaucoup de variables cachées. Notre stratégie consiste à fixer la heuristiquement la structure et à l'évaluer. Pour les modèles couplés, les structures de réseaux sont basées sur les critères suivants :

– le modèle doit avoir un nombre raisonnable de paramètres pour que la complexité des calculs reste abordable.
– aucune variable continue ne doit avoir de fils discret afin de pouvoir appliquer un algorithme d'inférence exacte.
– des liens doivent exister entre les variables cachées.

C.3.3.1 Modèles mono-flux

Les deux premiers modèles, appelés le *HMM vertical* et le *HMM horizontal*, reproduisent deux modèles HMMs mono-flux simples. Le HMM vertical a pour séquence d'entrée les colonnes de pixels du caractère, et le HMM horizontal observe les lignes de pixels. Les modèles HMMs sont des cas particuliers de RBDs. Une variable (ou vecteur) observée a pour parent une seule variable d'état discrète. Les observations sont modélisées par une distribution gaussienne. Nous avons utilisé des matrices de transition entre états gauche-droite pour ces modèles mono-flux.

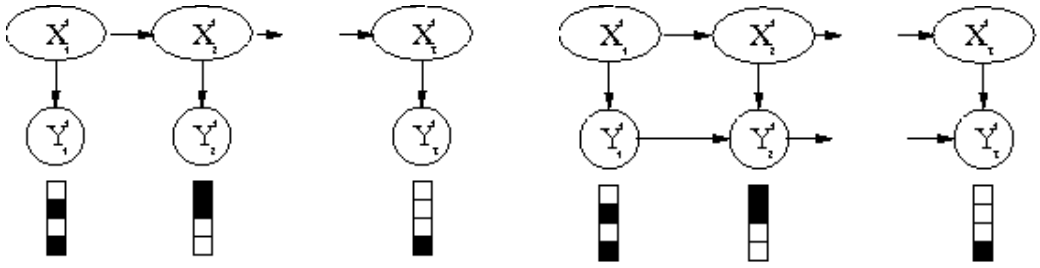


Figure 12. *Modèle mono-flux (HMM-vertical) et modèle mono-flux auto-régressif (AR-vertical)*

Les modèles deux suivants, le *AR-vertical* et le *AR-horizontal*, sont des modèles auto-régressifs mono-flux qui lient les observations entre elles entre deux pas de temps. Les modèles HMM auto-régressifs proposés appartiennent à la catégorie de modèles auto-régressifs de type *switching Markov models* d'ordre 1 [Hamilton 98]. La moyenne associée à un état k est déplacée en fonction de l'observation précédente et de la matrice de régression associée à cet état.
Le modèle auto-régressif vertical (AR-vertical) observe de même les colonnes de pixels et le AR-horizontal les lignes de pixels. Dans la structure du réseau, un lien est ajouté entre la variable d'observation au temps *t* et celle au temps *t+1*. Ceci permet de modéliser le fait que les observations ne sont pas indépendantes entre elles (les modèles HMMs supposent au contraire l'indépendance conditionnelle des observations).
Les paramètres d'un modèle auto-régressif associé à un flux *i* (i=1 ou 2, c-à-d horizontal ou vertical), et pour un nombre d'états *Q* du modèle sont: la distribution initiale Π des états associée au nœud caché initial X_1, la matrice *A* de transition entre états et la distribution des observations (CPD) Y_t:

$$\Pi_1^i = P(X_1^i = k) \quad \forall k, j \in [1,Q]$$
$$A_{j,k}^i = P(X_t^i = k \mid X_{t-1}^i = j) \quad \forall k, j \in [1,Q]$$
$$P(Y_t^i = y_t^i \mid X_t^i = k) = N(y_t^i; W_k^i \; y_{t-1}^i + \mu_k^i, \Sigma_k^i)$$

où W_k est la matrice de régression associée à l'état *k*, μ_k et Σ_k les vecteur moyenne et la matrice de covariance associés à l'état *k*. La matrice de transition entre états est contrainte gauche-droite comme pour les HMMs simples.

C.3.3.2 Modèles couplés

Nous présentons quatre modèles couplés (Fig. 13 et 15). Le premier modèle couplé ST-CPL (*state coupled*) couple les états des deux chaînes pour chaque pas de temps: les liens ajoutés entre les états des deux structures vont dans le sens de la chaîne verticale vers la chaîne horizontale. Ceci donne plus de poids à la chaîne verticale qui contrôle ainsi la chaîne horizontale. En effet, les expériences précédentes [Hallouli 2002] ont montré que le HMM vertical était plus performant que le HMM horizontal pour les caractères latins et les chiffres. La direction dominante des traits est en effet celle de la verticale. Le deuxième modèle est un modèle plus général (GNL_CPL) qui couple, de plus, les états de la chaîne horizontale aux observations de la chaîne verticale. Le troisième modèle est un modèle auto-régressif couplé (AR-CPL) : des liens sont ajoutés sur la première structure couplée ST_CPL pour lier les observations entre deux pas de temps.

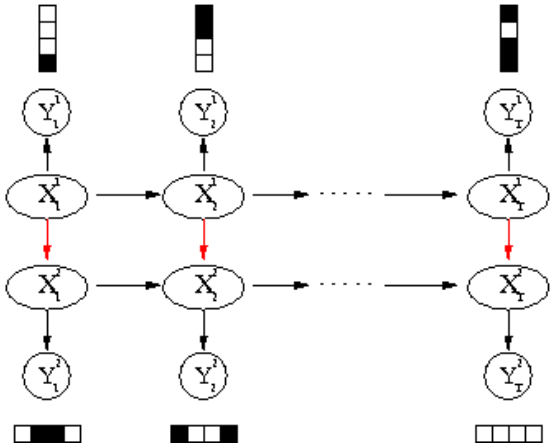


Figure 13. *Modèle couplé par états uniquement : ST_CPL*

La paramétrisation du modèle ST_CPL est définie par la donnée des CPTs initiales Π, et pour $t >= 2$ par les CPTs (matrices de transition entre états) A et U, et les distributions (CPDs) b^i liées aux observations de chaque flux i:

$$\Pi_k^1 = P(X_1^i = k) \quad \forall k \in [1,Q]$$

$$\Pi_{j,k}^2 = P(X_1^2 = k \mid X_1^1 = j) \quad \forall k, j \in [1,Q]$$

$$A_{k,j} = P(X_t^1 = j \mid X_{t-1}^1 = k)$$

$$U_{j,k,l} = P(X_t^2 = l \mid X_{t-1}^2 = j, X_t^1 = k) \qquad \forall k, j, l \in [1,Q]$$

$$b_k^i(y_t^i) = P(Y_t^i = y_t^i \mid X_t^i = k) = N(y_t^i; \mu_k^i, \Sigma_k^i) \qquad i = 1,2$$

La matrice *A* est une matrice de transition entre états gauche-droite classique. La matrice (CPT) *U* est par contre plus complexe et de plus grande dimension que *A*. C'est un ensemble de *Q* matrices (une pour chaque valeur possible de X_t^2) de taille *QxQ*. Pour réduire le nombre de paramètres de la CPT *U*, on impose que les transitions soient gauche droite entre valeurs d'états du flux horizontal (X_t^2) et ergodiques pour les transitions d'états entre le flux vertical et le flux horizontal (de X_t^1 à X_t^2). Ainsi on n'impose pas de synchronie entre les états des deux flux. La valeur X_t^2 doit être supérieure ou égale à celle de X_{t-1}^2, mais elle peut être supérieure, inférieure ou égale à celle de X_t^1 (cf. Fig. 14).

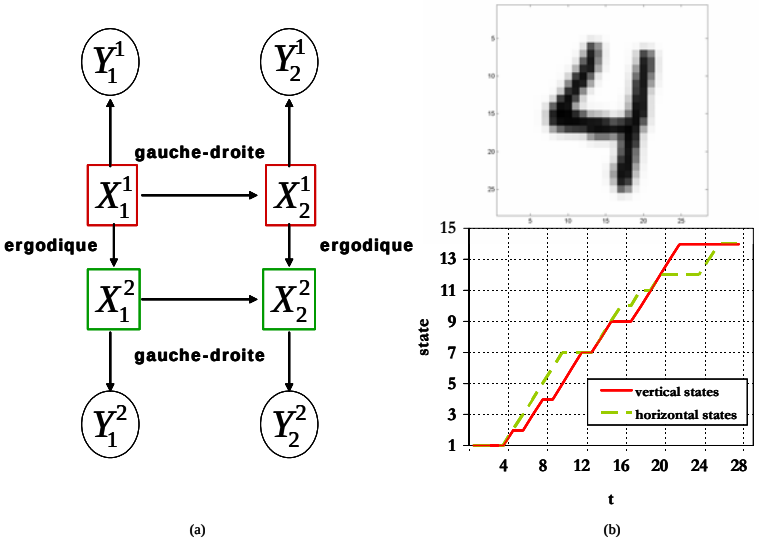


Figure 14. (a) *Transitions gauche-droite ou ergodiques entre variables d'états. (b) Deux séquences d'états (une pour le flux vertical, l'autre pour le flux horizontal) sont produites au décodage lors de la reconnaissance du chiffre 4.*

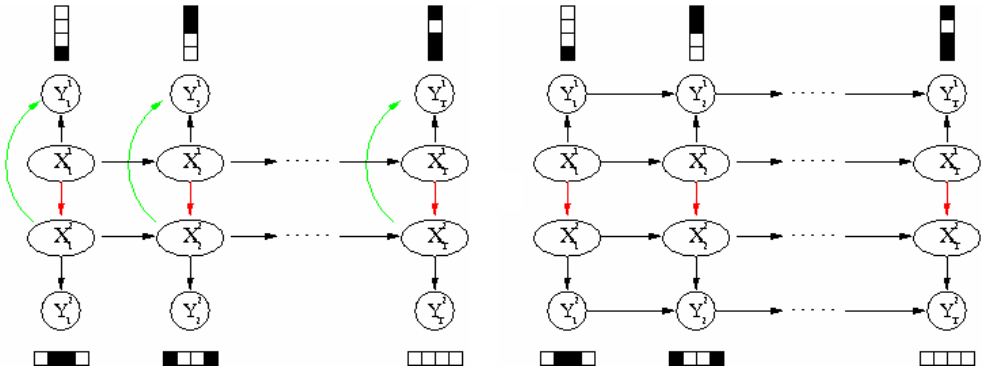


Figure 15. *Modèles couplés GN_CPL (gauche) et AR_CPL (droite)*

La paramétrisation du modèle GN_CPL diffère du modèle couplé ST_CPL par la CPD b^1 du flux vertical . Celle ci s'écrit :

$$b^1_{j,k}(y^1_t) = P(Y^1_t = y^1_t \,|\, X^1_t = j, X^2_t = k) = N(y^1_t; \mu^1_{j,k}, \Sigma^1_{j,k})$$

Cette paramétrisation induit un nombre supérieur de paramètres : *QxQ* vecteurs moyennes et *QxQ* matrices de covariance pour modéliser b^1.

Enfin, la paramétrisation du modèle auto-régressif couplé AR_CPL diffère du modèle couplé ST_CPL par les CPDs b^i , i=1,2 qui s'expriment pour $t \geq 2$ par:

$$b^i_k(y^i_t; y^i_{t-1}) = P(Y^i_t = y^i_t \,|\, Y^i_{t-1} = y^i_{t-1}, X^i_t = k) = N(y^i_t; W^i_k\ y^i_{t-1} + \mu^i_k, \Sigma^i_k) \quad i = 1,2$$

Ce modèle bénéficie à la fois du fait de coupler les observations dans un seul modèle et du côté prédictif de l'auto-régression. L'ensemble des modèles présentés est appliqué à des tâches de reconnaissance de caractères, notamment des caractères dégradés (cf. Section C.3.3.3).

C.3.3.3 Reconnaissance de caractères dégradés

Les tâches de reconnaissance sur lesquelles nous appliquons nos modèles à base de réseaux Bayésiens Dynamiques sont la reconnaissance de chiffres manuscrits et de caractères imprimés anciens. Ces caractères sont dégradés soit artificiellement, pour les chiffres, soit naturellement pour les caractères anciens.

Les chiffres manuscrits sont issus de la base MNIST [LeCun 1998] collectés lors d'une campagne de recensement aux Etats Unis. La base d'apprentissage est constituée de 60 000 chiffres, et la base de test de 10 000 chiffres centrés dans une fenêtre 28x28. Les images sont ensuite filtrées (filtrage gaussien 3x3) et les valeurs de pixels normalisées entre 0 et 1.
La dégradation artificielle consiste à créer des coupures (ici 5x5) dans les caractères par un modèle de dégradation. Celui ci s'inspire du modèle de dégradation de Baird [Baird 1992] et consiste à modifier la valeur des pixels des imagettes de caractères à l'intérieur d'une fenêtre de taille fixe. Ce modèle inclut trois paramètres :
- w : le nombre de fenêtres appliquées sur chaque caractère.
- la moyenne μ et l'écart type σ de la loi Gaussienne des valeurs de pixels modifiés.

Pour changer les valeurs originales normalisées de pixels en valeurs de fond, nous utilisons les paramètres : $\mu = 0$ et $\sigma = 0.015$. Le nombre de fenêtres varie : w=0,1,2 ou 3. Pour dégrader un caractère : on recherche aléatoirement le point central de la fenêtre. Si celui ci est dans le fond de l'image, le point d'écriture le plus proche est recherché et la fenêtre centrée sur ce nouveau point. On renouvelle ce procédé pour chaque nouvelle fenêtre (si $w \succ 1$).

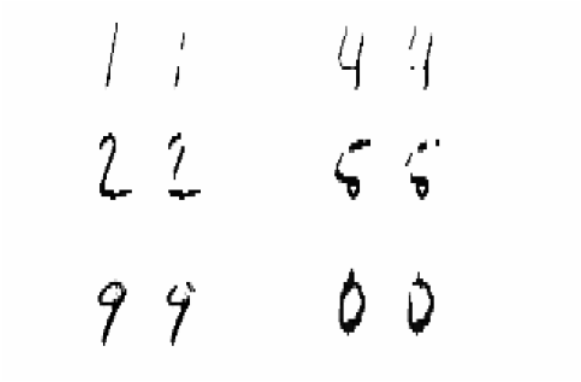


Figure 16. Caractères originaux (gauche) et caractères dégradés (droite). Deux zones de coupures sont crées sur chaque caractère (w=2)

Le nombre global d'états Q est fixé à Q=14. Cette valeur est obtenue sur une base intermédiaire de validation. L'apprentissage des paramètres des modèles est réalisé classe par classe, sur un ensemble de 5000 chiffres (500/classe) non dégradés de la base d'apprentissage. Les paramètres sont obtenus par validation croisée en découpant la base d'apprentissage en cinq sous-ensembles. Les paramètres conduisant au meilleur taux de reconnaissance sur un sous-ensemble de validation sont conservés et le modèle correspondant est évalué sur la base de test.

Table 3. Taux de reconnaissance pour les chiffres manuscrits suivant les modèles, couplés ou non, et pour différents niveaux de dégradation (w=0 –pas de dégradation, w=1, w=2).

Modèle	w=0	w=1	w=2
HMM vertical	90.2	86.9	83.8
HMM horizontal	87.4	82.8	75.3
AR-vertical	93.2	89.8	85.3
AR-horizontal	87.7	81.6	75.6
ST_CPL	92.4	90.8	87.4
GNL_CPL	93.4	90	86.2
AR_CPL	94.9	93.4	90.9

Les résultats en phase de test sont indiqués en Table 3. Les différents niveaux de dégradation correspondent au nombre de coupures créées dans chaque caractère. On remarque que les modèles observant les colonnes de pixels sont meilleurs que ceux observant les lignes, ce qui confirme nos résultats précédents concernant les HMMs classiques. Les modèles auto-régressifs sont généralement meilleurs que leur équivalent non auto-régressif. La supériorité des modèles couplés est surtout sensible pour les caractères dégradés (w=1,2). Le modèle couplé auto-régressif AR_CPL est le modèle le plus performant parmi tous les modèles réalisés.

Les caractères imprimés anciens sont extraits d'un livre imprimé du XVIIème siècle numérisés par la British Library. Ce livre est écrit en français et décrit la réception donnée en 1636 à Fontainebleau par Louis XIII en l'honneur du duc de Parme. Certaines pages sont bien contrastées. Sur d'autres, l'encre s'est estompée au cours du temps et les caractères sont fragmentés comme le montre la Figure 17.

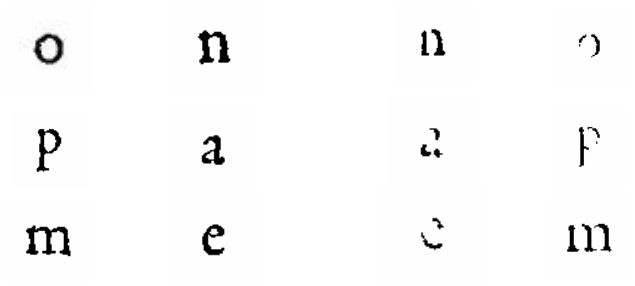


Figure 17 Caractères anciens standard (gauche) et dégradés (droite)

L'ensemble d'apprentissage inclut 16 classes de caractères et 50 caractères par classe. Deux ensembles de test sont collectés d'environ 1 000 caractères chacun: l'un issu des pages bien contrastées (test-s), l'autre des pages dégradées (test-d).

Table 4. Taux de reconnaissance pour les caractères imprimés anciens suivant les modèles, et pour deux niveaux de dégradation (test-s : ensemble standard non dégradé, test-d : ensemble dégradé).

Modèle	*test-s*	*test-d*
HMM vertical	98.3	93.8
HMM horizontal	93.7	88.1
AR-vertical	97.9	94.5
AR-horizontal	96.2	91.2
ST_CPL	98.7	95.5
GNL_CPL	98.6	94
AR_CPL	98.8	96

Comme précédemment, les modèles couplés sont meilleurs pour les caractères dégradés que les modèles non couplés, à l'exception du modèle GNL_CPL. Un défaut sur les observations du flux vertical entraîne des perturbations sur les deux séquences d'états horizontal et vertical. Tandis que pour les autres modèles couplés, la perturbation est limitée aux états du flux défectueux.

C.3.3.4 Implémentation

Les modèles de caractères ont été réalisés avec la boîte à outils BayesNet [Murphy, 2002] dans le cas dynamique. C'est un outil générique écrit en Matlab qui permet de représenter les structures de réseau, et d'effectuer l'apprentissage et l'inférence. L'écriture en Matlab implique des temps très longs de calculs. En particulier la reconnaissance d'un caractère prend plusieurs minutes alors qu'elle est quasiment instantanée pour les modèles de Markov avec HTK.

Les CPDs des modèles HMMs mono-flux sont initialisés linéairement (répartition linéaire des observations dans les états). Les CPDs des modèles auto-régressifs (mono-flux et couplés) sont initialisées aléatoirement pour chaque état avec des matrices de covariance pleines. Les CPDs des modèles ST_CPL et GNL_CPL sont initialisées avec la même matrice de covariance et la même moyenne : la moyenne et la covariance empirique calculée sur l'ensemble des observations.

C.4 Conclusion

L'approche stochastique, développée initialement pour la reconnaissance de la parole, est aussi très appropriée pour la reconnaissance de l'écriture. Nos premiers travaux ont utilisé les modèles de Markov cachés (HMMs) pour la reconnaissance de caractères en fusionnant deux modèles, l'un observant les lignes, l'autre les colonnes, au niveau décision et au niveau des données. Nous avons aussi appliqué les HMMs à la reconnaissance de mots manuscrits cursifs arabes en proposant un ensemble de caractéristiques robuste et performant pour cette tâche. Nous avons aussi fusionné plusieurs classifieurs au niveau décision pour améliorer encore les performances et tenir compte des positions erronées des marques diacritiques et de l'inclinaison possible de l'écriture.

Les résultats obtenus ont montré l'intérêt du couplage, au niveau décision ou au niveau des données. Nous avons aussi voulu étendre le cadre des HMMs pour tenir compte des dépendances entre observations. Les réseaux Bayésiens dynamiques sont un formalisme faisant partie des approches stochastiques qui permet de dépasser les hypothèses d'indépendance conditionnelle des observations. Dans ce cadre, nous avons tout d'abord construit des modèles mono-flux simples observant soit les lignes soit les colonnes de l'image. Puis nous avons lié les observations entre deux pas de temps pour obtenir des modèles mono-flux auto-régressifs qui ne supposent plus l'indépendance conditionnelle.

Plusieurs modèles couplant les deux flux d'observations ont été par la suite proposés. Ceux ci lient entre eux les états des deux modèles mono-flux. Un des modèles proposé, le modèle GNL_CPL, couple les états d'un premier flux avec les observations du second flux. De même que pour les modèles mono-flux, nous proposons aussi des modèles couplés auto-régressifs qui lient les observations d'un même flux entre elles.

Nous avons appliqué ces modèles à deux tâches de reconnaissance de caractères dégradés. La première tâche concerne la reconnaissance des chiffres manuscrits sur lesquels des coupures sont créées artificiellement suivant un modèle de dégradation. La deuxième tâche concerne la

reconnaissance de caractères imprimés anciens qui incluent de nombreuses coupures naturelles. Les modèles proposés permettent de compenser les défauts sur un flux d'observation par le deuxième flux d'information observé conjointement. Le modèle couplé auto-régressif qui lie les observations entre elles, améliore encore la modélisation et obtient les meilleures performances.

D'autres structures de réseaux Bayésiens, plus optimales, peuvent être trouvées : les dépendances entre variables cachées ou observables dans nos modèles ont été fixées a priori. L'apprentissage de la structure d'un réseau Bayésien à partir des données est un domaine de recherche qui permettra aussi de trouver la structure de réseau optimale pour chaque classe de caractère.

Bibliographie (sujet C)

publications issues du sujet C

Recht M. (1999), *Contribution à la reconnaissance de caractères imprimés dégradés par chaînes de Markov cachées*, mémoire de DEA ENSEA, Cergy Pontoise.

Godeau J. (2002), Reconnaissance de caractères manuscrits à l'aide des modèles de Markov Cachés, *mémoire de DEA Sciences et technologies des télécommunications*, Université de Rennes1 et Brest.

Hallouli K. (2002), *Utilisation de modèles markoviens pour la reconnaissance des caractères imprimés*, Rapport de recherche n°2002D002 ENST.

Hallouli K., Likforman-Sulem L., Sigelle M. (2002), A comparative study between decision fusion and data fusion in Markovian printed character recognition, *Proc. of International Conference on Pattern Recognition*, Québec, pp. 147-150.

Hallouli K., Likforman-Sulem L., Sigelle M. (2003), Réseaux Bayésiens Dynamiques pour la reconnaissance des caractères imprimés dégradés, actes du *GRETSI*, Paris, CD-ROM.

Hallouli K. (2004), *Reconnaissance de caractères imprimés par méthodes Markoviennes et réseaux Bayésiens*, Thèse de Doctorat ENST, 178 pages.

Hallouli K., Likforman-Sulem L., Sigelle M. (2004), Reconnaissance de caractères manuscrits par réseaux Bayésiens dynamiques, Actes de *Cifed'04*, La Rochelle.

El-Hajj R., Mokbel C., Likforman-Sulem L. (2005), HMM-Based Arabic Cursive Handwritten Recognition System, *Actes de Research Trends in Science and Technology, RTST 05*, March 7-9, Beyrout.

El-Hajj R., Likforman-Sulem L., Mokbel C. (2005), Arabic handwriting recognition using baseline dependent features and Hidden Markov Modeling, 8th *ICDAR 2005*, Seoul, August 2005.

Likforman-Sulem L., Sigelle M. (2005), Représentation et reconnaissance de caractères manuscrits par Réseaux Bayésiens Dynamiques, *Revue des Nouvelles Technologies de l'Information* - Extraction des connaissances : Etat et perspectives, Nov 2005, vol. RNTI, n° E-5, pp. 61-64.

El-Hajj R., Mokbel C., Likforman-Sulem L. (2006), Reconnaissance de l'écriture arabe cursive : combinaison de classifieurs MMCs à fenêtres orientées, Actes de CIFED'06, Fribourg (Suisse), septembre 2006.

El-Hajj R. (2007), *Reconnaissance de mots manuscrits cursifs par l'utilisation de systèmes hybrides et de techniques d'apprentissage automatique*, Thèse de l'Ecole Nationale Supérieure des Télécommunications, juillet 2007.

Likforman-Sulem L., Sigelle M., (2007a), Recognition of degraded handwritten digits using dynamic Bayesian networks, IST&SPIE Electronic Imaging 2007.

Likforman-Sulem L., Sigelle M. (2007b), Recognition of broken characters from historical printed books using Dynamic Bayesian Networks, *Actes de ICDAR'07*, Curitiba (Brésil), pp. 173-177.

Likforman-Sulem L., Sigelle M. (2008), Recognition of degraded characters using Dynamic Bayesian Networks, *Pattern Recognition*, DOI, 2008.

El-Hajj R., Mokbel C., Likforman-Sulem L. (2008), Recognition of Arabic handwritten words using contextual character models, IST&*SPIE Document Recognition and Retrieval XV*, San José (CA).

El-Hajj R., Likforman-Sulem L., Mokbel C. (2008), Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition, *IEEE PAMI*, accepté.

Autres références

Anigbogu J. C., Belaïd A. (1995), Hidden Markov Models in Text Recognition, *IJPRAI*, Vol. 9, no 6, 925-958.

Arica N., Yarman-Vural F.T., Optical character recognition for cursive handwriting. *IEEE PAMI*, 24(6):801–813, 2002.

Askilsrud A., Haralick R., Phillips I. T. (1993), A quick guide to UW English Document Database I ,CD ROM set, Intelligent Systems Lab., University of Washington.

Baird H.S (1992), Document Image Defect Models, in H.S. Baird, H. Bunke, K. Yamamoto Eds.,in *Structured Document Image Analysis*, Springer-Verlag, pp. 546-556, 1992.

Bazzi I., Schwartz R., Makhoul J. (1999), An omnifont open vocabulary OCR system for English and Arabic, *IEEE PAMI*, Vol. 21, No 6, pp. 495-504.

Becker A., Geiger D., Schaffer A. (1996), Automatic selection of Loop breakers for genetic linkage analysis, *Human Heredity*, Vol 48, no 1.

Belaid A., Saon G., Utilisation des processus Markoviens en reconnaissance de l'écriture, Traitement du signal, vol.14, no2, pp 161-178, (1997).

Ben Amara N., Bouslama F. (2003), Classification of Arabic script using multiple sources of information: state of the art and perspective, *IJDAR*, Vol. 5, pp. 195-212.

Brand M. (1997), *Coupled hidden Markov models for modeling interactive processes*, MIT Media Lab, Technical Report.

Cho S., Kim J.H. (2003), Bayesian Network Modeling of Hangul Characters for on-Line Handwriting Recognition, *Proc. of ICDAR'03*.

Choisy C. (2002), *Modélisation analytique de l'écriture manuscrite par une approche optimale sans segmentation basée sur les champs de Markov*, Thèse de l'Université de Nancy 2.

Daoudi K., Fohr D., Antoine C. (2000), A new approach for multiband speech recognition based on probabilistic graphical models, *International conference on Spoken Language Processing (ICSLP)*.

Duin R. (2002), The combining classifier: to train or not to train, *Proc. of ICPR'02*, Quebec, pp. 765-770.

Elms A., Procter S, Illingworth J. (1998), The advantage of using an HMM based approach for faxed word recognition, *IJDAR*, Vol 1, no 18-36.

Geoffrois E., Chevalier S., Prêteux F. (2004), Programmation dynamique 2D pour la reconnaissance de caractères manuscrits par champs de Markov, *Actes de RFIA.*

Gilloux M. (1994), Reconnaissance de chiffres manuscrits par modèles de Markov pseudo-2D, *actes de CNED*, pp. 11-17.

Heckerman D. (1999), *Learning in Graphical Models,* M. Jordan, ed., MIT Press, Cambridge, MA.

Ho T. K., Hull J., Srihari S. N. (1994), Decision Combination in Multiple Classifier Systems, *IEEE PAMI*, Vol. 16., No. 1 , pp. 66-75.

Jensen F.V. (1996), *An introduction to Bayesian Networks*. UCL Press.

Kittler J., Hatef M., Duin R., Matas J. (1998), On combining classifiers, *IEEE PAMI*, Vol 20, no 3, pp. 266-239.

Khorsheed M.S. (2003), Recognizing Arabic manuscripts using a single hidden Markov model, *Pattern Recognition Letters*, 24, pp. 2235-2242.

Lauritzen L.S., Spiegelhalter J., Cowell G.R., P.A. Dawid (1999), *Probabilistic Networks and Expert Systems*, Springer, 175 Fifth Avenue, New York NY 10010 USA, première édition.

Lecun Y. (1998), http://www.research.att.com/yann/ocr/mnist/, the MNIST handwritten digit database.

Miled H., Olivier C., Cheriet M., Lecourtier Y. (1997), Coupling observation/letter for a Markovian modelization applied to the recognition of Arabic handwriting, ICDAR 97,Ulm, pp 580 – 583.

Mokbel C., Abi Akl H., Greige H. (2002), Automatic speech Recognition of arabic digits over the telephone network, Proc. of Research Trends in Science and Technology RTST'02, Beyrouth.

Murphy K. (2001), The Bayes Net Toolbox for Matlab, http ://www.ai.mit.edu/ murphyk/.

Murphy K. (2002), *Dynamic Bayesian Networks : Representation, Inference and Learning*, PhD thesis, University of California, Berkeley.

Pearl J. (1988), *Probabilistic Reasoning in Intelligent Systems Networks of Plausible Inference*, Morgan, 2ème édition.

Pechwitz M., Maddouri S., Märgner V., Ellouze N. (2002), IFN/ENIT–DataBase for Handwritten Arabic words, CIFED'02, Hammamet, Tunisia, pp 129-136.

Pechwitz M., Märgner V. (2003), HMM Based approach for handwritten Arabic Word Recognition Using the IFN/ENIT– DataBase, ICDAR'03, Edinburgh, pp. 890-894.

Sicard R., Artières T., Petit E (2006), Modeling on-line handwriting using pairwise relational features, IWFHR, La Baule.

Souafi S. (2002), *Contribution à la reconnaissance des structures des documents écrits: Approche probabiliste*, Thèse de Doctorat, INSA de LYON.

Souafi S., Parizeau M., Lebourgeois F., Emptoz H., (2002), Bayesian networks classifiers applied to documents, ICPR 2002.

Vinciarelli A., Bengio S., Bunke H. (2004), Offline handwriting recognition of un-constrained handwritten texts using HMMs and statistical language models. *IEEE PAMI*, 26(6):709–720.

Wang T., Diao Q., Zhang Y., Song G., Lai C., Bradski G., A Dynamic Bayesian Network Approach to Multi-cue based Visual Tracking, ICPR 2004.

Xiao X., Leedham G. (2002), Signature verification using a modified Bayesian network, *Pattern Recognition* (5): 983-995.

Young S.J. (2001), HTK : Hidden Markov Model Toolkit V3.0, http://htk.enf.cam.uk.uk/

Zweig G. (1998), *Speech Recognition with Dynamic Bayesian networks*, PhD thesis, University of California, Berkeley.

D. PERSPECTIVES

Les méthodes développées pour la structuration (sujet A) peuvent être appliquées à divers types de documents : lettres de clients s'adressant à une entreprise ou une administration, manuscrits ou documents d'archives dans le cadre d'activités de numérisation. Des outils d'aide à la conversion en texte intégral des manuscrits, de recherche de mots ou de motifs à partir d'images vont se développer dans les années à venir. Ces outils peuvent s'appuyer sur la segmentation et la structuration de ces documents riches en informations graphiques.

Les méthodes issues du sujet B concernent l'analyse et compréhension de documents et peuvent se développer dans le domaine actuel de recherche d'information sur Internet. De nombreux documents image circulent, générés électroniquement (pdf notamment). Les méthodes de recherche de noms présentées seront d'autant plus efficaces que ces documents ne contiennent pas d'erreurs au sens OCR. D'autre part, cette méthode se généralise à l'extraction d'autres types d'information comme l'objet du message.

Le système HMM de reconnaissance de mots appliqué à l'écriture arabe (sujet C) peut être adapté à l'écriture latine (modification du dictionnaire et re-apprentissage) et utilisé pour des tâches d'extraction d'information. L'existence de la base de données de courriers manuscrits issue du projet TechnoVision Rimes va permettre de développer des systèmes de reconnaissance sur des pages de texte (au lieu de mots isolés) et pour un vocabulaire large, voire ouvert.

Enfin, les Réseaux Bayésiens peuvent être appliqués dans de nombreux domaines : notamment pour l'authentification biométrique. Nous envisageons de réaliser l'apprentissage de la structure de réseau pour l'authentification d'images de mains, pour rechercher les dépendances entre observations dans le cadre du projet Biomet. Dans le domaine de la reconnaissance d'écriture, nous cherchons à améliorer les outils existants en déterminant les meilleures conditions initiales qui permettront de réaliser un apprentissage optimal des paramètres. L'extension des réseaux Bayésiens à la modélisation des mots est aussi envisagée. Cette extension n'est cependant pas directe du fait de la normalisation effectuée sur les images de caractères pour que les séquences d'observations soient de même longueur.

E. SELECTION DE PUBLICATIONS

- ***Recognition of degraded characters using Dynamic Bayesian Networks***

L. Likforman-Sulem et M. Sigelle

Pattern Recognition, DOI 10.1016/j.patcog.2008.03.022, 2008.

- ***Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition***

R. El-Hajj, L. Likforman-Sulem, C. Mokbel

IEEE Transactions on Pattern Analysis and Machine Intelligence, accepté, 2008.

- ***Text line segmentation of Historical documents: A survey***

L. Likforman-Sulem, A . Zahour et B. Taconet

International Journal on Document Analysis and Recognition, Springer, DOI 10.1007/s10032-006-0023-z, 2007.

- ***Automatic name extraction from degraded document images***

L. Likforman-Sulem, P. Vaillant et A. de Bodard de la Jacopière

Pattern Analysis and Applications, Vol. 9, no 2-3, pp. 211-227, Springer, 2006

- ***A Reference Biometric System based on Hand Modality***

H. Dutagaci, G. Fouquier, E. Yoruk, B. Sankur, L. Likforman-Sulem, J. Darbon,

à paraître dans *Guide to biometric reference systems and performance evaluation*, Springer, 2008. (D. Petrovska, G. Chollet, B. Dorizzi eds), ISBN 978-1-84800-291-3.

- ***A comparative study between decision fusion and data fusion in Markovian printed character recognition***

  K. Hallouli, L. Likforman-Sulem, M. Sigelle

  International Conference on Pattern Recognition ICPR 2002, Québec, pp.147-150, 2002

- ***Image and text coupling for creating electronic books from manuscripts***

  L. Robert, L. Likforman-Sulem et E. Lecolinet

  Int. Conference on Document Analysis and Recognition, *ICDAR'97*, Ulm, pp. 823-826.

- ***Une méthode de résolution des conflits d'alignements pour la segmentation des documents manuscrits***

  L. Likforman-Sulem, C. Faure

  Traitement du Signal, Vol. 12, no 6, pp. 541-549, 1995.

- ***An expert vision system for analysis of Hebrew characters and authentication of manuscripts***

  L. Likforman-Sulem, H. Maitre et C. Sirat

  Pattern Recognition, Vol. 24, no 2, pp. 121-137, 1991.