



HAL
open science

AMBRES : une approche générique pour l'analyse et la reconnaissance de documents manuscrits

M. Lemaitre, E. Grosicki, E. Geoffrois, Françoise Prêteux

► **To cite this version:**

M. Lemaitre, E. Grosicki, E. Geoffrois, Françoise Prêteux. AMBRES : une approche générique pour l'analyse et la reconnaissance de documents manuscrits. Colloque International Francophone sur l'Écrit et le Document, Oct 2008, France. pp.195-196. hal-00334419

HAL Id: hal-00334419

<https://hal.science/hal-00334419v1>

Submitted on 26 Oct 2008

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

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

AMBRES : une approche générique pour l'analyse et la reconnaissance de documents manuscrits

Mélanie Lemaître¹ – Emmanuèle Grosicki² – Edouard Geoffrois² – Françoise Prêteux³

¹ ULCO/LIL - 50, rue Ferdinand Buisson 62228 Calais cedex - France

² DGA/Centre d'Expertise Parisien - 16 bis avenue Prieur de la côte d'Or 94114 Arcueil cedex - France

³ TELECOM Sud Paris, Dpt ARTEMIS - 9, rue Charles Fourier 91011 Evry cedex - France

mel.lemaître@wanadoo.fr, Emmanuele.Grosicki@etca.fr, Edouard.Geoffrois@etca.fr, Francoise.Preteux@it-sudparis.eu

Résumé : Dans cet article, nous présentons une approche bidimensionnelle markovienne générique pour l'analyse et la reconnaissance de documents manuscrits appelée *AMBRES* (Approche Markovienne Bidimensionnelle pour la Reconnaissance et la Segmentation d'images). Elle a permis de traiter avec succès la reconnaissance de l'écriture manuscrite et la structuration de documents manuscrits et a été validée au sein de la campagne d'évaluation française RIMES (Reconnaissance et Indexation de données Manuscrites et de fac similÉS).

Mots-clés : Reconnaissance, structuration, champs de Markov, approche générique.

1 Introduction

Dans cet article, nous proposons une approche générale de reconnaissance et de segmentation d'images appliquée avec succès à la reconnaissance de l'écriture manuscrite et à la structuration de documents. L'approche générale ainsi définie est appelée *AMBRES* [LEM 07b].

AMBRES est fondée sur les champs de Markov, la programmation dynamique 2D et une analyse bidimensionnelle de l'image. De plus, un modèle de position est introduit dans le modèle markovien afin de relier la configuration d'un site à sa position spatiale (intérêt pour la modélisation de certains types de documents tels que les courriers manuscrits).

Après avoir présenté le principe de l'approche *AMBRES*, nous verrons comment *AMBRES* a permis d'aborder la reconnaissance de l'écriture manuscrite dans un premier temps puis la structuration de documents manuscrits dans un second temps. Enfin, nous concluons sur nos travaux.

2 L'approche *AMBRES*

L'idée générale est de créer un modèle markovien bidimensionnel pour chaque classe d'image. Ces modèles seront utilisés lors de la phase de reconnaissance pour déterminer la classe la plus probable des différentes images de la base de test ou pendant la structuration pour déterminer la configuration optimale des étiquettes recherchées.

On considère une approche Bayésienne classique d'analyse d'image où l'image observée est supposée être générée par un processus caché que l'on cherche à déterminer. Les modèles correspondant à chaque classe d'images vont ainsi

relier les états cachés $\omega_{i,j}$ aux primitives $o_{i,j}$ extraites des images (scalaires ou vecteurs). L'objectif est de déterminer pour chaque image la configuration optimale $\hat{\omega}$ de la grille d'états connaissant les observations. Pour cela, on adopte une stratégie Bayésienne :

$$\hat{\omega} = \arg \max_{\omega \in \Omega} P(O|\omega)P(\omega),$$

où Ω est l'ensemble des configurations possibles ω et $O = \{o_{i,j}, (i,j) \in L\}$ est l'ensemble des primitives de l'image.

À l'information de dépendance contextuelle locale introduite par les champs aléatoires de Markov (ici, voisinage en 4-connexité), on ajoute une information de position exploitant des considérations spatiales. $P(\omega)$ va s'exprimer de la façon suivante :

$$P(\omega) = P^T(\omega) \times P^P(\omega).$$

Les paramètres du modèle sont ainsi ceux du terme d'attache aux données $P(O|\omega)$, ceux du terme de transition $P^T(\omega)$ et ceux du terme de position $P^P(\omega)$.

- $P(O|\omega)$ est modélisé par des multigaussiennes.
- $P^T(\omega)$ s'exprime de la façon suivante :

$$P^T(\omega) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} V_c(\omega)\right),$$

où \mathcal{C} est l'ensemble des cliques associées au voisinage, V_c est le potentiel associé à la clique c et Z est la constante de normalisation.

- $P^P(\omega)$ calcule les probabilités des sites d'être dans un certain état suivant la position spatiale du site (i,j) :

$$P^P(\omega) = \prod_{i,j} P(\omega_{i,j} | (i,j)),$$

où les $P(\omega_{i,j} | (i,j))$ sont modélisés par des multigaussiennes de dimension 2 (informations horizontale et verticale).

3 *AMBRES* pour la reconnaissance de l'écriture manuscrite

3.1 Modélisation de l'écriture manuscrite

L'écriture manuscrite est une forme constituée d'un ensemble de traits de différentes directions et situés à différents

endroits dans l'image. L'objectif va donc être de segmenter les images de caractères ou de mots en fonction de la direction et de la position des traits dans l'image. Les $n_x \times n_y$ zones ainsi obtenues correspondent aux différents états du champs de Markov qui seront donc caractéristiques d'une direction et d'une position dans l'image. Cette information de direction est extraite à partir de primitives directionnelles : nous avons pour cela choisi d'utiliser les primitives spectrales locales.

3.2 Reconnaissance de caractères isolés

Ici, nous traitons une tâche de reconnaissance de caractères isolés. On obtient, sur la base de données MNIST, un taux d'erreurs de 2.04%. Outre le taux d'erreurs, on peut également observer la bonne segmentation en états 5×7 des images de chiffre (figure 1 : par exemple l'état 7 est associé pour les deux images à la direction horizontale positionnée sur le haut du "3").

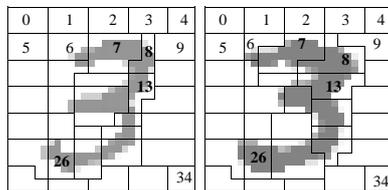


FIG. 1 – Exemple de segmentations en états obtenues pour deux images de la classe "3"

4 AMBRES pour la structuration de documents manuscrits

Pour réaliser un traitement automatique de documents non contraints, une phase préliminaire d'extraction de la structure physique est nécessaire afin d'extraire des informations clés. Comme première application, nous avons choisi d'aborder le cas des courriers manuscrits similaires à ceux envoyés par des personnes à des entreprises.

4.1 Structuration de courriers manuscrits

L'objectif de la structuration de lettres manuscrites est d'étiqueter les différents pixels de l'image suivant leur appartenance à un champ. Nous utilisons dans ce contexte 8 champs différents qui correspondent chacun à un état du champ de Markov : fond, coordonnées expéditeur, coordonnées destinataire, objet, ouverture, signature, corps de texte et date, lieu.

4.2 Mise en oeuvre

Pour extraire la structure physique des courriers manuscrits, nous n'utilisons ici que l'information de texture (présence de l'écriture) et de position spatiale des différents champs [LEM 07a]. En effet, l'écriture d'un courrier manuscrit est régie par certaines règles qui peuvent varier suivant le scripteur. Ces règles sont pour la plupart des règles de positionnement : par exemple l'adresse de l'expéditeur est en haut à gauche. Le modèle de position introduit dans AMBRES permet de tenir compte de la position spatiale des différents champs.

Les primitives utilisées pour cette application correspondent à la valeur moyenne des niveaux de gris de chaque fenêtre. Les paramètres du modèle sont appris directement sur les vérités terrains des images de la base de développement.

Un taux d'erreurs global de 10.7% a été atteint sur la base de données RIMES-1. Un exemple de structuration donnée par AMBRES est illustré figure 2. Dans cet exemple un taux d'erreurs de 5.3% a été obtenu.



FIG. 2 – Exemple de structuration obtenue avec AMBRES (Jaune : corps de texte ; Rose : signature ; Cyan : ouverture ; Rouge : expéditeur ; Vert : destinataire ; Gris : date, lieu ; Bleu : objet)

5 Conclusion

Nous avons présenté une approche générale bidimensionnelle markovienne appelée AMBRES appliquée avec succès à des tâches aussi diverses et variées que la reconnaissance de l'écriture manuscrite ou la structuration de documents manuscrits.

Les performances obtenues en reconnaissance de caractères manuscrits sur MNIST sont à l'état de l'art. De plus, AMBRES a obtenu les meilleurs résultats sur 2 des 3 tâches de reconnaissance de caractères isolés lors de la campagne d'évaluation RIMES-1.

La généralisation au niveau document a été abordée dans le cadre de la structuration de courriers manuscrits. La méthode mise en oeuvre permet déjà à ce stade d'obtenir de bonnes performances. L'introduction d'une phase préliminaire de reconnaissance de mots clés pourrait, par la suite, permettre d'améliorer significativement les performances.

Enfin, AMBRES pourrait être étendue à d'autres problématiques de l'analyse de documents voire du domaine de la vision.

Références

[LEM 07a] LEMAITRE M., GROSICKI E., GEOFFROIS E., , PRÊTEUX F., Preliminary experiments in layout analysis of handwritten letters based on textural and spatial information and a 2D Markovian approach, *ICDAR*, vol. 2, 2007, pp. 1023–1027.
 [LEM 07b] LEMAITRE M., Approche markovienne bidimensionnelle d'analyse et de reconnaissance de documents manuscrits, *Thèse de doctorat de l'Université Paris V*, 2007.