

L'écrit et le document

Reconnaissance de chiffres manuscrits par modèle de Markov pseudo-2D

Handwritten Digit Recognition Using Pseudo-2D Markov Models

par Michel GILLOUX

Service de Recherche Technique de la Poste SRTP/RD/RVA, Mabon
10, rue de l'Île, F-44063 Nantes cedex 02

Résumé

Nous présentons une nouvelle méthode de reconnaissance de caractères manuscrits fondée sur le principe des modèles de Markov pseudo-2D ou planaires (PHMM). Cette méthode se situe dans la classe des techniques qui tentent de rétablir la forme prototype d'un caractère à partir d'une version déformée de celle-ci. Contrairement aux approches souvent proposées dans ce domaine, notre méthode ne repose sur aucune extraction de traits explicite et produit des scores de reconnaissance qui sont des estimations des probabilités bayésiennes. Elle se prête à un apprentissage automatique. Une de ses caractéristiques distinctive est l'utilisation d'un modèle probabiliste (réseau de Markov) réellement bi-dimensionnel mais causal pour estimer les probabilités. Enfin, on montre comment on peut engendrer des images synthétiques des caractères possibles et ce suivant leur probabilité estimée par le modèle.

Mots clés : Reconnaissance de caractères manuscrits, champs de Markov, modèles de Markov cachés, modèles de Markov pseudo-2D, apprentissage automatique.

Abstract

We present a new method for recognising handwritten characters based on pseudo-2D Markov models. This method belongs to a class of techniques which attempt to recover the prototype shape of a character starting from a distorted image. By contrast with popular approaches in this field, our method does not need any explicit feature extraction and it outputs recognition scores which are estimates of Bayes probabilities. It also allows an automatic training. One of its distinctive characteristics is the use of a truly bi-dimensional but causal model for the estimation of probabilities. Lastly, we show how synthetic images of characters fitted to the statistical distribution of shapes in the training set can be generated by this model.

Key words : Handwritten characters recognition, Markov random fields, hidden Markov models, pseudo-2D hidden Markov models, automatic training.

1. Introduction

Les modèles de Markov cachés (HMM pour *Hidden Markov Model*) [Bahl] ont récemment été employés pour reconnaître l'écriture cursive [Chen], [Gilloux]. Ces travaux reposent pour une grande part sur l'expérience accumulée dans le domaine de la reconnaissance de la parole où les HMM sont fréquemment employés. L'intérêt suscité par les modèles de Markov cachés en reconnaissance des formes tient à ce qu'ils ne requièrent que peu de connaissance a priori sur le signal à classer. A cela s'ajoute l'existence de techniques d'apprentissage automatique des paramètres d'un HMM. Les travaux publiés dans le cadre de la reconnaissance de mots manuscrits reposent tous sur une représentation schématique des images en « segments d'écriture » destinée à traduire les formes initiales en un signal mono-dimensionnel (1D) formé d'une suite d'observations propre à être traitée par un modèle de Markov. En effet, un HMM est composé d'une chaîne de Markov dont chaque état traversé, non

directement observable, émet une observation. Les observations utilisées dans le contexte de la reconnaissance de l'écriture sont en général identifiables aux lettres ou à des portions de lettres mais elles peuvent aussi correspondre aux colonnes de pixels de l'image. On peut penser que les limites de ces approches sont dues à la non prise en compte du caractère bi-dimensionnel (2D) des images.

Il existe différentes extensions des modèles de Markov à la dimension 2. Les *champs de Markov* [Geman] constituent la classe la plus générale dans ce domaine. Ils reprennent la notion d'état caché et d'émission d'observations par ces états mais reposent sur des lois de probabilités conditionnelles entre états véritablement bi-dimensionnelles. Cependant, la non causalité des champs cachés (chaque valeur d'un champ dépend de celle des sites voisins qu'elle conditionne à son tour) rend difficile leur mise en pratique.

Une variante des champs de Markov est constituée des *réseaux de Markov* [Abend] dans lesquels les lois de probabilités sont bi-dimensionnelles mais causales et permettent la reconnaissance d'une image par un algorithme simple mais qui ne possèdent pas de méthode d'apprentissage efficace.

Une autre extension des HMM à la dimension 2, les modèles de Markov pseudo-2D ou planaires (PHMM pour *Planar Hidden Markov Model*) [Levin] a été proposée dans le contexte de la reconnaissance de caractères. Les PHMM permettent une reconnaissance des images par un algorithme simple mais aussi un apprentissage automatique des paramètres du modèle. Ces raisons font qu'il est intéressant de considérer des modèles de Markov pseudo-2D bien que leur pouvoir de génération soit inférieur à celui des champs et réseaux de Markov.

Les HMM pseudo-2D ont déjà été utilisés en reconnaissance de caractères [Levin], [Agazzi93a], [Agazzi93b], [Bose] mais seulement dans le contexte de textes imprimés. Dans ce domaine, leur principal intérêt est la possibilité d'intégrer reconnaissance et segmentation et aussi d'estimer certaines déformations déterministes d'amplitude limitée (dilatation, inclinaison, translation) bien que celles-ci ne soient pas les seules à affecter les caractères imprimés [Agazzi93a]. On observe cependant dans la pratique que, de par leur modélisation trop simple du contexte local, les PHMM estiment mal les scores de reconnaissance bien qu'ils réalisent correctement l'attribution des états cachés aux pixels de l'image.

La technique des PHMM est employée dans cet article pour définir une correspondance entre pixels de l'image d'un caractère et pixels d'un prototype d'une classe de caractères. Les paramètres des modèles sont appris automatiquement à partir d'exemples.

Notre approche est nouvelle en ce qu'elle tente d'appliquer des HMM pseudo-2D à la reconnaissance de caractères manuscrits. Dans ce cadre, les déformations ne sont jamais déterministes et ne peuvent donc pas être modélisées aisément. De plus, il existe de multiples façons d'écrire un caractère. Cette variabilité doit aussi être prise en compte dans le modèle de reconnaissance.

Notre méthode se distingue aussi par l'emploi, lors du calcul des scores de reconnaissance, d'un réseau de Markov dont on a vu que le pouvoir de représentation est supérieur à celui des PHMM. L'attribution des états cachés aux pixels de l'image étant effectuée le plus souvent correctement par un PHMM, nous utilisons ce dernier modèle dans cette tâche particulière puis un réseau de Markov (causal) pour estimer la probabilité de génération de l'image et des états associés.

Enfin, nous montrerons comment il est possible d'étudier qualitativement les capacités du modèle en lui faisant engendrer des images synthétiques de caractères manuscrits. Cette faculté distingue les approches markoviennes, orientées dans le sens de la génération de formes, des autres approches, en particulier neuronales, dont les paramètres internes ne sont pas aisément interprétables.

Cet article est organisé comme suit. La section 2 définit les différents modèles de Markov mono et bi-dimensionnels. La section

3 décrit le modèle particulier que nous avons utilisé. La section 4 rapporte les résultats obtenus en reconnaissance. La section 5 est relative à la synthèse de caractères.

2. HMM, HMM pseudo-2D, réseaux et champs de Markov

Nous rappelons dans ce chapitre les principales définitions relatives aux méthodes markoviennes considérées : HMM, PHMM, réseaux et champs de Markov.

Ces méthodes constituent une classe de modèles statistiques paramétriques utilisée en reconnaissance des formes bayésienne. Dans le contexte bayésien, la classification optimale des formes x dans des classes c est réalisée en estimant les probabilités $p(C = c|X = x)$ d'appartenance de x à c . Par application de la règle de Bayes,

$$p(C = c|X = x) = p(X = x|C = c) \times p(C = c) / p(X = x)$$

Comme la probabilité a priori $p(X = x)$ de la forme à classer ne départage pas les hypothèses de classe, la reconnaissance se fait à partir des vraisemblances

$$V(C = c|X = x) = p(X = x|C = c) \times p(C = c)$$

La reconnaissance bayésienne requiert des techniques capables d'estimer $p(X = x|C = c)$. L'ensemble des méthodes markoviennes parmi lesquelles figurent les HMM sont utilisées à cet effet par assimilation de chaque classe à un modèle.

2.1. CHAMPS DE MARKOV CACHÉS

Soit $X = \{X_{i,j}, 1 \leq i \leq M, 1 \leq j \leq N\}$ un ensemble de variables aléatoires observables définies sur une grille de taille $M \times N$ et $E = \{E_{i,j}, 1 \leq i \leq M, 1 \leq j \leq N\}$ un second ensemble de variables aléatoires non directement observables (*états cachés*) sur la même grille. On notera Ω l'ensemble des valeurs possibles de X et Σ celui des valeurs de E . Soit d'autre part un système de voisinage $V = \{V_{i,j}, 1 \leq i \leq M, 1 \leq j \leq N\}$ où $V_{i,j}$ est l'ensemble des voisins du site $(i, j) : V_{i,j} \subset [1, M] \times [1, N]$.

$V_{i,j}$ vérifie :

(1) $(i, j) \notin V_{i,j}$

et (2) $(i, j) \in V_{k,l}$ si et seulement si $(k, l) \in V_{i,j}$

X est un champ de Markov caché si et seulement si

1. $\forall e \in \Sigma, p(E = e) > 0$

2. $\forall i, j, \forall e,$

$$p(E_{i,j} = e_{i,j} | E_{k,l}, (k, l) \neq (i, j)) = p(E_{i,j} = e_{i,j} | E_{k,l} = e_{k,l}, \text{ou } (k, l) \in V_{i,j})$$

3. $\forall x \in \Omega,$

$$p(X = x | E = e) = \prod_{(i,j) \in S} p(X_{i,j} = x_{i,j} | E_{i,j} = e_{i,j}).$$

Les termes de forme $p(X_{i,j} = x_{i,j} | E_{i,j} = e_{i,j})$ sont appelés *probabilités d'émission*¹. La deuxième condition traduit les propriétés locales du champ de Markov des états : chaque site n'est conditionné que par ses voisins. La troisième condition indique que les observables X sont conditionnées site par site par les états cachés correspondants. Les systèmes de voisinage généralement utilisés dans un champ de Markov caché sont indiqués sur la figure 1.

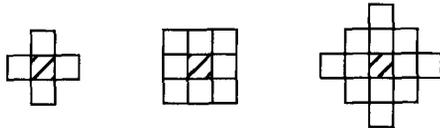


Fig. 1. – Différents systèmes de voisinages utilisés dans les champs de Markov cachés (le site courant est hachuré).

On peut remarquer que les systèmes de voisinage de la figure 1 ne permettent pas de calculer la probabilité d'une image par une formule analytique à cause de la *non causalité* du champ caché. Les réseaux de Markov ont été introduits afin de faire apparaître une causalité dans les lois de probabilité des états cachés.

2.2. RÉSEAUX DE MARKOV CACHÉS

Dans un réseau de Markov homogène d'ordre 3, le voisinage d'un site est tel que :

$$p(e_{i,j} | e_{k,l}, (k,l) \neq (i,j)) = p \left(e_{i,j} \begin{matrix} e_{i-1,j-1} e_{i-1,j} \\ e_{i,j-1} \end{matrix} \right) = f(e_{i,j}, e_{i-1,j-1}, e_{i-1,j}, e_{i,j-1})$$

Cette nouvelle définition permet le calcul de la probabilité d'une image x pourvu que le champ caché e soit connu. Cependant, sans connaissance a priori des états cachés, la probabilité de l'image est :

$$p(x) = \sum_{E \in \Sigma} \left(\prod_{1 \leq i \leq M, 1 \leq j \leq N} p(x_{i,j} | e_{i,j}) \right) \times \left(\prod_{1 \leq i \leq M, 1 \leq j \leq N} p \left(e_{i,j} \begin{matrix} e_{i-1,j-1} e_{i-1,j} \\ e_{i,j-1} \end{matrix} \right) \right)$$

Cette formule possède une complexité non polynômiale d'ordre $O(2MN \cdot C^{MN})$, C étant le nombre d'états possibles, et ne donne lieu à aucun calcul par récurrence de complexité polynômiale.

2.3. MODÈLES DE MARKOV PLANAIRES

Dans un HMM planaire, le champ caché est décomposé en états et *super-états*. Chaque site (i,j) recèle une valeur cachée composite

N.B. : Dans la suite de cet article et lorsque cette notation ne sera pas ambiguë, les expressions de probabilités d'événements de type $p(X = x)$ seront simplifiées en $p(x)$.

$(s_{i,j}, e_{i,j})$ formée du super-état $s_{i,j}$ et de l'état $e_{i,j}$. La loi de probabilité du champ caché est alors la suivante :

$$p((s_{i,j}, e_{i,j}) | (s_{k,l}, e_{k,l}), (k,l) \neq (i,j)) = \delta(s_{i,j}, s_{i,j-1}) \times p(e_{i,j} | e_{i,j-1}, s_{i,j-1})$$

avec les conditions aux limites.

$$p((s_{i,1}, e_{i,1}) | (s_{k,l}, e_{k,l}), (k,l) \neq (i,1)) = p(s_{i,1} | s_{i-1,1}) \times p(e_{i,1} | s_{i,1})$$

$$s_{M,j} = s_f \quad 1 \leq j \leq N \text{ (super-état final)}$$

$$e_{i,N} = e_f(s_{i,N}) \text{ (état final pour le super-état).}$$

La probabilité des états cachés est alors :

$$p(e) = p(s_1) \times \prod_{1 < i \leq M} p(s_{i,1} | s_{i-1,1}) \times \prod_{1 \leq i \leq M, 1 < j \leq N} \delta(s_{i,j}, s_{i,j-1}) \times p(e_{i,j} | e_{i,j-1}, s_{i,j})$$

On peut démontrer à partir de cette formule que non seulement le champ caché est causal mais aussi que la probabilité du champ caché, et par voie de conséquence celle de l'image x , peuvent être calculées par récurrence suivant un algorithme polynômial d'ordre $O(MNSE)$, S et E étant respectivement le nombre de super-états et états différents.

Une autre définition peut permettre d'éclairer la nature d'un HMM pseudo-2D. Un PHMM est un modèle de Markov caché dont les probabilités d'émission sont aussi modélisées par des HMM. Rappelons qu'étant donné un ensemble de variables aléatoires observables $X = (X_i, 1 \leq i \leq N)$ et un ensemble de variables aléatoires non observables (états) $E = \{E_i, 1 \leq i \leq N\}$, X et E sont engendrés par un modèle de Markov caché d'ordre 1 si et seulement si :

1. $p(X = x | E = e) = \prod_i p(X_i = x_i | E_i = e_i)$
2. $p(E_i = e_i | E_j = e_j, j \neq i) = p(E_i = e_i | E_{i-1} = e_{i-1}) = f(e_i, e_{i-1}) \quad 1 < i \leq N$
3. $p(E_i = e_i | E_j = e_j, j \neq i) = p(E_i = e_i), i = 1$
4. $E_N = e_f$ (état final).

Les trois dernières conditions indiquent que les états forment une chaîne de Markov d'ordre 1. Pour rapprocher ce dernier modèle et celui des PHMM décrit auparavant il suffit de remarquer que les super-états respectent les conditions 2 et 3 alors que pour un super-état donné, les états d'une ligne forment aussi une chaîne de Markov (voir figure 2).

Nous avons vu que les différents modèles de type markovien servent à estimer la probabilité qu'une image soit engendrée dans la classe modélisée. En parallèle, un modèle de ce type peut servir à estimer les états cachés les plus vraisemblables pour une image donnée. Cette opération est à la base des algorithmes d'apprentissage. Dans le cas d'un PHMM, cette estimation est faite au moyen d'une extension de l'algorithme de Viterbi. Le schéma algorithmique récursif utilisé à cet effet est décrit ci-dessous. Nous utilisons les notations suivantes :

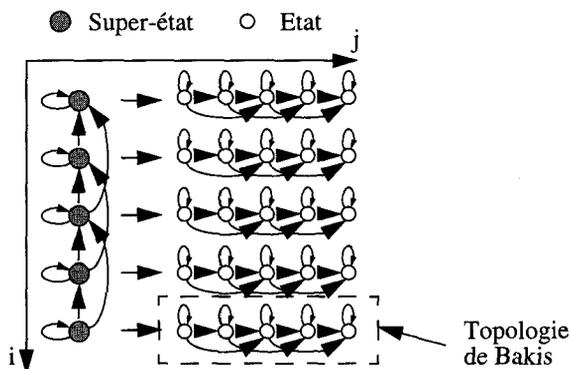


Fig. 2. – Structure d'un PHMM.

$$E_{i,j} = (s, e) \Rightarrow E_{i,j}|_s = s, E_{i,j} = (s, e) \Rightarrow E_{i,j}|_e = e$$

$$p(E_{1,1} = (s, e)) = C(s, e), p(X_{i,j} = x | E_{i,j} = (s, e)) = B(s, e, x)$$

$$p(E_{i,1}|_s = s | E_{i-1,1}|_s = s') = A(s, s')$$

$$p(E_{i,j} = (s, e) | E_{i,j-1} = (s, e')) = a(s, e, e')$$

L'algorithme est défini au moyen de deux types de fonctions. Les fonctions q et Q permettent de calculer la probabilité de la forme étant donné le modèle considéré. Les fonctions r et R servent à enregistrer les configurations cachées qui maximisent les quantités q et Q . L'algorithme est décrit par le schéma récursif suivant :

– Initialisation

$$q(s, e, i, 1) = c(s, e) \times B(s, e, x_{i,1})$$

$$r(s, e, i, j) = nil$$

– Récurrence

$$q(s, e, i, j) = \max_{e'} q(s, e', i, j-1) \times a(s, e, e')$$

$$\times B(s, e', x_{1,j})$$

$$r(s, e, i, j) = \operatorname{argmax}_{e'} q(s, e', i, j-1) \times a(s, e, e')$$

$$\times B(s, e', x_{1,j})$$

$$Q(s, 1) = C(s) \times q(s, e_f(s), 1, N)$$

$$Q(s, i) = \max_{s'} Q(s', i-1) \times A(s, s') \times q(s, e_f(s), i, N)$$

$$R(s, i) = \operatorname{argmax}_{s'} Q(s', i-1) \times A(s, s') \times q(s, e_f(s), i, N)$$

Finalement, les états expliquant la valeur de $p(x) = Q(s_f, M)$ sont retrouvés en partant de $R(s_f, M)$ en suivant les pointeurs formés par R et r .

Les paramètres qui définissent un PHMM ou un réseau de Markov sont nombreux et ne peuvent être déterminés que par apprentissage automatique à partir d'exemples. Il existe différentes manières de réaliser l'apprentissage des paramètres du modèle. Elles consistent toutes à maximiser la vraisemblance des exemples d'apprentissage. A cet effet, on peut avoir recours à des techniques d'optimisation par gradient ou à des variantes de l'algorithme EM [Devijver]. Ces solutions sont toutefois d'une complexité de calcul importante. Une autre famille d'algorithmes d'apprentissage est constituée de méthodes dirigées par les décisions (DD).

Dans ce cadre, les probabilités des événements caractéristiques du modèle sont estimées par les fréquences empiriques de ces événements pour les états attribués par l'algorithme de Viterbi aux exemples d'apprentissage. L'algorithme DD est itératif et débute avec des valeurs de probabilité issues d'un tirage aléatoire. C'est cet algorithme que nous avons utilisé dans notre application.

3. Modélisation des chiffres manuscrits

Dans le contexte de la reconnaissance de chiffres manuscrits, le PHMM est utilisé pour modéliser la déformation du caractère à reconnaître par rapport à des prototypes. Les probabilités calculées au moyen d'un réseau de Markov serviront ensuite à estimer la vraisemblance de la forme à classer pour chaque classe.

L'application du modèle PHMM à des images de caractères manuscrits consiste à utiliser les super états le long d'une des deux dimensions (ici l'axe vertical) tandis que ces super états émettent des observations (pixels d'une ligne) le long de la dimension complémentaire (ici les lignes). Les observations émises par les états sont les valeurs binaires des pixels. Les états et super-états sont destinés à coder les déformations ayant engendré le caractère à reconnaître.

La figure 2 représente la topologie que nous avons utilisée dans notre application. Les HMM employés ont tous une topologie de type Bakis [Bahl] qui permet de modéliser des distorsions locales d'un signal mono-dimensionnel (dilatation ou contraction). La topologie de Bakis suppose les états totalement ordonnés et n'autorise de transitions d'un état que vers lui-même, son successeur ou le successeur de celui-ci. Ce modèle est dit *gauche-droite* à cause de l'ordonnement total des états. La mise en correspondance du modèle et d'un signal temporel, que l'on appelle aussi alignement temporel, revient à numéroter chaque échantillon du signal par le numéro d'ordre de l'état associé. Lorsque plusieurs numéros successifs sont identiques, on peut considérer que les échantillons correspondants sont de même nature et constituent une répétition d'une portion élémentaire d'un signal prototype (dilatation). En revanche, dès qu'il existe un saut supérieur à 1 entre numéros d'état successifs, on considère qu'un ou plusieurs échantillons du prototype ont été omis (contraction).

Le modèle PHMM tente donc de rendre compte de la production de différentes formes de caractères en décomposant les déformations possibles en déformations locales affectant soit l'ensemble des pixels d'une ligne, soit chacun des pixels d'une ligne indépendamment des pixels des lignes voisines.

Un des principaux défauts des PHMM est l'hypothèse d'indépendance entre probabilités d'émission des lignes (ou colonnes), en particulier consécutives, dès lors que sont connus les super-états correspondants. Cette hypothèse n'est certainement pas réaliste si l'on se contente d'utiliser un modèle tel que celui de la figure 2.

Afin de rendre indépendantes autant que faire se peut les probabilités d'émission nous avons utilisé une topologie dérivée de

celle de la figure 2 dans laquelle les super états codent non seulement une information relative à la déformation suivant l'axe des i mais aussi la déformation suivant les j . Cette topologie est décrite sur la figure 3.

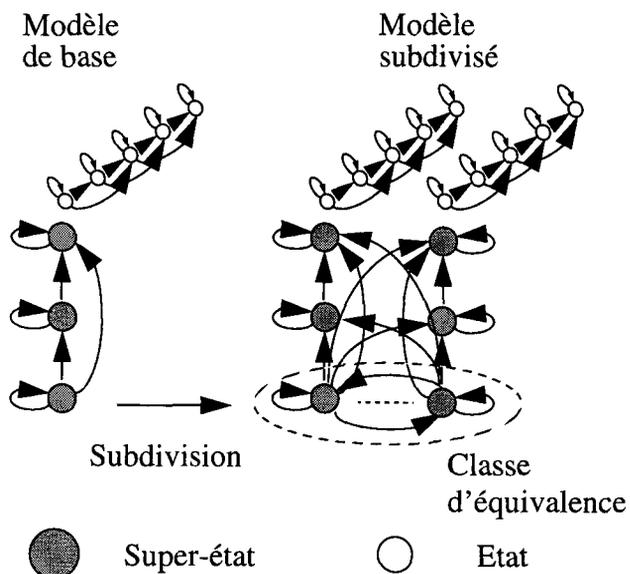


Fig. 3. – Topologie des PHMM de caractères.

Sur cette figure, les super-états s sont groupés en classes d'équivalence c . Autrement dit, un super-état du modèle initial est subdivisé en plusieurs super-états dans le nouveau modèle. Notons l_i et l_{i-1} deux lignes consécutives. La loi de probabilité des lignes conditionnelle aux classes correspondantes s'écrit en effet dans le modèle initial :

$$p(l_i, l_{i-1} | c_i, c_{i-1}) = p(l_i | c_i) \times p(l_{i-1} | c_{i-1})$$

Dans le nouveau modèle, elle devient :

$$p(l_i, l_{i-1} | c_i, c_{i-1}) = \sum_{s_i \in c_i, s_{i-1} \in c_{i-1}} p(l_i | s_i) \times p(l_{i-1} | s_{i-1}) \times p(s_i | s_{i-1}) \times p(s_{i-1})$$

On voit aisément qu'en subdivisant chaque super-état du modèle de Bakis de base en différents super-états regroupés dans une classe d'équivalence, les lignes consécutives sont corrélées au travers du terme $p(s_i, s_{i-1}) = p(s_i | s_{i-1}) \times p(s_{i-1})$.

Chaque classe d'équivalence est destinée à coder une déformation en i alors qu'à l'intérieur de la classe on espère coder différentes déformations en j . Par la même occasion, ce modèle devrait pouvoir rendre compte des multiples façons d'écrire un caractère en associant super-états dans les classes et variantes d'écriture. C'est cependant le rôle de l'apprentissage automatique de rendre compte à travers chaque super-état d'une classe des variations constatées sur l'ensemble d'apprentissage.

Comme le montre la figure 4, l'attribution des états cachés par le PHMM correspond bien à l'interprétation que l'on a voulu donner aux états et aux super-états.

Super-états	Etats
j	i
0	: 7666666666543210
1	: 7665432222221000XXXXXXXXXX...
2	: 765433333333210
2	: 765433333333210XX..XX..XXXX.
3	: 7654211111111110X.....
3	: 7654211111111110X.....
3	: 7654333211111110XXXXX.....
4	: 765433333321110XXX..XXXX.....
5	: 7666666543332100XXXXX.....
5	: 7666666665432100XXXX.....
6	: 766654433322100XX.....XX.....
6	: 766654433322100XX.....XX.....
6	: 766654433322100XX.....XX.....
7	: 76665433332100XXXXXXX.....
8	: 766665432111110XX.....
9	: 7654321000000000

Fig. 4. – Affectation des états et super-états (10 classes de 1 super-état, 8 états par classe).

En effet, des lignes consécutives identiques ou peu différentes dont le nombre peut varier sans changer la nature du caractère reçoivent le même super-état (lignes 11, 12 et 13 par exemple). De même, des pixels consécutifs d'une même ligne ayant le même état sont aussi ceux dont le nombre peut varier sans changement de signification de l'image. Aussi, le modèle du PHMM peut être considéré comme suffisant pour cette attribution. En revanche, il estime assez mal la probabilité de la forme et des états attribués à cause de l'hypothèse d'indépendance entre lignes, notamment consécutives. D'où l'idée d'utiliser un autre modèle pour estimer cette probabilité, ce second modèle opérant à la fois sur l'image et sur les états déterminés par le PHMM. Nous employons à cet effet des réseaux de Markov homogènes d'ordre 3 dont les états sont les couples état/super-état du PHMM et dont les paramètres sont appris au moyen de l'algorithme DD sur les états et super-états attribués par les PHMM.

4. Résultats

Ce modèle a été utilisé pour reconnaître des chiffres manuscrits omni scripteurs recueillis sur les codes postaux de lettres ordinaires. Les images utilisées sont binaires et de taille 16×16 . Les images binaires originelles ont subi une normalisation conservant le rapport d'aspect.

L'apprentissage du PHMM et du réseau de Markov ont été faits sur un ensemble de 4891 images au moyen respectivement de l'algorithme de Viterbi et de l'algorithme DD. Le test a été réalisé sur un ensemble distinct de 4891 images. Le modèle utilisé comporte, pour chaque chiffre, 5 classes d'équivalence de 10 super-états et 8 états par super-état. Ces nombres de classes, d'états et de super-états ont été choisis de manière empirique après essai de plusieurs configurations. La configuration retenue est celle qui donne les meilleurs résultats en reconnaissance. On constate cependant que les performances obtenues varient peu

lorsque les nombres d'états et super-états sont proches des valeurs retenues.

Le taux de reconnaissance sans rejet obtenu est supérieur à 90% ce qui constitue un premier résultat très moyen mais encourageant.

5. Application à la synthèse automatique de chiffres manuscrits

La modélisation par réseau de Markov se prête aisément à une génération automatique d'images de caractères par méthode de Monte Carlo. Le but de cette génération est plus de visualiser le modèle de caractère appris que de synthétiser de véritables images de caractères virtuels.

Les images obtenues (cf. figure 5) sont assez éloignées des exemples d'apprentissage.



Fig. 5. - Chiffres synthétiques.

Ces résultats montrent qu'une modélisation plus adéquate des caractères est à rechercher. En particulier, on s'aperçoit que les observations lignes sont engendrées de façon encore trop indépendante ce qui crée des distortions dans les formes observées. Ce problème pourrait être réglé par une topologie plus complexe et par l'usage d'images multi-spectrales dans lesquelles les pixels pourraient stocker des informations locales de nature topologique ou morphologique.

6. Conclusion

Nous avons présenté une méthode de reconnaissance de caractères fondée sur les modèles de Markov pseudo-2D. Cette méthode a pour avantage de prendre en compte la dimension 2 des images à traiter. Contrairement aux applications des champs de Markov, elle ne sert pas à rechercher une information cachée attachée à

chaque pixel mais plutôt à évaluer la probabilité suivant laquelle un modèle engendre une image. Cette évaluation est faite par un algorithme déterministe et non par relaxation. Un apprentissage automatique des modèles est possible de même que la génération automatique d'images de caractères. L'emploi d'un réseau de Markov permet une meilleure estimation des probabilités de reconnaissance que celle du modèle pseudo-2D. Enfin, les résultats obtenus en reconnaissance montrent que la méthode est viable bien que perfectible.

Une des voies d'amélioration envisageables est l'enrichissement des observations associées à chaque site (chaque pixel). Si notre modèle identifie les observations aux valeurs de luminance binaire des pixels, rien n'interdit d'ajouter à celles-ci des informations mesurées sur un voisinage du pixel considéré (orientation locale du contour, possibilité d'atteindre les bords de l'image sans rencontrer de pixel noir, existence locale d'un trait remarquable tel que hampe ou jambage, barre de t).

BIBLIOGRAPHIE

- [Abend] K. Abend, T.J. Harley, L.N. Kanal, « Classification of binary random patterns », *IEEE Trans. Inform. Theory*, 11, 196, pp. 538-544.
- [Agazzi93a] O.E. Agazzi, S.S. Kuo, « Hidden Markov Model Based Optical Character Recognition in the Presence of Deterministic Transformations », *Pattern Recognition*, vol. 26, n°12, 1993, pp. 1813-1826.
- [Agazzi93b] O.E. Agazzi, S.S. Kuo, E. Levin, R. Pieraccini, « Connected and degraded text recognition using planar hidden Markov models », *Proc. of ICASSP'93*, 1993, pp. V113-V116.
- [Bahl] L.R. Bahl, F. Jelinek, R.L. Mercer, « A Maximum Likelihood Approach to Speech Recognition », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, n°2, 1983, pp. 179-190.
- [Bose] C. Bose, S. Kuo, « Connected and degraded text recognition using a hidden Markov model », *Proc. of the 11th IAPR Int. Conf. on Pattern Recognition*, 1992.
- [Chen] M.-Y. Chen, A. Kundu, J. Zhou, « Off-line handwritten word recognition using a hidden Markov model type stochastic network », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, n°5, 1994, pp. 481-496.
- [Devijver] P.-A. Devijver, M.M. Dekesel, « Algorithmes d'apprentissage d'un modèle markovien d'image », *Actes du 7ème congrès RFIA*, 1987, p. 193-207.
- [Geman] S. Geman, D. Geman, « Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6, 1984, pp. 721-741.
- [Gilloux] M. Gilloux, J.-M. Bertille, M. Leroux, « Recognition of Handwritten Words in a Limited Dynamic Vocabulary », *Proc. of the 2nd International Workshop on Frontiers in Handwritten Recognition*, 1993, pp. 417-422.
- [Levin] E. Levin, R. Pieraccini, « Dynamic planar warping for optical character recognition », *Proc. of ICASSP'92*, 1992, pp. III149-III152.

Manuscrit reçu le 26 Octobre 1995.

LES AUTEURS

Michel GILLOUX



Michel Gilloux est Ingénieur en Chef des Télécommunications, ancien élève de l'École Polytechnique (1977) et ingénieur de l'École Nationale Supérieure des Télécommunications (1982). Il dirige le département Reconnaissance-Vision-Architecture du Service de Recherche Technique de la Poste situé à Nantes où il est responsable des projets de recherche et développement en lecture automatique, notamment dans le domaine de la reconnaissance des adresses postales.