

L'écrit et le document

Traitement des montants numériques de chèques postaux. Une méthode de segmentation basée sur la reconnaissance

Processing of Numerical Check Amounts A Segmentation-Based Approach

par E. LETHELIER, M. LEROUX, M. GILLOUX

SRTP – Département RVA,
10, rue de l'Île Mabon F-44063 Nantes cedex 02

Résumé

Nous présentons un système de lecture automatique des montants numériques des chèques dont le principe repose sur une technique de segmentation des caractères validée par la reconnaissance. Ce système est décrit depuis la phase de localisation du champ montant sur le document numérisé, jusqu'à la génération de la liste des hypothèses de montant. La segmentation, de type explicite, permet de déterminer des zones de coupure potentielles entre caractères et fournit une représentation spatiale des composantes segmentées. Le meilleur chemin de segmentation du montant est déterminé par la combinaison des scores de reconnaissance des caractères, de la vraisemblance de la segmentation et de la probabilité d'apparition de ce montant. Un apprentissage est effectué à partir d'un algorithme d'étiquetage automatique qui améliore nettement les performances des différents modules et a également permis l'intégration d'une classe rejet au niveau du module de reconnaissance. Afin de pouvoir quantifier la robustesse de ce système, nous avons mesuré ses performances à partir d'une base de 10 000 images de montants de chèques réels.

Mots clés : Reconnaissance manuscrite hors-ligne, Segmentation et reconnaissance de caractères, Application aux chèques bancaires.

Abstract

We present an automatic recognition system applied to handwritten numeral check amounts which is based on a segmentation-by-recognition probabilistic model. This system is described from the amount field localization on the document image to the generation of hypotheses. An explicit segmentation algorithm determines potential cuts between characters and provides a spatial representation of segmented components. The best path for the segmentation is determined by the combination of recognition scores, of segmentation likelihoods and of a priori probabilities of amounts. Training is done through a bootstrapping technique, which significantly improves the performances of the different algorithms. It also allows the use of a reject class at the recognition step. The robustness of the system was assessed on a database of 10,000 real cheques images.

Key words : *Off-line handwriting recognition, Characters segmentation and recognition, Application to bank checks.*

1. Introduction

Le domaine de la lecture automatique de l'écriture occupe bon nombre d'équipes de recherche à travers le monde depuis maintenant plusieurs décennies. Après la mise en œuvre de premières applications destinées au traitement de l'écriture dactylographiée, les techniques ont naturellement évolué vers le traitement de l'écriture manuscrite. L'un des premiers secteurs qui a investi dans cette nouvelle technologie est bien entendu celui du tri postal qui a ainsi pu maîtriser l'accroissement du flux de courrier tout en

diminuant les temps d'acheminement. Continuant d'améliorer les performances de ses machines de tri automatique sur le courrier manuscrit [El Yacoubi], La Poste française a également entrepris des études dans le domaine financier. Ces recherches ont permis de réaliser un premier système de lecture appliqué aux montants en lettres des chèques [Leroux] et aboutissent aujourd'hui à une technique de lecture des montants en chiffres. L'originalité de l'application mise en œuvre dérive de la première constatation que l'on peut faire au sujet des images de montants en chiffres de chèques. Contrairement à la représentation du code postal, celle

d'un montant possède une variabilité importante dont il est nécessaire de tenir compte pour la lecture automatique. Elle concerne à la fois le nombre de caractères présents (variable), leur type (numérique ou non numérique) et les liens qui peuvent les relier (voir figure 1).

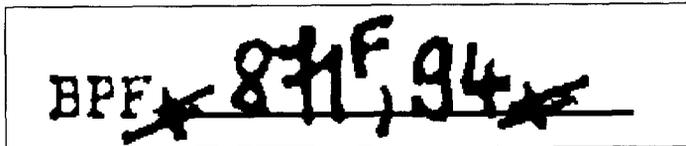


Fig. 1. – Exemple d'une image binarisée de montant chiffres.

La détermination des caractères présents sur l'image suppose une segmentation en caractères individualisés. L'examen de quelques exemples de liens entre caractères permet de constater que cette segmentation ne peut se faire correctement sans mettre en œuvre des connaissances liées à la forme individuelle des caractères (module de reconnaissance). Or, l'utilisation d'un module de reconnaissance ne peut se faire qu'après segmentation. Il existe donc un paradoxe qui s'énonce de la manière suivante : pour segmenter il faut avoir reconnu (les caractères) mais pour reconnaître il faut avoir segmenté. La seule façon de briser cette étreinte fatale consiste à mener de front segmentation et reconnaissance ou, autrement dit, de permettre à la reconnaissance de remettre en cause la segmentation et réciproquement. C'est de cette manière que fonctionne le système décrit dans cet article.

L'interaction entre segmentation et reconnaissance sera faite au travers d'un score qui agrège les vraisemblances des hypothèses de segmentation et de celles de reconnaissance. Par ailleurs, il est souhaitable d'intégrer à ce score toutes les sources de connaissance qui peuvent contribuer à la réalisation d'un choix entre les hypothèses de segmentation. Un certain nombre de régularités propres aux montants en chiffres des chèques (fréquence des montants, fréquence des différentes façons d'exprimer un montant, nature et fréquence des liens préférentiels entre caractères donnés) serviront à cet effet.

La nécessité d'intégrer diverses sources de connaissance dont la plupart correspondent à des régularités statistiques nous a conduits à réaliser le calcul des scores au moyen d'une approche probabiliste. Elle s'inspire de modèles de la théorie de l'information déjà utilisés dans le cadre de la reconnaissance de documents [Kopec], [Breuel], mais dérivés eux-mêmes de travaux appliqués à la reconnaissance de la parole [Leung]. Les paramètres qui définissent les différents modèles probabilistes employés sont appris automatiquement à partir d'une base d'images de chèques étiquetées par programme.

Cet article est organisé de la manière suivante. Après avoir détaillé le modèle statistique sur lequel est fondé notre système, nous en présentons les différentes étapes. Elles concernent la phase de pré-traitement qui a pour effet de normaliser l'image d'entrée, l'étape de segmentation qui fournit une représentation des segments élémentaires suivant les coupures potentielles, la phase de

reconnaissance qui détermine la classe d'appartenance des images des segments parmi un certain nombre de classes candidates et enfin le contrôleur syntaxique qui fournit les hypothèses montants et leurs expressions les plus probables. Nous présentons ensuite les tests réalisés à partir d'une base de 10 000 images de montants réels.

2. Modélisation statistique

L'objectif de la modélisation statistique est de pouvoir intégrer au mieux toute l'information contextuelle disponible de manière, en particulier, à déterminer la segmentation correcte des caractères. Le modèle que nous détaillons dans ce paragraphe repose sur le principe bayésien de la décision. Dans notre cas de figure, il s'agit de construire un estimateur dont l'entrée correspond à l'image I du montant localisé après extraction des données pré-imprimées et segmentation en segments d'images élémentaires. Cette image est enrichie d'annotations indiquant la nature des règles de coupure ayant isolé des segments. La valeur M à estimer représente, quant à elle, le montant réel inscrit sur le document. D'après la théorie bayésienne de la décision, la meilleure estimation du montant M est donnée par le critère du maximum a posteriori :

$$P(\hat{M}/I) = \max_M P(M/I)$$

Dans notre modèle l'image I sera formée de la suite des images des segments élémentaires formés des parties de l'image situés entre deux points de segmentation potentiels successifs. On intercale, entre les images, des symboles représentant les règles de segmentation qui ont conduit à effectuer des coupures : $I = i_1 r_1 i_2, \dots, r_{n-1} i_n r_n$.

On considère que le montant M est exprimé par le scripteur au travers d'une variante V choisie parmi les différentes façons d'exprimer M . Chaque variante est formée d'une suite de caractères : $V = v_1 \dots v_m$. On représente aussi au travers de la variante certains effets de la phase de pré-traitement. Ainsi, 10000 est une variante de 100 Frs dans la mesure où le pré-traitement peut supprimer la virgule en l'assimilant aux traits pré-imprimés du fond de chèque. Le montant lui-même sera une suite de caractères $M = m_1 \dots m_p$ représentant une forme canonique de celui-ci. Ce sera ici la représentation décimale (100,00 pour 100 Frs par exemple). La décomposition de l'image de la variante V en segments élémentaires permet d'introduire un terme de segmentation $S = s_1 \dots s_m$, où m correspond au nombre de caractères de V , qui définit la correspondance entre caractères de V et segments de I et indique pour chaque caractère de V la nature des règles de segmentation qui l'ont éventuellement fragmenté.

$$s_i = r_{d(i)} \dots r_{f(i)} \text{ ou } s = \lambda \text{ (séquence vide)}$$

avec :

$$r_1 \dots r_n = s_1 \cdot s_2 \dots s_m \cdot \text{est l'opérateur de concaténation}$$

En sommant à la fois sur toutes les segmentations et toutes les variantes, nous pouvons écrire :

$$P(M/I) = \sum_S \sum_V P(M, V, S/I)$$

En tout état de cause, nous pouvons établir l'hypothèse que le montant écrit s'explique essentiellement par une segmentation S et une variante V . Soit :

$$\exists V, S \text{ tels que } P(M/I) \approx P(M, V, S/I)$$

En appliquant le théorème de Bayes, nous pouvons écrire :

$$P(M/I) \approx \frac{P(I/M, V, S) \times P(M, V, S)}{P(I)}$$

Puisque $P(I)$ est identique pour tous les montants hypothèses :

$$P(M/I) \propto P(I/M, V, S) \times P(M, V, S)$$

Ce terme se décompose en :

$$P(M/I) \propto P(I/M, V, S) \times P(S/M, V) \times P(M, V)$$

Nous supposons que l'image I et le montant ne sont corrélés que par V et S , d'où :

$$P(M/I) \approx P(I/V, S) \times P(S/M, V) \times P(M, V)$$

Enfin, nous supposons aussi que la segmentation S ne dépend de la valeur du montant M qu'au travers de sa variante V puisqu'elle décrit la façon dont V (et non pas M) est répartie en segments élémentaires. D'où :

$$P(M/I) \approx P(I/V, S) \times P(S/V) \times P(M, V)$$

Cette expression explicite les trois termes à estimer : le terme de reconnaissance $P(I/V, S)$, le terme de segmentation $P(S/V)$ et le terme indiquant la fréquence d'apparition conjointe des montants et de leurs variantes $P(M, V)$.

Il est souhaitable que le module de reconnaissance individuelle des caractères s'applique aux images formées de un ou plusieurs segments élémentaires que détermine la segmentation S , sa tâche étant de fournir, pour chaque classe de caractère v et chaque segment formé d'une suite d'imagettes consécutives, une estimation de la probabilité $P(v/[i_j \dots i_k])$ où $[i_j \dots i_k]$ dénote l'image formée de l'union des imagettes consécutives de rang j à k .

Nous supposons ici en première approximation que la forme des caractères ne dépend que de leur classe et de leur segmentation. Cette hypothèse n'est qu'une approximation puisque, d'une part, ces caractères proviennent d'un même scripteur et sont donc corrélés et que d'autre part, certains chaînages produisent des variations contextuelles dans leur représentation.

Nous pouvons alors exprimer la probabilité conjointe $P(I/V, S)$ par des termes individuels suivant la formule :

$$P(I/V, S) = \prod_i P(i_{d(i)} r_{d(i)} \dots i_{f(i)} r_{f(i)} / v_i, s_i = r_{d(i)} \dots r_{f(i)})$$

Nous faisons une autre hypothèse qui consiste à supposer que la forme de chaque caractère ne dépend pas de la nature des règles de segmentation qui s'appliquent à lui de sorte que :

$$\begin{aligned} P(i_{d(i)} r_{d(i)} \dots i_{f(i)} r_{f(i)} / v_i, s_i = r_{d(i)} \dots r_{f(i)}) \\ \approx P([i_{d(i)} \dots i_{f(i)}] / v_i) \end{aligned}$$

En appliquant la règle de Bayes il vient finalement :

$$\begin{aligned} P(M/I) \approx \prod_i \frac{P(v_i/[i_{d(i)} \dots i_{f(i)}]) \times P([i_{d(i)} \dots i_{f(i)}])}{P(v_i)} \\ \times P(S/V) \times P(M, V) \end{aligned}$$

Cette décomposition fait appel comme souhaité à la probabilité $P(v_i/[i_{d(i)} \dots i_{f(i)}])$ fournie par le module de reconnaissance. Elle fait apparaître des termes de forme $P([i_{d(i)} \dots i_{f(i)}])$ qui définissent, pour chaque suite de segments consécutifs $i_{d(i)} \dots i_{f(i)}$, la probabilité qu'ils forment l'image d'un caractère.

On peut par ailleurs remarquer que la dérivation qui aboutit à la formule précédente ne préjuge pas de l'ensemble des classes v employées dans le module de reconnaissance. Aussi, à côté des classes « naturelles » qui forment le vocabulaire sur lequel sont formées les variantes $V(0, 1, \dots, 9, F, \text{ tiret}, \text{ virgule})$, on pourra demander au module de reconnaissance de reconnaître toute autre classe comme par exemple les suites de plusieurs segments s'étendant sur plusieurs caractères d'une variante. Bien évidemment, l'application correcte et cohérente de la formule de Bayes requiert alors que $P(I_{d(i)} \dots i_{f(i)})$ et $P(v_i, s_i)$ prennent en compte cette extension de la notion de classe de reconnaissance. Nous verrons lors de la description du module de reconnaissance comment se fait cette prise en compte. Nous supposons enfin que la segmentation S est conditionnée terme à terme par les éléments de la variante V , soit :

$$P(S/V) \approx \prod_i P(s_i/v_i)$$

On peut remarquer que cette hypothèse suppose qu'il n'existe pas de cas de sous-segmentation, c'est-à-dire que tous les liens entre caractères sont brisés par la segmentation.

Finalement, l'expression du score de reconnaissance d'une hypothèse (M, V, S) est la suivante :

$$\begin{aligned} P(M/I) \approx \prod_i \frac{P(v_i/[i_{d(i)} \dots i_{f(i)}]) \times P([i_{d(i)} \dots i_{f(i)}])}{P(v_i)} \\ \prod_i P(s_i/v_i) \times P(M, V) \end{aligned}$$

Nous retrouvons dans ce score les différentes sources de connaissance contribuant à la détermination de la segmentation correcte :

- $P(v_i/[i_{d(i)} \dots i_{f(i)}])$ scores de reconnaissance des caractères individuels.
- $P([i_{d(i)} \dots i_{f(i)}])$ probabilités a priori des images à reconnaître par le module de reconnaissance.
- $P(v_i)$ probabilités a priori des classes à reconnaître par le module de reconnaissance.

- $P(s_i/v_i)$ probabilité de la configuration de segmentation pour un caractère d'une variante.
- $P(M, V)$ probabilités a priori des couples variante/montant.

Nous présentons dans la suite de l'article les estimateurs des différentes probabilités entrant dans cette formule ainsi que la stratégie de calcul des scores de reconnaissance.

3. Présentation de la chaîne de traitement

Nous détaillons dans ce paragraphe les modules du système de reconnaissance. Après une description des fonctionnalités de chacun d'entre eux, nous expliquons la façon dont l'estimation probabiliste est mise en œuvre sur trois d'entre eux, à savoir la segmentation, la reconnaissance et la fréquence des couples montant-variante. Cette description applicative s'articule autour de cinq étapes :

- Les pré-traitements : après une localisation du champ montant sur le document, un module d'extraction permet de filtrer les formes qui ne sont pas considérées comme indispensables pour la reconnaissance du montant.
- La segmentation : elle effectue une détection de tous les liens potentiels pouvant correspondre à un chaînage de caractères. Elle fournit également une représentation ordonnée de l'image en une séquence de segments en fonction des points de coupure générés. Enfin, elle estime les termes de forme $P(s_i/v_i)$.
- La reconnaissance : scindée en deux fonctions, elle est en mesure de distinguer l'état de segmentation d'un caractère (pré-reconnaissance) suivant ses configurations métrique et topologique et affecte une probabilité d'appartenance à l'une des classes de son ensemble d'apprentissage pour chaque séquence image fournie par la segmentation (reconnaissance). Cette fonction remplit deux rôles : le premier est de choisir par ses scores la meilleure segmentation des caractères constituant le montant et le deuxième consiste à proposer la classe correspondant le mieux à un symbole en termes de probabilités.
- La fréquence des couples montant/variante : elle a pour objet de pondérer les scores de reconnaissance par la probabilité a priori d'apparition du couple considéré.
- Le filtrage syntaxique : il fournit en sortie du système une liste de montants syntaxiquement corrects relatifs aux composantes reconnues et à la configuration de la segmentation choisie.

Nous utilisons des images de type binaire obtenues à partir d'un numériseur rapide dont la résolution est de 12 points par mm (300 dpi). L'image d'entrée représente la globalité du document.

3.1. PRÉ-TRAITEMENTS

La première étape de ces pré-traitements consiste à localiser le champ montant sur l'image globale du document, au moyen d'un

fenêtrage dont les origines se situent au coin supérieur droit de l'image du chèque. Les dimensions en hauteur et largeur sont relativement sur-estimées afin de prendre en compte toute possibilité de débordement de l'écriture. La procédure d'extraction permet de traiter plusieurs types de formules de chèques. Les deux zones d'informations à localiser et à extraire correspondent à la mention « BPF », ainsi qu'à la ligne de base et/ou un cadre réservés à l'écriture du montant. L'élimination de cette ligne de base s'effectue à l'aide d'un algorithme de suivi de trait qui gère d'éventuelles superpositions avec les traits d'écriture. La mention « BPF » est recherchée et éliminée à proximité de la limite gauche de la ligne de base.

Considérant qu'il existe deux types de données (numérique et non numérique), la seconde étape effectue un filtrage des symboles dits non numériques. Ils concernent les fragments de symboles de très petite taille (bruit), les séparateurs de début et de fin de montant (le plus souvent représentés par un trait), les séparateurs de champs numériques (virgule ou point) et enfin, certains fragments particuliers de symboles telles que les barres de 5 ou de F.

La détection des symboles non numériques est réalisée grâce à une pseudo-reconnaissance utilisant des données structurales, topologiques et métriques des composantes connexes. La description structurale est obtenue par polygonalisation du squelette de la composante avec un codage suivant les directions de Freeman. Les points caractéristiques symbolisant la connexion de deux, trois ou quatre segments viennent enrichir la description. Ce type de description, complétée par les données topologiques et métriques, est tout à fait adapté à la localisation des traits.

La première sous-classe (bruit) est très facile à distinguer grâce à la surface moyenne des composantes connexes de l'image. Pour les autres sous-classes, l'expérience nous a montré que leurs composantes connexes sont caractérisées par des formes simples (trait horizontal ou trait courbé) et des positions particulières (début, fin, haute, basse). Cependant, pour des raisons de ressemblance importante avec les classes de symboles numériques, en terme de forme et de position, le symbole « F » ne fait pas partie d'une sous-classe non numérique.

La détection de la virgule demande, quant à elle, davantage d'informations tant au niveau de sa description que de sa localisation. Les critères sont plus nombreux que pour une détection de trait et, par conséquent, peuvent engendrer des cas de confusion plus nombreux. La localisation de la virgule s'effectue en plusieurs étapes. Une fois l'intégralité de la zone d'intérêt vérifiée, une décision hiérarchique choisit la composante à plus forte ressemblance. Cette démarche a l'avantage de pouvoir prendre en compte les cas où le séparateur francs-centimes est représenté par un point soit d'origine naturelle (scripteur), soit résultant d'une détérioration de la virgule après extraction.

Enfin, le dernier cas traité par ce module concerne les regroupements de parties de symboles fragmentés. Cette fragmentation a deux origines :

- soit naturelle donc dépendante du scripteur (lever de stylo),
- soit résultant des modules de traitement amonts.

La fragmentation dépend en majeure partie de la qualité de l'acquisition de l'image. Pour pallier ce problème, nous utilisons une méthode de détection de chevauchement entre deux composantes consécutives. Pour chaque paire de composantes, un taux de recouvrement est calculé puis comparé à un seuil que nous avons estimé sur un ensemble représentatif d'images de chèques. Cette méthode relativement fruste nous assure, dans la majorité des cas simples de fragmentation, une option de regroupement. Ce critère itératif permet d'effectuer successivement plusieurs regroupements.

3.2. SEGMENTATION DES CARACTÈRES

Deux axes se dégagent des techniques rencontrées dans la littérature abordant le problème de la segmentation : le premier concerne les techniques de segmentation de type implicite où les options de coupures dépendent directement de la reconnaissance. Autrement dit, la segmentation fait partie des primitives extraites du caractère et qui sont soumises à la reconnaissance. Ces méthodes ont souvent la particularité de séparer l'image en une succession de fenêtres horizontales [Gillies], [Martin], [Matan].

La seconde voie explore, quant à elle, des techniques où la décision de segmentation précède la reconnaissance. Dans ce cas, la segmentation est dite explicite du fait de son action antérieure à la reconnaissance [Shridhar], [Fenrich], [Fujisawa], [Kimura]. Son élaboration est généralement plus coûteuse, car elle requiert une description de primitives de niveau supérieur, nécessaire pour décider de la meilleure segmentation. Toutefois, les configurations de segmentation sont plus précises en terme de localisation. C'est une notion très importante lorsque les types de chaînages sont multiples. Ce qui est d'ailleurs le cas dans le contexte des chèques manuscrits. C'est la raison pour laquelle nous avons choisi de développer notre application autour de ce deuxième principe.

Notre méthode de segmentation repose sur la détection du tracé de l'écriture pouvant correspondre à un lien inter-caractères. Un processus de contrôle assure une séparation cohérente de la composante connexe d'origine en deux sous-composantes distinctes. Afin d'obtenir le maximum d'efficacité, nous avons recensé tous les types de chaînages rencontrés sur un montant manuscrit (voir figure 2). Il apparaît alors deux familles de liens :

- les liens simples, où il n'existe qu'une connexion entre deux traits d'écriture (A, B),
- les liens multiples (C).

En ce qui concerne la détection de ces liens, la technique employée consiste à déterminer les extrema locaux sur deux informations bi-dimensionnelles que sont le contour et le profil des composantes connexes. Ainsi, chaque vallée (resp. pic) produit une hypothèse de coupure de type descendant (resp. ascendant) autour de cet extremum [Lecolinet].

Disposant d'une image dont les composantes ont subi des segmentations multiples (fig. 3a), le rôle du filtrage consiste à limiter

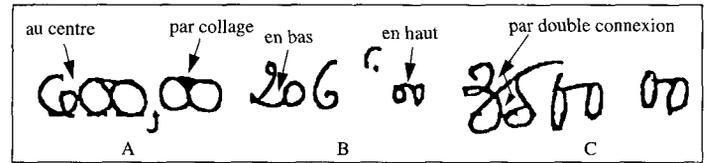


Fig. 2. - Types de chaînages.

les différentes propositions de coupure générées par le segmenteur (fig. 3b). Les critères de validation de points reposent essentiellement sur la valeur surfacique de chacune des sous-composantes et sur les valeurs moyennes (avant et après segmentation). Ce processus est réinitialisé après chaque regroupement effectué et permet de diminuer pratiquement de moitié le nombre d'hypothèses de coupures initiales. Enfin, une étape de normalisation de la représentation des sous-composantes est appliquée (fig. 3c). Elle permet de reproduire un ordre logique proche du tracé, tout en linéarisant la représentation du graphe des sous-composantes. Cette simplification du graphe est utile dans la suite des traitements, puisqu'elle conditionne le nombre final d'hypothèses de segmentation qui sont soumises à la reconnaissance. Le fait de linéariser le graphe limite par conséquent l'espace de recherche du chemin de segmentation optimal après la reconnaissance. Cette particularité n'est en fait qu'un compromis entre la qualité de la segmentation et le coût calculatoire qu'elle engendre.

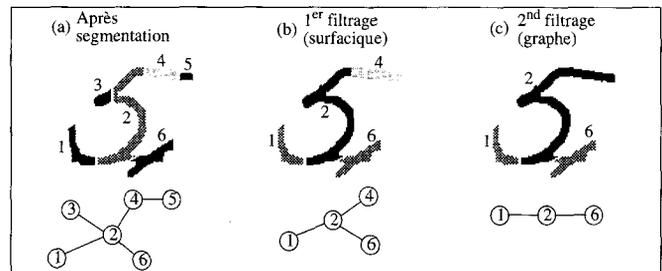


Fig. 3. - Filtrage de la segmentation.

C'est à partir du résultat de ce filtrage (fig. 3c) que les mesures de segmentation ont été effectuées. Grâce à l'algorithme d'étiquetage automatique que nous avons développé, nous avons estimé $P(s_i/v_i) = P(s_i, v_i)/P(v_i)$ sur l'ensemble des classes de caractères faisant partie de l'ensemble d'apprentissage du module de reconnaissance, soit plus de 10 000 images de montants, en assimilant ce terme à la fréquence des couples (s_i, v_i) . Pour une classe donnée, les mesures prennent en compte deux états :

- le nombre de points de coupure « internes » au caractère (1, dans le cas de la fig. 3c),
- la présence d'un point de coupure « externe » au caractère, dans le cas d'un chaînage avec le caractère (numérique ou non numérique) suivant.

Les valeurs mesurées sont très variables d'une classe à l'autre, suivant les configurations de segmentation et de chaînage. Par

conséquent, l'estimateur fournit une information contextuelle pertinente bien que nous ayons arbitrairement choisi de prédire la présence d'un point de segmentation par le caractère à gauche de celui-ci alors que les deux caractères de part et d'autre contribuent à cette prédiction.

3.3. PRÉ-RECONNAISSANCE ET RECONNAISSANCE

Disposant du graphe de segmentation détaillé au paragraphe précédent, le module de reconnaissance fournit une liste de scores suivant le nombre d'hypothèses de segmentation du montant. L'expérience a montré que la procédure de segmentation et les filtres associés ne sur-segmentaient pas plus de deux fois un caractère. Cette remarque nous a permis de limiter le nombre de regroupements successifs entre sous-composantes de même famille (ayant appartenu à la même composante avant le processus de segmentation). En dehors des points de coupure générés par le module de segmentation et ceux correspondant à une option de regroupement, tout point de segmentation naturel est considéré comme valide. Dans ce cas, il n'est pas remis en cause par la reconnaissance (cf fig. 4). A partir du graphe de représentation de l'image associant les composantes et les points de segmentation, le système est en mesure de générer les différentes combinaisons d'associations de sous-composantes segmentées.

Le premier processus, dit de pré-reconnaissance, exprime la probabilité $P(i_j)$ de la $j^{\text{ème}}$ sous-composante de l'image. Il permet de comparer deux types de sous-composantes :

- celles correspondant à une partie d'un caractère segmenté (sur-segmentation),
- et celles ayant l'aspect d'un caractère à part entière (segmentation correcte).

Les primitives utilisées sont pratiquement les mêmes que celles servant au filtrage de la segmentation, enrichies d'un degré d'appartenance à chacune des trois régions (supérieure, médiane et inférieure) de l'image. Cette séparation est obtenue en fonction de l'importance du corps de ligne (zone médiane). L'influence du terme $P(i_j)$ est prépondérante, notamment lorsque, pour un caractère incorrectement segmenté, le score fourni par la reconnaissance est ambigu. La probabilité du terme de pré-reconnaissance est calculée par comparaison à des mesures statistiques obtenues après regroupement des images dans des classes d'équivalence par la méthode des nuées dynamiques.

La fonction de reconnaissance comporte quant à elle deux modules distincts. Le premier utilise un vecteur codant les différentes concavités de l'image et leur surface relative et un réseau de neurones composé d'une base de fonctions radiales (RBF) [Lemarié]. Le deuxième est un réseau de neurones de type TDNN [Le Cun].

Ces deux réseaux ont été entraînés sur la même base d'apprentissage (10 000 symboles). Leurs sorties correspondent à 14 classes de caractères les plus représentés, à savoir :

- les 10 classes numériques,
- 4 classes non numériques : « F », Virgule, Trait, et Bruit.

De plus, une classe de rejet a également participé à la phase d'apprentissage; elle concerne en particulier les cas de sous-segmentation (n -uplets) de symboles.

Les sorties des réseaux de neurones sont assimilés aux probabilités recherchées [Lippman]. De plus, on ne prend en compte que les quatre meilleures valeurs du vecteur de sortie des réseaux. Les hypothèses de rang supérieur ne sont pas considérées dans l'exploration des différentes hypothèses de montant possibles.

Finalement, pour reproduire le modèle de la reconnaissance décrit au deuxième paragraphe, la valeur de la probabilité a priori de la classe $P(v_i)$ est calculée sur l'ensemble d'apprentissage de la reconnaissance. Son importance se justifie du fait de la non homogénéité des classes composant la base d'apprentissage du modèle de reconnaissance. Comme pour la segmentation, la base d'apprentissage a été obtenue au moyen de notre processus d'étiquetage automatique.

3.4. ESTIMATION DE LA PROBABILITÉ A PRIORI DE M ET V

L'estimateur des termes $P(M, V)$ doit opérer sur des couples de séquences de symboles. Rappelons que M est la forme décimale des montants $M = m_1 \dots m_n$ et que la variante V est constituée d'une suite d'étiquettes de caractères appartenant à l'alphabet $\{0, \dots, 9, F, \text{virgule}, \text{trait}, \text{bruit}\}$.

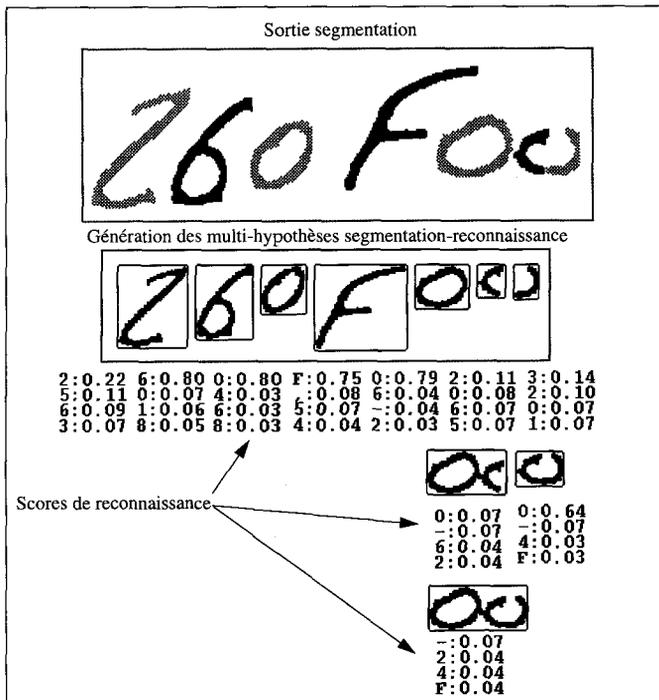


Fig. 4. – Scores de reconnaissance en fonction de la segmentation.

L'estimateur utilisé est un transducteur d'états finis stochastique [Gilloux], c'est-à-dire un transducteur d'états finis dont les transitions sont assorties de probabilités. Ces probabilités sont obtenues par apprentissage automatique à partir d'un ensemble de couples exemples. La technique d'apprentissage est une extension de celle employée pour les modèles de Markov cachés. Le nombre d'états du transducteur est fixé avant l'apprentissage mais sa valeur optimale est déterminée de manière empirique après plusieurs essais. Enfin, toutes les transitions entre états sont possibles dans la topologie retenue (modèle ergodique).

Si ce transducteur donne bien une estimation de $P(M, V)$, il peut cependant affecter une probabilité non nulle à des couples dont la composante M n'est pas une forme canonique de montant. C'est pourquoi il y a nécessité de mettre en œuvre un filtre syntaxique pour éliminer ce type d'hypothèse.

3.5. FILTRAGE SYNTAXIQUE

Le contrôle syntaxique s'opère en deux temps :

- une phase de construction des hypothèses montants,
- une phase de filtrage des montants syntaxiquement incorrects.

Le processus de construction des montants repose sur un algorithme de Viterbi modifié [Chen], [Forney], qui calcule les K meilleures configurations de segmentation-reconnaissance du graphe.

Les valeurs comparées correspondent à la combinaison des scores de segmentation et de reconnaissance de chaque imagerie, comme le montre la figure 5. Chaque nouvelle hypothèse montant construite est analysée par le filtre syntaxique qui fournit au transducteur le montant sous forme d'un couple d'expressions (canonique, variante). La probabilité fournie par le transducteur est ensuite combinée au score de segmentation-reconnaissance. On obtient ainsi la liste finale des meilleures hypothèses de reconnaissance du montant.

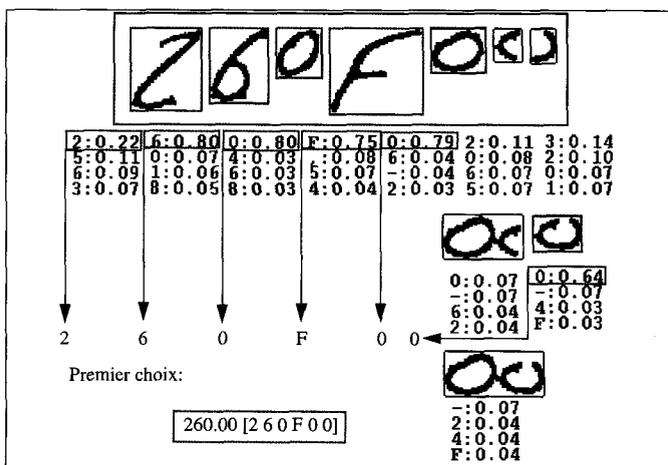


Fig. 5. - Génération de la meilleure hypothèse reconnaissance.

4. Evaluation de la chaîne

Les résultats que nous présentons ont été obtenus à partir d'une base de test de 10 000 images de montants de type « tout venant », d'une résolution de 300 dpi. Le tableau 1 présente les performances globales du système, où le taux de lecture (TL) est exprimé en fonction de la longueur de la liste des hypothèses. Etant donné l'absence de rejet global de reconnaissance, le faible taux de rejet reporté dans ce tableau ne concerne que le rejet issu des pré-traitements. Lorsque la liste d'hypothèses montant ne comporte pas la valeur correcte, cela correspond à une confusion de lecture. Le tableau 2 répertorie les principales confusions du système.

Tableau 1. - Performances de reconnaissance sur 10,000 images.

Reco.	$TL[1]$	$TL[2]$	$TL[5]$	$TL[10]$	$TL[16]$	Rejet	Conf.
%	60.4	72.8	81.2	85.5	87.1	0.19	12.7

Tableau 2. - Principales confusions.

Traitement	%
Extraction	7.5
Fragmentation	8.8
Sous-segmentation	10.5
Pré-reconnaissance	7.4
Reconnaissance	57.9
Non numér.	31.7
Numérique	26.2

5. Analyse

Le système de reconnaissance peut être évalué suivant deux critères :

- la capacité de générer le montant correct dans la liste d'hypothèses,
- la capacité de proposer le montant correct dans les premiers choix.

Il est important de pouvoir distinguer ces deux critères, car ils vont entraîner deux méthodologies d'améliorations différentes :

- une modification des algorithmes de traitements implémentés,
- une intégration des connaissances a priori du contexte.

Dans le premier cas, les erreurs proviennent principalement du module de reconnaissance totalisant plus de 57% des problèmes. Dans le second cas, les différences des scores de reconnaissance entre la classe correcte et les suivantes ne sont pas suffisantes pour pouvoir obtenir la meilleure position dans la liste des hypothèses montant après pondération syntaxique. Par exemple, une confusion du symbole 9 par le symbole 3 entraînera une hypothèse plus forte en terme de probabilité pour « 300,00F » que pour « 900,00F ». De plus, cet effet est accentué en cas de confusions entre les classes numériques et non numériques. Nous notons ici l'intérêt de la pondération du score de reconnaissance par celui de la pré-reconnaissance. Néanmoins, nous avons pu vérifier et valider le principe de segmentation-reconnaissance développé.

Un autre point important concerne la notion du rejet reconnaissance. Son implémentation permet de réduire considérablement les confusions en cas de sous-segmentation de symboles, mais n'autorise pas encore de rejet reconnaissance au niveau du montant global. Cette technique est d'ailleurs en cours de développement.

Enfin, nous avons pu remarquer la sensibilité de l'analyseur syntaxique en fonction du contexte, ce qui prouve l'importance d'une base d'apprentissage conséquente afin de prendre en compte la variabilité d'écriture des montants.

6. Conclusion et perspectives

Nous avons présenté notre système de segmentation-reconnaissance appliqué aux montants numériques des chèques. Notre approche consiste à combiner les sorties des différentes étapes de traitement suivant le modèle statistique que nous avons établi. Étendu à 14 classes de symboles numériques et non numériques, le moteur de reconnaissance est relativement bien adapté au contexte réel des chèques. Au vu des résultats encourageants de chacun des modules de traitement et par conséquent ceux relatifs au système global, l'analyse précise du comportement de ces processus permettra sans aucun doute d'améliorer les performances actuelles. En outre, cette méthode originale multi-hypothèses segmentation-reconnaissance, peut très facilement s'intégrer dans des applications de traitement de l'écriture manuscrite telle que la reconnaissance de mots cursifs ou bâton.

BIBLIOGRAPHIE

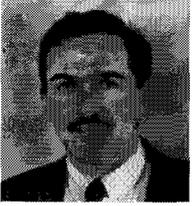
[Breuel] T.M. Breuel, « A System for the Off-Line Recognition of Handwritten Text », *Proc. of ICPR*, Vol. 2, Jerusalem, Israel, November 1994, pp. 129-133.

- [Chen] M.-Y. Chen & al., « Off-Line Handwritten Word Recognition Using Hidden Markov Model », *Proc. of USPS ATC*, Vol. 1, USA, 1992, pp. 563-577.
- [El Yacoubi] A. El Yacoubi, J.-M. Bertille, M. Gilloux, « Towards a More Effective Handwritten Word Recognition System », *Proc. of IWFHR-IV*, Taipei, Taiwan, Dec. 1994, pp. 378-385.
- [Fenrich] R. Fenrich & S. Krishnamoorthy, « Segmenting Diverse Quality Handwritten Digit Strings in Near Real-Time », *Proc. of USPS ATC*, Vol. 1, USA, 1990, pp. 523-537.
- [Forney] G.D. Forney Jr, « The Viterbi Algorithm », *Proc. of the IEEE*, Vol. 61, N°3, March 1973.
- [Fujisawa] H. Fujisawa, Y. Nakano & K. Kurino, « Segmentation Methods for Character Recognition : From Segmentation to Document Structure Analysis », *Proc. of the IEEE*, Vol. 80, N°7, July 1992, pp. 1079-1092.
- [Ganzberger] M. J. Ganzberger & al., « A System for Handwritten Address Interpretation », *Proc. of USPS ATC*, Vol. 1, 1992, pp. 337-351.
- [Gillies] A.M. Gillies, « Cursive Word Recognition Using Hidden Markov Models », *Proc. of USPS ATC*, Vol. 1, USA, 1992, pp. 557-562.
- [Gilloux] M. Gilloux, « Automatic Learning of Word Transducers from Examples », *Proc. of 5th CECACL*, Berlin, Germany, April 1991, pp. 107-112.
- [Hull] J.J. Hull, T.K. Ho, J.T. Favata, V.K. Govindaraju & S.N. Srihari, « Combination of segmentation-based and wholistic handwritten word recognition algorithms », *Proc. of IWFHR-2*, Bonas, France, September 1991, pp. 229-240.
- [Kimura] F. Kimura & M. Shridhar, « Segmentation-Recognition Algorithm for ZIP Code Field Recognition », *Machine Vision and Application*, 5, 1992, pp. 199-210.
- [Kopeck] G.E. Kopeck & P. A. Chou, « Document Image Decoding Using Markov Source Models », *PAMI*, Vol. 16, N°6, June 1994, pp. 602-617.
- [Lecolinet] E. Lecolinet, « A Grapheme-Based Segmentation Technique for Cursive Script Recognition », *Proc. of ICDAR'91*, Vol. 2, St Malo, France, 1991, pp. 740-748.
- [Le Cun] Y. Le Cun & al, « Handwritten digit recognition with back-propagation network », *Neural information processing systems*, Vol. 2, Morgan Kaufmann 1990.
- [Lemarié] B. Lemarié, « Practical implementation of a Radial Basis Function Network for Handwritten Digit Recognition », *Proc. of ICDAR'93*, Tsukuba, Japan, 1993, pp. 412-415.
- [Leroux] M. Leroux, J.-C. Salomé, J. Badard, « Recognition of Cursive Script Words in a small Lexicon », *Proc. of ICDAR'91*, Vol. 2, St Malo, France, 1991, pp. 774-482.
- [Lethelier] E. Lethelier, M. Leroux, J.-F. Couthouis, « Automatic Processing of Numerical Amounts on Postal Cheques », *Proc. of 1st European Conf. on Postal Technologies (JetPoste'93)*, Vol. 2, Nantes, France, June 1993, pp. 697-704.
- [Leung] H.C. Leung & al, « Speech Recognition Using Stochastic Explicit-Segment Modeling », *Proc. of 2nd European Conference on Speech Communication and Technology*, 1991, pp. 931-934.
- [Lippmann] R.P. Lippmann, « Neural Networks, Bayesian a posteriori Probabilities and Pattern Classification », *Proc. of NATO Advanced Studies, From Statistics to Neural Networks*, Les Arcs, France, 1993, pp. 1-7.
- [Martin] G.L. Martin & M. Rashid, « Recognizing Overlapping Hand-Printed Characters by Centered-Objects Integrated Segmentation and Recognition », *Proceedings of IJCNN*, 1991, pp. 504-511.
- [Matan] O. Matan & C.J.C. Burges, « Multi-Digit Recognition Using A Space Displacement Neural Network », *Proc. of IJCNN*, 1991, pp. 488-495.
- [Shridhar] M. Shridhar & A. Badreldin, « Recognition of Isolated and Simply Connected Handwritten Numerals », *Pattern Recognition*, Vol. 19, N°1, 1986, pp. 1-12.
- [Srihari] S. N. Srihari, E. Cohen, V. Govindaraju & A. Shekhawat, « Determining Delivery Point Codes on Handwritten Addresses », *Proc. of USPS ATC*, Vol. 1, USA, 1992, pp. 321-335.

Manuscrit reçu le 2 Octobre 1995.

LES AUTEURS

Michel GILLOUX



Michel Giloux est Ingénieur en Chef des Télécommunications, ancien élève de l'Ecole Polytechnique (1977) et ingénieur de l'Ecole 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.

Manuel LEROUX



Manuel Leroux, diplômé d'un doctorat d'Informatique en 1991, il occupe depuis 1992 un poste d'ingénieur recherche au Service de Recherche Technique de La Poste (SRTP). Au sein du département Reconnaissance-Vision-Architecture, il est responsable de la lecture automatique des chèques.

Edouard LETHELIER



Edouard LETHELIER prépare actuellement une thèse sur la reconnaissance de l'écriture manuscrite appliquée à la lecture des montants numériques des chèques.