

Une étude sur le choix des caractéristiques pour la représentation de caractères imprimés

Jean-Luc Henry

Equipe de Traitement d'Images, Vision et Parallélisme

Université des Antilles et de la Guyane

Département de Mathématiques et Informatique

97159 Pointe-à-Pitre, Guadeloupe

Jean-Luc.Henry@univ-ag.fr

Résumé

La représentation de caractères de façon fiable avant traitement (reconnaissance, apprentissage, classification, etc.) n'est pas totalement résolue. La difficulté subsiste dans le choix des critères pour caractériser une forme de façon non ambiguë. Aucune méthode aussi sophistiquée soit elle, ne peut à elle seule résoudre tous les problèmes. L'évolution de la recherche dans ce domaine s'oriente vers une tendance à la coopération de méthodes.

Cet article présente une étude sur le choix de la meilleure caractéristique pour représenter un caractère. Après avoir appliqué à un grand nombre de caractères nos algorithmes de recherche de contour, de génération de profils et de projections (partielles ou totales), nous arrivons à la conclusion qu'il est nécessaire de combiner plusieurs caractéristiques pour représenter un caractère.

Abstract

The representation of characters in a reliable way before processing (recognition, training, classification, etc.) is not completely resolved. The difficulty subsists in the choice of criteria to characterize a pattern in a non ambiguous way. No method as sophisticated is it, can not solve alone all problems. The evolution of the research in this field orient toward a tendency to the cooperation of methods. This paper presents a study on the choice of the best characteristic to represent a character. After having applied on a large number of characters our algorithms of contour research, of profiles and projection generation (partial or total), we arrive to the conclusion that it is necessary to combine several characteristics to represent a character.

1 Introduction

Dans cette étude, nous cherchons une description non ambiguë et la plus complète possible des formes de caractères. Nous avons exclu les primitives car elles nous paraissent moins robustes que les attributs et elles sont trop sensibles aux bruits et aux ruptures de tracés. Ce choix

s'avère prépondérant dans le cas où l'on envisage d'améliorer les critères de l'étape de reconnaissance du système. Les attributs ne nécessitent ni d'interprétations ni de phase de segmentation particulière ; nous allons les évaluer en nous appuyant sur une étude expérimentale. Nous avons limité notre étude aux caractéristiques issues du contour, des profils et des projections (partielles ou non). Nous fixons le même protocole pour tous les attributs afin de mesurer leur pouvoir discriminant. Nous utilisons pour chaque catégorie de tests un vecteur de caractéristiques de dimension 64.

La méthode de comparaison utilisée est celle des 5-ppv, car les k plus proches voisins sont des outils suffisamment éprouvés et robustes.

2 Recherche d'une description non ambiguë des formes de caractères

A partir d'une étude expérimentale, nous expliquons les choix des attributs utilisés. Nous considérons que les attributs sont plus robustes que les primitives. En effet, il n'y a pas d'erreurs dans l'extraction d'un attribut alors que la segmentation d'une primitive peut échouer et conduire à une fausse interprétation du caractère. Le choix des attributs par rapport aux primitives s'explique par l'objectif d'établir une coopération entre l'étape de reconnaissance et l'étape de correction contextuelle du système. En effet, il serait alors possible de modifier les décisions prises par l'étape de reconnaissance (à partir des attributs) en fonction des résultats fournis par l'étape d'analyse contextuelle (selon les échecs recensés). En outre, l'utilisation de primitives mal reconnues induirait de fausses corrections. En utilisant plutôt des attributs nous facilitons la recherche des causes d'erreurs dans la chaîne de traitement de l'information.

Notre expérimentation va être construite sur un échantillon de polices différentes (environ 200 fontes) sur plusieurs déclinaisons (gras, italique, inclinaison et homothéties différentes, ...) soit un total de 10700 caractères tous différents. Elles proviennent de pages tirées du livre de référence des peintres en lettres et des

imprimeurs. Les pages ont été numérisées, les caractères isolés puis contrôlés un à un et leur identification a été saisie manuellement avant d'être stockée dans un fichier. Chaque caractère, de part sa description, est particulier. En effet, une police donnée ne peut contenir deux fois la même représentation graphique pour un caractère ; ceci a pour conséquence, que tous les caractères de la base ont le même nombre de représentations différentes.

3 Un opérateur efficace pour différencier les formes - le contour

Le contour contient une information structurelle suffisamment complète pour permettre de décrire un caractère. Cette information est très sensible à des déformations et à des bruits dus à la numérisation, comme les ruptures de tracés, par exemple. Le contour a connu plusieurs regains d'intérêt avec l'appariement des contours codés sous forme de chaînes grâce aux travaux de Freeman [1].

Le contour est extrait en parcourant la frontière intérieure de la forme du caractère [2][3]. Afin d'éviter le bruit dû à la discrétisation, nous introduisons la notion de tangente sur plusieurs points du contour. Pour une meilleure compréhension des données, nous avons stocké les normales au contour (cf. Figure 1) plutôt que les tangentes. Le coefficient de lissage qui correspond au nombre de points du contour entrant dans le calcul de la tangente est fixé suivant le nombre de points du contour ; il dépend donc de la taille du caractère. Les grands caractères qui présentent des défauts importants sur le contour sont donc plus lissés. On ne retient dans la succession des normales que l'angle des tangentes avec le repère initial. Les valeurs des angles stockées en degré sont ramenées de 0 à 180° afin d'être sauvegardées, du point de vue informatique, dans des octets.

Pour pouvoir comparer des caractères de tailles différentes, les contours sont quantifiés sur 64 valeurs.

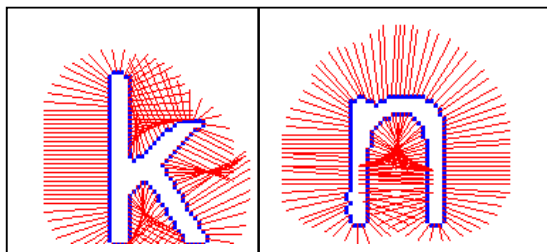


Figure 1. Normales d'un contour.

4 Une nouvelle approche pour les caractères dégradés - le profil externe

Les profils externes d'une forme, très utilisés pour le manuscrit [4] et les caractères chinois [2], sont sûrement des mesures plus robustes que le contour, mais on doit s'attendre à un pouvoir discriminant moindre puisque les cas d'ambiguïtés sont nombreux.

On obtient les profils haut, bas, gauche et droit du caractère en mesurant la distance séparant chaque pixel du caractère au bord du cadre (cf. Figure 2). Le profil est dit consolidé dans la mesure où l'on effectue un calcul d'histogrammes sur 16 valeurs, de chaque côté, de façon à obtenir un vecteur de caractéristiques de 64 composantes. Le profil, comme le contour, est quantifié sur 64 valeurs. Ainsi, les petites ruptures de tracés n'influenceront pas les valeurs moyennes (cf. Figure 3). On comparera cette mesure avec celle du contour réel.

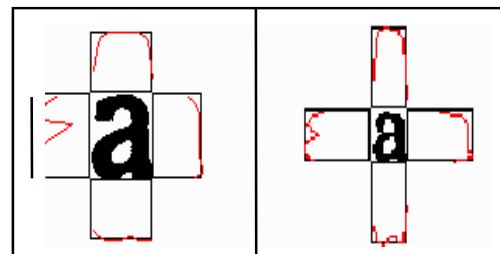


Figure 2. Profils nord, sud, est et ouest des caractères "a". Les profils ont peu varié bien que le deuxième "a" présente une liaison.

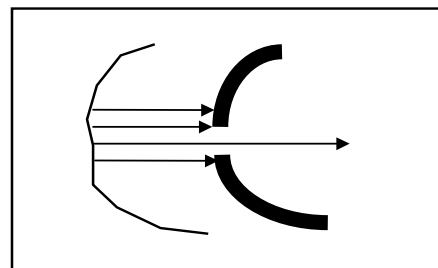


Figure 3. Le renforcement du profil par une moyenne sur plusieurs valeurs rend cette mesure insensible aux petites ruptures de tracé.

5 Gestion du cas spécifique des caractères mono-fontes - les projections totales

L'utilisation des projections après une initialisation, dans un premier temps, pour l'imprimé a été généralisée au chinois, à l'arabe et maintenant au manuscrit [5][6]. Cette mesure globale est prise sur 4 directions 0°, 45°, 90° et 135°, puis quantifiée sur 1 octet et ramenée sur 16 valeurs (cf. Figure 4). Elle consiste à projeter dans chaque

direction le caractère et à effectuer un calcul d'histogramme de la répartition des pixels sur l'axe orthogonal à la direction. Cette évaluation totalement indépendante de la structure peut présenter des ambiguïtés si on ne considère qu'une direction ("u","n") en projection verticale ou ("p","q","F") en projection horizontale). Les ambiguïtés en utilisant les deux directions principales horizontale et verticale sont plus rares, mais elles concernent les caractères qui possèdent majoritairement des traits obliques ("S","5"; "2","Z"; "X","K","Y"; ...); d'où la nécessité d'utiliser les projections obliques à 45° et 135° même si leur utilité ne peut être prouvée de façon formelle. Les substitutions possibles comme ("h","b") ("s","8","a") démontrent les limites de cette mesure. En effet, les projections ne donnent que très peu d'informations sur le nombre d'intersections et sur la position relative des traits alignés côte à côte. Ceci induit un nombre élevé de substitutions du type "h" et "b" ou "8" et "a" qui présentent les mêmes projections horizontales et verticales. Pour compléter les informations apportées par les projections, il faut réintroduire la notion d'intersection; ce qui est fait avec les projections partielles.

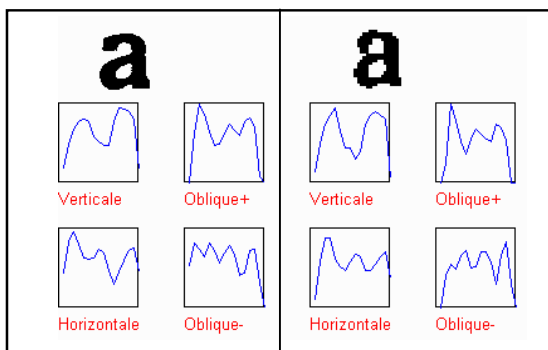


Figure 4. Projections sur 4 directions 0°, 45°, 90°, 135°.

6 Une technique d'extraction d'informations structurelles – les projections partielles

La combinaison des profils et des projections totales revient à mesurer la distance comprise entre le bord du cadre et le caractère ainsi que le nombre de pixels dans une direction donnée. La présence de boucles, la taille et la position de trous ne sont pas mesurées avec suffisamment de précision. L'épaisseur importante des traits peut minimiser la présence d'une boucle. Il faut donc réintroduire la notion de boucles par le nombre d'intersections.

Nous présentons deux méthodes simples et robustes qui permettent d'obtenir des informations structurelles par le biais des projections partielles.

La première consiste à effectuer des projections partielles suivant chaque intersection en fonction de la direction du parcours. On calcule le nombre de pixels dans une direction tant que l'on reste sur le même trait. Dès que l'on passe à un autre trait, on ouvre un autre histogramme et ceci pour chaque intersection (cf. figure 5). Pakker [7] a analysé les limites et les performances des projections partielles sur les caractères imprimés. Nous ne nous intéressons ici qu'aux traits et non aux espaces qui contiennent les lettres. On peut se demander si les projections horizontales et verticales partielles sont-elles suffisantes pour dissocier toutes les formes? Elles sont cependant, souvent utilisées dans l'extraction des caractéristiques.

La quantification, sur huit valeurs seulement, des histogrammes tend à lisser les pics locaux dus aux bruits ou à de petites intersections superflues. La quantification rend les mesures moins sensibles aux micro-ruptures du tracé ou aux jointures de traits.

L'autre méthode consiste à calculer les projections partielles sur des régions bien définies du caractère en découpant l'image en sous régions. On retrouve cette même idée dans les travaux sur le « shadow-code » pour l'authentification des signatures [8].

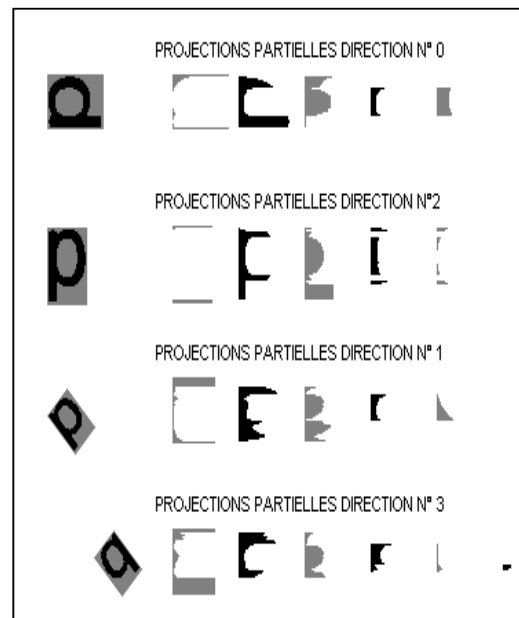


Figure 5. Projections partielles sur quatre directions normalisées sur huit valeurs de "p".

De nombreux tests ont permis de mettre en évidence l'intérêt d'un découpage en six régions caractéristiques pour les caractères imprimés simples. Les travaux de Pakker [7] localisent ce découpage, grâce à un calcul d'histogrammes sur les positions des segments de droites

dans les caractères. Ces histogrammes sont obtenus à partir d'une base importante de caractères imprimés. Ils montrent la distribution des segments par rapport à la hauteur et à la largeur des caractères.

7 Discussion et résultats

Le contour

Les calculs ont généré une table de substitutions, établie sur la base de tests de 10700 caractères. A partir des 5-ppv et en utilisant uniquement les contours, nous obtenons un taux de reconnaissance de 88% ce qui peut sembler efficace pour une telle mesure. Nous remarquons que les couples de caractères classiques comme ("5","S") et ("2","Z") présentent les taux de substitution les plus élevés (49% et 35%). Si nous faisons abstraction des substitutions élémentaires ("5","S") et ("2","Z") nous atteignons un taux de reconnaissance de 90% sur des formes totalement différentes. Cependant, des erreurs très grossières subsistent comme ("e","c"), ("l","j") et même sur des formes totalement différentes ("m","h","k") par exemple. Cela vient du fait qu'un même caractère peut présenter des contours radicalement différents si la structure est modifiée.

Le contour est donc un très bon opérateur pour différencier des formes, mais il peut être à l'origine d'erreurs très grossières que d'autres caractéristiques ne feraient pas apparaître. Les substitutions sont très nombreuses et proviennent essentiellement de la rupture des tracés ou de la jointure des traits (cf. Figure 6). De même, les formes composées nécessitent un prétraitement pour repérer la forme principale.

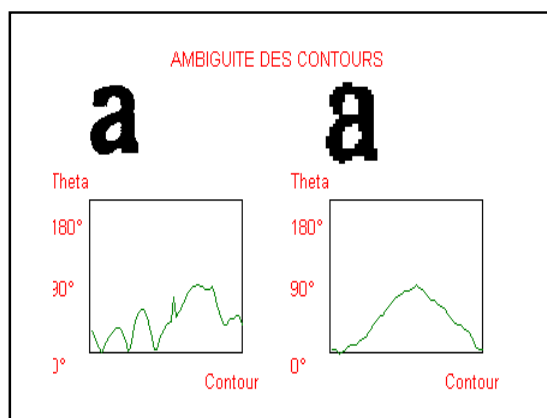


Figure 6. Sensibilité du contour aux défauts de structure (jointure).

Les profils externes

La table de substitutions calculée selon le même protocole que celle concernant le contour met bien en valeur les ambiguïtés. Une simple consultation de la base d'apprentissage démontre qu'il est facile de trouver des

caractères différents qui possèdent des profils presque équivalents.

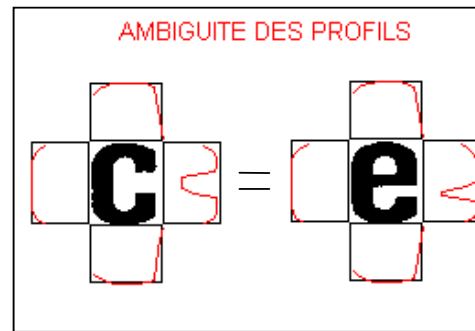


Figure 7. Caractères faiblement discriminés par la seule comparaison des profils externes.

Par exemple, les caractères "e" et "c" substituables par leurs contours externes restent encore ambigus (cf. Figure 7). Le profil et le contour ne sont pas des informations redondantes, mais bien complémentaires puisque leur efficacité diffère selon la qualité de l'image des caractères. Ainsi, les profils permettent de prendre une meilleure décision dans le cas des lettres dégradées.

Si les caractères ("5","S") et ("2","Z") sont mieux reconnus, beaucoup de substitutions sont obtenues pour la colonne du "o" substituable un grand nombre de fois avec les caractères "G","Q","6","8","9". Cela tient du fait que l'information permettant de distinguer ces caractères provient des boucles intérieures fermées et non du contour externe. De même, les nombreuses substitutions entre "h" et "b" proviennent des effets néfastes du renforcement du profil sur les caractères à forts empattements.

Les projections totales

La table de substitutions avec seulement 60 % de reconnaissance démontre le faible pouvoir discriminant de cette mesure, si elle est utilisée seule. De plus, cette caractéristique est trop sensible à l'épaisseur des traits. Il est alors souhaitable de ne pas analyser la distribution brute mais les différences premières de cette distribution ; donc, il est préférable de comparer les variations des épaisseurs des tracés plutôt que les épaisseurs elles-mêmes qui ne sont pas comparables directement (cf. Figure 8). Ceci peut être obtenu en remplaçant le calcul de la distance par une distance sur les différences de la distribution des projections. En effet, en comparant les changements d'épaisseur entre deux caractères, on aboutit à un taux de reconnaissance de 82% très supérieur au précédent. Les projections constituent une mesure robuste mais faiblement discriminante, elles sont utilisées en général pour la reconnaissance monofonte et seules, elles ne conviennent pas à une reconnaissance multifonte. Pour permettre la comparaison de deux caractères dont les traits sont d'épaisseurs différentes, nous avons constaté qu'il est préférable de comparer les différences premières de chacune des distributions.

Les projections totales sont insensibles aux variations de la structure mais restent sensibles aux épaisseurs des traits. C'est l'une des mesures les plus utilisées de nos jours, en Reconnaissance Optique de Caractères, associée à d'autres caractéristiques.

Les projections partielles

Le principal inconvénient de ces projections partielles par régions, reste le découpage arbitraire qui pénalise les caractères inclinés ; c'est le cas pour l'italique et les caractères de représentation plus libres dans leurs tracés, sur de trop nombreuses polices fantaisistes ou décoratives.



Figure 8. Comparaison directe des projections par rapport à une comparaison des variations des épaisseurs des tracés.

La mesure des espaces séparant le cadre du caractère est aussi importante que la mesure des traits. Dans la méthode conjointe des projections totales et des profils, on retrouve la même notion que la projection partielle des espaces et des traits. Ainsi, les profils gauche et droit ne sont que la mesure du premier et du dernier espace dans le sens horizontal. En utilisant une succession d'informations sur les traits et les espaces, on obtient une description complète du caractère que l'on peut reconstituer par simple sommation des histogrammes.

Comme toutes les projections, les projections partielles sont relativement dépendantes de l'inclinaison et des ruptures de continuité ou des collages entre les traits. Pour remédier à ce problème, il suffit de renforcer les projections en prenant la moyenne sur plusieurs valeurs. Si les caractères de forme complexe sont bien discriminés, en revanche les performances sont amoindries sur les caractères de forme simple comme ("l", "t", "j"). En effet, la comparaison effective n'a lieu que sur les trois premiers histogrammes (un histogramme pour l'espace à gauche entre le cadre et le caractère, un deuxième histogramme pour l'épaisseur de l'unique trait du caractère et un troisième pour l'espace à droite entre le caractère et le cadre). Ainsi, pour ces formes simples (une seule intersection dans les deux sens avec une droite horizontale), la mesure des projections partielles devient équivalente aux méthodes conjointes des projections et des profils. On peut dire que la projection partielle des espaces et des traits est bien une généralisation des méthodes classiques décrites précédemment.

Avec un taux de reconnaissance de 86% sur la même base de tests, les projections partielles des traits et des espaces entre les traits, présentent l'inconvénient de ne pas être économique en place mémoire pour un taux de reconnaissance encore insuffisant.

8 Conclusion

Cette étude montre qu'un attribut seul ne peut caractériser de façon non ambiguë un caractère. La combinaison de plusieurs types d'attributs pour représenter un caractère apparaît comme étant la solution incontournable. L'ensemble de ces attributs ajoutés à d'autres caractéristiques (surface, rapport, hauteur ou largeur, ...) permet la construction du vecteur de caractéristiques de chaque forme de caractère. L'information principale de ce vecteur est contenue dans les profils consolidés (insensibles aux ruptures de tracés) et dans les projections partielles des premiers traits. L'information sur les contours et sur les projections totales est en réserve pour distinguer des formes de caractères pour lesquelles les précédents attributs seraient inefficaces. La table de substitutions générée lors de la phase expérimentale montrent des taux de plus de 99 % de reconnaissance sur la base de tests.

Références

- [1] H. Freeman. On the encoding of arbitrary Geometric Configurations. *IRE Transaction on Electronic Computers*, 10 :260-268, 1961.
- [2] K. Yamamoto and S. Mori. Recognition of Handprinted Characters by an Outmost Point Method. *Pattern Recognition*, Vol. 12, pp. 229-236, 1980.
- [3] M.C. Rahier and P.G.A Jaspers. Dedicated LSI for a Microprocessor-Controlled Hand-Carried OCR System. *IEEE Trans.*, Vol. 29, N° 2, pp.225-229, 1980.
- [4] D. Gokana. Contribution à la Reconnaissance Automatique de Caractères Manuscrits. Application à la Lecture Optique de Caractères sur Supports Mobiles. *Thèse de Doctorat : Université de Paris Sud, Centre d'Orsay*, 1986.
- [5] L. Heutte, Y. Lecourtier et J.V. Moreau. Une Nouvelle Méthode d'Extraction de Projection et de Profils pour la Reconnaissance de Chiffres Manuscrits. In : *Actes du 3^{ème} Colloque National sur l'Ecrit et le Document*, Rouen, pp. 1-10, 6-8 Juillet 1994.
- [6] T.H. Hildebrandt and W. Liu. Optical Recognition of Handwritten Chinese Characters : Advances since 1980. *Pattern Recognition*, Vol. 26, No 2, pp. 205-225, 1993.
- [7] K. Pakker. Reconnaissance des Caractères Alphanumériques par une Analyse Structurale Hétérarchique. *Thèse de Doctorat : Institut National Polytechnique de Grenoble*, 1985.
- [8] R. Sabourin et G. Genest. Coopération de Classificateurs pour la vérification automatique des signatures. Dans : *Actes du 3^{ème} Colloque National sur l'Ecrit et le Document*, pp. 89-98, Rouen, 6-8 Juillet, 1994.