

Différenciation entre Alphabets dans des Textes Manuscrits

Audrey Seropian, Michel Grimaldi, Nicole Vincent

► **To cite this version:**

Audrey Seropian, Michel Grimaldi, Nicole Vincent. Différenciation entre Alphabets dans des Textes Manuscrits. Jun 2004, 2004. <sic_00001193>

HAL Id: sic_00001193

https://archivesic.ccsd.cnrs.fr/sic_00001193

Submitted on 7 Dec 2004

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

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

Différenciation entre Alphabets dans des Textes Manuscrits

Audrey Seropian¹ – Michel Grimaldi¹ – Nicole Vincent²

¹ Laboratoire d'Optique appliquée
Université du Sud – Toulon – Var
BP 20132, 83957 La Garde Cedex

² Laboratoire CRIP5-SIP
Université de Paris 5
45 Rue des Saints Pères, 75270 Paris Cedex 06

seropian grimaldi@univ-tln.fr, nicole.vincent@math-info.univ-paris5.fr

Résumé : *Notre but est de différencier et d'identifier des textes manuscrits écrits dans différents alphabets. Nous parvenons à notre but grâce à l'analyse fractale du style de l'écriture. Pour chaque alphabet, un ensemble de caractéristiques qui s'appuie sur les propriétés d'autosimilarité présentes dans l'écriture est extrait. Pour ce faire, des formes invariantes caractérisant l'écriture sont extraites par le processus de la compression fractale pendant la phase d'apprentissage. Elles sont ensuite organisées dans une base de référence qui peut être associée à un alphabet. L'étape d'identification d'alphabets est basée sur un processus de Pattern Matching utilisant successivement les différentes bases de références. Les résultats de cette analyse sont estimés par le coefficient de corrélation entre l'image initiale du texte et celle reconstruite à partir des différentes bases de références..*

Mots-clés : différenciation d'alphabets, compression fractale, extraction d'invariants, pattern matching.

1 Introduction

De jour en jour le nombre de documents créés est croissant et ces documents sont écrits dans différentes langues avec différents alphabets [HOC 99] suivant les auteurs et les destinataires. L'automatisation du traitement de ces documents est souvent liée à l'identification de l'alphabet afin d'appliquer le processus adéquat. Le cas des documents imprimés est assez bien résolu mais des améliorations restent à apporter dans le cas des documents manuscrits. Ici, nous abordons le problème de l'identification de l'alphabet utilisé dans des documents manuscrits [SER 03]. Cette opération fait partie du domaine de l'analyse et de la reconnaissance d'écriture. Connaître la langue d'un document permet en particulier d'adapter un processus de reconnaissance d'écriture.

Différentes méthodes existent [DIN 97, PAL 01, TAO 01]. Elles varient selon la manière dont l'alphabet est caractérisé. Notons la technique qui s'appuie sur la confiance associée aux résultats de quelques logiciels d'OCR ou celle ayant recours à des formes spécifiques extraites ou encore celle des processus globaux analysant la texture générale de l'image de texte.

Notre méthode est basée sur une approche globale tout en extrayant aussi des formes spécifiques invariantes dans l'écriture. Ces formes spécifiques sont trouvées pendant la phase d'apprentissage à partir d'un processus de compression fractale appliqué à l'image d'écritures représentatives de l'alphabet que l'on voudrait reconnaître.

Dans une première partie, les principes de la compression et décompression fractales sont rappelés. Dans une seconde partie, nous montrons comment les invariants dans l'écriture sont extraits. Ces invariants sont extraits afin d'être utilisés plus tard pour caractériser un alphabet. Enfin, nous présentons comment ces similarités entre les écritures et les formes enregistrées sont quantifiées pour différencier des alphabets. Quelques résultats sont montrés dans la dernière partie.

2 Compression fractale

2.1 Système de fonctions itératives

La compression fractale est une technique qui a été développée par Y. Fisher [FIS 95] à partir du formalisme des fractales développé par B. Mandelbrot. Reposant sur la présence d'autosimilarité dans un ensemble, le principe de la compression fractale est d'essayer de considérer une image donnée I comme le point fixe d'une transformée géométrique T définie sur l'ensemble

des images. Le plus souvent, la transformée T ne se limite pas à une transformation affine et l'image est définie comme l'attracteur d'un système de fonctions itératives (IFS). Le point fixe est obtenu comme la limite d'une séquence d'images :

$$I = \lim_{n \rightarrow \infty} T(I_n)$$

qui est définie itérativement par un premier terme aléatoire et une relation entre deux termes consécutifs de la suite : $I_{n+1} = T(I_n)$.

Alors, le problème de la compression est précisément d'obtenir, à partir d'une image connue, un système de transformées admettant précisément cette image comme point fixe.

Un IFS est un ensemble de transformées élémentaires géométriques souvent linéaires ou affines, contractantes qui permet de transformer une image en une autre image. Ces transformées que l'on suppose en nombre n rendent possible la définition d'une fonction T définie par :

$$T(I) = \bigcup_{i=1}^n T_i(I)$$

Le point fixe (image) qui est obtenu possède quelques caractéristiques spécifiques. En particulier, il est fait de copies d'image de lui-même mais modifiées par des transformations élémentaires. Dans l'exemple suivant, le système est fait de 3 transformées, une réduction, suivie d'une translation parmi trois qui assurent le repositionnement sous forme de triangle. Le point fixe, indépendant du motif de départ, est le triangle de Sierpinski.

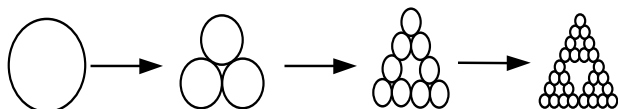


FIG. 1 - Génération d'une image fractale par IFS

Afin d'adapter cette méthode à des images naturelles, les transformées T_i sont limitées à des applications de similarités entre des parties de l'image, la réunion des ensembles d'arrivée assurant une partition du domaine de l'image. Alors, le modèle choisi est un système de fonctions itératives partitionnées (PIFS). Un PIFS définit une transformation T qui est l'union des transformées affines contractantes définies sur des domaines inclus dans l'image.

$$T(I) = T_1(I_1) \cup T_2(I_2) \dots \cup T_n(I_n)$$

L'ensemble de toutes les images obtenu à partir de toutes les transformées des imagettes I_i permet de partitionner le domaine spatial de I ; les domaines I_i , quant à eux ne constituent pas un recouvrement de l'image initiale. Ainsi, l'image initiale devient l'attracteur de l'IFS et pourrait être dérivée à partir de n'importe quelle image.

2.2 Compression fractale

Le but de l'étape de compression est de déterminer les transformées qui font partie du PIFS ayant l'image initiale I comme point fixe. Pour construire le PIFS, l'image est partitionnée en imagettes R appelées Ranges. Ces Ranges sont interprétés comme étant le résultat d'une transformation géométrique affine contractante définie à partir d'un domaine D . Ces Domaines sont eux-mêmes des imagettes de l'image initiale. Afin d'appliquer le théorème du point fixe et pour assurer la contractance de la transformation globale, chacune des transformations élémentaires est choisie contractante et plus précisément, les Domaines D sont choisis deux fois plus grands que les Ranges. Nous avons $R = T_i(D)$. D'un point de vue pratique, les parties R de l'image sont approximées à l'aide des transformées en minimisant une distance entre R et $T_i(D)$, deux sous ensembles de l'image, de même taille. Pour la distance, l'erreur quadratique moyenne entre les deux sous-images est utilisé.

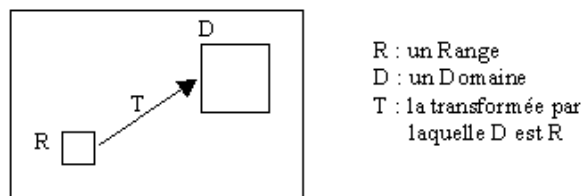


FIG. 2 - Principe de la compression fractale

La recherche de la meilleure transformée est restreinte, elle est limitée à un nombre fixe de transformées affines possibles qui ont la définition analytique suivante :

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} e \\ f \\ o \end{bmatrix}$$

Le paramètre (x,y) indique les coordonnées d'un pixel et z son niveau de gris. Les coefficients a, b, c et d déterminent la partie spatiale de la transformée, e et f déterminent la translation, o le contraste et s la luminosité. Il a été prouvé que pour obtenir un ensemble de transformations contractantes, le paramètre s doit être strictement inférieur à 1.

Le processus de compression fractale consiste à enregistrer le système de transformées qui sont définies chacune par 8 paramètres ainsi que la position des Domaines de définition concernés.

2.3 Décompression fractale

Dans le processus de décompression les transformées sont itérativement appliquées à toutes les imagettes d'une image quelconque jusqu'à obtenir le point fixe. Il est en fait supposé atteint quand la différence entre deux

images successives de la séquence est minimale. Pour quantifier la qualité de la compression fractale d'une image, en plus du taux de compression, le rapport signal/bruit (PSNR) est généralement utilisé. Dans notre cas, la compression n'est pas le but du processus, nous préférons donc l'utilisation du coefficient de corrélation calculé entre deux images x et y , car sans être idéal, il est plus représentatif de la perception visuelle.

On calcule le coefficient de corrélation $C(x,y)$ selon la formule suivante :

$$C(x, y) = \frac{1}{N} \sum_i \sum_j F(i, j)G(i + x, j + y)$$

Ce principe de sélection de domaines représentatifs d'une image est le point de départ à partir duquel le concept de notre méthode de différenciation d'alphabets est dérivé. Nous allons montrer comment une base constituée d'éléments invariants peut être construite pour chaque alphabet.

3 Apprentissage de l'alphabet

Dans cette section, la possibilité d'extraire des éléments invariants à partir de l'écriture [VIN 98] est démontrée. Chaque alphabet est considéré comme la combinaison de formes invariantes qui sont caractéristiques de l'alphabet et dont un certain nombre ne se retrouvent pas dans les autres alphabets. Nous considérons une image d'écriture manuscrite comme l'agrégation des lettres. Notre méthode est basée sur les autosimilarités qui existent dans l'écriture. Ces éléments sont extraits pendant le processus de compression fractale d'une image représentative de l'écriture manuscrite utilisant un alphabet donné, c'est-à-dire pendant la détermination des transformées du PIFS associé à l'image.

Pendant l'étape de compression, l'objectif est que tous les Ranges soient obtenus par une transformation d'un Domaine de l'image. Le critère pour choisir la meilleure transformée est le calcul de l'erreur quadratique moyenne (RMS) entre R et $T_i(D)$. Les Domaines sont choisis parmi toutes les imagerie de taille convenable contenues dans l'image.

En fait, nous ne sommes pas intéressés par les transformées elles-mêmes mais par les imagerie (Domaines) qui sont sélectionnées. Elles contiennent des éléments qui peuvent être retrouvés à deux échelles différentes dans l'image. Alors, pour des documents utilisant un même alphabet, une base de référence B_i est construite à partir de l'ensemble des Domaines qui ont été extraits et utilisés dans la sélection de la compression. Ces Domaines contiennent des caractéristiques sur l'écriture. Afin que cette base soit représentative de l'alphabet, le texte utilisé pendant la phase d'apprentissage doit être assez long et avoir été écrit par autant de scripteurs que possible pour que tous les styles d'écriture soient représentés. La base de

référence représente les similarités internes contenues dans l'écriture, ou plutôt dans les éléments de l'alphabet.

Lors de la recherche des meilleures transformées possibles pour associer les Ranges à des Domaines, nous avons limité cette recherche à un seul type de transformée : l'homothétie avec un rapport de $\frac{1}{2}$ associée à une translation. De plus, le choix d'une partition de l'image détermine la forme géométrique des imagerie qui constituent les Ranges et les Domaines. A cette étape nous avons choisi des fenêtres carrées et un rapport de contraction spatiale de $\frac{1}{2}$.

4 Différenciation d'alphabets

Bien sûr, afin de différencier un alphabet d'un autre, l'étape d'apprentissage concernant chaque alphabet doit avoir été faite et les bases de référence correspondantes doivent avoir été enregistrées. Si N alphabets doivent être différenciés, nous disposons *a priori* de N bases de références.

Il est à noter que la comparaison entre les écritures n'est possible que si les tailles des écritures sont relativement identiques. C'est pour cela qu'il est nécessaire d'appliquer une étape de normalisation sur les écritures. En fait, la taille est liée au choix de la dimension de la taille fixe des Ranges et doit être adaptée aux détails usuels de l'écriture, ainsi une taille du corps de l'écriture doit être associée à chaque base de référence. Les écritures sont alors normalisées par rapport au corps du texte. Plus précisément, une symétrie centrale scalaire est calculée selon la direction verticale et appliquée dans les deux directions de l'image. Le plus important est d'appliquer des processus équivalents au texte à étudier et aux textes ayant servi à l'apprentissage.

La longueur du texte à étudier n'est pas nécessairement aussi longue que celle du texte d'apprentissage. Dans tous les cas, il contient généralement seulement un style de scripteur alors que le texte d'apprentissage fait appel à plusieurs scripteurs.

Comme pendant la phase de compression fractale effectuée pendant l'apprentissage, l'image est partitionnée de la même manière en ranges. Alors, ce ne sont plus les autosimilarités qui sont recherchées, mais un processus de pattern matching entre les ranges du texte à identifier et les éléments de la base de référence qui est engagé. Les ranges ont la même forme que les imagerie de la base de référence. Pour chaque base et afin de quantifier les correspondances entre le texte et la base, une nouvelle image est construite à partir d'image blanche. Dans celle-ci sont copiées les imagerie pour lesquelles le coefficient de corrélation avec les ranges est le plus grand. La qualité de la correspondance est mesurée en comparant le texte initial et l'image de synthèse construite dans le processus de pattern matching (cf. FIG. 3).

Ainsi, avec chaque range R_i de l'image à étudier, une imagerie de la base B_j est associée en minimisant le critère des moindres carrés

$$d(R_i, T(D_k^j))$$

Ici encore, la transformée considérée est basée sur la seule homothétie de rapport un demi pour assurer un processus de pattern matching faisant intervenir deux échelles d'observation.

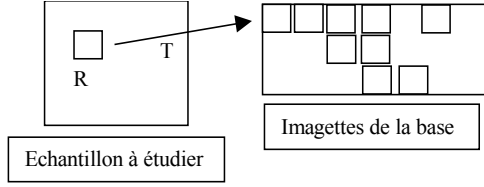


FIG. 3 - Etape de Différenciation

Le texte sera identifié comme utilisant l'alphabet associé à la base de référence qui a conduit au plus grand nombre de correspondances. La qualité de la similarité entre l'image initiale et l'image synthétisée est quantifiée par le coefficient de corrélation.

5 Résultats

Dans nos premières expérimentations, nous nous sommes limités à l'utilisation de deux alphabets différents et avons considéré deux textes d'apprentissage écrits en deux langues différentes. Nous avons pris des textes en arabe et des textes en français. Les deux textes d'apprentissage sont construits à partir d'échantillons d'écritures issus de différents scripteurs. Dans chaque cas, nous avons ici considéré 10 scripteurs différents. Un échantillon des textes est montré sur la FIG. 4.

Après l'étape d'apprentissage, nous obtenons deux bases (cf. FIG. 5) à partir des deux textes. La FIG. 5 (a) représente la base arabe avec 365 imagettes et la FIG. 5 (b) la base française avec 279 imagettes.

La méthode de différenciation a été testée sur un ensemble de 100 textes différents. Le tableau TAB. 1 indique les valeurs moyennes des coefficients de

corrélation obtenues sur les textes de la base de validation écrits par des scripteurs différents des textes de la base d'apprentissage. Tous les textes ont été triés avec succès selon l'alphabet utilisé et nous voyons que la différence entre les images synthétisées selon l'utilisation de la base de référence arabe ou latine est assez significative. Elle permet d'assurer une différenciation robuste entre les alphabets considérés.

	Images de tests	Texte écrit en arabe	Texte écrit en français
Bases			
Langue arabe		0,94	0,50
Langue française		0,47	0,87

TAB. 1 - Résultats de la différenciation

En fait, cette technique a déjà été utilisée pour identifier des scripteurs écrivant dans la même langue [SER 03]. Donc nous avons eu la curiosité de vérifier que nous n'identifions pas les scripteurs mais bien l'alphabet utilisé. Nous avons fait un test avec des personnes bilingues. Des textes ont été écrits par ces scripteurs à la fois en arabe et en français (cf. FIG. 6). Les résultats montrent une différenciation avec le même taux de reconnaissance parfait. Néanmoins, les résultats sont moins significatifs (cf. TAB. 2) que dans le cas général mais les formes obtenues dans les différents alphabets discriminent les écritures avec succès.

	Images de tests	Texte écrit en arabe	Texte écrit en français
Bases			
Langue arabe		0,93	0,74
Langue française		0,89	0,87

TAB. 2 - Résultats pour un scripteur bilingue et deux textes

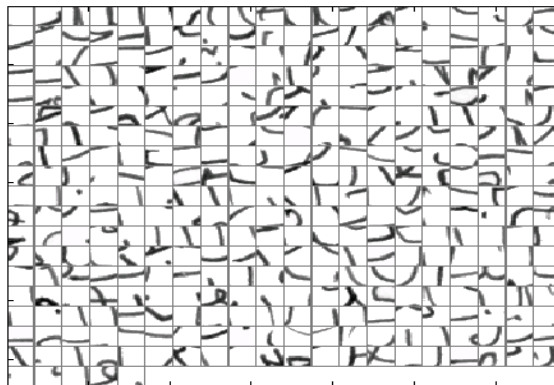
بالامس كانت الأمطار غزيرة وجلست برفقة زوجتي
 في فصل الربيع، السماء زرقاء والصفير
 الأطفال يلعبون في العلى كل يوم

(a) échantillon du texte d'apprentissage arabe, 3 scripteurs différents

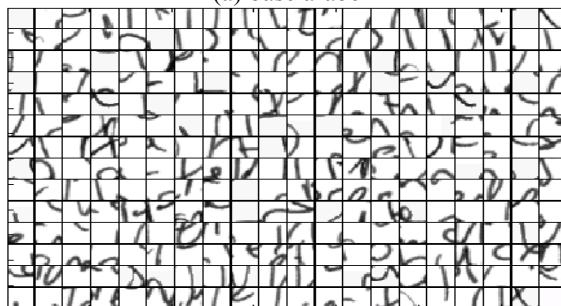
Nous rappelons dans cette section les principaux effets de la
 théorie des systèmes de fonctions itératives permettant le codage
 l'écriture suit celle de la règle de la géométrie fractale

(b) échantillon du texte d'apprentissage français, 3 scripteurs différents

FIG. 4 - Textes d'apprentissage des deux alphabets



(a) base arabe



(b) base latine

FIG. 5 - Exemples des bases de référence

اليوم الطقس جميل، نحن نغفل بالذوا بالاشاطرة

(a) texte écrit en arabe

Aujourd'hui il ne fait pas très beau

(b) texte écrit en français

FIG. 6 - Textes écrits par un scripteur bilingue

6 Conclusion

Nous avons montré que la différenciation d'alphabets pouvait être réalisée pour des textes manuscrits avant tout processus de reconnaissance par un processus qui s'apparente à de la différenciation de texture. Par contre,

nous n'avons pas utilisé des techniques de traitement d'images dont le niveau nous semble trop bas par rapport à la nature des images considérées. Les caractéristiques que nous avons extraites de manière automatique sont représentatives de l'alphabet utilisé et une information plus pertinente est contenue dans les éléments élémentaires extraits. Nous avons apporté une approche originale. L'image est traitée dans sa globalité en prenant en compte les caractéristiques spécifiques de chaque alphabet, nous avons travaillé sans utiliser de manière

explicite d'extraction de formes spécifiques reposant sur une connaissance *a priori* de ces formes discriminantes. La méthode n'est pas spécifique des alphabets pouvant être considérés. Les résultats sont convaincants et doivent être bien sûr confirmés avec un plus grand nombre de scripteurs et d'autres alphabets tels que l'alphabet cyrillique ou des alphabets asiatiques.

7 Bibliographie

- [HOC 99] Hochberg J., Bowers K., Cannon M., and Kelly P., « Script and Language Identification for Handwritten Document Images », *International Journal on Document Analysis and Recognition*, Volume 2, Number 2, pages 45-52, 1999.
- [DIN 97] Ding J., Lam L., Suen C.Y., « Classification of Oriental and European Scripts by using Characteristic Features », *ICDAR'97*, Ulm Germany, pp. 1023-1027, 1997.
- [PAL 01] Pal P.U., « Automatic identification of English, Chinese, Arabic, Devnagari and Bangala Script », *ICDAR*, Seattle, USA, 2001.
- [TAO 01] Tao Y., Tang Y.Y., « Discrimination of oriental and euromerican scripts using fractal feature », *ICDAR*, Seattle, USA, 2001
- [FIS 95] Fisher Y., « *Fractal Image Compression : Theory and Application* », Springer-Verlag, New York, 1995.
- [VIN 98] Vincent N., Barbezieux S., « Compression of handwriting images: a way to define a writing style », *Vision Interface*, 1998, pp. 347-354.
- [SER 03] Seropian A., Grimaldi M., Vincent N., « Writer Identification based on the fractal construction of a reference base », *ICDAR*, Edimburgh, pp.1163-1167, 2003.