

RECONNAISSANCE DES CARACTERES MANUSCRITS LATINS PAR LA METHODE DES MOMENTS

D. Saigaa⁽¹⁾ N. Benoudjit⁽²⁾ M. Boumehraz⁽¹⁾ K. Benmehamed⁽³⁾ M. Verleysen⁽²⁾

(1) :Département d'électronique Université Mohamed Khider Biskra (e-mail : d_Saigaa@yahoo.fr)

(2) : Laboratoire de Microélectronique - Unité DICE Batiment Maxwell Place du Levant, 3 B-1348,
Louvain-La-Neuve Belgium (e-mail: benoudjit@dice.ucl.ac.be)

(3) :Département d'électronique Université Farhat Abbes Sétif

RESUME - Le but assigné à ce travail est de faire la reconnaissance de caractères manuscrits latins. Pour cela, nous avons utilisé un modèle paramétrique. Ce modèle consiste à caractériser un objet (caractère) suivant un vecteur d'éléments caractéristiques tels que moments, descripteurs de Fourier, etc....

Pour notre application, nous avons utilisé deux méthodes d'extraction de paramètres caractéristiques, à savoir les moments invariants et les moments Zernike. Pour être efficaces dans leur exploitation ces paramètres caractéristiques doivent remplir les deux conditions suivantes : une petite invariance intra-classe et une large séparation inter-classe. Une fois que le vecteur caractéristique d'un caractère est extrait, l'étape suivante consiste à le présenter comme entrée au réseau de neurones, le MLP ; ce réseau est entraîné dans la phase d'apprentissage par l'algorithme de rétropropagation. Par la suite le taux de réussite est calculé pour chaque méthode. Enfin, nous avons comparé nos résultats avec ceux obtenus dans la littérature.

1-INTRODUCTION

La reconnaissance de formes est une partie essentielle de n'importe quel système d'analyse d'image de haut niveau. Le but d'un système typique de vision d'ordinateur est d'analyser des images d'une scène donnée et d'identifier le contenu de la scène. De tels systèmes sont maintenant en service dans différents domaines, tels que la robotique, la reconnaissance militaire, la télédétection, le traitement de documents, et l'automatisation industrielle. La plupart des systèmes de classification d'images partagent une structure générale qui se compose de quatre modules [1].

Le premier bloc concerne l'acquisition de l'image. Le deuxième bloc est celui du prétraitement. Le troisième bloc est celui de l'extraction des paramètres caractéristiques (features) par lesquels l'image est

représentée par un ensemble de caractéristiques numériques afin d'enlever la redondance des données et de réduire sa dimension. Le quatrième bloc concerne la classification, où une étiquette de classe est assignée à l'image/objet inconnu en examinant ses paramètres caractéristiques extraits et en les comparant aux représentations de classe, que le classifieur a appris pendant son étape d'apprentissage.

2-LES MOMENTS

L'extraction des paramètres caractéristiques est le dernier processus de réduction de l'information. De bons paramètres caractéristiques sont ceux satisfaisant deux conditions[2]:

- Une petite invariance **intra-classe** : des formes avec de légères différences appartenant à une même classe devraient avoir numériquement des valeurs proches.
- Une large séparation **inter-classe** : les paramètres caractéristiques de différentes classes devraient être tout à fait différents numériquement.

2-1 MOMENTS INVARIANTS

Les moments invariants sont un ensemble de fonctions non linéaires qui sont invariants à la translation, au changement d'échelle et à la rotation, et sont définis par les moments géométriques d'une image. Ils ont été présentés la première fois par Hu [3]. Dudani et al [4] les ont appliqués avec succès dans l'identification d'avions. Wong et Hall [5] ont utilisés les moments invariants pour coïncider des images de radar aux images optiques.

Soit une image bidimensionnelle binaire $M \times M$ $\{f(x, y); x, y = 0, \dots, M - 1\}$, le

moment géométrique d'ordre $(p + q)$ est défini par [1][6]

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} x^p y^q f(x, y), \quad (1)$$

avec $p, q = 0, 1, 2, \dots$

Pour maintenir l'intervalle dynamique du m_{pq} cohérent pour des images de tailles différentes, le plan d'image $M \times M$ est premièrement projeté sur un carré défini par $x \in [-1, +1]$ et $y \in [-1, +1]$.

Ceci implique que les emplacements de grille ne seront plus des nombres entiers mais auront des valeurs réelles dans l'intervalle $[-1, +1]$. Ainsi la définition du m_{pq} devient [2] :

$$m_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} x^p y^q f(x, y). \quad (2)$$

Le moment d'ordre 0, m_{00} , représente la surface du caractère (pixel noir); les moments d'ordre 1, m_{10} et m_{01} , définissent

les coordonnées du centre de gravité (\bar{x}, \bar{y}) de cette surface

$$\bar{x} = \frac{m_{10}}{m_{00}} \text{ et } \bar{y} = \frac{m_{01}}{m_{00}}. \quad (3)$$

Pour rendre ces moments invariables à la translation, on peut définir un moment central comme

$$\mu_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} (x - \bar{x})^p (y - \bar{y})^q f(x, y). \quad (4)$$

Les moments d'ordre 2, μ_{20} , μ_{02} et μ_{11} permettent de trouver les axes principaux d'inertie, et donc les allongements et les orientations du caractère. Les moments centraux peuvent être normalisés pour devenir invariants par rapport aux changements d'échelle en définissant

$$\eta_{pq} = \mu_{pq} / \mu_{00}^\gamma, \quad (5)$$

avec : $\gamma = (p + q) / 2 + 1$.

Un ensemble de fonctions non linéaires appelés moments invariants sont définis par les η_{pq} . Ces moments sont invariants à la rotation, à la translation et au changement d'échelle [3][6]. Il s'agit de :

$$\phi_1 = \eta_{20} + \eta_{02}, \quad (6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \quad (7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \quad (8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \quad (9)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[\frac{(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2}{3(\eta_{21} + \eta_{03})^2} \right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[\frac{3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2}{- (\eta_{21} + \eta_{03})^2} \right], \quad (10)$$

$$\phi_6 = (\eta_{20} - \eta_{02}) \left[\frac{(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2}{(\eta_{21} + \eta_{03})^2} \right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}). \quad (11)$$

Les valeurs numériques de ϕ_1 à ϕ_6 sont très petites. Pour éviter les problèmes de précisions les logarithmes des valeurs absolues de ces six fonctions sont choisis comme paramètres caractéristiques (features) représentant l'image.

2-2 MOMENTS ZERNIKE

La définition des moments géométriques réguliers a la forme d'une projection de la fonction $f(x, y)$ sur le monôme $x^p y^q$. Malheureusement, l'ensemble de base $x^p y^q$ n'est pas orthogonale. Par conséquent, les paramètres caractéristiques définis sur des fonctions de m_{pq} manquent d'optimisation dans la redondance de l'information et dans d'autres propriétés utiles qui pourraient résulter de l'utilisation des fonctions de bases orthogonales [2].

Zernike a présenté un ensemble de polynômes complexes qui forment un ensemble orthogonal complet à la limite de l'intérieur du cercle unité, c'est-à-dire $x^2 + y^2 = 1$.

Les moments Zernike invariants sont liés aux moments centraux normalisés. Ainsi une fois qu'on calcule les moments centraux normalisés d'un caractère en utilisant l'équation (5), on passe alors au calcul des A_{nm} en utilisant l'équation suivante[7] :

$$A_{nm} = \left[\left(n + \frac{1}{\pi} \right) \sum_{k=m}^n \sum_{j=0}^q \sum_{l=0}^m (-i)^l \binom{p}{j} \binom{m}{l} B_{nmk} \eta_{k-2j-m, 2j+m} \right], \quad (12)$$

avec :

$$B_{nmk} = \frac{(-1)^{(n-k)/2} [(n+k)/2]!}{[(n-k)/2]! [(m+k)/2]! [(k-m)/2]!},$$

$$q = \left(\frac{k-m}{2} \right) \text{ et } i = \sqrt{-1}.$$

Les moments Zernike invariants du 2^{ème} ordre au 12^{ème} ordre sont donnés par le tableau 1[2] :

Ordre	Moments	Nombre des moments
2	A ₂₀ , A ₂₂	2
3	A ₃₁ , A ₃₃	2
4	A ₄₀ , A ₄₂ , A ₄₄	3
5	A ₅₁ , A ₅₃ , A ₅₅	3
6	A ₆₀ , A ₆₂ , A ₆₄ , A ₆₆	4
7	A ₇₁ , A ₇₃ , A ₇₅ , A ₇₇	4
8	A ₈₀ , A ₈₂ , A ₈₄ , A ₈₆ , A ₈₈	5
9	A ₉₁ , A ₉₃ , A ₉₅ , A ₉₇ , A ₉₉	5
10	A _{10.0} , A _{10.2} , A _{10.4} , A _{10.6} , A _{10.8} , A _{10.10}	6
11	A _{11.1} , A _{11.3} , A _{11.5} , A _{11.7} , A _{11.9} , A _{11.11}	6
12	A _{12.0} , A _{12.2} , A _{12.4} , A _{12.6} , A _{12.8} , A _{12.10} , A _{12.12}	7

Tableau 1 : Liste des moments Zernike

3-RESULTATS EXPERIMENTAUX

La base de donnée consiste en des images binaires 100x100 des 26 caractères majuscules de l'alphabet latins (A à Z). 200 images sont générées pour chaque caractère ; chacun des 50 scripteurs a écrit trois fois l'alphabet de A à Z pour la phase d'apprentissage et une fois pour la phase de test c'est à dire au total 5200 caractères.

L'estimation du taux de réussite est effectuée en recherchant le rapport des échantillons de test bien identifiés au nombre total des échantillons testés.

$$\text{Taux de réussite} = 100 \times (N_c / N)$$

Avec

N_c : le nombre de caractères bien identifiés.

N : le nombre total des échantillons testés.

Les paramètres sélectionnés pour les deux classifieurs à réseaux de neurones [9] [10][11][12][13] MLP1 et MLP2 (ayant chacun deux couches) sont comme suit :

Pour le premier classifieur, MLP1 :

- Le coefficient d'apprentissage $\mu = 5 / (5 + \text{epoque})^{\frac{1}{2}}$, évolue en fonction du nombre d'itérations.
- Le nombre de neurones d'entrée est le même que le nombre de paramètres caractéristiques
- Le nombre de neurones de sortie 26 .

Pour le deuxième classifieur, MLP2 :

- La valeur initiale du coefficient d'apprentissage est $\mu = 0,2$. Ce coefficient est adapté au cours de la phase d'apprentissage selon la méthode heuristique [9][10]. En effet, en se basant sur le signe de la dérivée de l'erreur, le coefficient d'apprentissage est incrémenté ou décrémenté.
- Le momentum $\beta = 0,5$.

Les paramètres caractéristiques de l'ensemble d'apprentissage sont normalisés pour obtenir une moyenne zéro et une variance égale à l'unité. Ceci afin d'assurer qu'un sous groupe des paramètres caractéristiques ne domine pas le processus d'ajustement des poids durant l'apprentissage. Le m^{ème} paramètre caractéristique est normalisé par :

$$x'_m = (x_m - \bar{x}_m) / \sigma_{x_m},$$

où \bar{x}_m et σ_{x_m} sont respectivement la moyenne de l'échantillon et l'écart-type du $m^{\text{ème}}$ paramètre caractéristique de toutes les classes.

3-1 Expériences avec les moments invariants

Dans le premier ensemble d'expériences, les paramètres caractéristiques utilisés sont les 6 moments invariants pour chaque caractère. Le tableau 2 montre les résultats obtenus avec les deux MLP.

	N.N.C	T en %
MLP1	37	29,30
MLP2	34	33,30

Tableau 2 : Taux de réussite obtenu avec les deux MLP(cas des moments invariants)

N.N.C :Nombre de neurones dans la couche caché

3-2 Expériences avec les moments Zernike

Le tableau 3 illustre les résultats obtenus pour les deux MLP en utilisant des moments Zernike variant de l'ordre 5 à l'ordre 12.

	Ordre	N.N.C	T en %
MLP 1	5 (10 moments)	36	38,92
	8 (23 moments)	44	40,84
	9 (28 moments)	48	41,38
	12 (47 moments)	55	44,07
MLP2	5 (10 moments)	32	44
	8 (23 moments)	34	46,69
	9 (28 moments)	46	48,30
	12 (47 moments)	56	50,69

Tableau 3 : Taux de réussite obtenu avec les deux MLP(cas des moments Zernike)

Le choix du nombre de neurones dans la couche cachée de ces deux réseaux (tableau 2 et 3) est fait après que nous ayons réalisé plusieurs essais heuristiques. Ces valeurs donnent la meilleure performance pour les deux réseaux utilisés.

Nous remarquons que le taux de réussite le plus élevé est obtenu par la méthode des moments Zernike, ceci en utilisant les 47 moments Zernike d'ordre 12.

3-3 Commentaires et comparaison

Les résultats obtenus dans les deux expériences montrent que les moments Zernike sont meilleurs par rapport aux moments invariants, malgré que, ces résultats restent médiocres par rapport à

notre objectif qui est l'obtention d'un taux de reconnaissance élevé. Ceci s'explique par le fait que les paramètres caractéristiques correspondants ne remplissent pas les deux conditions

Pour la première condition : Ces paramètres caractéristiques ont une grande invariance intra-classe, ce qui veut dire que des caractères appartenant à une même classe et ayant de légères différences, ont des valeurs numériques trop espacées. Ceci est illustré par les figures (1) qui montre les distances entre les lettres A de différents scripteurs.

En effet, d'après la figure(1) nous remarquons que l'invariance intra-classe pour la lettre A varie entre 0 et 5,5. Comme il sera vu dans la suite (figure (2)), ces valeurs sont beaucoup trop grande par rapport aux plus petites distances inter-classes (de l'ordre de 0,5).

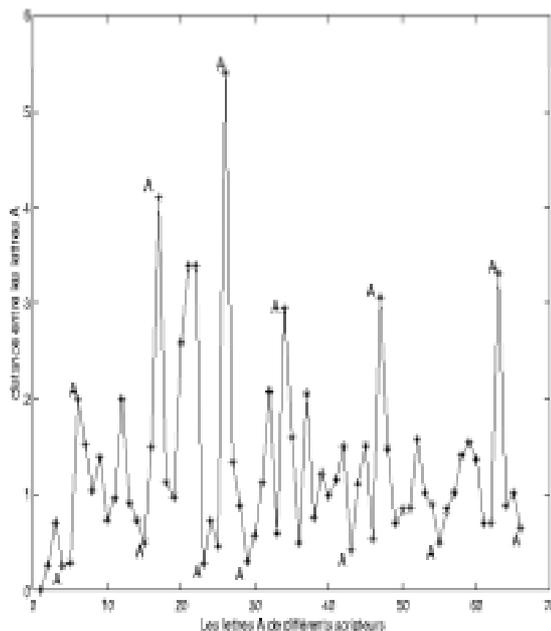


Figure (1) : Invariance **intra-classe** de la lettre **A** de différents scripteurs

Les valeurs numériques d'une même lettre pour différents scripteurs sont plus proches à celle des autres classes, ceci implique qu'ils possèdent une petite séparation inter-classe. La figure (2) montre les distances des paramètres caractéristiques de la lettre A par rapport à ceux des autres caractères.

Pour la deuxième condition : Les valeurs numériques d'une même lettre pour différents scripteurs sont plus proches à celle des autres classes, ceci implique qu'ils possèdent une petite séparation inter-classe. La figure (2) montre les distances des paramètres caractéristiques de la lettre A par rapport à ceux des autres caractères.

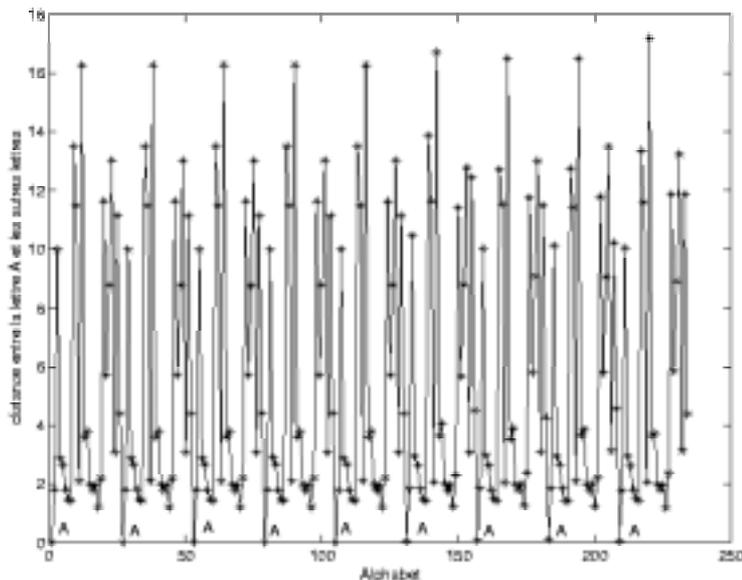


Figure (2) : Invariance **inter-classe** de la lettre A par rapport aux autres lettres des différents scripteurs

L'auteur Khotanzad [2] a utilisé les mêmes paramètres caractéristiques (moments Zernike) dans la reconnaissance des caractères latins. La base de données qu'il a utilisée consiste en des images binaires 64x64 des 26 caractères latins (A à Z). 24 images différentes sont générées à partir de 4 silhouettes légèrement différentes pour chaque caractère

Le taux de réussite obtenu par l'auteur Khotanzad est de 99,8%, ceci en utilisant un MLP avec les paramètres suivants :

- Nombre de données d'entrée 50.
- Nombre de neurones dans la couche caché 50.
- Nombre de neurones dans la couche de sortie 26.

- Le coefficient d'apprentissage est égal à 0,7.
- Nombre d'échantillons pour la base d'apprentissage 312.

Nous remarquons, d'après ces paramètres que le nombre d'échantillons utilisés pour la phase d'apprentissage est 312.

Ce nombre est très inférieur au nombre de paramètres du réseau qui est 3726 ; l'auteur fait de l'overfitting.

L'auteur Khotanzad ne s'est pas rendu compte de l'overfitting parce que les caractères utilisés pour le test sont presque identiques à ceux utilisés lors de la phase d'apprentissage ; ceci a donné une erreur de généralisation trop petite ou un taux de réussite trop élevé.

De notre côté, nous avons fait les expériences sur une base de données constituée de 468 caractères obtenus par des changements de tailles, des rotations et des translations arbitraires des 26 caractères d'un même scripteur. Le taux de réussite est de 98.29% en utilisant les moments invariants avec 22 neurone dans la couche caché et il est de 99.14% en utilisant les moments Zernikh (ordre 12) avec 38 neurones dans la couche caché.

4-CONCLUSION

L'objectif de notre travail était de faire la reconnaissance des caractères manuscrits latins tout en obtenant un taux de réussite élevé.

D'après les résultats des expériences que nous avons réalisées, nous pouvons conclure que les moments invariants et les moments Zernike ne sont pas de très bons paramètres caractéristiques pour la reconnaissance des caractères latins manuscrits ; ceci est prouvé par un taux de reconnaissance médiocre, 50,69 %.

Les résultats obtenus montrent que ces paramètres caractéristiques ne remplissent

pas les deux conditions nécessaires mentionnées auparavant.

Ces paramètres caractéristique ont une grande invariance **intra-classe**, ce qui veut dire que les paramètres caractéristiques des caractères appartenant à une même classe et ayant de légères différences, ont des valeurs numériques trop espacées.

Les valeurs numériques des paramètres caractéristiques d'une même lettre pour différents scripteurs se rapprochent des valeurs numériques d'autres lettres ; ceci implique qu'ils possèdent une petite séparation **inter-classe**.

Par contre, nous pouvons dire que ces paramètres caractéristiques sont invariants par rapport au changement de taille, à la rotation et à la translation .

Le taux de reconnaissance obtenu par l'auteur Khotanzad 99,8 %, pour une base de données obtenue par des rotations, des translations et des changements de taille de quelques manuscrits, n'est pas favorable en réalité dans la reconnaissance réelle.

A notre avis, l'auteur Khotanzad ainsi que d'autres travaux réalisés par la suite [8][14][15] n'ont pas fait une reconnaissance de caractères proprement dite, qui doit vérifier les deux conditions nécessaires déjà mentionnées ci-dessus , concernant l'invariance intra-classe et inter-classe, mais ils ont plutôt montré que ces paramètres caractéristiques (moments invariants et moments Zernike) sont invariants par rapport au changement de taille, à la rotation et à la translation.

REFERENCES BIBLIOGRAPHIQUES

- [1] S. Mori, H. Nishida and H. Yamada, 'Optical Character Recognition', Copyright © 1999 by John Wiley and Sons.
[2] A. Khotanzad and J. Lu, 'Classification of Invariant Image Representation Using a Neural Network', IEEE, Trans. On Acoust. Speech and Signal Processing, Vol. 38 N°. 6, June 1990, pp. 1028-1038.
[3] M. Hu, 'Visual Pattern Recognition by Moment Invariants', IRE Trans. Inf. Theory 8, pp.179-187, 1962.

- [4] S. A. Dudani, K. J. Breeding and R. B. Mcghee, 'Aircraft identification by moment-invariants', IEEE Trans. Comput., vol. C.26, pp. 39-45, jan 1977.
[5] R. Y. Wong and E. L. Hall, 'Scene Matching With Invariant Moment', Computer Graphics and Image Processing, vol. 8, pp. 16-24, 1978
[6] R. Mukundan and K. R. Ramakrishnan, 'Moment functions in Image Analysis : Theory and Applications', Copyright © 1998 by World Scientific Publishing Co. Pte. Ltd.
[7] M. R. Teague, 'Image Analysis Via the General Theory of Moments', J. Opt. Soc. Am, Vol. 70, N°. 8, August 1980.
[8] O. D. Trier, A. K. Jain and T. Taxt, 'Feature Extraction Methods For Character Recognition – A Survey', Pattern Recognition, Vol. 29, No. 4, pp. 641-662, 1996.
[9] Jacek M. Zurada, 'Introduction to artificial Neural systems', West Publishing Company, 1992.
[10] Richar P. Lippmann, 'An Introduction to Computing With Neural Nets', IEEE ASSP Magazine, Vol.4, N°.2, pp 44-22, April 1987.
[11] P. J. Werbos, 'Backpropagation Through Time : What it does and how to do it', Proceeding of IEEE, Vol. 78, N°. 10, October 1990.
[12] Don. R. Hush and Bill G. Horne, 'Progress in Supervised Neural networks, What's New Since Lippmann?', IEEE Signal Processing Magazine, pp. 8-38, january 1993.
[13] M. Weinfeld, 'Réseaux de neurones', Techniques de l'Ingénieur, Traite Informatique.
[14] A. Khotanzad and Y. H. Hong, 'Rotation Invariant Image Recognition Using Features Selected Via a Systematic Method', Pattern Recognition, Vol. 23, N°. 10. pp. 1089-1101. 1990.
[15] A. Khotanzad and Y. H. Hong, 'Invariant Image Recognition by Zernike Moments', IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. 12, 1990.

