

Une méthode de classification de textures par extraction linéaire non paramétrique de caractéristiques

A non parametric linear feature extraction

approach to texture classification



Alain HILLION

Groupe Traitement d'Images, Département Mathématiques et Systèmes de Communication, École Nationale Supérieure des Télécommunications de Bretagne, BP n° 832, 29285 BREST CEDEX.

Ancien élève de l'École Normale Supérieure, Alain Hillion est agrégé de Mathématiques et Docteur d'État. Il est actuellement Professeur à l'École Nationale Supérieure des Télécommunications de Bretagne et dirige le département Mathématiques et Systèmes de Communication où il poursuit ses recherches dans le domaine des méthodes statistiques de reconnaissance des formes.



Pascale MASSON

Groupe Traitement d'Images, Département Mathématiques et Systèmes de Communication, École Nationale Supérieure des Télécommunications de Bretagne, BP n° 832, 29285 BREST CEDEX.

Après une maîtrise EEA et un DEA d'Électronique obtenus à l'Université de Bretagne Occidentale, Pascale Masson prépare actuellement une thèse au sein du Groupe Traitement d'Images de l'École Nationale Supérieure des Télécommunications de Bretagne. Sa recherche porte sur la classification contextuelle appliquée à la reconnaissance des formes.



Christian ROUX

Groupe Traitement d'Images, Département Mathématiques et Systèmes de Communication, École Nationale Supérieure des Télécommunications de Bretagne, BP n° 832, 29285 BREST CEDEX.

Ancien élève de l'École Normale Supérieure de Cachan, Christian Roux est agrégé de Sciences Physiques et Docteur de troisième cycle. Actuellement Professeur à l'École Nationale Supérieure des Télécommunications de Bretagne, il est responsable du Groupe Traitement d'Images où il poursuit ses recherches sur la caractérisation des textures et la segmentation des images.

RÉSUMÉ

On présente une approche non paramétrique de l'extraction linéaire de caractéristiques et son application à la classification de textures. Le cadre théorique de l'étude est rappelé et on donne une nouvelle présentation de l'extracteur optimal de caractéristiques selon la distance de Patrick-Fischer [17]. Les grandes lignes de la mise en œuvre de cette méthode sont présentées. La classification de textures synthétiques binaires ayant un aspect visuel naturel [15] est ensuite abordée; sur les exemples étudiés, on constate que la méthode proposée est meilleure, en terme de taux de bonne classification, que le classifieur basé sur l'analyse discriminante de Fisher [7].

MOTS CLÉS

Extraction de caractéristiques, classification non paramétrique, caractérisation et classification de textures.

SUMMARY

A non parametric approach to linear feature extraction is presented. The theoretical background is introduced with a new derivation of the equation that gives the best scalar extractor according to the Patrick-Fischer distance [17]. The main characteristics of the implementation are given. The application of the method to the classification of some binary synthetic textures with a natural visual aspect [15] leads to results better than those based on the Fisher discriminant analysis [7].

KEY WORDS

Feature extraction, non parametric classification, texture characterization and classification.

1. Introduction

La caractérisation de textures reste un problème largement ouvert dans le domaine de l'analyse d'images. Diverses approches, syntaxiques ou structurelles, statistiques, se côtoient dans la littérature [22]; elles ont chacune leur domaine d'application. En ce qui concerne les méthodes statistiques, principalement basées sur les matrices de cooccurrence, on a montré qu'elles étaient bien adaptées à l'analyse d'images représentant des scènes naturelles comme, par exemple, en télédétection par satellite [20, 22]. Un problème majeur se pose lorsqu'il s'agit d'utiliser l'ensemble des caractéristiques de texture définies par Haralick *et al.* [11]: ces paramètres caractéristiques extraits des matrices de cooccurrence sont en nombre important; ils ne sont pas indépendants et ils nécessitent de nombreux calculs, ce qui, malgré leur bon pouvoir discriminant, les rend d'un emploi assez lourd, surtout pour des images de taille importante.

C'est pourquoi nous proposons d'utiliser les méthodes d'extraction de caractéristiques. La richesse, la diversité et la complexité des données intervenant dans la segmentation des images font que cette étape d'extraction de caractéristiques est nécessaire. Depuis les premiers travaux de R. A. Fisher en 1936 dans le contexte de l'analyse discriminante [7], de nombreux efforts ont porté sur la définition de procédures efficaces et de mise en œuvre simple. Leur cadre théorique couvre un spectre large, allant des méthodes de l'analyse des données (nuées dynamiques [6], classification automatique [14]) à la minimisation effective de l'erreur de Bayes (dans le cas gaussien [16]) ou l'utilisation de considérations plus heuristiques (dans le cadre de l'analyse de texture [12]).

Dans le contexte de la segmentation des images de télédétection, il a été récemment proposé, pour résoudre ce problème, d'utiliser l'analyse discriminante de Fisher pour mener à bien une phase d'extraction linéaire de caractéristiques [13]. Parmi les nombreuses mesures de séparabilité [3, 5], nous avons choisi, à cause de sa relative simplicité mathématique, la distance introduite par E. A. Patrick et F. P. Fischer [17], que nous appellerons dans la suite la distance de Patrick-Fischer.

Nous présenterons le cadre théorique de l'étude, et plus particulièrement de l'extraction linéaire de caractéristiques basée sur la distance de Patrick-Fischer dans le paragraphe suivant.

Dans le paragraphe 3, on donnera les grandes lignes de la mise en œuvre de la méthode; on précisera en

particulier la notion de meilleur extracteur linéaire, on définira le meilleur classifieur et on étudiera ses performances.

Des résultats seront présentés et commentés dans le paragraphe 4, où on réalisera, selon la méthode proposée, la classification de textures synthétiques binaires ayant un aspect visuel « naturel ».

On donnera en conclusion des indications pour des applications et des travaux futurs.

2. Cadre théorique de l'étude

LES CARACTÉRISTIQUES DE TEXTURE DANS LA MÉTHODE DES MATRICES DE COOCCURRENCE

Soit G l'ensemble des n_g niveaux de gris : $G = \{k \in \mathbb{N}; 0 \leq k \leq n_g - 1\}$. Une image est une application de $DC \mathbb{N}^2$ dans G qui, à un point n de D (un pixel), associe une valeur (un niveau de gris) $X(n)$. On dira que l'image est une *texture* si, pour certaines translations $\delta \in \mathbb{N}^2$, les distributions de $(X(n), X(n+\delta))$ sont homogènes, c'est-à-dire indépendantes de n , de telle sorte que l'on peut poser :

$$\{p_\delta(i, j)\}_{i \in G, j \in G} = P_\delta = \{P[X(n)=i, X(n+\delta)=j]\}_{i \in G, j \in G}.$$

Les probabilités jointes $p_\delta(i, j)$ sont estimées par la méthode de la δ -matrice de cooccurrence [11, 20]. On calcule la fréquence de transition entre le niveau de gris i et le niveau de gris j , par rapport à la translation δ dans l'image :

$$f_\delta(i, j) = \sum_{m \text{ et } m+\delta \in D} 1_{\{i, j\}}(X(m), X(m+\delta))$$

où

$$1_{\{i, j\}}(X(m), X(m+\delta)) = 1 \quad \text{si } X(m)=i \text{ et } X(m+\delta)=j \\ = 0 \quad \text{sinon.}$$

Pour obtenir une matrice symétrique, on calcule aussi $f_{-\delta}(i, j)$ que l'on ajoute à $f_\delta(i, j)$; une estimation de la probabilité jointe $p_\delta(i, j)$ est alors donnée par :

$$\hat{p}_\delta(i, j) = \frac{1}{N} (f_\delta(i, j) + f_{-\delta}(i, j))$$

avec :

$$N = \text{card} \{D \cap D_\delta\} + \text{card} \{D \cap D_{-\delta}\} \\ = 2 \text{ card} \{D \cap D_\delta\} \\ D_\delta = \{m/m+\delta \in D\}$$

Haralick [11] propose d'extraire de chaque δ -matrice de cooccurrence 14 paramètres caractéristiques dont les plus classiquement utilisés sont des moments et des entropies [20].

EXTRACTION DE CARACTÉRISTIQUES

Le nombre de paramètres caractéristiques précédent est trop élevé pour une mise en œuvre pratique; c'est pourquoi on procède à une étape d'extraction de caractéristiques.

Un *extracteur de caractéristiques* est un opérateur T défini sur l'espace X des observations et à valeurs dans l'espace X_0 des caractéristiques. T remplace l'observation brute x de X (en général compliquée, X est un espace euclidien de dimension élevée) par un vecteur $T(x)$ d'un espace de dimension inférieure X_0 . $T(x)$ est appelé caractéristique (ou paramètre, ou primitive, ou « feature » par les anglo-saxons ou bien encore, dans le contexte de l'image, indice visuel). Toute application de X dans X_0 est un extracteur; dans la pratique, on se restreint à des sous-ensembles τ d'applications linéaires, voire scalaires si $X_0 = \mathbb{R}$. Si T est une projection de X dans X_0 , on effectue une sélection ou un choix de caractéristiques. Dans un sous-ensemble donné τ d'extracteurs de caractéristiques on cherche l'extracteur T^* qui permette « au mieux » la reconnaissance des formes (la classification) à partir de l'espace des caractéristiques X_0 . Dans le contexte des approches statistiques (que nous adoptons par la suite) la précision d'un problème de reconnaissance des formes est caractérisée par l'erreur de Bayes, e (qui s'écrit en fonction de π_1 et π_2 , probabilités *a priori* des classes 1 et 2 et de f_1 et f_2 , densités des deux lois possibles P_1 et P_2 des observations). Précisément

$$(1) \quad e = \frac{1}{2} - \frac{1}{2} \int_{X_0} |\pi_1 f_1(x) - \pi_2 f_2(x)| dx$$

L'erreur de Bayes $e(T)$, attachée au problème de reconnaissance de formes fondé sur l'extracteur T , s'écrit

$$(2) \quad e(T) = \frac{1}{2} - \frac{1}{2} \int_{X_0} |\pi_1 f_{(1, T)}(x) - \pi_2 f_{(2, T)}(x)| dx$$

[où $f_{(1, T)}$ et $f_{(2, T)}$ sont les densités des deux probabilités induites par T sur X_0].

La formulation mathématique de la recherche d'un extracteur optimal est alors facile :

$$(3) \quad \text{Trouver } T^* \in \tau, \text{ tel que } e(T^*) = \inf_{T \in \tau} e(T)$$

MESURES DE SÉPARABILITÉ

En fait, la minimisation de $e(T)$ [cf. (3)] étant impossible dans le cas général et difficile même dans des cas particuliers importants (voir [16] pour le cas gaussien), on est contraint à une procédure sous-optimale qui consiste à maximiser des *mesures de séparabilité* (ou de divergences) entre lois de probabilité. Une mesure de séparabilité [3, 5] attachée à un problème de

reconnaissance des formes est une quantité numérique $\Delta = \Delta(P_1, P_2; \pi_1, \pi_2)$ qui est maximale quand l'erreur de Bayes est nulle.

D. E. Boeke et J. C. A. Van Der Lubbe [3] par exemple, ont proposé un traitement général de la plupart de ces mesures de divergence (distance en variation, distance d'Hellinger, distance de Matusita, ...) en indiquant le lien mathématique entre Δ et e .

On remplace alors la résolution de l'équation (3) par :

$$(3') \quad \text{Trouver } T^* \in \tau, \text{ tel que } \Delta(T^*) = \sup_{T \in \tau} \Delta(T)$$

[La forme mathématique de Δ est telle que la résolution de (3') est plus facile que (3).]

Dans la suite, nous nous intéresserons à la « distance » introduite par E. A. Patrick et F. P. Fischer [17].

ANALYSE DISCRIMINANTE

Si la méthode des mesures de séparabilité a l'avantage de conduire (soit par des formules exactes soit par des méthodes d'analyse numérique de maximisation) à une solution sous-optimale de (3), elle présente l'inconvénient d'utiliser les densités $f_1, f_2, f_{(1, T)}, f_{(2, T)}$. En règle générale on ne dispose que d'échantillons d'apprentissage C_1 et C_2 , d'exemplaires des classes 1 et 2. On utilise alors ces échantillons pour estimer $\Delta(T)$, en remplaçant les densités par des estimations classiques (noyau, fonctions orthogonales, ...) et on maximise l'estimation $\Delta(T)$.

Une méthode très simple et très ancienne (puisqu'elle remonte en fait à R. A. Fisher [7] en 1936) et qui évite tout calcul de densité consiste à définir directement sur le nuage d'apprentissage une mesure de séparabilité : le critère à maximiser est le rapport de l'inertie interclasse sur l'inertie intraclasse. La mesure de séparabilité sera d'autant plus grande que les deux nuages seront d'une part éloignés et d'autre part concentrés.

C'est la méthode classique d'*analyse discriminante* (ou *méthode de Fisher*) que nous comparerons dans la suite à la maximisation d'une estimation de la distance de Patrick-Fischer.

DÉFINITION ET ESTIMATION DE LA DISTANCE DE PATRICK-FISCHER

Étant données deux probabilités P_1 et P_2 [définies sur un espace euclidien $X = \mathbb{R}^d$, admettant pour densités $f_1(x)$ et $f_2(x)$], E. A. Patrick et F. P. Fischer [17] ont introduit une mesure de séparabilité probabiliste (appelée *distance de Patrick-Fischer*) par :

$$(4) \quad \Delta(P_1, P_2; \pi_1, \pi_2) = \int_X |\pi_1 f_1(x) - \pi_2 f_2(x)|^2 dx$$

où π_1 et π_2 sont les probabilités *a priori* des classes 1 et 2.

Pour estimer la distance de Patrick-Fischer à partir d'un ensemble supervisé d'échantillons d'apprentissage, C , divisé en deux nuages, C_1 (n_1 échantillons

de la classe 1) et C_2 (n_2 échantillons de la classe 2), on remplace les quantités intervenant dans la formule (4) par des estimateurs $\hat{\pi}_1, \hat{f}_1, \hat{\pi}_2, \hat{f}_2$, définis classiquement par :

$$(5) \quad \hat{\pi}_1 = n_1^{-1} \cdot n_1, \quad \hat{\pi}_2 = n_2^{-1} \cdot n_2$$

$$(6) \quad \hat{f}_1(x, h) = n_1^{-1} \cdot h^{-d} \cdot \sum_{x_i \in C_1} K(h^{-1} \cdot (x - x_i))$$

$$(6') \quad \hat{f}_2(x, h) = n_2^{-1} \cdot h^{-d} \cdot \sum_{y_j \in C_2} K(h^{-1} \cdot (x - y_j))$$

où h est un nombre positif et $K(u)$ un noyau de Parzen pair, c'est-à-dire une densité de probabilité vérifiant $K(u) = K(-u)$, $u \in \mathbb{R}^d$ [18, 9].
L'estimateur de $\Delta(P_1, P_2; \pi_1, \pi_2)$ s'écrit

$$(7) \quad \hat{\Delta}_K(h) = \int_X |\hat{\pi}_1 \cdot \hat{f}_1(x, h) - \hat{\pi}_2 \cdot \hat{f}_2(x, h)|^2 dx.$$

Le développement de l'intégrale dans (7) conduit à évaluer des quantités telles que

$$\int_X K(h^{-1} \cdot (x - x_i)) K(h^{-1} \cdot (x - y_j)) dx$$

qui, K étant pair, sont, en fait, des produits de convolution. En posant, pour $m \in \mathbb{R}^d$,

$$(8) \quad D_h(m) = \int_X K(v) K(h^{-1} \cdot m) dv = (K * K)(h^{-1} \cdot m)$$

$$(9) \quad I_{K, h}(C_1) = \sum_{x_i, x_j \in C_1} D_h(x_i - x_j)$$

$$(9') \quad I_{K, h}(C_1, C_2) = \sum_{x_i \in C_1, y_j \in C_2} D_h(x_i - y_j)$$

On écrit (7) sous la forme :

$$(10) \quad \hat{\Delta}_K(h) = n^{-2} \cdot h^{-d} \{ I_{K, h}(C_1) + I_{K, h}(C_2) - 2 \cdot I_{K, h}(C_1, C_2) \}$$

En choisissant, dans le cas où $d=1$, $K(u) = 1_{[-1/2, 1/2]}(u)$ (noyau uniforme), on obtient $D_h(m) = 1 - h^{-1} \cdot |m|$ pour $m \leq h$. $I_{K, h}$ est la version non linéaire de l'inertie classique du nuage C_1 .

En choisissant, dans le cas général, pour $K(u)$ un noyau gaussien,

$$K(u) = (2\pi)^{-d/2} \cdot \exp \{ -2^{-1} \cdot \|u\|^2 \},$$

on obtient

$$(11) \quad D_h(m) = 2^{-d} \cdot \pi^{-d/2} \cdot \exp \{ -4^{-1} \cdot \|h^{-1} \cdot m\|^2 \}$$

ce qui est l'expression donnée par E. A. Patrick et F. P. Fischer [17]. Dans la suite, nous nous limiterons, comme dans [17], au cas de noyaux gaussiens.

EXTRACTION SCALAIRE DE CARACTÉRISTIQUES

Nous nous limiterons au cas de l'extraction scalaire de caractéristiques ($\dim X_0 = 1$, T est défini par $T(x) = \langle w, x \rangle$, $\langle \cdot, \cdot \rangle$ est le produit scalaire sur X) et nous chercherons à maximiser la distance de Patrick-Fischer des deux probabilités induites ou, pratique-

ment, à maximiser un estimateur de cette distance qui sera noté $\Delta(h, w)$. De manière plus précise, pour un noyau standard gaussien, la formule (10) appliquée aux deux nuages induits

$$C_1^w = \{ \langle w, x_i \rangle, 1 \leq i \leq n_1 \}$$

et

$$C_2^w = \{ \langle w, y_j \rangle, 1 \leq j \leq n_2 \}$$

se réduit à une combinaison linéaire de termes en $\exp \{ -4^{-1} \cdot h^{-2} \cdot \langle w, x_i - y_j \rangle^2 \}$.

$\Delta(h, w)$ étant essentiellement invariant par homothétie sur w et h [on a, en effet, la relation $\Delta(\lambda h, w) = \lambda^{-1} \Delta(h, \lambda^{-1} w)$], nous rechercherons $w^* \in X$ tel que

$$(12) \quad \Delta(1, w^*) = \sup_{w \in X} \Delta(1, w)$$

w^* sera le meilleur extracteur scalaire, au sens de la distance de Patrick-Fischer.

3. Mise en œuvre

Le meilleur extracteur scalaire w^* défini par la relation (12) est calculé par un algorithme classique quasi Newtonien d'optimisation, qui se trouve dans la bibliothèque mathématique NAG (routine E04KBF). Cette technique nécessite le calcul du gradient, par rapport à w , de $\Delta(h, w)$, qui, dans notre cas, se réduit à une combinaison linéaire de quantités telles que

$$w \cdot \langle w, x_i - y_j \rangle \exp \{ -4^{-1} \cdot h^{-2} \cdot \langle w, x_i - y_j \rangle^2 \}$$

Une fois w^* calculé, on définit alors une famille de classifieurs $\{\hat{\omega}_\sigma, \sigma > 0\}$, en remplaçant dans la règle (idéale) de Bayes les densités inconnues des deux lois de probabilité (selon P_1 et P_2) de la variable aléatoire $\langle w^*, x \rangle$ par des estimées. Le classifieur $\hat{\omega}_\sigma$ attribuera x à la classe 1 si et seulement si :

$$(13) \quad \sum_{i=1}^{n_1} \exp \{ -2^{-1} \cdot \sigma^{-2} \cdot \langle w^*, x - x_i \rangle^2 \} \geq \sum_{j=1}^{n_2} \exp \{ -2^{-1} \cdot \sigma^{-2} \cdot \langle w^*, x - y_j \rangle^2 \}$$

En prenant en compte un nuage test supervisé C' (indépendant du nuage d'apprentissage C), nous caractérisons les performances du classifieur $\hat{\omega}_\sigma$ en comptant le nombre d'échantillons mal classés [10, 21], noté $e(\sigma)$. Nous définissons enfin le meilleur classifieur scalaire en déterminant σ^* tel que

$$(14) \quad e(\sigma^*) = \inf_{\sigma} e(\sigma)$$

Dans le paragraphe suivant, nous comparerons les performances du classifieur que nous proposons ($e(\sigma^*)$) aux performances du classifieur basé sur l'analyse discriminante de R. A. Fisher, en les appliquant à des images texturées.

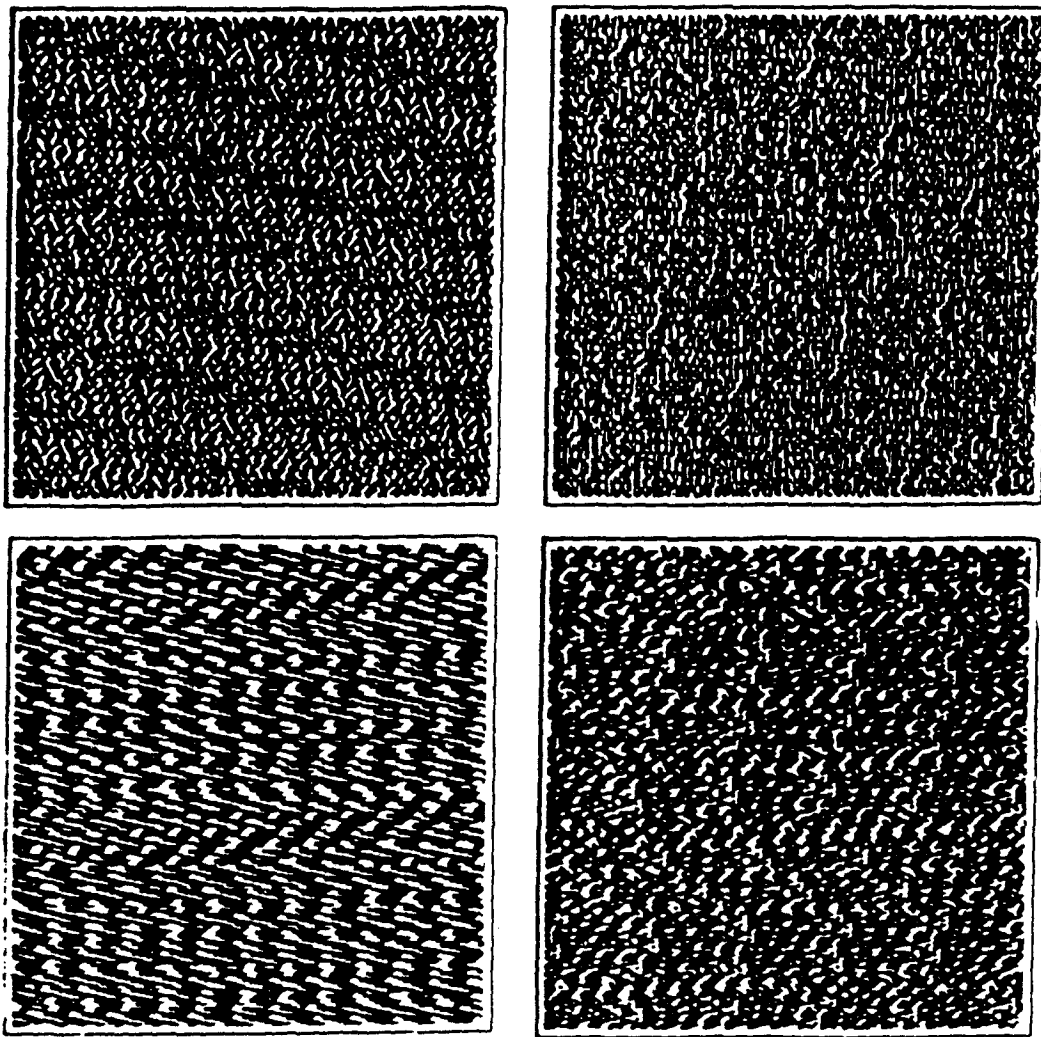


Fig. 1. — Les quatre textures binaires synthétiques utilisées
image 1 en haut à gauche, image 2 en haut à droite, image 3 en bas
à gauche, image 4 enfin).

L'extracteur scalaire, basé sur l'analyse de R. A. Fisher, sera noté w^\dagger , et le nombre d'échantillons mal classés sera noté e^\dagger .

4. Résultats

Pour évaluer les performances de la méthode, on utilise quatre textures binaires synthétisées par la méthode de substitution développée par Michel et al. [15] que nous allons décrire brièvement.

Si P est un ensemble de symboles, toute application de P dans P^* (ensemble formé par tous les blocs de deux symboles au moins de P) est appelée substitution. On définit une substitution sur un bloc de symboles par concaténation des substitutions sur chaque symbole formant le bloc. L'itération de ce processus conduit à une suite unidimensionnelle de symboles. Cette suite unidimensionnelle sert de base à la génération de l'image qui est construite ligne par ligne. Selon la substitution de départ, la forme de l'image et sa taille, l'aspect visuel de la texture ainsi synthétisée peut varier d'un aspect très structuré (rigoureusement

périodique) à un aspect plus « naturel ». Les symboles utilisés peuvent être des niveaux de gris, des nuances colorées ou des compositions graphiques à base de nuances de gris ou de couleurs.

Les images, obtenues à partir de deux symboles 0 et 1 (blanc et noir), sont montrées sur la figure.

Chaque image, de taille 256 par 256, est divisée en 144 fenêtres de 21 par 21 pixels, sans recouvrement. Dans cet ensemble de 144 échantillons, 72 servent à l'apprentissage de la méthode; les 72 restant sont utilisés pour la phase de test. Dans chaque fenêtre, la δ -matrice de cooccurrence est calculée selon quatre valeurs de δ , (1,0), (1,1), (0,1), (-1,1), c'est-à-dire pour une distance égale à 1 dans les quatre directions 0, 45, 90, 135°.

Comme $n_g = 2$ et que la somme de tous ses éléments est égale à 1, cette matrice est formée de deux éléments indépendants α et β , les éléments diagonaux. Le vecteur de caractéristiques est défini par $x = (\alpha, \beta)^t$. Le meilleur extracteur linéaire de caractéristiques w^* fournira le scalaire $\langle w^*, x \rangle$.

L'expression de l'extracteur scalaire a été calculée pour $h=1$, puis comparée à l'extracteur scalaire w^\dagger

MÉTHODE DE CLASSIFICATION DE TEXTURES

TABLEAU I.

Résultats de la classification des images 1 et 4.

Translation	Extracteur (Patrik-Fischer)		Nombre d'échantillons mal classés sur 144 en fonction de σ											
			0,01	0,1	0,2	0,3	0,4	0,5	1	1,5	2	3	10	100
0°	20,50	21,40	0	0	0	0	0	0	0	0	0	0	0	0
45°	37,42	42,11	43	28	25	25	25	25	25	31	36	41	46	47
90°	103,00	66,02	64	52	49	50	49	46	46	46	45	45	45	45
135°	25,74	47,32	23	12	12	12	12	12	9	10	11	11	11	11
	Extracteur (Fisher)													
0°	20,50	21,39	0	0	0	0	0	0	0	0	0	0	0	0
45°	49,46	-26,93	59	52	47	46	46	46	48	47	48	49	49	49
90°	78,42	93,20	58	51	49	46	43	48	46	44	44	43	43	43
135°	23,49	48,48	23	12	12	12	12	12	9	10	11	11	11	11

TABLEAU II.

Résultats de la classification des images 1 et 3.

Translation	Extracteur (Patrik-Fischer)		Nombre d'échantillons mal classés sur 144 en fonction de σ											
			0,01	0,1	0,2	0,3	0,4	0,5	1	1,5	2	3	10	100
0°	23,36	23,74	0	0	0	0	0	0	0	0	0	0	0	0
45°	36,07	36,59	53	43	33	29	29	29	31	35	37	39	42	42
90°	55,86	53,49	11	8	8	8	8	8	8	9	9	9	9	9
135°	34,75	32,15	34	40	39	40	42	43	43	43	43	43	41	41
	Extracteur (Fisher)													
0°	23,38	23,71	0	0	0	0	0	0	0	0	0	0	0	0
45°	36,07	36,58	53	43	33	29	29	29	31	35	37	39	42	42
90°	54,91	54,14	9	8	8	8	8	8	9	9	9	9	9	9
135°	33,52	33,42	41	40	40	42	42	43	41	42	44	44	43	43

TABLEAU III

Résultats de la classification des images 2 et 3.

Translation	Extracteur (Patrik-Fischer)		Nombre d'échantillons mal classés sur 144 en fonction de σ											
			0,01	0,1	0,2	0,3	0,4	0,5	1	1,5	2	3	10	100
0°	16,12	16,04	1	0	0	0	0	0	0	0	0	0	0	0
45°	-76,32	78,68	55	27	28	28	28	28	30	36	39	49	71	72
90°	30,90	31,13	7	0	0	0	0	0	0	0	0	1	1	1
135°	77,75	-73,80	49	30	26	26	28	25	28	34	37	48	71	72
	Extracteur (Fisher)													
0°	16,12	16,04	0	0	0	0	0	0	0	0	0	0	0	0
45°	82,20	72,34	81	75	72	76	77	73	73	71	69	71	72	72
90°	30,79	31,23	7	0	0	0	0	0	0	0	0	1	1	1
135°	77,18	73,97	67	62	59	52	53	51	50	48	46	60	70	72

donné par le critère de Fisher. Le taux de mauvaise classification a été évalué comme indiqué dans le paragraphe 3, pour différentes valeurs de σ allant de 0,01 à 100. L'ensemble des résultats obtenus est présenté dans les tableaux I à III.

Le tableau I présente les résultats obtenus lors de la comparaison des images 1 et 4 (voir fig. 1). L'extracteur scalaire se révèle être la trace de la matrice de cooccurrence pour les directions 0 et 45°, dans ces deux cas, les deux textures peuvent être discriminées par la trace de leurs matrices de cooccurrence. Ceci n'est pas vérifié pour les autres directions. Par exemple, dans la direction 90°, les éléments diagonaux

(α β) de la matrice de cooccurrence sont (0,25 0,25) pour l'image 1 et (0,24 0,26) pour l'image 2, ce qui conduit bien évidemment à deux valeurs égales pour les traces, qui ne constituent plus alors un paramètre discriminant.

En ce qui concerne le taux de mauvaise classification, les valeurs calculées sont à peu près identiques pour les deux critères dans les directions 0, 90 et 135°. A l'inverse, il est de 25 sur 144 pour la méthode proposée contre 46 sur 144 pour la méthode basée sur le critère de Fisher, dans la direction 45°. Le meilleur classifieur, selon (14), est donc celui basé sur la distance de Patrick-Fischer. De plus, les taux de mau-

vaise classification les plus bas sont obtenus pour les valeurs de σ allant de 0,1 à 1.

Le tableau II présente les résultats concernant les images 1 et 3, le tableau III les images 2 et 3. Des commentaires semblables peuvent être faits pour interpréter les résultats; dans certains cas, l'extracteur est proportionnel à la trace de la matrice de cooccurrence; lorsque ceci n'est pas vrai, on peut remarquer que les valeurs des traces sont très proches. En outre, le classifieur basé sur la distance de Patrick-Fischer conduit à des valeurs du même ordre de grandeur ou meilleures, mais en aucun cas moins bonnes, que le classifieur de référence.

Si, dans tous les cas qui sont présentés ici, le meilleur extracteur scalaire est celui basé sur la distance de Patrick-Fischer, une expérimentation systématique reste à être menée sur d'autres types de textures.

5. Conclusion

Nous avons présenté la première étape d'une approche non paramétrique de l'extraction scalaire de caractéristiques, en prenant comme distance entre lois de probabilités celle de Patrick-Fischer. Cette approche est réellement non paramétrique, dans le sens où les densités inconnues sont estimées à partir des données. Dans le cas d'application choisi (des images binaires synthétiques texturées), la méthode proposée, malgré une complexité de calcul plus élevée, se révèle plus intéressante, en terme de taux de bonne classification, que l'analyse discriminante de Fisher. La trace de la matrice de cooccurrence apparaît comme une caractéristique discriminante dans la classification de deux textures, dès lors que les valeurs de ces traces sont différentes.

Nous avons l'intention de développer cette étude dans plusieurs directions : dans le contexte des images texturées, nous chercherons à utiliser simultanément les quatre matrices de cooccurrence pour déterminer laquelle est la plus discriminante. Dans un esprit plus théorique, nous allons faire porter nos efforts sur la recherche du noyau de Parzen apportant le biais le plus bas et le meilleur taux de convergence à la distance de Patrick-Fischer.

Manuscrit reçu le 26 février 1988.

BIBLIOGRAPHIE

- [1] J. P. BENZÉCRI, *L'analyse des données : la taxinomie*, Dunod, Paris, 1980.
- [2] H. B. BIDASARIA, Least desirable feature elimination in a general pattern recognition problem, *Pattern Recognition*, 20, 1987, p. 365-370.
- [3] D. E. BOEKEE et J. C. A. VAN DER LUBBE, Some aspects of error bounds in feature selection, *Pattern Recognition*, 11, 1979, p. 353-360.
- [4] H. P. DECELL et L. F. GUSEMAN, Linear feature selection with applications, *Pattern Recognition*, 11, 1979, p. 55-63.
- [5] P. A. DEVIJVER et J. KITTLER, *Pattern recognition: a statistical approach*, Prentice-Hall, London, 1982.
- [6] E. DIDAY et J. C. SIMON, *Cluster analysis. In Digital Pattern Recognition*, K. S. FU éd., Springer Verlag, Berlin, 1976.
- [7] R. A. FISHER, The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, 7, 1936, p. 179-188.
- [8] K. S. FU, PYUNG JUNE MIN et T. J. LI, Feature selection in *Pattern Recognition*, *IEEE Trans. Sys. Man Cybern.*, 6, 1970, p. 33-39.
- [9] D. J. HAND, *Kernel discriminant analysis*, Research studies Press, 1982.
- [10] D. J. HAND, Recent advances in error rate estimation, *Pattern Recognition Letters*, 4, 1986, p. 335-346.
- [11] R. M. HARALICK, K. SHANMUGAN et I. DINSTEN, Textural features for image classification, *IEEE Trans. Sys. Man Cybern.*, 3, 1973, p. 610-621.
- [12] D. C. HE, L. WANG et J. GUIBERT, Texture feature extraction, *Pattern Recognition Letters*, 6, 1987, p. 269-273.
- [13] J. C. LEJEUNE, F. BONN, C. ROUX et B. ODOUX, Apport de l'information de texture à la segmentation des images du satellite SPOT en milieu urbain, *Comprint'87, International IEEE Montech'87 Conference*, November 10-12, 1987, Montréal.
- [14] I. C. LERMAN, *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.
- [15] P. MICHEL C. ROUX et M. DAUD, A new model of texture based on substitution sequences, *SIAM Conf. on Applied Geometry*, Albany, NY, 1987.
- [16] S. D. MORGERA et L. DATTA, Toward a fundamental theory of optimal feature, *IEEE Trans. Pattern Analysis Machine Intellig.*, 6, 1984, p. 601-616.
- [17] E. A. PATRICK et F. P. FISCHER, Non parametric feature selection, *IEEE Trans. Information Theory*, 15, 1969, p. 557-587.
- [18] E. PARZEN, On estimation of a probability density function and mode, *Ann. Math. Statist.*, 33, 1962, p. 1065-1076.
- [19] J. C. SIMON, *Patterns and operators: the foundations of data representation*, North Oxford, London, 1986.
- [20] M. SLIMANI, C. ROUX et A. HILLION, Image segmentation by cluster analysis of high resolution SPOT images. *Architectures and Algorithms for Digital Image Processing*, J. B. DUFF, H. J. SIEGEL et F. J. CORBEIT éd., Proc. SPIE 596, 1986, p. 112-119.
- [21] G. T. TOUSSAINT, Bibliography on estimation of misclassification, *IEEE Trans. Information Theory*, 20, 1974, p. 472-479.
- [22] L. VAN GOOL, P. DEWAELE et A. OOSTERLINCK, Texture analysis anno 1983, *Comput. Vision Graphics and Image Process.*, 29, 1985, p. 336-357.