Descripteurs de Fourier généralisés pour la reconnaissance d'objets couleur : performances de classification

Fethi Smach[1,2], Johel Miteran[1], Mohamed Atri[3,] et Jean-Paul Gauthier[1]

[1] Laboratoire Le2i, Université de Bourgogne, Faculté Mirande BP 47870, 21078 Dijon, FRANCE
Fethi.smach@u-bourgogne.fr, johel.miteran@u-bourgogne.fr

2 Laboratoire CES, Université de Sfax, ENIS, Tunisie,

3 Laboratoire EµE, Faculté des sciences de Monastir, Tunisie

Résumé Un des problèmes majeurs rencontrés en analyse d'images est la classification d'objets dans une scène. De nombreuses approches de reconnaissance sont proposées dans la littérature, le plus souvent basées sur le calcul d'invariants utilisés pour alimenter un classifieur. Nous proposons ici une extension de la méthode des descripteurs de Fourier généralisées à des images couleurs, et une étude comparative avec les moments de Zernike notamment grâce à l'usage d'une méthode de classification basée sur les SVM (Support Vector Machines). Nous évaluons dans ce papier la robustesse et les performances de notre approche appliquée sur différentes bases d'images.

Mots clés Descripteurs de Fourier généralisés, Reconnaissance d'objets couleur, Calcul d'invariants, SVM.

1 Introduction

Ce Les applications de reconnaissance de formes représentent un nombre croissant d'applications, de la vision artificielle au domaine industriel. La littérature abonde de techniques en reconnaissance de formes. Bon nombre de ces études ont pour objectifs de résumer une image à un nombre réduit de valeurs qui ont comme propriété d'être invariantes par rapport aux déplacements (translation, rotation). Plusieurs travaux ont été consacrés à la définition de descripteurs de formes invariants par un groupe de transformations [1,2]. Considérant les déplacements dans le plan, Gauthier et al [3] ont proposé une famille d'invariants, appelés descripteurs de mouvement, appliqués aux images binaires, qui sont invariants par translation, par rotation, et insensibles à l'effet miroir. H. Fonga [4] a étendu l'utilisation des descripteurs de mouvement, définis de manière similaire et appliqués aux images en niveau de gris. Notre but, ici, est de montrer que de tels descripteurs peuvent être utilisés de manière souple et robuste pour la reconnaissance de formes basée sur des images couleur, en les associant avec les SVM [5], et de proposer une implantation matérielle sur une architecture reconfigurable qui permet d'accélérer le calcul de ces descripteurs. Dans la section 2 de cet article, nous introduisons un bref rappel des descripteurs de Fourier généralisés et nous présentons une famille d'invariants complets dans le cas des groupes compacts. Cette complétude donnera une base théorique solide pour l'utilisation de ces descripteurs aussi bien en approche globale qu'en approche locale. Dans la section 3, nous rappelons la définition des moments de Zernike. La méthodologie que nous avons adoptée pour la reconnaissance d'objets couleur, combinant les SVM et les différents invariants, est présentée dans la section 4. Nous présentons ensuite quelques résultats obtenus sur des bases standards, COIL et AR-faces [6,7]. La section 6, présente l'architecture matérielle de l'unité de calcul des descripteurs de Fourier.

2 Rappel sur les descripteurs de Fourier généralisés

2.1 Les premiers invariants

2.1.1 Définition

La transformée de Fourier $\hat{f}$ est une grandeur physique définie par l'équation suivante:

$$\hat{f}(\xi) = \int_{\mathbb{R}^2} f(x) \exp(-j\langle x \mid \xi \rangle)\, dx \tag{1}$$

Où f est une fonction sommable sur le plan et $\langle . \mid . \rangle$ est le produit scalaire sur $\mathbb{R}^2$.

En coordonnées polaires, $\hat{f}(\lambda, \theta)$ désigne la transformée de Fourier d'un point ξ de coordonnées (λ, θ). Gauthier a défini les Descripteurs de Fourier Généralisés (DF1) par D_f de $\mathbb{R}^*_+$ dans $\mathbb{R}^*_+$ par :

$$D_f(\lambda) = \int_0^{2\pi} \left| \hat{f}(\lambda, \theta) \right|^2 d\theta \tag{2}$$

Dans le domaine discret, ces descripteurs sont remplacés par un ensemble fini de valeurs formant les composantes d'un vecteur qu'il est possible d'utiliser à l'entrée d'un classifieur.

2.1.2 Propriétés

Les descripteurs de Fourier généralisés définis par l'équation (2) ont d'importantes propriétés. Ils sont invariants par déplacement et réflexion :

- Si M est un déplacement tel que $g(x) = f \circ M(x)$, alors $\forall x \in \mathbb{R}^2, D_g(\lambda) = D_f(\lambda), \forall \lambda \in \mathbb{R}^2$ (3)
- Si $\Re$ est une réflexion tel que $g(x) = f \circ \Re(x)$, alors $\forall x \in \mathbb{R}^2, D_g(\lambda) = D_f(\lambda), \forall \lambda \in \mathbb{R}^2$ (4)
- Soit $f \in L^2(\mathbb{R}^2)$, si $g(x,y) = f(kx, ky)$ $K > 0$, alors $D_g(\lambda) = \frac{1}{k^4} D_f(\lambda)$ (5)

2.2 Les invariants complets

Les premiers descripteurs de Fourier généralisés ne sont pas complets (i.e. deux images différentes peuvent avoir le même DF1). Ceci n'est pas un inconvénient majeur car pour une très large classe d'applications, ces descripteurs sont suffisants. Toutefois, il est possible de définir un second ensemble d'invariants qui apporteront des améliorations dans certains cas, comme nous le verrons dans la section consacrée à l'étude des performances.

Ces seconds descripteurs sont complets dans les cas des groupes commutatifs et dans une large classe de fonctions, dans le cas des groupes compacts. Fonga [4] a montré que ces invariants ne sont pas complets dans le cas de groupe de déplacements dans le plan. Cependant, ils sont plus riches que les premiers descripteurs de Fourier généralisés [8].

2.2.1 Définition

Soit $\hat{f}$ la transformée de Fourier de f, Ω L'action de S_1 dans $\mathbb{R}^2$ définie par :

$$\Omega(\theta)(x,y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) \tag{6}$$

$\forall \xi_1 \in \mathbb{R}^2$, $\dot{\xi}$ est le cercle de centre (0,0) et de rayon $\|\xi_1\|$. Les orbites sont donc paramétrées par les rayons des cercles de centre (0,0). Les invariants sont définis par l'équation suivante :

$$I^{\xi_1,\xi_2} = \int_{S_1} \hat{f}\left[\Omega(\theta+\omega)\xi_1 + \Omega(\theta)\xi_2\right]\overline{\hat{f}}\left[\Omega(\theta+\omega)\xi_1\right]\overline{\hat{f}}\left[\Omega(\theta)\xi_2\right]d\theta \tag{7}$$

Avec $\|\xi_1\| = \lambda_1$, $\|\xi_2\| = \lambda_2$, $\omega \in S_1$ $\lambda_1, \lambda_2 \in \mathbb{R}^*_+$

Où ξ_1 et ξ_2 sont les points de coordonnées polaires respectives $(\lambda_1, 0)$ et $(\lambda_2, 0)$.

Donc $\forall \omega \in S_1$ fixé, il est possible de déterminer ces invariants :

$$\begin{aligned} & I_f^w(\lambda_1,\lambda_2) = \\ & \int_{S_1} \hat{f}\left(-\lambda_1 \sin(\theta+\omega) - \lambda_2 \sin\theta, \lambda_1 \cos(\theta+\omega) + \lambda_2 \cos\theta\right) \\ & \overline{\hat{f}}\left(-\lambda_1 \sin(\theta+\omega), \lambda_1 \cos(\theta+\omega)\right) \\ & \overline{\hat{f}}\left(-\lambda_2 \sin\theta, \lambda_2 \cos\theta\right) d\theta \end{aligned} \tag{8}$$

Cette quantité est équivalente à : $I_f^w(\lambda_1,\lambda_2) = \int_{S_1} \hat{f}(\Omega(\theta)(V_1+V_2))\overline{\hat{f}}(\Omega(\theta)(V_1))\overline{\hat{f}}(\Omega(\theta)(V_2))d\theta$ (9)

Où V1 et V2 sont deux vecteurs fixes dans l'espace de Fourier et V3=V1+V2 (fig.1). En pratique, on fixe l'angle entre V1 et V2 et on applique l'action $\Omega(\theta)$. Pour chaque valeur de l'angle θ, on accumule les valeurs du produit $\hat{f}(V_1)\hat{f}(V_2)\overline{\hat{f}}(V_3)$, où $\hat{f}$ est la transformée de Fourier et $\overline{\hat{f}}$ sa transformée conjuguée.

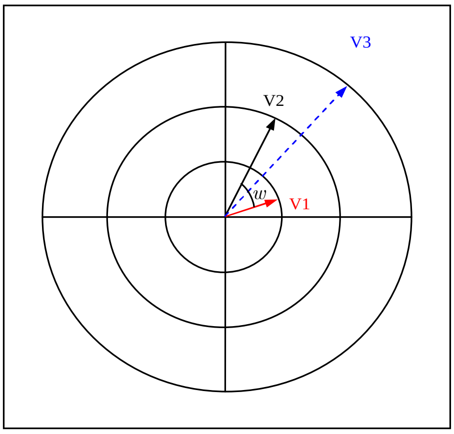


Figure 1 : Position des vecteurs dans l'espace de Fourier

2.2.2 Propriétés

– Soit $f \in L^2(\mathbb{R}^2)$ à valeurs réelles. $\forall \lambda_1, \lambda_2 \in \mathbb{R}_+$, $\forall \omega \in [0, 2\pi[$, alors est réel.

Pour ξ_1 (point du cercle de rayon λ_1) et ξ_2 (resp. λ_2), Les invariants dépendent uniquement de $\|\xi_1\|$ et $\|\xi_2\|$, et de l'écart angulaire signé entre ξ_1 et ξ_2. La quantité est symétrique en ξ_1 et ξ_2, on a donc :

– $I_f^{\omega}(\lambda_1, \lambda_2) = I_f^{\omega}(\lambda_2, \lambda_1) = I_f^{-\omega}(\lambda_1, \lambda_2)$ (10)

Cette famille d'invariants stables, que nous avons nommés DF2, est adaptée au groupe M_2 des déplacements dans le plan et la complétion de cette famille est assurée dans le groupe $M_{2,N}$ des déplacements d'angle $\frac{2k\pi}{N}$, lorsque est impair. Cette complétion permet d'assurer que deux objets différents auront des DF2 différents. Cette dernière propriété donne une justification théorique solide à leur utilisation en tant que paramètres discriminants pour la reconnaissance de forme.

3 Les moments de Zernike

Les moments de Zernike (MZ) de l'image ont été introduits par Teague [9]. Il a proposé d'utiliser les polynômes complexes de Zernike, orthogonaux à l'intérieur du cercle unité, que l'on peut écrire sous la forme :

$$V_{nm}(r, \theta) = R_{nm}(r) e^{-im\theta}$$

$$\text{Avec } \begin{cases} n \in \mathbb{N} \\ m \in \mathbb{Z} \text{ tel que n-}|m \leq n| \end{cases}$$

Avec $$R_{nm}(r) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n-2s+|m|}{2}\right)!\left(\frac{n-2s-|m|}{2}\right)!} r^{n-2s} \quad (12)$$

Les moments de Zernike $A_f(n,m)$ d'une image sont les projections de l'image f sur la base $\{V_{n,m}(r,\theta)\}$ et donc :

$$A_f(n,m) = \frac{n+1}{\pi} \int_0^1 f_m(r) \mathrm{R}_{n,m}(r) dr \tag{13}$$

Les modules des moments de Zernike sont des descripteurs invariants par rotation. O. Kubler et Wallin [10] ont proposé une famille de descripteurs composée d'un ensemble d'invariants par rotation et d'un ensemble de pseudo-invariants par réflexion.

Raveendran [11] introduit des invariants en translation composés par des combinaisons de moments de Zernike.

4 Méthodologie

Le classifieur retenu dans cette étude est basé sur les SVM [5], méthode supervisée très utilisée ces dernières années en reconnaissance des formes, pour les bonnes performances qu'elle procure, tant en termes de vitesse et souplesse d'apprentissage que de performances de classification.

4.1 Rappel des principes de base des SVM

Les SVM ont été introduits par Vladimir Vapnik en 1992 [5], et c'est seulement depuis une petite dizaine d'années qu'une large communauté s'est axée sur ce domaine. Il existe différentes familles de SVM selon la séparabilité des données.

L'idée de base des SVM est la détermination d'un hyperplan séparateur entre des échantillons appartenant à deux classes distinctes en maximisant la marge inter classe.

Soit une base d'apprentissage définis par n couples , $\{x_i, y_i\}$ où $x_i \in \mathbb{R}^d, y_i = \pm 1$ xi est le point décrivant le prototype i dans l'espace de description de dimension d, et yi l'étiquette de la classe de ce prototype. Supposons connu un hyperplan séparant les valeurs négatives des valeurs positives. L'équation d'un tel hyperplan est : $Wx + b = 0$ avec W la normale à cet hyperplan $\frac{b}{\|W\|}$, la distance perpendiculaire du plan à l'origine et où $\|\cdots\|$ définit la norme euclidienne. Posons d+ (resp. d̃) la plus petite distance entre l'hyperplan séparateur et l'élément de la classe {+1} (resp. {-1}) le plus proche. On définit la marge de cet hyperplan par l'équation suivante:

$$y_i(x_i W + b) - 1 \geq 0 \tag{14}$$

Déterminer les équations des deux hyperplans qui maximisent cette marge, revient à minimiser $\|W^2\|$ sous la contrainte de l'équation (14).

Il s'agit d'une résolution quadratique sous contrainte, l'utilisation des multiplicateurs de Lagrange (α_i) nous conduit à :

$$L_p = \frac{\|W\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i (x_i W + b) + \sum_{i=1}^{n} \alpha_i \tag{15}$$

La résolution de ce problème peut aussi s'effectuer à l'aide du dual de Lp, le dual de Wolfe, LD:

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \tag{16}$$

Sous la contrainte : $\sum_{i=1}^{n} \alpha_i y_i = 0$ (17)

Dans le cas où les classes ne sont pas linéairement séparables il est possible de résoudre ce problème d'une manière simple et efficace [12]. Les équations (eq. 15, eq. 16 et eq. 17) font intervenir les données (les prototypes xi et leurs étiquettes de classe yi correspondante) sous la forme d'un produit scalaire. Si on projette l'espace de description $\mathbb{R}^d$ dans un espace euclidien $\mathbb{H} = \mathbb{R}^e$

à l'aide d'une fonction telle que : $\phi : \mathbb{R}^d \to \mathbb{R}^e, \phi(x) = \tilde{x}$ (18)

Il est possible que dans cet espace les données initiales soient linéairement séparables [13]. Si ces espaces sont munis d'un produit scalaire, noté (.) ; on peut définir sur $\mathbb{R}^d$ une fonction noyau K telle que : $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^e \times \mathbb{R}^e, K(x, y) = \phi(x).\phi(y) = \tilde{x}\tilde{y}$ La connaissance explicite de la fonction ϕ n'est plus nécessaire et seul K(.,.) apparaît et la fonction de décision devient :

$$y = signe\left(\sum_{k=1}^{N_s} \alpha_k y_k K(S_k, x) + b \right) \tag{19}$$

Où S_i les sont les vecteurs supports issus de la phase d'apprentissage et sont déterminés de la même façon que précédemment. En revanche, le choix de la fonction noyau K, doit répondre au critère de Mercer, et il a été montré que toute fonction de la forme :

$$K(x, y) = \sum_{j=0}^{\infty} a_j (x, y)^j \forall a_j \in \mathbb{R}^+ \tag{20}$$

répond à ces critères.

Les noyaux les plus fréquemment utilisées sont :

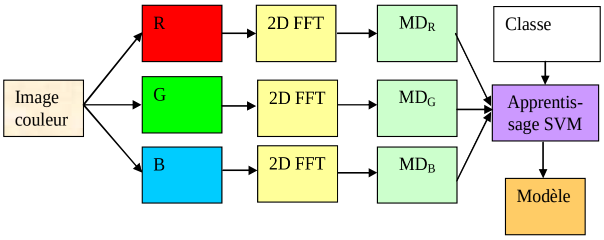
- Le noyau polynomial : $K(\mathrm{x}, \mathrm{y}) = (\mathrm{x.y} + 1)^p$ (21)
- Le noyau exponentiel ou RBF : $K(\mathrm{x}, \mathrm{y}) = e^{\left(\frac{-\|\mathrm{x}-\mathrm{y}\|^2}{2\sigma^2}\right)}$ (22)
- Le noyau tangente hyperbolique : $K(\mathrm{x}, \mathrm{y}) = Tanh\{k(\mathrm{x.y}) - \Theta\}$ (23)

4.2 Méthodologie d'application

Nous appliquons les phases classiques d'apprentissage (Figure2) et de décision (Figure3) de la manière suivante :

- L'image est redimensionnée au format 128x128 pixels.
- La FFT est calculée pour les 3 canaux rouge, vert et bleu.
- Les descripteurs de Fourier généralisés sont calculés pour chaque canal dans l'espace de Fourier. Les moments de Zernike sont calculés à partir de l'image pour chaque canal de couleur.

– L’ensemble des descripteurs (soit 3x64=192 valeurs pour les descripteurs de Fourier et 14x3=42 pour les moments de Zernike) est utilisé pour alimenter le classifieur qui génère un modèle.



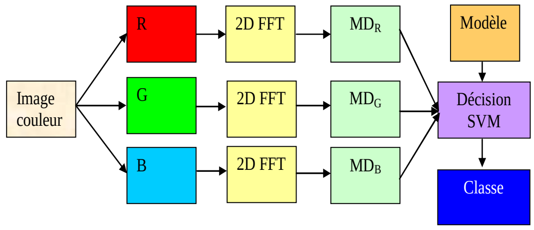


Figure 2 : Phase d'apprentissage **Figure 3 :** Phase de décision

Au cours de la phase de décision, les descripteurs sont calculés de la même manière que pendant la phase d’apprentissage. La classification s’effectue directement grâce à la fonction de décision des SVM.

5 Evaluation des performances

Nous présentons dans cette partie les résultats obtenus à partir des descripteurs décrits précédemment. Ceux-ci ont été appliqués sur la base COIL, AR-faces et une base d’images représentant des téléphones mobiles. Les performances ont été systématiquement évaluées par validation croisée d’ordre 5.

5.1 Base COIL-100

La base COIL [6] (Figure 4) est une base de 100 objets d’images couleur. Pour chaque objet, on dispose de 72 images d’objet en rotation (5° entre chaque image).

Figure 4: Différent objets de la base COIL-100

Les résultats de classification sont présentés au tableau (Table1). Le meilleur taux de reconnaissance est obtenu pour une combinaison des descripteurs de Fourier et des moments de Zernike avec une erreur de 0.01%. Les descripteurs de Fourier permettent de mieux discriminer les objets que les moments de Zernike. Seuls les résultats obtenus avec les premiers descripteurs généralisés ont été reportés ici, car dans ce cas les seconds descripteurs généralisés n'apportaient pas d'amélioration significative.

Table 1: Erreur de classification obtenue sur la base COIL pour chaque famille de descripteurs.

SVM Kernel RBF (σ)	Moments de Zernike	Invariants Fourier (DF1)	Moments de Zernike et Invariants Fourier (DF1)
100	23.33 %	1.69 %	3.10 %
10	1.89 %	3.87 %	1.00 %
1	0.22 %	0.09 %	0.01 %
0.1	1.89 %	38.48 %	16.47 %

Nous avons également étudié l'influence du nombre d'images utilisées pendant la phase d'apprentissage. En effet, dans les applications réelles (par exemple en biométrie), il est courant de ne disposer que d'un très faible nombre d'images par individu. Il est donc important que la méthode employée soit suffisamment performante dans ce cas.

La figure (Figure 5) présente l'erreur ainsi obtenue. Avec seulement 20% d'images d'apprentissage, les descripteurs de Fourier présentent une erreur de seulement 2% alors que l'erreur est supérieure de 4% avec les moments de Zernike. La meilleure convergence est obtenue avec une combinaison des deux approches : l'erreur est alors inférieure à 2%.

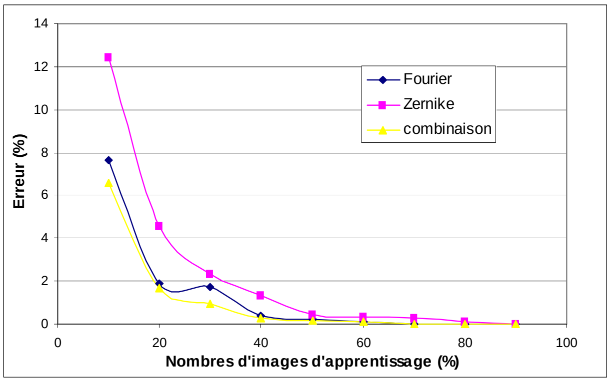


Figure 5: Erreur de classification en fonction du nombre d'image d'apprentissage pour la base COIL-100

5.2 Base AR-Face

La deuxième base utilisée pour comparer les descripteurs est la base AR-faces [7] (Figure 6). C'est une base de 4000 images couleur de 126 visages (70 hommes et 56 femmes), la base AR-Face est constituée d'images prises avec différents éclairages et les individus y portent parfois des lunettes.

Figure 6:Visages de la base AR-faces

Sur cette base, les résultats sont encore plus probants puisque avec les deuxièmes descripteurs de Fourier, nous obtenons une erreur de e=0.46% contre e=2.31% pour la première famille des descripteurs de Fourier. Pour les invariants de Zernike, l'erreur est e=10.6% ce qui montre notamment la meilleure robustesse des descripteurs de Fourier généralisées par rapport aux changements d'éclairages. La combinaison des MZ et des DF1 ou DF2 ne donne ici pas de meilleurs résultats que les DF2 seuls.

Table 2: Erreur de classification obtenue pour chaque famille de descripteurs (base AR-faces).

SVM Kernel RBF	MZ	DF1	*DF2*	MZ+DF2
$\sigma = 0.1$	10.6%	2..31%	0.46%	0,46%

6 Conclusion

Nous avons évalué sur des bases d'image standard les descripteurs généralisés de Fourier appliqués à la reconnaissance d'objets couleur, en les combinant avec les SVM. Nous avons notamment montré la robustesse de notre approche par rapport aux variations d'éclairage, et nous l'avons comparée avec d'autres méthodes appliquées aux mêmes bases. Nous avons réalisé une implantation PC complète qui permet de réaliser par exemple des opérations de contrôle qualité par vision à la cadence de 25 images/s. Une implantation sur Pocket PC à base de Windows mobile a également été réalisée et permet une reconnaissance embarquée à la cadence d'une image/s.

Plus généralement, nous avons montré le pouvoir discriminant de descripteurs de Fourier Généralisés. Ceci pourra être appliqué dans le cas de la reconnaissance d'objets réalisée à partir de régions ou points d'intérêts, qui permettent la reconnaissance de plusieurs objets dans une même scène, en présence de fonds complexes ou d'occlusions d'objets partielles. En effet, les descripteurs pourront alors être calculés dans des fenêtres de plus petites tailles (16x16 par exemple) autour des points d'intérêt afin de caractériser les régions considérées.

Références

[1] M. Petrou and A. Kadrov. “Affine invariant features from the trace transform”. IEEE Transaction on Pattern Analysis and Machine Intelligence, 26(1) (2004), pp.:30-44..

[2] F. Ghorbel. “A complete invariant description for gray-level images by the harmonic analysis approach”, Pattern Recognition Letters, 15, (1994), pp. 1043-1051.

[3] J. Turski,: Geometric Fourier analysis of the conformal camera for active vision. SIAM Rev. 46(1) (2004)

[3] J.P Gauthier, G. Bornard, M. Silbermann. “Motion and pattern analysis: harmonic analysis on motion groups and their homogeneous spaces. IEEE-trans SMC, vol. 21, no 1, Feb. 1991, pp. 159-172.

[4] H.Fonga, “Pattern recognition in gray-level images by Fourier analysis”, Pattern Recognition Letters 17 (14) (1996), pp. 1477-1489 (1996).

[5] B.E Boser, I.M. Guyon, V. Vapnik, “A Training Algorithm for Optimal Margin classifiers”, proc. Fifth Ann. Workshop Computational Learning theory, ACM, Press, 1992, pp. 144-152.

[6] http://www.cs.columbia.edu/CAVE/

[7] A. M. Martinez, “Recognition of Partially Occluded and/ or Imprecisely Localized Faces Using Probabilistic Approach”, Proceeding of IEEE Computer Vision and Pattern Recognition, CVPR’2000, pp 712, 717.

[8] J.P. Gauthier, F.Smach, C. Lemaitre, J.Mietran, “A finding invariants of group action on function spaces, a general Methodology from non-Abelian harmonic analysis”, Workshop on Mathematical control theory, Lisbon, Portugal, 2007.

[9] M. Teague, “Image analysis via the general theory of moments”, J. of the Optical Society of America, vol. 70, no. 80, (1980), pp. 920-930.

[10] A. Wallin et O. Kubler, “Complete sets of complex Zernike invariants and the role of pseudo invariants”, IEEE trans. On Pattern Analysis and Machine Intelligence, vol. 10, no. 6, (1988), pp. 811-821.

[11] C.W Chong, P. Raveendran, R. Mukundan, “Translation in Variants of Zernike moments”, Pattern Recognition, vol. 36, (2003), pp. 1765-1773.

[12] B.E.Boser, I.M.Guyon and V.N.Vapnik, "A training algorithm for optimal margin classifiers" in Proc. of Fifth annual workshop on computional learning theory, (1992) pp. 144-152.

[13] R.O.Duda, P.E.Hart and D.G.Stork, Chap. n° 5 "Linear Discriminant functions" in Pattern classification, John Wiley & Sons, 2001, pp. 264-265.