A comparative study of Motion Descriptors and Zernike moments in color object recognition

Cedric Lemaitre*, Fethi Smach**, Johel Miteran*, Jean Paul Gauthier* and Mohamed ATRI**

* *Université de Bourgogne, Laboratoire LE2 BP 47870 21078 Dijon - France*

cedric.lemaitre@u-bourgogne.fr

miteranj@u-bourgogne.fr

gauthier@u-bourgogne.fr

** *Université de Sfax, Laboratoire CES ENIS, 3000 Sfax- Tunisie*

fethi.smach@ieee.org

Abstract: Classification and object recognition is one of the most important tasks in image processing. Most applications deal with the classification of definite shapes, for example identifying a particular type aircraft. In these applications, compact visual descriptors are necessary to describe image content. Fourier descriptors are widely used in image processing to describe and classify object. Several techniques have proved useful moment's invariants. In this paper, we studied Motion descriptors (MD) introduced recently by Gauthier et al.; combined with Zernike Moments (ZM). Experiments are conducted using three databases: COIL-100, which consists of 3D objects, A R faces and cellular phones database. Recognition is performed by a Support Vector Machine as supervised classification method.

Key words: Motion Descriptors, Zernike Moments, SVM, Color Object Recognition

INTRODUCTION

Color and invariant object recognition is a critical problem in image processing. Numerous approaches are proposed in the literature, often based on the computation of invariants followed by a classification method. In this paper, we extend the notion of Fourier Descriptors to color images, and we use the descriptors as an input of a SVM based classifier. Considering the group of motions in the plane, Gauthier et al. [1] proposed a family of invariants, called Motion Descriptors, which are invariants in translation, rotations, scale and reflexions. H. Fonga [2] extended the Motion Descriptors, defining Similarity Descriptors and applying them to grey level images.

Our aim is to demonstrate here empirically the ability of such descriptors to be used successfully in color pattern recognition, and also combined with another well known set of descriptors: the Zernike Moments [3], [4]. We present results obtained testing our method with standard databases in the object recognition community: the COIL databases [5], [6] which contain images from 100 objects, A R face databases [7] (126 people) and a self made cellular phones database (20 phones).

In section 2 and 3, we review the Motions Descriptors and Zernike Moments. Then in section 4, the basic theory of support vector machines is reviewed. The obtained experimental and numerical results are illustrated in section 5. Finally the conclusion is given in section 6.

1. Review of Motion Descriptors

1.1. Definition

Motion Descriptors (MD) are defined as follows. Let f be a square summable function on the plane, and $\hat{f}$ its Fourier transform:

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

Where $\langle . \mid . \rangle$ is the scalar product in $\mathbb{R}^2$.

If (λ, θ) are polar coordinates of the point ξ, we shall denote again $\hat{f}(\lambda, \theta)$ the Fourier transform of f at the point (λ, θ). Gauthier defined the mapping D_f from $\mathbb{R}_+$ into $\mathbb{R}_+$ by

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

So, D_f is the feature vector which describes each image and will be used as an input of the supervised classification method.

1.2. Properties

Fourier descriptors, calculated according to equation (2), have several properties useful for invariant object recognition [1]:

Motion descriptors are motion and reflexion-invariant:

- If M is a "Motion" such as $g(x) = f o M(x)$, so for any x in $\mathbb{R}^2$,

$D_g(\lambda) = D_f(\lambda), \forall\lambda \in \mathbb{R}^2$	(3)

- If there exists a reflexions $\Re$ such that $g(x) = f o \Re(x)$, so for any x in $\mathbb{R}^2$,

$D_g(\lambda) = D_f(\lambda), \forall\lambda \in \mathbb{R}^2$	(4)

Motion descriptors are scaling-invariant:

- if *k* is a real constant such as $g(x) = kf(x)$, for any x in $\mathbb{R}^2$,

$D_g(\lambda) = \frac{1}{k^4} D_f(\frac{\lambda}{k}), \forall\lambda \in \mathbb{R}^2$	(5)

2. Zernike Moments

The kernel of Zernike Moments is the set of orthogonal Zernike polynomials defined over the polar coordinate space inside a unit circle. The two dimensional Zernike Moments of an image intensity function $f(r,\theta)$ are defined as [8]

$z_{pq} = \frac{p+1}{\pi} \int_0^1 \int_{-\pi}^{\pi} V_{pq}(r,\theta) r dr d\theta,$ $\mid r \mid \leq 1$	(6)

where the Zernike polynomials are defined as:

$R_{pq}(r) = \sum_{S=0}^{\frac{p-\mid q \mid}{2}} (-1)^s \frac{(p-s)!}{s!(\frac{p-2s+\mid q \mid}{2})!(\frac{p-2s-\mid q \mid}{2})!} r^{p-2s}$	(7)

$V_{pq}(r,\theta) = R_{pq}(r) e^{-jq\theta}$	(9)

The real-valued radial polynomials:

$R_{pq}(r) = \sum_{S=0}^{\frac{p-\mid q \mid}{2}} (-1)^s \frac{(p-s)!}{s!(\frac{p-2s+\mid q \mid}{2})!(\frac{p-2s-\mid q \mid}{2})!} r^{p-2s}$	(10)

Zernike moments are rotation-invariant: the image rotation in spatial domain simply implies a phase shift to the Zernike moments.

Mukandan et al [3], and Khotanzad [4], have shown that translation- invariance of Zernike moments can be achieved using image normalization method. In [8], Chee-Way chong, presents a mathematical framework for the derivation of translation invariants of radial moments defined in polar form.

3. Review of SVM based classification

A Support Vector Machine (SVM) is a universal learning machine developed by Vladimir Vapnik [9], [10]. A review of the basic principles follows, considering a 2-class problem (whatever the number of classes, it can be reduced, by a "one-against-others" method, to a 2-class problem).

The SVM performs a mapping of the input vectors (objects) from the input space (initial feature space) R_d into a high dimensional feature space *Q*; the mapping is determined by a kernel function *K*. It finds a linear (or non-linear) decision rule in the feature space *Q* in the form of an optimal separating boundary, which leaves the widest margin between the decision boundary and the input vector mapped into *Q*. This boundary is found by solving the following constrained quadratic programming problem: maximize

$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$	(12)

under the constraints

$\sum_{i=1}^{n} \alpha_i y_i = 0$	(13)

and $0 \leq \alpha_i \leq T$ for i=1, 2, …, n where $x_i \in R_d$ are the training sample set vectors, and $y_i \in \{-1,+1\}$ the corresponding class label. *T* is a constant needed for non-separable classes. $K(u,v)$ is an inner product in the feature space Q which may be defined as a kernel function in the input space. The condition required is that the kernel $K(u,v)$ be a symmetric function which satisfies the following general positive constraint:

$\iint_{R_d} K(u,v) g(u) g(v) du dv > 0$	(14)

which is valid for all *g≠0* for which

$\int g^2(u)\ du < \infty$ (Mercer's theorem).

The choice of the kernel *K(**u**, **v**)* determines the structure of the feature space *Q*. A kernel that satisfies (11) may be presented in the form:

$$K(u,v) = \sum_k a_k \Phi_k(u)\Phi_k(v) \qquad (15)$$

where a_k are positive scalars and the functions Φ_k represent a basis in the space *Q*. Vapnik considered three types of SVMs [10]:
Polynomial SVM:

$$K(x,y) = (x.y + 1)^p \qquad (16)$$

Radial Basis Function SVM (RBF):

$$K(x,y) = e^{\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)} \qquad (17)$$

Two-layer neural network SVM:

$$K(x,y) = Tanh\{k(x.y) - \Theta\} \qquad (18)$$

The kernel should be chosen *a priori*. Other parameters of the decision rule (16) are determined by calculating (9), i.e. the set of numerical parameters $\{\alpha_i\}_1^n$ which determines the support vectors and the scalar *b*.
The separating plane is constructed from those input vectors, for which $\alpha_i \neq 0$. These vectors are called *support vectors* and reside on the boundary margin. The number *Ns* of support vectors determines the accuracy and the speed of the SVM. Mapping the separating plane back into the input space R_d, gives a separating surface which forms the following nonlinear decision rules:

$$C(x) = \mathrm{Sgn}\left(\sum_{i=1}^{Ns} y_i \alpha_i \cdot K(s_i, x) + b\right) \qquad (19)$$

Where s_i belongs to the set of *Ns* support vectors defined in the training step.
SVM based classifier condenses all the information contained in the training set relevant to classification in the support vectors. This reduces the size of training set identifying the most important points. Moreover, SVM are quite naturally designed to perform classification in high dimensional spaces [11].

4. Object Recognition Process and experimental Results

4.1. Test Protocol

In order to validate our approach, we performed a cross validation test using two public databases: the COIL-100 [4] and the A R face color database [7] and one self made database of similar objects (cellular phones).

4.1.1. *Training Step*

During the training step (Fig. 1), the data flow is as follows: the input image is resample to 128x128 pixels, and a standard FFT is computed for each color channel (Red, Green, and Blue). The three corresponding Motion Descriptors are computed from the FFT values and the Zernike moments are computed from the 3 color channels. The final size of the vector used for SVM training is d=63x3=189 for Motion Descriptors, and d=14x3=42 for Zernike Moments. The result of the training step is the model (set of support vectors) determined by the SVM based method.

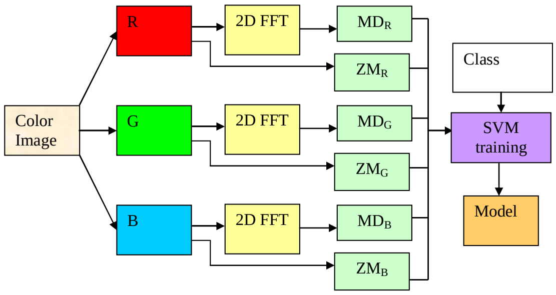

Fig. 1. Training Process

4.1.2. *Decision Step*

During the decision step, the Motion Descriptors or Zernike Moments are computed using the same way, and the model determined during the training step is used to perform the SVM prediction. The output is the image class (Fig. 2).

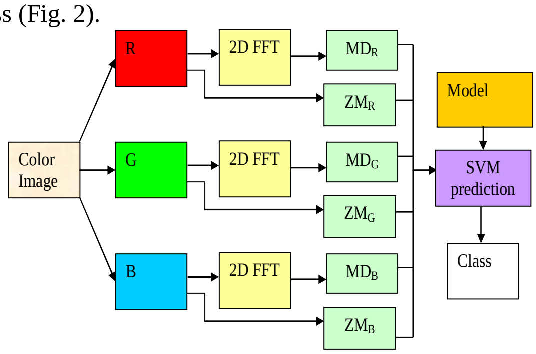

Fig.2. decision Process

The classification error rate was evaluated using cross-validation. The training step was performed using a training subset of samples B, and a test step was performed using a test subset of samples Γ, with $\Gamma \cup B = D$ and $\Gamma \cap B = \varnothing$ where *D* is the set of every available images in the database. For each database, we evaluated separately the classification error obtain using the Motion Descriptors, the Zernike Moment, and the mixing of both feature vectors. In this

case, the dimension of the feature space is d=189+42=231.
Since we used the RBF kernel in the SVM classification process, we have to tune the kernel size, i.e. the value of σ in the equation (14). This has been done empirically for each database, choosing the kernel value σ_{opt} which gave the minimum error rate.

4.2. Numerical results

4.2.1. *COIL-100 database*

COIL-100, the Columbia Object Image Library (COIL-100, Fig. 3) [5] is a database of colour images of 100 different objects, where 72 images of each object were taken at pose intervals of 5°. The images were pre-processed so that either the object's with or height (whatever is larger) fits the image size of 128 pixels.

Fig. 3. Sample Objects of COIL-100 database

a) Classification performance

Table 1 presents results obtained testing our object recognition method with the COIL-100. Tests have been performed using 5-fold cross validation (58 images used for training, 14 images used for testing, for each validation step). Optimum error values are depicted in red. In this case, Motion Descriptors outperforms Zernike Moment, and the combination of both descriptors improve significantly the global performance of the system.
Other methods testing the COIL-100 database, in the literature provide error rates from 12.5% to 0.1%. Testing is performed using different protocols [12].
In our global approach, the error e=0.01% corresponds to only 1/7200 image classified faulty.

Table 1: Cross validated error rate on COIL-100 database

	Zernike Moments	Motion Descriptors	*Motion-Descriptors and Zernike Moments*
$\sigma_{opt} = 0.1$	0.22 %	0.09 %	0.01 %

We studied the influence of the number of image samples used during the training step. Results are depicted in Fig. 4. The faster convergence is obtained for the combination of both descriptors.

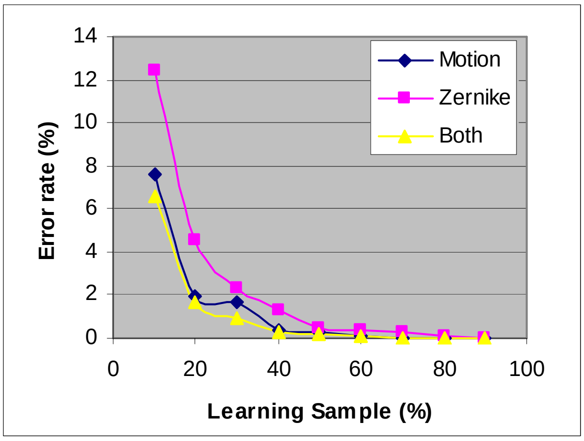

Fig.4. Influence of number of training samples of COIL

b) Robustness against noise

In order to study Zernike moments and Motion descriptors noise robustness, we evaluated the classification error obtained using a noisy database. This database has been created adding Gaussian noise to the COIL images. In order to test several noise levels, we created databases with different standard deviation Sd (0.0004<Sd<0.23). Some examples of noisy images are depicted in Fig. 5.
Table 2. achieves results of our method with noisy databases. Tests have been done using 9-fold cross validation and the best set SVM parameters kernel obtained in the section 5.3.1 a). Results show noise has little influence on classification performance as much when we use Zernike moments or Motion descriptors. Nevertheless Zernike seems to be more robust to additive noise, while combining descriptors is not really efficient here.

Fig. 5. Sample of COIL noisy object

Table 2: Error rate on COIL-100 noisy database

St. Dev. of Gaussian noise	Zernike Moments	Motion Descriptors	*Motion-Descriptors And Zernike Moments*
0.04	0.40 %	0.29 %	0.4 %
0.08	0.29 %	0.36 %	0.54 %
0.12	0.27 %	0.38 %	0.51 %
0.16	0.34 %	0.40 %	0.42 %
0.19	0.26 %	0.47 %	0.48 %
0.23	0.43 %	0.38 %	0.61 %

5.1.2 A R face database

Face detection is a difficult problem for which a lot of methods have been studied [13], [14], [15], [16], [17]. The face database we used to validate our approach (Fig. 6) was created by Martinez in the computer vision center [7]. It contains over 4.000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). Each image in the database consists of a 786x576 array of pixels, and each pixel is represented by 24 bits of RGB color.

Fig. 6. Face samples from the A R database

The third database tested with our approach is the A R face. For the experiments reported, images were morphed to a final 512x512 pixel size array. The best performance obtained is e=3.4%, using a 10-fold cross validation and Motions Descriptors. In this case, the addition of Zernike Moment to the Motion Descriptors does not improve performance, since the error is e=3.5%. However, our approach gives better results than in [7], where Martinez focuses on solving the localization error and occlusions. The error in this case is range to 15-5%.

Table 3: Error rate on A R face database

SVM RBF Kernel (CV 10)	Zernike Moments	Motion Descriptors	*Motion-Descriptors and Zernike Moments*
$\sigma_{opt} = 0.1$	35.12%	3.4%	3.5%

5.1.3 Cellular phone database

This cellular phones (Fig. 7) database has been created in our laboratory in order to illustrate the ability of Motion Descriptors and Zernike Moments to recognize similar objects. The database contains thus 20 objects (phones) and 300 images by object. The acquisition protocol is similar to the COIL acquisition, since each object is put on a turntable in order to perform an acquisition each 1.2 degree.

Applied on cellular phone database, Motion Descriptors and Zernike Moments (and combination) give both a null error using cross validation.

Fig. 7. Sample objects of the cellular phone database

We also studied the influence of the number of samples used during the learning step. The results are reported in the Fig. 8. Motion Descriptors are globally more efficient than Zernike Moments, and one can note, as in the COIL case, that the combination of both descriptors allows converging faster, since the error e<2% is obtained when only 3% of available samples are used during the training step.

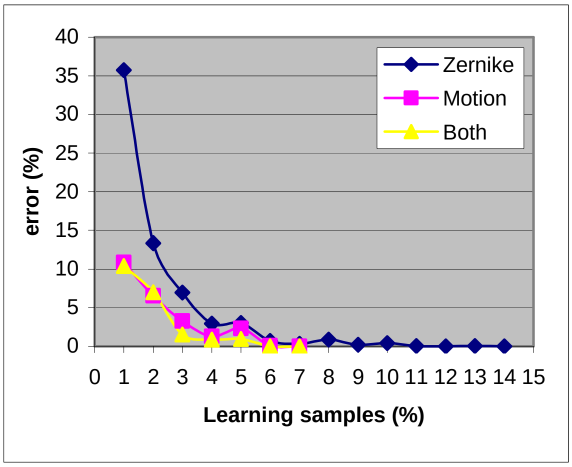

Fig.8. Influence of number of training samples of COIL

5. Conclusion

We proposed in this paper an evaluation of performance of Motion Descriptors compared and combined with Zernike Moments applied to color object recognition. The descriptors have been defined and their properties reviewed. Using standard databases of pattern recognition we shown that Motion Descriptors often outperform Zermike Moments, and that combination of both descriptors allows limiting the number of samples used during the training step of the classification process. These descriptors can be used successfully in a pattern recognition task for which rotation, scale and translation invariant is important.

We built software working in real-time using a standard PC architecture. During the training step, the user as to record a few images of the object to be recognized. The decision step (including resampling, Motion Descriptors computation and SVM prediction) is performed in 50ms on a Pentium IV, 1.5 GHz. Moreover, it is also possible to compute Zernike Moments in real time [18].

In future work, we intend to add a new family of invariants, and cooperation between local and global approaches will be tested for shape indexing.

REFERENCES

1. 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.
2. H. Fonga, G. Bornard, J.P Gauthier, "Analyse harmonique sur les groupes et reconnaissance des formes : calcul des descripteurs de Fourier Généralisés", 23ème congrès national d'analyse numérique, Royan 26-29 mai 1991.
3. R. Mukundan, K.R. Ramakarishnan,"Moment Functions in Image Analysis-Theory and Applications", *World Scientific*, Singapore, 1998.
4. A. Khotanzad, H.H. Yaw, "Invariant image recognition by Zernike moments", *IEEE Trans. PAMI*, vol. 12, no. 5, 1990, pp 489-497.
5. http://www.cs.columbia.edu/CAVE/
6. H. Murase and S. K. Nayar, "Visual learning and recognition of 3D objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, 1995, pp. 5-24.
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. Chee-Way Chong, P. Raveendran, R. Mukundan, "Translation invariants of Zernike moments", *Pattern Recognition* 36, 2003, pp 1765-1773.
9. V. N. Vapnik, *The statistical Learning Theory*, Springer, 1998.
10. 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.
11. V. N. Vapnik and A. Ja. Chervonenkis, "On the uniform convergences of relative frequencies of events to their probabilities" Theory Probab Appl., vol. 16, 1971, pp. 264-280.
12. S. Obrzalek and J. Matas, "Object recognition using local affine frames on distinguished regions," *in Electronic Proceeding of the 13th british Machine Vision Conference*, University of Cardiff, 2002, pp 113, 122.
13. M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3 no. 1, 1991, pp 71-86.
14. Erik Hjelmas, "face detection: A Survey", *Computer Vision and Image Understanding*, no. 83, 2001, pp 236-274.
15. R. Huang, V. Pavlovic, D. N. Metxas, "A hybrid Face Recognition Method using Markov random Fields », in *Proceedings of ICPR 2004*, pp. 157-160.
16. P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Trans-PAMI*, vol. 19, no. 7, 1997, pp 711,720.
17. S. Lawrence, C. L. Giles, A. C Tsoi, and A. D. Back, "Face recognition: A convolutional neural network approach", *IEEE trans. Neural Networks*, vol. 8, 1997, pp 98-113.
18. R. Mukundan, K. R. Ramakrishnan, "Fast Computation of Legendre and Zernike Moments", Pattern Recognition, vol. 28, no. 9, 1995, pp. 1433-1442.