Colour Object recognition combining Motion Descriptors, Zernike Moments and Support Vector Machine

Fethi SMACH
fethi.smach@ieee.org

Cédric LEMAITRE
Cedric.lemaitre@u-bourgogne.fr

Johel MITERAN
Johel.miteran@u-bourgogne.fr

Jean Paul GAUTHIER
gauthier@u-bourgogne.fr

Université de Bourgogne, Laboratoire Le2i
BP 47870   21078 Dijon
France

Abstract – **Fourier descriptors have been used successfully in the past to grey-level images, rigid bodied object. Here we used Motion Descriptors (MD) introduced recently by Gauthier et al., combined with Zernike Moments (ZM), in order to perform a recognition task in colour images. The feature vector for the MD obtained for each object appears to be unique and can be used for shape recognition. The MD, alone or combined with ZM, are used as an input of a Support Vector Machine (SVM) based classifier. We illustrate results on three available datasets: ORL faces database, COIL-100, which consists of 3D objects and A R faces.**

I. INTRODUCTION

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 colour 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 theoretically and practically the ability of such descriptors to be used successfully in colour pattern recognition, and also combined with another well known set of descriptors: the Zernike Moments [3], [4]. We present results on experiments done with standard databases in the object recognition community: the COIL databases [5], [6] which contain images from 100 objects rotated on a turntable (72 images for each object, i.e. images taken every 5 degree). ORL and A R face databases.

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.

II. REVIEW OF MOTION DESCRIPTORS

A. 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.

B. 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 $g(x) = foM(x)$, 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 $g(x) = fo\Re(x)$, 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)

III. 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 [7]

$$z_{pq} = \frac{p+1}{\pi}\int_{0}^{1}\int_{-\pi}^{\pi} V_{pq}(r,\theta) r dr d\theta, \quad | r |\leq 1 , \qquad (6)$$

where the Zernike polynomials are defined as:

$$V_{pq}(r,\theta) = R_{pq}(r)e^{-jq\theta} \qquad (7)$$

The real-valued radial polynomials:

$$R_{pq}(r) = \sum_{S=0}^{\frac{p-|q|}{2}} (-1)^{s} \frac{(p-s)!}{s!(\frac{p-2s+|q|}{2})!(\frac{p-2s-|q|}{2})!} r^{p-2s} \quad (8)$$

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 [7], Chee-Way chong, presents a mathematical framework for the derivation of translation invariants of radial moments defined in polar form.

IV. REVIEW OF SVM CLASSIFICATION

A Support Vector Machine (SVM) is a universal learning machine developed by Vladimir Vapnik [8]. In 1979, 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

$$\mathrm{W}(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathrm{x_i},\mathrm{x_j}), \qquad (9)$$

under the constraints

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \qquad (10)$$

and $0 \leq \alpha_i \leq T$ for i=1, 2, …, n where $\mathrm{x_i} \in R_d$ are the training sample set vectors, and $y_\mathrm{i} \in \{-1,+1\}$ the corresponding class label. *T* is a constant needed for nonseparable classes. $K(\mathrm{u},\mathrm{v})$ 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 $K(\mathrm{u},\mathrm{v})$ be a symmetric function which satisfies the following general positive constraint:

$$\iint_{R_d} K(\mathrm{u},\mathrm{v})\mathrm{g}(\mathrm{u})\mathrm{g}(\mathrm{v})\mathrm{d}\mathrm{u}\mathrm{d}\mathrm{v} > 0 , \qquad (11)$$

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

$\int g^2(\mathrm{u})\ \mathrm{du} < \infty$ (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:

$$K(\mathrm{u},\mathrm{v}) = \sum_{k} a_k\Phi_k(\mathrm{u})\Phi_k(\mathrm{v}), \qquad (12)$$

where a_k are positive scalars and the functions Φ_k represent a basis in the space *Q*. Vapnik considered three types of SVMs [9]:

Polynomial SVM:

$$K(\mathrm{x},\mathrm{y}) = (\mathrm{x.y} + 1)^p \qquad (13)$$

Radial Basis Function SVM (RBF):

$$K(\mathrm{x},\mathrm{y}) = e^{\left(\frac{-\|\mathrm{x}-\mathrm{y}\|^2}{2\sigma^2}\right)} \qquad (14)$$

Two-layer neural network SVM:

$$K(\mathrm{x},\mathrm{y}) = Tanh\{k(\mathrm{x.y}) - \Theta\} \qquad (15)$$

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:

$$\mathrm{C}(\mathrm{x}) = \mathrm{Sgn}\left(\sum_{i=1}^{Ns} y_i\alpha_i \cdot K(\mathrm{s}_i,\mathrm{x}) + b\right), \qquad (16)$$

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 [10].

V. OBJECT RECOGNITION PROCESS AND EXPERIMENTAL RESULTS

In order to validate our approach, we performed a cross validation test using three distinct databases: the ORL [11], the COIL-100 [4] and the A R face color database [12].

A. Databases

1) ORL database

The ORL database (Fig. 1) used in this paper is composed of 400 grey level images of size 112x92; there are 40 persons with ten images per person. The images are taken at different time instances, with varying lighting conditions, facial expressions (open/closed eyes, smiling/no-smiling), and facial details (glasses/no glasses). All the subjects are in up-right, frontal position (with tolerance for some pose variation)

Fig. 1. Face samples from the ORL database

2) COIL-100 database

COIL-100, the Columbia Object Image Library (COIL-100, Fig. 2) [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. 2. Several objects from COIL-100 database

3) AR Face database

This face database (Fig. 3) was created in the computer vision center. It contains over 4.000 colour 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), [12]. Each image in the database consists of a 786x576 array of pixels, and each pixel is represented by 24 bits of RGB colour.

Fig. 3. Face samples from the A R database

B. Test protocol

1) Training step

During the training step (Fig. 4), 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.

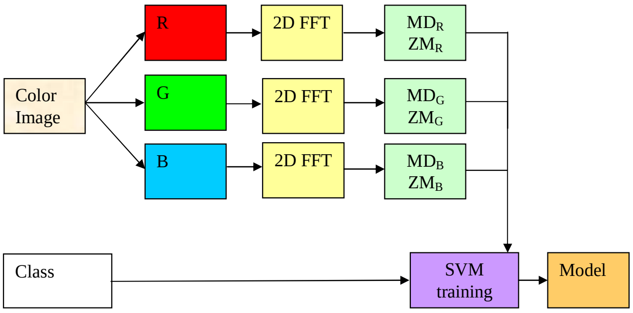

Fig. 4. Training process

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. 5).

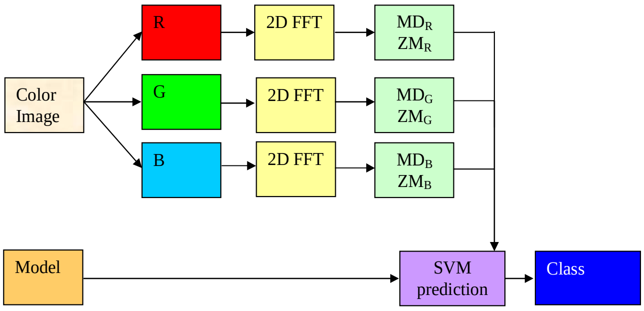

Fig. 5. Decision process

The classification error rate was obtained using a ten-fold based 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 space is d=189+42=231.

Since we used the RBF kernel in the SVM classification process, we have to choose 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 which gave the minimum error rate.

C. Numerical Results

1) ORL database

Published results in the literature range from 7.5% to 0% error rate [13], [12]. The protocol used for learning and testing is different from one paper to another. In [14], Hjelmas reported a classification error rate e=15% using the ORL dataset and feature vector consisting of Gabor coefficients. In [15], the PCA based method [16], LDA-based method [17], and a nearest neigbor-based method where tested for comparisons. With 10 images of each subject for training, the error rate is 6.25% with LDA-based method and the best performance is an error of 2.1% with NN-based method.

In [18], a hidden Markov model (HMM) based approach is used, and the best model resulted in a 13% error rate. Lawrence et al [19] takes the convolutional neural network approach for the classification of ORL database, and the best error is 3.83%.

We performed experiments on the ORL database using the Zernike moments and motion descriptors. The result is shown in table 1.

Experimental results show that the performance of our approach is comparable and sometimes better than HMM and LDA based methods, and that the Motion Descriptors and the Zermike Moment are complementary, since the use of both feature vectors allows dividing the error by two. However, best results will be obtained with colour databases.

Table 1: Error rate on ORL database

SVM Kernel RBF	Zernike moments	Motion descriptors	*Motion-Descriptors and Zernike Moments*
$\sigma = 0.1$	25%	9.5%	4.25%

2) COIL-100 database

Table 2 achieve the result obtained testing our object recognition method with the COIL-100. Tests have been done using 2-fold cross validation and 5-fold cross validation. Optimum error values are depicted in red. These two experiments illustrate the fact that increasing the number of sample images during the training step improves the performance from e=1% to e=0.01%.

Table 2: Error rate on COIL-100 database

SVM Kernel RBF (CV/ σ)	Zernike Moments	Motion Descriptors	*Motion-Descriptors and Zernike Moments*
2/0.1	1.89 %	38.48 %	16.47 %
2/1	0.78 %	15.41 %	3.40 %
2/10	1.89 %	3.87 %	1.00 %
2/100	23.33 %	1.69 %	3.10 %
5/1	0.22 %	0.09 %	0.01 %

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 [20].

In our global approach, the error e=0.01% corresponds to only 1/7200 image classified faulty.

3) AR Face 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=2.35%, 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=2.6%. However, our approach gives better results than in [12], 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 Kernel RBF	Zernike moments	Motion descriptors	*Motion-Descriptors and Zernike Moments*
$\sigma = 0.1$	35.12%	2.35%	2.6%

VI. CONCLUSION

We proposed in this paper an evaluation of performance of Motion Descriptors combined with Zernike Moments applied to colour object recognition. The descriptors have been defined and their properties reviewed. Using standard databases of pattern recognition we shown that 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 [21].

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.

VII. REFERENCES

[l] 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. BORNNARD, J.P GAUTHIER. « Analyse harmonique sur les groupes et reconnaissance des formes : calcul des descripteurs de fourier Généralisés. 23éme 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] Chee-Way Chong, P. Raveendran, R. Mukundan, "Translation invariants of Zernike moments", *Pattern Recognition* 36, 2003, pp 1765-1773.

[8] 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.

[9] V. N. Vapnik. *The statistical Learning Theory*. Springer, 1998.

[10] V. N. Vapnik and A. Ja. Chervonenkis, "On the uniform co,vergences of relative frequencies of events to their probabilities," *Theory Probab Appl.*, Vol. 16, 1971, pp. 264-280.

[11] ORL face database, AT&T Laboratories, Cambridge, U.K. http://www.cam-orl.co.uk/facedatabase.html

[12] 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.

[13] G. –D. Guo, S. Li, and k. Chan, "Face recognition by support vector machines," in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2000, pp 196-201.

[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] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3 no. 1, 1991, pp 71-86.

[17] 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.

[18] F. S. Samaria and A. C. Harter, "Parametrization of a stochastic model for human face identification" in *Proceedings of the 2nd IEEE workshop on Applications of Computer Vision*, 1994.

[19] 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.

[20] 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.

[21] R. Mukundan, K. R. Ramakrishnan, "Fast Computation of Legendre and Zernike Moments", Pattern Recognition, vol. 28, no. 9, 1995, pp. 1433-1442.