GENERIC TARGET RECOGNITION

C. Rosenberger[1] *A. Rakotomamonjy*[2] *B. Emile*[1]

[1] Laboratoire Vision et Robotique, UPRES EA 2078
ENSI de Bourges - Université d'Orléans
10 boulevard Lahitolle, 18020 Bourges, France
phone: +33 2248484000, fax: +33 2248484000
email:christophe.rosenberger@ensi-bourges.fr
[2] Laboratoire Perception, Systèmes et Informations, FRE CNRS 2645
INSA de Rouen, Avenue de l'Université, 76801 Saint Etienne du Rouvray
phone: +33 232959703 , fax: +33 232959708
email:alain.rakoto@insa-rouen.fr

ABSTRACT

We present in this paper a study on target recognition. The goal of this work is to determine and compare different methods from the pattern recognition domain in order to be able to recognize some objects in an image. We suppose having detected by a segmentation process a candidate object appearing with an unknown scale or rotation. To be able to recognize this object, we have first to describe it by some features having the property to be invariant by rotation, translation or scale. Second, we have to realize a supervised classification in order to compare this unknown object with one from the knowledge database. We present some experimental results for target recognition by comparing several features, classification methods and methodologies.

1. INTRODUCTION

The interpretation of images is still a complex problem and is primordial in lots of applications (spatial, military, industrial and medical). Since the beginning of the 80's, lots of research works have been achieved for the conception of vision systems in order to recognize objects in an image [15]. An essential stage concerns the strategy of object recognition because an object can appear at different places in the image or at different orientations and scales [21]. Several works have been dedicated to the definition of shape descriptors invariant by rotation and scale [22].

In order to identify an unknown object, a supervised classification method is generally achieved by taking into account these descriptors and a knowledge database. Object recognition consists therefore in extracting from the image a set of characteristic features. These features as well as the type of the different objects to discriminate are provided then to a classifier, in order to estimate the similarity of the unknown object with another one in the knowledge database. This approach, combining extraction of descriptors and discrimination, have been already extensively studied. Typically, in applications of image interpretation, classifiers such as neural networks [18], the nearest neighbor method [17] or methods based on probabilistic models [1] are often used. These methods also showed their efficiency on character recognition [4] and have been for a long time the most efficient methods. Support Vector Machines (SVM), based on theoretical concepts developed by Vapnik [20] become currently to maturity. Effectively, they proved to be very efficient for real problems such as color image recognition [6], face recognition [16].

The objective of this communication is to evaluate some features, supervised classification methods and methodologies for the conception of a target recognition system. First of all, the performance of some shape descriptors, invariant by rotation, translation and scale are studied. Second, we compare different supervised classification methods in order to define a generic method for target recognition having these properties. We show the interest of this combination in order to get a good property of invariance. We illustrate this methodology by some experimental results for target recognition, and more especially for planes. We also evaluated different methodologies: computation of the features on the contour of the object, on the binary region or by using the gray-level of the object.

2. PATTERN RECOGNITION

In order to recognize an object in an image, we need to make two choices. The first one concerns the selection of a characteristic feature of each object. This feature must have in general some properties as invariance by rotation, scale or translation of the object. It can be directly computed on the original image or afterwards a segmentation result as for example a contour detection. The second choice concerns the decision criteria for the object recognition among one of its known objects in the knowledge database using the previous features.

2.1 Features invariant by translation, rotation and scale

Lots of works have been achieved to solve the more general problem of object recognition invariant by rotation and scale. There exists several approaches. On the one hand, we find non-parametric methods based on the projection of objects on an appropriated basis of functions. These methods are extensively used in the field of character recognition. Hu's moments [14], Zernike's moments [19] or the Fourier-Mellin parameters [12] are such examples of non parametric object features. On the other side, there exists methods based on the use of neural networks for the object recognition invariant under a transformation group [22]. The idea is to present to the neural network during its training different orientations and scales of the same object.

In this communication, we are going to evaluate the performance of two non parametric features : Fourier-Mellin parameters and Zernike's moments. These features showed their efficiency in previous studies [12]. We also evaluate the ability of SVM to achieve an invariant object recognition under a transformation group.

Translation invariance

In order to guarantee the translation invariance, these features are computed after a preliminary stage on the image. We suppose in this study, to process small images containing only one object. It is then possible to compute the barycentre of the object. We can bring back the object on the center of the region of interest by translation.

Rotation and scale invariance

• Zernike's moments are calculated as following, from the object $g(\rho,\theta)$ expressed, after interpolation, in polar coordinates:

$$A_{g[p,q]}(\rho,\theta) = \frac{p+1}{\pi}\int_{\theta=0}^{2\pi}\int_{\rho=0}^{\infty} g(\rho,\theta)\ \mathscr{Z}^{*}_{[p,q]}(\rho,\theta)\ \rho\ d\rho\ d\theta \quad (1)$$

where

$$\mathscr{Z}^{*}_{[p,q]} = \sum_{s=0}^{(p-|q|)/2} \frac{(-1)^s\ [(p-s)!]\ \rho^{p-2s}\ e^{-jq\theta}}{s!\ ((p-|q|)/2-s)!\ ((p-|q|)/2-s)!} \quad (2)$$

Complex moments $A_{g[p,q]}$ are known to be invariant by rotation and scale. We used the ©Matlab implementation achieved in [5].

• Fourier-Mellin transform (FMT), that corresponds to the generalization of the Fourier transform with the group of positive similitude is applied on the image in polar coordinates. More precisely, we use the analytic extension of the FMT [12] given by:

$$M_g(\nu,q) = \int_{\theta=0}^{2\pi}\int_{\rho=0}^{+\infty} \rho^{-i\nu+\sigma_0}\ qe^{-i\theta}\ g(\rho,\theta)\ \frac{d\rho}{\rho}\ d\theta \quad (3)$$

$$\text{with } q\in\mathbb{Z},\ \nu\in\mathbb{R},\ \textit{and}\ \sigma_0\in\mathbb{R}^{*}_{+}$$

This particular transformation avoids the divergence of the Fourier-Mellin integral that appears in most of the practical situations. In [12], the following set of scale and rotation invariant features has been proposed :

$$I_g(\nu,q) = M_g(\nu,q)K_g(\nu,q) \quad (4)$$

$$K_g(\nu,q) = \left[M_g(0,0)\right]^{-1+i\frac{\nu}{\sigma_0}}\left[M_g(0,1)\right]^{-q}\left|M_g(0,1)\right|^{q} \quad (5)$$

2.2 Supervised classification

For target recognition, in this work, we will focus on the context of supervised learning. Our training set $\{\mathbf{x}_i, y_i\}_{i=1\cdots\ell}$, where each $\mathbf{x}_i\in\mathbb{R}^d$ and $y_i\in\{1,\cdots,N\}$ in the case where we try to recognize N different classes, consists in all the previous features and the class of each object in the knowledge database. Our objective is to determine a function $f(\mathbf{x})$ that estimates dependencies between the $\mathbf{x}_i$ and the y_i and that minimizes the risk of error classification for a given point $\mathbf{x}$ not belonging to the training database. There exists several supervised learning methods among which we can mention distance minimization based algorithms, Bayesian methods [8] and connexionnist methods [2].

We present in this communication some results of object recognition by using these three kinds of classification methods.

2.2.1 *Minimization distance*

Every object class C_i is represented by the mean vector of features denoted $E[\mathbf{a}_i]$. An object with a vector of features $\mathbf{a}$ is affected to the C_i class if and only if:

$$i = \arg\min_{j=1..n} D(\mathbf{a}, E[\mathbf{a}_j]) \quad (6)$$

where D corresponds to a distance (in this communication, it is the Euclidean distance). This classifier corresponds to a maximal *a posteriori* classifier in the case where each class can be modelled by a Gaussian distribution probability and where the *a priori* probability of each class is the same [9].

2.2.2 *Fuzzy classifier based on the k nearest neighbors*

This fuzzy classification method defined by Charroux *et al.* [7] provides the definition of a degree of belonging of an unknown object without having any *a priori* knowledge on the distribution of observations. This method is particularly adapted when distributions of the observations are unknown. This method is based on the Grenier's algorithm [7], whose principle is to calculate a potential vector where each component gives a degree of belonging to one of the classes.

2.2.3 *Support Vector Machines*

For two classes problems, $y_i\in\{-1,1\}$, the Support Vector Machines implement the following algorithm. First of all, the training points $\mathbf{x}_i$ are projected in a space $\mathscr{H}$ (of possibly infinite dimension) thanks to a function $\Phi(\cdot)$. Then, the goal is of to find in this space, an optimal decision hyperplane, in the sense of a criterion that we are going to define [20]. Note that for a same training set, different transformations $\Phi(\cdot)$ lead to different decision function. The transformation is achieved in an implicit manner thanks to a kernel $K(\cdot,\cdot)$ and the decision function is defined as :

$$f(\mathbf{x}) = \langle w, \Phi(\mathbf{x})\rangle + b = \sum_{i=1}^{\ell}\alpha_i^{*} y_i K(\mathbf{x}_i,\mathbf{x}) + b \quad (7)$$

with $\alpha_i^{*}\in\mathbb{R}\ \forall i$. In SVMs, the optimality criterion to maximize is the margin, that is the distance between the hyperplane and the nearest point $\Phi(\mathbf{x}_i)$ of the training set (see Figure 1 in the case where $\Phi(\cdot)$ is the identity function). The α_i^{*} allowing to optimize this criterion are defined by solving the following problem:

$$\begin{cases} \max_{\alpha_i}\sum_{i=1}^{\ell}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{\ell}\alpha_i\alpha_j y_i K(\mathbf{x}_i,\mathbf{x}_j y_j) \\ \text{with constraints,} \\ 0\ \le\ \alpha_i\ \le\ C\,, \\ \sum_{i=1}^{\ell}\alpha_i y_i\ =\ 0\,. \end{cases} \quad (8)$$

where C is a coefficient penalizing examples located in or beyond the margin and providing to reach a compromise between their numbers and the width of the margin.

Originally, SVMs have essentially been developed for the two classes problems, however several approaches can be used for extending to multiclass problems [13]. The method we use in this communication, is called *one against one*. Instead of learning N decision functions, each class is here discriminated from another one. Thus, $\frac{N(N-1)}{2}$ decision functions are learned and each of them makes a vote for the affectation of a new point $\mathbf{x}$. The class of this point $\mathbf{x}$ becomes then the majority class after the voting [11].

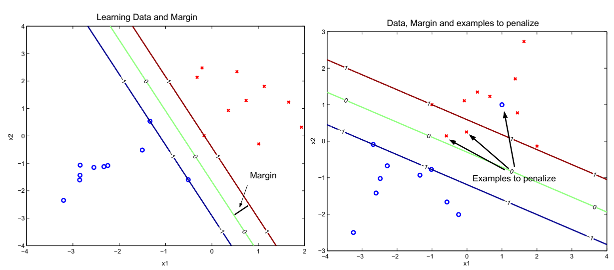


Figure 1: Illustration of the SVM discrimination for linearly separable (on the left) and no separable data (on the right).

3. EXPERIMENTAL RESULTS

In order to evaluate the efficiency of these different techniques, we achieved different experiments for target recognition applications. The first one consists in recognizing some views of planes at different orientations and scales. The second experimentation consists to evaluate the performance of these techniques for the recognition of 3D planes from different 2D views.

3.1 Recognition of 2D objects

We used first of all, an image database composed of 10 planes [10] (see figure 2). We generated a set of 360 images of size 65 $\times$ 65 pixels for each plane by successive rotations of one degree angle. Then, for each orientation, two scales modification are applied (20% and 50%). We had for our experiments a database of 10800 images for each plane.

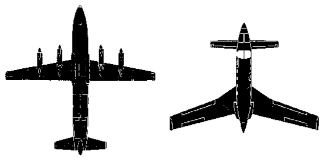
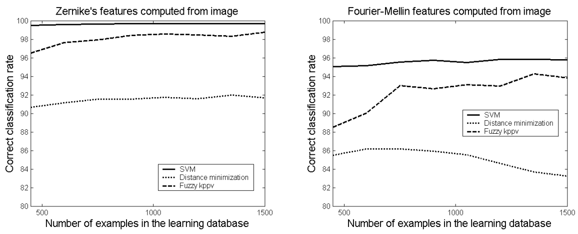


Figure 2: 2 examples in the target database

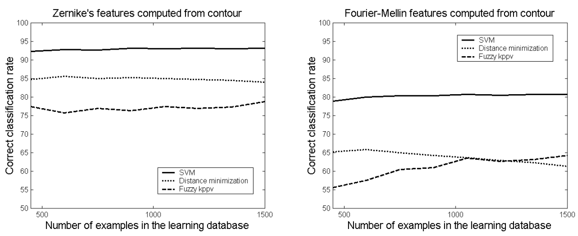
At the beginning, each image has been segmented to obtain the shape of the plane. Then, we calculate the Fourier-Mellin parameters and Zernike's moments for each picture. We have thus one vector of 36 Zernike's moments and 33 Fourier-Mellin parameters for each plane.

Our methodology for the discrimination of the classes is the following : from the previous database, a training set and a test set have been created. The training database is composed, for each class, of a number of N examples distributed uniformly on the three scales while the rest of database is used as test. Our experiments consist in studying the performance of the different classifiers according to the number of training examples.



(a) Zernike (b) Fourier-Mellin

Figure 3: Performance of classifiers according to the number of examples for each class with features computed on the image



(a) Zernike (b) Fourier-Mellin

Figure 4: Performance of classifiers according to the number of examples for each class with features computed on the contour

A multiclass SVM based on the *one-against-one* with a linear kernel was used first. The penalization parameter C of badly classified training points has been fixed to 1000. This value permits to guarantee that the number of errors during the training phase remains low. The best parametrization of the fuzzy classifier is achieved for a high value of k, we used in our experiments $k = 13$.

Recognition of binary objects

First of all, we used a binary segmentation result of this database for the training and recognition steps. An unknown object is represented by its contour or a binary region. Figures 3 and 4 present the classification results of targets by using the two types of invariants and the three supervised classification methods described previously. One can note on the one hand, the efficiency of the SVM for the recognition compared to the two other methods.

On the other hand, Zernike's moments give in this case a better recognition rate (increase of 4 to 5%) compared to Fourier-Mellin parameters. By using Zernike's moments, we reach an excellent recognition rate equal to 99% with only 15 examples for each plane (with different orientations and scales) in the training phase.

As we could wait for it (in the measure where the quantity of information is less important), the classification results are worst when we use only the contour of planes (decrease in performance of 5 % for the recognition rate). It shows the interest to exploit all the information in an image compared to the use of a segmentation result.

3.2 Recognition of 3D objects

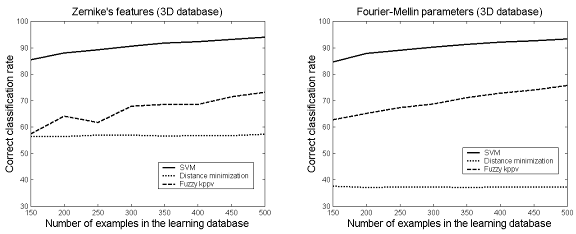
For the recognition of 3D planes from 2D views, the training database is composed of 648 different views for each plane (see in Figure 5 some binary segmentation results).

Figure 5: Different views of a MIG29 plane for the training phase

As previously, the image database is separated in a training set and a test set. We are going to use the same features and the three previous supervised classification methods. For this problem, the SVM used is based on a polynomial kernel of order 2 (we have therefore a non-linear decision border) and for the same reasons that previously, the penalization coefficient of C is still fixed to 1000. For the fuzzy classification algorithm, we chose $k = 11$ (for computation time reasons). The Figure 6 presents the average rate of correct classification obtained on 20 tests. Each curve is function of the number of examples, in the training database. We remark this time that the two types of features have an equivalent behavior and we reach a correct classification rate equal to 95% with 500 examples.

Contribution of the combination

In order to identify the contribution of the combination between the features and the SVM, we achieved one experimentation of recognition on the same database by using as a characteristic of an object, the image itself. In this case, each target is represented by a vector of size 65×65 corresponding to the value grey-levels of each pixel. Classification results are not better than 20% of good classification rate putting in evidence the interest of the combination of these two approaches instead using only the SVM.



(a) Zernike (b) Fourier-Mellin

Figure 6: Recognition of 3D objects

Recognition of grey-levels objects

We also tried to recognize the same targets by using the grey-levels of each plane. We present in the figure 7 the classification results by using the Fourrier-Mellin parameters (Zernike's moments give similar results) with the 3 classification methods. We obtained a good increase of the performance of recognition of order 5%.

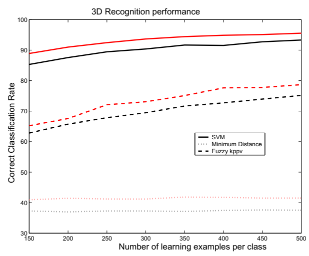


Figure 7: Recognition of 3D objects by using grey-levels targets

4. CONCLUSIONS AND PERSPECTIVES

We studied in this paper different methods for target recognition. We combined invariant features and one supervised classification method. We put in evidence on the one hand, the efficiency of Zernike's moments and Fourier-Mellin parameters. On the other hand, the performance of the SVM has been highlighted with regard to a fuzzy classifier using the K nearest neighbors and the minimization of a distance. We obtained on a significant image database excellent recognition results of 2D planes and good results from 2D views of 3D planes.

Perspectives of this study concerns the quantification of the impact of a background all around the object and screening problems for the target recognition.

REFERENCES

[1] R. Bailey and M. Srinath. Orthogonal moment features for use with parametric and non-parametric classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):389–399, 1996.

[2] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995.

[3] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *5th Annual ACM Workshop on COLT*, pages 144–152, Pittsburgh, PA, 1992. ACM Press.

[4] C. Cariou, J.M. Ogier, S. Adam, R. Mullot, J. Gardes, and Y. Lecourtier. A multi-scale and multi-orientation recognition technique applied to document interpretation : Application to french telephone network maps. *International Journal Pattern Recognition Artificial Intelligence : special number on invariants*, 1:1201–1218, 1999.

[5] Kui-yu Chang and J. Ghosh. Three-dimensional model-based object recognition and pose estimation using probabilistic principal surfaces. In *SPIE:Applications of Artificial Neural Networks in Image Processing V*, pages 192–203, 2000.

[6] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. *IEEE Transaction on Neural Networks*, 10(5):1055–1064, 1999.

[7] B. Charroux. *Analyse d'images : coopération d'opérateurs de segmentation guidée par l'interprétation*. PhD thesis, universit de Paris XI Orsay, 1996.

[8] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley Interscience, 2001.

[9] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley Interscience, 2001.

[10] G. Endres and M. Gething. *Jane's Aicraft Recognition Guide*. Harper Collins publishers, 2002.

[11] J. Friedman. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, 1996.

[12] F. Ghorbel. A complete invariant description for gray-level images by the harmonic analysis approach. *Pattern Recognition Letters*, 15:1043–1051, 1994.

[13] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425, 2002.

[14] M. K. Hu. Visual pattern recognition by moment invariants. *IEEE Transactions Information Theory*, 8:179–187, 1962.

[15] D. Marr. *Vision*. W.H. Freeman, 1982.

[16] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. IEEE International Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997.

[17] P. Simard, Y. LeCun, and J. Denker. Efficient pattern recognition using a new transformation distance. In Morgan Kaufman, editor, *Advances in Neural Information Processing Systems 5*, pages 50–58, 1993.

[18] S. Singh and A. Amin. Neural network recognition of hand-printed characters. *Neural Computing and Applications*, 8(1):67–76, 1999.

[19] M. Teague. Image analysis via the general theory of moments. *Journal Optical Society of America*, 70:920–930, 1980.

[20] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[21] L.-C. Wang, S. Z. Der, and N. M. Nasrabadi. Automatic target recognition using a feature-decomposition and data-decomposition modular neural network. In *in Proceedings International Conference SPIE*, volume 3307, pages 02–13, San Diego, USA, 1998.

[22] J. Wood. Invariant pattern recognition: A review. *Pattern Recognition*, 1(29):1–17, 1996.