

Pattern Recognition and feature extraction: a comparative study

VINCENZO NIOLA

Department of Mechanical Engineering for Energetics
University of Naples "Federico II"
Via Claudio, 21 - 80125 Napoli
ITALY

GIUSEPPE QUAREMBA

University of Naples "Federico II"
Via S. Pansini, 5 - 80131 Napoli
ITALY

Abstract: - The selection of features for classifying a pattern by means a fuzzy reasoning, is fundamental in order to obtain a reliable and significative response. The scope of this work is to compare three methods specialized for the extraction of features from images and, consequently, to study the ability of classification performed by applying a fuzzy inference system. The methods to be compared were: Fourier descriptors, Zernike moments and Wavelet coefficients. The best result, in terms of the best performances obtained both as classification reliability and computational time, was represented by the application of wavelet transform.

Key-words: - Fuzzy logic, Image analysis, Pattern recognition, Feature selection, Fourier descriptors, Zernike moments, Wavelet coefficients.

1 Introduction

The *feature extraction* is a typical problem of Pattern Recognition (PR), in which the events to be classified are represented by means of images where, in general, the information is distorted for the presence of noise [1]. Sometimes the information is extracted by performing algorithms of Image Analysis (IA) in the case of system based on the vision [2].

Generally, the problem of PR consists of two sub-problems:

- to provide an inner representation of input data (i.e., pattern);
- to perform a decisional process as well as to assign the pattern to its belonging dominion.

In this way, the first problem is to find the pattern, starting from the input data (e.g., image, data stream produced by sensors, etc.) and to produce an inner representation of the pattern as output.

This inner representation (often a lot of features extracted from the input data), have to be classified into belonging classes (if there exists).

The problem of recognition, as said before, is faced by means of techniques of IA.

The activity of interpretation and analysis of images is a complex operation indeed; in fact the image is simply like a table of gray values (i.e., the gray levels, represented as a matrix of pixel), with added disturbs, called noise, due to electrical instruments.

Usually, the activity of interpreting an image is influenced by the typology of the image to be recognized, consequently the choice of algorithms, specifically designed, must take into account such a problem.

The aim of this work is to compare three different methods for the calculation of the features to be submitted to a neural network for its training.

The data set was composed of 270 images (i.e., 200 pins and 70 nuts) of several dimensions and superficial characteristics (i.e., zined and burnished), placed in various positions. The 70% of original images, randomly selected, were used for the training set, while 30% for testing set. The images have been made by a digital camera with 3.2 million of pixel, without the use of flash and filter, with a dimension of 640x480 pixel. Calculations were made using Matlab 7.0, The MathWorks, Inc, Natick, Mass.

2 Image Analysis

The scope of the Image Analysis is to execute quantitative measures on image in order to obtain a description and, starting from that description, to interpret its content [3].

In general, an image can be thought as a two-dimensional continuous function that associates to a scene its two-dimensional representation. For instance, a monochromatic B/W photo can be associated to a function $f(x, y)$; the value assumed by the function in a point, represents the gray level of the image in that point. By means the technique of digitalization to each pixel is associated a value, that corresponds to the average value of gray in the corresponding area (i.e., quantization).

The aim of PR is to represent a digitalized image by a finite set of features extracted from the same image.

A feature, as said before, could be an object of high level: a geometrical descriptor of a region of the image or a geometrical object in 3D. The features can be represented as continuous, discrete or binary functions.

However it must notice:

- the process of feature extraction requires time consuming for the calculation;
- the extracted features could contain errors, due to noise or an erroneous application of algorithms.

The process of selection of features is one of the key problems for every system dedicated to the pattern recognition and it consists to decide which features to use, among all those available ones, for a specific problem.

It is important, for a reliable solution of such systems, to choose (and therefore to extract) features such that they:

- are important for the problem;
- are computing realizable;
- give the minimum number of misclassifications;
- reduce the problem to a minimum number of data without loss of information.

Sometimes a mathematical approach can help us to choose the most appropriate features, in other cases it could be more useful to perform some simulations. However, many times the choice is based on the experience.

As the features are grouped in array of n dimensions the space, in which the problem is involved, is represented by R^n .

The key point of IA is the automatic recognition of an object in a scene, independently from its position, dimension and direction. Learning Vector Quantisation (LVQ) is a supervised version of vector quantisation, similar to Self-Organizing Maps (SOM) based on work of [4], [5] and [6] for a comprehensive overview). It can be applied to pattern

recognition, multi-class classification and data compression tasks, e.g. speech recognition, image processing or customer classification. As supervised method, LVQ uses known target output classifications for each input pattern of the form . The main idea is to cover the input space of samples with 'codebook vectors' (CVs), each representing a region labelled with a class. A CV can be seen as a prototype of a class member, localized in the centre of a class or decision region ('Voronoi cell') in the input space. As a result, the space is partitioned by a 'Voronoi net' of hyperplanes perpendicular to the linking line of two CVs (mid-planes of the lines forming the 'Delaunay net'. A class can be represented by an arbitrarily number of CVs, but one CV represents one class only. In terms of neural networks a LVQ is a feed-forward net with one hidden layer of neurons, fully connected with the input layer. A CV can be seen as a hidden neuron ('Kohonen neuron') or a weight vector of the weights between all input neurons and the regarded Kohonen neuron respectively

3 Mathematical background

3.1 Zernike moments

The moments and functions of moments have been used in order to obtain features in numerous PR applications applied to two-dimensional images [7]. Since 1961 the moment invariants were introduced by using a linear combination of regular moments [8]. These not linear functions have the property of being invariant for translation, rotation and scale. Since such features capture total information on the image, than they have to be applied to the entire image.

Let be $f(x, y)$ a continuous intensity function of the image in a point (x, y) ; the regular moments m_{pq} of order $(p + q)$ are defined as [7]:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (1)$$

where $\{p, q = 0, 1, 2, \dots, \infty\}$.

In order to make the moments invariant to translation, central moments are defined as follows:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (2)$$

where

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}, \quad (p, q = 0, 1, 2, \dots, \infty)$$

The point (\bar{x}, \bar{y}) , calculated with (1), is the centroid of the image. By means of linear combination of moments several not linear functions can be derived, invariant for translation, scale and rotation [9]. The central moments normalized become invariant for scale [8], [9]:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = \frac{p+q}{1} + 1$$

where $\{p+q = 3, 4, 5, \dots, \infty\}$.

A non linear set of functions, invariant for rotation, translation and scale, are the following [8]:

$$\begin{aligned} \zeta_1 &= \eta_{20} + \eta_{02} \\ \zeta_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \zeta_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \zeta_4 &= (\eta_{30} - \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \end{aligned}$$

In many applications, in order to avoid the overflow, the logarithm is assumed: $\{\log(|\zeta_i|), i = 1, \dots, 4\}$.

On the base of Jain Theorem [3] the moments $\{m_{pq} | p, q = 0, 1, \dots, \infty\}$ show the property to represent uniquely the image.

Teague [10] suggested the use of orthogonal moments, based on the theory of orthogonal polynomials in order to eliminate the problem associated with regular moments, which show a redundancy due to the fact that the basis $x^p y^q$ is not orthogonal. He introduced the Zernike Moments as:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nm}(r, \theta)] * f(r \cos \theta, r \sin \theta) r dr d\theta$$

where

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right), \quad -1 < x, y < 1,$$

with m, n integer values;

$$0 \leq |m| \leq n, \quad n - |m| = \text{even}, \quad n > 0$$

3.2 Fourier descriptors

In the other side, the Fourier descriptors [3], are applied exclusively to the contour of the object to be recognized; it makes more sensitive to the noise and the perturbations showed by the contour. Based on a Fourier analysis technique applied to the boundary coordinates of an object expressed as complex numbers, Fourier descriptors are widely used in image processing to describe and classify shapes. The shape descriptors generated from the Fourier coefficients numerically describe shapes and can be normalised to make them independent of translation, scale and rotation. Classification is performed by comparing descriptors of the unknown object with those of a set of standard shapes, finding the closest match [11]. [12] introduced Fourier descriptors using complex representation in 1972. This method ensures that a closed curve will correspond to any set of descriptors.

The shape is described by a set of N vertices $\{z(i) : i = 1, \dots, N\}$ corresponding to N points of the outline.

The Fourier descriptors $\{c(k) : k = -N/2 + 1, \dots, N\}$ are the coefficients of the Fourier transform of z :

$$z_i = \sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} c_k \exp\left(2\pi j \frac{ki}{N}\right)$$

The inverse relationship exists between $c(k)$ and $z(i)$:

$$c_k = \frac{1}{N} \sum_{i=1}^N z_i \exp\left(-2\pi j \frac{ik}{N}\right)$$

The range of k can be restricted to $\{-N/2 + 1, N/2\}$: according to Shannon's theorem, the highest frequency is obtained for $k = N/2$, and any $c(k)$ with k greater than $N/2$ would be redundant since we use a discrete representation of the outline.

3.3 Continuous Wavelet Transform

The wavelets used in this paper are those proposed by [12] She constructed a series of mother wavelets (indexed by N and denoted by dbN) with each mother in the series having regularity proportional to N [14]. Each Daubechies' wavelet are compactly supported in the time domain. Typically wavelets of class mr are specifically constructed so that some properties are verified [15].

A mother wavelet ψ is a function of zero h-th moment:

$$\int_{-\infty}^{+\infty} x^h \psi(x) dx = 0, \quad h \in \mathbb{N}.$$

From this definition, it follows that, if ψ is a wavelet whose all moments are zero, also the function ψ_{jk} is a wavelet, where

$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k).$$

In fact, we have

$$\begin{aligned} & \int_{-\infty}^{+\infty} 2^{j/2} x^h \psi(2^j x - k) dx = \\ & = 2^{j/2} \int_{-\infty}^{+\infty} \frac{1}{2^j} \left(\frac{y+k}{2^j} \right)^h \psi(y) dy = \\ & = \frac{2^{j/2}}{2^{j(h+1)}} \int_{-\infty}^{+\infty} (y+k)^h \psi(y) dy = \\ & = \frac{2^{j/2}}{2^{j(h+1)}} \sum_{m=0}^h \binom{h}{m} k^{h-m} \int_{-\infty}^{+\infty} y^m \psi(y) dy = 0. \end{aligned}$$

Moreover, consider a wavelet ψ and a function ϕ such that $\{\{\phi_{j_0 k}\}, \{\psi_{jk}\}, k \in \mathbb{Z}, j = 0, 1, 2, \dots\}$ is a complete orthonormal system. By Parseval theorem, for every $f \in L^2(\mathbb{R})$, it follows that

$$f(t) = \sum_k a_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{j_1} \sum_k d_{jk} \psi_{jk}(t)$$

The decomposition of a function $f(t)$ by wavelet (i.e., the CWT) is represented by the following detail function coefficients:

$$d_{jk} = \int_{-\infty}^{+\infty} s(\tau) \cdot \frac{1}{\sqrt{2^j}} \psi\left(\frac{\tau - k}{2^j}\right) d\tau$$

and by the approximating scaling coefficients

$$a_{j_0 k} = \int_{-\infty}^{+\infty} s(\tau) \cdot \psi(\tau - k) d\tau$$

Note that d_{jk} can be regarded, for any j , as a function of k . Consequently, it is constant if the function $f(t)$ is a smooth function, having considered that a wavelet has zero moments.

To show the above mentioned property, it is sufficient to expand the function in Taylor series.

An example of wavelets is given by Daubechies' family $\{\text{dbN}, N = 1, 2, \dots\}$ [16]. It is

$$\text{supp } \phi \subseteq [0, 2N - 1], \quad \text{supp } \psi \subseteq [0, 2N - 1]$$

and

$$\int_{-\infty}^{+\infty} x^h \psi(x) dx = 0, \quad h = 0, 1, \dots, N - 1.$$

Moreover, there is the following smoothness property: for any $N > 2$, the D2N wavelets verify

$$\phi, \psi \in H\lambda N, \quad 0.1936 \leq \lambda \leq 0.2075,$$

where $H\lambda N$ is the Hölder smoothness class with parameter λ .

3.3.1 Two-dimensional Wavelets

Let us consider a two dimensional function $f(x, y)$ which are square integrable over the real plane: $f(x, y) \in L^2(\mathbb{R}^2)$. A wavelet basis for $L^2(\mathbb{R})$ is to take the simple product of one-dimensional wavelet:

$$\Psi_{j_1 j_2 k_1 k_2}(x, y) = \psi_{j_1 k_1}(x) \psi_{j_2 k_2}(y)$$

It is easy to show that Ψ 's as defined above are indeed wavelets and that they form an orthonormal basis for $L^2(\mathbb{R}^2)$.

It can be shown that the "detail space" W_j is itself made up of three orthogonal subspaces as follows:

$$\Psi^1(x, y) = \phi(x) \psi(y);$$

$$\Psi^2(x, y) = \psi(x) \phi(y);$$

$$\Psi^3(x, y) = \psi(x) \psi(y).$$

Mallat [16] notes that the three sets of wavelets correspond to specific spatial orientations: the wavelet Ψ^1 corresponds to the horizontal direction, the wavelet Ψ^2 with the vertical direction and Ψ^3 with the diagonal.

4 Results

The hardware used for the study was not particularly expensive or specialized for the image analysis. For every image of training set the features were calculated by applying the methods exposed in the previous paragraph. For giving a

total appraisal on the reliability of recognition, several testing sessions have been executed; the following parameters were estimated:

- recognition factor, denoted with ρ , estimates the efficiency of the recognizer, calculated as: # objects correctly recognized/# objects analyzed;
 - computational time: expressed in seconds, the time occurring for the identification of the object;
 - memory occupation;
 - number of features necessary for a complete inner representation of the pattern, indicated as F_{tot} .
- Since the first three parameters are depending from F_{tot} , it is important to investigate mainly on F_{tot} and to express the results of the parameters that we want to estimate by means this value.

An example of grouped objects is shown in the Fig.1. Note that the background was black, the objects were randomly positioned while their surface was zined or burnished.



Fig.1 An example of grouped pins and nuts

The methods were tested in several conditions: number of neurons and learning rate of LVQ; the responses are reported in the Fig.2 and 3 respectively. The total percentage of recognition for each method is reported in Fig.4.

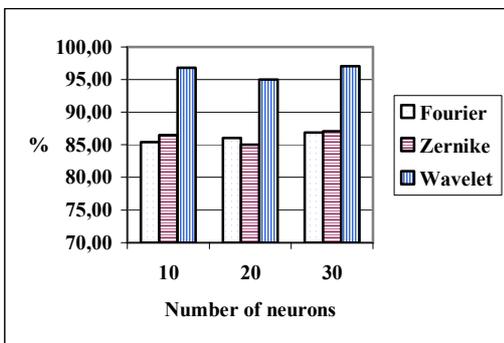


Fig.2 Percentage of recognition for various neurons ($\alpha = 0.1$)

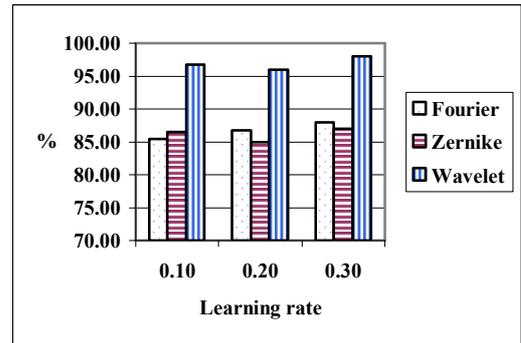


Fig.3 Percentage of recognition for various learning rate (10 neurons)

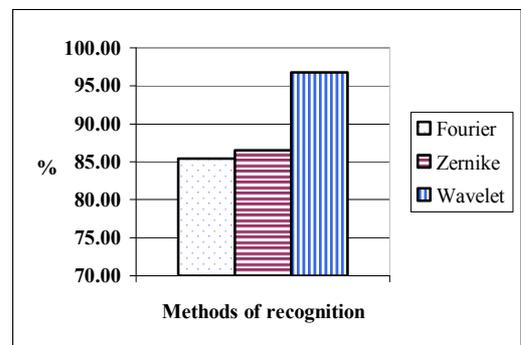


Fig.4 Test Set ($\alpha = 0.1$): percentage of recognition

The results presented, were obtained by using a vector composed of 10 features.

In the Tab. 1 below, are compared the computational time occurred for the identification of 30 objects, disposed in several positions, belonging to the Training Set, with reference to the applied method.

Method	Computational time (s)
Fourier descriptors	72
Zernike moments	108
Wavelet coefficients	215

Tab. 1 Comparison of computational time

5 Conclusions

A comparative study (both analytical and experimental) on the recognition rate for different extraction methods was performed. The comparison of three methods for detecting the pattern of several objects, showed the reliability and the robustness of wavelet coefficients in order to the pattern recognition. The wavelet application seems to represent the

best compromise among the factors indicating the performance of a complete system of Image Analysis. The results reached, compared with the ones obtained by means of other methods, demonstrates the validity of this choice. Since the training session is fast and easy, the neural networks are ideal in order to solve problems of pattern recognition. At this stage of the study we considered not important to choose a better neural network, in fact, in particular it should be more stimulating and interesting to study, in the future, the problem of pattern recognition in the case of imprecise or noised images in conjunction with the wavelet transform. The response, in terms of percentage of recognition, was also experimentally studied for several learning rates and number of neurons of LVQ. An application of the proposed net for object extraction based on noisy scenes is also tested. The recognition rate is quite stable if compared with various learning rate and number of neurons employed in the LVQ. Probably this is due to the fact that the noise was not added to the scenes. A complete investigation will be need in order to take into account higher dimensional feature space in presence of noised images.

References:

- [1] Jeffrey D. Ullman: *Pattern Recognition Techniques*. Butterworths, London, 1973.
- [2] Richard Watson: A survey of gesture recognition techniques. Technical report, *Dep. of Computer Science, Trinity College*, Dublin, 1993.
- [3] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.
- [4] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84--95, Jan. 1980
- [5] Gersho A. and Gray R.M., (1992), *Vector Quantization and Signal Compression*. Kluwer Academic Publishers.
- [6] Kohonen, T.; *The Self-Organizing Map*, *Procs. IEEE*, 78, 1464 ss, 1990.
- [7] C. Teh and R. T. Chin. On image analysis by the methods of moments. *IEEE Trans. on Pattern Analysis and machine Intellig.*, 10:496-512, July 1988.
- [7] M. K. Hu. Pattern recognition by moment invariants. In *Proc. IRE*, number 49, Sept. 1961.
- [8] T. H. Reiss. *Recognizing Planar Objects Using Invariant Image Features*. Springer-Verlag, 1993.
- [9] M. Teague. Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, August 1980
- [10] Keyes, L. and Winstanley, Adam C. (1999) *Fourier Descriptors as A General Classification Tool for*

Topographic Shapes. In Whelan, Paul F., Eds. *Proceedings Irish Machine Vision and Image Processing Conference*, pages 193-203, Dublin City University.

- [11] G.H. Granlund, "Fourier Pre-processing for hand print character recognition", *IEEE Trans. Computers*, Vol C-21, Febr. 1972, pp. 195-201.
- [12] Daubechies I., *Ten Lectures on Wavelets*, SIAM, 1992.
- [13] Strang G. and Nguyen T.: *Wavelets and Filter Banks*. Wellelsley-Cambridge Press, 1996.
- [14] Meyer Y.; *Wevelets and Operators* – Cambridge: Cambridge University Press, 1992.
- [15] Härdle, W. et al 1998. *Lecture Notes in Statistics: Wavelets, approximation and statistical applications*. New York: Springer.
- [16] S. Mallat (1989), "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Pattern Anal. And Machine Intell.* Vol. 11 n. 7 pp 674-693.