



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Iterative Feature Selection for Classification

T.B. Stambouli, M. Keche and A. Ouamri
 Laboratoire Signaux et Images, USTO, Spain

Abstract: In this study, we address the problem of image classification by using Bayes distance. We focus on the feature selection and we propose an iterative method which give significant improvements and allow the use of features provided by Gabor Transform in spite of its implicit redundancy. Only, a few numbers of features were necessary to reach this performance. Error rate falls from 71.48 to 27.29% for Gabor transform and from 31.25 to 16.8% for Wavelet transform. Thus, we show that choosing the best combination of features enhances classification.

Key words: Image classification, multiresolution analysis, Gabor transform, Separable Wavelet transform, Bayes rule

INTRODUCTION

Surface and object identification is fundamental to many applications such as industrial monitoring of product quality, remote sensing of earth resources and medical diagnosis.

Important characteristics can be provided by textures analysis which is still an interesting problem in image processing.

Gaussian Markov random field (Chellappa, 1985; Cross and Jain, 1983) and Gibbs distribution (Derin, 1986) where developed in the 1980's. These methods take into account only one scale. Later, methods based on multiresolution analysis (Bouman and Liu, 1991; Bovik *et al.*, 1990; Unser and Eden, 1989; El-Ramsisi and Khalil *et al.*, 2007) have been exploited. Classification using distance measures based on Bayes decision rule is often applied with success (Mojsilovic *et al.*, 1998; Ghafoor *et al.*, 2005). But it is not suitable when the number of features is too large. Thus, when using Gabor transform (Imine and Belbachir, 2006), it seems that Bayes distance is not adequate. We show that selecting a small set of features, allow a significant enhancement of the classification performance, even when applied to Separable Wavelet transform (Abdullah *et al.*, 2008).

FEATURE EXTRACTION

2D Gabor transform: The 2D Gabor Transform (Lee, 1996) of a given function $f(x, y)$ consists in determining coefficients obtained by carrying out a scalar product given by:

$$TGF(x_0, y_0, \theta) = \iint f(x, y)g(x - x_0, y - y_0, \theta)dx dy \quad (1)$$

where, $g(x, y, \theta)$ is the Gabor function of elliptic shape with θ as orientation:

$$g(x, y, \theta) = \frac{\omega_0}{\sqrt{2\pi\kappa}} e^{-\frac{\omega_0}{2\kappa} [4(x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2]} [e^{i(x\omega_0 \cos\theta + y\omega_0 \sin\theta)} - e^{-\kappa^2}] \quad (2)$$

In this Eq. $\kappa = \omega_0 \sigma$ depends on the band-width of the Gabor function (ω_0 : Radial frequency, σ : Standard deviation) and is used to eliminate DC component. One thus obtains a function which is admissible as a Wavelet, for a multi resolution analysis.

This transform is based on a 2D directional Gabor function. The latter consists of a directional elliptic Gaussian envelope modulated, according to the short axis of the ellipse, by oscillations. So, one can consider the Gabor transform as a kind of bandwidth filtering, according to the short axis, combined with a low pass filtering, according to the long axis. This characteristic makes it possible for isolating details according to one privileged direction.

A multi resolution analysis is made by introducing a parameter of dilation a . In this case, the Gabor transform may be used as a Wavelet transform:

$$TGf(x_0, y_0, \theta, a) = \frac{1}{a} \iint f(x, y)g\left(\frac{x-x_0}{a}, \frac{y-y_0}{a}, \theta\right) dx dy \quad (3)$$

This can be calculated by using discrete values for, a, x_0, y_0 and θ .

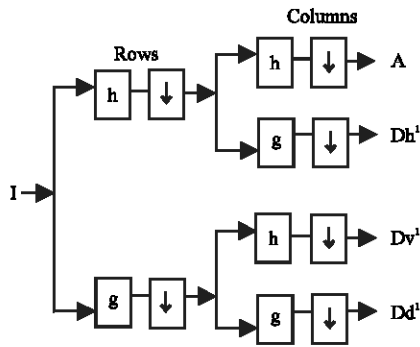


Fig. 1: Mallat Multi-Resolution Algorithm (resolution level 1). (h: low pass filter, g: high pass filter, Down arrow: dyadic undersampling operator, I: Image, A: Approximation, Dh¹, Dv¹, Dd¹: respectively Horizontal, Vertical, Diagonal details at level 1)

Separable wavelet transform: To carry out a Separable Wavelet Transform on a digital image, one successively operates on the lines, then, on the columns. The decomposition is generally carried out in a pyramidal way according to Mallat (1989) Algorithm (Fig. 1).

Feature extraction and bayes distance: Each decomposition produce sub-images. For every sub-image, the variance of the intensity of the pixels is the most significant feature (the mean is generally close to zero).

Gabor transform is achieved with 6 orientations per resolution level ($\theta = n\pi/6, n = 0..5$) and 4 resolution levels with dyadic values ($\alpha = 2^m, m = 0..3$). This produces a set of 24 features per image.

Three resolution levels are used for Separable Wavelet Transform so that we obtain 12 features.

Bayes distance is given by:

$$DB(x,i) = (x-\mu_i)^T C_i^{-1} (x-\mu_i) + \log [\det(C_i)] \quad (4)$$

Where:

x = Feature vector

μ_i = Mean vector of class i

C_i = Covariance Matrix of class i

This distance assumes that the feature vectors of class i follow a normal distribution of average μ_i and matrix covariance C_i .

FEATURE SELECTION

Gabor functions redundancy: The nonorthogonality of the Gabor functions implies that there is redundant information in the filtered images. This redundancy has a

positive impact when it increases the difference between the classes. But it is not always compatible with the Bayes distance because of the covariance matrix which can become non invertible. Thus, one cannot use all the features produced to evaluate the Bayes distance and a feature selection is necessary.

Adaptive feature selection: An Energy Based adaptive filter selection was proposed by Manjunathi and Ma (1996). The aim of the study was not to improve the performance but to reduce the image processing time, while keeping acceptable error rates. The selection scheme uses the spectral information in conjunction with the average database image properties to select a subset of filters. First, it calculates the difference between the spectrum of the input image pattern and the average spectrum. Let:

$$D(u, v) = [F_{input}(u, v) - F_{mean}(u, v)]^2 / F_{var}(u, v) \quad (5)$$

where, $F_{input}(u, v)$ is the Fourier transform of the input image pattern. $F_{mean}(u, v)$ and $F_{var}(u, v)$ are, respectively the mean and variance associated with the distribution of Fourier transforms of all image patterns in the database.

Then, each filter is evaluated based on the total difference energy within its spectral coverage:

$$C_{mn} = \sum_u \sum_v D(u, v) |G_{mn}(u, v)|^2 \quad (6)$$

where, $G_{mn}(u, v)$ is the frequency response of the filter $g_{mn}(x, y)$. The larger the value C_{mn} is, the better the performance of the filter.

Thus, the filters can be ordered based on their C_{mn} and the selection consists in using only the top filters.

Notice that this method is applied with Gabor transform and not with Separable transform.

Iterative feature selection: This method is relatively simple. We begin counting error rate using one feature (it is preferred to use one from those corresponding to the highest resolution level). Then, we add successively the other features counting each time the error rate. If a feature increases the error rate, it is removed, otherwise, it is kept.

RESULTS

The textures used are all selected from the Brodatz album (112 images of size 640×640 pixels) and will be numbered from 1 to 112. From this set, a total number of

Table 1: Error rates using Gabor transform (3 strategies) and Separable Wavelet transform (2 strategies). f stands for feature number

Transforms	First features	Adaptive selection	Iterative selection
Gabor transform	71.48% (f: 1 to 14)	68.25% (10 features)	27.29% (f: 1, 3 to 7, 20)
Separable Wavelet transform	31.25% (all features)	Adaptive strategy is not designed for mallat algorithm	16.8% (f: 1 to 6)

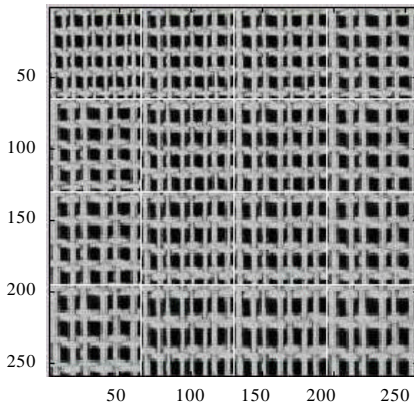


Fig. 2: Example of Sub-Images obtained by division (x and y axis, respectively horizontal and vertical position of pixels, division lines separate between sub-images)

112×16 = 1792 subimages is created by dividing each image into 16 nonoverlapping subimages of size 160×160. An example is given in Fig. 2.

A comparison between the strategies used in our study is presented in this section. The results are shown in Table 1.

In all cases shown here, we evaluate the performances using a leave-one-out method which consists in removing the image to classify from the data base and incrementing the error count if it is not classified in its own class. An image is considered to belong to a given class if this class is nearest to it in term of Bayes distance.

DISCUSSION

Manjunathi and Ma (1996) proposed an Energy Based Adaptive filter selection.

The main result of this method is CPU time reduction without affecting performance. It is a great advantage for Texture Browsing and Gabor Transform seemed to provide the most representative features. It has also been noticed that some of the features contribute to lead in a decrease in performance. All of the experiments were done using Euclidian distance. Bayes distance has an interesting property since it takes into account a covariance matrix which is a disparity characteristic of the members of a given class. That’s why it can be applied

with success in medical image classification (Mojsilovic *et al.*, 1998; Ghafoor *et al.*, 2005).

As mentioned in §2-3, with Gabor transform, we get a set of 24 features. Each class containing 16 images, when we use the leave-one-out method, we have to compute some covariance matrices with 15 vectors. This is not enough to represent all the features and these covariance matrices may not be invertible which makes Bayes distance evaluation impossible. Thus, we started using only the first features and 7 features were necessary to give the lowest error rate. Next we used the adaptive feature selection as described in § 3-2. The best results were obtained with 10 features. Then, we selected the features iteratively beginning from only one feature. Finally, we got a set of 7 features which are not all the first ones (5 correspond to the highest resolution level, 1 to a middle resolution level and one to the lowest resolution level).

The Separable Wavelet transform provides a set of 12 features. Two experiments were done. One with all the features and one by selecting iteratively each feature as for Gabor Transform.

These results show that the classification performance doesn’t rely on the number of features. Since the adaptive selection doesn’t give effective improvement, it seems that the representativeness of a feature is not sufficient to discriminate between the classes. Nevertheless, the iterative selection reveals the necessity of choosing correctly the features and this make us notice that it’s either a combination of some features which best characterise a class. We also see that the iterative selection make the results provided by Gabor transform become close to those obtained with Separable Wavelet transform.

CONCLUSION

By this study, we made a contribution to improve the results of Bayes Classification. The most important remark is that the feature selection is a very determinant factor for the performance. The results (error rates) obtained here are valid for our data base. Because of the computing time which takes more than one hour with this set of images, the iterative selection method described is usable with relatively small data bases. For large data bases, the computing time may be too prohibitive and it would be of interest to develop methods which find the best

combination of features without calculating the error rates. We conclude that improvement of results doesn't rely on the number of features or only on difference energy as used in adaptive feature selection but on a combination of features which has to be optimized.

REFERENCES

- Abdullah, S., S.N. Sahadan, M.Z. Nuawi and Z.M. Nopiah, 2008. Fatigue road signal denoising process using the 4th order of daubechies wavelet transforms. *J. Applied Sci.*, 8: 2496-2509.
- Bouman, C.A. and B. Liu, 1991. Multiple resolution segmentation of textured images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13: 99-113.
- Bovik, A., M. Clark and W.S. Geisler, 1990. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Pattern Anal. Machine Intel.*, 12: 55-73.
- Chellappa, R., 1985. Two-dimensional discrete Gaussian Markov random field models for image processing. *Pattern Recognition*, 2: 79-112.
- Cross, G.R. and A.K. Jain, 1983. Markov random field texture models *IEEE Trans. Pattern Anal. Machine Intell.*, 5: 25-39.
- Derin, H., 1986. Segmentation of textured images using Gibbs random fields. *Comput. Vision Graphics Image Processing*, 35: 72-98.
- El-Ramsisi, A.M. and H.A. Khalil, 2007. Diagnosis system based on wavelet transform, fractal dimension and neural network. *J. Applied Sci.*, 7: 3971-3976.
- Ghafoor, A., F. Muhammad and I.A. Arshad, 2005. Bayesian regression with prior non-sample information on mash yield. *J. Applied Sci.*, 5: 187-191.
- Imine, R. and M.F. Belbachir, 2006. Analyze images, the coefficients of gabor: A simple method for calculation of the coefficients of gabor. *J. Applied Sci.*, 6: 94-97.
- Lee, T.S., 1996. Image representation using 2D gabor wavelets. *IEEE Trans. Pattern Anal. Machine Intell.*, 18: 959-971.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. *Trans. Pattern Anal. Mach. Intell.*, 11: 674-693.
- Manjunathi, B.S. and W.Y. Ma, 1996. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis Machine Intell.*, 31: 837-842.
- Mojsilovic, A., M. Popovic, S. Markovic and M. Krstic, 1998. Characterization of visually similar diffuse diseases from b-scan liver images using nonseparable wavelet transform. *IEEE Trans. Med. Imaging*, 17: 541-549.
- Unser, M. and M. Eden, 1989. Multiresolution feature extraction and selection for texture segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11: 717-728.