International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:4, No:10, 2010

A Constrained Clustering Algorithm for the Classification of Industrial Ores

Luciano Nieddu, Giuseppe Manfredi

Abstract—In this paper a Pattern Recognition algorithm based on a constrained version of the k-means clustering algorithm will be presented. The proposed algorithm is a non parametric supervised statistical pattern recognition algorithm, i.e. it works under very mild assumptions on the dataset. The performance of the algorithm will be tested, togheter with a feature extraction technique that captures the information on the closed two-dimensional contour of an image, on images of industrial mineral ores.

Keywords—K-means, Industrial ores classification, Invariant Features, Supervised Classification.

I. Introduction

AUTOMATIC classification techniques have become a very important and effective tool in many real industrial and medical problems, where objects need to be classified according to some features and an automated technique is necessary to determine in a fast, reliable and efficient way, to which class an object belongs [2],[3], [1].

The aim of this paper is to present a supervised statistical pattern recognition technique [4], [5] suitable to discriminate between objects with respect to their 2D contour.

The proposed technique will be applied on industrial ores but is general and can be applied on various 2D objects or on the two dimensional projection of 3D objects.

The main assumption of this paper is that the class of the object can be determined via the pattern in the 2D contour of the object, i.e. only the information in the 2D shape is necessary to classify the object. This assumption is valid in a variety of real classification problems, such as, e.g. geometrical shapes, handwritten characters, OCR etc. [6], [7], [2], and is particulary valid in the case of mineral ores where the contour of each piece of mineral should be influenced by the chemical structure of the particle itself.

In Section II the proposed algorithm will be presented. In Section III the dataset used in the experiments and the feature extraction technique will be discussed while in Section IV the results of the application of the proposed technique will be presented. Then in Section V some conclusions will be drawn.

II. The Algorithm

The algorithm presented in this paper is a supervised classification algorithm [5], i.e. a data set of elements with known classes is supposed to be available. As any supervised learning technique it is composed of two phases:

1) a *training phase*: a training set of elements of known classes is used to fine-tune the algorithm for the particular problem at hand.
2) a *classification phase*: once trained the algorithm is used to classify elements of unknown classes. These elements are usually referred to as *query points*

The performance of the algorithm is assessed via cross-validation [5], [8], i.e. the dataset is split into two independent subsets, both composed of elements of known classes, one is used to train the algorithm, and the other to assess the performance of the algorithm, comparing the real class of each elements to the class assigned by the algorithm.

Given a data set of n pattern vectors in $\Re^p$, assume a partition defined on the dataset, i.e. each pattern vector is assigned to one and only one of k known classes. Let assume a Euclidean norm defined on the dataset and let ψ be a function from $\Re^p$ onto the set $\mathcal{C}= \{1, 2, \ldots, k\}$ which maps each pattern vector $\mathbf{x}_j$, $j = 1, \ldots n$ into the class $c \in \mathcal{C}$ that it belongs to.

The algorithm presented begins computing the barycenter of each class, yielding an initial set of k barycenters. Then the Euclidean distance of each pattern vector from each barycenter is computed. If each pattern vector is closer to the barycenter of its class the algorithm stops, otherwise there will be a non empty set $\mathcal{M}$ of pattern vectors which belong to a class and are closer to a barycenter of a different class. In $\mathcal{M}$ select the pattern vector $\mathbf{x}_w$ that is farthest from the barycenter of its class. This pattern vector will be used as a seed for a new barycenter for class $\psi(\mathbf{x}_w)$.

A k-means algorithm [9] will then be performed for all the pattern vectors in class $\psi(\mathbf{x}_w)$ using, as starting points, the set of barycenters for class $\psi(\mathbf{x}_w)$ and the vector $\mathbf{x}_w$. Once the k-means has been performed the set of barycenters will be composed of $k + 1$ elements.

The barycenters at the new iterations need not be computed for all classes, but only for class $\psi(\mathbf{x}_w)$ since the barycenters for the other classes have remained unchanged. In the following step the distance of each pattern vector from all the barycenters is computed anew, and so is the set $\mathcal{M}$ (see figure 1).

If $\mathcal{M}$ is not empty then the pattern vector in $\mathcal{M}$ which is farthest from a barycenter of its own class is once again selected to serve as a seed for a new barycenter. This procedure iterates until the set $\mathcal{M}$ is empty. The convergence of the algorithm in a finite number of steps has been proved in various ways (see [1],[10]).

Upon convergence, the algorithm yields a set of barycenters which, in the worst case, are in a number equal to the number

L. Nieddu is Associate Professor of Statistics with the Faculty of Economics, LUSPIO University, Rome, 00147 Italy e-mail: l.nieddu@luspio.it.
G. Manfredi is Full Professor of Statistics with the Faculty of Political Sciences, LUSPIO University, Rome, 00147 Italy.

Step1 **Let**
- $\mathbf{x}_j$, $j = 1, \ldots, n$ be the pattern vectors in the training set
- $\mathbf{B}_0$ be the set of k initial barycenters $\mathbf{b}_i$, $i = 1, \ldots, k$

Step2 **Compute** the distances of each x_j from all the $b_i \in B_t$
Let $\mathcal{M}$ be the set of $\mathbf{x}_w$ that are closer to a barycenter of a class different from their own.
$t \leftarrow 0$

Step3 **while** $M \neq \emptyset$
- **Let** $\mathbf{x}_s$, $s \in \mathcal{M}$ be the vector with the greatest distance from its own barycenter.
- $c \leftarrow \psi(\mathbf{x}_s)$
- **Let** $B_{t+1} \leftarrow B_t \cup \mathbf{x}_s$
- for all the elements of class c perform a k-means routine using as starting points the barycenters of B_{t+1} that belong to class c
- $t \leftarrow t + 1$
- **Compute** the distances of each $\mathbf{x}_j$ from all the $\mathbf{b}_i \in B_t$
- **Let** $\mathcal{M}$ be the set of $\mathbf{x}_w$ that are closer to a barycenter of a class different from their own.

end

Fig. 1. : The proposed algorithm in meta-language

of elements in the dataset and which has a lower bound in the number of classes.

The aim of this algorithm is to find subclasses in the dataset which can be used to classify new vectors of unknown class. It is worth noticing that if the partition defined on the dataset is consistent with the features considered, i.e. if the pattern vectors are linearly separable, then the algorithm generates a number of barycenters equal to the number of classes. On the other hand, if the dataset is not linearly separable, then the algorithm continues splitting the classes until the subclasses obtained are linearly separable. It is obvious that it can continue splitting until all the subclasses are composed of only one vector (singleton). It will not converge only if two vectors in the dataset belong to different classes and are represented by the same pattern vector [10], [1]. This problem can be easily overcome increasing the dimension of the vector space.

Once the algorithm has converged the sets of barycenters can be used to classify new query points assigning the new element to the class of the barycenter it is closest to. It is worth noticing that if elements from the training set are used as query points than the algorithm always classify them correctly because, once converged, all pattern vectors in the training set are closer to a centroid of their own class.

One interesting consequence of this technique is that the *apparent recognition rate* is 100%. This is a direct consequence of the convergence criterion, i.e. upon convergence all the elements in the training set are closer to a barycenter of their own class and therefore if the elements of the training set are classified using the barycenters generated during the training phase, then all the element are classified correctly. Therefore, if the training set is representative of the the population under study then the algorithm should be able to perform fairly well on unknown query points. On the other hand if the performance of the algorithm on a particular problem is not satisfactory then two cases are possible:

- the pattern vector does not contain enough information to discriminate between objects
- the training set is not representative of the population under study.

In the first case other features need to be determined from the objects, in the latter the size of the training set should be increased in order to obtain a representative training set.

Therefore if the training set is large enough to represent all the possible prototypes of the objects under consideration and if the features considered for each object are sufficient enough to assure coherence of the dataset, then the algorithm should be able to correctly classify any new query points.

III. The Pattern Vector

A. Data Collection

The aim of this study is to verify if the contour of a 2D picture of a mineral can be used to determine which class the mineral belongs to. The external shape of the mineral is influenced both by external factors and by its internal structure, i.e. the arrangement of atoms that determine the internal lattice of the mineral. To reduce the effect of external factors on the shape of the mineral and to increase the effects of the internal structure, the mineral have been crushed into small pieces, assuming that the way a mineral breaks up into pieces is influenced by the internal lattice. Three types of mineral have been considered, namely: Chalcopyrite, Galena, Hematite.

Chalcopyrite is a copper iron sulfide. It is the most commonly encountered copper mineral and is the most important ore of copper. Chalcopyrite deposits are found in hydrothermal veins, void fillings and replacements in limestones, contact metamorphic deposits and magmatic separations. It has the chemical composition $CuFeS_2$ and crystallizes in the tetragonal system.

Galena is a lead sulfide mineral commonly found in hydrothermal veins, or as fracture fillings, cavity fillings and replacements in limestone. Galena is the most important ore of lead and is often mined for its silver content (silver substitutes for lead within the galena structure). Its perfect cleavage, silver color and very high specific gravity make it very easy to identify. Silver is often produced as a by-product. It has chemical composition PbS, is one of the most abundant and widely distributed sulfide minerals and crystallizes in the cubic crystal system often showing octahedral forms.

Hematite is a relatively hard oxide mineral, ferric oxide (chemical composition Fe_2O_3), that constitutes the most important iron ore because of its high iron content and its abundance. Hematite crystallizes in the rhombohedral system.

Once all the speciments of the three ores have been collected and crushed, a picture of 512x512 pixels of each piece has been taken and then converted into a 256 gray scale image. Only the contour of the image is of interest in this study, therefore the image has been segmented into a black and white image via thresholding. Of such an image only the contour has been retained.

In Table I the distribution of the 2928 images by type of mineral has been displayed. The average radius and average perimeter (in number of pixels) for the images of each class of mineral have also been displayed.

The average size of Hematite is significantly ($\alpha = 0.05$) different from that of Galena and Chalcopyrite, i.e. the dimension of the single piece of mineral could be considered

TABLE I
NUMBER OF IMAGES FOR EACH MINERAL TOGETHER WITH AVERAGE RADIUS AND PERIMETER (IN PIXELS)

Mineral	Images	Av. radius (pxs)	Av. perim. (pxs)
Galena	942	60.91	437.79
Chalcopyrite	968	69.83	516.94
Hematite	1018	82.80	637.72

as a discriminating feature in determining the class the piece belongs to. Unfortunately the dimension of each single piece of mineral is not a characteristic feature of the mineral but only a byproduct of the crushing process and should not be taken into account during the classification process. Therefore the information that will be extracted from each image should be invariant to variations in size, to rotations of the image and to shifting of location. In the next subsection the feature extraction technique, based on the Discrete Fourier Transform [11], that has been used to obtain invariant features will be described.

B. Feature Extraction

A fundamental part of every Pattern Recognition algorithm consists in extracting features from the pattern vector, both to get rid of non important features and to reduce the dimension of the vectors the algorithm has to work with [12]. In the following the feature extraction technique used to map each image into a pattern vector and then into a feature vector will be described. The final mapping should result in a feature vector which should retain as much as possible of the information in the contour of the image and get rid of spurious information. Assuming that only the shape of the contour is relevant to discriminate between minerals, the feature vector so obtained should be invariant to changes in position, size and rotation.

Let's consider the 2D closed contour of each image as obtained in the previous section. Set a polar coordinate system with pole at the barycenter of the picture. Considering a constant increment equal to $2\pi/n$ for the angle of the polar coordinate system, n points are sampled counterclockwise from the contour.

Therefore the object is represented by a set of n radiuses (pattern vector) $\rho(i)$, $i = 0, \ldots, n-1$. On this sequence of n real numbers the Discrete Fourier Transform [11] has been applied, obtaining another sequence $a(k)$, $k = 0, \ldots, n-1$, of n complex numbers describing the structure of the frequencies that make up the contour. To speed up the process the Fast Fourier Transform (FFT) [13] has been used requiring only $O(n \log_2 n)$ operations.

The feature vector obtained according to this procedure is invariant to change in position because the origin of the coordinate system located at the barycenter of the object. To assure invariance to rotation, considering that a rotation in the object influences only the phase of each element of the FFT [14], only the modulus $(\Im^2[a(k)]+\Re^2[a(k)])$, $k = 0, \ldots, n-1$ of each complex element of the FFT will be considered, where $\Im$ and $\Re$ represent the immaginary and the real part of a complex number respectively. Invariance to changes in dimension will be attained considering the equivalent radius r_{eq} for each object [2], which is the radius of circle with the same area of the considered object. Invariance will be attained dividing by the equivalent radius. Therefore the new feature vector will be:

$$\left[\frac{a(0)}{r_{eq}}; \frac{\Im^2[a(j)] + \Re^2[a(j)]}{r_{eq}}; \ldots; \frac{\Im^2\left[a(\frac{n}{2})\right] + \Re^2\left[a(\frac{n}{2})\right]}{r_{eq}}\right].$$

The information in this feature vector is then transformed using a Karhunen-Loéve [19] expansion in order to reduce the size of the vector, retaining only the information contained in the first few components of the expansion. This expansion is particularly useful when the dataset is affected by noise, which usually can be filtered out considering only the information in the first few components.

IV. EXPERIMENTAL RESULTS

In this section the results of the application of the proposed technique will be presented. The performance of the proposed technique will be compared with the results of Fisher Discriminant Analysis (FDA) [4], [15].

The performance of each technique has been assessed in a cross-validation scheme [15], [1], i.e. the dataset has been split into two subsets: 2567 images have been used for training and 320 images randomly selected from the 2887 have been used for testing. The same test dataset has been used throughout all the experiments to make results immediately comparable.

The FFT has been applied on the pattern vector and then data have been made invariant to rotation and size as described in the previous Section. The Karhunen-Loéve (KL) transform has then been applied. For the experiments only the first 3, 5, 7, 9 e 11 elements of the KL transform have been retained.

Both the proposed algorithm and FDA have been trained on the training set of 2567 images and then tested on the set of 320 images. It is worth noticing that, should the algorithm be tested on the 2567 images in the training set, it would obtain a 100% correct recognition, whereas the performance of FDA on the training set is usually better or at least comparable to the one obtained on the test set, but in general does not reach 100% correct recognition.

The performance has been analyzed considering the confusion matrix

	PREDICTED			
TRUE	Galena	Chalcopyrite	Hematite	Tot
Galena	n_{11}	n_{12}	n_{13}	$n_{1.}$
Chalcopyrite	n_{21}	n_{22}	n_{23}	$n_{2.}$
Hematite	n_{31}	n_{32}	n_{33}	$n_{3.}$
Tot	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

where the quantity n_{ij}, $i, j = 1, 2, 3$ represents the number of element of class i in the test set that have been classified as belonging to class j.

The structure of a confusion matrix can be studied using the cross-product ratios or odd-ratios [16],[17] defined as

$$q_{(ii'),(jj')} = \frac{m_{ij} m_{i'j'}}{m_{ij'} m_{i'j}} \quad i, i', j, j' = 1, 2, 3$$

TABLE II
PERFORMANCE OF THE PROPOSED ALGORITHM ON INVARIANT FEATURES

Feat. Vect.	Recg. Rate	$\hat{\tau}_{(13)(13)}$	$\hat{\tau}_{(23)(13)}$	$\hat{\tau}_{(13)(23)}$	$\hat{\tau}_{(23)(23)}$
3	37.5	-0.26(0.31)	-0.05(0.33)	-0.62(0.42)	0.76(0.37)
5	37.8	-0.24(0.31)	-0.12(0.33)	-0.69(0.43)	0.76(0.37)
7	39.1	-0.16(0.31)	-0.29(0.33)	-0.39(0.41)	0.77(0.36)
9	38.1	-0.30(0.32)	-0.25(0.33)	-0.48(0.41)	0.78(0.37)
11	41.6	0.06(0.31)	-0.08(0.33)	-0.36(0.42)	1.02(0.36)

TABLE III
PERFORMANCE OF FDA ON INVARIANT FEATURES

Feat. Vect.	Recg. Rate	$\hat{\tau}_{(13)(13)}$	$\hat{\tau}_{(23)(13)}$	$\hat{\tau}_{(13)(23)}$	$\hat{\tau}_{(23)(23)}$
32	39.7	-0.09(0.31)	0.20(0.34)	-0.45(0.42)	1.09(0.38)

where m_{ij} is the expected value of element of class i that is classified in class j. The quantity

$$\tau_{(ii'),(jj')} = \ln q_{(ii'),(jj')}$$

is known as log-cross-product ratio and can be estimated [16] using the quantity

$$\hat{\tau}_{(ii'),(jj')} = \ln n_{ij} - \ln n_{ij'} - \ln n_{i'j} + \ln n_{i'j'}$$

which has an asymptotically normal distribution with mean $\tau_{(ii'),(jj')}$ and standard deviation equal to $\left(\frac{1}{m_{ij}} + \frac{1}{m_{i'j}} + \frac{1}{m_{ij'}} + \frac{1}{m_{i'j'}}\right)^{1/2}$ that can be estimated by $\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}} + \frac{1}{n_{ij'}} + \frac{1}{n_{i'j'}}\right)^{1/2}$.

In a $r \times s$ contingency table, as suggested by Agresti [18], only $(r-1)(s-1)$ cross-products out of the total $\binom{r}{2}\binom{s}{2}$ cross-products are independent. Therefore in a 3×3 table only 4 cross products are necessary.

On each cross-product-ratio the null hypothesis $H_0 : \tau = 0$ will be tested at an $\alpha = 5\%$ significance level, i.e. the null hypothesis of independence will be statistically tested.

In Table II the performance of the proposed techinque on invariant feature vectors of 3, 5, 7, 9 and 11 elements, together with the estimated log-cross-product ratios and standard deviations (in brackets), has been reported.

All the log-cross-products, except for $\tau_{(23)(23)}$ are not significantly different from zero ($\alpha = 5\%$), meaning that, except for the way that chalcopyrite and hematite have been classified as chalcopyrite and hematite, there is not significant difference from a random allocation for the other cell of the confusion matrix. Anyhow, the hypothesis of random allocation for all the minerals must be rejected because at least one log-cross-product is significantly different from zero, although the level of correct classification is close to the random recognition level equal to 33.3%.

In Table III the results of FDA applied on the invariant feature vector have been reported. FDA has been applied on the feature vector of 32 element without applying KL dimensional reduction because FDA itself is based on an optimal dimensionality reduction technique [15].

TABLE IV
PERFORMANCE OF THE PROPOSED ALGORITHM ON NON INVARIANT DATA (COMPLEX DATA)

Feat. Vect.	Recg. Rate	$\hat{\tau}_{(13)(13)}$	$\hat{\tau}_{(23)(13)}$	$\hat{\tau}_{(13)(23)}$	$\hat{\tau}_{(23)(23)}$
3	52.2	3.13(0.44)	1.36(0.39)	1.92(0.40)	1.06(0.31)
5	53.4	2.90(0.40)	1.33(0.38)	1.02(0.40)	0.87(0.31)
7	54.4	2.40(0.37)	1.36(0.38)	0.50(0.39)	1.21(0.31)
9	53.4	2.95(0.40)	1.45(0.39)	0.50(0.41)	0.70(0.31)
11	53.6	2.91(0.41)	1.14(0.39)	0.85(0.39)	0.78(0.30)

TABLE V
PERFORMANCE OF FDA ON NON INVARIANT FEATURES (COMPLEX DATA)

Feat. Vect.	Recg. Rate	$\hat{\tau}_{(13)(13)}$	$\hat{\tau}_{(23)(13)}$	$\hat{\tau}_{(13)(23)}$	$\hat{\tau}_{(23)(23)}$
32	56.2	3.09(0.40)	1.44(0.36)	1.26(0.43)	1.19(0.32)

The results of FDA confirm what has already been obtained with the proposed technique, i.e. the minerals have been classified with a performance significantly different from the random recognition rate and the only log-cross-product significantly different from zero is again $\tau_{(23)(23)}$. The performances of the two algorithms are comparable and, being the test set always the same in all the experiments, the performance of the proposed technique is significantly better than FDA when feature vectors of 11 elements are used.

In Table IV the result of the proposed algorithm on the FFT complex coefficients have been reported, i.e. the data are now invariant only to position, but not to rotation and size. Almost all cross-products are now significantly different from zero except for $\tau_{(13)(23)}$ with feature vectors of 7 and 9 elements. This implies that galena and hematite are equally likely to be classified as chalcopyrite and hematite. Contrary to what happened on invariant features, $\tau_{(13)(13)}$ is now always significanlty different from zero.

In Table V the analogous results using FDA have been reported. All cross-products are significantly different from zero, indicating a non random allocation of the vectors.

The performance of the proposed algorithm on non invariant features is worse than that obtained with FDA. This could be ascribed to the fact that maybe the information retained with feature vectors of 11 elements is not sufficient when dealing with non invariant features. Therefore the experiments have been extended considering feature vectors of 15, 20, 25, 30 and 32 elements and the performance of the proposed algorithm, together with the log-cross-products, have been displayed in Table VI

With feature vectors of 32 elements the proposed technique performs better than FDA. The performance obtained with non invariant features is always better than the one obtained with invariant features. This could be ascribed to the fact that some features, such as dimension of the various pieces, can influence the recognition process, i.e. once taken into consideration could improve the recognition rate. It has been already pointed out (see Table I) that the pieces of hematite are significantly larger than galena and chalcopyrite, but unfortunately that

TABLE VI
PERFORMANCE OF THE PROPOSED ALGORITHM ON NON INVARIANT DATA AND EXTENDED FEATURE VECTOR DIMENSION (COMPLEX DATA)

Feat. Vect.	Recog. Rate	$\hat{\tau}_{(13)(13)}$	$\hat{\tau}_{(23)(13)}$	$\hat{\tau}_{(13)(23)}$	$\hat{\tau}_{(23)(23)}$
15	57.2	3.36(0.44)	1.66(0.43)	0.80(0.41)	1.04(0.30)
20	53.4	3.05(0.42)	1.38(0.40)	0.94(0.40)	0.78(0.30)
25	54.1	3.16(0.43)	1.38(0.41)	0.83(0.39)	0.73(0.30)
30	56.6	3.41(0.46)	1.43(0.41)	1.65(0.41)	1.19(0.31)
32	58.1	3.43(0.44)	1.26(0.41)	1.15(0.42)	1.05(0.30)

should be considered only a confounding factor because the size is not an intrinsic characteristic of the class the mineral belongs to, but only of the crushing process and therefore should not be considered in the classification process.

Nonetheless the recognition rate on invariant features is significantly different from the random recognition rate, implying that there is a "signature" in the profile of the image that can be used to classify each element into one of 3 classes.

V. CONCLUSIONS

In this paper a non parametric statistical patter recognition algorithm has been presented together with a technique to obtain invariant features from the 2D contour of mineral images. The proposed algorithm, when compared with FDA, more than holds its own. The experiments have been carried out on invariant and non invariant features to ascertain the effect of size on the classification process. On this regard: $\tau_{(13)(13)}$ on non invariant features is always different from zero, implying that it is possible to discriminate between galena and hematite. In the experiments on non invariant features, on the other hand, $\tau_{(13)(13)}$ is always non significantly different from zero, indicating that, using invariant features, such a discrimination is not obtainable. Using invariant features galena and hematite tend to be misclassified and therefore have a negative effect on the performance of the algorithm. The use of non invariant features reduces this effect but this can be ascribed to the use of the size of each image in the classification and, once again, the size of the pieces is a non characteristic element that can be used to classify this type of minerals. As future agenda the use of other features (e.g. color), will be considered in the recognition process to increase the separation of the classes and to obtain a more coherent dataset.

ACKNOWLEDGMENT

The authors would like to thank prof. Prof. P. Massacci and Prof. G. Bonifazi from the Dept. of Chemical Engineering, University of Rome "Sapienza" for their help and suggestions and for providing the images.

REFERENCES

[1] L. Nieddu and G. Patrizi, Formal methods in pattern recognition: a review. *European Journal of Operational Research*. v120. 459-495, 2000.
[2] J. Yokono and T. Poggio, Oriented filters for objects recognition: an empirical study, *Proceedings of the IEEE Conference on Face and Gesture Recognition* (FG2004). Seoul, Korea, 2004.
[3] V. N. Vapnik. *The nature of statistical learning theory* Springer-Verlag, Berlin, 1995.
[4] R. O. Duda and P. E. Hart. *Pattern Classification and scene analysis* Wiley, New York, 1973.
[5] S. Watanabe. *Pattern Recognition: human and mechanical* Wiley, New York, 1985.
[6] A. Khotanzad and Y. H. Hong. Rotation invariant image recognition using features selected via a systematic method. *Pattern Recognition*, 23(10) pp. 1089-1101, 1990.
[7] D. G. Lowe. Object recongition from local scale invariant fetures. In *ICCV*, pp. 1150-1157, 1999.
[8] K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, 1990.
[9] D. Gordon. *Classification.* Chapman & Hall Ltd., London, New York, 1999
[10] L. Nieddu and G. Patrizi. Optimization and algebraic techniques for image analysis. In M. Lassonde, editor, *Approximation, optimization and Mathematical Economics*, pp. 235-242. Physica-Verlag, 2001
[11] J. W. Cooley, P. A. Lewis and P. D. Welch. The finite Fourier transform. *IEEE Trans. Audio electroacoustics*, 17(2) pp. 77-85, 1969.
[12] R. Bellman. *Adaptive control Processes: a guided tour.* Princeton University Press, 1961.
[13] J. W. Cooley, P. A. Lewis and P. D. Welch. The fast Fourier transform and its application to time series analysis. In *Statistical Methods for digital computers*, Wiley, New York, 1977.
[14] E. O. Brigham. *the fast Fourier transform and its applications.* Englewood Cliffs, N.J.: Prentice hall, 1988.
[15] G. J. McLachlan. *Discriminant Analysis ans Statistical Pattern Recognition.* John wiley & Sons, 1992.
[16] S. J. Haberman. *Analysis of qualitative data. Volume 1: introductory topics*, New York, Academic Press Inc, 1978.
[17] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland. *Discrete myultivariate analysis: theory and practice*, Cambridge, Massachusset, The MIT press, 1975.
[18] A. Agresti. *Analysis of ordinal data* Wiley, New York, 1984.
[19] S. watnabe. Karhunen-Loéve expansion and factor analysis. Theoretical remarks and applications. *Transactions of the fourth Prague conference on Information Theory, Statistical Decision functions, Random Processes*, pp. 635-660, 1965