Bayesian Classification for Inspection of Industrial Products

Petia Radeva, Marco Bressan, A. Tovar, and Jordi Vitrià

Computer Vision Center, Depart. Informatica, Universitat Autònoma de Barcelona,
Ed. O Campus UAB, 08190 Cerdanyola (Barcelona), Spain
`{petia,marco,atobar,jordi}@cvc.uab.es`

Abstract. In this paper, a real time application for visual inspection and classification of cork stoppers is presented. Each cork stopper is represented by a high dimensional set of characteristics corresponding to relevant visual features. We have applied a set of non-parametric and parametric methods in order to compare and evaluate their performance for this real problem. The best results have been achieved using Bayesian classification through probabilistic modeling in a high dimensional space. In this context, it is well known that high dimensionality does not allow precision in the density estimation. We propose a Class-Conditional Independent Component Analysis (CC-ICA) representation of the data that even in low dimensions, performs comparably to standard classification techniques. The method has achieved a success of 98% of correct classification. Our prototype is able to inspect the cork stoppers and classify in 5 quality groups with a speed of 3 objects per second.

1 Introduction

Cork inspection is the least automated task in the production cycle of the cork stopper. Due to the inspection difficulty of the natural cork material and the high production rates even the most experienced quality inspection operators frequently make mistakes. In addition, it is increasingly difficult to find labor willing and able to do a job that is at the same time both skilled and highly repetitive. On the other hand, human inspection leads to a lack of objectivity and uniform rules applied by different people at different time. As a result, there is a urgent need to modernize the cork industry in this direction. In this paper, we consider a real industrial computer vision application of classification of natural (cork) products. Cork products in the manufacture are inspected for different faults like small holes due to insect attacks, channels due to imprecise stopper cutting, stopper breaking, cracks and woody surfaces (see fig. 1). Although it does not seem difficult for human beings to detect different faults in the cork material, it turns out difficult to precisely formulate the features of the cork faults due to the porosity of the natural material. It is difficult even for the cork quality experts to exactly define all cork features that they take into account in the process of stopper inspection, the feature values and ranges in order to define whether there is a fault in the cork stopper or the stopper is of poor quality.

M.T. Escrig Monferrer and F. Toledo Lobo (Eds.): CCIA 2002, LNAI 2504, pp. 399–407, 2002.

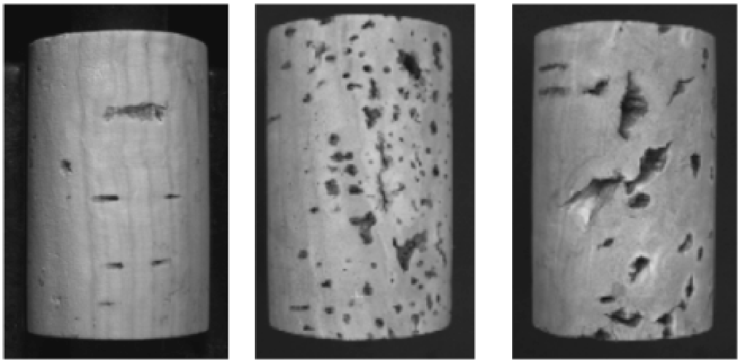

Fig. 1. Cork stopper without fault, woody stopper and a stopper with a crack.

There have been different attempts to develop vision cork inspection systems in the manufacture where the people working in the manufacture should define the values and ranges of the image features and elaborate the decision rules in the process of the stopper inspection. Given that people in the manufacture work with rather qualitative than quantitative information to classify the quality of a stopper, managing such vision cork inspection systems represent a tedious and time-consuming task. The problem of the classification of the cork product in different (in this case, five) quality groups additionally difficult the problem. This fact prevents cork stopper industry from defining and assuring the quality of the products in front of the providers. In order to cope with the problem of subjectivity in the process of cork inspection and quality classification, we study different techniques from the fields of Computer Vision and Pattern Recognition. We propose to apply statistic algorithms in order to analyze a set of 43 different features of the cork (e.g. number of cork holes, average stopper gray level, average holes gray level, holes gray level deviation, length of largest cork hole, etc), that are considered by the operators during the cork analysis.

The problem of high dimensional data classification is presented in section 2, where our approach is exposed. Section 3 describes the visual features used on the characterization of cork stoppers. Section 4 introduces our approach to high dimensional data classification: Class-Conditional Independent Component Analysis. Results and comparisons to other methods are presented in section 5. Finally, in section 6 we expose our conclusions.

2 Classification of High Dimensional Data

High dimensional data appears in many pattern recognition problems such as remote sensing, appearance-based object recognition, text categorization, etc. A stochastic approach for the classification of high dimensional data is always a delicate issue. For linear or quadratic classifiers the number of training samples depends linearly or quadratically on the data dimensionality subset of features, an exhaustive sequential feature selection procedure is required, so the size of the problem grows combinatorially on the dimension. Furthermore, the training sample size needs to increase exponentially in order to effectively estimate the multivariate densities needed to perform nonparametric classification. To avoid the problem of dimensionality, the most common approach is the implementation of feature extraction or dimensionality reduction algorithms. Principal Component Analysis (PCA) [5] is widely used due to its noise reduction properties. PCA treats the data as if they belong to a single distribution, so it has nothing to do with discriminative features optimal for classification. Linear Discriminant Analysis (LDA) [1] can be used to derive a discriminative transformation that is optimal for certain cases. LDA makes use of only second-order statistical information and this causes that if the difference in the class mean vectors is small, the features chosen will not be reliable. More importantly, LDA can not produce more features than the number of classes involved. Feature Subset Selection [6] is yet another perspective on feature extraction. This problem considers a subset of all linear combinations of the original feature set, according to a certain criterion. In order to produce an optimal subset of features, an exhaustive sequential feature selection procedure is required, so the size of the problem grows combinatorially on the dimension.

The approach proposed on this paper does not seek dimensionality reduction followed by the implementation of parametric or nonparametric techniques for density estimation. Instead, it focuses on the higher level statistical properties of the data, which is transformed in such a way that density estimation in the transformed space is simplified and more accurate. For this purpose, we consider an Independent Component Analysis (ICA) [2] representation for each class. This representation projects our data into a space where the components have maximized statistical independence, and in many real problems, sparse distributions. Independence turns an M-dimensional density estimation into M one-dimensional estimations.

3 Cork Stopper Feature Extraction

Our objective is to construct robust algorithms to classify cork stoppers in 5 quality groups (see fig. 2). To this purpose the operators have provided training examples. In order to classify the cork stoppers we extract 43 image features. A blob analysis is done and blob features are considered as follows: stopper area, number of blobs, average blob area, average blob elongation, average blob grey-level, average compactness, average roughness, features of the blob with largest area (area, length, perimeter, convex perimeter, compactness, roughness, elongation, length, width, average blob gray-level, position with respect to the center of the stopper), features of the longest blob, etc.

4 Cork Stopper Feature Classification

The classification problem of cork material is stated as follows: given a set of cork stoppers, each one represented by its image features $H = (h_1, ..., h_L)$, labeled by an operator (the learning set) and a unlabeled set of feature representations

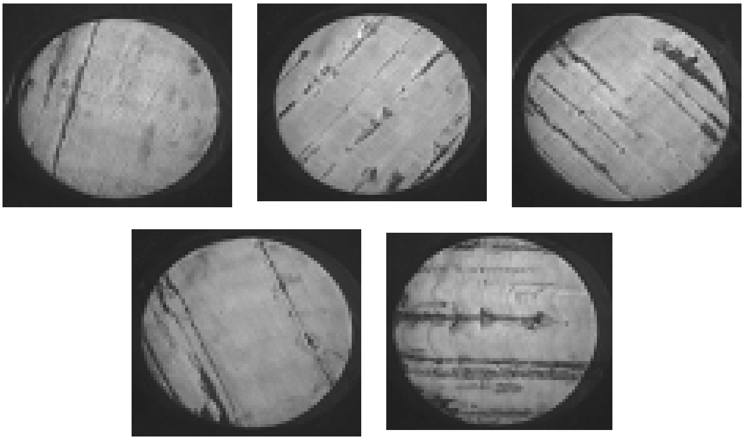

Fig. 2. Cork stoppers of 5 quality groups ordered from best to worst quality (from left to right and from top to down).

from the same group of stoppers (the test set), identify each stopper in the test image set. Under a Bayesian framework, we wish to assign a sample object H_t to a particular class using the probability of misclassification as an error measure. It can be seen that the solution to this problem is to assign H_t to the class that maximizes the *posterior probability*. This is called the Maximum a Posteriori or MAP solution. Using the Bayes rule we can formulate the posterior probability in terms of quantities which are easier to estimate, and the MAP solution takes the form:

$$C_{MAP} = \arg_{k=1,\dots,K} \max\{P(H_t/C_k)P(C_k)\}$$

where $P(C_k)$ is usually called the *prior* probability, $P(H_t/C_k)$ is referred to as the *class-conditional probability* or, when seen as a function of the parameters, the *likelihood*. In practice, the class-conditional probabilities can be modeled parametrically or non-parametrically from our training set. The priors, as their name indicate, are estimated from our prior knowledge of the problem, and if such knowledge is not available, equiprobable priors are usually assumed.

In our problem, the test object H_t is represented by its feature vector $h_1, ...h_L$. If we additionally assume conditional independence in the occurrence of a particular value for the feature vector, the MAP solution takes a new form commonly known as the naive Bayes rule,

$$C_{Naive} = \arg_{k=1,\dots K} \max \prod_{l=1}^{L} P(h_l/C_k)$$

We still have to model the class-conditional probabilities for the feature vectors. A first approach can be done using the Gaussian Kernel Estimator. The likelihood $P(h_l/C_k)$ is estimated by summing the contribution from a mixture of N Gaussians:

$$P(h_l/C_k) = \sum_{n=1}^{N} \omega_n G(h_l - \mu_n; \Sigma_n/C_k)$$

and a Kernel method is used for parameter selection. (the Kernel method positions a Gaussian on each sample of the distribution).

The precision in the estimation of the class-conditional probabilities $P(H_t/C_k)$ is decisive on the performance of the classifier. This precision is not easy to attain, specially in the case of high dimensional data. In the next subsection we introduce an alternative representation for the data and show how this representation both simplifies and improves the density estimation.

4.1 Class-Conditional ICA for Bayesian Estimation

We choose to represent the data from each class using the transform provided by Independent Component Análisis (ICA) [2]. This linear transform represents our data in a space where the statistical dependence between the components is minimized. After introducing the ICA model, we show how the assumption of independence simplifies density estimation for high dimensional data and analyze the consequences in Bayesian Decision.

The ICA of a D dimensional random vector is a linear transform that minimizes the statistical dependence between its components. This representation in terms of independence proves useful in an important number of applications such as data analysis and compression, blind source separation, blind deconvolution, denoising, etc. Assuming the random vector we wish to represent through ICA has no noise and is zero-centered, the ICA Model can be expressed as,

$$Wh = s$$

where h is the random vector representing our data, s is the random vector of *independent components* with dimension $M \leq D$, and W is called the *filter* or *projection matrix*. To avoid ambiguities the filter matrix is chosen such that the independent components have unit variance (they already are zero-centered). The pseudoinverse of W which we will represent as A is called the *mixture matrix*, and it provides an alternative representation of the ICA Model. Given K classes $C - 1, ...C_k$ in a D dimensional space, the ICA model is estimated from the training set for each class. If W_k and s_k are the projection matrix and the independent components for class C_k with dimensions $M_k \times D$ and M_k respectively, then:

$$s_k = W_k(h - \bar{h}^k)$$

Where $h \in C_k$ and $\bar{h}^k$ is the mean of the class, estimated from the training set. We use the following notation for the density distribution of the independent components: $P^k(s)$. The density distribution of the projected data can be rewritten using the independence assumption for the independent components:

$$P^k(s) = \prod_{m=1}^{M_k} P^k(s_m)$$

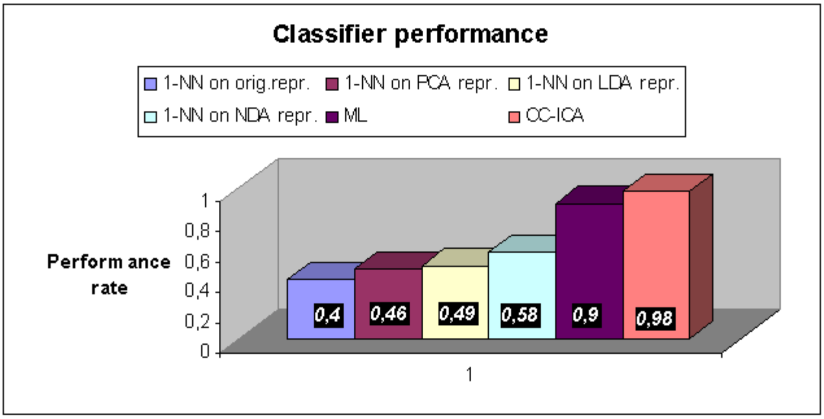

Fig. 3. Classifier Performance.

If H_l is a feature vector from a test object, we can project it into the ICA model learnt for class C_k and obtain the independent components s_l^k. The likelihood of the representative feature vector is obtained from the easier to calculate likelihood of the transformed feature. Using the log-likelihood to turn the products into sums, the Naive Bayes rule can be rewritten as,

$$C_{Naive} = \arg_{k=1,\ldots K} \max \sum_{l=1}^{L} \left(\sum_{m=1}^{M_k} \log P^k(s_{l_m}) \right)$$

5 Results

In order to compare our approach and to assess which is its performance for the problem of cork stopper classification, we have implemented and tested the following methods:

- The most simple classification technique, the Nearest Neighbor classifier (NN) [3], which classifies each cork representation -on the original space- to the class of the nearest representation of a stopper from the learning set;
- Principal Component Analysis (PCA) of the data for dimensionality reduction, followed by NN classification [3];
- Linear Discriminant Analysis (LDA), as described in [1], followed by NN classification;
- Nonparametric Discriminant Analysis (NDA), as described in [4];
- Maximum Likelihood Classification (ML) using a Gaussian distribution [3];
- Class-Conditional ICA, as described in section 4.

Fig.3 shows the performance of these methods, measured from test data using cross-validation. It can be observed that for this problem, parametric methods are clearly superior to the nonparametric ones. Class-Conditional ICA gets the maximum score.

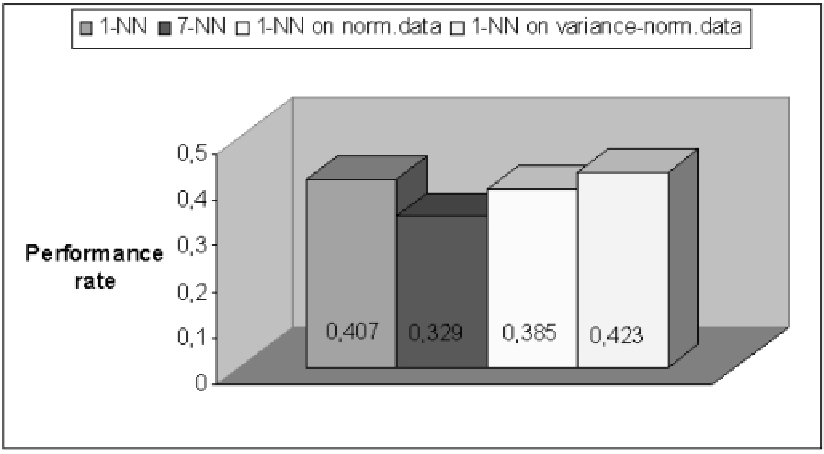

Fig. 4. Dependence of k-NN classifier on representation of data.

In fig.4 the results corresponding to various implementations of the nearest neighbor approach are shown: a) 1-NN classification on original data, b) 7-NN classification on original data, c) 1-NN on linearly normalized data, e) 1-NN on data normalized on variance. Performance of the method is not greatly affected by this fact.

Fig. 5 shows the dependence of the methods on the data representations. Different tests have been run on original data, normalized data and representation on reduced feature spaces by PCA and LDA. Different classifiers have been tested on original and normalized data. The figure shows the final results of applying Nonparametric Discriminant Analysis that gave a success rate of 52% and 58%. After reducing the feature space by PCA to different dimensions the final results were quite similar e.g. classifying a reduced space of 42 and 11 dimensions the success rate was 40% and 46%, respectively. When LDA has been applied to obtain an optimal subspace for class discrimination (the space has been reduced to R4) applying two different training sets, the results achieved 50% of success. Summarizing, the results show that the classification did not depend significantly on the data representation.

The results differed meaningfully when parametric classifiers are applied (see fig. 6). If the mean result of classification success was about 45% with nonparametric classifiers, parametric classifiers doubled the success rate achieving performance rate up to 98%. Furthermore, the results on the graphics show that keeping the full dimensionality of data is important for the classification performance. Maximum likelihood classification of original data got 90% of success while Naive Bayes classification on CC-ICA got 98%. The results from CC-ICA can be explained by a better estimate of the probability density function of the different classes of cork stoppers thanks to complexity reduction when we transform the estimate of an N-dimensional density function to N estimates of 1-dimensional density functions.

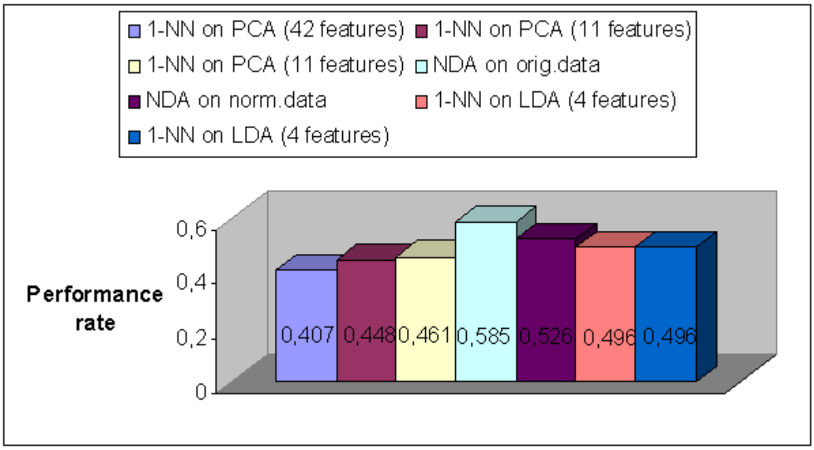

Fig. 5. Dependence on data representation.

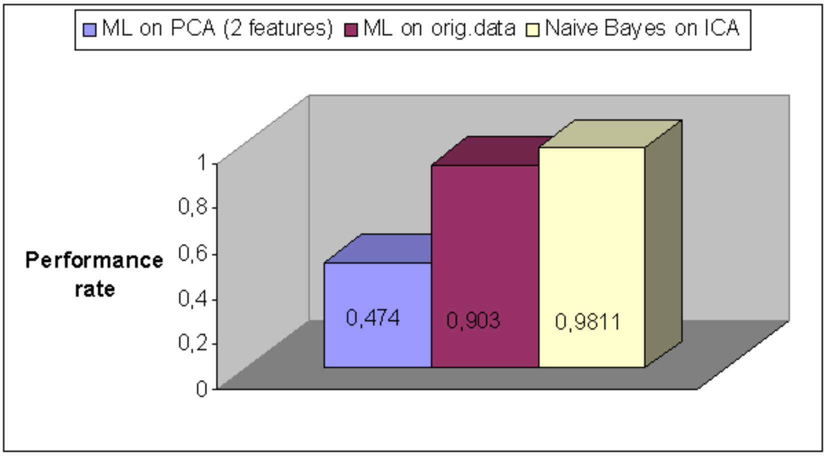

Fig. 6. Results of parametric classification.

6 Conclusions

We have tested the prototype in more than 2000 cork stoppers applying a set of parametric and non-parametric classifiers. Special attention has been paid to data representation. As a result, we have obtained that data representation of feature space of cork stoppers is important to the purposes of the correct classification. In particular, standard techniques for reducing feature space (in terms of blob characteristics of cork stoppers) like PCA and LDA hide the danger of loosing important information for the following process of classification. Our conclusion is that a sufficient representation of the probability density function of full dimensional cork data is the best approach. When faced with this problem, the CC-ICA method proved its advantages as a robust estimate of this function. After testing an extensive set of classifier methods, we concluded that non-parametric group of classifiers showed a low performance rate (mean of 45% of success). In contrast, Bayesian classification achieved high performance rate (mean 94%) in different tests. Although the computational complexity of the selected method of CC-ICA during the learning phase can be high, the process of classification of new examples can be implemented in real-time achieving the best performance rate.

Acknowledgements. This work was supported by IST project IST-1999-20188-CORKINSPECT sponsored by the European Comission. This project was also partially supported by "Ministerio de Ciencia y Tecnologia" grants TIC2000-1635-C04-04 and TIC2000-0399-C02-01. The work developed by M. Bressan has been supported by the Secretaría de Estado de Educación y Universidades of the Ministerio de Educación, Cultura y Deportes de España.

References

1. P. Belhumer, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on PAMI*, 19(7):711–720, 1997.
2. P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.
3. R. Duda, P. Hart, and D. Stork. *Pattern Classification (2nd Ed).* Wiley, New York,, 2000.
4. K. Fukunaga and J.M. Mantock. Nonparametric discriminant analysis. *IEEE Transactions on PAMI*, 5(6):671–678, November 1983.
5. I.T. Jolliffe. *Principal Component Analysis.* The MIT Press, 1986.
6. R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence Journal, Special Issue On Relevance*, 97(1):273–324, 1995.