Recent Issues in Pattern Analysis and Recognition
ALTERNATIVE FEATURE SELECTION PROCEDURES FOR PARTICLE CLASSIFICATION BY
PATTERN RECOGNITION TECHNIQUES.

Giuseppe Bonifazi,
Dipartimento di Ingegneria Chimica, dei Materiali, delle Materie Prime e
Metallurgia, Università degli Studi di Roma "La Sapienza".

Paolo Massacci,
Dipartimento di Ingegneria Chimica, dei Materiali, delle Materie Prime e
Metallurgia, Università degli Studi di Roma "La Sapienza".

Giacomo Patrizi,
Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università degli Studi di Roma "La Sapienza".

Abstract:
This paper examines a number of feature extraction algorithms to
improve the classification of mineral particles when the pattern
vector is defined as a series of Fourier coefficients. Considerable improvement can be obtained by using nonlinear
feature extraction algorithms, and this approach provides a
viable alternative to classification procedures of extremely
small samples of particles based on their physical and chemical
properties.

1. INTRODUCTION

Particle shape is being recognized as an important factor in
determining the behaviour of particulate solids in manufacturing
processes. The analysis, classification and the recognition of the
shape of a particle in suitable categories are therefore considered
important problems for which no general procedures have as yet been
formulated. A suitable method to effect this recognition seems to be
through pattern recognition techniques and Fourier analysis. In the case
of minerals, their crystal structure tends to give to the particles a
signature which can be recognized. Because of noise, a decision
theoretic approach to pattern recognition seems better suited for this
implementation than a syntactic approach, see Watanabe(1975).

Ideally, given a particle, it is desired to determine, with a
high degree of probability, the type of material to which it belongs and
its shape, without considering its chemical and physical properties but
only its geometrical properties, as obtained by automatic measurements
and their subsequent mathematical transformation. If this type of
classification is possible with a sufficient accuracy, then it may
become an important classification tool, since it is not destructive and
just a single particle would be sufficient. This would lead to a very
economic method of classification and control.
The aim of this paper is to examine some alternative feature extraction algorithms for the analysis of the characteristics of shape obtained through image techniques, with a view to the recognition of the type of materials considered and their size-class. A comparison will be made by applying the Karhunen-Loève feature extraction procedure, while retaining different number of elements, so that the recognition capabilities of each variant can be evaluated. The analysis will be based on a sample of over 1500 particles of three basic types of materials: Iron grit, Shale rock and Riolitic rock, classified in various sizes. The contour of the two dimensional projection of each particle has been obtained through image processing techniques, see Bonifazi et al. (1988).

The pattern recognition approach adopted here is a global or nonlinear classification approach. The objects considered have similar geometrical shapes, so that their classification must be based on suitable transformations of their shape characteristics, by defining higher moments, or Fourier transforms. This becomes particularly important when it is remembered that all measurements are corrupted by noise and that often the particles considered are not crystals.

In this paper the available patterns will be taken as given and their obtainment will be briefly summarized in section 2, as it has been fully discussed in Bonifazi (1987), Bonifazi & Massacci (1987), Bonifazi et al. (1988), together with the feature extraction algorithms. In section 3, the description of the classification algorithm applied will be given and its properties illustrated. In section 4, experimental results will be examined for the various alternative feature extraction procedures used, while in section 5, the formulateable conclusions will be presented.

2: PATTERN SELECTION AND FEATURE EXTRACTION:

The alternative definitions and the numerous procedures suggested, or utilized, to classify the dimensional and morphological properties of particulate solids, indicate the difficulties which may be encountered to determine these aspects, as discussed in Aravamudhan et al. (1984); Clark et al. (1984); Meloy et al. (1985), (1986).

2.1. Pattern Selection Procedures:

A set of particles can be characterized individually by their physical properties, by their chemical composition and their geometrical
characteristics, which may be analysed through a three-dimensional representation of the particles defined over the domain through suitable functions.

For the shape analysis and the morphology of a particle it is sufficient to obtain a black and white image of it and the shape itself may be evaluated by taking into account a projection of the image on a plane. The analysis of the gray scale pixels over the whole image allows the enhancement of the grays of the particle image from those of the background, through binarization techniques, see Aggarwal et al. (1977). The direct characterization of the morphology of a particle requires the recognition of the boundary of its image. In the procedure adopted, this has been done completely automatically, see Bonifazi (1987); Bonifazi & Massacci (1987); Bonifazi et al. (1988). Thus for this representation, choosing a fixed angular coordinate interval, say $2\pi/N$ radians, the distances of the origin from the boundary at these $N$ points is determined so that the shape of the image of the particle is represented by these distances and their angular coordinates.

These values may be represented in a straightforward way by a suitable Fourier series, so that the distances $R(j)$ $j=1, \ldots , N$ will be represented by:

$$R(j) = \sum_{k=0}^{N-1} c_k \exp \left(-j(2\pi/N)ik\right) \quad (2.1)$$

It is possible to generate the coefficients $c_k$ from knowledge of the distances $R(j)$ by the inverse relationship:

$$c_k = (1/N) \sum_{j=0}^{N-1} R(j) \exp \left(-j(2\pi/N)ik\right) \quad (2.2)$$

It is evident that both series are periodic with period $N$. Thus from the $N$ measurements determined on the image of the particle, the series of coefficients $c_k$ $k=0, \ldots , N-1$ can be estimated, which are called Fourier coefficients. Concerning the discrete Fourier series and transforms, convergence and approximation results are well known and have been extensively used in signal processing and image analysis, see: Cappellini (1985); Duda & Hart (1973); Edwards (1979); Papoulis (1977).

To calculate the Fourier coefficients, the Fast Fourier Transform (FFT) was adopted, as indicated by Luerkens (1981). Further, if the boundary presents marked concavities, the distance between the origin and the boundary becomes indeterminate. In this case for this procedure the smallest distance is considered.

Thus for the particles considered and for the experiments to be described below, a series of pattern measurements were obtained, in natural length units, indicating for the $N$ points the distance from the barycentre to the boundary at a fixed angular interval of $2\pi/N$ radians.
From this series a set of Fourier coefficients were determined, as indicated in (2.2), which, as is well known, some coefficients are the complex conjugates of others. See Edwards (1979). Specifically, the following relation holds:

\[ c_j = c_{n-j} \quad j = 1, 2, \ldots, N/2 \]  

(2.3)

where \( c_{n-j} \) is the complex conjugate of \( c_{n-j} \), so that only half of the series of measurements need be retained.

More precisely, for this polar Fourier implementation, the pattern vectors are formed by interleaving the imaginary coefficients with the real coefficients, and recalling that \( c_0 \) and \( c_{N/2} \) are always real, from the original \( N \) complex terms, pattern vectors of \( N \) elements are retained. No loss of generality is incurred in considering these patterns as consisting of real numbers, see Birkoff & MacClaine (1967).

2.2. Feature Selection Procedure:

Feature selection consists of reducing the dimensions of the pattern vector by discarding some elements of the vector or by defining a suitable linear or nonlinear transformation which results in a set of vectors more suitable for classification. Feature selection is essentially a redundant operation, as, Young & Calvert (1974) point out, since a suitable classifier can act directly on the patterns, without the intermediary of a feature vector. Nevertheless, feature extraction is considered an important part of Pattern Recognition, since it may contribute more insight into the classification procedure.

It is to be noticed that, in contraposition to the above remarks, any useful feature extraction algorithm may be used as a pattern selection algorithm, i.e. to decide what aspects of an object should be selected for measurement; see: Meisel (1972); Ullman (1973). However, pattern vectors are usually considered as given, constituting part of the problem data, which will depend on the data collection techniques, the experimental setup and other factors, not readily amenable to modification, so there is some point to define appropriate feature extraction procedures.

The feature extraction algorithm considered is the familiar Karhunen-Loève expansion method, see: Duda & Hart (1973); Ullmann (1973); Young & Calvert (1974). Essentially the procedure consists of minimizing a mean square error criterion of misclassification, when the mean vector of the class is used as a criterion of classification.

To do this, the covariance matrix of the pattern vectors is calculated and its eigenvalues and eigenvectors are estimated. Then
retaining a certain number of eigenvectors corresponding to the $k$ largest eigenvalues, the pattern vectors are premultiplied by a matrix formed from these eigenvectors. The resulting vector is the feature vector and is obtained from the original pattern vector by premultiplying it by the matrix of the selected eigenvectors, see Young & Calvert (1974). The resulting selection of features is optimal in the least squares sense, given $k$.

This procedure was implemented using the standard EISPACK routines for the calculation of the eigenvalues and eigenvectors, see Smith et al. (1976), Garbow et al. (1977).

3 CLASSIFICATION AND RECOGNITION:

The classification algorithm proposed in this application is a variant of one adopted for a number of recognition and classification problems, see: Patrizi (1979a), (1981), Calamassi & Patrizi (1988). It is based on the recognition algorithm of Firschein & Fischler (1963), whose convergence properties have been studied in detail by Patrizi (1979b).

The algorithm has two modes of operation: a training mode and a classification mode. In the training mode, a set of feature vectors are available belonging to known classes and the training set, formed by these vectors, includes members of every class considered. In the classification mode, a feature vector is submitted to the classifier and its class membership is determined by assigning it to that class corresponding to the barycentric vector which results nearest to that feature vector.

The algorithm in the training mode proceeds in the following way. Given the training set, an initial barycentric matrix is formed by finding the average feature vector from all the feature vectors of a class. The distance of each feature vector from each barycentric vector of the classes is determined. According to applications, the distance may be a generalized distance with different weights assigned to different components. All the feature vectors, which are nearer to a barycentric vector of a different class to that which it should belong to, are marked and among these, the one that is furthest from one of the barycentric vectors of its class is selected to form the seed to a new barycentric vector. All the feature vectors are reassigned to the barycentric vectors belonging to their own class. Thus at every iteration all the barycentric vectors remain the same, except for the class for which a new seed barycentric vector has been defined. In this
class, the feature vectors of the class are assigned among the different barycentric vectors according to their distance from each, to form subclasses. Once assigned, the barycentric vectors of this class are recalculated as the average of the feature vectors in each subclass. The barycentric matrix will contain one more column and the algorithm proceeds to another iteration, by recalculating the distances. The algorithm will eventually converge, if it is assumed that the training set satisfies certain mild conditions, see: Patrizi (1979b), Calamassi & Patrizi (1988). The algorithm in structural form follows:

CLASSIFICATION ALGORITHM (Training mode):

DCL NN number of features in the vectors,
DCL T number of objects in the training set,
DCL MM(j) number of barycentric vectors of class j,
DCL NC number of classes
DCL Data() Vector of features of the object i,
DCL Tracer(j) barycentric vector j.

Begin;
  For i = 1, ..., T Do;
    For j = 1, ..., MM(NC) Do;
      DD(i) = Min, ( Distance [data(i), tracer(j)] );
    End;
    CLTC(i) = arg ( Min, ( Distance [data(i), tracer(j)] ));
  End;
  For class_of(Data(i)) | CLTC(i) Do;
    I^* = arg( max, ( DD(I) ));
    Tracer(MM(NC) + 1) = Data(I^*) U class_of(Data(I^*));
  End;
Sort( Tracer per class);
Update ( number of barycentres MM(j) j = 1, ..., NC);
For i = 1, ..., T Do;
  P = class_of(Data(i));
  For j = MM(P-1)+1, ..., MM(P) Do;
    J^* = arg( min, ( Distance ( Data(i), Tracer(j) ) ));
  End;
  NTracer(J^*) = [N(J^*) * NTracer(J^*) + Data(J^*)] / ((N(J^*) + 1);
  N(J^*) = N(J^*) + 1;
End;
Tracer <- NTracer;
End;
Otherwise;
  Barycentric_matrix = Tracer;
  Terminate;
End;

CLASSIFICATION ALGORITHM (classifier mode):

DCL v the feature vector to be classified,

Begin;
  For j = 1, ..., MM(NC) Do;
    J^* = arg( min, ( DD(v) ));
    class_of(v) = class_of(Tracer(J^*));
  End;

End;
It is just worth noticing that the algorithm may be slow in the training mode, since many iterations may be required to reach convergence, but it will always be fast in the classificatory mode, as only matrix-vector product is involved and the selection of a least element.

4 PARTICLE CLASSIFICATION EXPERIMENTS:

To test the various feature selection algorithms for the classification procedures, various granular materials were chosen:

1) Material G: consists of iron grit used in the cutting and polishing of ornamental stone slabs. Its morphology greatly affects the velocity of cutting and polishing slabs. The material G1 is mainly composed of spheroidal particles, while G2 are particles of irregular shapes.

2) Material S: obtained by grinding Shale rock and has been divided into four size-classes:

   i) S1 (-4.000 + 2.000 mm)
   ii) S2 (-2.000 + 1.000 mm)
   iii) S3 (-1.000 + 0.500 mm)
   iv) S4 (-0.500 + 0.250 mm).

3) Material R: obtained by grinding riolitic rock and is used in the production of glass. It has been subdivided into three size-classes:

   i) R1 (-4.000 +2.000 mm),
   ii) R2 (-2.000 +1.000 mm),
   iii) R3 (-1.000 +0.500 mm)

For each group, at least 85 images were generated and in total 1589 images were formed, belonging to the nine classes of minerals described above.

The results have been classified in a number of ways to bring out certain aspects of this research. Thus in table 1, the performance of the classification algorithm is presented, for the cases in which the patterns were classified in the natural units (32 elements) and with 3, 5, 7, 9, 15 features. In all cases the training phase was completed successfully, in accordance with the results deduced in Patrizi (1979b), while independent samples were used in the classificatory phase to verify the performance of the various implementations.

In the table, for each class the percentage of misclassified patterns which resulted in the classification stage are presented. These patterns could have been used in the training phase, as knowledge of their membership is available. They were however sacrificed to have an
independent sample for testing. In table 1, only the percentages are given, as different numbers of patterns were used in the tests.

Table 1: Performance of the algorithm for Polar Fourier patterns under a natural classificatory experiment and feature extraction methods. Percentage misclassified.

<table>
<thead>
<tr>
<th>Class</th>
<th>Natural classifier</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.6</td>
<td>58.3</td>
<td>58.3</td>
<td>50.0</td>
<td>50.0</td>
<td>41.7</td>
</tr>
<tr>
<td>2</td>
<td>90.8</td>
<td>60.0</td>
<td>60.0</td>
<td>55.0</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>3</td>
<td>50.0</td>
<td>64.3</td>
<td>64.3</td>
<td>64.3</td>
<td>57.1</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>68.3</td>
<td>71.4</td>
<td>67.9</td>
<td>67.9</td>
<td>67.9</td>
<td>60.7</td>
</tr>
<tr>
<td>5</td>
<td>70.0</td>
<td>66.7</td>
<td>64.1</td>
<td>61.5</td>
<td>59.0</td>
<td>56.4</td>
</tr>
<tr>
<td>6</td>
<td>41.5</td>
<td>43.5</td>
<td>41.3</td>
<td>39.1</td>
<td>39.1</td>
<td>37.0</td>
</tr>
<tr>
<td>7</td>
<td>74.8</td>
<td>57.1</td>
<td>57.1</td>
<td>54.8</td>
<td>54.8</td>
<td>52.4</td>
</tr>
<tr>
<td>8</td>
<td>75.0</td>
<td>64.0</td>
<td>64.0</td>
<td>64.0</td>
<td>64.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>67.7</td>
<td>42.9</td>
<td>42.9</td>
<td>42.9</td>
<td>42.9</td>
<td>42.9</td>
</tr>
</tbody>
</table>

Mean 66.6 57.9 56.8 55.0 53.2 49.6

In table 2, exactly the same results are given for the three main groups of materials: Riolitic rock (R), Iron Grit (G) and Shale rock (S). The purpose of this experiment is to determine the performance of a two stage classifier. The first stage of the classifier selects the type of material, while the second selects the size class of the material.

Table 2: Performance of the algorithm for Polar Fourier patterns under a natural classificatory experiment and feature extraction methods. Material Groups, Percentage misclassified.

<table>
<thead>
<tr>
<th>Class</th>
<th>Natural classifier</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>61.8</td>
<td>43.1</td>
<td>41.7</td>
<td>41.7</td>
<td>40.2</td>
<td>37.5</td>
</tr>
<tr>
<td>G</td>
<td>50.6</td>
<td>37.3</td>
<td>35.8</td>
<td>35.8</td>
<td>35.8</td>
<td>32.8</td>
</tr>
<tr>
<td>S</td>
<td>34.1</td>
<td>31.9</td>
<td>30.5</td>
<td>29.8</td>
<td>29.0</td>
<td>27.7</td>
</tr>
</tbody>
</table>

Mean 45.7 36.1 34.6 34.3 33.6 31.4

For the Riolitic rock material group, the sample available for classification consisted of 74 elements of which 58 were correctly classified while 16 were misclassified. The result of the experiments for each material group are presented respectively in tables 3, 4, 5. On the diagonal of each table, the number of correctly classified elements are recorded for classification with 7 features and the results for 15 features are given in brackets, while the off-diagonal entries indicate the number of misclassified patterns. Assigned classes are read column-wise, while the correct classification is read by rows.
### Table 3: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Riolitic Material Group.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>R3</th>
<th>R2</th>
<th>R1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>22 (22)</td>
<td>3 (3)</td>
<td>..</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (1)</td>
<td>14 (16)</td>
<td>3 (3)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>..</td>
<td>7 (4)</td>
<td>22 (25)</td>
<td>29</td>
</tr>
</tbody>
</table>

### Table 4: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Iron Grit Material Group.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>G1</th>
<th>G2</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>22 (23)</td>
<td>6 (5)</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (2)</td>
<td>35 (37)</td>
<td>39</td>
</tr>
</tbody>
</table>

### Table 5: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Shale rock Material Group.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>S4</th>
<th>S3</th>
<th>S2</th>
<th>S1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>40 (41)</td>
<td>3 (2)</td>
<td>..</td>
<td>..</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (2)</td>
<td>22 (23)</td>
<td>1 (0)</td>
<td>..</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>..</td>
<td>1 (2)</td>
<td>10 (9)</td>
<td>..</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>..</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>20 (21)</td>
<td>23</td>
</tr>
</tbody>
</table>

For classification of elements belonging to the same material group, the precision of classification increases markedly. Thus, for the Riolitic rock group and with 7 and 15 features, a correct classification is obtained respectively in 77% and 82% of the attempts, while for the Iron Grit group, these become 85% and 89%. Finally for the Shale rock group, the percentage correct classification is respectively 89% and 91%. It is to be noticed that in all cases, except for one instance, the misclassification occurs with an element of an adjacent class.

It is seen, therefore, that if a classification approach based on limited comparison seems to give a much better performance than a classification based on a wide set of classes. However, the problem remains on how to select the material group.

In tables 6, 7, 8 a comparison is made by examining the classification performance of the algorithm for materials of different groups, but of the same size class. Thus, R3 is compared to S3, R2 to S2 and R1 to S1.
Table 6: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Samples from R3 and S3.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>R3</th>
<th>S3</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>17 (17)</td>
<td>7 (7)</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 (4)</td>
<td>37 (38)</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 7: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Samples from R2 and S2.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>R2</th>
<th>S2</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>15 (16)</td>
<td>5 (4)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 (6)</td>
<td>19 (19)</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 8: Performance of the algorithm for Polar Fourier patterns with 7 & 15 features (in parentheses). Sample from R1 and S1.

<table>
<thead>
<tr>
<th>Class</th>
<th>Assigned class</th>
<th>R1</th>
<th>S1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>16 (16)</td>
<td>12 (12)</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 (7)</td>
<td>21 (21)</td>
<td>28</td>
</tr>
</tbody>
</table>

The overall correct classification percentage for samples from different groups and from the same size class is, respectively, for 7 features and 15 features, 74.9% and 76.1%. Finally, it may be of interest to inspect how the correct classification varies with an increasing number of features. In table 9, these results are reported for a sample composed of minerals of the type R3 and S3.

Table 9: Performance of the algorithm for Polar Fourier patterns with 3, 5, 7, 9, 15 features. Sample from R3 and S3.

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>R3</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>14</td>
</tr>
<tr>
<td>S3</td>
<td>15</td>
</tr>
</tbody>
</table>

The evidence available in the tables seems to indicate that nonlinear characteristics have a strong influence on the classification. This is in line with the previous findings, see Bonifazi et al. (1988). Further, detailed comparisons of the behaviour of the number of misclassified patterns, as the number of features increase, indicate
that a misclassified patterns tend to be insensitive to the number of features between a lower threshold and an upper one. As the number of features increase, the random noise component seems to predominate, leading to a deterioration in performance. This is in accordance with similar experiments, Bonifazi et al. (1988) and the ugly duckling theorem, as Watanabe (1985) has so aptly demonstrated.

5: CONCLUSIONS:

The results presented indicate that a pattern recognition procedure to recognize and classify particulate solids is not only possible, but that it compares favorably with more traditional forms of analysis, based on chemical and physical methods. The great advantage of this approach is that it can be implemented completely by machine, and that it is not destructive, so a considerable scope for this method may arise from the special materials under development, such as ceramics, where the purity, the size and the shape of these materials is very important for the technical characteristics of the manufactures.

From a general point of view, the nonlinearities in the pattern elements do not seem to be fully analyzed, so probably more powerful methods of feature extraction are required than those inherent in the Karhunen-Loève procedure.

6: REFERENCES:


