Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains

P. LI, W. J. TRELOAR, J. R. FLENLEY* and L. EMPSON
Geography Programme, School of People, Environment and Planning, Massey University, Palmerston North, New Zealand

ABSTRACT: The automation of palynology (the identification and counting of pollen grains and spores) will be a small step for image recognition, but a giant stride for palynology. Here we show the first successful automated identification, with 100% accuracy, of a realistic number of taxa. The technique used involves a neural network classifier applied to surface texture data from light microscope images. A further significance of the technique is that it could be adapted for the identification of a wide range of biological objects, both microscopic and macroscopic. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: pollen; classification; automation; texture measures; neural networks.

Introduction

Palynology is used in Quaternary science, asthma and allergy research, the search for oil and gas, climate change research, palaeoecology, archaeology, forensic science, plant pathology and apiculture. It is at present a laborious and subjective process. The potential advantages of automating palynology have already been pointed out (Stillman and Flenley, 1996). They include: faster results, greater objectivity, finer taxonomic determinations, increased count size, increased sampling frequency and wider applications. The automation of palynology has therefore been called for as an urgent priority (Stillman and Flenley, 1996; Green, 1997).

Although the idea of automating palynology was suggested over 30 years ago (Flenley, 1968), it proved difficult to implement. An early attempt to use holography (Mirkin and Bagdasaryan, 1972) was doomed because it was, in effect, using template matching on material with far too much variability for it to succeed. This biological variability also overcame a method applying multivariate statistical analysis to data on size and shape of pollen grains (Witte, 1988), which achieved only 68% accuracy. Size and shape measurements vary by ±15% or more, with shape being particularly variable because the pollen grain wall is made of a flexible polymer, sporopollenin (Brooks and Shaw, 1968), so that fossilised pollen grains are frequently distorted or completely flattened. A further complication introduced by light microscopy (LM) is that the depth of focus at the high magnifications needed (typically × 400 to × 1000) is far less than the diameter of the grain, so that the grain surface cannot be entirely in focus at any one focal level.

The use of texture analysis (Langford et al., 1990) overcame the problem of variable size and shape. Various measures of surface texture (Haralick, 1979; Pietikainen et al., 1983) were used to characterise the surface pattern, and these data were subjected to Fisher’s linear discriminant analysis. This was done on scanning electron microscope (SEM) images, thus overcoming the problem of depth of focus. The result was 98.4% successful with six taxa (Langford et al., 1990).

Since SEM is too costly and elaborate for routine use, the next hurdle was to find a technique suitable for LM (light microscopy) and apply it to a larger number of taxa. Texture measures were found to be diagnostic even on optical images not in perfect focus, but multivariate analysis became inadequate with increased numbers of taxa (Treloar, 1993). The work described in this paper was undertaken to solve these problems.

Three different experiments are described. The first of these develops the use of texture variables derived from light microscope images. This is used with shape analysis (Treloar et al., 2004) to classify four pollen taxa by linear discriminant analysis.
The second experiment also uses texture and shape features, incorporates a neural network classifier, and succeeds in classifying 13 pollen taxa.

In the third experiment, texture alone is used, and the same 13 pollen taxa are successfully classified.

**Experiment 1**

**Materials and methods**

**Pollen sample**

The following four fresh pollen taxa were used:

- Hoheria populnea (H.pop);
- Phormium tenax (P.ten);
- Phymatosorus novaerzelandiae (P.nov);
- Podocarpus totara (P.tot).

The letters in brackets represent a code for each taxon and will be used from here on. These samples were taken from reference material originating from the New Zealand flora.

**Image capture**

Eighteen images of each species were captured by a framestore within a PC connected to the video output from a Philips CCD camera, mounted on a Zeiss Axiopt microscope (see Fig. 1).

The framestore employed was a Data Translation DT2855 with 256 grey-levels, image dimensions of 768 x 512 pixels and a 1:1 aspect ratio. The images, once in the framestore, were analysed directly using a custom-built interface and programs written with a Borland Pascal compiler.

Sub-images were taken for subsequent texture analysis. To avoid bias in the choice of sub-images, areas were selected in a standardised manner. The area selected was quite simply the largest square region that could fit completely inside the pollen grain at the maximum magnification possible. This eliminates some of the problem areas of a pollen grain, for example the edges of the grains that curve away from the viewer and change their texture. It also averages out the changes in texture in intensity between adjacent pixels, resulting in edge enhancement. The convolution is repeated for the other three operators (Fig. 2(c), (d), (e)) and the maximum value at each point is stored in the image. Texture analysis procedures were then carried out on the resulting edge-enhanced image, as well as on the unmodified image.

**Texture measures**

Haralick et al. (1973) introduced a set of second-order statistics to analyse the grey-tone spatial dependence of texture. It has been used in various ways. Don et al. (1984) used it to measure surface roughness in metals and Weszka et al. (1976) classified terrain from aerial and satellite images.

The starting point of this type of analysis is the construction of co-occurrence matrices containing information on the spatial organisation of grey-levels of an image. This is best illustrated with the use of an example, see Fig. 3, where a digital
image has been equalised to four-grey-levels, Fig. 3a. Then a
grey-level distribution matrix is constructed, Fig. 3b, using
the displacement vector \((d, \theta)\) where \(d\) is the distance and \(\theta\)
the angle between pixel pairs. In the case shown in Fig. 3
\((d, \theta)\) is \((1, 0^\circ)\). This means that each pixel in the equalised image
is compared with the pixel found at a distance of 1 pixel away
and at an angle of 0° from the x-axis. The elements within the
matrix that are indexed by the grey-levels of the pixel pairs are
incremented. The matrix is then made symmetrical, Fig. 3c,
about the main diagonal by averaging equivalent elements,
e.g. elements \((1,2)\) and \((2,1)\). The co-occurrence matrix,
Fig. 3d, is then easily formed by dividing each element by
the sum of all the elements in the symmetrical grey-level distri-
bution matrix. Each element \(P(i, j|d, \theta)\) is the probability of
finding grey-levels \(i\) and \(j\) separated by a displacement vector
of \((d, \theta)\).

When co-occurrence matrices are used to analyse texture, a
variety of displacement vectors are used, in this case from 1 to
17 pixels. To reduce the effects of rotation of a texture in this
application, the displacement vectors over the angles 0°, 45°,
90° and 135° are usually averaged.

Haralick et al. (1973) described 14 statistical measures of
distribution within the matrix that relates to textural informa-
tion. Five of these measures are used in this paper. These are
as follows:

1. **Angular second moment**, or homogeneity of the image, is a
measure of how evenly the values are distributed through-
out the matrix; the lower the value, the greater the homo-
geney.

   \[
   A_{d, \theta} = \sum_{i=0}^{L} \sum_{j=0}^{L} P(i, j)^2 
   \]

   where \(L\) is the maximum grey-level.

2. **Contrast or local variation** (coarseness) of grey-levels is a
measure of the moment of inertia about the main diagonal
of the matrix. When the probabilities are concentrated
about this main diagonal the contrast is low.

   \[
   C_{d, \theta} = \sum_{i=0}^{L} \sum_{j=0}^{L} (i - \mu)^2 \cdot P(i, j) 
   \]

3. **Variance** is a measure of the spread of the probabilities in
the co-occurrence matrix. The closer the probabilities are
to the centre of the matrix, the lower the variance.

   \[
   V_{d, \theta} = \sum_{i=0}^{L} \sum_{j=0}^{L} (i - \mu)^2 \cdot P(i, j) 
   \]

4. **Inverse difference moment** or local homogeneity is given by:

   \[
   I_{d, \theta} = \frac{\sum_{i=0}^{L} \sum_{j=0}^{L} P(i, j)}{1 + (i - j)^2}
   \]
(5) Entropy or texture non-uniformity gives high values when the elements in the matrix are large and low values when they are unequal.

\[ E_{ab} = -\sum_{i=0}^{k} \sum_{j=0}^{k} P(i, j) \cdot \log P(i, j) \]  

Classification

The measures produced for each of the pollen taxa are used to determine a set of classification rules. These are then applied to unknown samples to provide a measure of the classifier efficiency. As one variable is highly unlikely to separate the different pollen classes, a combination of variables is selected. A feature vector is constructed and applied to a multivariate statistical classification scheme.

The classifier employs Fisher’s linear discriminant function to find the optimal linear direction in multivariate space that separates feature vectors from two classes. The separation between classes is defined as the distance between the class feature variable means, standardised for within-class variance. Consider two classes \( C_i \) and \( C_j \) each with a set of feature vectors \( x_{i1}, x_{i2}, \ldots, x_{ik} \) and \( x_{j1}, x_{j2}, \ldots, x_{jk} \). Their optimal linear direction \( \alpha_{ij} \) is given by:

\[ \alpha_{ij} = \left( \Sigma_i + \Sigma_j \right)^{-1} (\mu_i - \mu_j) \]

where \( \mu_i \) is the sample means of class \( i \), and \( \Sigma_i \) is the sample covariance of class \( i \).

This is calculated for all the samples in a training set of features, and then used to construct a decision boundary, defined by:

\[ B_{ij} = \frac{\mu_i \sigma_j + \mu_j \sigma_i}{\sigma_i + \sigma_j} \]

where \( \mu_i \) is \( \alpha \times \) sample means of class \( i \), and \( \sigma_i \) is \( \alpha \times \) sample standard deviations of class \( i \).

To determine the class to which an unknown sample \( x \) belongs, \( \alpha \) is calculated, compared to the boundary and assigned to a class using the following rules:

- If \( \alpha_{ij} \cdot x > B_{ij} \) then assign to \( C_i \) (10)
- If \( \alpha_{ij} \cdot x < B_{ij} \) then assign to \( C_j \) (11)
- If \( \alpha_{ij} \cdot x = B_{ij} \) then indeterminate (12)

In practice, the final case (12) is very unlikely.

As there are more than two possible classes into which an unknown may fall, a voting system is used. The unknown feature vector is compared with each of the known class pair boundaries. In each case, the one to which it is assigned has its vote counter incremented. As soon as all the class pairs have been considered, the vote counter with the highest value represents the class into which the unknown is classified. Consider a case where an unknown sample \( x \) is from class 1, and there are four classes. When \( x \) is compared between class pairs 1–2, 1–3, and 1–4, class 1 should be selected every time. But when it is compared between class pairs 2–3, 2–4, and 3–4, its selection should be essentially random. Thus, the class 1 vote counter will have the largest value, and therefore class 1 is selected.

A problem may occur if two or more class vote counters are drawn with the maximum value. In this case, the above classification procedure is repeated with only the drawn class pairs used. If classification is still not possible, the system awards an equal probability to each drawn class.

A leave-one-out (cross-validation) scheme was used to test the true success rate of the classifier. This involves taking each sample in turn and classifying it using a discriminant function built from the remaining samples. This ensures that the training set is as large as possible, and provides a good assessment of the true success rate. It does, however, significantly increase the computational cost, as a new classifier is constructed for each sample.

Although many texture variables were calculated, it is neither necessary nor practical to use them all to construct the final discriminant functions of the classifier. A subset of variables often exists that produces comparable results, but with less calculation. Also, increasing the dimensionality of the classifier can lead to diminished classifier performance (Hand, 1981). Thus, a variable selection procedure was used to select the optimal subset of variables.

Hotelling’s \( T^2 \) statistic, a multivariate extension of the standard single variable test, was employed as a measure of separability between two classes. The \( T^2 \) statistic is useful for this type of search as it is strictly monotonic, i.e., the \( T^2 \) value of a subset of variables is less than or equal to that for the full variable set. It indicates whether any two pollen classes are statistically inseparable using the available data set, by producing a measure of the distance between sample means, normalised to dispersion within the samples:

\[ T^2 = \frac{s_j}{s_i} \left( \frac{N_i - 1}{s_i} \right) \left( \frac{N_j - 1}{s_j} \right) \cdot \left( \mu_i - \mu_j \right) \]

where \( s_j \) is the number of samples in class \( i \), \( s_i \) is the mean centroid vector of class \( i \) (sample mean for each variable), and \( s \) is the assumed common variance–covariance matrix.

In practice, however, \( s_i = s_j \) and the within-class scatter matrix \( W \) is used in place of \( s \) as it is easier to calculate, thus:

\[ T^2_n = \frac{1}{2} \cdot \left( \frac{N_i - 1}{n_i} \right) \cdot \left( \mu_i - \mu_j \right) \cdot \left( \mu_i - \mu_j \right)^\top \]

where

\[ W = \sum_{j=1}^{n_i} \sum_{k=1}^{n_j} (x_j - \mu_j) \cdot (x_j - \mu_j)^\top \]

and \( s \) is the number of samples in class \( n_i \), \( n_j \) is the number of sample points in class \( j \), \( x_j \) is the \( \text{th} \) point from class \( j \) and \( \mu_j \) is the sample mean for class \( j \).

The variable selection procedure employed was a sub-optimal search, using sequential backward elimination (Hand, 1981). This was applied to several sets of texture variables. The most efficient combinations of variables for each class pair were noted. These were recombined and passed through the selection procedure again to find the optimum variable set to be used in the classifier.

Sequential backward elimination starts with a complete variable set and removes the variable that reduces the value of \( T^2 \) the least. In other words, \( T^2 \) is calculated for the complete data set with consecutive variables removed. The variable omitted for the maximum value of \( T^2 \) is then permanently removed from the data set. This procedure is repeated until the value of \( T^2 \) falls below the absolute separability threshold value given by:

\[ T^2_{\text{min}} = \frac{(2s - 2) \cdot N \cdot F}{(2s - 1 - N)} \]

where \( N \) is the number of variables in feature vectors, \( F \) is the 5% level of the F-distribution with \( N \) & \((2s - 1 - N)\) degrees of freedom and \( s \) is greater than \( N \).

Copyright © 2004 John Wiley & Sons, Ltd.

Table 1  Confusion matrix of classification results

<table>
<thead>
<tr>
<th>Class assigned</th>
<th>H.pop</th>
<th>P.ten</th>
<th>P.nov</th>
<th>P.tot</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.pop</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P.ten</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P.nov</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>P.tot</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Results

Table 1 shows the confusion matrix produced by the leave-one-out classifier, using the texture variables discussed earlier. The vertical matrix index relates to the class entered into the classifier, and the horizontal matrix index relates to the class to which it was assigned. The elements within the matrix are expressed as percentages.

The classifier performed very well, with all taxa identified 100% of the time. This was achieved using only eight variables: angular second moments with displacement vectors of 1, 2, 3 and 4 pixels, angular second moments of Sobel-enhanced images with displacements of 2, 3 and 6 pixels, and finally, contrast of Sobel-enhanced images with a displacement of 9 pixels.

Conclusions

These results are very promising, as they demonstrate that texture classification of light microscope images is feasible using relatively few variables. The high classification rates, compared with earlier studies, are mainly due to the fact that the scheme was specifically designed for the pollen in question. Although only the angular second moment and contrast variables were used in this case, other Haralick measures may prove more effective for other pollen groupings.

Experiment 2 (Li et al., 1998)

Materials

The materials used were 13 taxa of modern pollen and spore types found in New Zealand, many of which are of European origin (see Table 2). These taxa were chosen to give a representative range of major pollen classes and sculpture types. Pollen and spores were prepared using standard pollen preparation techniques (Moore et al., 1991; Faegri et al., 1989), and mounted in glycerine jelly. Images of 18 different specimens of each pollen taxon were digitised using a Zeiss microscope equipped with a monochrome TV camera giving images of 256 grey-levels. Typical surface textures of the 13 types are shown in Fig. 4.

Method

This experiment relied partially on shape analysis. In modern reference material, shape is commonly preserved as a usable feature. The 13 types of pollen were first classified into groups according to their geometric features: area, perimeter length, and compaction index (Treloar, 1992; Treloar et al., 2004). This was done using Fisher’s linear discriminant analysis.

Further analysis was carried out using Haralick’s texture measures (Haralick et al., 1973) and Law’s masks (Pietikainen et al., 1983). The Haralick measures were extended by the Sobel edge-enhancing operators and the $\sqrt{2}$ operation (see Experiment 1), and the displacement was chosen to be from 1 to 32 pixels. Neural network classification of these surface texture measures was then carried out using a feedforward multilayer perceptron (MLP) trained with back-propagation of error. The MLP consists of layers of artificial neurons with weighted connections joining all neurons in successive layers. Input data from the calculated features derived from pollen images are fed into the input layer of the network, and the network computes the activations of the neurons based on the weighted connections and the inputs. At the output level of the network, is a layer of nodes, with each node representing a pollen taxon. The node with the highest activation in the output layer is selected as the classification.

Training consists of introducing inputs from a pollen example allowing the network to compute its classification and then, if the classification is wrong, systematically updating the weighted connections between nodes so that the network learns about that input. Training of the MLP proceeds over many cycles where all of the data in the training set, together with their correct classification, is presented in each cycle. Following training, an independent set of test images is used to evaluate the accuracy of the network.

Leave-one-out testing, as was used in Experiment 1, could also have been used here, but the neural network took a long time to train, making the technique infeasible.

Results and conclusion

The pre-classification based on shape divided the images into six groups. The neural network classification then successfully identified all of the images. In no case were more than 74% of the images needed for the training, and the average value for this was 54%. This 100% successful classification of 13 pollen taxa is clearly superior to the successful classification of four pollen taxa in Experiment 1.

Experiment 3 (Flenley et al., 1999)

Materials

Given that fossil pollen grains are frequently distorted, flattened and damaged, the criteria of size and shape (as used in the pre-classification in Experiment 2) are clearly potentially unreliable. On the other hand, surface texture is frequently characteristic, even when grains are damaged or fragmented. The next step was therefore to leave out the pre-classification based on shape.

Method

The same 13 pollen taxa as in Experiment 2 were again used. The same 18 images of each taxon were used, giving 234 images in all.

Method and results

The same type of MLP neural network as in Experiment 2 was used. Again, 100% classification rates were achieved for the
13 pollen taxa. Several variants in the design of the neural network all produced very similar results (see Table 3).

The set of images was split into training and testing sets, with at least 76 images (32%) in the test set, making the results sound.

Conclusion

The result suggests that surface texture is a reliable feature for identification. This may mean that even distorted or fragmented grains might be classifiable by the methods described here for automatic identification.

General discussion and conclusion

We consider that MLP neural networks, trained using surface texture data, will be successful in classifying pollen assemblages into their constituent taxa with great efficiency. Increasing the 13 types to the 50 or so needed in routine palynology in
Table 2  List of pollen taxa used

<table>
<thead>
<tr>
<th>Species</th>
<th>Family</th>
<th>Identification (Fig. 4)</th>
<th>Major pollen class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascarina lucida</td>
<td>Chloanthaceae</td>
<td>A</td>
<td>Monocolpate</td>
</tr>
<tr>
<td>Coprosma lucida</td>
<td>Rubiaceae</td>
<td>B</td>
<td>Tricolpate</td>
</tr>
<tr>
<td>Dodonaea viscosa</td>
<td>Sapindaceae</td>
<td>C</td>
<td>Tricorpus</td>
</tr>
<tr>
<td>Hoheria populnea</td>
<td>Malvaceae</td>
<td>D</td>
<td>Periporate</td>
</tr>
<tr>
<td>Holcus lanatus</td>
<td>Poaceae</td>
<td>E</td>
<td>Monoporate</td>
</tr>
<tr>
<td>Knightia excelsa</td>
<td>Proteaceae</td>
<td>F</td>
<td>Tricolporate</td>
</tr>
<tr>
<td>Leptospermum scoparium</td>
<td>Myrtaceae</td>
<td>G</td>
<td>Syncolpate</td>
</tr>
<tr>
<td>Notholagus tusca</td>
<td>Fagaceae</td>
<td>H</td>
<td>Stephanocolpate</td>
</tr>
<tr>
<td>Phormium tenax</td>
<td>Agavaceae</td>
<td>I</td>
<td>Trichotecmolpate</td>
</tr>
<tr>
<td>Phymatosorus novaeezelandiae</td>
<td>Polyopoideae</td>
<td>J</td>
<td>Monolete</td>
</tr>
<tr>
<td>Plantago lanceolata</td>
<td>Plantaginaceae</td>
<td>K</td>
<td>Periporate</td>
</tr>
<tr>
<td>Podocarpus totara</td>
<td>Podocarpaceae</td>
<td>L</td>
<td>Vesiculate</td>
</tr>
<tr>
<td>Taraxacum officinale</td>
<td>Asteraceae</td>
<td>M</td>
<td>Fenestrate</td>
</tr>
</tbody>
</table>

Acknowledgements  We thank the British Science and Engineering Research Council for a grant at Hull University, and Massey University Research Fund for postdoctoral fellowships at Massey University for W. J. Treloar and Ping Li. We also thank Sarah Metcalfe, Annette Parkes, David Chesmore, Tony Sinclair, Bernd Striewski, Bob Hodgson, Stephen Marsland and Ollie Harris for their advice and help.

References


Forster RM, Flenley JR. 1989. The application of density gradient centrifugation to palynology. School of Geography & Earth Resources, University of Hull: Miscellaneous Series No. 35.


Table 3  The neural networks used and the results obtained

<table>
<thead>
<tr>
<th>Network identification</th>
<th>Structure</th>
<th>Images trained</th>
<th>Images tested</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N13_6-20-13+AI.net</td>
<td>6-20-13</td>
<td>153</td>
<td>81</td>
<td>234</td>
</tr>
<tr>
<td>N13_6-25-13+AI.net</td>
<td>6-25-13</td>
<td>149</td>
<td>85</td>
<td>234</td>
</tr>
<tr>
<td>N13_6-30-13+Co.net</td>
<td>6-30-13</td>
<td>158</td>
<td>76</td>
<td>234</td>
</tr>
<tr>
<td>N13_6-35-13+DI.net</td>
<td>6-35-13</td>
<td>146</td>
<td>88</td>
<td>234</td>
</tr>
</tbody>
</table>

Computer technology will massively increase the training time needed. The work reported here was carried out on a 486 dx2/33 computer. A modern computer has been found to perform satisfactorily even for such large data sets. There are also techniques (see, for example, Demuth and Beale, 1996) for greatly reducing the training time. In any case, the training has to be done only once for each region of the world, and actual identification and counting of assemblages will then be extremely rapid. As an interim stage, easy pollen types could be counted automatically, and difficult ones identified mechanically for the palynologist to identify manually. Work is envisaged on training neural networks to identify Easter Island pollen types, as this island has a total pollen flora of ~ 40 types (Flenley et al., 1991). We will then test this using some fossil assemblages from Easter Island.

Techniques already exist for the automated scanning of slides (Watanabe, 1974), for finding objects on them (Gelsema and Landerege, 1982), for improving the purity of pollen preparations by density gradient centrifugation (Forster and Flenley, 1989, 1993), and for automatic preparation (Stillman and Flenley, 1996). Therefore, the way is now clear for the full automation of palynology, with all the advantages listed above. We also believe that a neural network, as described here, could easily be adapted for the classification of a wide range of microscopic and macroscopic biological objects.


