9 Automatic recognition and classification of archaeological charcoals

Raquel Pique i Huerta & Josep Miquel Pique i Huerta

9.1 INTRODUCTION

Anthracology is an analytical technique that allows us to identify specifically charcoal fossils found in archaeological excavations. This identification is done through microscope observation of their anatomy. The data provided by anthracology allow us to reconstruct the vegetation of trees and shrubs surrounding the excavation and their evolution. They also provide information related to the use of combustible materials and the exploitation of natural resources.

The use of image processing in Archaeology provides many possibilities that have not been explored yet. In general, it allows us to process the main data of the image in an objective way, thereby improving the analytical process. In the case of Anthracology, we intend to use it in order to enhance images, to quantify anatomic characteristics of wood and to classify samples by using an interactive expert system.

9.2 ANTHRACOLOGICAL ANALYSIS METHOD

The anthracological analytical method is based on the observation of anatomical characteristics of carbonised wood by using an optical reflection microscope. As charcoal preserves the internal structure of wood, its anatomy can be easily recognised. In order to classify it, it is necessary to study charcoal from the three standard anatomical sections of wood (Figure 9.1): transversal, longitudinal tangential and longitudinal radial, all of them obtained through manual fracture of the charcoal with no previous preparation. In doing so, three defining characteristics can be observed. These characteristics allow us to differentiate vegetal types and, in some cases, to determine the species. Distribution, size and number of trachea are some of the characteristics that allow us to make a distinction between species. In this way, classification is quite trustworthy if the charcoal is in good condition. In order to identify the characteristics we use a reflected light optical microscope and lenses of 40× to 400×.

In order to determine the type and species from which the charcoal stems, we compare the cell structure of fossil charcoals with that of today's wood. We can do that by using either a reference collection of carbonised wood or various atlases of wood anatomy (Carlquist 1988; Jacquiot 1955; Jacquiot *et al.*, 1973; Schweingruber, 1978).

9.3 APPLICATION OF DIGITALISATION TO ANTHRACOLOGY

Image processing does not solve problems derived from the quality of preservation of samples. In our opinion, however, digitalisation and image processing can help to solve problems inherent to the determination process, such as the anatomical resemblance between species of the same type, making differentiation difficult.

It is on the technical side that digitalised image processing can help the most: improving the quality of charcoal images would make their identification easier, and biometrical analysis would be carried out more rapidly. It would also aid the development of specific studies on determining biometrical characteristics, fundamental to the determination of some species.

Anthracology is technically based on the identification of images using a microscope. Digitalized image processing leads to the improvement, quantification and statistic study of the images. This can be combined with a database later on, as we will explain below, by means of an open, in-

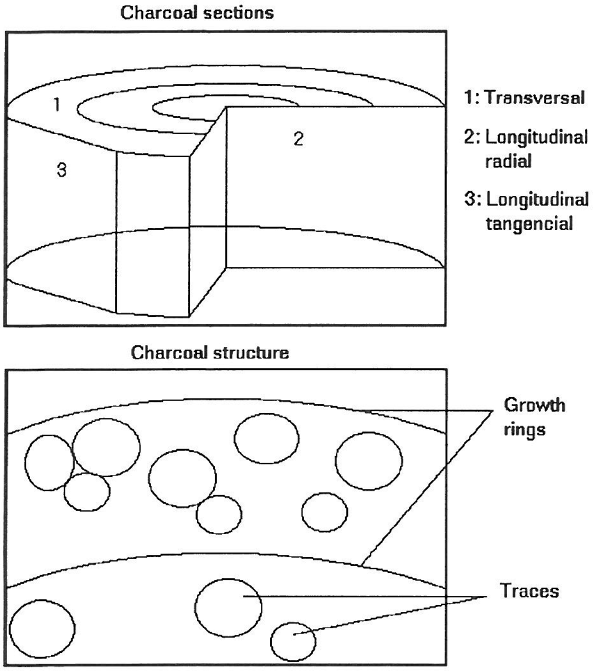

Figure 9.1: Illustration of charcoal sections and charcoal structure.

teractive system. The advantages of combining both the analytical systems are therefore obvious.

In the present case, we intend to perfect a technique for determination of fossil charcoals in the laboratory. The process is basically as follows:

1) Digitalizing images from a reference collection, treating and storing them in order to create an image base;
2) Isolating determining characteristics;
3) Achieving automation in the process of determination of taxons, by using digitalisation and processing of those images visualised in the microscope and by referring interactively to a database.

In order to obtain an initial experimental database, we have selected the most common trees and shrubs in the Mediterranean which are those normally found in archaeological excavations. In the future, the database can be improved with other European and American species. To start with, we have only used the Angiospermae. As we have pointed out above, the determination of charcoals is based on the existence of characteristics that make the differentiation of vegetal species possible. These characteristics can be observed in different anatomical sections.

To begin this study, we have focused on one of these sections, the transversal one. Our decision is based on technical reasons and on the fact that it allows quicker recognition of the vegetal species.

To implement image processing in the field of Anthracological analysis we have only taken into consideration those characteristics relating to distribution, size and number of trachea in every annual ring.

9.4 DESCRIPTORS USED

In the first place, we have established a group of descriptors or determining features that we have used in the design of our reference database. These descriptors will give us the information relating to the similarity of the charcoal we are studying with the reference charcoals. We base our study on the association forms and the size of the trachea (we have used data from the works by Schweingruber 1978; Jacquiot, Trenard & Dirol 1973, Carlquist 1988).

The determining characteristics for Angiospermae in their transversal section are with relation to the general disposition of the trachea:

1) species with a porous zone: those in which the larger trachea are grouped at the beginning of the annual ring;
2) species with a spread porosity and a semi–porous zone: those in which the tracheae are homogeneously distributed throughout the woody fibre and whose size do not vary much;
3) species with a flamed distribution of pores: tracheas are disposed in the shape of a flame;

Other morphological characteristics to be taken into account are those referring to the particular disposition, shape and quantity of the tracheas:

1) mainly isolated, some of them in groups of two;
2) grouped in two or three, some of them isolated;
3) grouped in radial rows;
4) radial disposition of the tracheas;
5) grouped in tangential bands;
6) ring aligned in the annual ring;
7) decreasing size of the trachea from the beginning to the end of the annual ring;
8) the same size of the trachea from the beginning to the end of the annual ring;
9) ring concentration in the initial woody fibre.

Relating to the tracheas' biometrical characteristics, we establish the following classes in microns:

10) very small, 10 to 25;
11) small, 25 to 55;
12) medium, 55 to 100;
13) large, 100 to 250;
14) very large, 250 to ...

And according to quantity:

15) many;
16) few.

9.5 PLANNING

The process to be followed to classify and assign the charcoal consists of several steps, briefly explained in the succeeding sections.

Recalling the restriction to only the transversal sections of angiosperms, this study has, in a first experimental stage, drawn attention to the tracheas' disposition and diameter. Nevertheless, this will be easily generalised, since this work is universally structured around charcoal recognition.

The algorithms implemented will help us to determine the charcoal species not only through morphologic and biometric data but also through the help of a database which will permit the specialised user to delimit its range of possibilities: by introducing data either about specific morphologic features or about associated aspects such as the location of the site from which the samples come. Therefore emphasis is placed on the interactive aspects of the software user–dialogue.

This study uses the following hardware: a reflected light optical microscope, a video camera, an image digitiser card and a 80386 CPU computer. Nevertheless, it can easily be adopted to other configurations, and it is open to improvement.

9.6 CHARCOAL IMAGE DIGITALISATION

The charcoal image is observed through a 100x microscope, which is used throughout the study, since its biometric features provide outstanding information.

Two digitising methods may be used:

The first consists of the microscope image captured on black and white film to obtain sharper contrast. Once the film has been developed, the images are scanned with a video camera, and this data is introduced into the computer via a digitiser. A file of the charcoal image results, with 512 by 512 pixels, 8 bits per pixel, and at most 256 grey levels.

The second process consists of direct image digitalisation through a video camera mounted on the microscope allowing us to work in real time.

In both cases, once the Charcoal Digitalised Image (CDI) is obtained, the way is open to subsequent computer file processing.

9.7 FEATURE EXTRACTION AND ENHANCEMENT

The recorded image goes through a whole series of processes in order to obtain the location and average radius of charcoal tracheas and the necessary descriptors for final comparison with known charcoals.

To enhance CDI tracheas we use a number of algorithms (González & Wintz 1988; Jain 1989). The CDI is an image with a concentrated histogram, so we have to use low–level image algorithms (Deriche 1990). The steps are:

9.7.1 Edge detection for low–level vision

We use a recursive filter which drastically reduces the computational effort required for smoothing the Laplacian of an image. These operations are done with a fixed number of multiplication's and additions per output point independently of the size of the neighbourhood. The key to this is first of all the use of an exponentially based family of filters and then the use of recursive filtering.

The algorithm used provides a degree of freedom, "alpha", ($0<\text{alpha}<1$) through which we are able to optimise the resulting image. In order to determine the optimum parameter, a statistical study has been carried out about its effect on the observed CDI. This show that the most suitable alpha value is around 0.25.

Through the derivative rule of convolution, smoothing the input image and calculating the Laplacian of the smoothed result can be done in one step by convolving the input image with the second directional derivatives in X and Y of the smoothing filter.

9.7.2 Trachea extraction

The data provided by this process are the X and Y co–ordinates and the radius of each of the tracheas. We have opted for a process consisting of a CDI scanning which detects pairs of points of tracheas' edges. If in consecutive scans we detect that the distances separating each pair of points is maximum, we say that a trachea has been detected. Its radius will be half of the separation of these points, the Y co–ordinate of its centre will

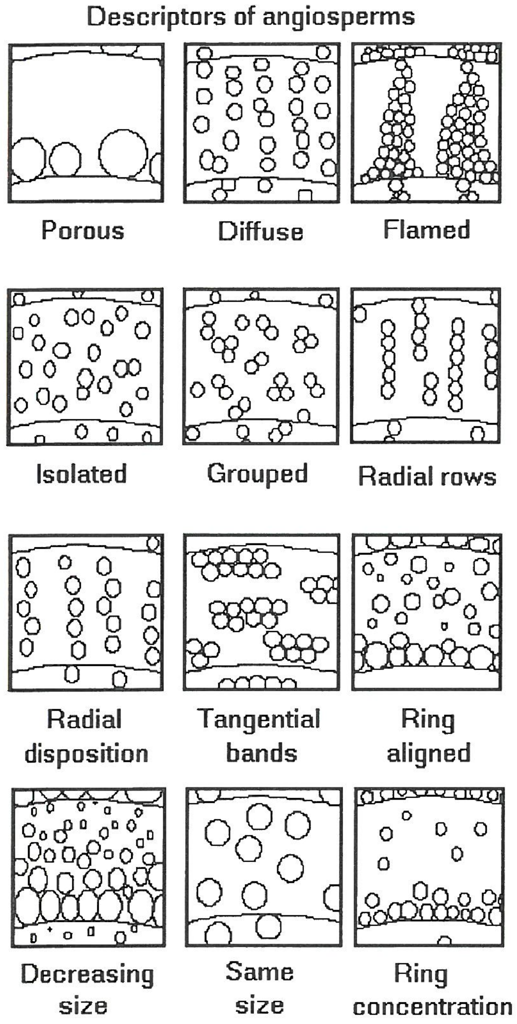

Figure 9.2: Different arrangements of angiosperms.

be the line number where the maximum separation has been detected, and the *X* co–ordinate the arithmetic average of the *X* co–ordinate of the two detected points.

The analysis is complemented by a vertical scanning so as to reduce trachea irregularity. The final result of the trachea extraction process is then the average of the co–ordinates and radii of both scans.

9.7.3 Codification

The data obtained is saved on disk. The file has a header and the co–ordinates of the tracheas. In the header we record the number of growth rings and their corresponding co–ordinates, 1 byte for the number of rings and 2 bytes for each co–ordinate of each ring. Thus we have the *X* and *Y* co–ordinates and the radius of each trachea with 5 bytes (two bytes for each of the co–ordinates and one for the radius).

After this feature extraction and enhancement process, we obtain the TCDI (Charcoal Digitalised Image with extracted Tracheas).

It results in strong data compression from a 256 Kb original image to a variable size with each TDCI oscillating around 1 Kb.

9.8 AUTOMATIC RECOGNITION AND CLASSIFICATION

This section analyses the TCDI (extracted Trachea Charcoal Digitalised Image) in order to:

- calculate the descriptors (Figure 9.2) of the transverse plane of the wood (Figure 9.1);
- create the TCDIV (TCDI Vector) of the charcoal;
- Compare the TCDIV of the charcoal under study with the descriptors of other, previously studied charcoals, that constitute the reference Vector Patterns (VP);
- assign the charcoal under study to one of the existent VPs or, if there is no corresponding VP, catalogue it as a new VP.

9.8.1 Descriptors

In order to study the images, as we have seen in section 9.4, a group of descriptors has been defined. These are based on the disposition, quantity and size of the trachea (Figure 2).

These descriptors will be treated as vectors, which, taken as a group, create an n-dimensional vector called TCDIV for each charcoal under study. To sum up, a TCDIV is a linear combination of the various descriptors of the TCDI.

9.8.2 Structure of the data and calculation of descriptors

The information provided by the process of adaptation & improvement is given by the TCDI. This file contains information on:

- number of growth rings;
- location of growth rings (approximating them to a straight line);
- location of each trachea (position *X* & *Y*);
- average radius associated with each trachea.

Each descriptor is calculated by means of an algorithm that detects it and weights it with a result between 0 and 1. This result constitutes the vector module that indicates the quantity of the TCDI's descriptor (which is greater the closer it gets to 1).

9.8.3 Selection of parameters

In order to get an optimum analysis, a pre-selection of the characteristics that will be used for comparison can be made. The parameters to be taken into consideration are the number of Vector Patterns (VP) of the Reference Data Base against which the image under study will be compared, and the Descriptors.

Vector Patterns: among the species (VP) that make up the Reference Data Base it is necessary to eliminate those that exclude the charcoal under study because of their geographical location (continent, climatic zone).

Descriptors: On the basis of the Descriptors obtained the charcoal under study, the expert anthracologist can make a selection of the descriptors that will form part of the process of comparison. In this way an orientation of the analysis toward the taxons most likely to be associated with the image under study can be achieved.

9.8.4 Comparison

On the basis of the *n* Descriptors selected, an *n*-dimensional vector is created for each element of the Reference Data Base that participated in the analysis, as well as a vector of the charcoal under study (TCDIV). By means of the scalar product, the similarity of the TCDIV vector and each of the VP can be obtained. The cosine of the angle formed by the two vectors serves as comparison value.

Once all of the comparisons have been made, the values obtained are ordered, from largest to smallest, and stored for subsequent interpretation.

9.9 INTERPRETATION OF RESULTS

The interpretation starts by taking the highest value obtained during the comparison process. If it does not exceed the minimum the expert has previously set up, we consider that, according to the chosen descriptors, none of the pattern image looks like the studied image. This means we have an element whose features are independent of the element selected from the reference database. Here we suggest that the expert add the studied sample features as a new element within the database for subsequent in depth study.

In case one or more than one exceeds the threshold, we try to guess which one belongs to the unknown sample. If there is only one element, this becomes straight forward. Otherwise we will have to check to see if there is an outstanding element within the margin of error the expert has set up (taken into account throughout the analysis). If this is the case, we will associate the VP with the image. If not, we will make use of complementary data from the tangential longitudinal section so as to achieve an assignment between a reference element and the studied image.

As a final result we must point out the features of the identified charcoal and bring the database results of the studied site up to date.

9.10 PERSPECTIVES

Using digital image processing and the automatic classification of the vegetal charcoals that derive from it offers a series of advantages when compared with traditional methods of classification.

By enhancing images and automating biometrical data we obviously accelerate and facilitate the classification process. The automation of the processing biometrical characteristics can be of great importance when establishing the determining quantitative criteria necessary to classify certain species. In the same way, one of its most positive aspects is the fact that it allows us to accelerate the first phase of the analysis, that is, the determination of taxons, helping us to interpret the results more efficiently. We have to bear in mind that the objective of anthracological analysis is not so much the listing of species rather it is the acquisition of data relating to the environment and to the maximisation of natural resources by the human community.

On the other hand, the automation of the process makes the study of a larger number of samples possible, making the results more trustworthy at two levels: the individual level, allowing us to analyse the samples of an excavation; and the general level, where the analysis of more excavations confirms the validity of regional paleoclimatic synthesis.

To sum up, using digital image processing in anthracological analysis is another example of the possibilities offered by the incorporation of new technology for the auxiliary disciplines of archaeology.

References

Carlquist, S.
1988 *Comparative wood anatomy*, Springer-Verlag, Berlin.

Deriche, R.
1990 Transactions on Pattern Analysis and Machine Intelligence. *IEEE* Vol 12, n. 1.

Gonzalez, R.C. & Wintz, P.
1988 *Digital Image Processing*, Addison–Wesley.

Jacquiot, C., Y. Trenard, & D. Dirol,
1973 *Atlas d'anatomie des bois des Angiospermes*, Centre Technique du bois, Paris, vol 2.

Jain, A.K.
1989 *Fundamentals of Digital Image Processing*, Prentice–Hall.

Lasierra, V.
1991 *Mejora y acondicionamiento de carbones vegetales*, Memoria de Fin de Carrera, EUETT La Salle Bonanova.

Pique, R. & J.M. Pique
1991 Aplicación del tratamiento de imágenes digitalizadas al análisis antracológico: un ensayo de determinación automática. in Vila, A. (ed.) *Arqueología Colección Nuevas Tendencias*, vol. 19 Consejo Superior de Investigaciones Científicas. pp. 115–129.

Schweingruber, F.H.
1978 *Mikroskopische holzanatomie* Zürcher A.G. Zug.

Yraolagoitia, J. de
1990 El formato de los ficheros DBF *PC.World*:207–224.

Authors' addresses

Raquel Pique i Huerta
Laboratori de Paleoeconomia i Paleoecologia Humana
Universitat Autònoma de Barcelona
c/Benlliure,2, 3–3
E–08290 Cerdanyola del Valles (Barcelona)

Josep Miquel Pique i Huerta
Ramon Arxer i Galabert
Departament de Processat de la Imatge
Telecomunicacions La Salle Bonanova
c/Radas,67 2–1
E–08004 Barcelona
Tel:34+3+443.06.89
Fax:34+3+430.84.64