1

Introduction

1.1 Definition and Motivation

Objects are everywhere – natural and man-made. Advances in technology have led to the routine collection of geometrical information and the study of the shape of objects is increasingly important. In particular, locating points on objects is often straightforward and this course primarily deals with the statistical shape analysis of such point set data. Shape analysis is of great interest in a wide variety of disciplines. Some specific applications follow in Section 1.2 from biology, medicine, image analysis, archaeology, geography, geology, agriculture and genetics.

The word 'shape' is very commonly used in everyday language, usually referring to the appearance of an object. Following D.G. Kendall (1977) the definition of shape that we consider is intuitive.

Definition 1.1 Shape *is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.*

So, an object's shape is invariant under the Euclidean similarity transformations of translation, scaling and rotation. For example, the shape of a human skull consists of all the geometrical properties of the skull that are unchanged when it is translated, rescaled or rotated in an arbitrary coordinate system. Two objects have the same shape if they can be translated, rescaled and rotated to each other so that they match exactly, i.e. if the objects are similar. In Figure 1 the two mouse vertebrae outlines have the same shape. In practice we are interested in comparing objects with different shapes and so we require a way of measuring shape, some notion of distance between two shapes and methods for the statistical analysis of shape.

Sometimes we are also interested in retaining scale information (size) as well as the shape of the object.

Definition 1.2 Size-and-shape *is all the geometrical information that remains when location and rotational effects are filtered out from an object.*

Two objects have the same size-and-shape if they can be translated and rotated to each other so that they match exactly, i.e. if the objects are rigid-body transformations of each other.

A common theme throughout the course is the geometrical transformation of objects. The terms superimposition, superposition, registration, transformation, pose and matching are often used equivalently for operations which involve transforming objects, either with respect to each other or into a specified reference frame.

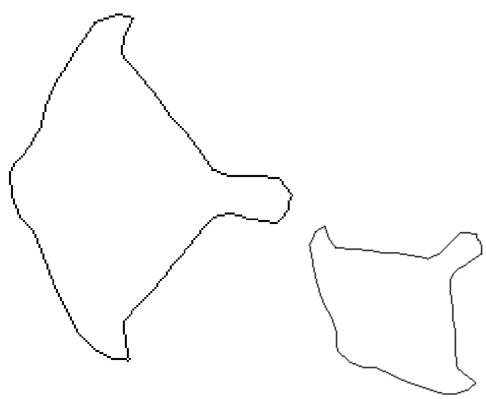


Figure 1 Two outlines of the same second thoracic (T2) vertebra of a mouse, which have different locations, rotations and scales but the same shape.

An early writing on shape was by Galileo (1638). Galileo was aware that bones in larger animals are not purely scaled up versions of those in smaller animals; there is a shape difference too. A bone has to become proportionally thicker so that it does not break under the increased weight of the heavier animal, see Figure 2. The field of geometrical shape analysis was initially developed from a biological point of view by D'Arcy Thompson (1917), who also discussed this application.

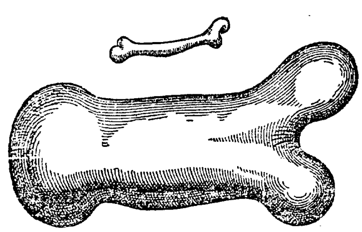


Figure 2 From Galileo (1638) illustrating the differences in shapes of the bones of small and large animals.

How should a biologist wishing to investigate a shape change proceed? Even describing an object's shape is difficult. In everyday conversation an object's shape is usually described by naming a second more familiar shape which it looks like, e.g. a map of Italy is 'boot shaped'. This leads to very subjective descriptions that are clearly unsuitable for any serious application. A practical way forward is to locate a set of points on each object.

1.1.1 Landmarks

We will describe shape by locating a finite number of points on each specimen which are called landmarks.

Definition 1.3 *A* **landmark** *is a point of correspondence on each object that matches between and within populations.*

In the literature there have been various synonyms for landmarks, including vertices, anchor points, control points, sites, profile points, ‘sampling’ points, design points, key points, facets, nodes, model points, markers, fiducial markers, markers, etc.

There are three basic types of landmarks in our applications: anatomical, mathematical and pseudo-landmarks.

An **anatomical landmark** is a point assigned by an expert that corresponds between organisms in some biologically meaningful way, e.g. the corner of an eye or the meeting of two sutures on a skull. Anatomical landmarks designate parts of an organism that correspond in terms of biological derivation and these parts are called homologous (for example, see Jardine, 1969). In Figure 3 we see some anatomical landmarks located on the skull of a macaque monkey, viewed from the side. This application is described further in Section 1.2.8.

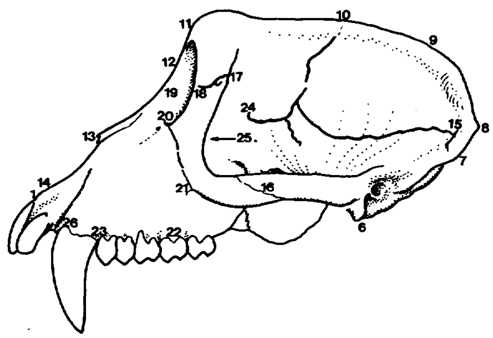


Figure 3 Anatomical landmarks located on the side view of a macaque monkey skull.

Mathematical landmarks are points located on an object according to some mathematical or geometrical property of the figure, e.g. at a point of high curvature or at an extreme point. The use of mathematical landmarks is particularly useful in automatic recognition and analysis.

Pseudo-landmarks are constructed points on an organism, located either around the outline or in between anatomical or mathematical landmarks. For example, Lohmann (1983) took equally spaced points on the outlines of micro-fossils. In Figure 4 we see six mathematical landmarks at points of high curvature and 42 pseudo-landmarks marked on the outline of a second thoracic (T2) mouse vertebra. Continuous curves can be approximated by a large number of pseudo-landmarks along the curve. Hence, continuous data can also be studied by the methods of this course, although one needs to work with discrete approximations. Examples of such approaches include the analysis of hand shapes in Grenander et al. (1991) and Mardia et al. (1991), the resistors in Cootes et al. (1994) and the mitochondrial outlines in Grenander and Miller (1994). Also, pseudo-landmarks are useful in matching surfaces, when points can be located on a regular grid over each surface.

We can also demark landmarks into three further types.

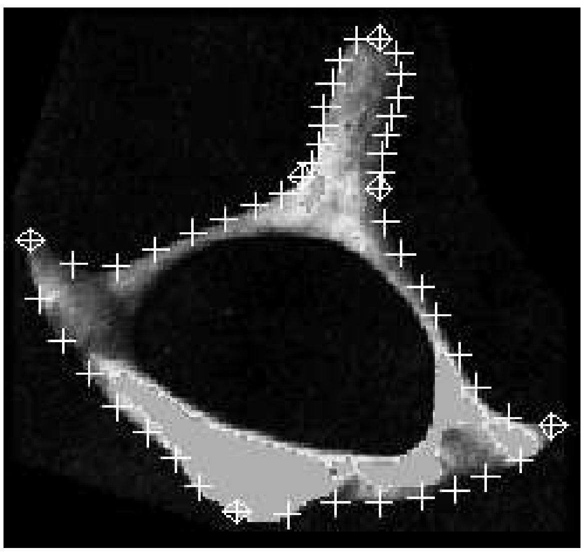


Figure 4 Grey level image of a T2 mouse vertebra with six mathematical landmarks (diamond round a +) and 42 pseudo-landmarks (+).

Definition 1.4 Type I landmarks *occur at the joins of tissues/bones;* **type II landmarks** *are defined by local properties such as maximal curvatures and* **type III landmarks** *occur at extremal points or constructed landmarks, such as maximal diameters and centroids.*

Anatomical landmarks are usually of type I or II and mathematical landmarks are usually of type II or III. Pseudo-landmarks are commonly taken as equi-spaced along outlines between pairs of landmarks of type I or II, and in this case the pseudo-landmarks are type III landmarks. Type I landmarks are usually the easiest and most reliable to locate and type III are the most difficult and least reliable to locate.

A further type of landmark is the **semi-landmark** which is a point located on a curve and allowed to slip a small distance with respect to another corresponding curve. The term ‘semi-’ is used because the landmark lies in a lower number of dimensions than other types of landmarks, e.g. along a one dimensional curve in a two dimensional image.

A further situation that may arise is the combination of landmarks and geometrical curves. For example, the pupil of the eye may be represented by a landmark at the centre surrounded by a circle, with the radius as an additional parameter. Yuille (1991) and Phillips and Smith (1993, 1994) considered such representations for analysing images of the human face.

Throughout most of this course the methodology is appropriate for landmark data or other point set data. Following Kendall (1984) our notation will be that there are k landmarks in m dimensions, where we usually have

$k \geq 3$ and $m = 2$ or $m = 3$.

Definition 1.5 *A* **label** *is a name or number associated with a landmark, and identifies which pairs of landmarks correspond when comparing two objects. Such landmarks are called* **labelled landmarks***.*

The landmark with, say, label 1 on one specimen corresponds in some meaningful way with landmark 1 on another specimen. A labelling is either naturally apparent or an objective method of relabelling could be used in certain situations. For example, in labelling the anatomical landmarks on a skull the labelling follows from the definition of the points. Alternatively, in the study of micro-fossils Lohmann (1983) obtained corresponding labels for the pseudo-landmarks on the outline by minimizing a cross-correlation function. Most of the methods considered in this course are for labelled configurations and when we refer to just ‘shape’ we implicitly mean the shape of labelled landmarks.

Unlabelled landmarks are those where there is no natural labelling correspondence between points on different specimens. In this case a shape analysis of the data is invariant under permutations of any labelling. Current research topics include the development of methods to deal with comparison of non-labelled landmarks.

Example 1.1 Consider the simple example in Figure 5. The six triangles (A, B, C, D, E and F) are constructed from triples of labelled points (1,2,3). Triangles A and B have the same size and the same labelled shape because they can be translated and rotated to be coincident. Triangle C has the same labelled shape as A and B (but has a larger size) because it can be translated, rotated and rescaled to be coincident with A and B. Triangle D has a different labelled shape but, if ignoring the labelling, it has the same unlabelled shape as A, B and C. Triangle E has a different shape to D but it can be reflected and translated to be coincident, and so D and E have the same reflection shape. Triangle F has a different shape from all the rest. □

1.1.2 Traditional methods

To perform a shape analysis, a biologist traditionally selects ratios of distances between landmarks or angles, and then submits these to a multivariate analysis (e.g. Rao, 1948). This approach has been called 'multivariate morphometrics' in biology and a review is given by

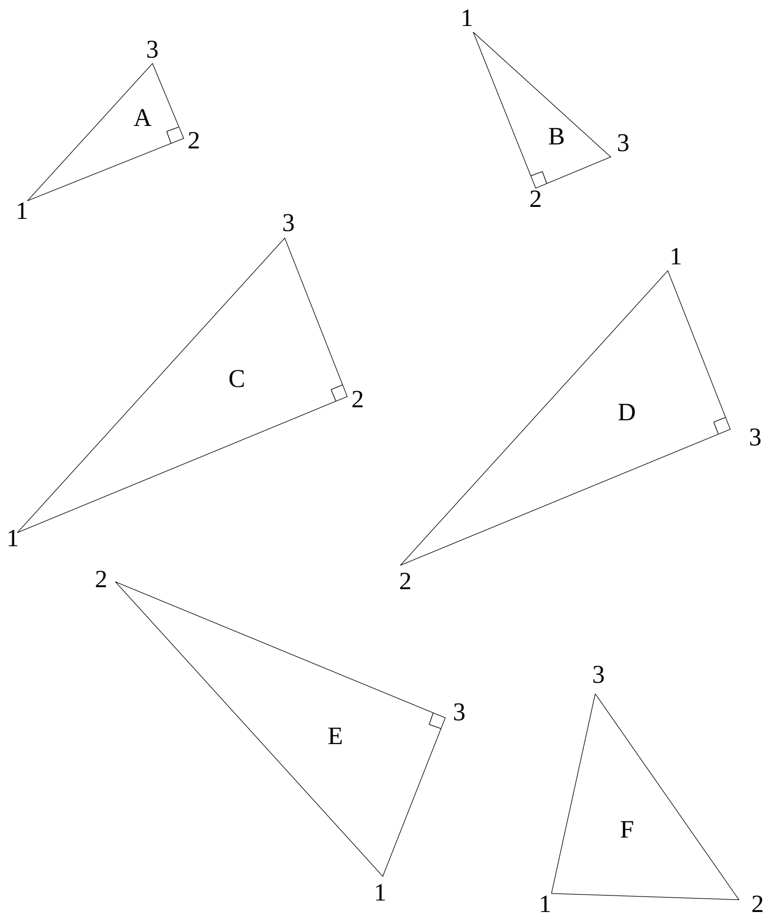


Figure 5 Six labelled triangles: A and B have the same size and labelled shape; C has the same shape as A and B (but larger size); D has a different shape but its labels can be permuted to give the same shape as A, B, C; triangle E can be reflected to have the same shape as D; triangle F has a different shape from A,B,C,D and E.

Reyment et al. (1984, pp.120).

In the studies of multivariate morphometrics one deals exclusively with positive variables (lengths, angles and ratios of lengths). However, to consider just distances and angles can be inferior to using the actual coordinates of the landmarks, because the geometry is often thrown away when using the former. Ratios of distances can easily be calculated from coordinates whereas the converse is not generally true. However, if enough distances are taken, then a configuration can be reconstructed up to a reflection.

1.1.3 Geometrical methods

In the last two decades there have been many key developments in shape analysis that allow us to work on the landmark coordinates directly. Also, the advances in technology of measuring landmarks have been helpful, e.g. landmarks from digitized objects. Of course if there were no constraints on the landmarks, then we could use standard multivariate analysis, but in general the statistical methodology for shape is inherently non-Euclidean.

The idea is that, rather than working with quantities derived from organisms, one works with the complete geometrical object itself (up to similarity transformations). The approach is very much in the spirit of D'Arcy Thompson (1917) who considered the geometric transformations of one species to another (see, for example, Figure 6). D'Arcy Thompson's key ideas will be discussed in more detail in Chapter 10, but the important point to note is that he worked with geometrical pictures of organisms rather than derived quantities. Throughout the course it will be observed that pictures of the organisms or objects under study can always be easily constructed and it is this that embodies our geometrical approach to shape analysis. In many biological applications the statistical goal is inference, for example testing for shape difference. However, the biological goal is to depict or describe the morphological changes in a study, and this is a major strength of the geometrical methods that we describe.

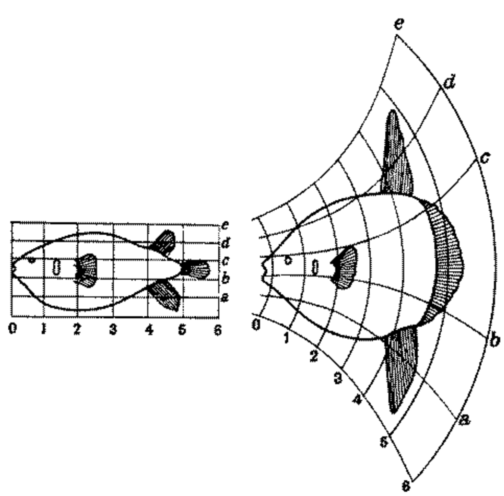


Figure 6 D'Arcy Thompson's (1917) famous example of a species of fish *Diodon* being geometrically transformed into another species *Orthagoriscus*.

In particular, we shall consider a shape space obtained directly from the landmark coordinates, which retains the geometry of a point configuration. This approach to shape analysis has been called ‘geometric shape analysis’ by various authors and the subject progressed rapidly around the late 1970s/early 1980s. Earlier work following D’Arcy Thompson (1917) included Medawar (1944) and Sneath (1967), but credit for the major developments in the last 20 years should go to D.G. Kendall and F.L. Bookstein, who independently developed many of the key ideas, in very different styles. Bookstein (1992) has summarized his view of the history of geometrical shape analysis, mainly through applications in biology. Kendall (1989) has reviewed shape theory and its development from a different, more theoretical, viewpoint, with applications in archaeology, astronomy and geography. Kendall (1989,1995) provides some historical remarks on his development of shape theory, and some of the contributions of his colleagues. The first articles on the subject were by Kendall (1977) and Bookstein (1978), and some work in the area was also given by Ziezold (1977). Some key papers in the field include Kendall (1984, 1989), Bookstein (1986), Goodall (1991), Le and Kendall (1993) and Kent (1994). Also, development in non-i.i.d. distribution theory for shape started with Mardia and Dryden (1989a).

1.2 Practical Applications

We now consider the description of several specific applications that will be used throughout the course to illustrate the methodology. In biology and medicine one wishes to study how shape changes during growth; how shape changes during evolution; how shape is related to size; how shape is affected by disease; how shape is related to other covariates such as sex, age or environmental conditions; how to discriminate and classify using shape; and how to describe shape variability. Various methodologies of multivariate analysis have been used to answer such questions over the last 60 years or so. Many of the questions in biology are the same as they have always been and many of the techniques of shape analysis are closely related to those in multivariate analysis. One of the problems is that small sample sizes are often available with a large number of variables. We shall describe some new techniques that are not part of the general multivariate toolkit, especially techniques concerned with visualization. As well as traditional biological applications many new problems can be tackled with statistical shape analysis, including automatic object recognition in image analysis, where shape analysis is useful for prior modelling of objects.

1.2.1 Biology: Mouse vertebrae

In an experiment to assess the effects of selection for body weight on the shape of mouse vertebrae, three groups of mice were obtained: Control, Large and Small. The Control group contains unselected mice, the Large group contains mice selected at each generation according to large body weight and the Small group was selected for small body weight. The bones form part of a much larger study and these bones are from replicate E of the study

(Falconer, 1973; Truslove, 1976; Johnson et al., 1985, 1988; Mardia and Dryden, 1989b).

We consider the second thoracic vertebra T2. There are 30 Control, 23 Large and 23 Small bones. The aims are to assess whether there is a difference in size and shape between the three groups and to provide descriptions of any differences. Each vertebra was placed under a microscope and digitized using a video camera to give a grey level image, see Figure 4. The outline of the bone is then extracted using standard image processing techniques (for further details see Johnson et al., 1985) to give a stream of about 300 coordinates around the outline. Six landmarks were taken from the outline using a semi-automatic procedure described by Mardia (1989a) and Dryden (1989, Chapter 5) – an approximate curvature function of the smoothed outline is derived and the mathematical landmarks are placed at points of high curvature as measured by this function. In the example of Figure 4 there are 6 landmarks and 42 pseudo-landmarks located on the outline. In a second example in Figure 7 we see again 6 landmarks but this time in between each pair of landmarks 9 equally spaced pseudo-landmarks are placed. We return to this application in Examples 2.2, 3.3, 4.1, 4.2, 5.2, 6.1, 6.4, 8.2, 10.13 and 12.4. The landmark data are given in the Appendix.

1.2.2 Biology: Gorilla skulls

In an investigation to assess the cranial differences between the sexes of apes, 29 male and 30 female adult gorilla skulls were taken. The data are described in detail by O'Higgins (1989) and O'Higgins and Dryden (1993).

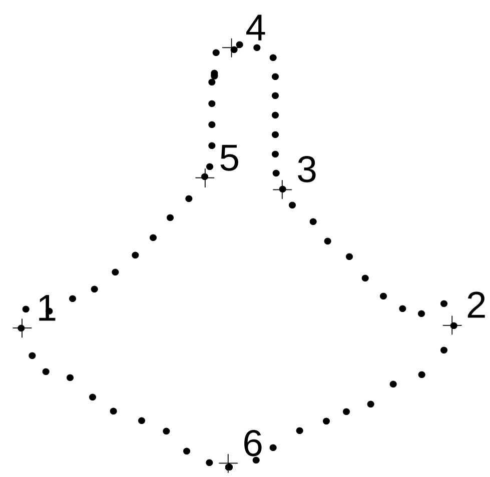


Figure 7 Six mathematical landmarks (+) on a second thoracic mouse vertebra, together with 54 pseudo-landmarks around the outline, approximately equally spaced between pairs of landmarks. The landmarks are 1 and 2 at maximum points of approximate curvature function (usually at the widest part of the vertebra rather than on the tips), 3 and 5 at the extreme points of negative curvature at the base of the spinous process, 4 at the tip of the spinous process, 6 at the maximal curvature point on the opposite side of the bone from 4.

Eight landmarks are chosen in the midline plane of each skull as shown in Figure 8. The landmarks are anatomical landmarks and are located by an expert biologist.

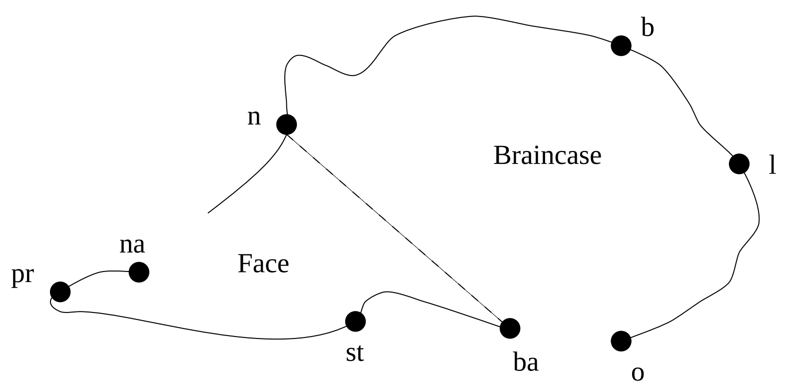


Figure 8 Eight landmarks on the midline section of the ape cranium. The face region is taken to be comprised of landmarks 7: *nasion* (n), 4: *basion* (ba), 5: *staphylion* (st), 1: *prosthion* (pr) and 6: *nariale* (na). The braincase region is taken to be comprised of landmarks 7, 4 and 8: *bregma* (b), 2: *lambda* (l) and 3: *opisthion* (o).

It is of interest to assess whether there is a size difference between the sexes and whether there are any shape differences between the sexes in the face and braincase regions. A biologist would also be interested in geometrical descriptions of the shape difference, and how shape relates to size and other covariates. We consider this application in Examples 3.2, 7.2, 7.5, 10.6, 10.7, 10.8 and 10.9. The landmark data are given in the Appendix.

1.2.3 Medicine: Brain MR scans of schizophrenic patients

Bookstein (1996b) considers 13 landmarks taken on near midsagittal two dimensional slices from Magnetic Resonance (MR) brain scans of 14 schizophrenic patients and 14 normal patients. It is of interest to study any shape differences in the brain between the two groups, either in average shape or in shape variability. If morphometric differences between the two groups can be established, then this should enable researchers to gain an increased understanding about the condition. In Figure 9 we see the 13 landmarks on a two dimensional slice from a scan of a schizophrenic patient. We return to this application in Examples 6.3, 7.4 and 10.2.

1.2.4 Image analysis: Postcode recognition

A random sample of handwritten British postcodes has been collected and digitized. An example digit ‘3’ is shown in Figure 10, taken from the data described by Anderson (1997). It is of interest to automatically classify each of the handwritten characters so that mail can be automatically sorted. The problem is a classic one in image analysis and many methods have been suggested, with varying degrees of success; e.g. see Hull (1990). The location and size of the characters are not so important for recognition but orientation information may be crucial, e.g. an ‘M’ must not be confused with a ‘W’. Some recent successful attempts at reading handwritten numbers include Hastie

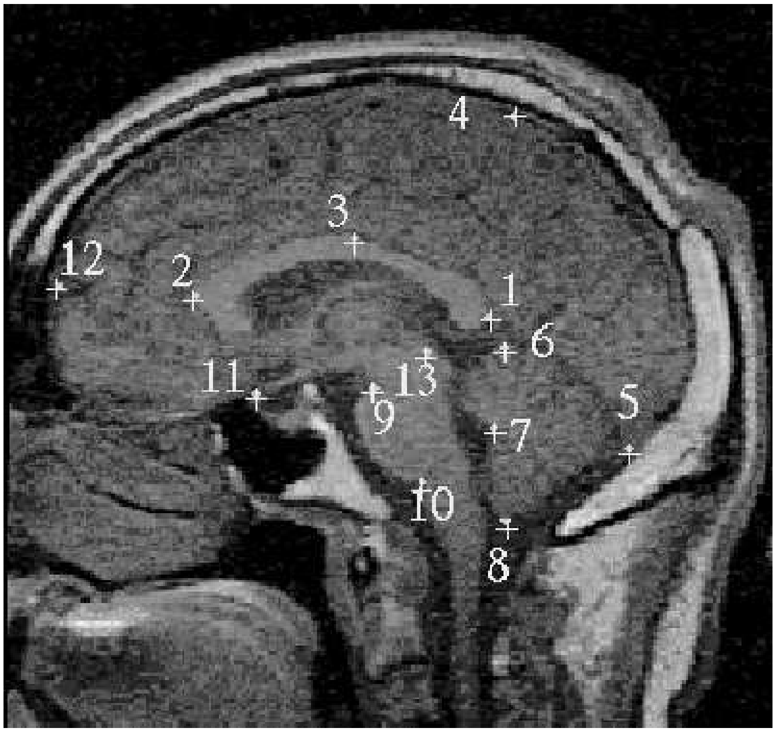


Figure 9 The 13 landmarks on a near midsagittal brain scan of a schizophrenic patient (after Bookstein, 1996b). The landmark positions are approximately located at each cross (+): 1: splenium, posteriormost point on corpus callosum, 2: genu, anteriormost point on corpus callosum, 3: top of corpus callosum, uppermost point on arch of callosum (all three landmarks registered to the diameter of the callosum), 4: top of head, a point relaxed from a standard landmark along the apparent margin of the dura, 5: tentorium of cerebellum at dura, 6: top of cerebellum, 7: tip of fourth ventricle, 8: bottom of cerebellum, 9: top of pons, anterior margin, 10: bottom of pons, anterior margin, 11: optic chiasm, 12: frontal pole, extension of a line from 1 through 2 until it intersects the dura, 13: superior colliculus.

and Tibshirani (1992) and Simard et al. (1993). A related topic is hand-drawn gesture recognition (see, for example, Mardia et al., 1993).

Anderson (1997) obtained mathematical landmarks and pseudo-landmarks on the digital images by hand, and in particular for the digit 3 there were 13 landmarks, as shown in Figure 10. It is of interest to examine the average shape and variability in shape in the data, which can then be used as a prior model for digit recognition from images of handwritten postcodes. We return to this application in Examples 5.3 and 7.1. The landmark data are given in the Appendix.

1.2.5 Archaeology: Alignments of standing stones

Consider a map of the 52 megalithic sites that form the ‘Old Stones of Land’s End’ in Cornwall, England, given

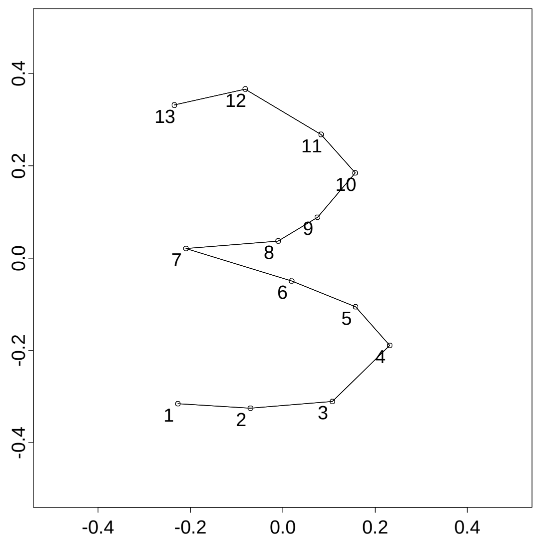


Figure 10 A handwritten digit ‘3’ from the postcode dataset, with 13 labelled mathematical landmarks. Landmark 1 is at the extreme bottom left, 4 is at the maximum curvature of the bottom arc, 7 is at the extreme end of the central protrusion, 10 is at the maximum curvature of the top arc and 13 is the extreme top left point. Landmarks 2, 3, 5, 6, 8, 9, 11 and 12 are pseudo-landmarks at approximately equal intervals between the mathematical landmarks.

in Figure 11. It was proposed by Alfred Watkins, in the early 1920s, that these and other megalithic sites were placed in deliberate straight lines, called ley lines. One approach is to consider the shapes of all possible triangles and to see if there are more ‘flat’ triangles (triangles with the largest angle close to 180 degrees) than expected under a randomness hypothesis. A major difference here compared with previous examples is that the points are unlabelled, and in this dataset there are $\binom{52}{3}$ triangles in two dimensions.

This dataset is particularly important in the history of shape analysis because it motivated D.G. Kendall’s pioneering work. Analysis of these data is considered by Broadbent (1980), Kendall and Kendall (1980), Small (1988) and Stoyan et al. (1995) among others.

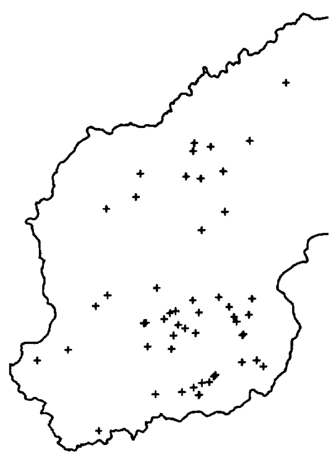


Figure 11 The map of 52 megalithic sites (+) that form the ‘Old Stones of Land’s End’ in Cornwall (from Stoyan et al., 1995).

1.2.6 Geography: Central Place Theory

Central Place Theory was postulated by Christaller (1933) and is the situation where towns are distributed on a regular hexagonal lattice over a homogeneous area (with towns at hexagon centres, see Figure 12). Mardia et al. (1977) consider this hypothesis for a map of 44 places in 6 counties in Iowa, namely Union, Ringgold, Clarke, Decatur, Lucas and Wayne Counties.

In order to examine whether Central Place Theory holds, one could examine the shapes of the triangles formed by a town and its neighbours to see if they are more equilateral than expected under a randomness hypothesis. A convenient triangulation of the towns is a Delaunay triangulation (Mardia et al., 1977; Green and Sibson, 1978). In Figure 12 we see that Voronoi polygons for ideal central places would be hexagons and Delaunay triangles would be equilateral triangles (see Okabe et al., 1992, for a detailed description of such tessellations). In Figure 13 we see a Delaunay triangulation and the Voronoi polygons for the Iowa data. An important question to ask is: are the Delaunay triangles more equilateral than expected by chance?

The points here are unlabelled (there is no correspondence in the vertices of the triangles). Also the triangles are correlated due to neighbouring triangles sharing points. The work of Mardia et al. (1977) led Kendall (1983, 1989) to further study shape in Delaunay triangulations, in order to investigate the Central Place Theory hypothesis. We return to this application in Example 12.7.

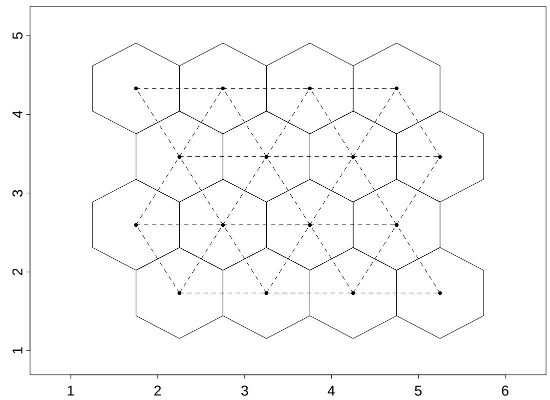


Figure 12 The Voronoi polygons (unbroken lines) and Delaunay triangulation (broken lines) for a completely regular configuration, i.e. ideal central places.

1.2.7 Geology: Microfossils

The microfossil *Globorotalia truncatulinoides* is a microscopic planktonic found in the ooze on the ocean bed. Lohmann (1983) published 21 mean outlines of the microfossil which were based on random samples of organisms taken at different latitudes in the South Indian Ocean. Figure 14 shows the three mathematical landmarks selected

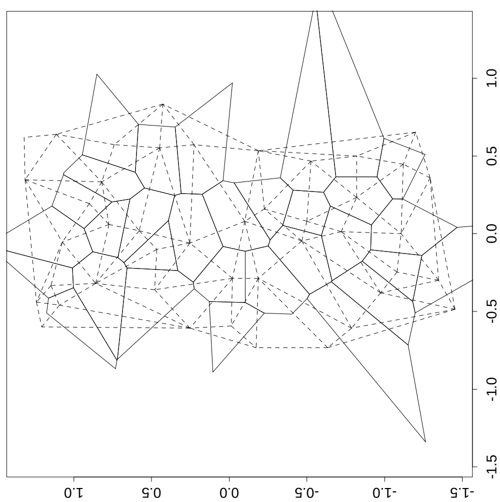


Figure 13 The Voronoi polygons (unbroken lines) and Delaunay triangles (broken lines) for the Iowa towns. The Voronoi polygons at the edges are not shown fully.

on the outline. The coordinates of the 21 landmarks are extracted from Figure 7 of Bookstein (1986). It is of interest to examine whether the size of the organisms is related to the shape, and whether size or shape are related to the covariate of latitude. A more basic problem would be to obtain an estimate of the average shape of the fossils and to describe the structure of the shape variability. We return to this application in Examples 6.2 and 8.1.

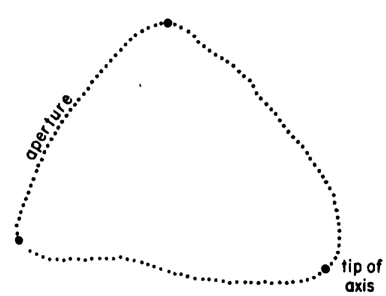
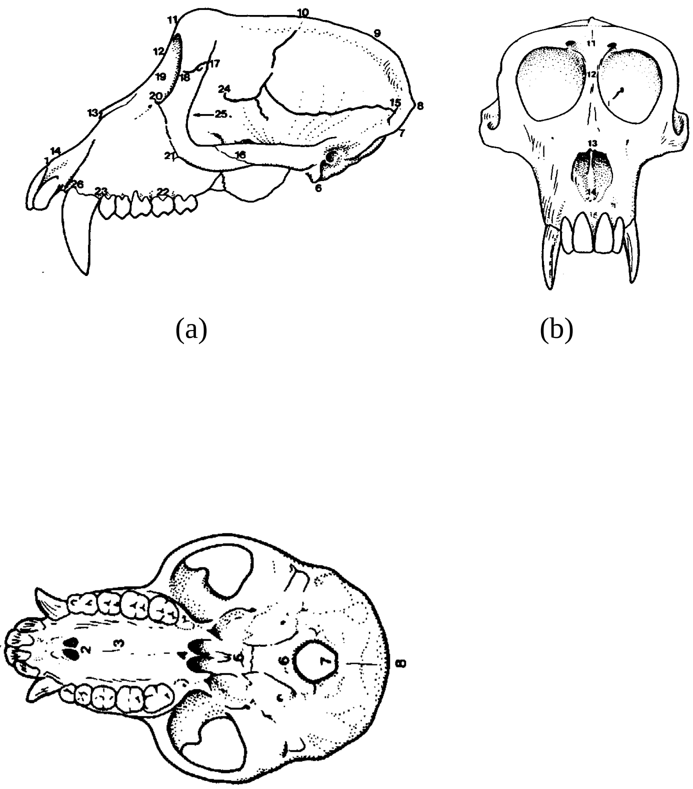


Figure 14 The outline of a microfossil with three landmarks (from Bookstein, 1986).

1.2.8 Biology: Macaque skulls

In an investigation into sex differences in the crania of a species of macaque *Macaca fascicularis* (a type of monkey) random samples of 9 male and 9 female skulls were obtained (Dryden and Mardia, 1993). A subset of seven anatomical landmarks was located on each cranium and the three dimensional coordinates of each point were recorded.

It is of interest to assess whether there are any size and shape differences between the sexes. If there are any differences, then a description of the differences is required. An artist's impression of the three dimensional skull with the anatomical landmarks is given in Figure 15. We return to this application in Examples 4.3, 5.1, 5.4 and 7.3.



(a) (b) (c)

Figure 15 A three dimensional macaque skull: (a) side view, (b) frontal view, (c) bottom view. A total of 26 landmarks are displayed on the skull and a subset of 7 was taken for the analysis. The seven chosen landmarks are 1. *prosthion*, 7. *opisthion*, 10. *bregma*, 12. *nasion*, 15. *asterion*, 16. *midpoint of zyg/temp suture*, 17. *interfrontomalare*.

1.2.9 Biology: Sooty mangabeys

Twelve landmarks are taken from the midline of the skulls of another type of monkey, sooty mangabey (*Cercocebus atys*), in a study described by O'Higgins and Dryden (1992), see Figure 16. The specimens ranged from young juveniles to an adult female and an adult male. The objective is to describe the size and shape differences in the individuals in the series from the young juveniles to the older juveniles, and then to the adults. A further problem is to examine whether the individuals can be modelled by a regression line in shape space. We return to this application in Examples 3.1 and 10.4.

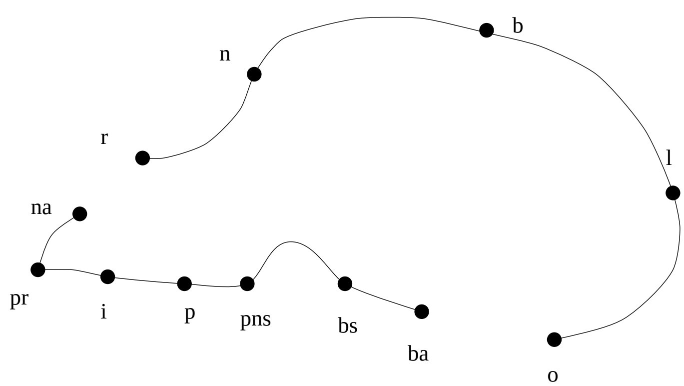


Figure 16 The 12 landmarks on the midline of the skull of a juvenile sooty mangabey. The chosen landmarks are *nasion* (n), *rhinion* (r), *nariale* (na), *prosthion* (pr), *incisive canal* (i), *palatine junction* (p), *posterior nasal spine* (pns), *basisphenoid* (bs), *basion* (ba), *opisthion* (o), *lambda* (l), *bregma* (b).

1.2.10 Agriculture: Fish recognition

Mardia et al. (1996f) consider the automatic location of fish in images. An underwater camera captures images of the fish and the aim is to identify the fish in the images and to automatically obtain information about the size and shape distribution of the fish. An example image is given in Figure 17.

Figure 17 An underwater image of fish in a tank.

By using a geometrical template, prior information about the average shape and shape variability of fish can be modelled to assist with recognition. We return to this application in Example 11.2.

1.2.11 Agriculture: Robotic harvesting of mushrooms

Mardia et al. (1996g) consider the problem of automatically recognizing mushrooms in an image such as that in Figure 18. The ultimate aim is to design a robot to harvest the mushrooms and an essential component will be the successful recognition of the mushrooms in the image, using prior size and shape information. We consider this application in Examples 11.3 and 11.5.

1.2.12 Genetics: Electrophoretic gels

A technique for the identification of proteins involves the comparison of electrophoretic gel images. Two examples

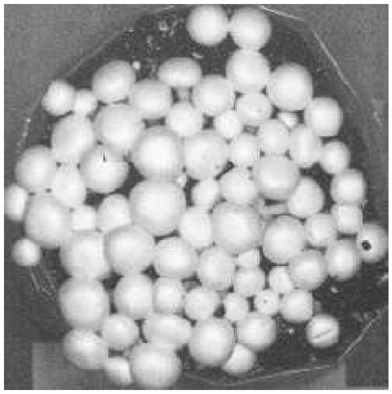
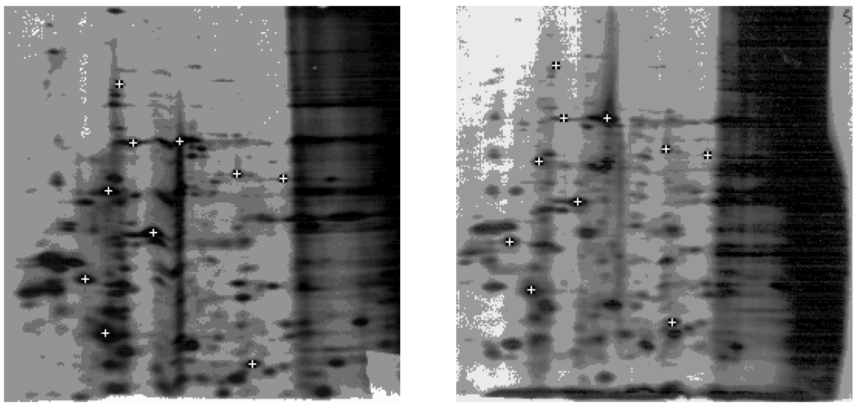


Figure 18 An image of mushrooms ready for harvesting.

of such images are gel A and gel B shown in Figure 19. The images were obtained from particular strains of parasites which carry malaria. The objective is to use the gel image to be able to identify the strain of parasite.

In each gel there are a number of black spots, where each spot can be one of two types – invariant or variant. The invariant spots are present for all parasites and the arrangement of variant spots enables identification of the parasite. A problem with the technique when used in the



(a) (b)

Figure 19 The electrophoretic gel images from (a) gel A and (b) gel B (after Horgan et al., 1992). The invariant spots are marked with a '+' in both images.

field is that the gels are prone to deformations and so the gel images first need to be 'registered' (transformed) so that direct comparisons can be made.

In this application, 10 invariant spots have been picked out by an expert, as shown in Figure 19. The invariant spots are used to match gel A to gel B, either by a similarity transformation or by a more complicated transformation. A question of interest is: can a matching procedure be made resistant to some outlier points, e.g. mislabelled points? We return to this application in Examples 12.2 and 12.3.